Real-time Analysis and Control of Plasma Etching via Full Wafer Interferometry

by

Ka Shun Wong

B.S., Rensselaer Polytechnic Institute, 1988
S.M., Massachusetts Institute of Technology, 1990

Submitted to the Department of Mechanical Engineering
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1996

© Massachusetts Institute of Technology, 1996. All Rights Reserved.

Author ....
Department of Mechanical Engineering
August 16, 1996

Certified by ....
Duane S. Boning
Assistant Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Certified by ....
Stephanie W. Butler
Manager, Advanced Process Control and Metrology Branch, Texas Instruments, Inc.
Thesis Supervisor

Accepted by ....
Ain A. Sonin
Chairman, Department Committee on Graduate Students

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

DEC 0 3 1996

ARCHIVES

LIBRARIES

Blank Page

Real-time Analysis and Control of Plasma Etching via Full Wafer Interferometry

by

Ka Shun Wong

Submitted to the Department of Mechanical Engineering
on August 16, 1996 in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Mechanical Engineering

Abstract

As device dimensions of microelectronic circuits continue to shrink, process control becomes even more important to ensure high yields. Plasma etching is a critical step in the microelectronic fabrication process, and key concerns are spatial uniformity of etch and reduction in over-etching. Recently, a full wafer interferometry system has been developed which enables simultaneous monitoring of many sites on the wafer during an etch in real-time. The sensor relies on the plasma emission to provide the interferometric signals. Since the average intensity of the light emission varies from one process run to another, there is a need to develop specific algorithms to utilize these signals for process control. This work is aimed at extracting key information from the signals, in particular, etch rate, endpoint detection (clearing of etching film), and endpoint prediction (or film thickness estimation), which is essential for implementing real-time process control.

Etch rate is found to be readily obtained from the periodicity of the signals. Four algorithms, namely the Fourier Transform, model matching, Kumaresan-Prony, and multiple signal classification methods, are examined for the accuracy of their estimates under various conditions. It is found that the Fourier Transform can provide an accurate result despite its simplicity and computational efficiency.

Two endpoint detection algorithms have been examined. The first algorithm relies on modeling matching and is applicable for etching non-absorptive films. The second algorithm is applicable for films with high refractive indices, and takes advantage of reduction in signal variations at endpoint. Both algorithms are found to be effective experimentally.

A new model associating the phase of interferometric signals to film thickness is proposed. The accuracy of the remaining film thickness estimates is examined with respect to drifting etch rates, variation in film stack structure, and choice of signal wavelengths. Simulation results show that short wavelength signals are more robust. Experimental results also verify the robustness of the algorithm for monitoring spatial uniformity of an etch in real-time. The accuracy of film thickness estimation is about ± 100Å for etching polysilicon.

Thesis Supervisor: Duane S. Boning
Title: Assistant Professor of Electrical Engineering and Computer Science

Thesis Co-supervisor: Stephanie W. Butler
Title: Manager, Advanced Process Control and Metrology Branch, Texas Instruments, Inc.

Acknowledgments

First of all, I have to thank Professor Duane Boning for his encouragement and insight. His guidance throughout my study, especially when I got lost in my direction, is essential for the success of this work. He has inspired a lot of interest in my project. In addition, I am tremendously appreciative that he is always patient with my last-minute requests.

I would like to thank Dr. Stephanie Butler for being my co-supervisor, and for taking the time to come to MIT so many times. Her enthusiasm has always made my project an exciting one, and her insight has been very valuable for the completion of this work.

I would like to thank Professor Emanuel Sachs for being the committee chairman. Eight years ago, he first introduced me to the field of process control. I have learned a lot from him all these years, and I will always appreciate the encouragement he has given me.

I would like to thank Professor Herbert Sawin for serving on my committee and making much of this research possible. His advice and insight have added to the value of this research. I also want to thank him for the access to the etching and the sensor equipments.

Many early experiments were performed by Timothy Dalton, and I want to thank him for his time.

I received a lot of help and suggestions from Sungdo Ha and Jonathan Claman at the beginning of the project, and I want to thank them.

Many thanks to all the members of the staff at the Microsystems Technology Laboratories. Without them, it would have taken me much longer to finish the experiments. In particular, please pardon my lengthy questions, Joe Walsh, Brian Foley, Pat Burkhart, Paul Tierney, Tim Tyson, and Bernard Alamariu.

I would also like to thank Bill Connor and Igor Tepermeister of Low Entropy Systems for their expertise on the full wafer interferometry system.

I would like to thank my officemates whom I worked with (played with) nearly everyday. Thanks, Aaron, for the help in all the experiments, and for sharing the worries for the etcher. Thanks, Taber, for all the jokes and for fixing my car's brakes. Thanks, Chantal, for livening up our office. Thanks, William, Minh and David, for making the office life in MIT more enjoyable. Also, I want to thank Scotti for her helpfulness in the office.

To my past housemates and pseudo-housemate, Alex, William, and Mike. Your presence added so much fun to my life at Boston.

Thanks, Sarah and Lam, for your friendship, and all your prayers for Amy and I.

Especially, I would like to thank my parents. They are always behind me, and they have provided an environment for me to pursue my own goal. I also want to thank my

sisters and brother for their support and their influence on my life.

With all my love, I want to thank my wife, Amy, for being with me, for putting up with all my work, and for believing in me. Finally, this work is over.

I would like to thank the financial support from ARPA under contract N00174-93-C0035 and from SEMATECH under contract 94-MC-503.

Contents

Abstract **3**

Acknowledgments **4**

List of Figures **9**

List of Tables **13**

1 Introduction **14**

1.1 Motivation 14

1.2 Plasma etching 15

1.3 Previous work 16

1.3.1 VLSI fabrication process control 16

1.3.2 Sensor and process control for plasma etching 18

1.4 Sensor and control 20

1.5 Thesis outline 20

2 Full wafer interferometry **23**

2.1 Optical emission interferometry 23

2.2 System setup of the current full wafer interferometer 26

2.3 Applied Materials Precision 5000 etcher 28

2.4 Summary 28

3 In-situ etch rate extraction from interferometric signals **29**

3.1 Introduction 29

3.2 Etch rate estimation algorithms 30

3.2.1 Fourier Transform Method 31

3.2.2 Model Matching Method 32

3.2.3 Kumaresan-Prony Method 33

3.2.4 Multiple Signal Classification Method 35

3.3 Simulation results 36

3.4 Experimental results 40

3.5 Effect of etch rate estimation accuracy 41
3.5.1 Neural network etch model 42
3.5.2 1-mid-course correction control 44
3.5.3 2-mid-course correction control 48
3.6 Analysis of etch rate estimation for mid-course correction control 52
3.7 Summary 54

4 Endpoint detection of plasma etching 55

4.1 Introduction 55
4.2 Model deviation endpoint algorithm 55
4.2.1 New model for interferometric signal 56
4.2.2 Endpoint detection algorithm 61
4.3 Slope detection endpoint algorithm 65
4.4 Summary 70

5 Endpoint prediction via phase angle of interferometric signals 71

5.1 Introduction 71
5.2 Film thickness estimation using interferometric phase information 71
5.3 Sensitivity analysis of endpoint phase angle with respect to etch rate drifting 76
5.3.1 Drifting rate effect and window length effect 77
5.3.2 Signal wavelength selection 81
5.3.3 Signal absorption rate on least square estimation of phase 85
5.4 Sensitivity analysis of endpoint phase angle with respect to film structure variations 87
5.5 Practical issues 90
5.6 Experiments on film thickness prediction 92
5.6.1 Endpoint prediction versus timed etching 95
5.7 Experiments on endpoint prediction 96

5.7.1 Endpoint prediction with phase
(without emission level normalization) 99
5.7.2 Endpoint prediction with phase
(with emission level normalization) 109
5.7.3 Etch chemistry switching 112
5.7.4 Film thickness estimation and time-to-endpoint 115
5.7.5 Modified algorithm for noisy data 118
5.7.6 Etch rate drifting 123
5.8 Full wafer endpoint prediction 125
5.9 Endpoint prediction versus endpoint detection 126
5.10 Summary 128

6 Conclusions and future work 129

6.1 Conclusions 129
6.2 Future work 130

Bibliography 132

Appendices

A Etch rate data used for neural network model building 138
B Neural network etch rate model 139
C Sensitivity analysis of film thickness estimation error at film clearing
due to film structure variation 140
D A Matlab script file for endpoint prediction 144

List of Figures

1-1 The etch process 15

1-2 Some undesired etch results 16

2-1 Full wafer interferometry for a 2-film structure 23

2-2 Examples of different interferometric reflectance 25

2-3 Simplified system setup for full wafer interferometry 26

2-4 CCD image of an etching wafer 27

3-1 Examples of low sampling rates of interferometric signals 30

3-2 Reflectivity of ideal interferometric signal and simulated noisy readings 36

3-3 Example of moving window etch rate estimates for different algorithms 38

3-4 Etch rate estimation from an etching with changing control settings 41

3-5 Pattern of wafer selected for etch model building 43

3-6 Structure of an artificial neural network for modeling etch rates 43

3-7 Radial etch rate non-uniformity from experiments 45

3-8 Flow chart for 1-mid-course correction control 46

3-9 RN of total etch for 1-mid-course control w/ non-optimal start
as a function of etch rate estimation error 47

3-10 RN of total etch for 1-mid-course control 2/ optimal start
as a function of etch rate estimation error 48

3-11 Goal of RN and TN for 2-mid-course control w/ compensation 49

3-12 An example of film surface under 2-mid-course control 50

3-13 Flow chart for 2-mid-course correction control etch simulation 50

3-14 RN of total etch for 2-mid-course control w/ non-optimal start
as a function of etch rate estimation error 51

3-15 RN of total etch for 2-mid-course control w/ optimal start
as a function of etch rate estimation error 52

4-1 7534Å signals for etching 5000Å of bare polysilicon on oxide 56

4-2 Spectrum of non-absorptive interferometric signal 58

4-3 Different model fitting of interferometric reflectance for etching 5000Å
of polysilicon on 1000Å of oxide 59

4-4 Endpoint for etching 5000Å patterned poly on 1100Å oxide (Model Dev.) 62
4-5 Normalized 7534Å signal and endpoint for etching patterned poly on 450Å oxide 64
4-6 Endpoint for etch of 5000Å patterned poly on 250Å oxide (Model Dev.) 64
4-7 Idea interferometric reflectance for etching 5000Å poly on 1000Å oxide 65
4-8 7534Å signal and endpoint for patterned area of poly (Slope Detection) 67
4-9 7534Å signals for etching 5000Å patterned poly on 450Å oxide (Slope Detection) 68
4-10 7534Å signals for etching 5000Å patterned poly on 250Å oxide (Slope Detection) 68
4-11 4780Å signals and endpoint for an unpatterned site when etching 5000Å patterned polysilicon on 1100Å oxide (Slope Detection) 69
5-1 Interferometric reflectance and the corresponding phase angles 72
5-2 Phase estimation from interferometric signal via different methods 73
5-3 Reflectance of etching poly over 1000Å of oxide and the sinusoidal approximation 74
5-4 Experimental data for film phase estimation 75
5-5 Fitting error of the unwrapped phase 76
5-6 Example of phase error for a drifting sinusoid 77
5-7 Qualitative analysis of fitted phase 78
5-8 4800Å reflectance for etching 5000Å poly on 1000Å oxide (etch rate increasing) 78
5-9 Normalized phase estimation error with different numbers of data cycles 79
5-10 4800Å signal for etching polysilicon over oxide (etch accelerates at different rates) 80
5-11 Phase estimation error due to different etch acceleration 81
5-12 Reflectance for etching 5000Å poly on 1000Å oxide at different wavelengths 82
5-13 Phase estimation error using model matching method 83
5-14 Phase estimation error using FFT method 84
5-15 Qualitative analysis of fitted phase for absorptive and non-absorptive

signals ... 86
5-16 Phase estimation error for data window beginning (model matching) ... 86
5-17 Interferometric reflectance for etching 5000Å of polysilicon on different oxide thicknesses over silicon substrate ... 88
5-18 Selection of initial phase using allowable thickness bounds ... 91
5-19 Design parameters for film thickness experiments ... 92
5-20 Pixel pattern selected for five-pixel averaging ... 93
5-21 CCD signals for the calibrating experiments ... 94
5-22 Experimental data and fitted model for polysilicon thickness estimation ... 94
5-23 CCD image of an etching wafer ... 99
5-24 Selected endpoint analysis dice ... 100
5-25 Endpoint prediction for bare poly areas I ... 100
5-26 Endpoint prediction for bare poly areas II ... 101
5-27 Endpoint prediction for via areas I ... 102
5-28 Endpoint prediction for via areas II ... 103
5-29 Endpoint prediction for line areas I ... 103
5-30 Endpoint prediction for line areas II ... 104
5-31 Endpoint prediction for serpentine areas I ... 104
5-32 Endpoint prediction for serpentine areas II ... 105
5-33 Endpoint prediction for diffraction gratings I ... 105
5-34 Endpoint prediction for diffraction gratings II ... 106
5-35 Endpoint prediction for noisy data I ... 106
5-36 Endpoint prediction for noisy data II ... 107
5-37 Endpoint prediction for noisy data III ... 107
5-38 Endpoint prediction for noisy data IV ... 108
5-39 Endpoint prediction for noisy data V ... 108
5-40 Endpoint time for line regions from Exp. 10 ... 109
5-41 Comparison of original and normalized data ... 110
5-42 Endpoint prediction for Exp. 13 using normalized data ... 111
5-43 Predicted etch switching time for bare poly areas I ... 113
5-44 Predicted etch switching time for bare poly areas II ... 113

5-45 Predicted etch switching time for normalized signals 114
5-46 Film thickness and time-to-endpoint estimates for line area in Site #5 116
5-47 Etch rate and time-to-endpoint estimates for a drifting etch 116
5-48 Etch rate and time-to-endpoint estimates for a "more" steady etch 117
5-49 Error in time-to-endpoint estimates for a "more" steady etch 118
5-50 Example for incorrect phase estimation using 2-cycle window 118
5-51 Example for correct phase estimation using all-data window 119
5-52 Incorrect initial cycle counting due to strong photoresist interference 122
5-53 Incorrect phase estimation due to photoresist interference 122
5-54 Ratio of the second half etch to initial half etch for different areas 124
5-55 Full wafer film thickness estimation 125
5-56 Endpoint detection of etching 5000Å poly over 1000Å oxide (Model dev.) 126
5-57 Reflectance of etching 5000Å polysilicon over 250Å oxide 127
5-58 Predicted endpoint for etching 5000Å polysilicon over 250Å oxide 127
5-59 Normalized signal and predicted endpoint for etching 1000Å polysilicon over 250Å oxide 128

List of Tables

3-1 Normalized mean and (std. dev.) of etch rate estimates 39

4-1 Statistics from the model fitting 60

5-1 Phase estimation error for interferometric signals 74

5-2 Phase estimation error for different signal wavelengths (model matching) 83

5-3 Phase estimation error for different signal wavelengths (FFT) 85

5-4 Thickness estimation error at beginning of data window for different wavelengths 87

5-5 Error of film thickness estimation from designed experiments 88

5-6 Parameter levels for the design of experiments 89

5-7 Maximum errors in endpoint prediction with respect to film structure variations (Å) 89

5-8 Fitting and estimation errors in film thickness using phase of interferometric signals 95

5-9 Design levels for the four control parameters 97

5-10 Design array for 4-parameter Box-Behnken design 98

5-11 Selected equipment settings from 4-parameter Box-Behnken design 101

5-12 Number of endpoint prediction failure using all-data window 120

5-13 Number of endpoint prediction failure using 2-cycle-signal sliding window 121

5-14 Percent failure of endpoint prediction using all-data window method 121

5-15 Statistics of etch rate ratio 124

A-1 Selected etch rate data from AME-5000 138

C-1 Film structures for sensitivity analysis 140

C-2 4-parameter 2-level full factorial design array 141

C-3 Film thickness estimation error due to film structure variation 142

Chapter 1

Introduction

1.3 Motivation

Plasma etching is an important part of microelectronic fabrication processes. In plasma etching, one of the important objectives is to etch away the film while minimizing undercut or other effects of overetch, which is typically used to ensure complete etch in the face of topography differences and wafer nonuniformity. In general, a uniform etch will result in uniform clearing of the film, so that overetch can be reduced. In the past, effort has focused on designing etch chambers that could provide a uniform plasma density over the wafer; one result of uniformity requirements has been the shift from batch fabrication to single wafer processing where better control is possible. Off-line design of experiments have been used to optimize the productivity and robustness of the process, and statistical process control charts are used for rapid detection of process deviations [49]. As competition for high quality and variety increases, the design and implementation of new and flexible fabrication processes becomes an important issue.

In recent years, more emphasis has been placed on controlling the process on-line. Contrary to robust design, on-line control requires that the control parameters can modify the process outputs effectively. Various run-by-run control methodologies have been developed, such as an exponentially weighted moving average control or predictor corrector control, which adjust equipment settings to compensate for process drifts and/or process shifts [13,62,67]. In addition, real-time control has been investigated for providing consistent equipment states, such as the control of the RF power [77]. However, most of these efforts have not been focused on the monitoring and control of wafer states measured in real-time. Recently, a full wafer interferometric system has been developed [16,17]. It facilitates simultaneous monitoring of etch performance at many sites over the wafer surface. However, since the sensor utilizes the plasma emission as the source of signals, a number of challenging issues need to be tackled before the sensor can be fully used for real-time process control. In this thesis, topics regarding etch rate extraction, endpoint detection, and endpoint prediction (or film thickness estimation) will be examined, with the goal of obtain-

ing accurate feedback for real-time control of the etch process.

1.2 Plasma etching

Plasma etch is an important step for patterning films on a wafer [74]. Before an etch, a layer of patterned photoresist is first coated on the film surface to protect specific areas from being removed. During the etch, a glow discharge is generated by an applied electric field, which produces highly energetic electrons, reactive ions, and reactive atoms from the etching chemicals. Combining the effects of chemical reaction of the reactive species, and the physical bombardment of the ions, various degrees of anisotropic etch can be produced. In addition, emission at various wavelengths can be observed from the glow discharge when orbital electrons or atoms return from excited states to lower energy levels. These emissions are unique with respect to the etch, and can be used for process diagnostics. Figure 1-1 shows the etch process.

Figure 1-1: The etch process

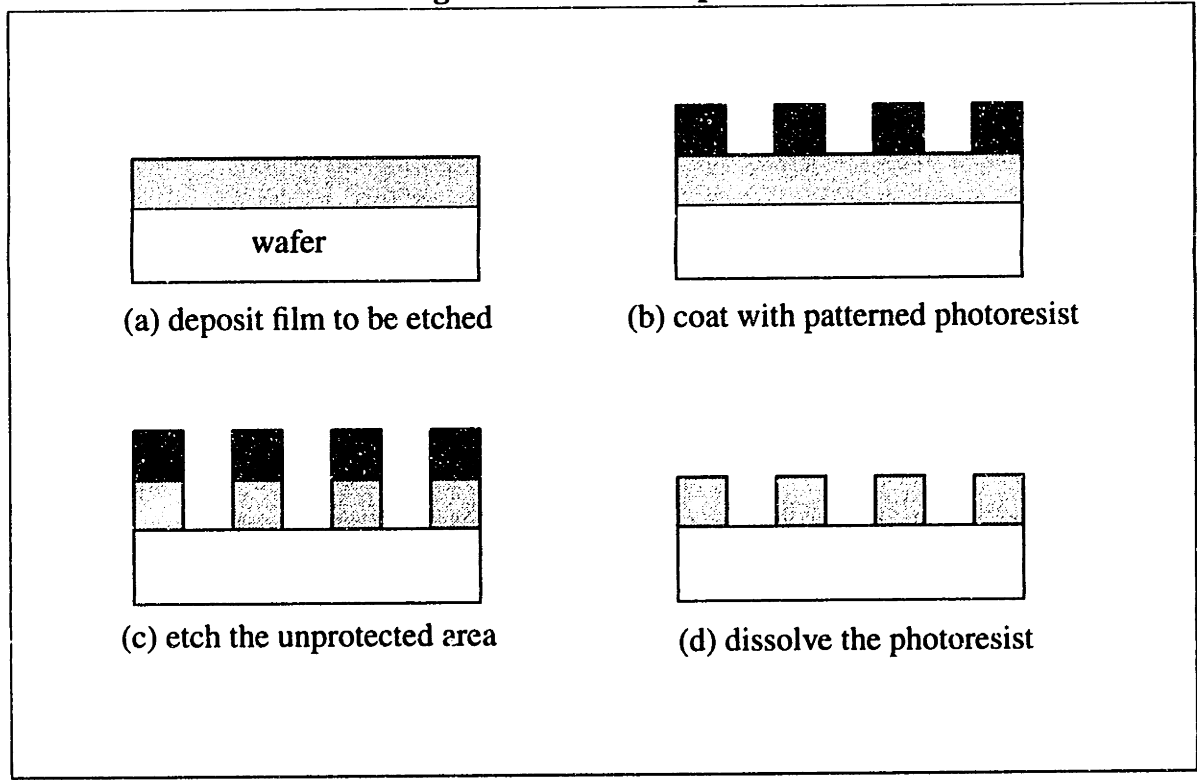
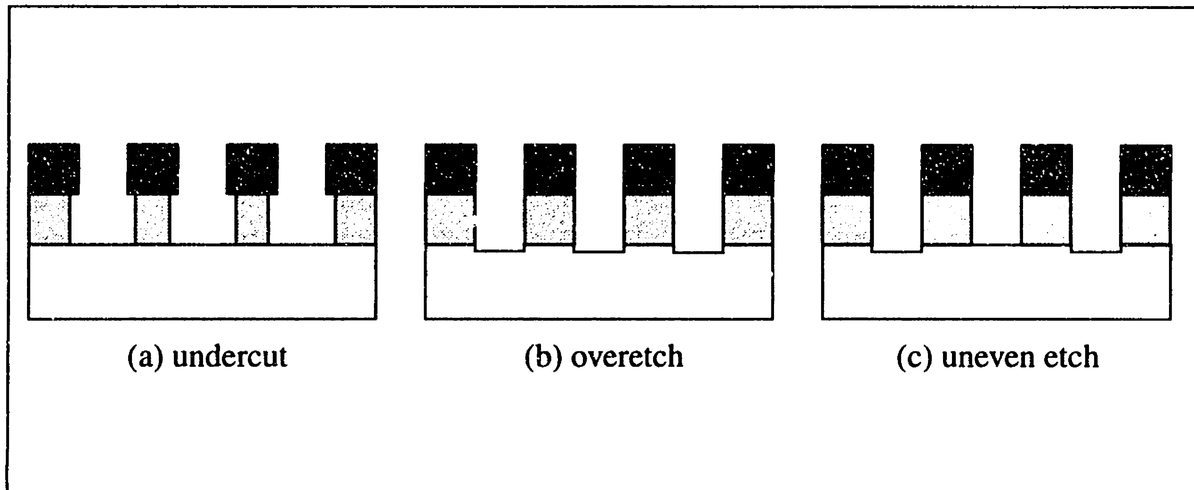


Figure 1-2: Some undesired etch results



If an etch is isotropic, part of the film underneath the protective photoresist layer will react with the plasma, resulting in line width reduction, as shown in Figure 1-2*a*. Other undesirable etch results include over-etching (Figure 1-2*b*), and uneven etching (Figure 1-2*c*) in which the etching rate could be higher (or lower) towards the edge of a wafer.

1.3 Previous work

Process control methodology in the field of semiconductor fabrication has been progressing from static process optimization to feedforward/feedback control. In this section, relevant work in semiconductor process optimization, run-by-run control, and real-time control will first be discussed. Specific work regarding the control of plasma etching will then be reviewed.

1.3.1 VLSI fabrication process control

In process optimization, response surface methods have been used to model process outputs as polynomial functions of process parameters, so that the conditions for attaining target outputs can be determined [10,50]. To reduce process variation, Taguchi methods have been applied [56]. For fabrication processes where spatial variation of product output is important, various spatial models have been developed. Guo and Sachs examined the use of a multiple response surface method. By combining separate low-order site models, they

are able to predict the spatial uniformity of an etch, which is usually a very complicated function of the control parameters [23]. Mozumder and Loewenstein developed a two-layered hierarchical spatial model for etch uniformity and selectivity. The first layer creates a spatial model of the etch, while a second layer relates the coefficients of the spatial models to the equipment settings [51].

In recent years, sensors and algorithms for both run-by-run monitoring and modification of control settings have been developed. Sachs *et al.* have proposed a modular framework for process control systems, which divides the process control domain into three modules: the flexible recipe generator, the run-by-run controller, and the real-time controller [61]. Each module is responsible for providing control action at different time scales. Also, sequential optimization techniques, such as evolutionary operation [9] and Ultramax [71], have been applied to improve process outputs on-line by refitting polynomial models after each run. Ha suggested using design of experiments to classify the process variations into both tunable variations and non-tunable variations [24]. Non-tunable variations are reduced by an off-line experimental approach, while tunable variations are regulated by adjusting suitably chosen control parameters on a run-by-run basis. Hu proposed using a Bayesian statistical method to detect process drifting and shifting for on-line process improvement [37]. Control algorithms utilizing monitor wafers combined with site models for uniformity improvement has been shown in [63]. Exponentially weighted moving average [67] and predictor corrector control [13] have been proposed to compensate for process shifts and process drifts given noisy feedback signals. Because of the complexity of the microelectronic fabrication processes, in particular for plasma etch, artificial neural networks are suggested for process modeling, which will be described in the next section. Methodologies for applying expert system for process control can also be found [14,47].

In all cases, the application of run-by-run or real-time process control depends both on the availability of feedback signals and the flexibility of modifying processing control parameters accordingly. Real-time statistical process control and diagnosis can be implemented once good sensor signals become available; an example can be found in [41]. Examples of real-time control in the field of plasma etching and rapid thermal processing can be found in [20,64].

1.3.2 Sensor and process control for plasma etching

Reliable sensor information is the fundamental key to plasma process control. An introduction to plasma characterization and process control diagnostics can be found in [35]. In general, feedback signals can be categorized into machine state, process state, and wafer state information. Throttle valve opening size is an example of machine state information. Process state and wafer state information can be connected to the process output more directly, and so they have been the main research focus.

In plasma etch, process state feedback can be obtained from optical emission spectroscopy (OES), which is widely used for endpoint detection of an etch [74]. This technique relies on the changes in the optical emission spectrum of the plasma. In general, a specific signal wavelength will be observed and a sudden shift in the emission intensity will suggest the clearing of a film. Mozumder and Barna considered etch rate and etch uniformity extraction from OES signals [52]. The etch rate is determined from the endpoint time, while the etch uniformity is determined from the slope of the OES signals at endpoint. This information is then incorporated into a control system for statistical process control and run-by-run process adjustment. For the analysis of non-uniform etch, a spatially resolved optical emission system has also been developed [5].

Other common process state sensors include pressure gauges, and RF power monitor. Use of these sensors for predicting etch properties can be found in [41]. In particular, for the purpose of endpoint detection (an important etch quality), other sensor possibilities include mass spectroscopy, impedance monitoring and Langmuir probe. Since the Langmuir probe may disrupt the plasma density, it is less preferred for on-line production monitoring. Marcoux and Foo [43] provide a general introduction to endpoint detection.

Wafer state feedback can be obtained from sensors such as interferometers or ellipsometers. Interferometer uses the constructive and destructive properties of light signals created due to the reduction of film thickness. The light source can come from the plasma emission or a laser. The intensity of laser interferometric signals is more steady, but it can only be applied to a small area over a wafer. The period of an interferometric signal can be used for etch rate determination, and a sudden change in the signal pattern will indicate etch endpoint. More details regarding interferometry will be given in Chapter 2. Since the sensor is non-intrusive, it is very suitable for process control, and has been extensively studied

[1,3,4,18,31,32,69].

By measuring the polarization of transmitted and reflected laser signals, an ellipsometer can provide both film thickness and refractive index information. Haverlag *et al.* applied ellipsometry to measure film thickness [26], and Haverlag and Oehrlein examined endpoint detection with it [27]. In [33], Henck developed real-time ellipsometry for film thickness measurement and etch control with accuracy up to ±4 Å. Further work has been done to extend the method for multilayer film stacks [34]. By measuring the etch rates at the wafer center using the real-time ellipsometer, etch rates at other sites can be inferred from spatial response surface models [68]. In addition, by combining with run-by-run control, the sensor can provide real-time endpointing at a targeted film thickness, while maintaining a maximum non-uniformity in the etch [13]. Currently, other real-time control systems for plasma etch focus mainly on process or equipment variables. Some examples of feedback control for power and DC bias can be found in [8,12,19,75]. A multivariable control system for reactive ion etching has been implemented for etch rate control [20]. A dynamic model was first constructed relating throttle positions and RF power to DC bias voltage and etchant concentration (fluorine in [20]), which are then used for real-time control of etch rate. It was shown that such a system will result in more stable etch rates even in the presence of various disturbances.

The control of plasma etching is a complicated process because many parameters are involved. Since artificial neural networks are capable of approximating functions with unknown structures, it is increasingly popular for both process modeling and process characterization. Himmel and May investigated the learning and prediction capability of neural network process models for plasma etch rate and selectivity based on RF power, pressure, and gas flow rates [36]. Huang *et al.* focused on constructing neural network models for etch rate and selectivity given a limited amount of experimental data. Their results indicate that neural network models can predict process behavior more accurately than classical regression methods at the limits of the operating range [38]. In general, the structure of a neural network model is very arbitrary. In [39], Kim and May suggest a systematic procedure for proper adjustment of neural network parameters for general modeling applications. In addition to process modeling, the pattern recognition property of neural networks has been used for real-time process control. Rietman [57,58], and Rietman and Patel [60] sug-

gested using neural network models to relate process parameters, such as applied RF power, chamber pressure, flow rate, and the optical emission trace to predict the thickness of film left, and determine the time for overetch. However, in their cases, a large amount of data are available for modeling building. To reduce the number of data required for training a neural network, Mundt developed a simple endpoint model to generate optical emission training data [53], and Allen *et al.* developed a multiresolution training method for model building [2]. In terms of the monitoring of RF power, Maynard *et al.* developed neural network models for endpoint detection. The output of the network is a classification of whether the process has reached endpoint. However, they indicate that the model has to be retrained if there is a change in the process recipe [46].

1.4 Sensor and control

Application of real-time control depends on the availability and accuracy of sensor signals. In general, without the necessary feedback signals, it could be more productive to monitor a process by statistical quality control charts and interrupt the process only when out-of-control status is recorded. In designing a control system for a process, such as plasma etching, the quality of the sensor information should always be examined before choosing the type of control strategy. Hence, the focus of this thesis is on the extraction and the accuracy of feedback information that would be useful for real-time control of plasma etching. The implication of accuracy for simple control strategies, such as adjusting the control settings once in a single process run, will be considered in Sections 3.5 and 3.6.

1.5 Thesis outline

In Chapter 2, the physics behind full wafer interferometry will be described. The sensor relies on the light source from plasma emission. Because the intensity of the plasma emission may fluctuate from one run to another, and strong signal noise is present, significant difficulties are encountered in the extraction of wafer state information from the interferometric signals. For the purpose of both run-by-by control and real-time adjustment of

control parameters, accurate etch rate extraction from the signal is necessary. In Chapter 3, four methods for determining the etch rate will be examined, including the Fourier Transform, model matching, Kumaresan-Prony and multiple signal classification methods. The influence of the different lengths of signal and sampling rates on estimation accuracy will be considered. Simulations regarding the application of accurate etch rates to real-time process control will also be illustrated.

Since the reduction of overetch is an important goal of etch control, two new algorithms for detecting endpoint will be proposed in Chapter 4. The first one is applicable for non-absorptive films, and relies on the deviation of the interferometric signal from the periodic signal trend. A new model based on fitting the interferometric signal with two sinusoids is suggested. The second algorithm is applicable for etching films with a high refractive index. It exploits the increase in periodicity of the interferometric signal when the etching film is cleared. Both algorithms are proposed because of their low computation requirements, enabling their use in real-time computation. Experimental data are used to show the efficiency of the algorithms.

One very important etch sequence today is to perform a bulk etch, followed by a final selective etch. This is particularly common among polysilicon gate etching. Thus, a method of measuring film thickness to determine when to switch chemistries is extremely useful for improved product quality. In Chapter 5, the phase of the interferometric signals will be used for estimating the film thickness left or for endpoint prediction. Sensitivity analysis with respect to small drifting etch rate and small variation in film structure will be examined. The choice of signal wavelengths for robust estimation will also be considered. Experimental data will be used to verify the accuracy of film thickness prediction, the reliability of the algorithm, and the possible drifts in the etch rate. Finally, the algorithm is applied to the signals from the full wafer interferometer for real-time monitoring of film thickness.

A summary of this thesis and future research suggestions will be given in Chapter 6. To facilitate understanding this thesis, the listing on the next page summarizes the main points in each chapter.

CHAPTER	SUMMARY
2	Full wafer interferometry – a sensor capable of monitoring etch on entire wafer in real-time – signals can be utilized for obtaining etch rate estimation, endpoint detection, and endpoint prediction
3	Etch rate estimation Methods: – Fourier transform – Model matching – Kumaresan-Prony method – Multiple signal classification Method recommended: Fourier transform
4	Endpoint detection Methods: – model deviation (for non-absorptive films) – slope fitting (for films with high refractive index)
5	Endpoint prediction (film thickness estimation) Required parameters: – signal phase at endpoint (a prior information) – film thickness etched per signal cycle – instantaneous phase – number of signal cycles remaining Phase calculation method: Fast Fourier Transform Short wavelength signals: robust with respect to film structure variation Long wavelength signals: robust with respect to initial thickness variation

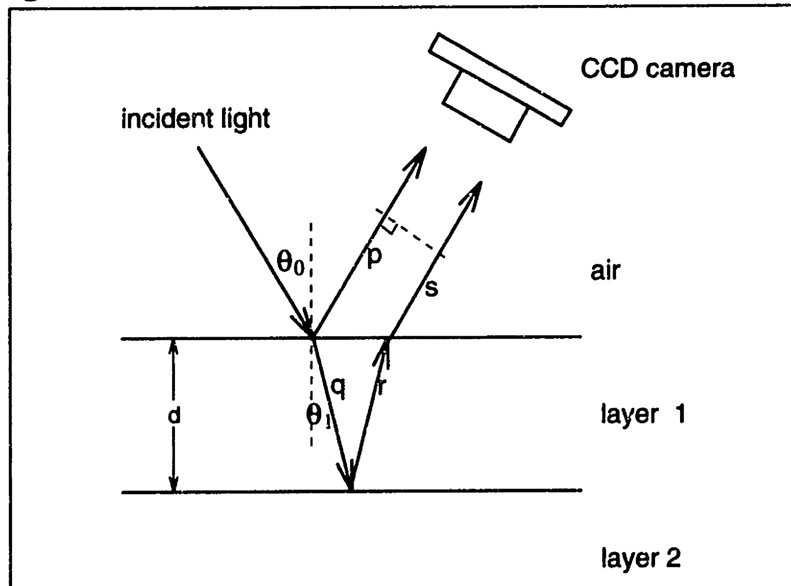
Chapter 2

Full wafer interferometry

2.1 Optical emission interferometry

During an etch, light emission from the plasma strikes the etching film surface. Part of this illumination is reflected at the top surface of the film, while part of it is transmitted into the film and reflected at the lower surfaces. As the thickness of the film decreases due to the etch, the difference in the length of the optical paths, as shown in Figure 2-1, varies and results in different levels of constructive or destructive interference. When a CCD camera with an appropriate band pass filter is used to monitor the intensity of light at each location on the wafer, each pixel acts as an individual interferometer for that location.

Figure 2-1: Full wafer interferometry for a 2-film structure



The optical path difference, δ, between the two light rays can be calculated as [30]:

$$\delta = q + r + s - p = 2n_1 d\cos(\theta_1). \qquad (2\text{-}1)$$

where n_1 is the refractive index of layer 1, d is the thickness of layer 1, and θ_1 is the angle of refraction. Other successively reflected signals are weak comparatively and can be

neglected. The interference varies with a period of λ (the signal wavelength) in the optical path difference, and leads to a cyclic signal, as shown in Figure 2-2*a*. The thickness of film etched corresponding to two successive signal maxima or minima is:

$$\Delta d = \frac{\lambda}{2n_1\cos(\theta_1)}. \tag{2-2}$$

Assuming a constant etch rate, if the time elapsed between the two successive maxima or minima (Δt) is known, the etch rate (ER) can be determined:

$$ER = \frac{\Delta d}{\Delta t} = \frac{\lambda}{2n_1\cos(\theta_1)}\frac{1}{\Delta t}. \tag{2-3}$$

In terms of the angle of incidence, θ_0, Equation 2-3 can be rearranged as:

$$ER = \frac{\lambda}{2n_1\Delta t}\left(\sqrt{1-[\sin(\theta_0)/n_1]^2}\right)^{-1}. \tag{2-4}$$

For a two-layer film structure as shown in Figure 2-1, if the film is non-absorptive, the reflectance (R) will be [29]:

$$R = \frac{(n_0^2+n_1^2)(n_1^2+n_2^2)-4n_0n_1^2n_2+(n_0^2-n_1^2)(n_1^2-n_2^2)\cos\left(\frac{4\pi}{\lambda}n_1d\cos(\theta_1)\right)}{(n_0^2+n_1^2)(n_1^2+n_2^2)+4n_0n_1^2n_2+(n_0^2-n_1^2)(n_1^2-n_2^2)\cos\left(\frac{4\pi}{\lambda}n_1d\cos(\theta_1)\right)} \tag{2-5}$$

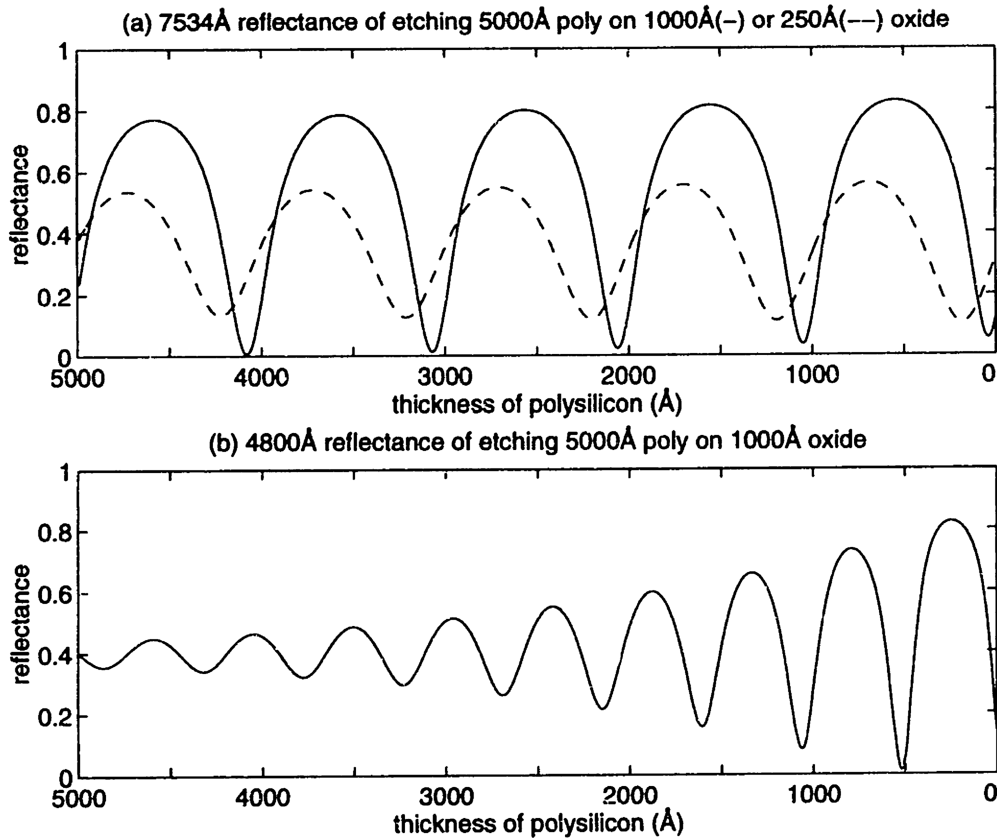
where n_0 and n_2 are the refractive indices for layer 0 and layer 2 respectively. Following [17], the correction factor for the non-zero angle of refraction is not critical, and for our analysis normal incidence will be assumed. At endpoint, $d = 0$, and

$$\frac{\partial}{\partial d}R(d=0) = 0, \tag{2-6a}$$

$$\frac{\partial^2}{\partial d^2}R(d=0) = \frac{-64\pi^2 n_0 n_1^4 n_2(n_0^2-n_1^2)(n_1^2-n_2^2)}{\lambda^2[(n_0^2+n_1^2)(n_1^2+n_2^2)+4n_0n_1^2n_2+(n_0^2-n_1^2)(n_1^2-n_2^2)]^2}. \tag{2-6b}$$

Since $n_1 > n_0$, $\frac{\partial^2}{\partial d^2}R(d=0)$ will be positive if $n_1 < n_2$, and negative if $n_1 > n_2$. This

Figure 2-2: Examples of different interferometric reflectance



means that the endpoint will occur at the maximum of signal or minimum of signal depending on the relative refractive index of the two-layer film structure.

For a multiple-layer film structure, the general shape of the interferometric signal is the same but the endpoint is not necessarily at the maximum or minimum of the signal, as pointed out in [6,15]. In [29], a matrix approach for deriving the interference reflection for multiple-layer film structure has been derived. In Figure 2-2*a*, the interferometric reflectance for etching 5000Å of polysilicon over 1000Å and 250Å of silicon oxide over a silicon substrate is shown. It can be observed that the endpoint for the 1000Å-oxide case is almost at a signal minimum while the 250Å-oxide is not.

In addition to the difference in underlying film structure, the signal absorption rate of the etching layer also modifies the signal pattern. For example, the absorption rate of a 4780Å wavelength signal by a polysilicon layer is substantial. Hence, as shown in Figure 2-2b, the amplitude of reflectance is very low when the film is thick, and it increases as the film thickness reduces.

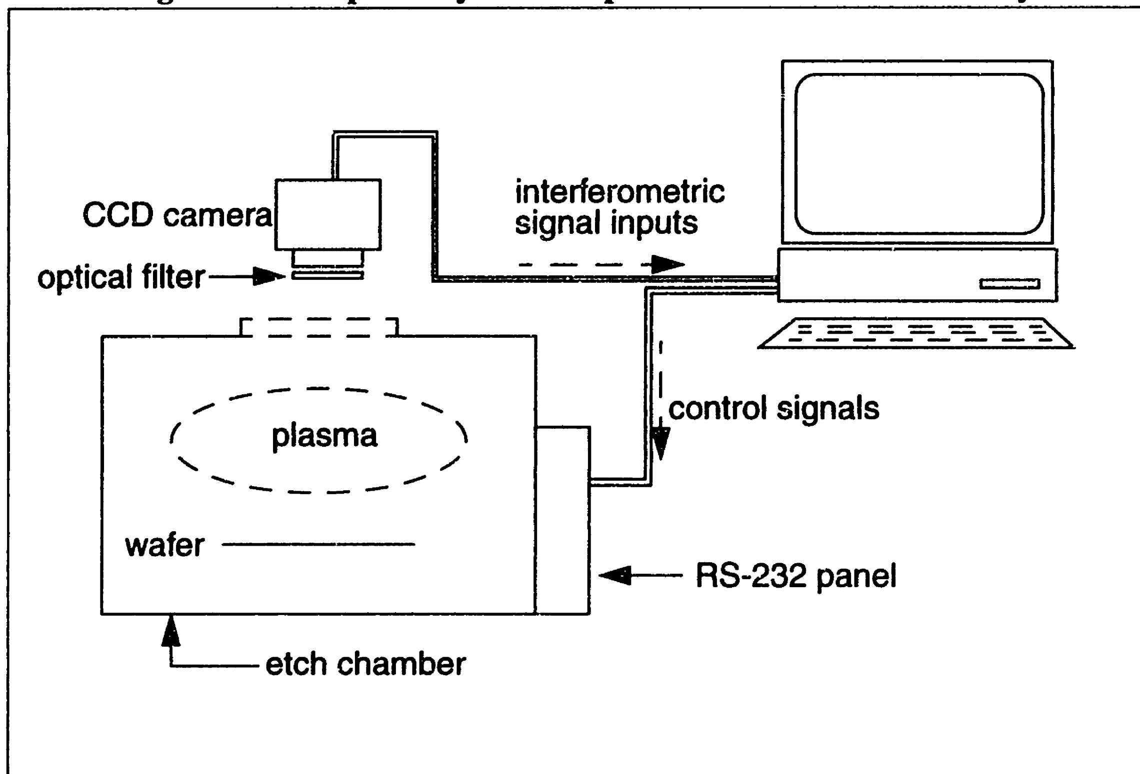
Another aspect of interferometric signals involves lateral interference caused by light

reflection from surfaces with different heights, e.g. the photoresist surface and the etching layer surface. If the light source is coherent, such as a laser, the signals as shown in Figures 2.2*a*,*b* will be distorted. Other concerns include non-uniform underlying film thickness and diffraction at step edges. Discussion of these topics can be found in [6,15,31,32].

2.2 System setup of the current full wafer interferometer

System setup of the full wafer interferometer for this work is very similar to that described in [17]. Figure 2-3 shows a rough outline of the system. A CCD camera (Hamamatsu C4742) is located over the viewport on the top of the etcher. There are a total of 1040×1024 pixels. When the CCD camera is used to monitor a 100mm wafer, each pixel approximately corresponds to a 100μm×100μm area on the wafer. The camera can capture signals at the wavelength from 400nm to 1000nm, and can produce ten-bit outputs per pixel. However, in order to reduce the demand on storage memory and sensor noise, only the highest eight bits are used. From the experimental data, it is found that such a resolution is sufficient.

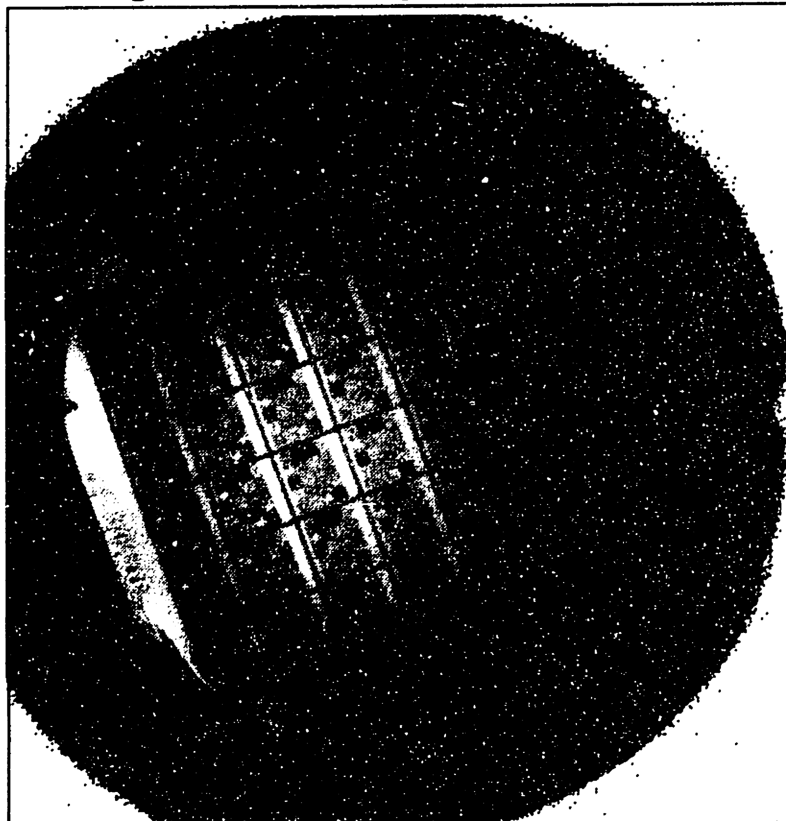
Figure 2-3: Simplified system setup for full wafer interferometry



After getting a frame, the CCD camera will transfer the interferometric data to an Alacron acquisition system via DMA (direct memory access). The Alacron system is situated inside an 80486-based host computer. It consists of an acquisition board, a processing board with two Intel i860 processors, and a display card. It is used for controlling the CCD camera and analyzing the interferometric signals. When the etch chemistry needs to be switched, or when endpoint is detected, the host computer can send control signals to the etcher. A detailed description of the machine/sensor communication system can be found in [15,22].

An optical band pass filter is required in front of the CCD camera lens so that only a narrow wavelength signal is captured. The experiments in this thesis will focus on the etch of polysilicon, and the etching chemistry will be a mixture of HBr and Cl_2. Hence, the band pass filters selected are centered at 4800Å, 6100Å and 7534Å, with half bandwidths of at least 150Å [17]. These signals capture bromine, chlorine and bromine lines in the plasma emission spectrum respectively. An example CCD image of an etching wafer is shown in Figure 2-4.

Figure 2-4: CCD image of an etching wafer



With the current setup, the maximum frame rate is approximately eight frames per second. However, as described in [17], the sampling rate of the CCD camera should be synchronized with the etcher's magnetic field rotation to reduce extra frequencies in the interferometric signals. In our experiments, the magnetic field of the etcher rotates at about 0.72 second per revolution. For simplicity, 0.72 sec will be considered as one "time unit" in this thesis.

2.3 Applied Materials Precision 5000 etcher

The Applied Materials Precision 5000 plasma etcher (AME 5000) is built for low pressure reactive ion etching. It is a multi-chamber system for single wafer processing. On the top of each processing chamber, a 50mm diameter viewport has been incorporated so that the full wafer interferometer can be set up for real-time monitoring. The etch is performed with a rotating magnetic field for improved etch uniformity. Some of the important control parameters of the etcher include chamber pressure, RF power, magnetic field strength, and the flow rate and composition of etching gases. The etcher can be connected to a host computer using SECS II (SEMI equipment communication standard II) protocol [66] via an RS-232 line for equipment monitoring and etch recipe download. In [22], the use of SECS II for run-by-run control and recipe update on the AME 5000 is examined.

2.4 Summary

Full wafer interferometry is an exciting new sensor that provides wafer state information across the entire wafer in real-time. The physics behind the sensor is similar to that of the laser interferometer. However, laser interferometry can be used only to obtain wafer state information for a limited portion of the wafer. On the other hand, the CCD camera in full wafer interferometry has thousands of pixels; each pixel functions like an individual interferometer. At present, full wafer interferometry is used for off-line analysis of etch rate only. In the following chapters, the focus will be on extracting essential information from these interferometric signals to facilitate real-time control of plasma etching, namely, the etch rate, endpoint detection and endpoint prediction.

Chapter 3

In-situ etch rate extraction from interferometric signals

3.1 Introduction

In plasma etching, accurate *in-situ* estimates of etch rates are needed for run-by-run and real-time control. The full wafer imaging system developed by Dalton *et al.* [16] facilitates *in-situ* etch rate extraction on multiple sites over the wafer surface. This chapter examines the accuracy and robustness of four algorithms for etch rate extraction from the interferometric signals. The algorithms considered are described in Section 3.2, and include Fourier transform, model matching, Kumaresan-Prony, and multiple signal classification methods. To directly improve etch uniformity with real-time control, instantaneous information on etch rate is indispensable, and so short lengths of signals are emphasized in this chapter. Simulation studies will be presented in Section 3.3 to compare method robustness against sensor noise, short data windows, and low sampling rates. In Section 3.4, experimental results of *in-situ* etch rate estimates corresponding to real-time modifications of control settings are given. The necessity of accurate etch rate determination for successful real-time process adjustments will be illustrated by simulations in Section 3.5, and an example justifying mid-course adjustment of control parameters will be provided in Section 3.6.

Instantaneous etch rate estimation from laser interferometric signals has been examined by Mishurada *et al.* [48] and Bushman *et al.* [11]. When the signal source is a laser, the baseline signal intensity is known and constant. However, applying laser interferometry over the whole surface of the wafer would be a difficult task. Full wafer interferometry using the emission from the plasma can obtain simultaneous interferometric signals at many wafer sites. However, the intensity of the plasma emission is weaker and can vary from run to run. In addition, any modification of the control parameters in the mid-course of a single process run will alter the emission intensity. Hence, we concentrate on algorithms that impose fewer assumptions on the sensor signal than are typically true in laser interferometry (e.g., strong constant light source intensity).

3.2 Etch rate estimation algorithms

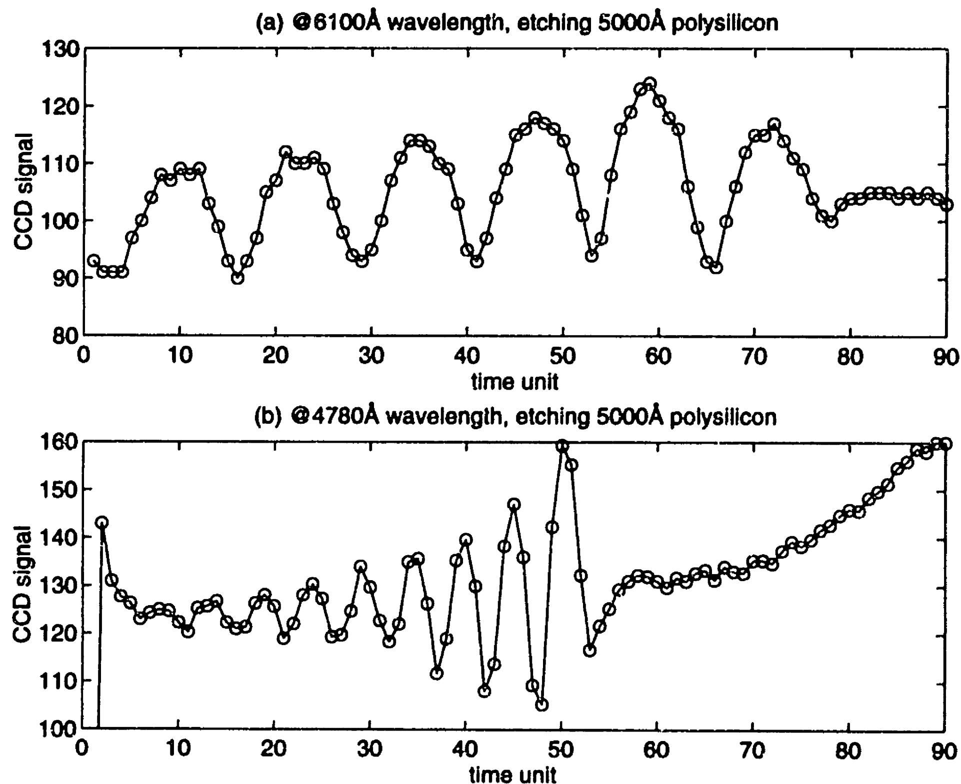
In Chapter 2, it has been shown that the etch rate (*ER*) is inversely proportional to the periodicity (Δt) of the interferometric signal:

$$ER = \frac{\lambda}{2n_1\Delta t}\left(\sqrt{1-[\sin(\theta_0)/n_1]^2}\right)^{-1} \quad (3\text{-}1)$$

where λ (signal wavelength), n_1 (film refractive index), and θ_0 (angle of incidence) can be pre-determined before an etch. Hence, the problem of etch rate estimation is transformed into that of determining the periodicity of the signal.

Signals from individual pixels in full wafer interferometry suffer from three complications compared to laser interferometry. First, the baseline intensity depends on plasma emission conditions. Second, care must be taken to deal with noise due to sensor and multiple light reflections within the chamber. Third, sampling rates could be very low (e.g. <15 samples/cycle) as shown in Figures 3-1a,b. Due to these issues, direct selection of coarsely

Figure 3-1: Examples of low sampling rates of interferometric signals



sampled extrema points is often ineffective. Therefore, effective estimation algorithms are needed to overcome these difficulties in full wafer interferometry. The algorithms considered here are the Fourier transform, model matching, Kumaresan-Prony (or Kumaresan-Tufts), and multiple signal classification (MUSIC) methods. These estimation methods all seek to determine the dominant frequency of the signal which can then be used for calculating etch rates. In particular, we compare the robustness of different algorithms against noise, their accuracy for small data sets, and their effectiveness for low sampling rates.

3.2.1 Fourier Transform Method

The Fourier transform is a well-known non-parametric tool for reconstructing the power spectrum of a signal [55]. As used here, we first subtract a linear trend from the signal. The data set is then multiplied by an appropriate window shaping to reduce the windowing edge effect. In this work the Hamming window is used [55]. The discrete Fourier transform of the data $x[1]$, ..., $x[N]$ can then be calculated as:

$$X[k+1] = \sum_{n=0}^{N-1} x[n+1] e^{(-j2\pi kn/K)}, \tag{3-2}$$

where K is the number of desired divisions of the angular frequency axis. This work uses $K = 2048$. Squaring the magnitude of $X[k]$ and dividing it by N gives the power spectrum of the data, and the fundamental frequency can then be identified as the frequency with the highest power. However, a limitation is that the reconstructed spectrum can be distorted if based on only a small amount of data. Let F be the fundamental frequency number, and T be the sampling interval, the periodicity (Δt) can be determined as:

$$\Delta t = \frac{KT}{F}. \tag{3-3}$$

In Chapter 5, the phase of the fundamental frequency (θ_N) at $x[N]$ will be needed for film thickness estimation, and its calculation is included here also. The phase at $x[1]$ (θ_1) is:

$$\theta_1 = \operatorname{atan}\left(\frac{\operatorname{imag}(X(F))}{\operatorname{real}(X(F))}\right), \tag{3-4}$$

where *imag* and *real* represent the imaginary and real part of $X(F)$ respectively. Let w_f be the fundamental frequency in radians per time unit,

$$\theta_N = \theta_1 + w_f(N-1). \tag{3-5}$$

3.2.2 Model Matching Method

In model matching, an ideal interferometric signal is first generated by using a multi-layer interferometric model [29]. The ideal signal will be an array of reflectance $R(t)$ with a known fundamental frequency (Q) in radian per time unit, and a known initial phase (θ_R). Then, the array is fitted to the acquired signal ($x[1]$, ..., $x[N]$) to determine the relative frequency parameter (w_r), bias (b), amplitude (A), and relative phase shift (θ_r), by minimizing the sum of square difference (D):

$$D(w_r, b, A, \theta_r) = \sum_{n=1}^{N} \left| x[n] - AR(w_r n + \theta_r) - b \right|^2. \tag{3-6}$$

A very coarse Fourier Transform solution can be used to provide the initial guesses of the signal frequency and phase for this algorithm. Initial guesses of the signal bias and amplitude can be obtained from the maximum and minimum value of the signal stream. To determine the fit coefficients, a non-linear optimization is required. Given a fixed length data array, we use a simplex method to determine the parameters [21]. In generating the ideal signal array, the refractive index of the film to be etched, and the thicknesses and refractive indices of the underlying layers must be known.

The estimated fundamental frequency (w_f) is:

$$w_f = Qw_r, \tag{3-7}$$

and the phase angle of $x[N]$ can be determined easily as:

$$\theta_N = \theta_R + \theta_r + w_f N. \tag{3-8}$$

3.2.3 Kumaresan-Prony Method

Both the Kumaresan-Prony method and multiple signal classification are regarded as modern spectral estimation methods [44]. These methods are of interest because their focus is on estimating the power spectrum with high resolutions from short length data sets. One main group of these methods models the data as autoregressive and moving average processes, of which the Kumaresan-Prony method is one example. This method is considered here because it is capable of modeling sums of damped exponentials, and our signals often exhibit a growing amplitude envelope (Figure 2-2*b*). Another type of spectral estimator is based on eigenanalysis of the autocorrelation matrix. The multiple signal classification approach is one such example, and will be considered in Section 3.2.4. These methods are used for frequency detection but do not reconstruct the real power spectrum for the signal. A detailed list of references for spectrum estimation can be found in [70].

The Kumaresan-Prony method [40,44] can be applied to determine the frequencies of the harmonics by assuming that the interferometric signals ($x[1]$, ..., $x[N]$) consist of a sum of m growing complex exponentials:

$$x[n] = \sum_{i=1}^{m} A_{ci} z_i^{\,n}, \tag{3-9}$$

$$\text{where } A_{ci} = A_i \exp(j\phi_i),$$

$$z_i = \exp[(\alpha_i + j2\pi f_i)T].$$

A_{ci} contains information of the initial amplitude and phase of the signals, while z_i is related to the frequencies and growth rate of the signals. Equation 3-9 is in the form of a solution for a constant coefficient difference equation:

$$x[n] = -\sum_{k=1}^{m} a[k]x[n-k]\ . \tag{3-10}$$

Substituting $x[n] = z^n$ into Equation 3-10 and dividing both sides by z^n, we get:

$$1 = -\sum_{k=1}^{m} a[k]z^{-k}. \tag{3-11}$$

Moving the right side of Equation 3-11 to the left side and factoring, $a[k]$ and z_i are found to be related by:

$$1 + \sum_{k=1}^{m} a[k] z^{-k} = \prod_{i=1}^{m} (1 - z_i z^{-1}) \cdot \tag{3-12}$$

The autoregressive coefficients $a[k]$ for the data can be determined by arranging the data in the form of:

$$\begin{bmatrix} x[1] & x[2] & \cdots & x[p] \\ x[2] & \ddots & & \vdots \\ \vdots & & \ddots & \\ x[N-p+1] & \cdots & \cdots & x[N-1] \end{bmatrix} \begin{bmatrix} a[p] \\ a[p-1] \\ \vdots \\ a[1] \end{bmatrix} = - \begin{bmatrix} x[p+1] \\ x[p+2] \\ \vdots \\ x[N] \end{bmatrix}, \tag{3-13}$$

where $p > m$ is arbitrary. The autoregressive coefficients $a[k]$ are determined by applying singular value decomposition (SVD) to the matrix above and retaining the components corresponding to the m largest eigenvalues, which filters out more noise. After solving for $a[k]$, the frequencies are:

$$f_i = \frac{\arg(z_i)}{2\pi T}, \tag{3-14a}$$

$$\text{or } w_i = \frac{\arg(z_i)}{T}, \tag{3-14b}$$

where z_i are the zeros from Equation 3-12 and T is the sampling interval. A rough estimate of the fundamental frequency can first be obtained based on a coarse scale FFT result. The correct fundamental frequency (w_f), which should be close to the rough estimate, can then be selected easily.

After determining z_i's, A_{ci} can be obtained from Equation 3-9 using linear regression. The initial phases corresponding to f_i (ϕ_i) are the angle of A_{ci}, and the phase of the

fundamental frequency (θ_N) at $x[N]$, can be calculated easily:

$$\theta_N = \phi_f + w_f N. \tag{3-15}$$

3.2.4 Multiple Signal Classification Method

The multiple signal classification (MUSIC) method [44,65] starts by obtaining an autocorrelation matrix, R_p, for the data ($x[1]$, ..., $x[N]$). We use the sample autocorrelation:

$$R_p = \frac{1}{p}\sum_{k=1}^{p} X_k X_k^{\perp}, \tag{3-16}$$

where $X_k = \left[x[1]\ x[2]\ \ldots\ x[N-p+1]\right]^{\perp}$ and

$\perp$ is used to indicate vector transpose.

Applying eigendecomposition on R_p gives:

$$R_p = \sum_{i=1}^{m} \lambda_i v_i v_i^{H} + \sum_{i=m+1}^{p} \sigma_i v_i v_i^{H}. \tag{3-17}$$

where H is used to indicate complex conjugate vector transpose. The λ_i's are the m largest eigenvalues corresponding to both the signal and the noise subspace, while the σ_i's are the p-m smaller eigenvalues (which are identical theoretically) corresponding only to the noise subspace. The eigenvectors (v_i) corresponding to the m largest eigenvalues (λ_i) span the signal subspace. With $r(f)$ defined as the eigenvectors for exponentials:

$$r(f) = [1\ e^{(j2\pi fT)}\ e^{(j4\pi fT)}\ \ldots\ e^{(j2\pi f(p-1)T)}]^{\perp}, \tag{3-18}$$

$$P(f) = \sum_{i=1}^{m} \left| r^{H} v_i \right|^2, \tag{3-19}$$

where T is the sampling interval. $P(f)$ can be treated as the sum of the magnitude of the projection of $r(f)$ onto the signal subspace. The fundamental frequency and its harmonics would correspond to peaks of $P(f)$, and $P(f)$ at various frequencies can be computed effi-

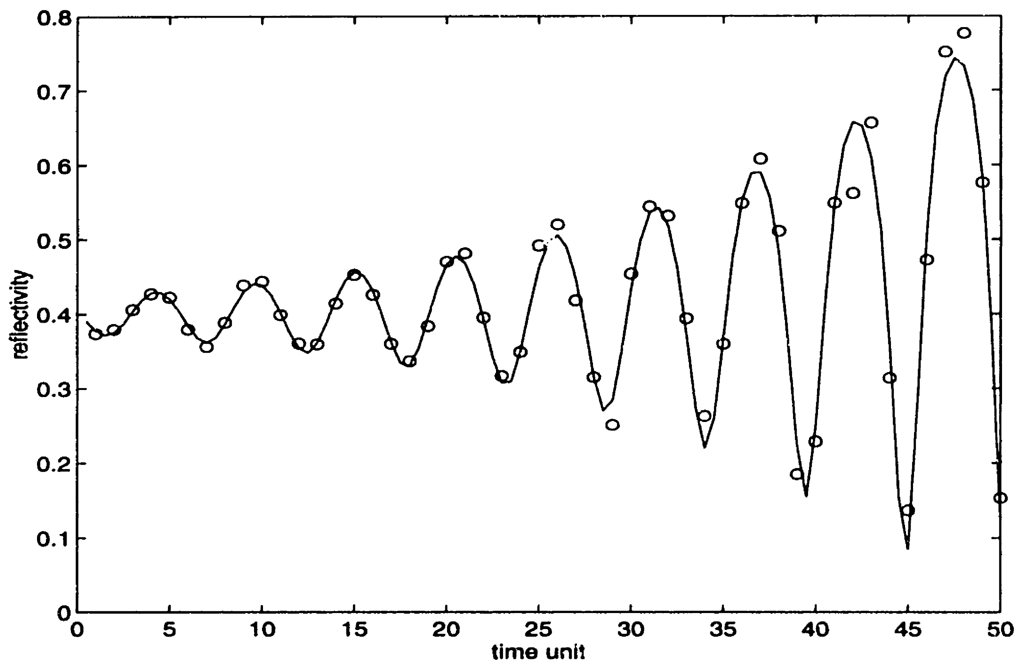
ciently with the Fast Fourier Transform. Multiple signal classification method is useful only for frequency detection, and cannot be used for phase estimation.

Both the Kumaresan-Prony and the multiple signal classification methods give fundamental frequency estimates and those of the higher harmonics. A rough estimate of the fundamental frequency can be obtained from prior knowledge, or from coarse FFT results.

3.3 Simulation results

Simulations have been performed to compare the above algorithms. Figure 3-2 shows an ideal and a noise contaminated simulation signal during the etch of 5000Å of polysilicon over 1000Å of silicon dioxide on a silicon substrate. The 4780Å-wavelength signal is considered because this gives more cycles of signal during the etch of a thin film, which will be shown to improve the accuracy of the estimates. In addition, the strong absorption rate at this wavelength gives rise to a more complicated signal, and enables study of estimation in a more difficult scenario. Sensor noise is added to the interferometric signal to examine the robustness of different estimation algorithms. Noise due to the sensor hardware and its output quantization may have a constant variance; this is already very small and can be mini-

Figure 3-2: Reflectivity of ideal interferometric signal and simulated noisy readings



mized with more sophisticated camera hardware [17]. Noise due to internal reflection within the chamber, on the other hand, is significant, and may be proportional to the intensity of the signal. In this chapter, the type of noise examined is normally distributed with a standard deviation proportional to signal amplitude.

In Figure 3-3, the dependence of estimates based on different number of cycles of signals is shown (with constant sampling rate). Windows of both 14 data points (2.4 cycles of signal) and 29 data points (3.3 cycles of signal) are slid through the interferometric signal as shown in Figure 3-2. All frequencies of the signal within each data window are assumed to be constant. The outputs are the dominant frequency normalized by the actual value (known for the simulation); thus 1 is the ideal output. The figures show that the longer the data window, the more accurate the outputs.

To obtain statistics for the estimators, a total of 50 simulations similar to those shown in Figure 3-3*a* to *d* are performed, and the means and standard deviations of the estimates are obtained under different scenarios. Table 3-1 summarizes the results of the simulations. Case A is a baseline reference in which a 2.4 cycle window is used. Case B extends the signal window to 3.3 cycles. In case C, the noise level is increased. In case D, both the number of cycles is reduced and the noise level is increased. In case E, the sampling rate is doubled; the data window still contains 2.4 cycles of signal.

Results show that the mean etch rate estimates of all four methods are very close to the actual value. Although reducing the data window size decreases the etch rate acquisition delay, accuracy will deteriorate greatly. Among the estimators, the model matching method gives a smaller variance. However, since the physical model used in this method requires values for various properties of the films underneath the etching film, its application can be limited. Although the Fourier transform method has low resolution and is computationally simple, its accuracy is comparable to the other two methods, particularly when the data length is sufficient. In addition, it can be observed that the accuracy of all methods increases with the sampling rate. The total number of cycles of signal is a function of the light source and the film properties and cannot be increased. Therefore, increasing the sampling rate is an important avenue for improved estimation.

Figure 3-3: Example of moving window etch rate estimates for different algorithms. Shown are normalized etch rate (1 is perfect estimate).

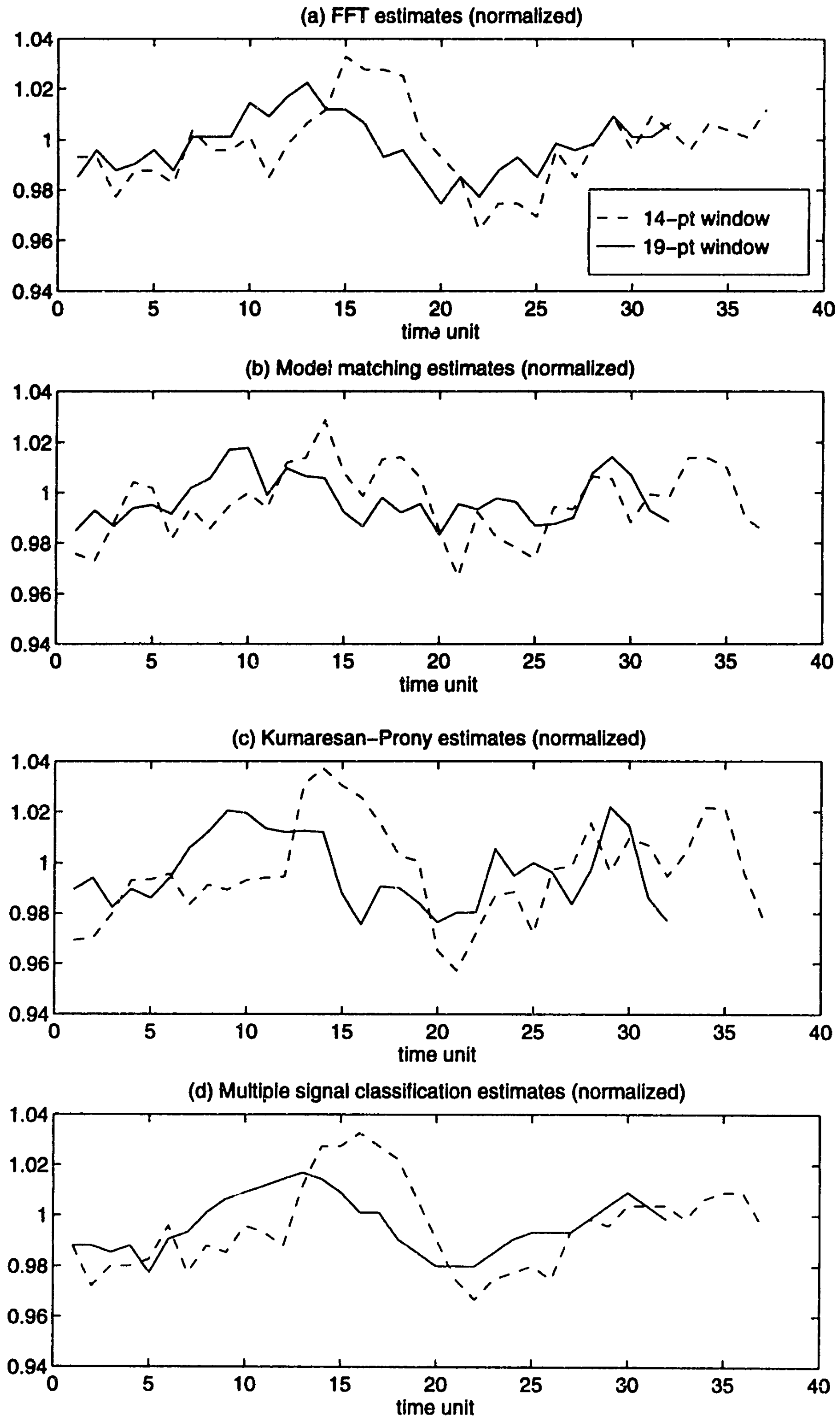


Table 3-1: Normalized mean and (std. dev.) of etch rate estimates (ideal mean=1)

	case A	case B	case C	case D	case E
	ref.	lg. win.	more noise	sh. win., more noise	samp. rate doubled
FFT	0.9997 (0.0170)	0.9994 (0.0116)	0.9993 (0.0331)	0.9925 (0.0565)	0.9990 (0.0119)
model matching	0.9983 (0.0157)	0.9984 (0.0106)	0.9981 (0.0289)	0.9993 (0.0470)	0.9994 (0.0135)
Kumaresan-Prony	0.9988 (0.0200)	0.9976 (0.0149)	0.9995 (0.0398)	0.9992 (0.0624)	1.0006 (0.0155)
multiple sig. class	0.9998 (0.0173)	0.9987 (0.0113)	0.9996 (0.0348)	0.9957 (0.0530)	1.0005 (0.0125)
sq. rt. of CRLB	0.0096	0.0058	0.0192	0.0355	0.0070

Case: A (reference) .... 14 data point window (~2.4 cycle window)

.... noise ~ 0.15*N(0,1)*signal amplitude

B (large data window) .... 19 data point window (~3.3 cycle window)

.... noise ~ 0.15*N(0,1)*signal amplitude

C (more noise) .... 14 data point window (~2.4 cycle window)

.... noise ~ 0.3*N(0,1)*signal amplitude

D (shorter window, more noise) ..... 10 data point window (~1.7 cycle window)

.... noise ~ 0.3*N(0,1)*signal amplitude

E (sampling rate doubled) .... 27 data point window (~2.4 cycle window)

.... noise ~ 0.15*N(0,1)*signal amplitude

The Cramer-Rao lower bound (CRLB) provides a theoretical limit on the smallest variance that can be achieved from any unbiased estimation. Consider a process $y[n]$, which is composed of m complex exponentials,

$$y[n] = x[n] + \varepsilon[n] \tag{3-20}$$

where $x[n]$ is defined in Equation 3-9, and $\varepsilon[n]$ is normally distributed independent noise

with variance $\sigma[n]^2$. The logarithmic likelihood function is defined as:

$$L(Y|\Gamma) = -N\log(2\pi) - 2\log(\sigma[1]\times\sigma[2]\times\ldots\times\sigma[N]) - (Y-X)R^{-1}(Y-X)^H, \quad (3\text{-}21)$$

where $Y = [y[1]\ \ y[2]\ \ \ldots\ \ y[N]]$,

$X = [x[1]\ \ x[2]\ \ \ldots\ \ x[N]]$,

$\Gamma = (A_1, \alpha_1, \phi_1, f_1, \ldots, A_m, \alpha_m, \phi_m, f_m)$,

$R = \mathrm{diag}(\sigma[1]^2, \sigma[2]^2, \ldots, \sigma[N]^2)$.

The CRLB is defined as the inverse of the Fisher information matrix (I):

$$I(\Gamma) = var\left(\frac{\partial}{\partial\Gamma}L(Y|\Gamma)\right). \quad (3\text{-}22)$$

The diagonal terms of the resulting inverse are the lower bounds on the variance of the estimates. Yao and Pandit provide a detailed development of the bound in [76]. By assuming that the interferometric signal is composed of two growing sinusoids, the square-root of the CRLBs calculated for various scenarios are provided in Table 3-1. Since the second and higher order harmonics of an interferometric signal are of very low power (less than ~1%), including these terms will only increase the CRLBs. It can be seen that the Fourier transform method gives estimates with variance less than twice the lower bound.

3.4 Experimental results

Figure 3-4a shows a plot of experimental data during the etch of 5000Å of polysilicon. The experiment was performed on an Applied Materials Precision 5000 plasma etcher. The wavelength of the signal is at 4780Å. An important control parameter is the magnetic field strength. Starting with a setting of 50 gauss (50 gauss), the magnetic field strength was modified to 79 gauss and then to 59 gauss during the course of the etch. The average intensity of the signal changes with the magnetic field strength and so does its periodicity.

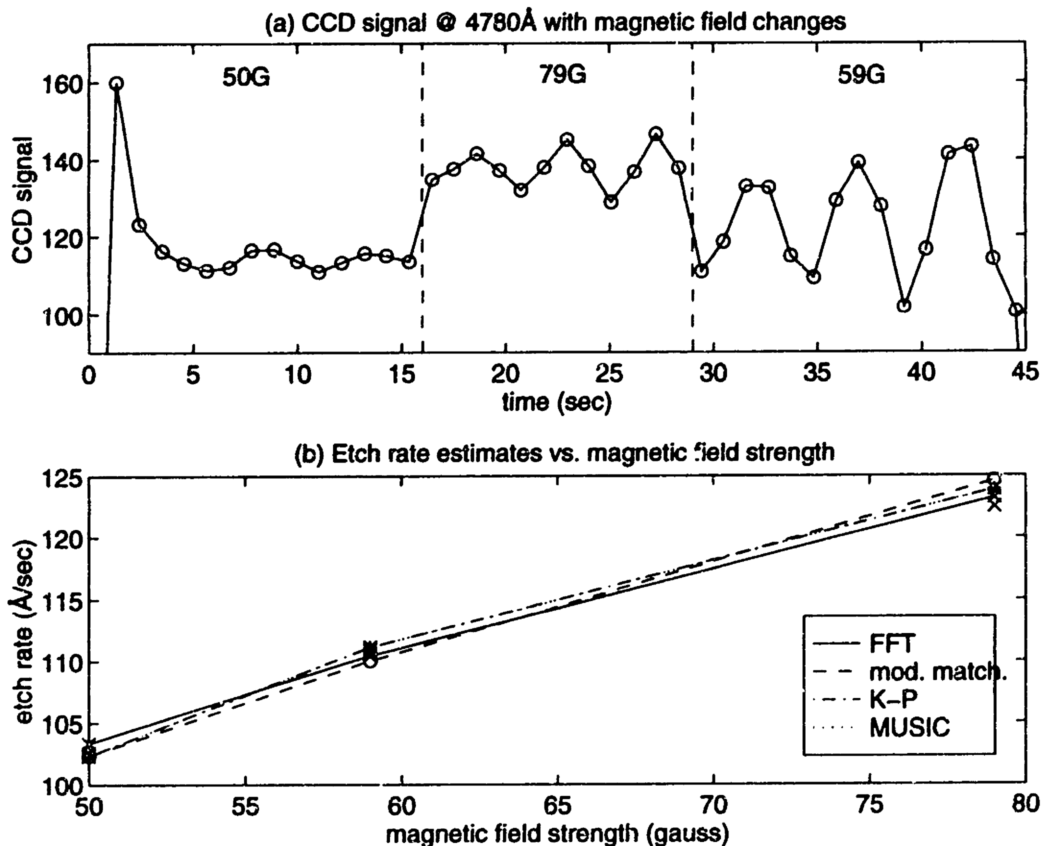
By examining local signal extrema from the interferometric signal, a rough approximation of the etch rate can be obtained. During the first etching stage, minima of the signals

occur at 5.66 sec and 11.04 sec. Therefore, the etch rate is approximately:

$$\frac{\lambda}{2n_1}\frac{1}{\Delta t} = \frac{4780}{2\times 4.416}\frac{1}{11.04-5.66} = 100.60\text{Å}/\text{sec}, \tag{3-23}$$

which is close to the estimates obtained from the four methods considered. In Figure 3-4b, the calculated etched rate using the four methods examined is plotted against magnetic field strength. It can be seen that the performances of the methods are comparable.

Figure 3-4: Etch rate estimation from an etch with changing control settings



3.5 Effect of etch rate estimation accuracy

In general, an etch process can be carried out as a bulk etch and an over-etch. To improve the etch uniformity so that damage to the underlying film can be minimized, etching chemistry may be switched from an anisotropic and fast etch to a more selective and

slower etch. A second approach could be a change of control settings in the mid-course of an etch to improve the etch uniformity. Chapter 5 will focus on sensor-based methods to determine the appropriate time for chemistry switching or predicting endpoint; these methods facilitate the bulk/selective etch approach. On the other hand, it will be shown in this section, by simulations, that accurate etch rate estimation will be essential for successful mid-course modification of control settings. In Section 3.5.1, the model used for simulating etch performance will be described, and in Sections 3.5.2 and 3.5.3, simulation results to evaluate control alternatives will be presented.

3.5.1 Neural network etch model

A neural network etch model is built which relates the single control (recipe) parameter to the etch output for an Applied Materials Precision 5000 etcher. The data used for building the model was obtained from [24]. Some of the key control parameters of the etch are the magnetic field strength, the RF plasma power, the chamber pressure, and the etching chemistry. In the original experiment design, the power was fixed at 375 watts, the pressure was fixed at 100 mtorr, and the etchants were HBr/Cl_2 at 10/30 sccm. Magnetic field strength was adjusted from 20 gauss to 100 gauss to produce different etch outputs. Figure 3-5 shows the patterns on the wafer. Average etch rates within the holes of the numbered dice were obtained by pre-etch and post-etch thickness measurements of incompletely etched polysilicon films [24], and are listed in Appendix A.

One way of building an etch model for the process is training an artificial neural network with the data [28,36]. An artificial neural network is a form of non-parametric data fitting which consists of an input layer, an output layer, and selected levels of hidden layers. Figure 3-6 shows the structure of the neural network used in this exercise. The magnetic field is first multiplied with the weights w_{11} and w_{12}. The output is then added with the bias term b_{11} and b_{12}, and passed through the function Γ. The computation for the second layer is performed similarly. By using error backpropagation to minimize the sum of squared error between the experimental data and the model output, the weight constants can be determined. The resulting network predicts the etch rates for each of die 1 through die 12 in Figure 3-5.

Figure 3-5: Pattern of wafer selected for etch model building

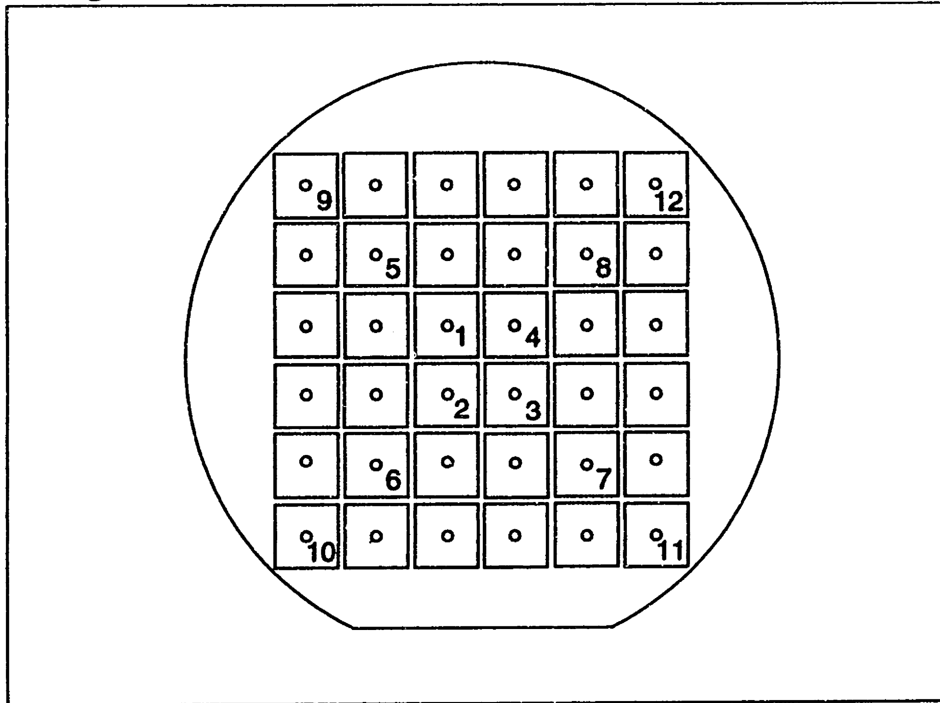
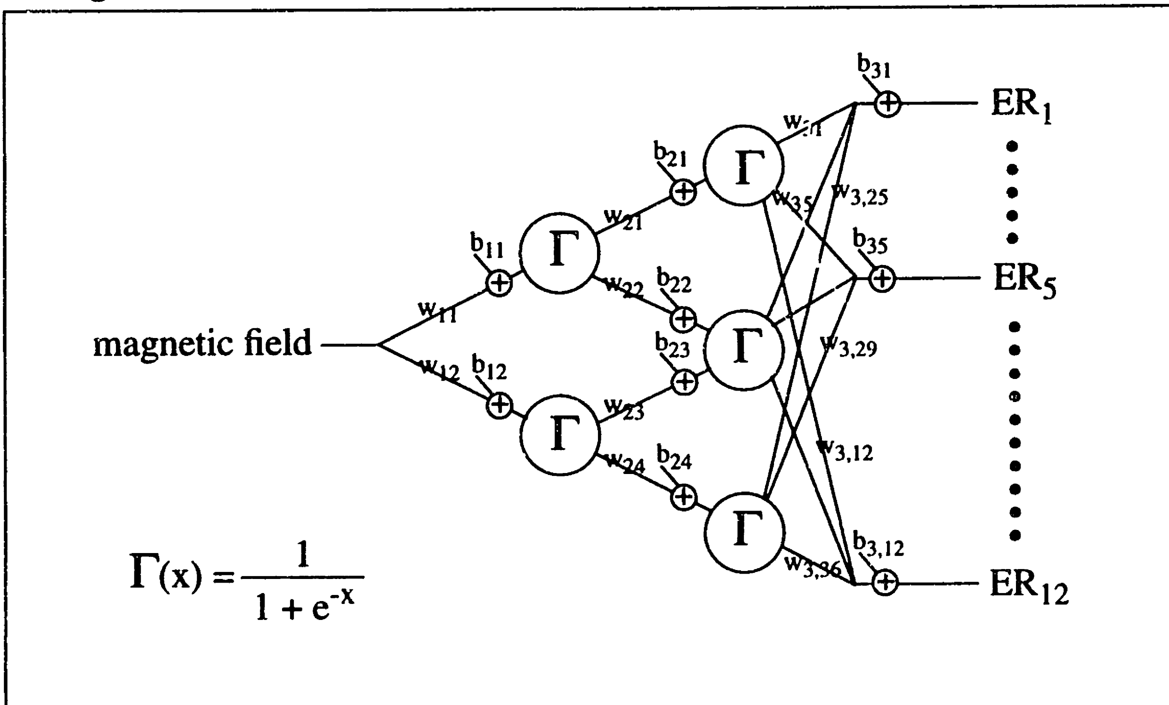


Figure 3-6: Structure of an artificial neural network for modeling etch rates



3.5.2 1-mid-course correction control

In the following simulations, radial etch rate and etch thickness non-uniformity will be considered. Define Y_1, Y_2 and Y_3 as the average etch rate of ER_1 to ER_4, ER_5 to ER_8, and ER_9 to ER_{12} respectively. Also, define Y_m to be the overall average of ER_1 to ER_{12}. Then the radial etch *rate non-uniformity* (*RN*) can be defined as:

$$RN = \frac{Y_3 - Y_1}{2Y_m} \times 100. \tag{3-24}$$

Correspondingly, the radial etch *thickness non-uniformity* (*TN*) for an etch from time 0 to time t_m can be defined as:

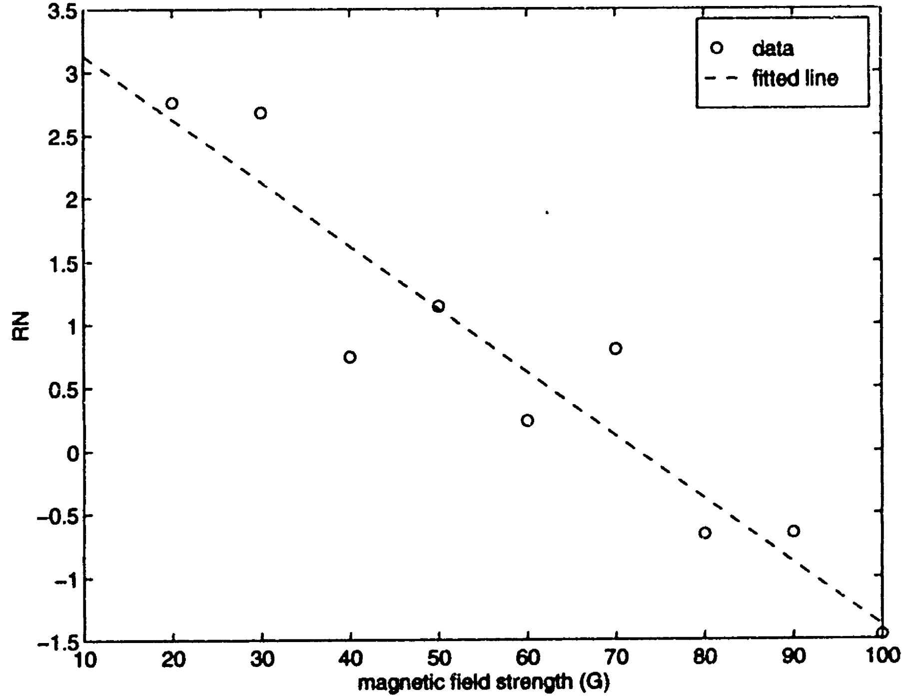
$$TN = \frac{\int_0^{t_m}(Y_3 - Y_1)dt}{2\int_0^{t_m} Y_m dt} \times 100. \tag{3-25}$$

Here, *etch thickness* is the amount of material removed as opposed to remaining film thickness. For simplicity, following [24], it can be shown that *RN* can be modeled as a first order function of the magnetic field strength (*B*), as illustrated in Figure 3-7. The fitted equation is:

$$RN(B) = 3.6297 - 0.0501B. \tag{3-26}$$

If the etch rate does not change, *RN* and *TN* are equivalent. However, they are different if control parameters are to be modified in the middle of a run. By using new incoming data during a process to update Equation 3-26, radial etch rate uniformity can be improved to reduce final overall etch thickness non-uniformity. For the simulations in this section, mid-course correction will be used, i.e., the control setting will be adjusted only very few times during a process run. In particular, only one process adjustment will be considered.

Figure 3-7: Radial etch rate non-uniformity from experiments



Two scenarios are examined. In the first case, assume that the etch was initially run for 25 seconds. From the etch rates received, the radial etch rate non-uniformity is determined and the constant coefficient in Equation 3-26 is updated. Then, the magnetic field strength is modified so that the subsequent radial etch non-uniformity will be zero, i.e. the new setting can be calculated as:

$$B_{new} = \frac{(RN_1 + 0.0501 B_1)}{0.0501} \qquad (3\text{-}27)$$

where RN_1 is the radial etch rate non-uniformity determined from the initial etch period, and B_1 is the initial magnetic field strength.

The second case is similar to the first case, except that the final magnetic field strength is to be selected to compensate for part of the non-uniform etch thickness. After the first etch period, the constant coefficient is modified. The thickness of the remaining film at the same twelve selected sites is obtained. The determination of the unetched film thickness will be discussed in more detail in Chapter 5. From the film thicknesses left (TH_1, TH_2, ..., TH_{12}), the radial film thickness *left non-uniformity* (*LN*) is calculated:

$$LN = \frac{mean(TH_9, \ldots, TH_{12}) - mean(TH_1, \ldots, TH_4)}{2 \times mean(TH_1, \ldots, TH_{12})} \times 100. \tag{3-28}$$

The final magnetic field is adjusted to achieve *LN* in *RN* or:

$$B_{new} = \frac{-(LN - RN_1 - 0.0501 B_1)}{0.0501}, \tag{3-29}$$

so that the film will be cleared uniformly. During the etch simulation, machine noise is incorporated into the control setting, which is equal to 1.5×N(0,1) (a normal random variable with variance 2.25). Also, to simulate the effect of measurement noise and etch rate estimation error, various levels of gaussian noise are added. A flow diagram for the simulation is shown in Figure 3-8. Ideally, the overall etch thickness non-uniformity should be zero. However, the output deteriorates because of the presence of noise.

Figure 3-8: Flow chart for 1-mid-course correction control

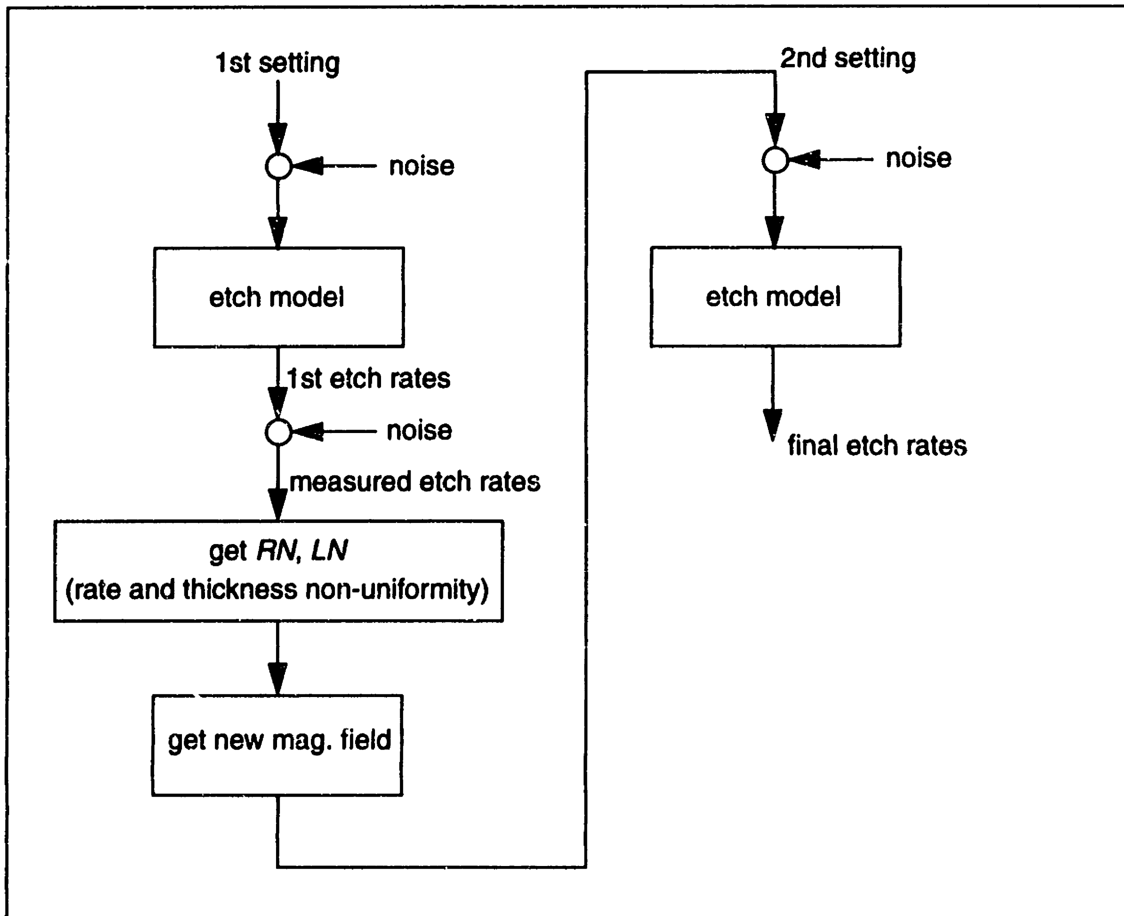


Figure 3-9 shows the results for the no-control case, 1-mid-course correction without compensation, and 1-mid-course with compensation. It is assumed here that the starting film thickness is uniform, and that all simulations were started with the same non-optimal setting at 45 gauss. The first etch lasts for 25 seconds, while the final etch is stopped when the mean of the total etch thickness is 5000Å. At each level of measurement noise, 40 simulations were performed and the upper bounds, the means, and the lower bounds of *TN* (removed thickness) are shown. For the no-control case, variation about the mean is due solely to fluctuation in equipment settings. At low measurement noise levels, both 1-mid-course correction algorithms are able to reduce the radial etch thickness non-uniformity. Furthermore, mid-course correction control with compensation is able to bring the process closer to the zero target value. However, if the measurement noise is too high, mid-course adjustment of the control setting may result in poor etch uniformity. In Figure 3-10, the simulations were started at the optimal setting. It can be observed that even though the performance of the initial process is fine, large errors in etch rate information have driven the etch to be non-uniform. Hence, accurate estimates of etch rate are essential for effective real-

Figure 3-9: RN of total etch for 1-mid-course control w/ non-optimal start as a function of etch rate estimation error. (showing means, upper bounds, and lower bounds)

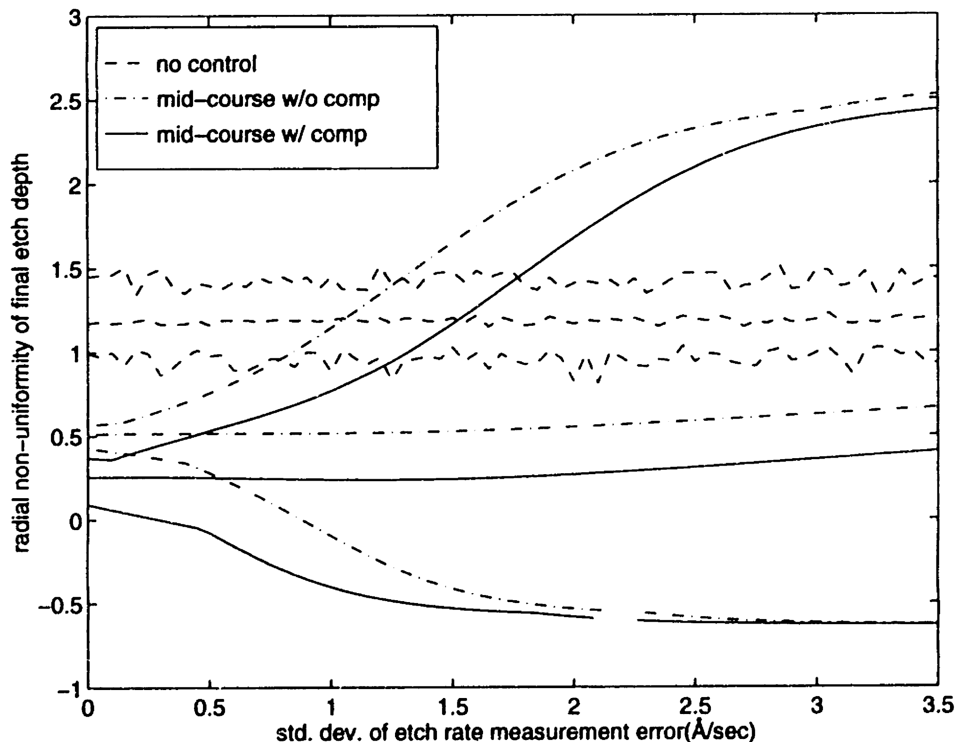
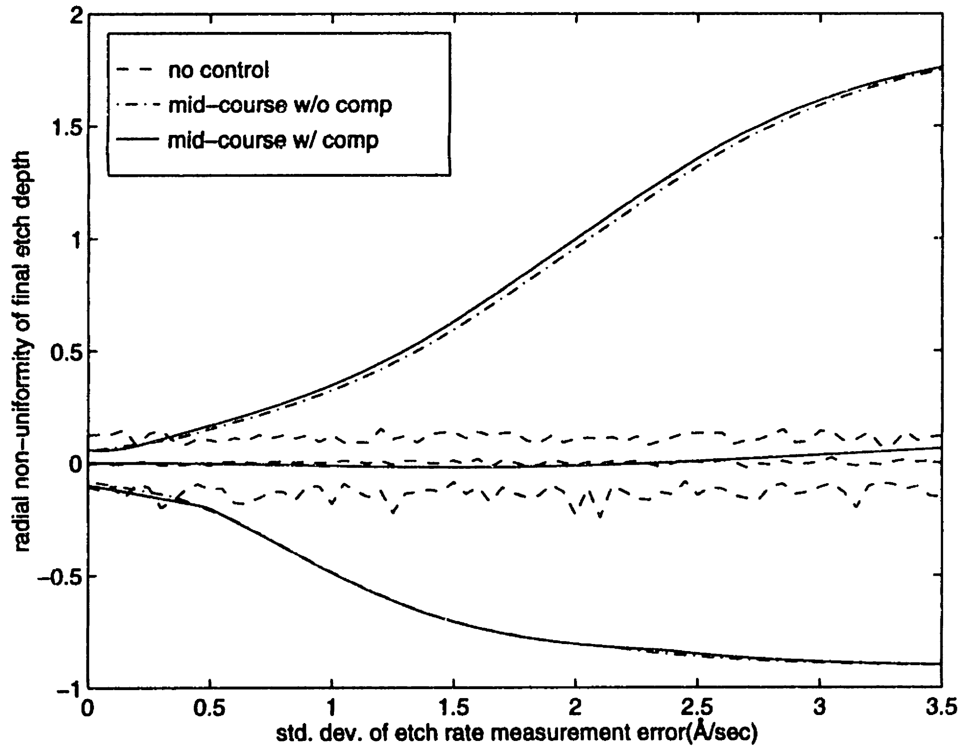


Figure 3-10: RN of total etch for 1-mid-course control w/ optimal start as a function of etch rate estimation error (showing means, upper bounds, and lower bounds)



time control of plasma etching. In Section 3.5.3, simulation results regarding 2-mid-course correction control (where process conditions are modified in two stages during the etch) are discussed.

3.5.3 2-mid-course correction control

In 2-mid-course correction control, the control settings of a process are adjusted only twice during a single run. Similar to the last section, the radial etch rate non-uniformity (*RN*) is modeled as a linear function of the magnetic field strength (*B*):

$$RN = C_0 + C_1 B . \tag{3-30}$$

After the first etch, based on *RN* determined from the interferometric signals, C_0 is updated. The magnetic field strength for the second stage is chosen to achieve more uniform etch rates. After the second etch stage, both C_0 and C_1 are updated, and a new magnetic field strength is again selected to achieve an overall uniform etch after the last stage.

In selecting new operation settings for a mid-course adjustment, non-uniform etching at the initial stages within the same run could also be compensated if such settings are allowed. For example, the required magnetic field must not go beyond the equipment capability, nor should it be high enough to damage the gate oxide. In our simulations, the 2-mid-course correction control with compensation is also examined. Here, both the first and second etch segments are of the same duration. The first mid-course adjustment is chosen such that the *RN* is set to $-RN_1$ determined from the first etch stage. After the second etch stage, since both C_0 and C_1 are more accurately determined, the final magnetic field strength is then selected such that *RN* equals *LN* of the remaining film thickness, i.e., the etch rate will be higher where the film is thicker. Figure 3-11 illustrates the trends of *RN* and *TN* for 2-mid-course correction control with compensation, and Figure 3-12 shows what the film may look like under a 2-mid-course correction control.

Similar to the simulations shown in Section 3.5.2, the magnetic field inputs are superimposed with gaussian noise to simulate the equipment parameter fluctuation, while the outputs are superimposed with gaussian noise to simulate the etch rate information errors. The flow chart for the simulation is shown in Figure 3-13. The control parameter fluctuation is selected to be 1.5×N(0,1), where N(0,1) is a normally distributed random variable. Each

Figure 3-11: Goal of RN and TN for 2-mid-course control w/ compensation

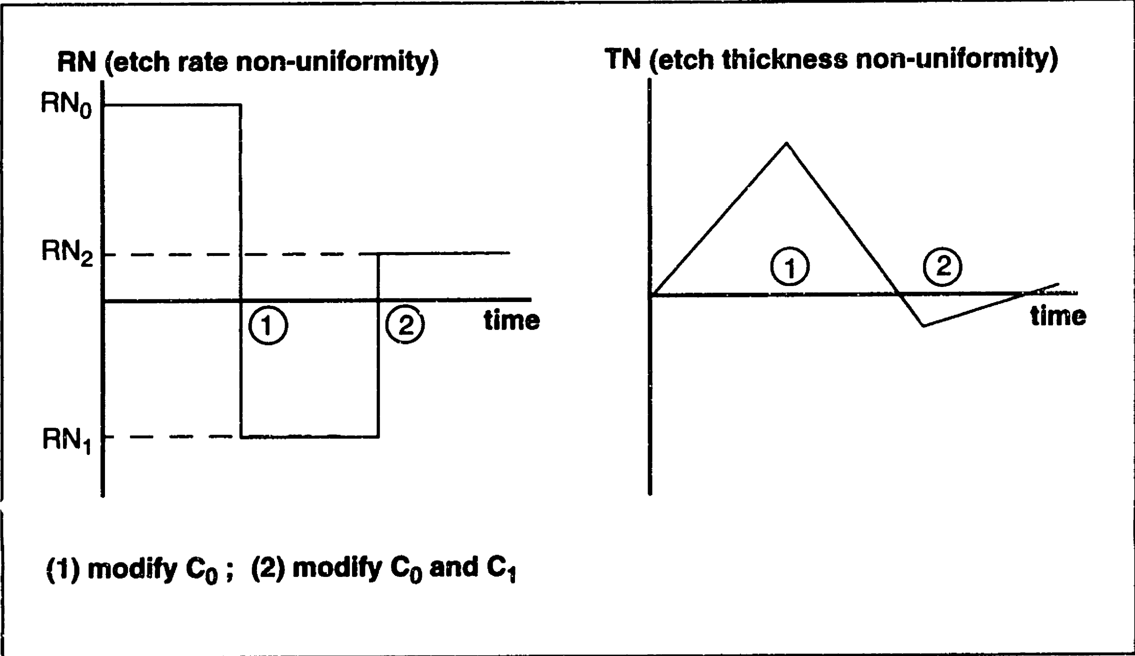


Figure 3-12: An example of film surface under 2-mid-course control

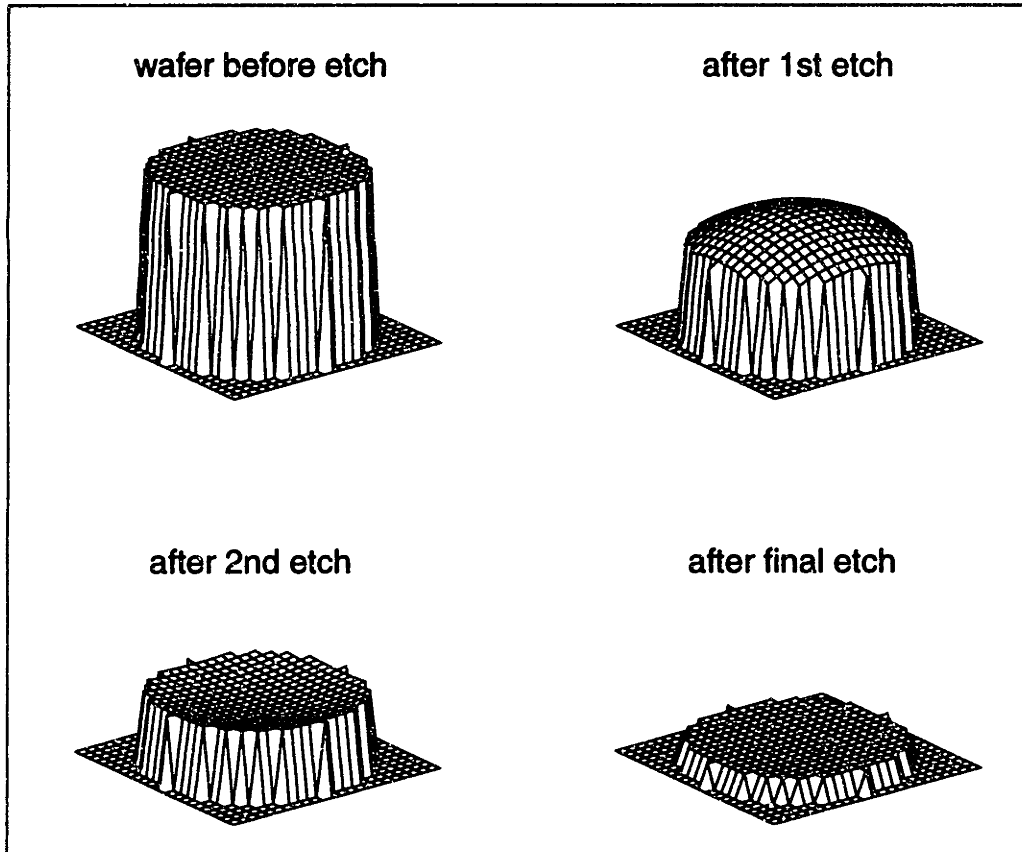
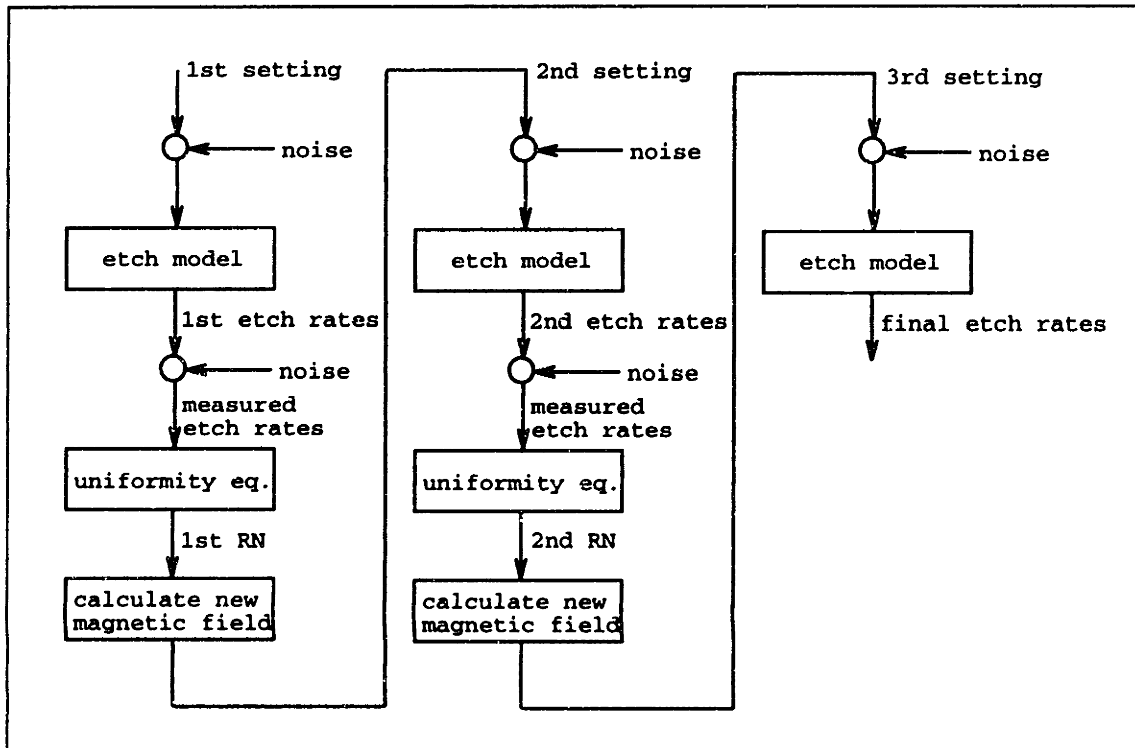


Figure 3-13: Flow chart for 2-mid-course correction control etch simulation



of the first etch stage and the second etch stage last for 25 seconds, while the final etch is stopped when the mean of the total etch thickness is 5000Å. For each level of etch rate variation, a total of 40 simulations are performed. Figure 3-14 shows the upper bound, the mean, and the lower bound of the overall *TN* when the initial magnetic field strength is non-optimal (selected to be 45 gauss). Figure 3-15 shows the performance for the case when the initial magnetic field strength is optimal. It can be observed from the figures that accurate etch rate information is essential for successful mid-course correction control. In fact, if the etch rate estimates are extremely inaccurate, it is shown that mid-course correction control with 100% compensation can performance worse than the corresponding non-compensating case.

Figure 3-14: RN of total etch for 2-mid-course control w/ non-optimal start as a function of etch rate estimation error (showing means, upper bounds, and lower bounds)

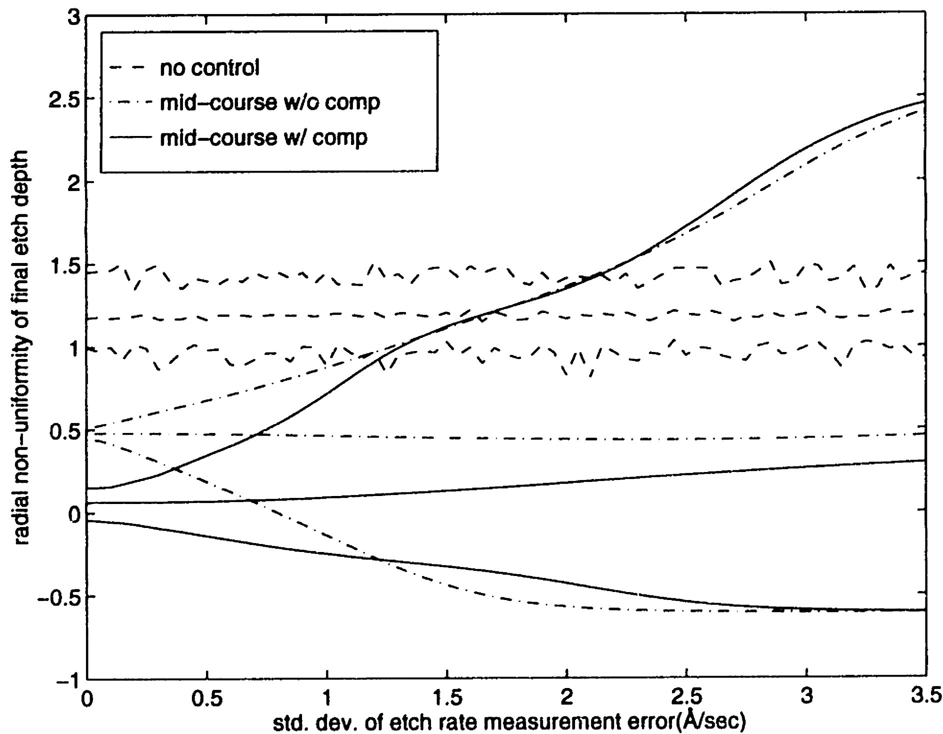
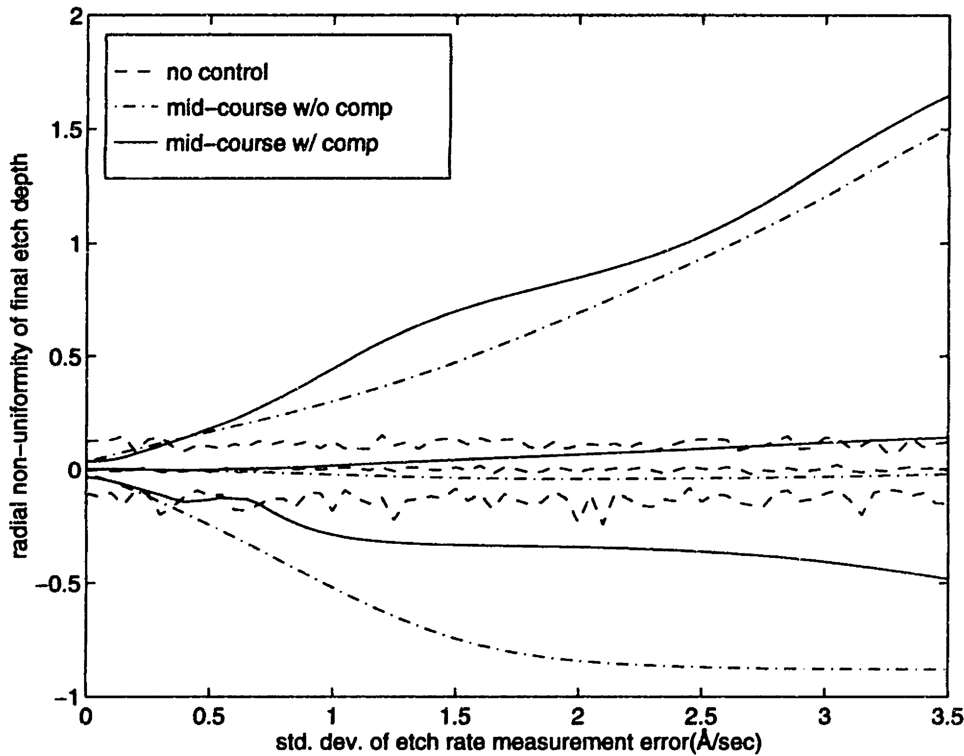


Figure 3-15: RN of total etch for 2-mid-course control w/ optimal start as a function of etch rate estimation error (showing means, upper bounds, and lower bounds)



3.6 Analysis of etch rate estimation for mid-course correction control

Based on the simulation results from Section 3.3, it was found that the accuracy of etch rate estimation is a function of the number of cycles of signals available, the amount of noise in the signal, and the sampling rate. The usefulness of this etch rate information for real-time control is determined by its accuracy, as illustrated in the last section. If the estimates are not reliable, static process optimization may be more appropriate. In this section, an example justifying mid-course correction control for achievable etch rate estimates is presented.

Consider the etch of a polysilicon film over silicon oxide similar to the case in the last section. Assume that the polysilicon film thickness is 3500Å, and the signal wavelength for full wafer interferometry is 4780Å. The film thickness etched off corresponding to a cycle of the interferometric signal is about 545Å. As mentioned in Section 2.2, the sampling rate should be synchronized with the rotating magnetic field of the etcher and is at 0.72 sec.

Using the neural network model constructed in Section 3.5, the radial etch rate non-uniformity is zero at a magnetic field of 72 gauss, which corresponds to a nominal etch rate of approximately 70Å/sec. The total number of signal cycles available is:

$$\frac{3500\text{Å}}{544\text{Å/cycle}} = 6.4\ \text{cycles}. \qquad (3\text{-}31)$$

The total number of signals cycles available suggests a 1-mid-course control adjustment after about 3 to 3.5 cycles of signals are received. The number of samples per cycles is:

$$544\ \text{Å/cycle} \div 70\ \text{Å/sec} \div 0.72\text{sec/sample} = 10.8\ \text{sample/cycle}. \qquad (3\text{-}32)$$

Compared to the simulation performed in Section 3.3, this rate is close to the doubled sampling case. However, the estimation accuracy of using 3.3 signal cycles with a doubled sampling rate has not been specifically examined. By doubling the sample, the standard deviation in estimation error for 2.4-cycle-window decreases by about 20%, as shown from case A and case E of Table 3-1. Similar reduction in error is assumed for the case with 3.3 signal cycles. With a nominal etch rate of 70Å/second, the standard deviation in etch rate estimation (using FFT method) is approximately:

$$0.0116 \times 80\% \times 70\text{Å/sec} = 0.65\text{Å/sec}. \qquad (3\text{-}33)$$

If the etch rate started non-uniformly as shown in Figure 3-9, the 1-mid-course correction control could improve the final etch, at the etch rate estimation error level of 0.65Å/sec. However, if the etch started at the optimal setting as in Figure 3-10, a mid-course adjustment of control parameters would deteriorate the final performance, resulting in a radial etch depth non-uniformity of 0.2. Based on the definition of Equation 3-25, this corresponds to an etch depth variation of:

$$0.2 \times 3500\text{Å} \div 50 = 14\text{Å}. \qquad (3\text{-}34)$$

If this non-uniformity can be tolerated, 1-mid-course correction control can still be applied since the etch will be adjusted in case of a process drift.

Mid-course control adjustment will be more fruitful if the error in etch rate estimation can be reduced. Its application also depends on the variations in the control parameters. In addition, a no-action zone can be defined such that the control parameters are not adjusted if the etch rate non-uniformity is within some prescribed values.

3.7 Summary

Etch rate estimation from interferometric signals is becoming more important if the etching process is to be automated and improved. In this chapter, four different etch rate estimation methods are first examined that are applicable to the signals generated by *in-situ* full wafer interferometry. It is shown that despite its simplicity, a method based on the Fourier transform is robust and efficient. Although the estimates can be improved by increasing the number of cycles of signal, the number of cycles is limited by the thickness and properties of the films to be etched. On the other hand, it is also shown that increasing the sampling rate is another alternative to achieve better estimates. By using the Cramer-Rao lower bound, a measure of the possible improvement in estimation accuracy can be determined. Experimental data verify the feasibility of obtaining etch rates while the control parameters are modified during a process run.

Finally, the potential application and accuracy requirements for etch rate estimation to be used in a feedback control scheme have been explored, indicating that sufficient estimation accuracy exists for future use in run-by-run and real-time control, in particular, when a sufficient number of signal cycles are available. However, for the etch of very thin films, mid-course estimation of etch rates for process adjustment may not be feasible and the chemistry-switching type of etch must be used; process control for such an etch will be described in Chapter 5.

To constitute a good etch, both uniformity in etch rate and minimization of excess overetch are necessary. While the extraction of etch rate and its application has been examined in this chapter, detection of the clearing of the etching film will be considered in Chapter 4.

Chapter 4

Endpoint detection of plasma etching

4.1 Introduction

In plasma etching, the detection of film clearing is important for consistent performance. Even if an etch rate is uniform, excess amount of overetch may increase undercut of the etching layer and may damage the underlying film structure. An efficient endpoint detection scheme reduces over-etching, and the time to endpoint can be used for process improvement via run-by-run control [7]. Various endpoint detection algorithms using laser or plasma emission signals have been reported [1,2,3,4,59]. Power has also been monitored for endpoint detection [46]; a general introduction to endpoint detection can be found in [43]. In general, these methods all focus on obtaining a single endpoint signal that represents the overall etch. Using full wafer interferometry, it is possible to perform multiple endpoint detection in real-time at many sites over the wafer. Compared with laser interferometric signals, the signals generated from plasma emission are often contaminated with high levels of noise, and the average signal intensity may vary from run to run. In this chapter, the use of phase information for endpoint detection is suggested to reduce false endpoint signals. Two endpoint detection methods are presented. For the model deviation approach, a simple model is proposed that is appropriate for interferometric signals with minimal absorption rate. The second algorithm is applicable for etching films when the refractive index of the etching film is higher than that of the film immediately underneath it, and can accommodate a large optical absorption rate.

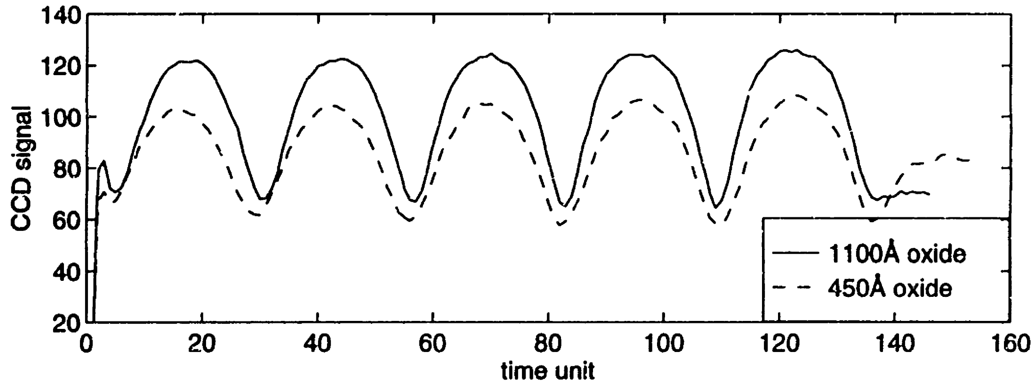
In the following sections, the two proposed endpoint detection algorithms will be described, and verification with experimental data presented.

4.2 Model deviation endpoint algorithm

Interferometric signals resulting from film etch are cyclical, as shown in Figure 4-1. For etching a specific film structure, the number of signal cycles is constant from one run to

another. In typical endpoint detection practice, one desires to focus attention on the last stage of the etch (to minimize early "false alarms"), and thus "cycle counting" by observing the maxima or minima of the signal is often employed. A more robust approach is proposed here which uses the phase of the dominant frequency of the signals, as obtained by a Fast Fourier transform. The number of signal cycles is then counted by noting the phase changes from $-\pi$ to π. From previous simulations, we find that approximately two cycles of signal are sufficient for a relatively accurate determination of the phase. In Chapter 5, the actual phase will be used to predict film thickness which could then also be used as a measure for endpoint prediction. In the next subsection, a new model for the interferometric signal will be proposed, followed by a description of the model deviation endpoint algorithm.

Figure 4-1: 7534Å signals for etching 5000Å of bare polysilicon on oxide



4.2.1 New model for interferometric signal

Since the interferometric signals for etching a non-absorptive film are periodic, endpoir. can be detected if the signal deviates tremendously from its periodic pattern. As suggested by [42], some potential models for the interferometric signals are:

$$y_t = A + B\cos(Ct + D) \qquad (4-1a)$$

$$y_t = A + B\cos(Ct + D) + Et \qquad (4-1b)$$

$$y_t = (A + B\cos(Ct + D))/(1 + E\cos(Ct + D)) \qquad (4-1c)$$

where y_t is the fitted value of the signal obtained from the CCD camera. These three functions are applicable for modeling signals with minimal absorption rate (i.e., small imaginary component of the refractive index). In general, the coefficients for models $1a$-c are deter-

mined initially by Fast Fourier transform, and then more accurately by non-linear regression methods. Despite their simplicity, models 4-1a and 4-1b can be used only for films with relatively low refractive index. On the other hand, fitting to model 4-1c is difficult and computationally inefficient for real-time use. Since interferometric signals for films with relatively low absorption rates are very close to sinusoids, they can be well approximated by the dominant frequency and its first harmonic. Hence, we propose to use the following model:

$$y_t = A + Bt + C\cos(Et + F) + D\cos(2Et + 2F). \tag{4-2}$$

Initial values for E and F can be obtained by applying a Fast Fourier transform to the signals and extracting the dominant frequency. The remaining initial values can then be determined by simple linear regression. The initial estimates of the parameters obtained in this way should be close to the overall least square solution. This solution can be refined as follows.

Define J as the Jacobian of the least square fitting problem:

$$J = \begin{bmatrix} \frac{\partial y_1}{\partial A} & \frac{\partial y_1}{\partial B} & \cdots & \frac{\partial y_1}{\partial F} \\ \frac{\partial y_2}{\partial A} & \frac{\partial y_2}{\partial B} & \cdots & \frac{\partial y_2}{\partial F} \\ \vdots & \vdots & \ddots & \\ \frac{\partial y_n}{\partial A} & \frac{\partial y_n}{\partial B} & & \frac{\partial y_n}{\partial F} \end{bmatrix}. \tag{4-3}$$

Using the model suggested in Equation 4-2, the Jacobian (J) can be computed relatively easily. Then, the least square solution can further be improved by performing a few (e.g., 2 or 3) Gauss-Newton iterations [21], i.e.

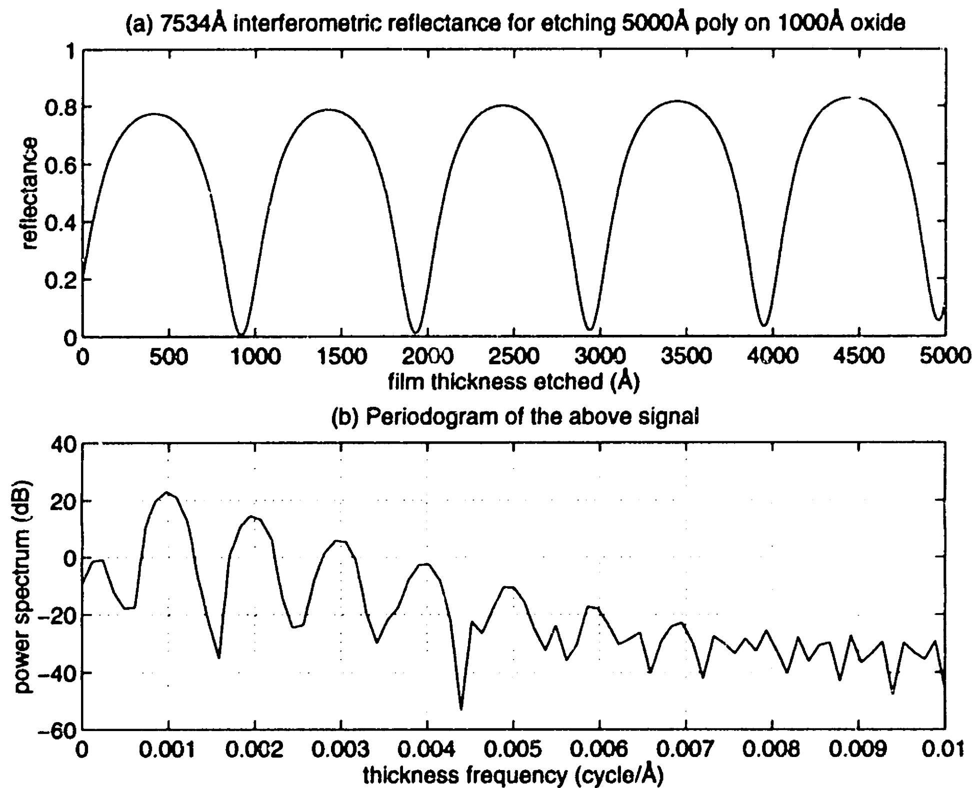
$$\alpha(new) = \alpha(old) - (J^T J)^{-1} J^{\perp}(Y - D), \tag{4-4}$$

where: $\alpha = \begin{bmatrix} A & B & C & D & E & F \end{bmatrix}^{\perp}$, Y and D are column vectors of y_t and the data d_t respectively, and $\perp$ is used to indicate vector transpose.

Figure 4-2a shows the 7534Å signal for etching 5000Å of polysilicon on 1000Å of

silicon oxide and Figure 4-2b shows the periodogram of the reflectance curve with its mean removed. From Fourier analysis, it is well know that a periodic signal can be approximated by a sum of sinusoids. In particular, if a signal is close to a sinusoid, only a small number of sinusoids is required to well-represent that signal. From Figure 4-2b, it can be seen that only two or three of the frequencies should be required for accurate approximation, since only three main frequencies are above 0 dB.

Figure 4-2: Spectrum of a non-absorptive interferometric signal



In the following, the reflectance curve is to be approximated with four models:

$$\text{A: } y_t = \frac{A + B\cos(Ct + D)}{1 + E\cos(Ct + D)} + Ft + G; \tag{4-5a}$$

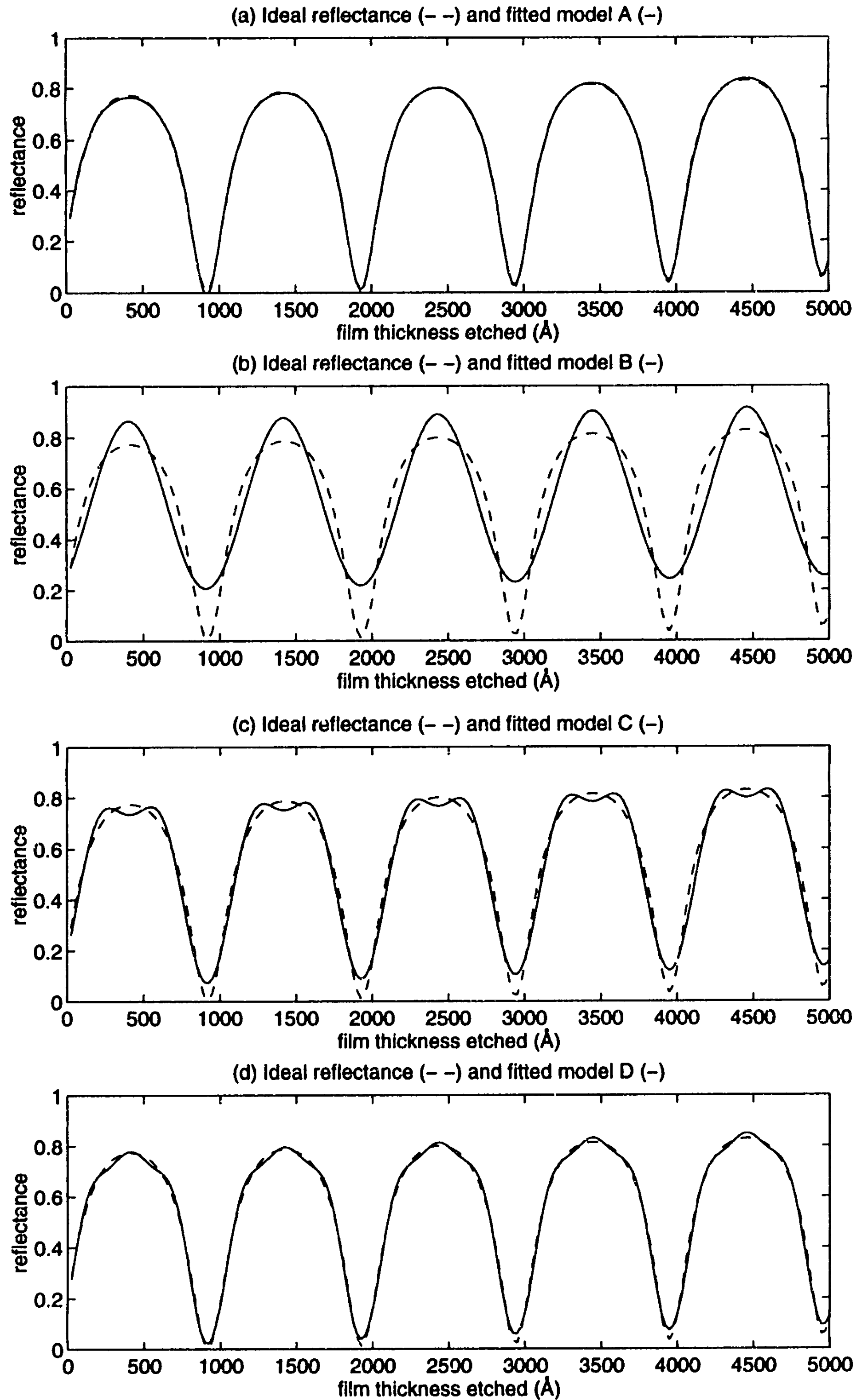
$$\text{B: } y_t = A + B\cos(Ct + D) + Et; \tag{4-5b}$$

$$\text{C: } y_t = A + Bt + C\cos(Et + F) + D\cos(2Et + 2F); \tag{4-5c}$$

$$\text{D: } y_t = A + Bt + C\cos(Et + F) + D\cos(2Et + 2F) + G\cos(3Et + 3F). \tag{4-5d}$$

To simplify the computation, only the reflectance at 25Å, 50Å, . . . , 5000Å will be fitted.

Figure 4-3: Different model fitting of interferometric reflectance for etching 5000Å of polysilicon on 1000Å of oxide



Figures 4-3 *a* to *d* show the original reflectance curve and the fitted models. Table 4-1 shows statistics regarding the fit of the models to the interferometric reflectance. The statistics are defined as follows:

$$SSTO = \sum_i (d_i - \bar{d})^2, \tag{4-6a}$$

$$SSE = \sum_i (d_i - \hat{d}_i)^2, \tag{4-6b}$$

$$R^2 = 1 - \frac{SSE}{SSTO}, \tag{4-6c}$$

$$R_a^2 = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE}{SSTO}. \tag{4-6d}$$

where d_i, $\bar{d}$, $\hat{d}_i$ are the reflectance data, the mean of the reflectance, and the fitted values of the reflectance respectively; $n = 200$ is the total number of reflectance data; and p is the number of parameters in the model. *SSTO* measures the magnitude of the variation about the mean of the signals, and R^2 measures the proportionate reduction of total variation due to model fitting. R_a^2 is sometimes preferred over R^2 because it adjusts for the different number of parameters in each model [54].

Since model A is derived from physics, it results in the best fitting. Model B uses only one cosine term in the model and so the error is the largest of all the models. Model C

Table 4-1: Statistics from the model fitting

	Model A	Model B	Model C	Model D
SSTO	12.8335	12.8335	12.8335	12.8335
SSE	0.0061	1.8777	0.2816	0.0472
R^2	0.9995	0.8537	0.9781	0.9963
R_a^2	0.9995	0.8491	0.9774	0.9962
# of coef.	7	5	6	7

is capable of capturing a large percentage of the function. Model D is still better in terms of the reduction in the fitting error, and its performance is close to the physically derived model A. However, it requires more computation because it has one more parameter than model C. When more accuracy is needed, model D is preferred. In the following, model C is considered because it should be sufficient in most cases and enables extremely fast computation.

4.2.2 Endpoint detection algorithm

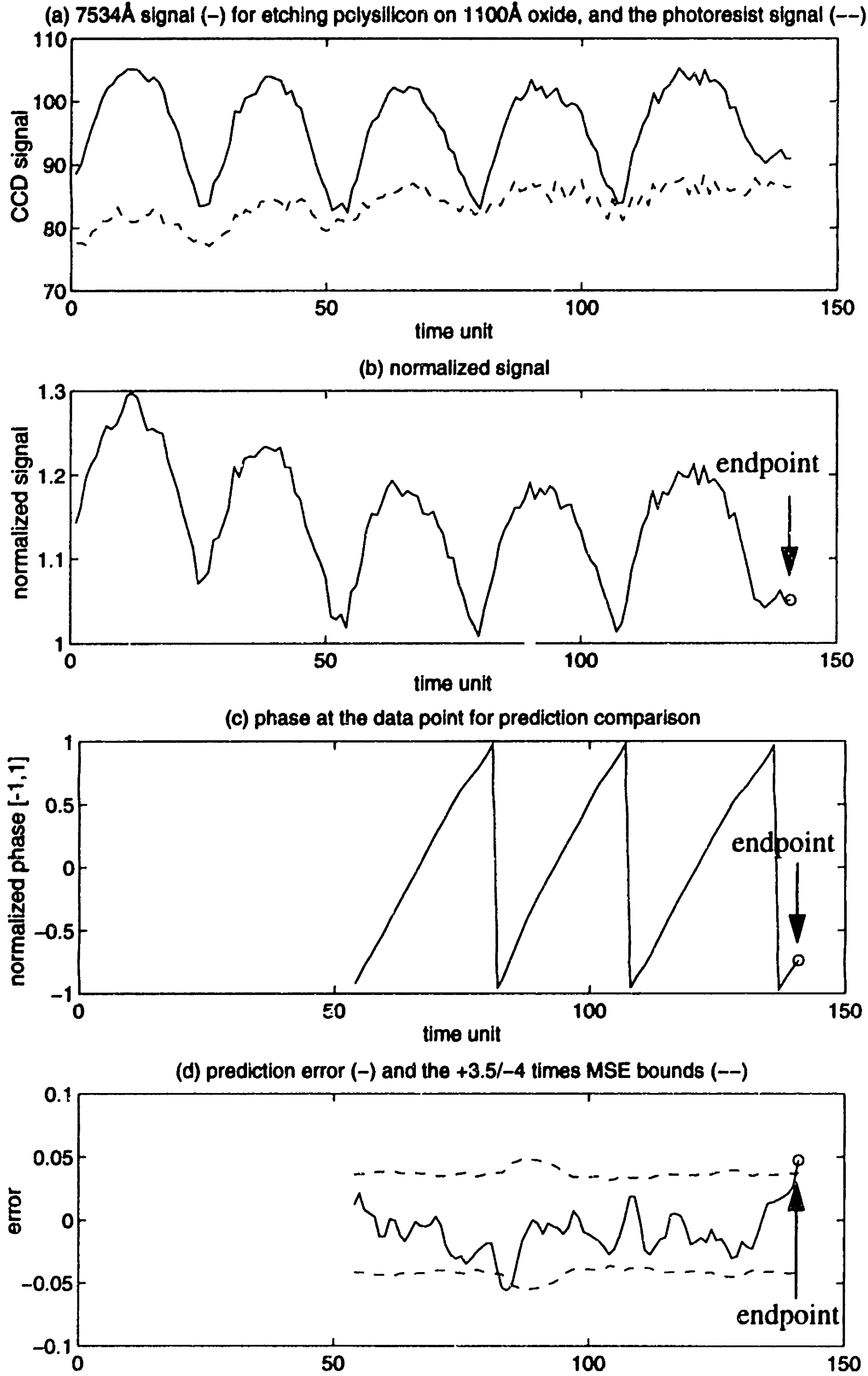
Given a set of data x_1 to $x_{n+ahead}$, x_1 to x_n are used for model building, while the deviation between $x_{n+ahead}$ and the fitted model is examined for endpoint. To reduce the effect of noise in data, the smoothed data at $x_{n+ahead}$ is preferred, i.e.

$$\hat{x}_{n+ahead} = \frac{x_{n+ahead-1}}{4} + \frac{x_{n+ahead}}{2} + \frac{x_{n+ahead+1}}{4} \qquad (4\text{-}7)$$

The use of a smoothed version may delay the detection of endpoint. However, the reduction of false alarms justifies this procedure.

Assuming that the error from fitting data to Equation 4-2 can be modeled as normal white noise, the mean of squared errors (*MSE*) is an estimate of the variance of the noise. The $\pm 3.5\sqrt{MSE}$ bounds should thus include 99.96% of the random error. If the observed error between $\hat{x}_{n+ahead}$ and the fitted model is outside the allowable error bounds, endpoint is determined to have been reached. To reduce the effect of rising overall plasma emission intensity near the end of an etch, the interferometric signal at the site of interest can be normalized by the signal at a non-etching site [42], or at a site with very slow etching rate, such as a large photoresist area.

Figure 4-4: Endpoint for etching 5000Å patterned poly on 1100Å oxide (Model Dev.)



The model deviation endpoint algorithm can be summarized as:

a) Smooth background signal (at an unetched site) with 3-time-point averaging;

b) Normalize the signal of interest by the smoothed background signal;

c) Perform Fast Fourier Transform (FFT) to obtain the dominant frequency and its phase;

d) Apply linear regression to determine the remaining parameters of Equation 4-2, and the allowable prediction bounds;

e) Improve parameter estimates with Gauss-Newton iterations.

d) If the required number of signal cycles have been received, and $\widehat{x}_{n+ahead}$ is outside the allowable error bounds, endpoint is detected. Otherwise, repeat from step (a) for new incoming data.

Two sets of experimental signals are examined with the suggested approach. These data were obtained by etching 5000Å of patterned polysilicon over 1100Å and 450Å of silicon oxide. The signal wavelength is selected to be at 7534Å for its low absorption rate. For the cases considered here, a moving window of about 2 cycles of data is used for signal modeling, and the value of *ahead* is 2. Figures 4-4*a,b,c,d* show the normalized signal, the estimated phase, the allowable error bounds, and the endpoint for the etch with 1100Å of underlying oxide. In Figure 4-5, only the normalized signal and the detected endpoints are shown for the etch with 450Å of underlying oxide. From these two cases, it can be observed that the model deviation endpoint algorithm works well for low absorptive film signals with moderate amount of noise. If the noise level is extremely high, then the endpoint detection algorithm may be less effective. For example, Figures 4-6*a,b,c* depict a 7534Å signal for etching patterned polysilicon on 250Å oxide. Both the original and the normalized signals contain a high level of noise. No endpoint was detected initially by the original model deviation endpoint algorithm, due to the high level of noise and the steep sloping trend of the signal. A rough endpoint was obtained by detrending the data within the 2-cycle data window with a linear polynomial before determining the initial estimates of *E* and *F*. Nevertheless, from the error plot of Figure 4-6*c* it can be seen that the algorithm is not as effective if the noise level is extremely high.

Figure 4-5: Normalized 7534Å signal and endpoint for etching patterned poly on 450Å oxide

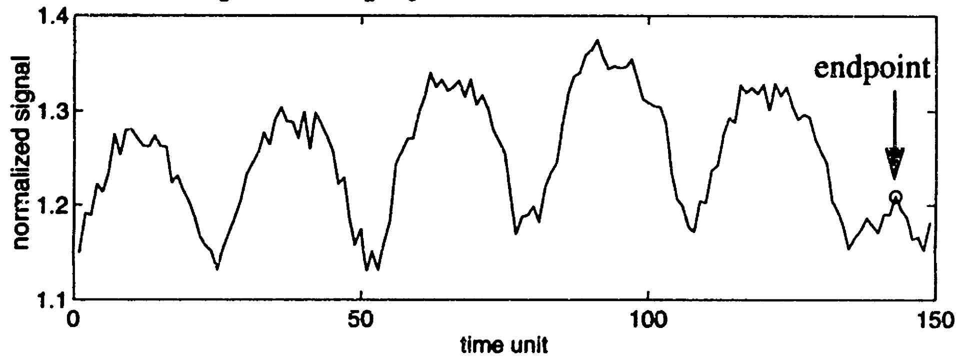
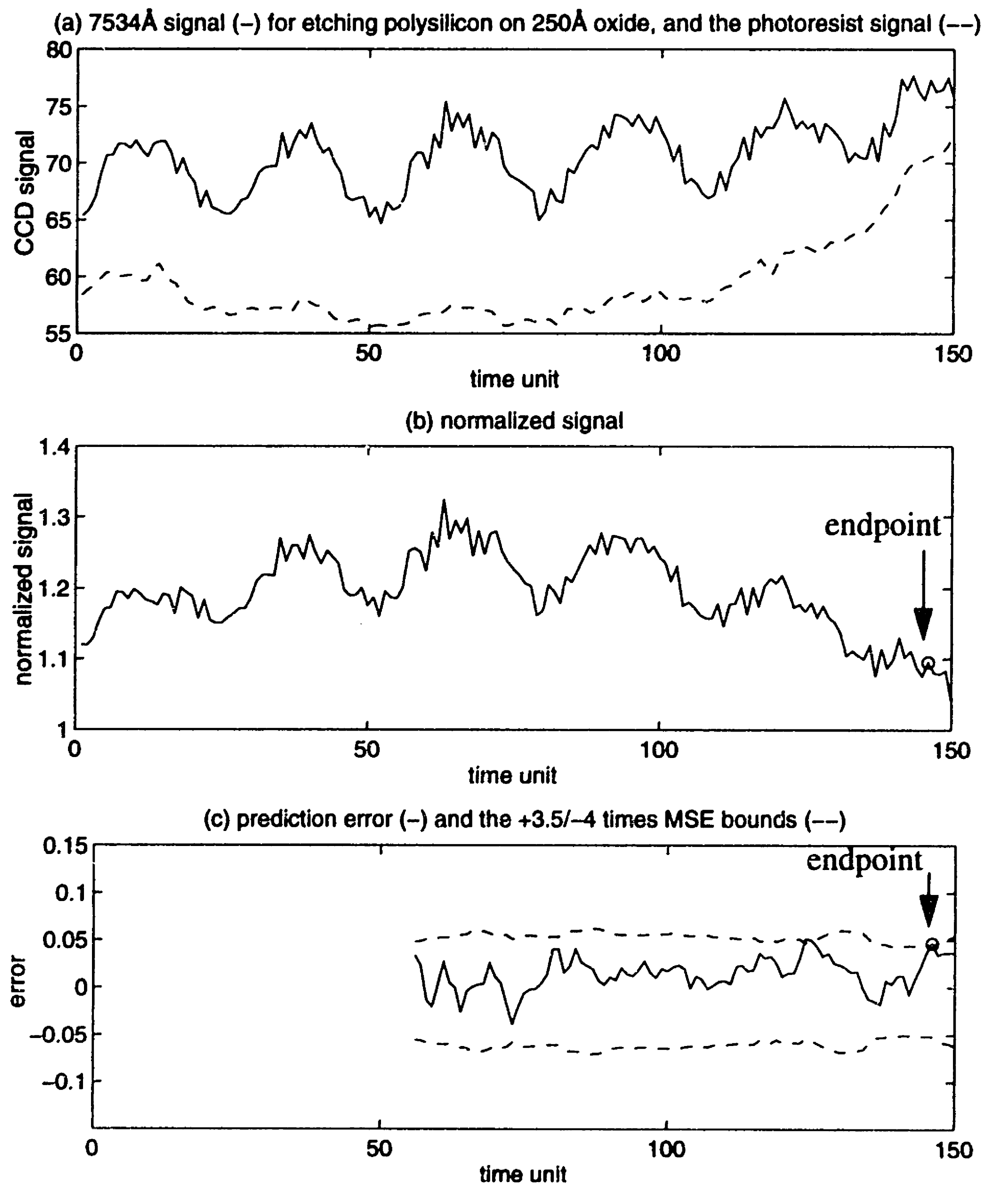


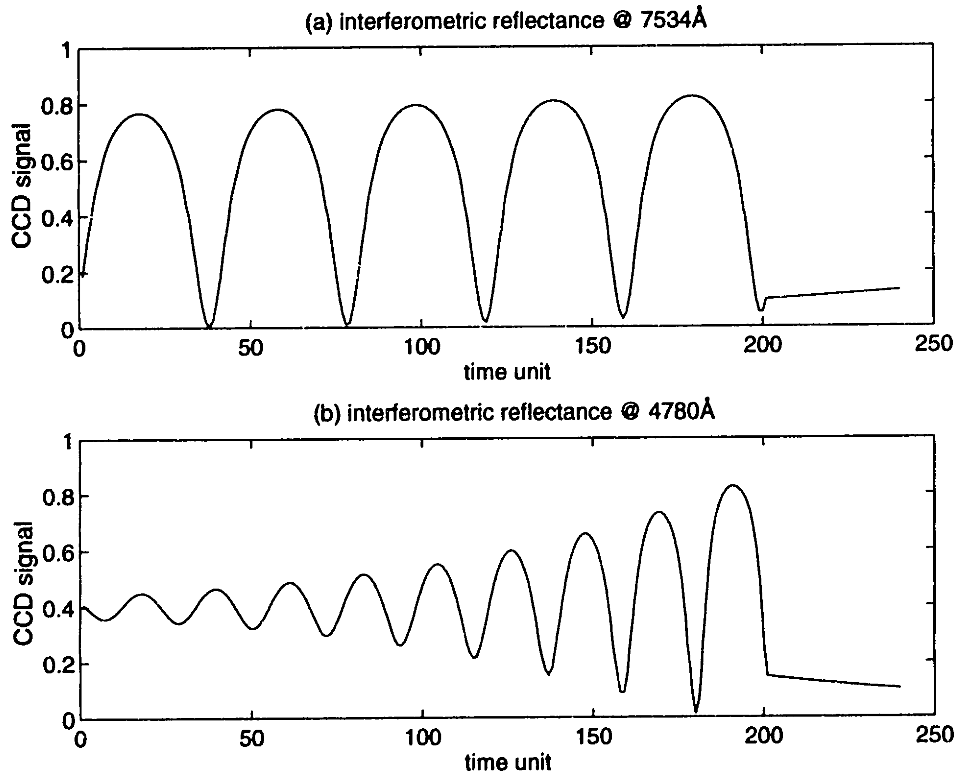
Figure 4-6: Endpoint for etch of 5000Å patterned poly on 250Å oxide (Model Dev.)



4.3 Slope detection endpoint algorithm

The model deviation endpoint algorithm in the previous section is proposed for signals with uniform amplitude envelopes over time. The suggested model is not well-suited for high absorption cases (where light is attenuated rapidly within the film), such as occurs for a short wavelength signal obtained from etching polysilicon. In this section, an endpoint detection algorithm is introduced for monitoring films having high refractive indices. Figures 4-7*a*,*b* show the theoretical reflectance of light from a wafer surface during an etch of 5000Å of polysilicon over 1000Å of silicon oxide, highlighting the effect of high absorption at the 4780Å wavelength. The reflectance of the signal will be constant after the endpoint if the oxide layer is not etched at all. In practice, despite the high selectivity of an etch the film immediately beneath the cleared layer could still be etched away slowly. However, the slow etch rate, combined with a lower refractive index, will usually guarantee a relatively flat signal pattern immediately after the etching layer is completely consumed. Here, similar to the model deviation endpoint algorithm, the total number of expected signal cycles is used to reduce false endpoint alarms.

Figure 4-7: Idea interferometric reflectance for etching 5000Å poly on 1000Å oxide



The algorithm can be summarized as:

a) Count the number of cycles of signals.
b) If the signal cycle equals the known number of cycles, stop etching when the signal flattens out.

The flattening of the signal can be detected by fitting local first order linear models for the most recent incoming signals, e.g., the last five data points. A number of criteria can be used to determine if the signal slope is flat. One simple method is to keep a record of the maximum slope, and regard the current signal as flat if it is less than a specified fraction of the maximum slope, e.g., one sixth.

Another precaution is also proposed to reduce false alarms. The maximum sum of squared errors during the determination of the slope is also recorded. Endpoint is assumed to occur with a small sum of squared errors, e.g., one quarter of the maximum value. This criterion is used to reduce false alarms at the minima of the signals where the sum of squared errors are usually quite large.

Figures 4-8 to 4-10 show the interferometric signals for etching 5000Å of polysilicon on 1100Å, 450Å and 250Å of silicon oxide respectively. These signals correspond to the patterned areas examined previously. However, instead of the normalized data, the original data is considered here because a low noise level is preferred for the current algorithm. If the noise level is too high, premature endpoint may be detected. Figure 4-11 shows a 4780Å signal for etching 5000Å of bare polysilicon over 1100Å of oxide. Since the sampling rate is rather low in this experiment, the slopes are determined from every consecutive four data points. It is found that the slope detection endpoint algorithm still performs well in this highly absorptive case.

On the other hand, if the etching layer has a lower refractive index than the underlying layer, the flattening of the signal may not occur. This is because the slower etch rate may be counteracted by the high refractive index of the underlying layer, and the slope of the signal will not be flat as a result. In cases where endpoint is not easily determined from interferometric signals, [42] suggests using the change of the overall emission level from a non-etching site to decide endpoint.

Figure 4-8: 7534Å signal and endpoint for patterned area of poly (Slope Detection)

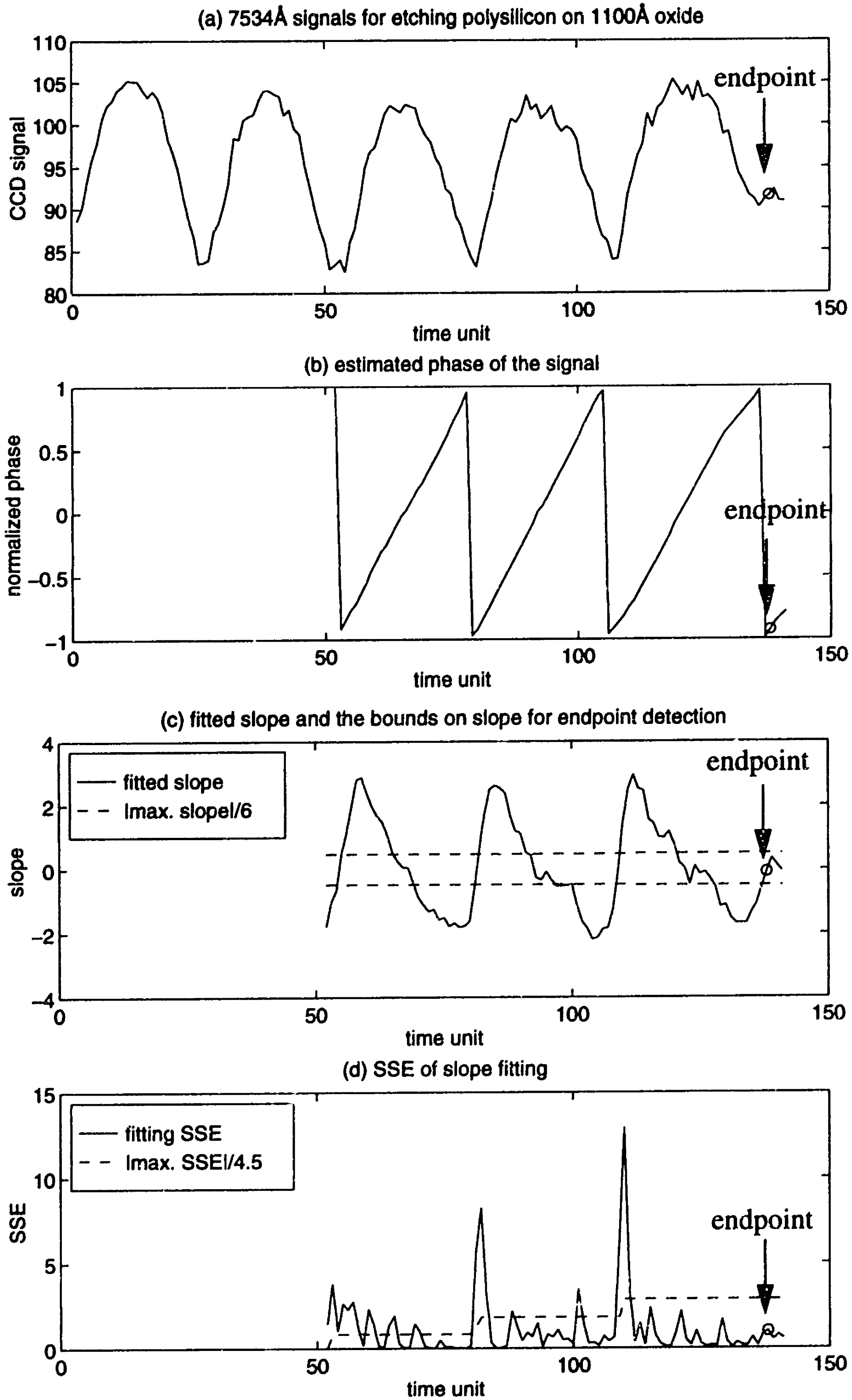


Figure 4-9: 7534Å signals for etching 5000Å patterned poly on 450Å oxide (Slope Detection)

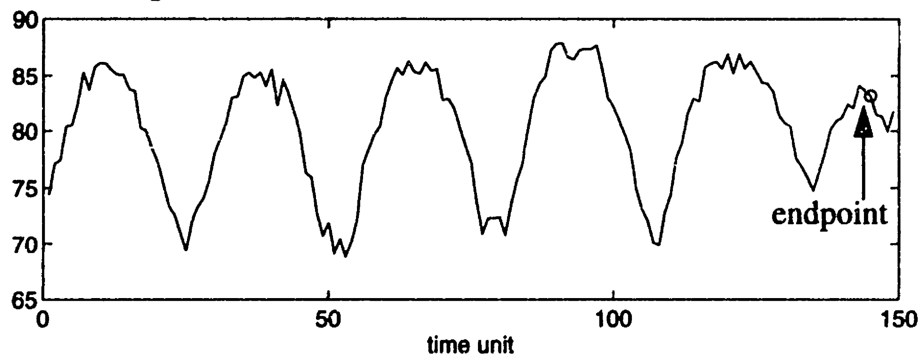


Figure 4-10: 7534Å signals for etching 5000Å patterned poly on 250Å oxide (Slope Detection)

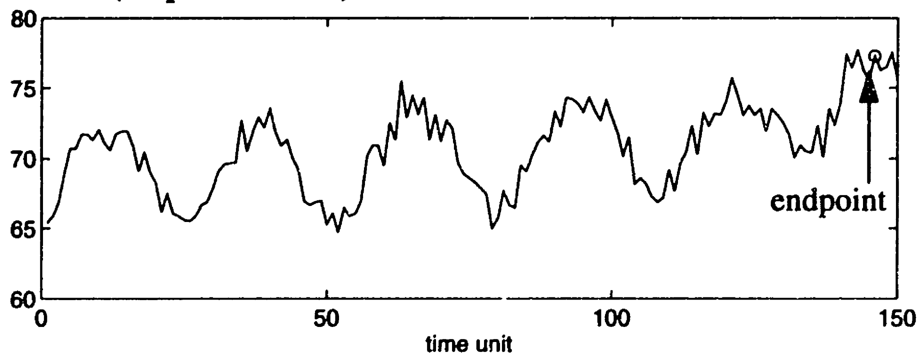
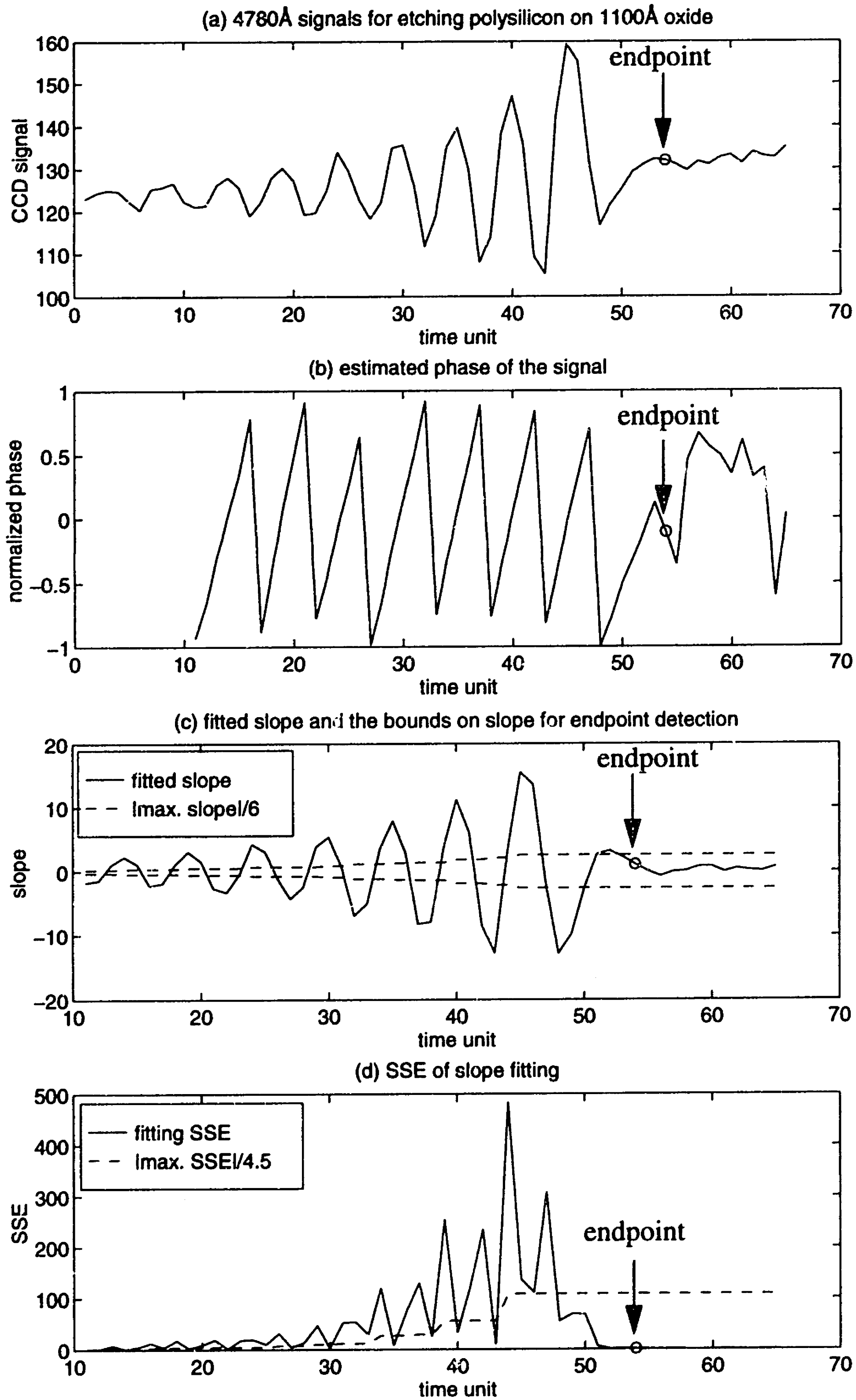


Figure 4-11: 4780Å signal and endpoint for an unpatterned site when etching 5000Å polysilicon on 1100Å oxide (Slope Detection)



4.4 Summary

Two endpoint detection algorithms for optical emission interferometric signals have been suggested. These algorithms can also be applied to laser inteferometry. They are especially important, however, for use in full wafer interferometry where the average emission signal varies from run to run and may suffer from substantial noise. In both algorithms, the phase of the signal is used to reduce false endpoint alarms. The model deviation endpoint algorithm is useful for detection in low-absorption cases. It is based on fitting a double-sinusoid curve to the interferometric signal, and endpoint is reached when the data deviate from the fitted model. The second algorithm is based on slope detection, and is applicable when etching a film with a higher refractive index than that of the layer immediate underneath it. From the experimental data considered, these two endpoint detection approaches are found to be effective for full wafer interferometry with polysilicon etch processes.

In cases where the noise in the sensor signal is exceedingly large, the interferometric signal obtained after endpoint could be very distorted and the detection of an endpoint signal is not feasible. In addition, for some etch processes, the etch chemistry is to be switched right before the film clears. Hence, algorithms are needed for estimating the approximate film thickness left, or the time-to-endpoint, and this topic will be examined in Chapter 5.

Chapter 5

Endpoint prediction via phase angle of interferometric signals

5.1 Introduction

In previous chapters, works relating to the determination of etch rates and detection of endpoint from optical emission interferometric signals have been analyzed. The emphasis in this chapter will be on film thickness estimation (or endpoint prediction). Currently, the etch of some material, such as polysilicon, involves multiple steps. For etching polysilicon, the first of a three-step etch is the descum step which removes the top thin oxide layer. The second step is the bulk etch. In general, the bulk etch is a timed process so that when a thin layer of the film is left, the chemicals are switched to a slower but more selective final etch. If the thickness of the film left for the final etch can be controlled more precisely, overetch of the underlying film could be minimized, which is particularly useful for incoming films with varying film thicknesses. In the next section, the use of phase angle as a measure of film thickness is discussed. In [25], the use of number of signal fringes for etched film thickness and endpoint has been suggested. However, the association of signal phase to extract film thickness has not been analyzed. Sensitivity of the measurement with respect to non-constant etch rate and variations in the film structure is discussed in Sections 5.3 and 5.4. Some practical issues for implementation will be discussed in Section 5.5. In Section 5.6, the proposed film thickness measure will be verified with experimental results. In Section 5.7 the use of signal phase for endpoint prediction will be examined, and in Section 5.8 the use of endpoint prediction versus endpoint detection will be compared. The conclusion will be given in the last section.

5.2 Film thickness estimation using interferometric phase information

Figure 5-1*a* shows the ideal reflectance of a 4780Å interferometric signal for etching 5000Å of polysilicon over 1000Å of silicon oxide at a constant etch rate. A short wave-

length signal is considered here because more cycles of signal are available for more accurate etch rate determination. The signal is sinusoidal-like with an exponentially growing amplitude. Following the definition of cosine, the phase is defined as zero at the local maxima, and it increases linearly. In Figure 5-1*b* and the subsequent analysis, the phase is normalized to the range (-1,1). In etching a fixed film structure, the pattern of the interferometric signal at any specific wavelength should be unique, i.e, with the same number of signal cycles and the same phase at endpoint. Since the thickness of film corresponding to a cycle of signal can be pre-determined, the amount of film unetched can be revealed by the number of signal cycles detected and the instantaneous phase estimates. The Fast Fourier Transform (FFT), the model matching, and the Kurmaresan-Prony (KP) methods are some of the methods that can be applied to calculate the phase. In this section, the performance of these methods will be examined with respect to different noise levels and data window sizes by simulations.

Figure 5-1: Interferometric reflectance and the corresponding phase angles

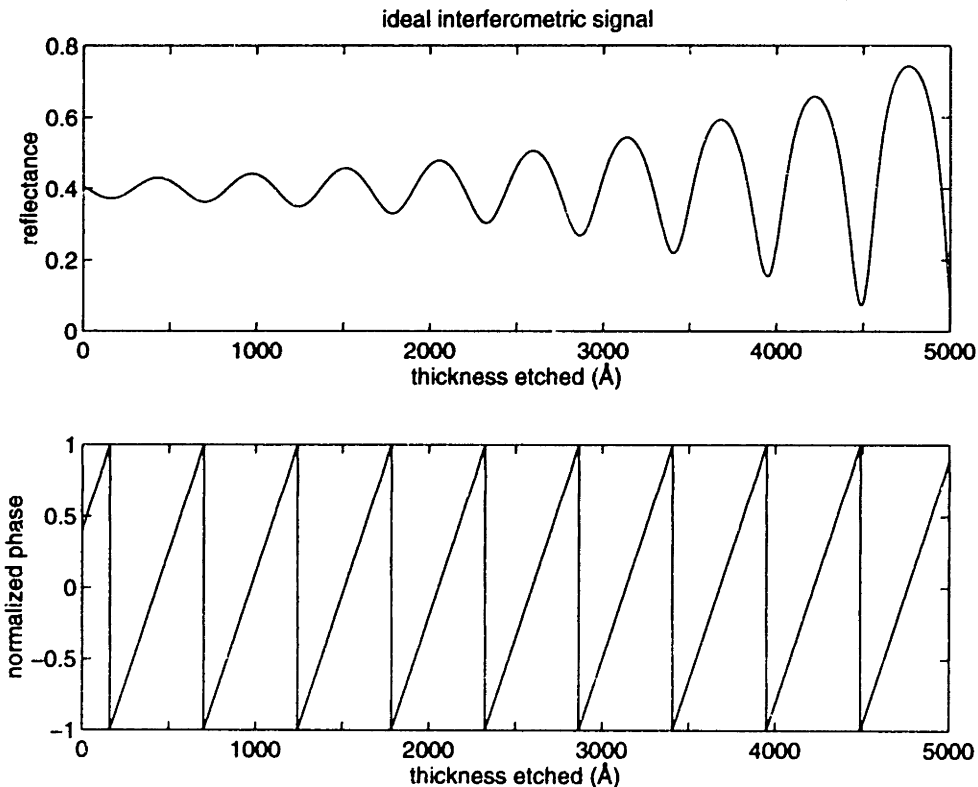
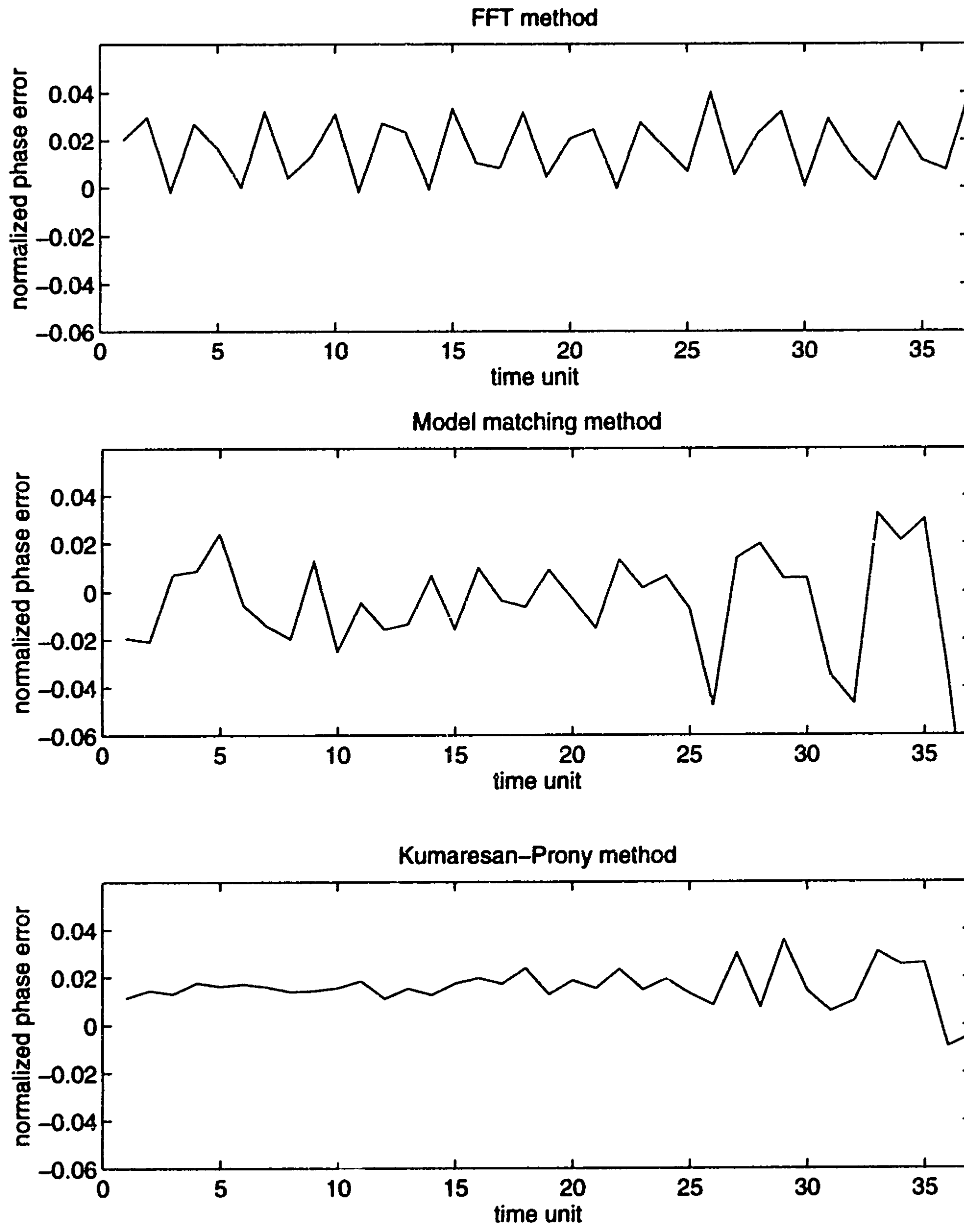


Figure 5-2: Phase estimation from interferometric signal via different methods



First, phases from the noiseless data from Figure 5-1*a* are obtained, and are shown in Figures 5-2*a*,*b*,*c*. The variance of errors in these estimates is due either to insufficient cycles of data used for determining each estimate or to numerical errors. For the FFT and the KP methods, there is a bias of approximately 0.01. In Figure 5-3, the same reflectance curve is plotted with a sine curve with similar periodicity and the same sine curve multiplied with an exponential envelope. The maxima of the normal sine curve occur at time units 427, 968, 1509 ... while those of the sine with exponential envelope are at time units 431, 972, 1513

... With a period of 541 time units, this shift corresponds to a normalized phase shift of $\frac{4}{541+2} = 0.015$, which is very similar to the bias found in Figures 5-2*a*,*c*. Therefore, it can be seen that the FFT and KP methods are more robust against exponential growth in signal, and are more suitable for general applications.

Figure 5-3: Reflectance of etching poly over 1000Å of oxide and the sinusoidal approximation

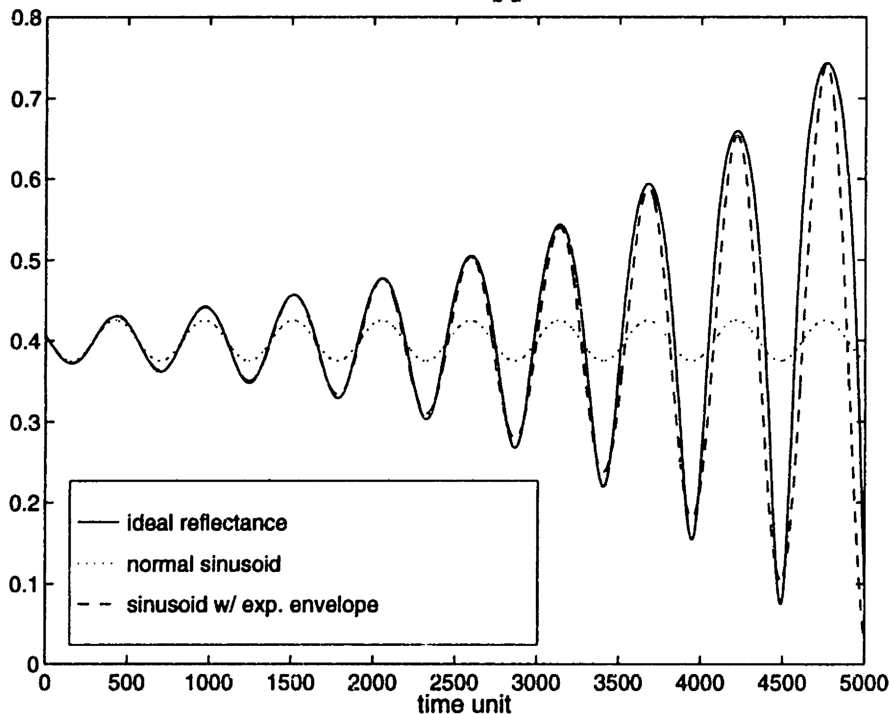


Table 5-1: Phase estimation error for interferometric signals

data window size (equivalent thickness)	2.5 cycles (1350Å)	3.5 cycles (1950Å)	4.5 cycles (2450Å)	
FFT method	0.0186 (0.0584)	0.0176 (0.0505)	0.0171 (0.0453)	← mean ← std. dev.
Model matching	-0.0082 (0.0548)	-0.0093 (0.0506)	-0.0108 (0.0453)	
K-P method	0.0174 (0.0613)	0.0163 (0.0507)	0.0165 (0.0465)	

Table 5-1 summarizes the results of the phase estimates when the interferometric signal has a signal-to-noise ratio of approximately 11dB. The noise added to the simulation is normal white noise. After the 0.01 constant bias for the FFT and KP method is taken into consideration, FFT is shown to give estimates comparable to the other two methods. In addition, the more cycles of data used for calculation, the smaller the variance.

Figure 5-4: Experimental data for film phase estimation

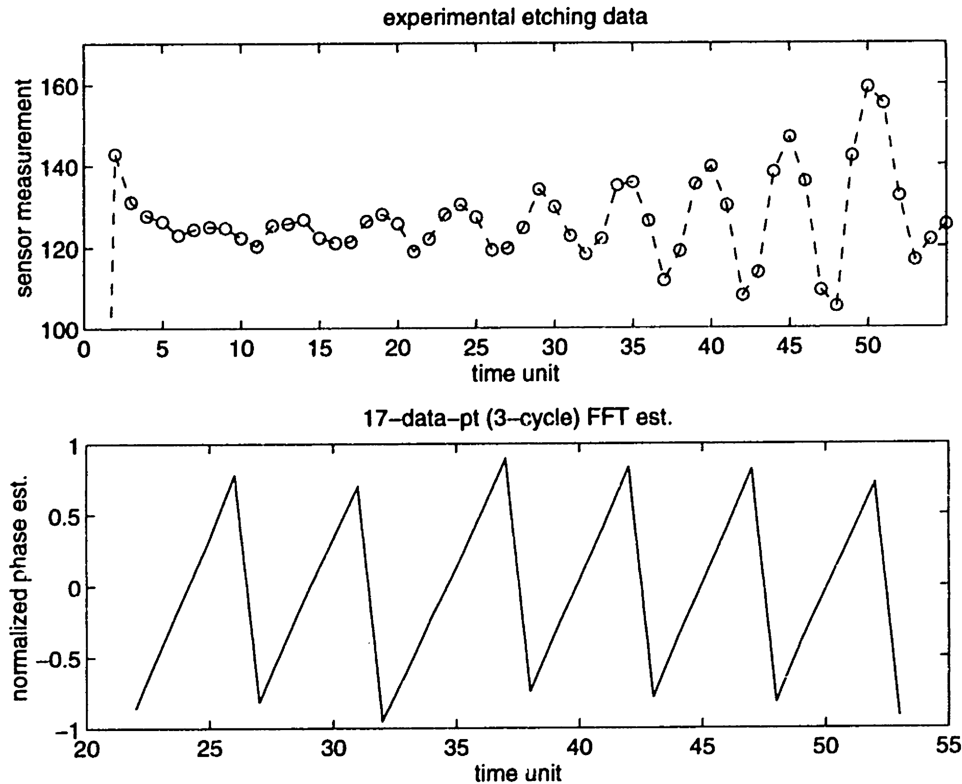
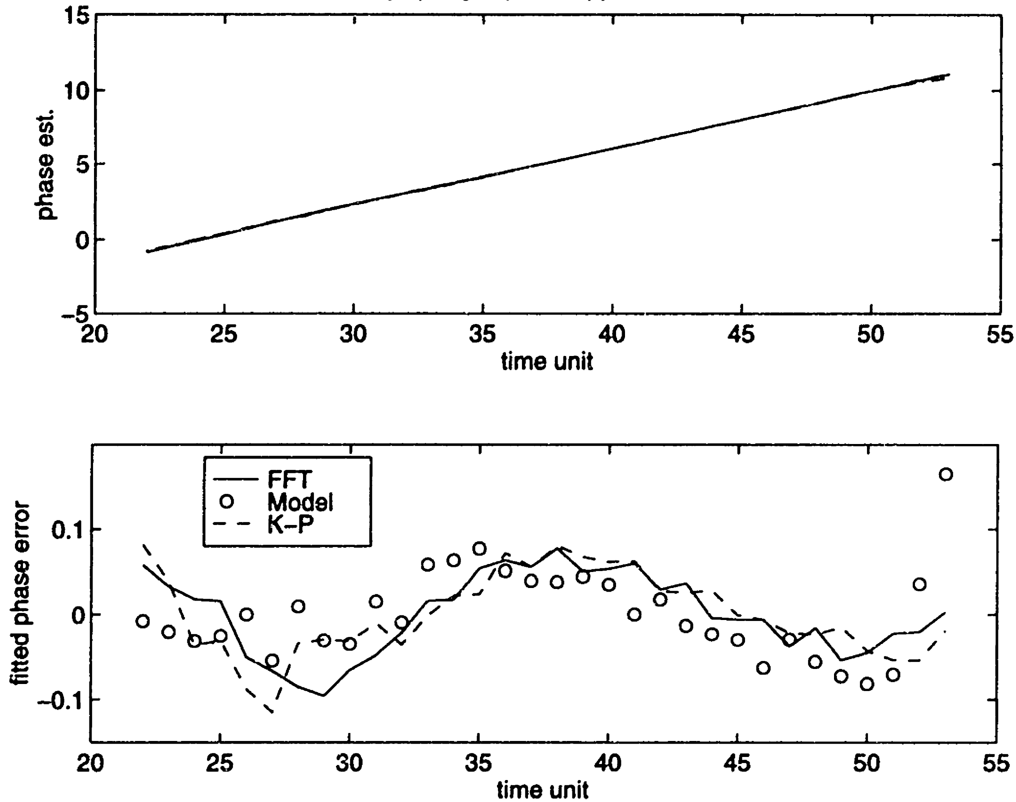


Figure 5-4*a* shows a set of experimental data resulting from etching 5000Å of polysilicon over 1000Å of silicon oxide. By using the FFT method, and using a 17-data (about 3 signal-cycle) moving window, the normalized phases of the signal at the end of the moving window are calculated as shown in Figure 5-4*b*. Assuming that the etch rate is relatively constant, a linear model is fitted for the unwrapped phase as shown in Figure 5-5*a*. Figure 5-5*b* show the differences between the fitted phase and the original values determined by three different methods. The maximum absolute phase difference is approximately 0.1, which corresponds to ±28Å error in thickness estimation for the 4780Å signal under consideration. In the next sections, the accuracy of using the phase of the signal as a measure of film thickness will be examined.

Figure 5-5: Fitting error of the unwrapped phase



5.3 Sensitivity analysis of endpoint phase angle with respect to etch rate drifting

In this section, the accuracy of obtaining the phase as a measurement of film thickness is examined when the etching rate is drifting. In particular, it is assumed here that the etch rate is increasing linearly. The case of a linear decreasing etch rate can be deduced directly. First, a preliminary analysis will be performed on a simple sinusoid to analyze the general change in phase angle estimation due to etch rate drifting. A graphical method will be used to assess the relative magnitude of phase error with respect to different data window lengths, different levels of etch rate drifting, and different signal wavelengths qualitatively. Then, phase estimation error for interferometric signal will be examined quantitatively by using simulations. In our simulations, the signals at a wavelength of 4800Å will be our focus because the high signal absorption rate facilitates a more general study. Signals at the wavelengths of 7534Å and 6100Å will also be examined for comparison.

5.3.1 Drifting rate effect and data window length effect

Consider Figure 5-6*a*. The dashed line corresponds to a sinusoid with constant angular frequency, and the solid line corresponds to a sinusoid with increasing angular frequency. Using a 20-data-point moving window, the argument (or phase) for the last data is estimated by least square method (or model matching method). In Figure 5-6b, the errors between the estimated and true phase angle are shown. The mean of estimated phases are smaller than the true phase. In the following paragraph, the bias in mean of phase estimation error will be examined via a graphical method qualitatively.

Figure 5-7 shows the phase of a sinusoid with increasing angular frequency for a single data window. The fitted phase, denoted by the dashed line, must have a constant slope larger than the minimum slope and smaller than the maximum slope of the drifting frequency phase. Also, it can be expected that the true phase and the fitted phase must be equal at some point within the data window. Figure 5-7 then shows that at both ends of the data

Figure 5-6: Example of phase error for a drifting sinusoid

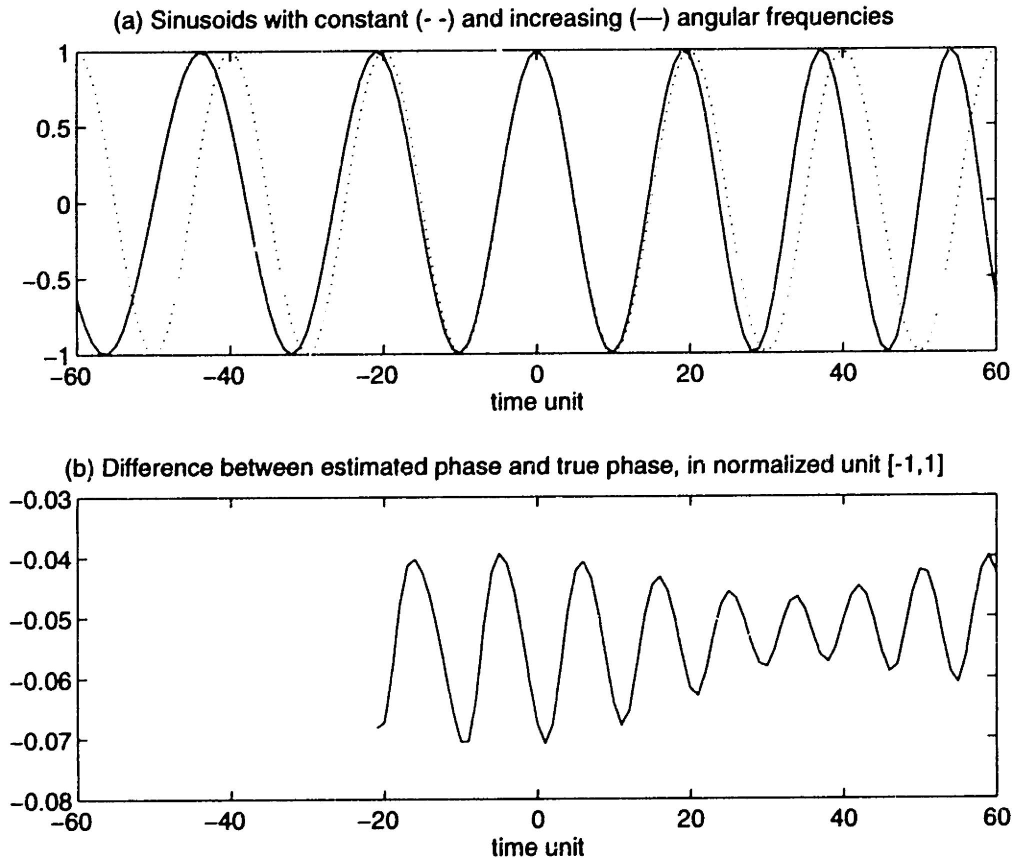


Figure 5-7: Qualitative analysis of fitted phase

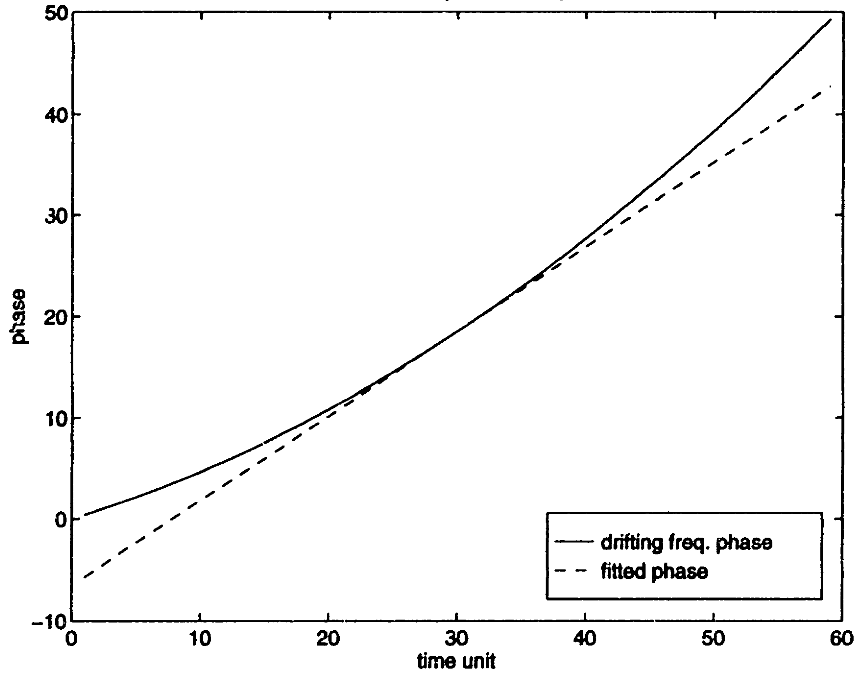
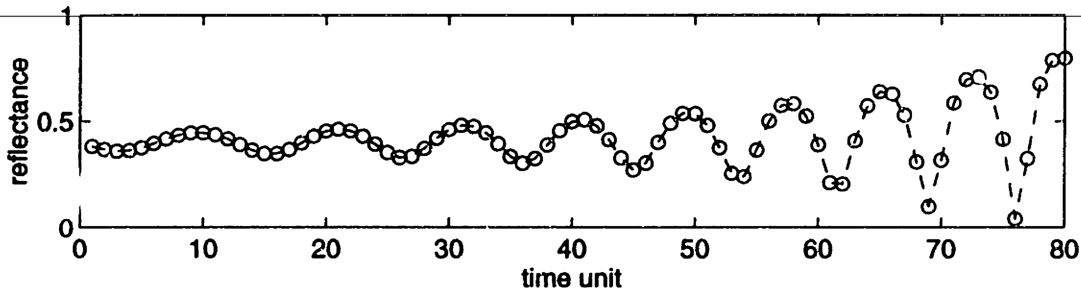


Figure 5-8: 4800Å reflectance for etching 5000Å poly on 1000Å oxide (etch rate increasing)

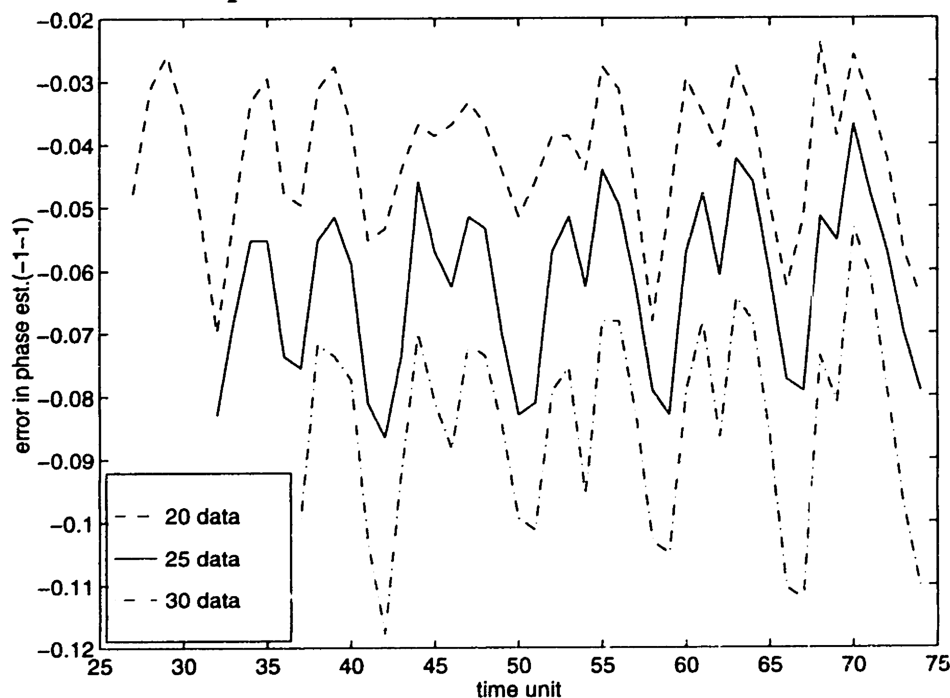


window, the fitted phase must be smaller than the true phase. It also suggests that the longer the data window, the larger the deviation; and the larger the frequency drifting rate, the larger the deviation also. Since the interferometric signals can be decomposed into a sum of sinusoids using Fourier analysis, such results are also expected for the interferometric signals. However, this simple graphical method does not show the preference of any particular signal wavelengths. Similarly, for a decreasingly drifting etch rate, the true phase in Figure 5-7 will be curving in the opposite direction, i.e., with a negative second derivative. Hence, the fitted phase is expected to be larger than the true phase.

Figure 5-8 shows a 4800Å wavelength interferometric reflectance for etching 5000Å

of polysilicon over 1000Å of silicon oxide where the etch rate is increasing. Model matching is used to fit a constant etch rate model to the data to determine the phase. Figure 5-9 shows that the phase estimated by using different numbers of data cycles will lead to different means of errors. As predicted by the simple sinusoidal model, the mean error is negative and it increases with the length of the data window.

Figure 5-9: Normalized phase estimation error with different numbers of data cycles



Figures 5-10*a*,*b*,*c* show three different 4800Å interferometric reflectances. The etch rates accelerate at different rates. Figure 5-11*a* shows the error in phase estimates using a fixed length of data window. It can be seen that the faster the etch accelerates, the larger the phase error. However, the phase error is both due to different etch accelerations as well as different numbers of data cycles included in the fixed-length data window for phase estimation. In order to decouple the results from both effects, Figure 5-11*b* shows the phase estimation error at the selected sites, which are numbered in Figure 5-10*a*,*b*,*c*. For each selected site, approximately 2.5 cycles of signal are used for phase determination using the model matching method. It can be seen that although different etch accelerations may produce different error means, the data length effect is more responsible for magnifying the error.

Figure 5-10: 4800Å signal for etching polysilicon over oxide (etch accelerates at different rates)

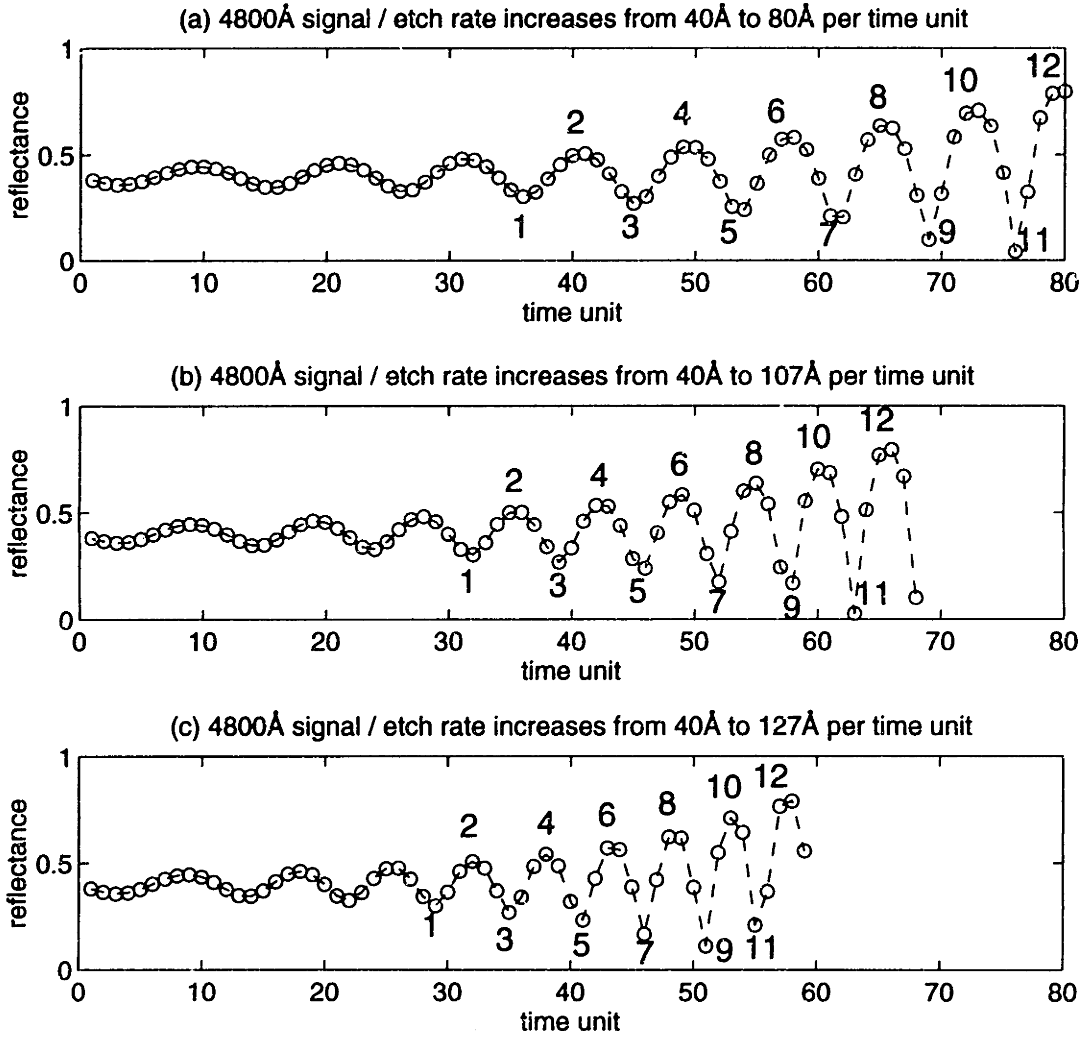
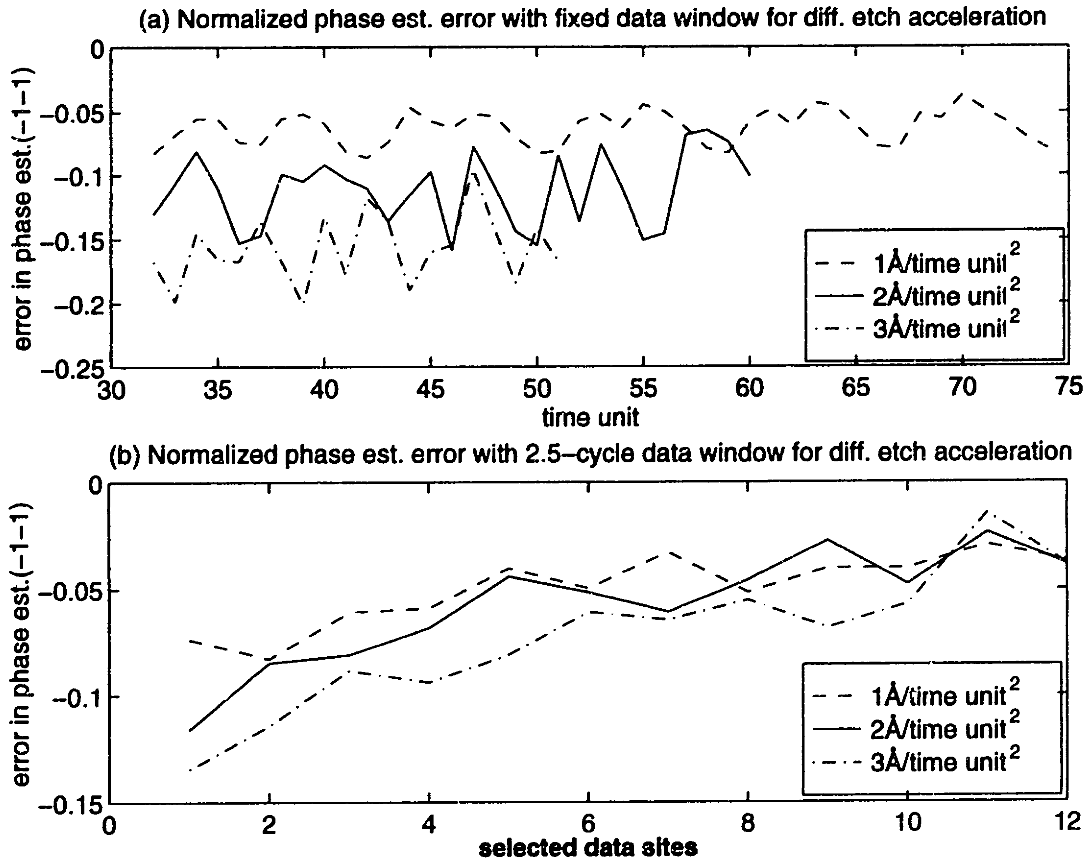


Figure 5-11: Phase estimation error due to different etch acceleration



5.3.2 Signal wavelength selection

In etch rate estimation, it was shown that short wavelength signals are preferred because they provide more cycles of signals which will improve the accuracy of the estimates. In this section, phase estimation using different wavelengths will be examined. Figure 5-12*a*,*b*,*c* shows the reflectance for etching 5000Å of polysilicon over 1000Å of oxide for the signals at 7534Å, 6100Å, and 4800Å respectively for etches with the same drifting etch rate. For the 7534Å and 6100Å signals, the polysilicon layer is minimally absorptive. However, the film is very absorptive for the 4800Å signal. Therefore, the reflectance for a 4800Å signal without signal absorption is also considered, which is shown in Figure 5-12*d* so the effect due to the high absorption rate can be decoupled. In all these cases, the etch rate accelerations are the same. Model matching method will first be applied to examine the phase estimation error. Then, the use of Fast Fourier Transform method will be considered.

Figure 5-12: Reflectance for etching 5000Å poly on 1000Å oxide at different wavelengths

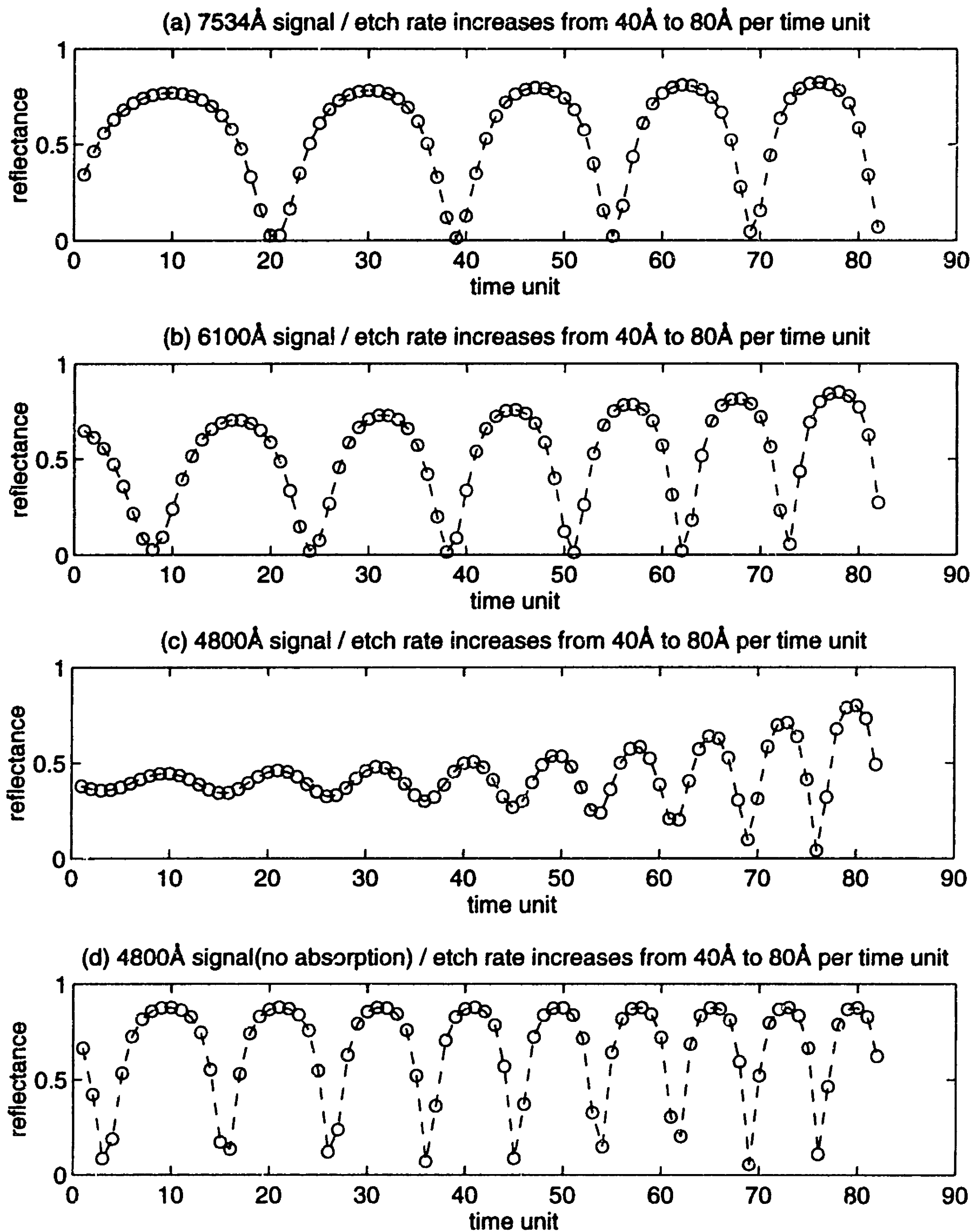


Figure 5-13*a*,*b* shows the error of phase estimates determined using different signal wavelengths. Since the phases for different wavelengths correspond to different film thicknesses, they must be converted back to the actual thickness. The mean and standard deviation of the thickness estimation error are shown in Table 5-2. It can be seen that the cases with minimal signal absorption give very close means of error. The standard deviations of

the error is smaller for the short wavelength signals because more cycles of signals are included for a fixed data window. The short wavelength signal with high absorption rate gives the curious result of having the smallest mean of error and the smallest error standard deviation. This may partly result from the minimization of mean square data fitting error at the high amplitude end of the data window, and will be examined in the next section. Since the FFT method is more robust against exponential growths in signal amplitude, as shown

Figure 5-13: Phase estimation error using model matching method

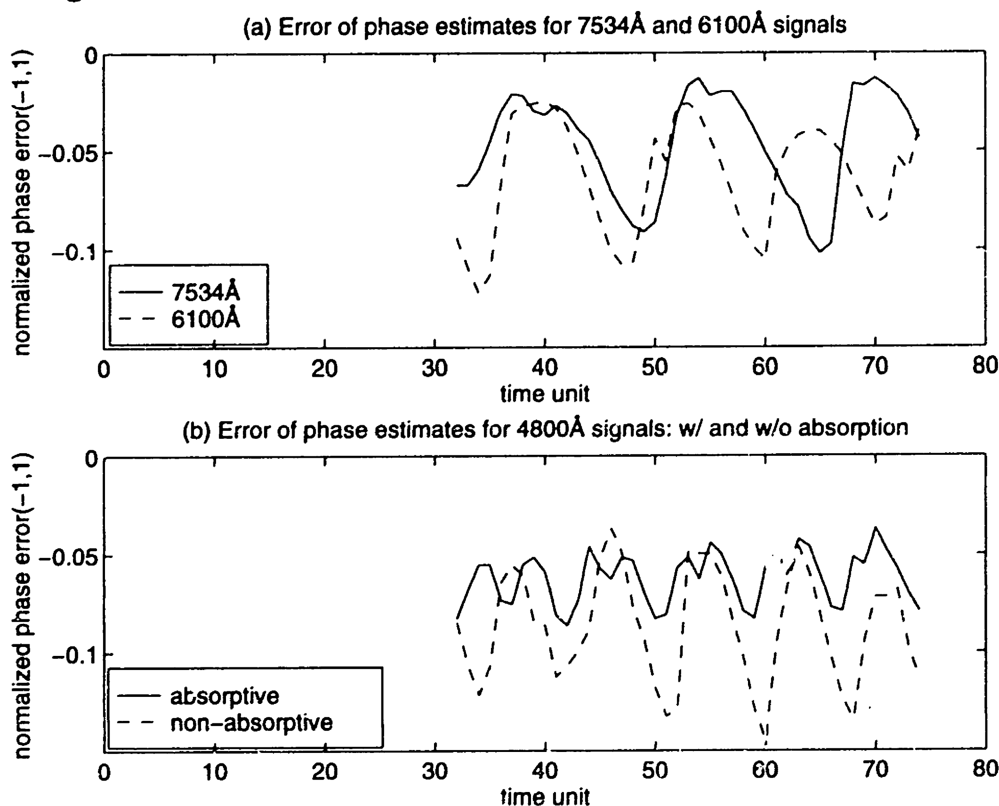


Table 5-2: Phase estimation error for different signal wavelengths (model matching)

signal wavelength (Å)	mean of thickness error (Å)	std. dev. of thickness error (Å)	film etched in one signal cycle (Å)
7534	-23.51	13.56	1010
6100	-24.58	11.21	777
4800(w/ abs.)	-16.93	3.64	544
4800(w/o abs.)	-23.76	7.78	544

in Section 5-2, the phase estimation is repeated with the FFT method and the results are shown in Figure 5-14*a,b*. The corresponding errors in thickness estimates are listed in Table 5-3. It can be seen that the mean of thickness estimation error is the same for all signals. Similarly, the error variance for longer wavelength signals are larger because fewer cycles of signals are included for a fixed data window.

From the simulations, it was shown that although the means of thickness prediction error may be the same for all signal wavelengths, shorter wavelengths are preferred because the error variances are smaller. In particular, estimation with more cycles of data are more robust when the data are contaminated with sensor errors. For a short wavelength signal with a high absorption rate, the model matching method seems to provide more accurate estimates; however, the number of parameters that need to be determined may not make it as widely applicable as the Fast Fourier Transform method.

Figure 5-14: Phase estimation error using FFT method

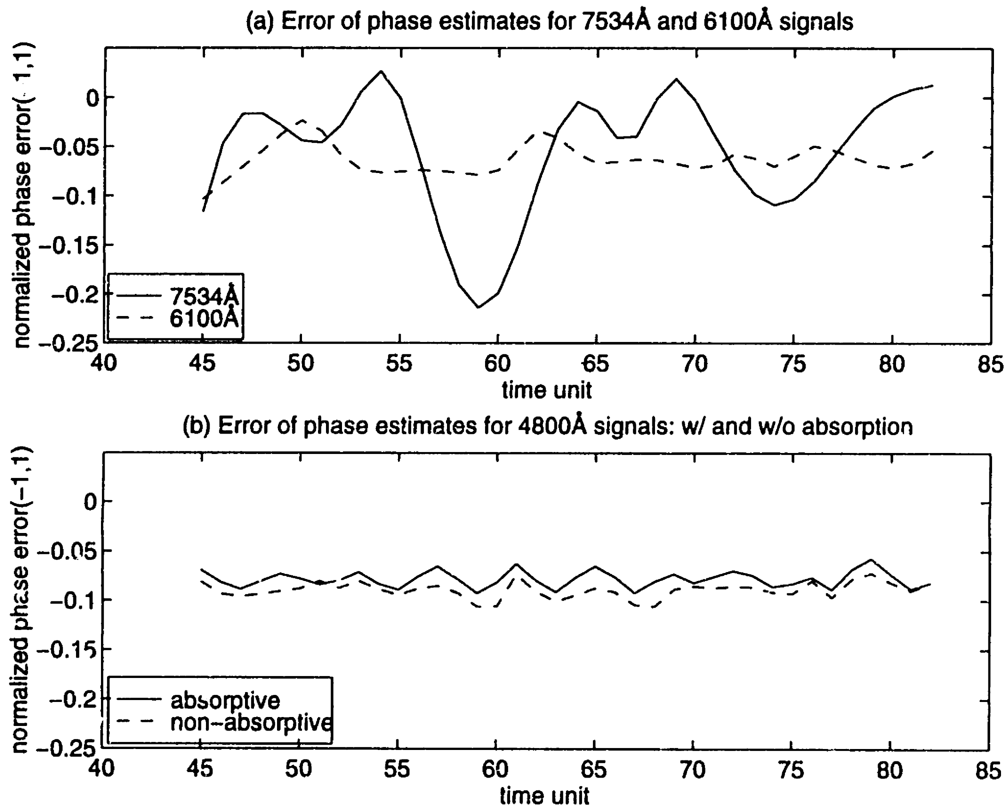


Table 5-3: Phase estimation error for different signal wavelengths (FFT)

signal wavelength (Å)	mean of thickness error (Å)	std. dev. of thickness error (Å)
7534	-27.30	31.26
6100	-24.48	6.05
4800(w/ abs.)	-21.24	2.32
4800(w/o abs.)	-24.30	2.26

5.3.3 Signal absorption rate on least square estimation of phase

In using model matching (or least square method) for fitting data, if the data values are doubled, the variance of the fitting error will be quadrupled. Therefore, if part of the data and the fitting model is multiplied by an exponential envelope, the effect is like increasing the magnitude of the error where the multiplier is larger. This is equivalent to a weighted least square fitting. In an extreme case, if the multipliers for various data differ greatly, the result will be similar to neglecting the data with small multipliers. Therefore, using the simple graphical analysis as in Section 5.3.1, the fitted phase should be modified as shown by the arrows in Figure 5-15 because the first part of the signals, which are associated with smaller multipliers (larger attenuation), have reduced contribution to the phase fitting. Hence, it is shown that the estimated phase at the end of the data window, via least square fitting, should be more accurate for the absorptive films than for the non-absorptive film. On the other hand, the error for the estimated phase at the beginning of the data window should be larger for the absorptive case.

In Section 5.3.2, errors of phase estimates at the end of a moving data window were examined. Using the same set of data, Figures 5-16*a*,*b* show the errors in phase estimate at the beginning of the moving data window, which has a length of 25 time units. In order to compare the errors between different wavelength signals, the mean and standard deviation of the errors in thickness estimation are shown in Table 5-4. It is found that the means of the phase errors for the non-absorptive cases remains relatively unchanged. However, the mean error of 4800Å absorptive case increases and is the largest, as expected.

Figure 5-15: Qualitative analysis of fitted phase for absorptive and non-absorptive signals

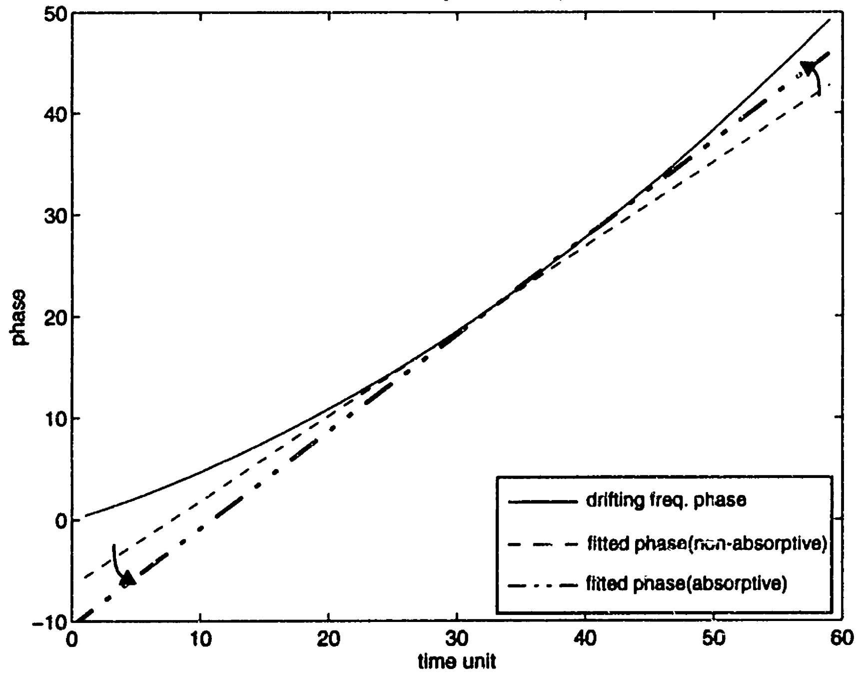


Figure 5-16: Phase estimation error for data window beginning (model matching)

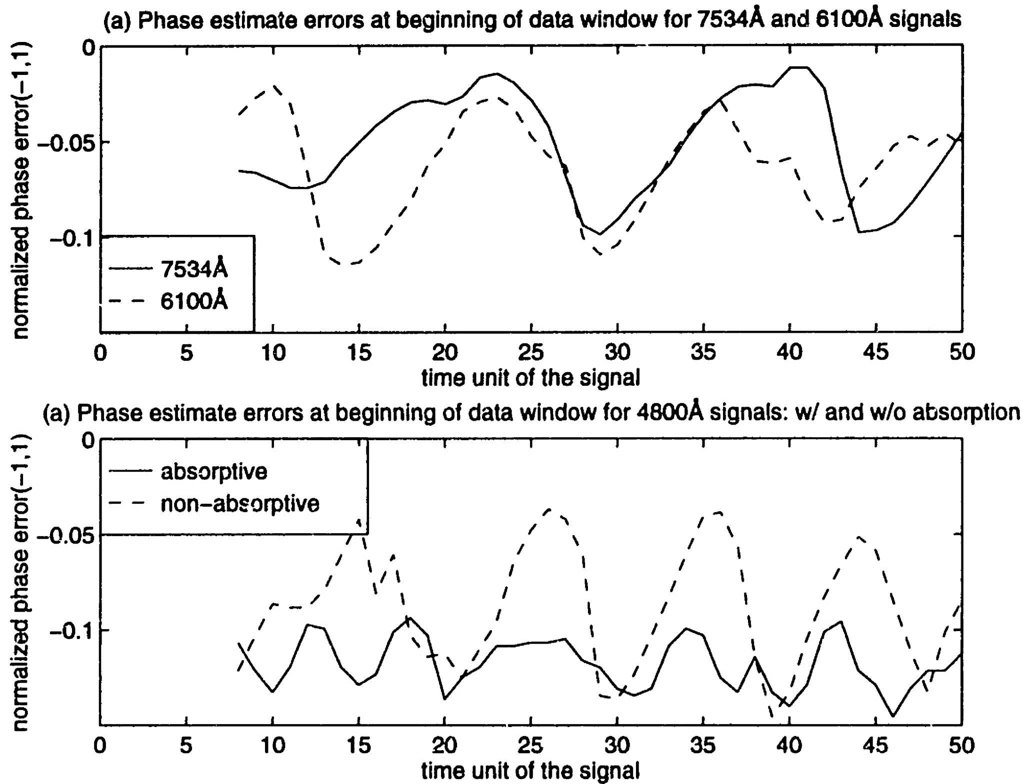


Table 5-4: Thickness estimation error at beginning of data window for different wavelengths

signal wavelength (Å)	mean of thickness error (Å)	std. dev. of thickness error (Å)
7534	-26.43	13.77
6100	-24.73	10.62
4800(w/ abs.)	-31.95	3.64
4800(w/o abs.)	-23.83	8.39

5.4 Sensitivity analysis of endpoint phase angle with respect to film structure variations

In this section, the accuracy of using the phase as an endpoint prediction measurement is examined with respect to variations in film structure. Figure 5-17 shows the interferometric reflectance for etching 5000Å of polysilicon over slightly different thicknesses of oxide. Since the reflectance curves differ, the phases at endpoints will differ. The sample interferometric reflectance is calculated from the etch of polysilicon over silicon oxide on a silicon substrate. The variations to be considered are the refractive indices of the etching films, and the reflective indices and thicknesses of the underlying films. Similar to the previous sections, the performances of three different wavelengths are considered.

As an example, consider the interferometric signals at 7534Å with a thick oxide layer with the following film layer structure:

- refractive index of polysilicon: 3.73
- refractive index of oxide: 1.45
- thickness of oxide: 1000Å
- refractive index of silicon substrate: 3.73

Assuming a 5% variation in the nominal values of the film structure, and assuming the absence of measurement error, endpoint phase is determined at both the high (H) and low (L) levels of the properties in a 2^4 factorial design, as shown in Table 5-5. The parameters'

levels are shown in Table 5-6. The largest prediction error using the fixed endpoint phase will lead to an error of 17.5Å. Other simulation results are included in the appendices. Table 5-7 summarizes results from all the cases considered.

Figure 5-17: Interferometric reflectance for etching 5000Å of polysilicon on different oxide thicknesses over silicon substrate

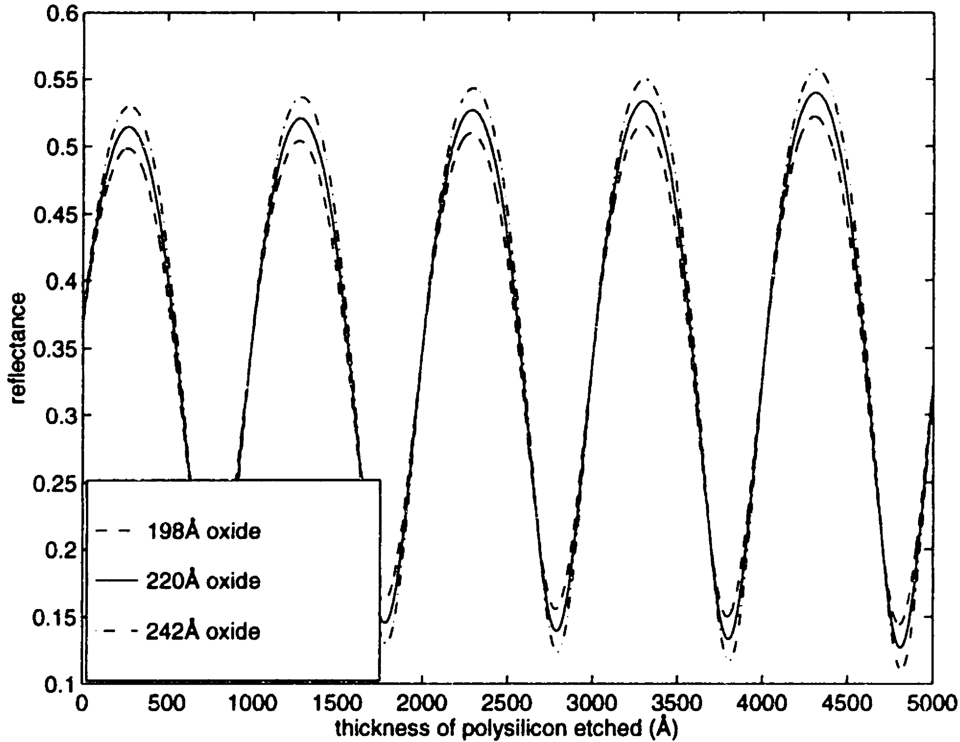


Table 5-5: Error of film thickness estimation from designed experiments

Experiment number	Poly refrac. index	oxide thickness	oxide refrac. index	sub. refrac. index	Δ thick. pred. (Å)
1	L	L	L	L	15.5
2	H	L	L	L	8.4
3	L	H	L	L	0.5
4	H	H	L	L	-3.1
5	L	L	H	L	4.0
6	H	L	H	L	-1.1
7	L	H	H	L	-12.5
8	H	H	H	L	-0.3

Table 5-5: Error of film thickness estimation from designed experiments (cont.)

Experiment number	Poly refrac. index	oxide thickness	oxide refrac. index	sub. refrac. index	Δ thick. pred. (Å)
9	L	L	L	H	17.5
10	H	L	L	H	9.9
11	L	H	L	H	2.5
12	H	H	L	H	-2.1
13	L	L	H	H	5.5
14	H	L	H	H	0.4
15	L	H	H	H	-11.5
16	H	H	H	H	13.1

Table 5-6: Parameter levels for the design of experiments

	level L	level H
poly. refrac. index	3.5435	3.9165
oxide thickness	950Å	1050Å
oxide refrac. index	1.3775	1.5225
sub. refrac. index	3.5435	3.9165

Table 5-7: Maximum errors in endpoint prediction w.r.t. film structure variations (Å)

		nominal oxide thickness		
		80Å	220Å	1000Å
signal wavelength	4800Å	20.8	11.63	-18.9
	6100Å	45.8	20.8	13.5
	7534Å	77.5	33.52	17.5

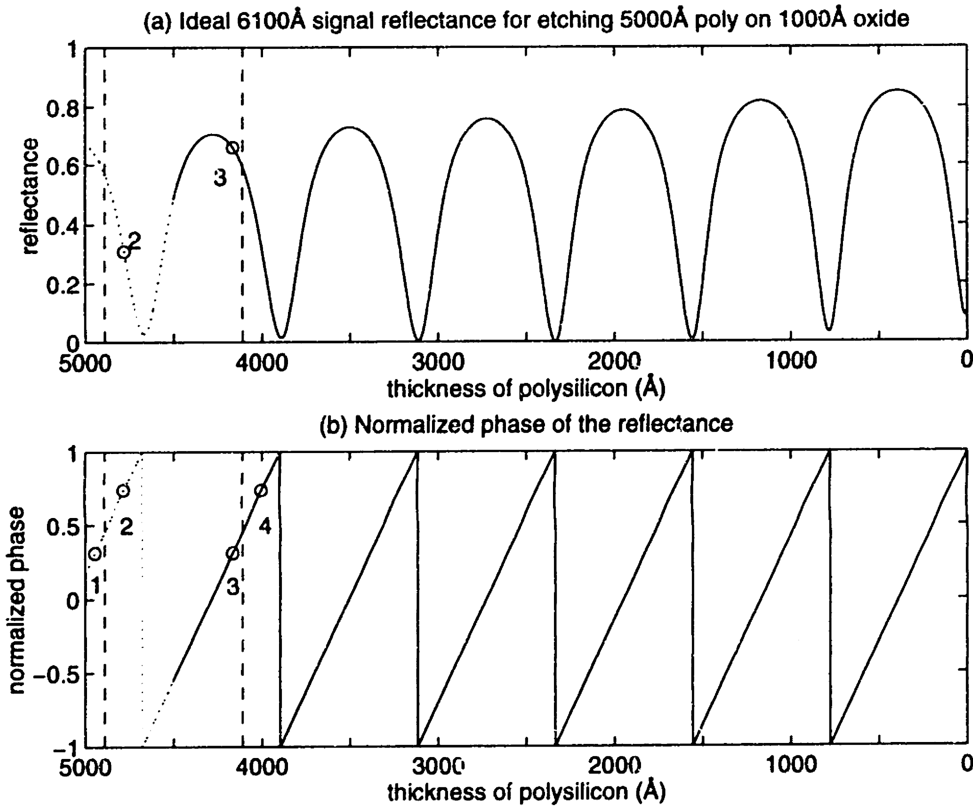
From the simulation results, it can be seen that when the oxide layer is thick, such as 1000Å, 5% variation in film properties will introduce only a small error in endpoint prediction. However, when the oxide layer is extremely thin, the short wavelength signal is preferred because the error introduced by the 5% variation is much smaller. Response surfaces have been fitted to the different cases, and interactions among the four parameters are found to be small. The primary factor responsible for the variations in endpoint phase is the refractive index of the etching layer, which will lead to errors in endpoint prediction.

5.5 Practical issues

Given the endpoint phase, and the film thickness per cycle of signal (which can be determined experimentally or by simulation), the total number of cycles and the initial phase can be calculated. Determining the right number of cycles of signals to be received is essential for the correct film thickness estimation or endpoint prediction. If the total number of cycles is counted incorrectly, the estimated film thickness or endpoint position will be off by a thickness equivalent to at least one cycle of signal. One proposed method is counting the sudden change of the wrapped phase from $+\pi$ to $-\pi$. This is equivalent to observing the number of minima of the signals. However, because of variation in initial film thickness, the total number of signal minima may vary also. Hence, it will be more robust to assume that the initial phase of the signal could be different from the expected value by $\pm\pi$. As a result, the allowable variation in the initial film thickness is then equivalent to one cycle of signal. Since a cycle of short wavelength signals corresponds to a thinner layer of films, it can accommodate a smaller variation in initial film thickness. In an extreme case, if one cycle of signals corresponds to a very small thickness of film, a slight film thickness difference will result in false endpoint prediction.

Consider Figure 5-18*a*. The solid line represents the 6100Å interferometric reflectance for etching 4500Å of polysilicon on 1000Å of oxide, while the dotted line corresponds to the etching of an extra 500Å of polysilicon (i.e. a total thickness of 5000Å). Assuming that the expected initial polysilicon film thickness is 4500Å, the normalized initial phase should be -0.5448. The ±0.5-cycle bounds are at the normalized phase of (1-0.5448)=0.4552, which is shown as the dashed lines in Figures 5-18*a*,*b*. For a particular

Figure 5-18: Selection of initial phase using allowable thickness bounds



film, if the initial normalized phase is determined to be 0.3130, it could represent locations 1 or 3 as shown in Figure 5-18*b*. However, since location 3 is within the allowable bounds of the initial film thickness, it is preferred and the film is assumed to be thinner than expected. Similarly, consider an initial normalized phase of 0.7359. Although it could be at locations 2 or 4, location 2 is assumed because it is within the allowable film thickness bounds. For location 2, six $+\pi$ to $-\pi$ phase transitions will be counted before predicting endpoint, while only five transitions are assumed if the film is exactly 4500Å thick. For the experiments in this chapter, the 6100Å wavelength signal for etching 5000Å of polysilicon on 1000Å of oxide will be examined. This wavelength is selected so that enough cycles of signals will be obtained, while the signal absorption rate will not be too large.

A MATLAB script file [45] for endpoint prediction is included in Appendix D. From past experiments, two cycles of signals were found to be essential for moderately accurate estimation of frequency and phase. Hence, starting with approximately a cycle of signal data, the algorithm first determines the number of data point equivalent to two cycles of signals before calculating the phase. This step is used for setting up a frequency bound on the

periodicity of the etch signal, which will improve the robustness of endpoint prediction. The 2-cycle specification is selected because its only effect is incrementing the total number of cycle count by an integer of 2. When new data are received, the subsequent phases will be extracted from a sliding data-window of about 2-signal-cycle (or from all the previous data as explained in Section 5.7.5). If the initial available signal is less than one cycle, noise in the signal may lead to incorrect frequency and phase calculation. However, in routine production, the approximate length of data for one cycle of signal should be available.

5.6 Experiments on film thickness prediction

In this section, experiments are performed to verify the prediction capability of the proposed algorithm. These experiments are performed in an Applied Materials Precision 5000 etcher. They are designed for verifying the estimation of film thickness left after an etch. In later sections, experiments for testing endpoint prediction when the estimated film thickness is zero will also be considered.

For film thickness estimation, a total of eight experiments are performed. Four experiments are designed for film thickness calibration and four other experiments will be used for testing purpose. Some of the main control parameters for etching of polysilicon film include power, pressure, magnetic field strength, and etchant gas ratio. In our experiments,

Figure 5-19: Design parameters for film thickness experiments

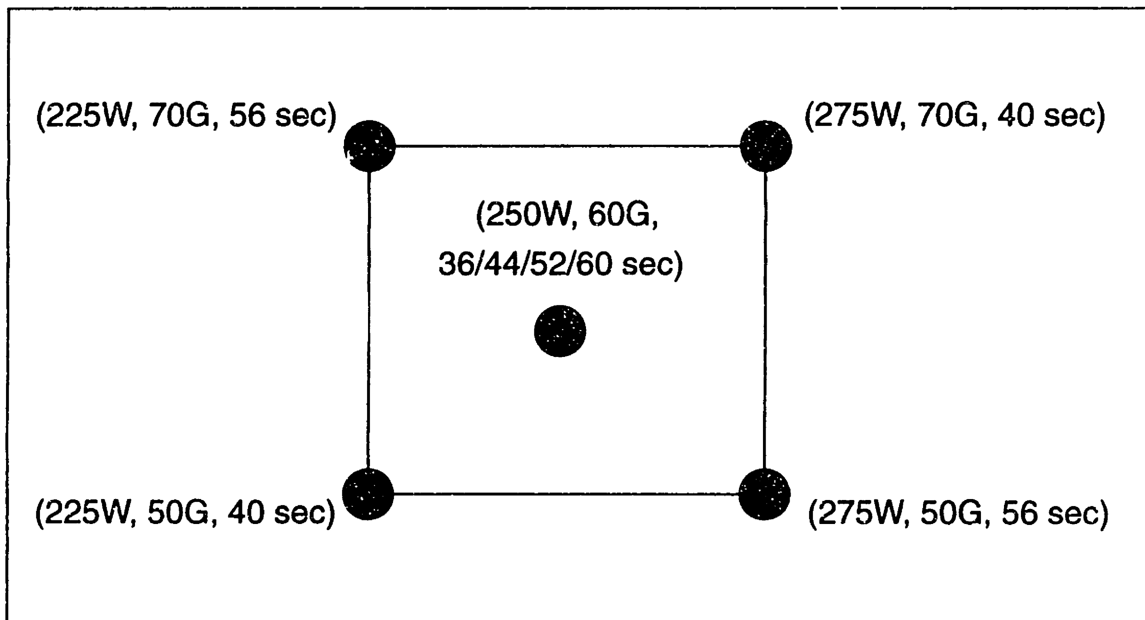
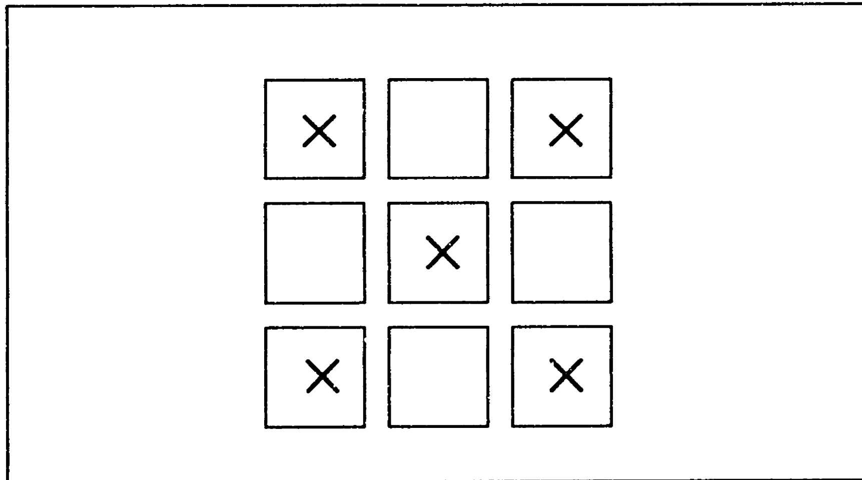


Figure 5-20: Pixel pattern selected for five-pixel averaging



the pressure and etchant gas are fixed at 100 mtorr and 10 sccm HBr / 30 sccm Cl_2 respectively. The initial four calibration experiments correspond to the design center in Figure 5-19. They are run at the power level of 250W and magnetic field strength of 60G. The etch times are selected to be at 36, 44, 52 and 60 seconds to produce different lengths of interferometric signals. In order to reduce the level of noise in the interferometric signal, five-pixel-averaged signals are used, which corresponds to areas of $300\mu m \times 300\mu m$ on the wafer approximately. For a group of 3×3 pixels, the five pixels selected for averaging are marked by "x" as shown in Figure 5-20. Figure 5-21 shows the interferometric signals corresponding to the four cases at a bare polysilicon site; the phase at the last data points were calculated via Fast Fourier Transform. From the Nanospec, thicknesses of polysilicon left after the etches are measured. A linear function can be established to estimate film thickness left from the calculated phase, as shown in Figure 5-22.

After the phase/film thickness relation has been established, four other testing experiments are performed. The power, magnetic field strength, and etch times were selected as the four corner designs in Figure 5-19. The power, magnetic field strength, and etch time are varied to obtain interferometric signals with different periodicities and signal lengths. Similarly, for the same four selected bare polysilicon sites, the phase at the end of the interferometric signals are calculated, and the film thicknesses are obtained. All eight film thickness data and the fitted phase-to-thickness linear model are shown in Figure 5-22. Table 5-8 shows the fitting errors for the modeling building experiments and the estimation error for the testing experiments. The largest error for film thickness estimation is about 100Å.

Figure 5-21: CCD signals for the calibrating experiments

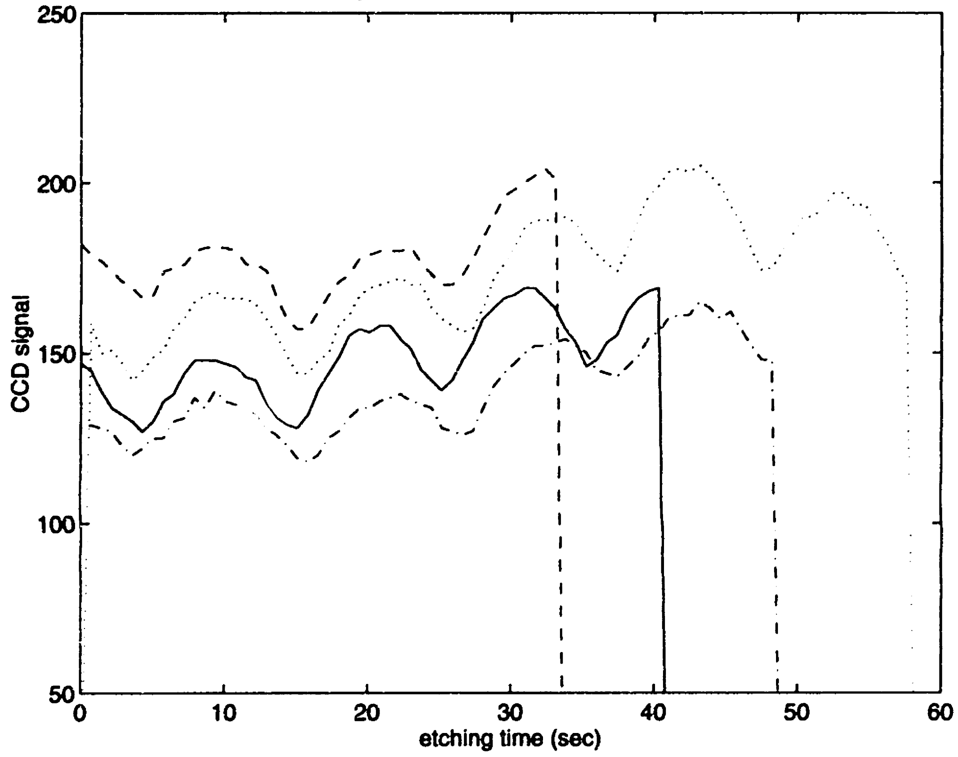


Figure 5-22: Experimental data and fitted model for polysilicon thickness estimation

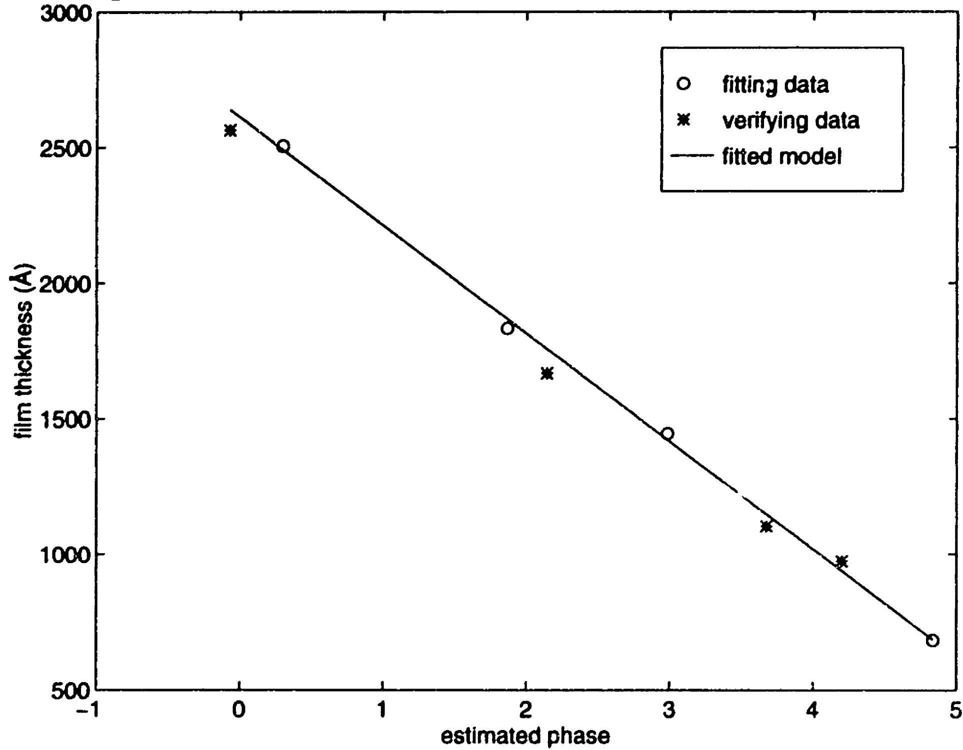


Table 5-8: Fitting and estimation errors in film thickness using phase of interferometric signals

	pt. 1	pt. 2	pt. 3	pt. 4
fitting error (Å)	-13.37	-5.83	1.79	2.30
	34.29	36.71	4.32	13.00
	-22.22	-46.52	-11.73	-26.97
	1.30	15.64	5.62	11.67
verification error (Å)	75.76	88.66	70.72	78.76
	89.67	106.12	85.19	86.06
	44.75	25.45	34.70	44.40
	-34.33	3.85	-44.99	-14.55

5.6.1 Endpoint prediction versus timed etching

In this section we compare the effect of conventional timed etch with endpoint prediction on film thickness variation. First, consider the possible variation in a timed etch. Assume that the nominal etch thickness is Th_0, the nominal etch rate is ER_0, the fractional variation in etch rate is f_{er}, and the fractional variation in initial film thickness is f_{th}. To estimate typical etch rate variations (f_{er}), consider the four experiments used for modeling building in Figure 5-19. Using Fast Fourier Transform method, the etch rates at pt. 1 are found to be 52.74, 51.03, 50.84, and 52.36 Å/time unit. The amount of variation, with process mean centered, is approximately:

$$\frac{1}{2} \times \frac{range(52.74, 51.03, 50.84, 52.36)}{mean(52.74, 51.03, 50.84, 52.36)} = 0.0184 = 1.84\%. \quad (5\text{-}1)$$

The nominal etch time is:

$$t_0 = \frac{Th_0}{ER_0}. \quad (5\text{-}2)$$

Due to the variation in etch rate, the actual etch thickness (Th) is:

$$Th = t_0 \times ER_0 \times (1 \pm f_{er}). \tag{5-3}$$

The variation in etching thickness (ΔTh) due to the fluctuations in etch rate is:

$$\Delta Th = Th - Th_0 = \pm f_{er} Th_0. \tag{5-4}$$

Including the initial film thickness variation ($\Delta Th_{incoming}$) of the incoming wafer, the total possible error in etched endpointing is:

$$\Delta Th_{endpoint} = \Delta Th_{incoming} \pm f_{er} Th_0. \tag{5-5}$$

Consider the etch of 2500Å of polysilicon. It is not unreasonable to assume a 2% variation in the etch rate and the incoming film thickness, which leads to a possible etch thickness error of:

$$\begin{aligned} \Delta Th_{endpoint} &= \pm 0.02 \times 2500 \pm 0.02 \times 2500 \text{ (Å)}, \\ &= \pm 100 \text{ (Å)}. \end{aligned} \tag{5-6}$$

In this case, the error is comparable to the worst-case of endpoint prediction error presented in Table 5-8. However, the experiments for film thickness estimation in Table 5-8 were performed at very different equipment settings and for very different film thickness. For mass production with the same equipment setting and similar film thickness, the film thickness estimation error should be much smaller. In addition, the use of endpoint prediction should be more robust against larger incoming film thickness variations or larger etch rate drift or step variations after routine maintenance work.

5.7 Experiments on endpoint prediction

In this section, the use of endpoint phase to predict the end of an etch is examined. For etching 5000Å of polysilicon on 1000Å of silicon oxide, it has been found from theory that six cycles of signals will be received and the endpoint occurs at a normalized phase angle of 0.9851. Or, the endpoint phase can be determined from a number of representative runs to accommodate more complex layer stacks. Then, the following endpoint prediction

algorithm is applied to predict endpoint of an etch:

a) Count the number of cycles of signals;

b) Signal endpoint if the number of required signal cycles, and the endpoint phase angle are reached.

A linear trend was subtracted from the signals before the phase calculation to reduce interference caused by the photoresist. In some cases, the interferometric signals are first normalized by an overall emission level, obtained from a non-etching area such as the clamping pins, before prediction algorithms are applied. However, the crude data will be emphasized in this thesis, which will be explained in Section 5.7.2.

A set of designed experiments are performed for analyzing different aspects of endpoint prediction using signal phase angle. The same four control parameters are considered, i.e., RF power, magnetic field strength, pressure and HBR/Cl_2 gas flow rates. A Box-Behnken design for the four variables was selected because the design is suitable for exploring a broad operating region using only a few experiments. The design levels for the four parameters and the design array are listed in Tables 5-9 and 5-10 respectively. The levels of the control parameters are deliberately selected to cover wide operating ranges.

The wafers were deposited with 5000Å of polysilicon on 1000Å of silicon oxide. A CCD snapshot of the pattern on the wafer is shown in Figure 5-23. The five features on each of the dice are bare polysilicon area, vias (rows of square holes), lines with various line widths, serpentines (meandering lines), and series of diffraction gratings. The stepper used for patterning could print features as low as 1 µm. Refer to [17] for a detailed description of the mask. Similar to Section 5.6, the signals used for phase determination are five-pixel averaged signals as shown in Figure 5-20 to reduce sensor noise.

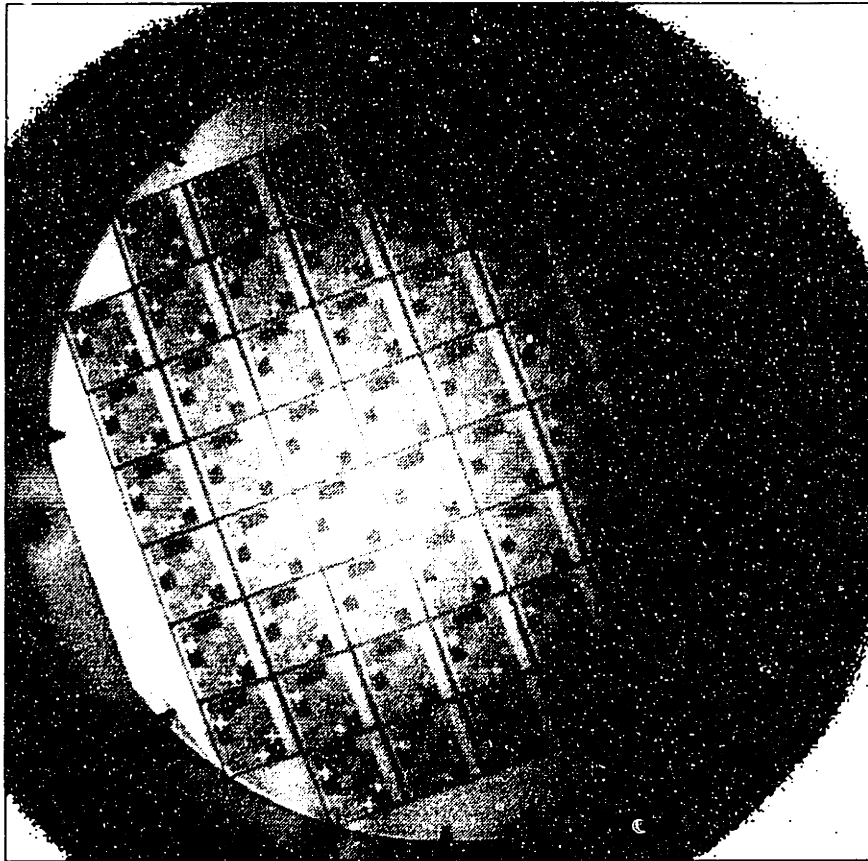
Table 5-9: Design levels for the four control parameters

design levels	RF power (W)	HBr/Cl_2 (sccm/sccm)	pressure (mtorr)	mag. field (gauss)
-1	150	5/35	50	20
0	250	10/30	100	60
1	350	15/15	150	100

Table 5-10: Design array for 4-parameter Box-Behnken design

Experiment #	RF Power	HBr/Cl_2	Pressure	Mag. Field
1	-1	-1	0	0
2	1	-1	0	0
3	-1	1	0	0
4	1	1	0	0
5	0	0	-1	-1
6	0	0	1	-1
7	0	0	-1	1
8	0	0	1	1
9	0	0	0	0
10	-1	0	0	-1
11	1	0	0	-1
12	-1	0	0	1
13	1	0	0	1
14	0	-1	-1	0
15	0	1	-1	0
16	0	-1	1	0
17	0	1	1	0
18	0	0	0	0
19	-1	0	-1	0
20	1	0	-1	0
21	-1	0	1	0
22	1	0	1	0
23	0	-1	0	-1
24	0	1	0	-1
25	0	-1	0	1
26	0	1	0	1
27	0	0	0	0

Figure 5-23: CCD image of an etching wafer



5.7.1 Endpoint prediction with phase (without emission level normalization)

In Section 5.7.3, the statistics for verifying the endpoint prediction algorithm with the design of experiments will be given. For each wafer a total of thirty-six measurement sites are considered as shown in Figure 5-24. However, specific properties regarding the algorithm are first discussed here and in the following sub-section.

Figures 5-25*a*,*b* and 5-26*a*,*b* show the etching signal and the corresponding endpoint for the bare polysilicon at site #15 for experiments #10-13. The settings for these four experiments are repeated in Table 5-11 for easy reference. These four experiments are selected for illustration because the RF power and magnetic field strength are the primary factors affecting the etch rate, and they vary substantially in these experiments. The phase angle of the signals were determined using a sliding data window of two signal cycles. The computations repeat until condition (b) of the above algorithm was reached. For each etch, the plasma was first ignited for a few seconds so that the desired CCD pixels can be selected; in this case we seek to predict or signal endpoint at 0Å thickness remaining. As a

Figure 5-24: Selected endpoint analysis dice

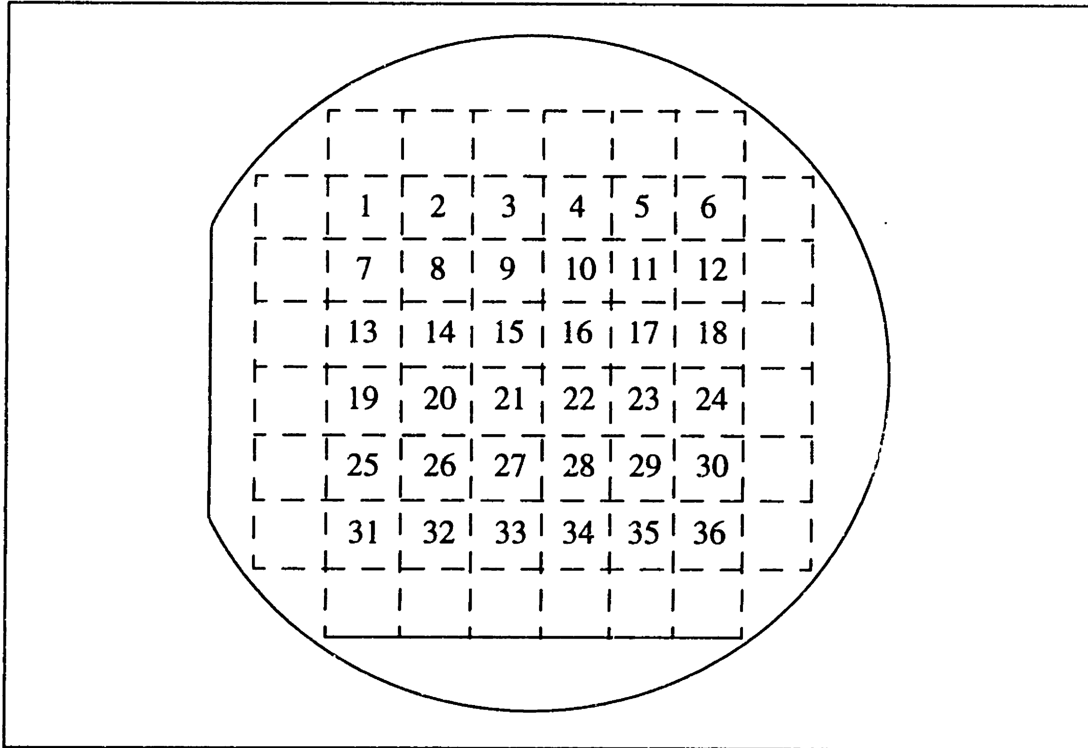


Figure 5-25: Endpoint prediction for bare poly areas I

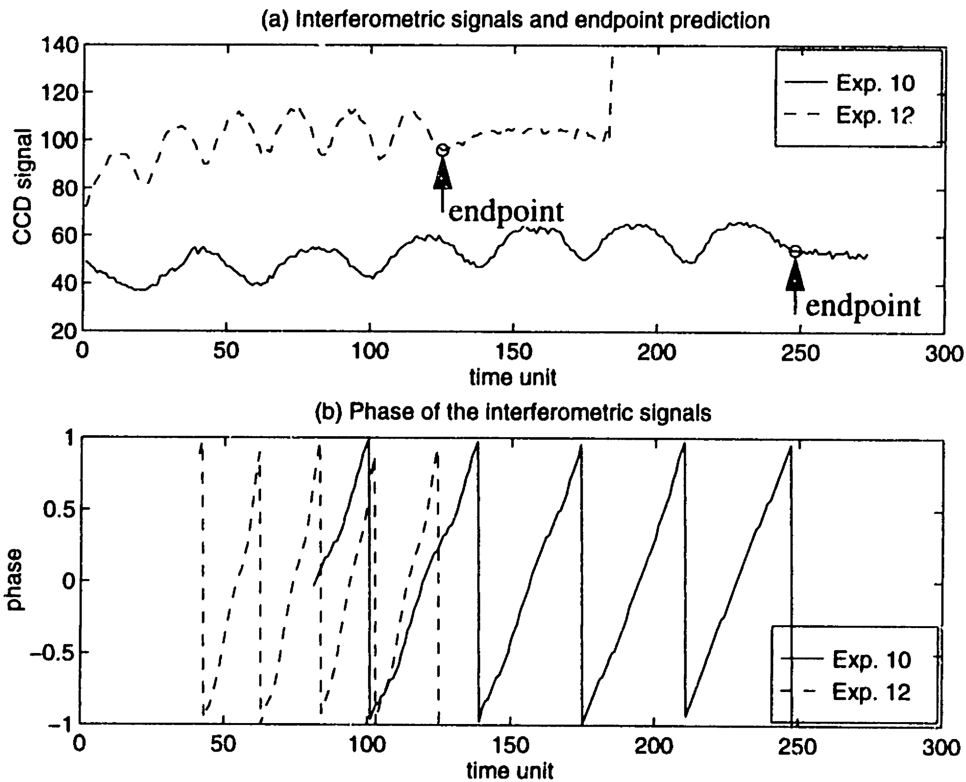


Figure 5-26: Endpoint prediction for bare poly areas II

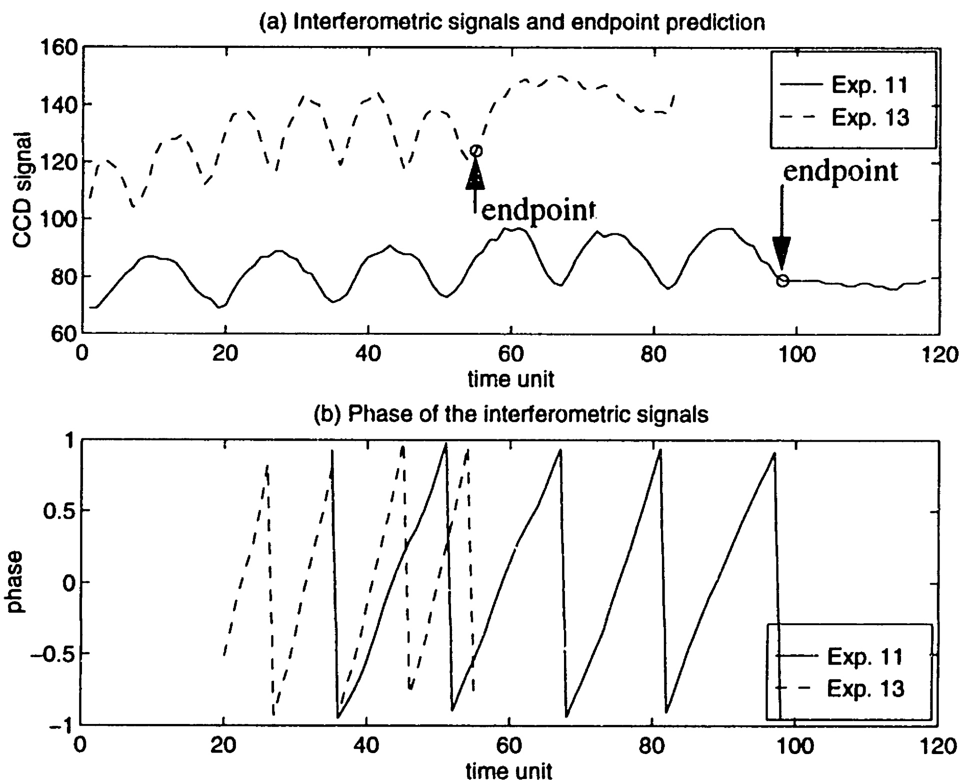


Table 5-11: Selected equipment settings from 4-parameter Box-Behnken design

Experiment #	RF power (W)	HBr/Cl_2 (sccm / sccm)	Pressure (mtorr)	Mag. Field (G)
10	150	10 / 30	100	20
11	350	10 / 30	100	20
12	150	10 / 30	100	100
13	350	10 / 30	100	100

result, the initial phases of the signal are shown to be different due to the initial short period of etching. In addition, the periodicities of the four interferometric signals are not the same due to different etch rates. From the plots, the endpoints were shown to be predicted with very good accuracy for the case of experiment #10-12 by using the endpoint phase. For experiment #13, although the endpoint selected satisfies the cycle-number and phase requirement, it is not obvious that endpoint is predicted correctly. Such phenomenon is due to the change in overall optical emission level when the film starts clearing up, and its effect on endpoint prediction will be explained in the next section.

For etching 5000Å of polysilicon, an error of one time unit for experiment #13 corresponds to the etching of about 100Å of polysilicon. On the other hand, one time unit error for experiment #10 corresponds to etching 20Å of polysilicon. Similarly, Figures 5-27 to 5-34 show the predicted endpoints and estimated phases for different features for the same dice. It can be observed that the endpoints can be predicted fairly accurately for the case of experiments #10-12 also. In addition, Figures 5-35 to 5-39 show additional signals and the predicted instance at zero film thickness. These examples are difficult cases and are selected to illustrate the robustness of using phase for endpoint prediction.

In Figure 5-40, the endpoint time for the line regions of the selected thirty-six dice from experiment #10 is shown to illustrate an application of the endpoint prediction algorithm and how non-uniform the etch is.

Figure 5-27: Endpoint prediction for via areas I

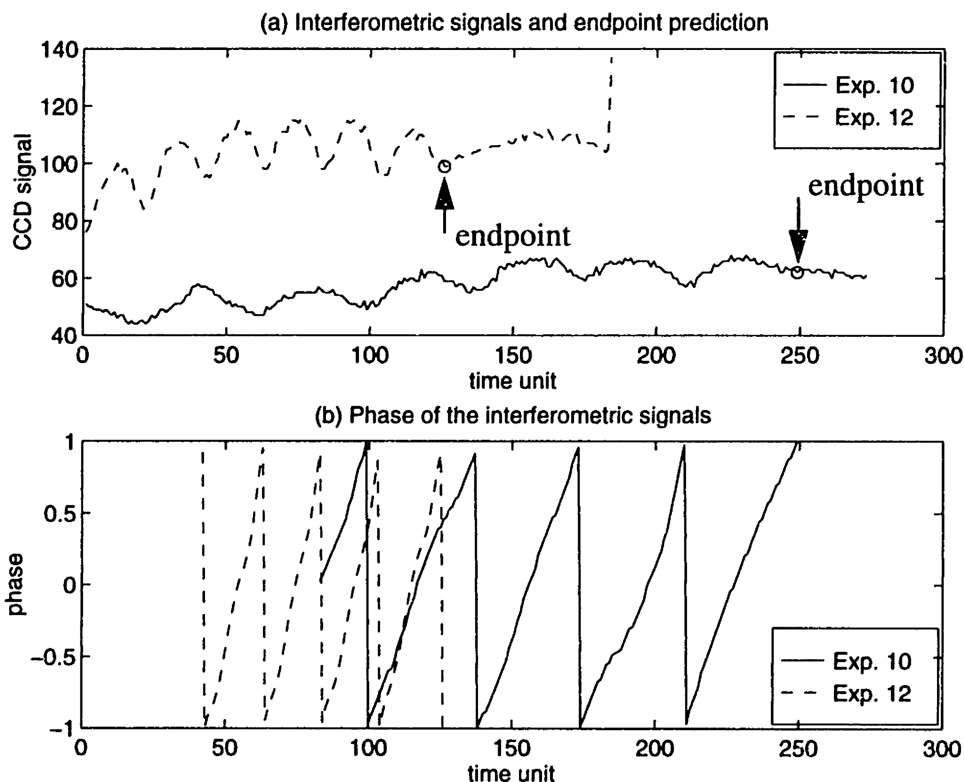


Figure 5-28: Endpoint prediction for via areas II

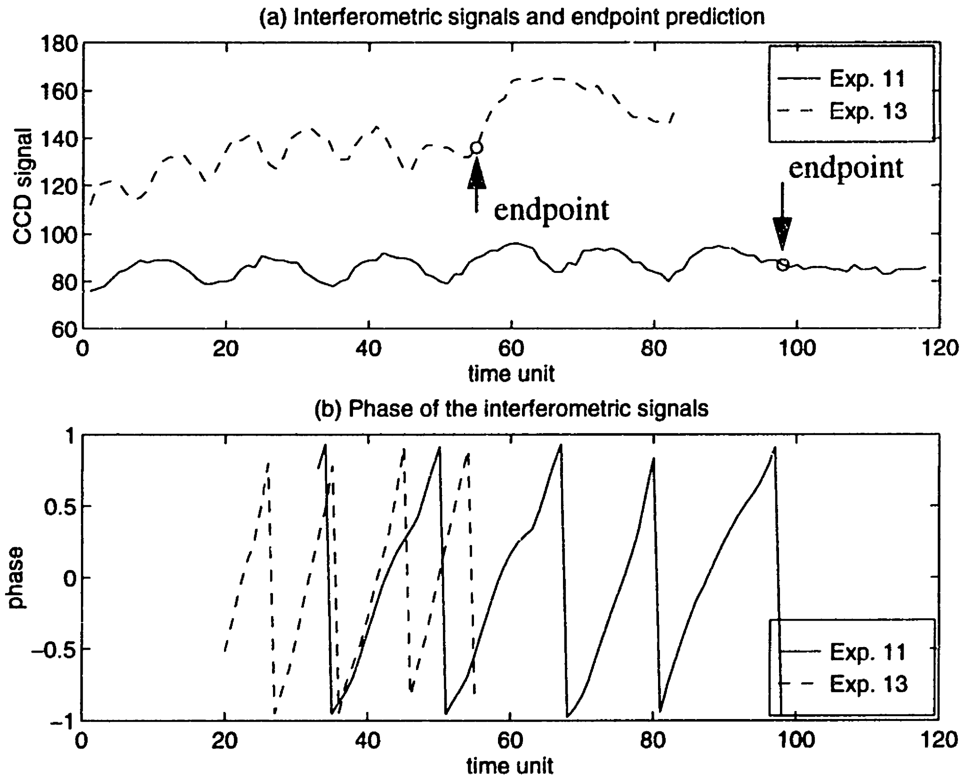


Figure 5-29: Endpoint prediction for line area I

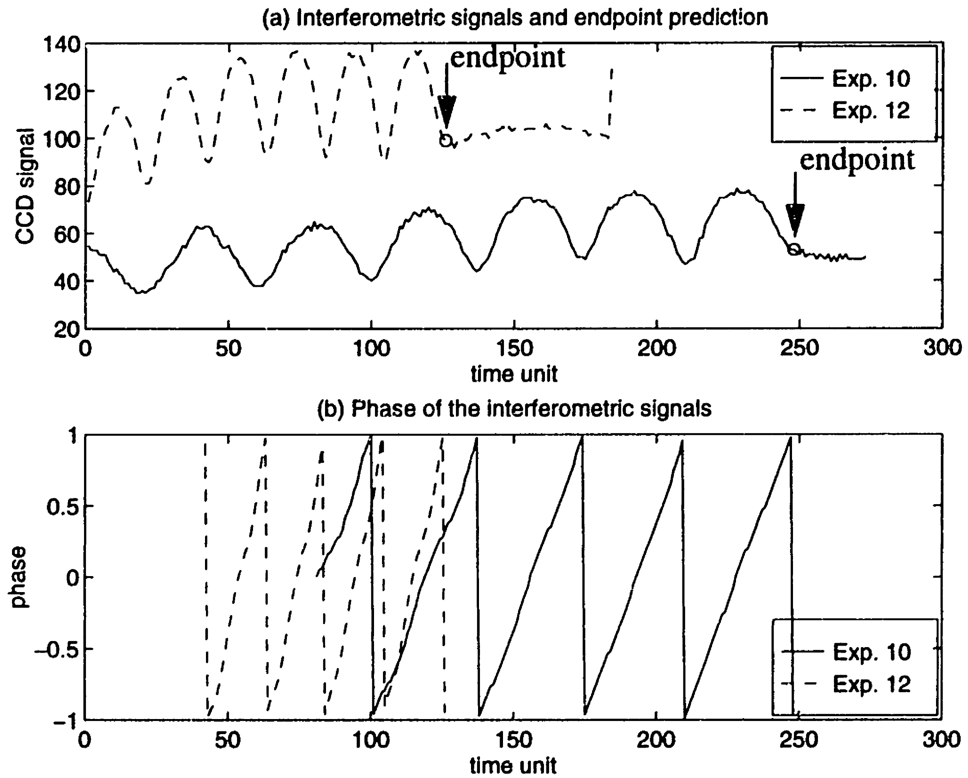


Figure 5-30: Endpoint prediction for line area II

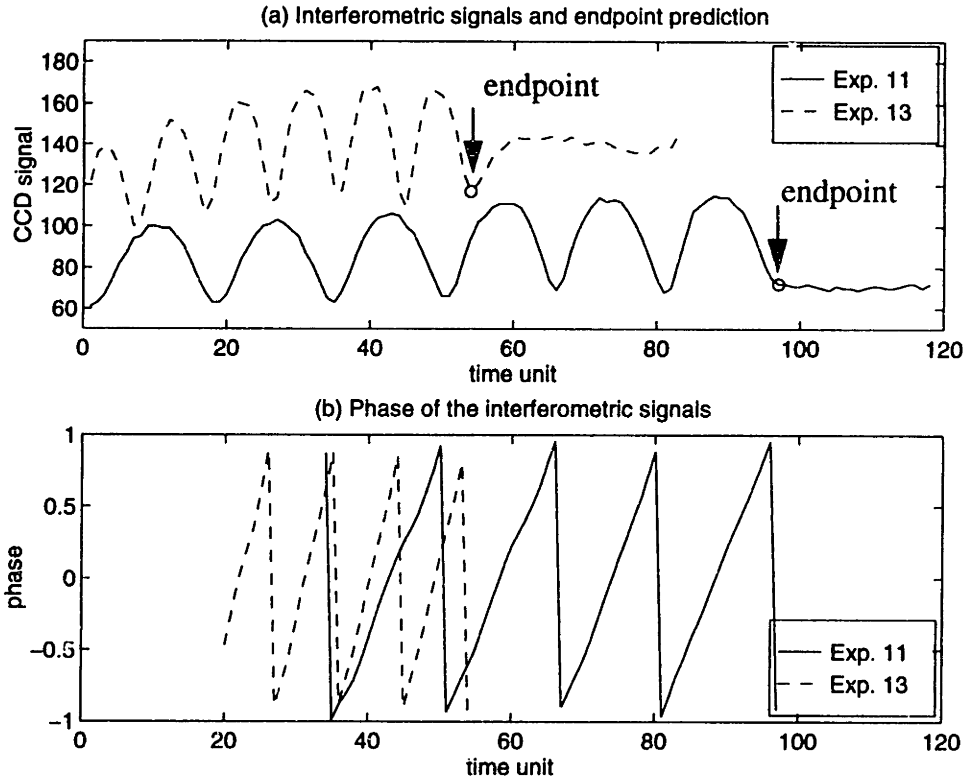


Figure 5-31: Endpoint prediction for serpentine areas I

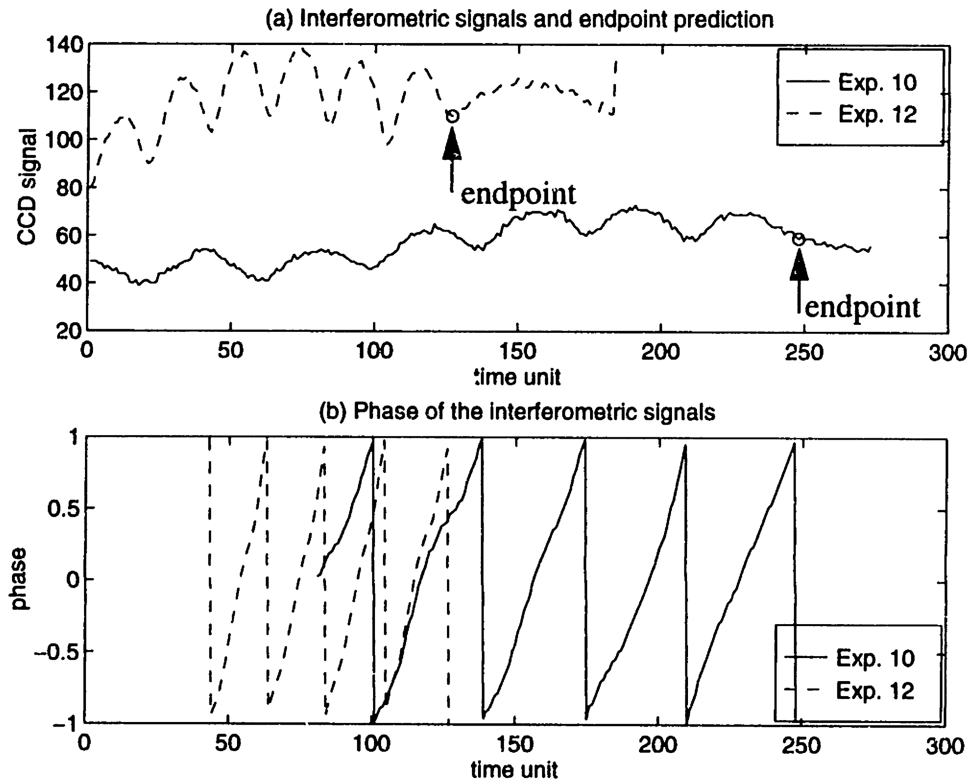


Figure 5-32: Endpoint prediction for serpentine areas II

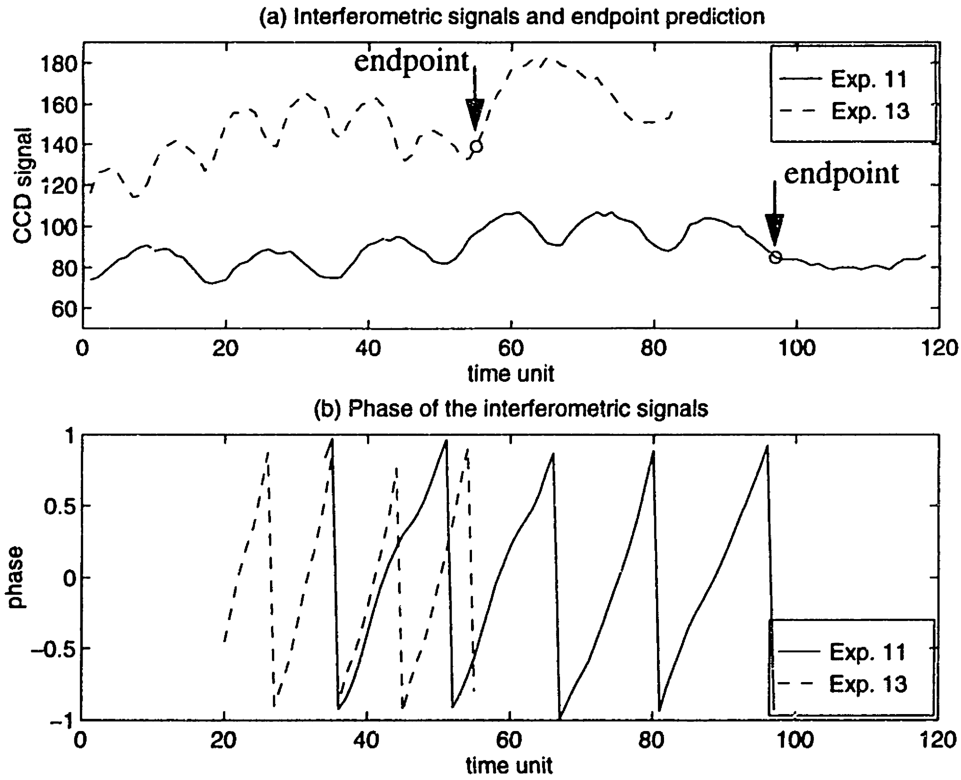


Figure 5-33: Endpoint prediction for diffraction gratings I

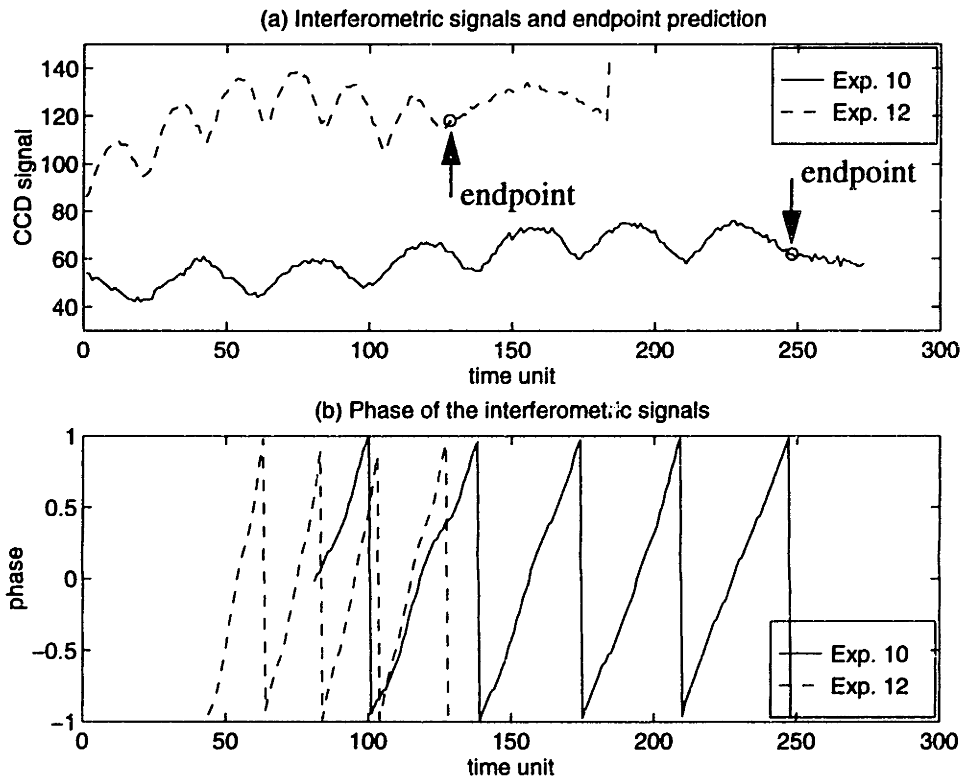


Figure 5-34: Endpoint prediction for diffraction gratings II

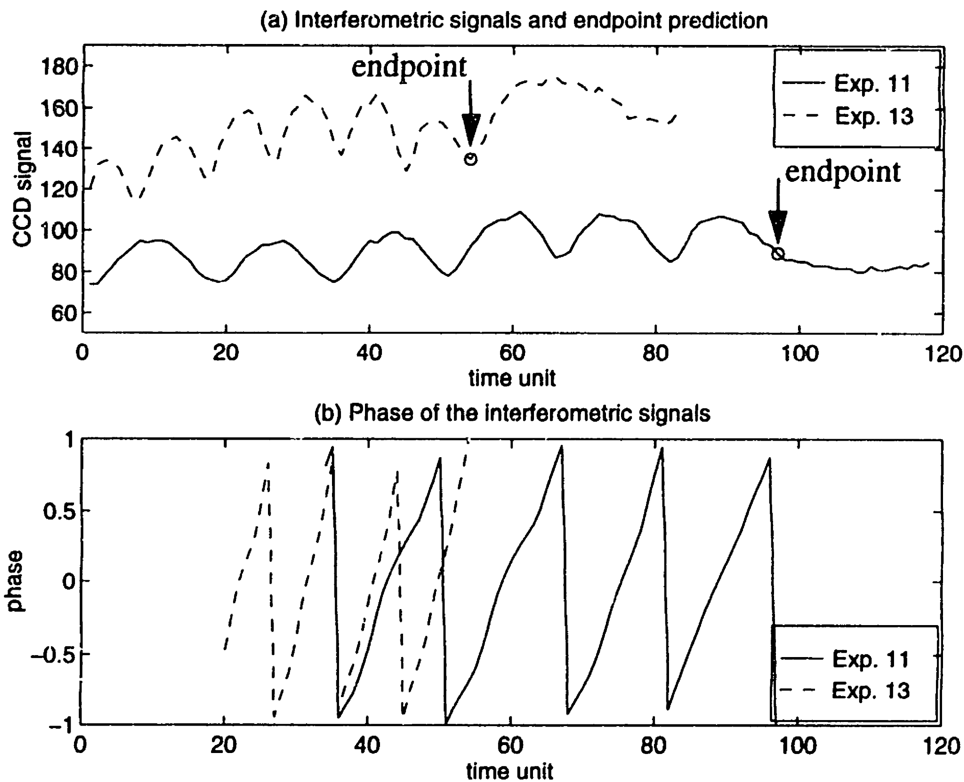


Figure 5-35: Endpoint prediction for noisy data I

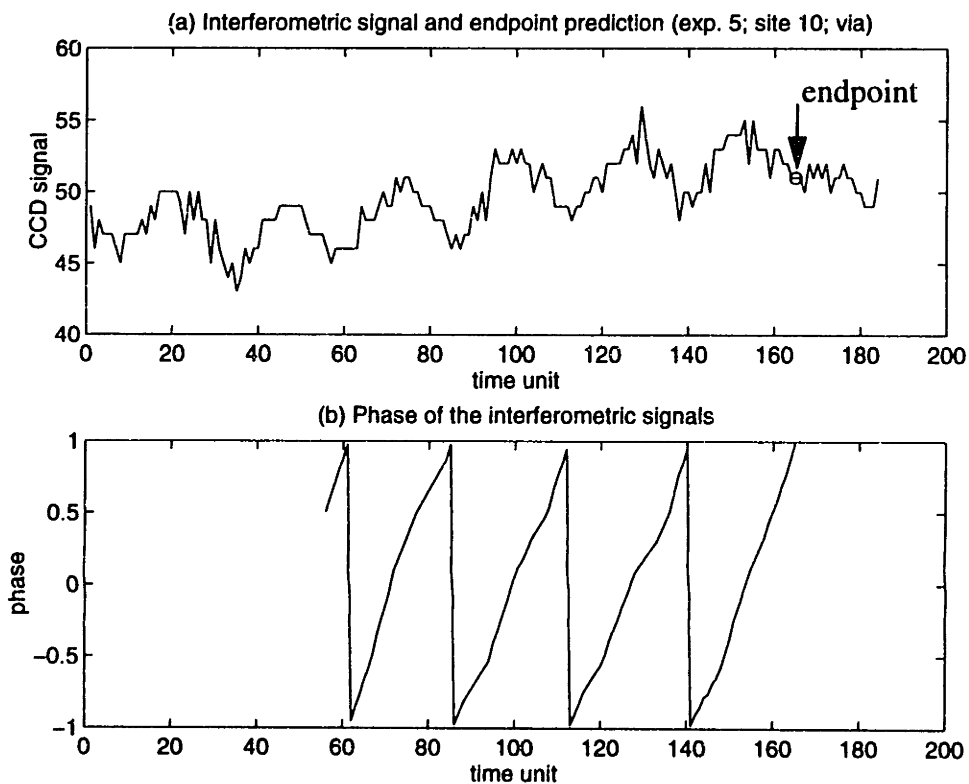


Figure 5-36: Endpoint prediction for noisy data II

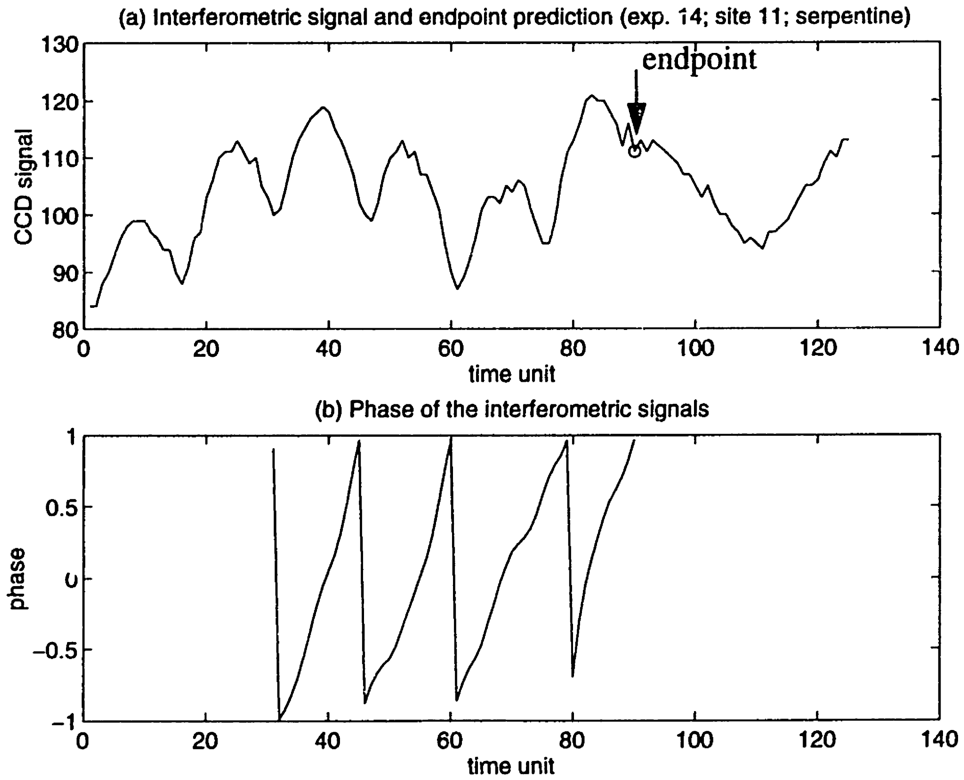


Figure 5-37: Endpoint prediction for noisy data III

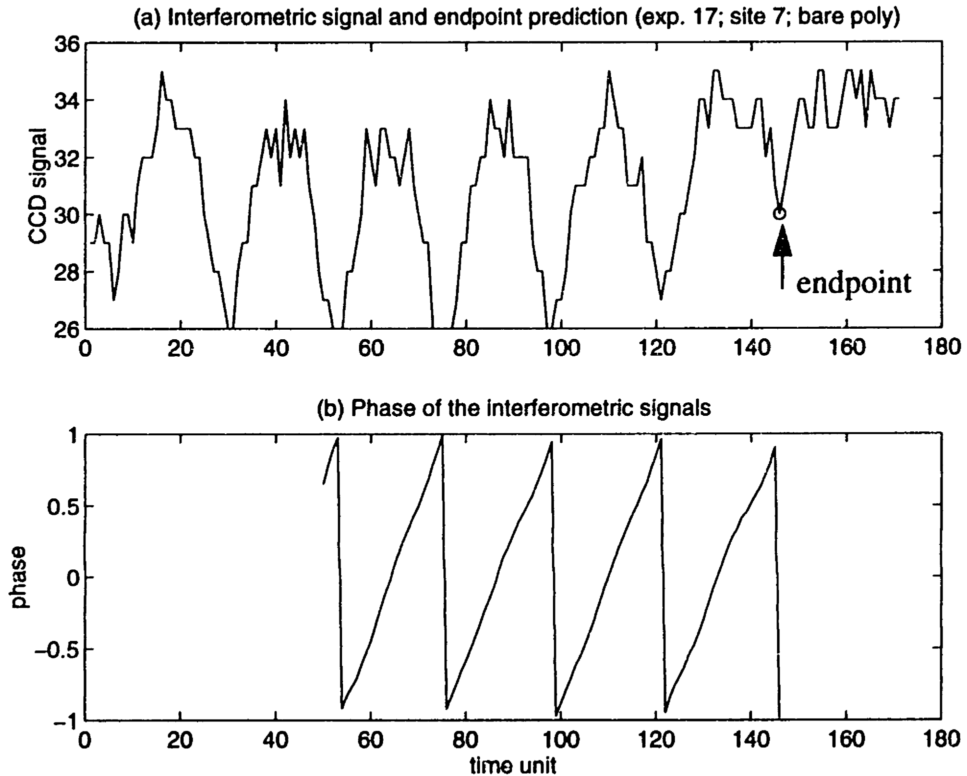
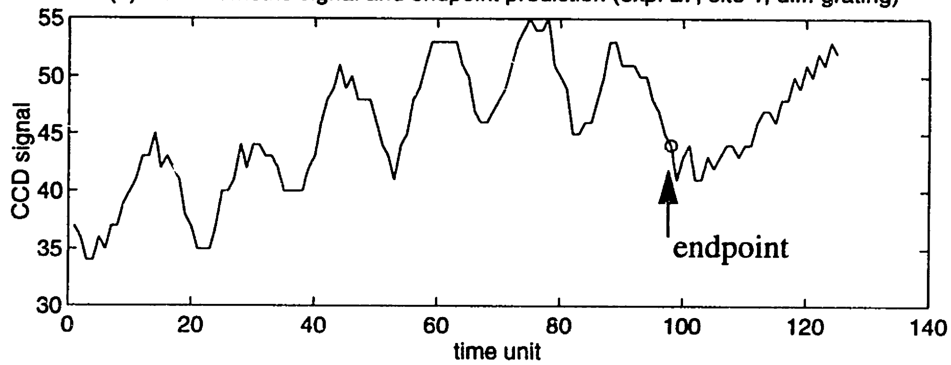


Figure 5-38: Endpoint prediction for noisy data IV



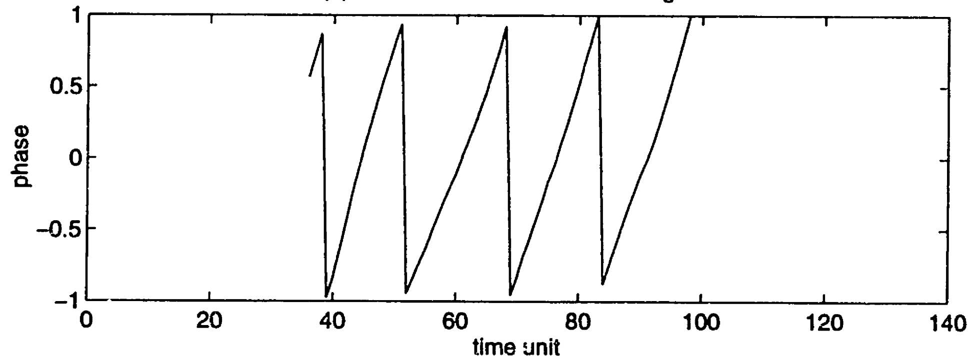
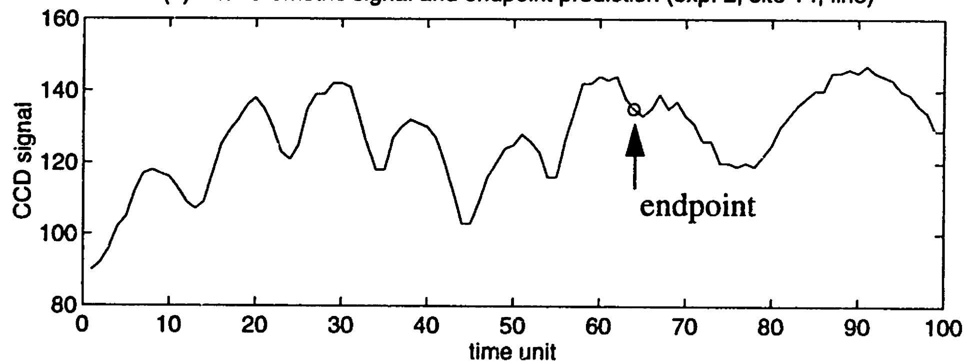


Figure 5-39: Endpoint prediction for noisy data V



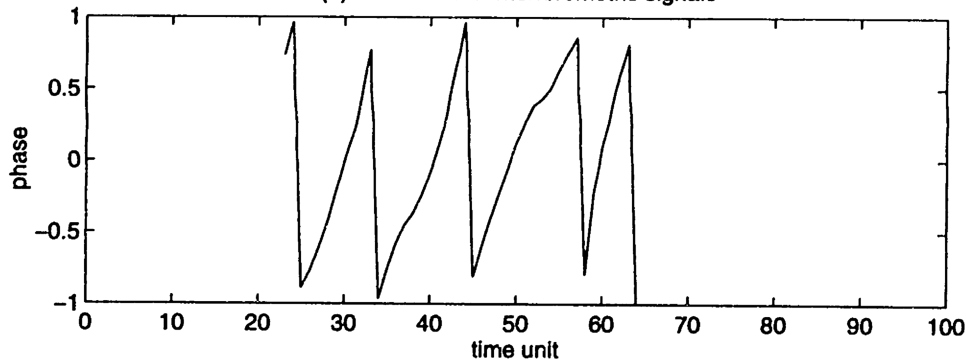
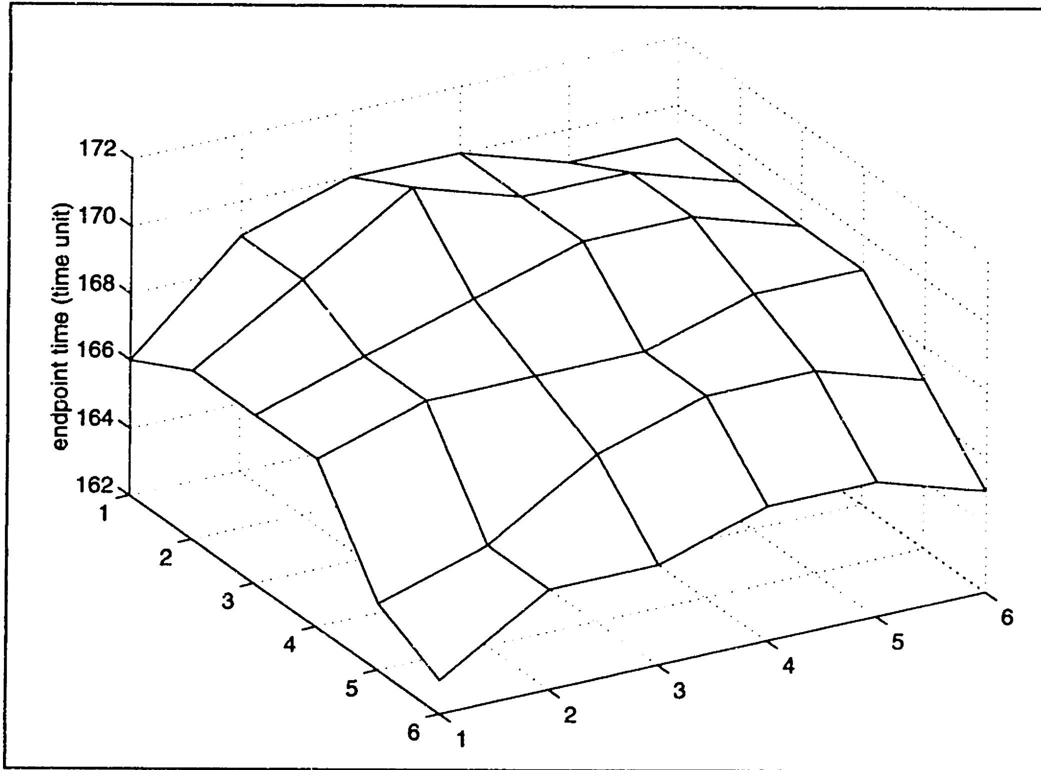


Figure 5-40: Endpoint time for line regions from Exp. 10

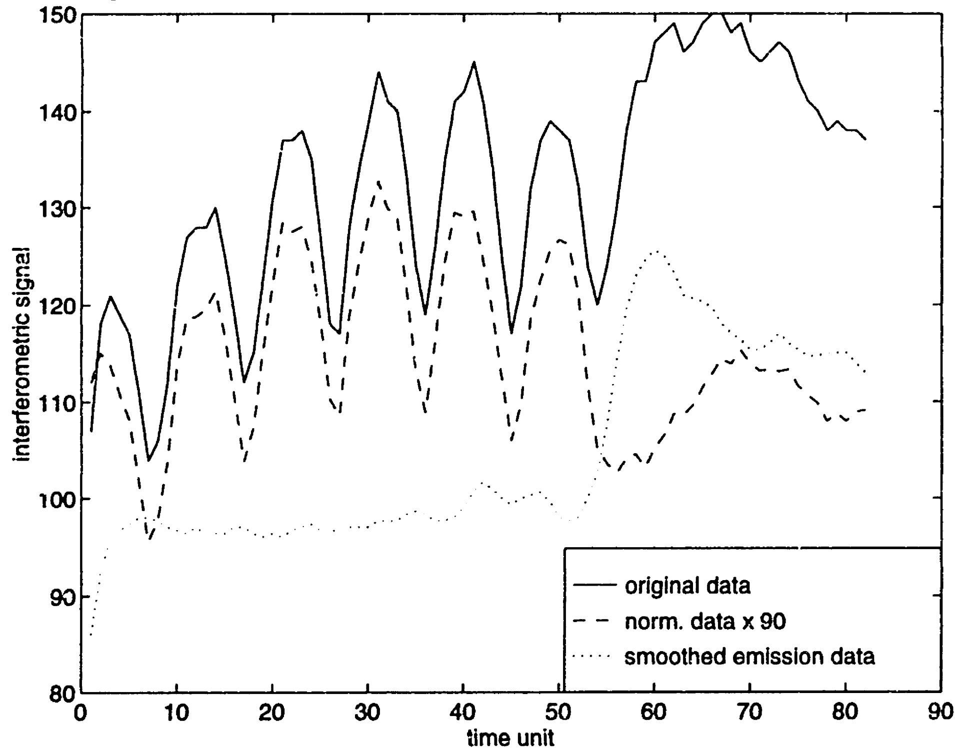


5.7.2 Endpoint prediction with phase (with emission level normalization)

When a film is cleared at the end of an etch, the plasma emission level varies because of a change in the composition of the plasma. Such a change of plasma emission has been used by optical emission spectroscopy for endpoint detection [74]. However, since the emission level starts to change even when only part of the film is clear, such a variation will distort the signals at other part of the wafer. The uncertainty of the endpoint locations in the interferometric signals are caused by the rise in emission level. As suggested in [42], the distortion of the signal can be reduced by dividing it with an estimate of the overall emission level, namely, signals obtained from an unetched site such as a wafer clamping pin. In order to reduce the noise from the emission level. Three-time-point averaging is applied to the emission level (s_n) before the division:

$$\hat{s}_n = \frac{(s_{n-1} + s_n + s_{n+1})}{3}. \qquad (5\text{-}7)$$

Figure 5-41 compares the original signal from a bare polysilicon etching area (experiment



#13 site # 15) to its normalized version. From the normalized data, endpoint can be recognized more easily. The normalized data and the predicted endpoints (*) for experiment #13 site #15 are shown in Figure 5-42. The endpoint times predicted from the original data (o) are also shown for reference. It can be noted that the predicted endpoint times determined from the original and normalized data are at most one time unit apart for the cases considered. If an etch is very non-uniform, the interferometric signals at the slowly etching area will be affected to a greater extend. However, signal normalization may not always be necessary for the following reasons:

i) An etch will not be very non-uniform;

ii) A major use of endpoint prediction is for facilitating etching chemistry switching, where a film does not clear;

iii) Division by emission signal introduces more noise to the interferometric signals.

Hence, for the rest of the analysis, the original signals will be considered instead.

Figure 5-42: Endpoint prediction for Exp. 13 using normalized data

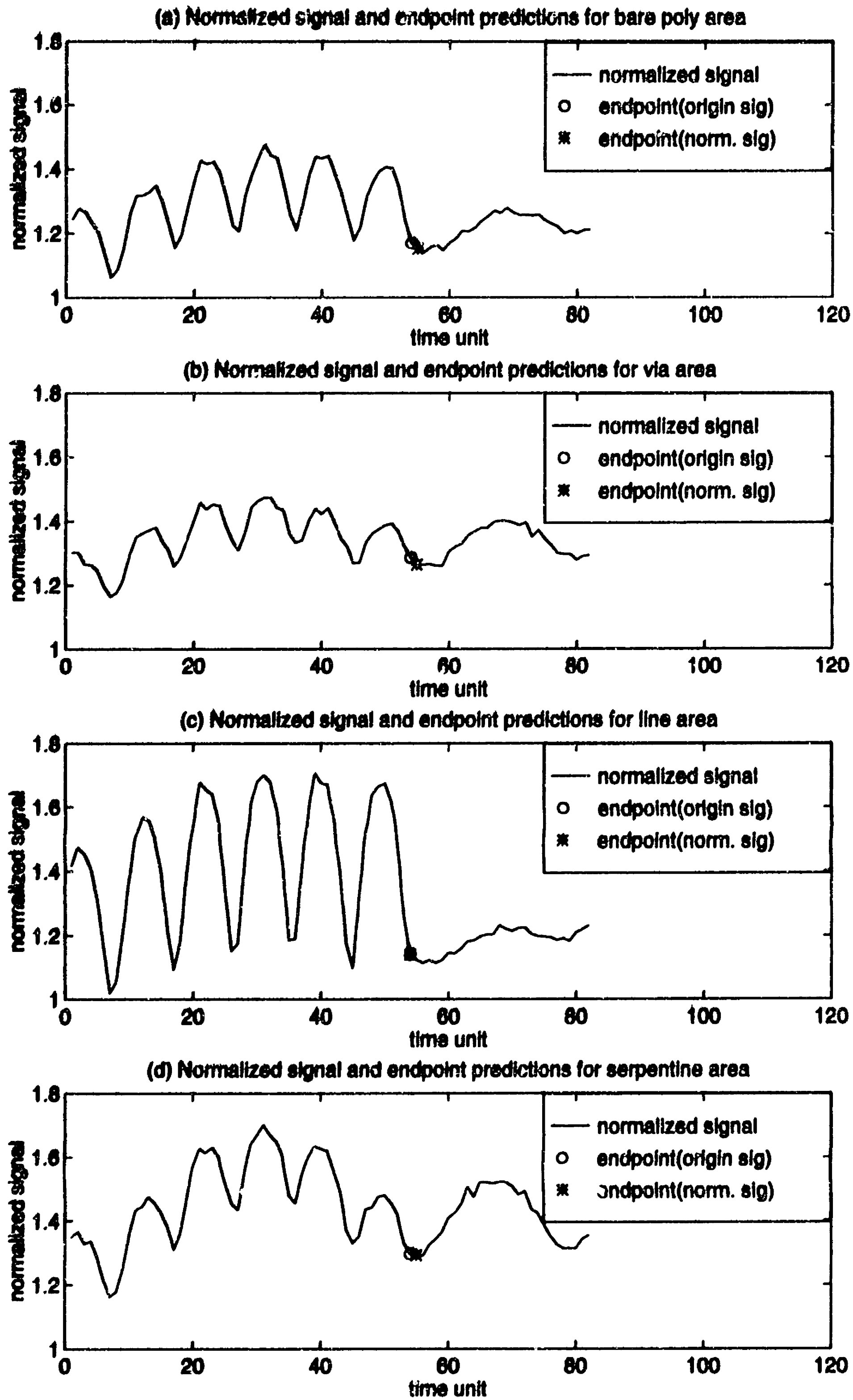
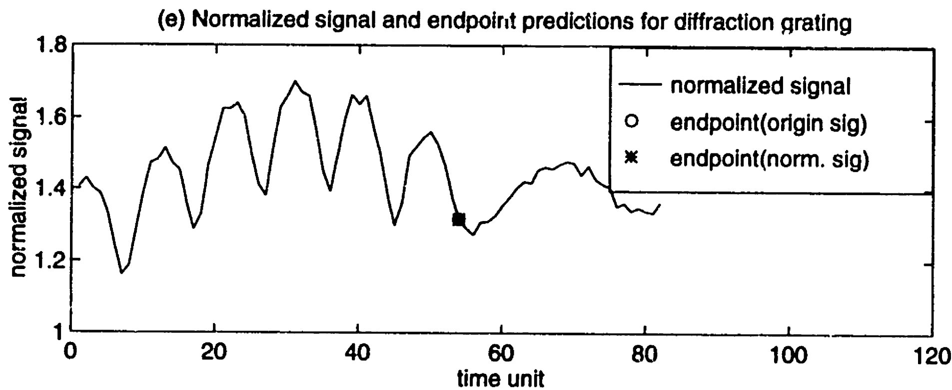


Figure 5-42: Endpoint prediction for Exp. 13 using normalized data (cont.)



5.7.3 Etching chemistry switching

Instead of stopping an etch when the etching film is cleared, a change of chemistry is sometimes desired for a more selective etch for the last few hundred angstroms of the film, e.g., Angell and Oehrlein has proposed a "flexible endpoint" method to facilitate the switching of etch chemistry [3]. For our experiments, by translating the specific thickness to the total number of signal cycles and phase angles, the switching of chemistry can be performed more consistently. For illustration purpose, it is intended to signal a chemistry-switching spot when the film thickness left is 382.65Å. This specific spot corresponds to the last maxima of the interferometric traces. Figures 5-43 and 5-44 show the selected chemistry-switching spot for the bare poly areas in site #15 for experiments #10 to #13. In these experiments, the processes were not changed so that the selected chemistry-switching spot could be observed more easily. It can be seen that the spots are detected very consistently, although the algorithm is unaware of and does not use the signals afterward the switching point.

To further verify that signal normalization by optical emission level may not be necessary, Figure 5-45 gives the predicted chemistry switching point based on both the original signals and the normalized signals of experiment #13 site #15. It can be observed that the predicted times from these two different kinds of data are at most one time unit apart.

Figure 5-43: Predicted etch switching time for bare poly areas I

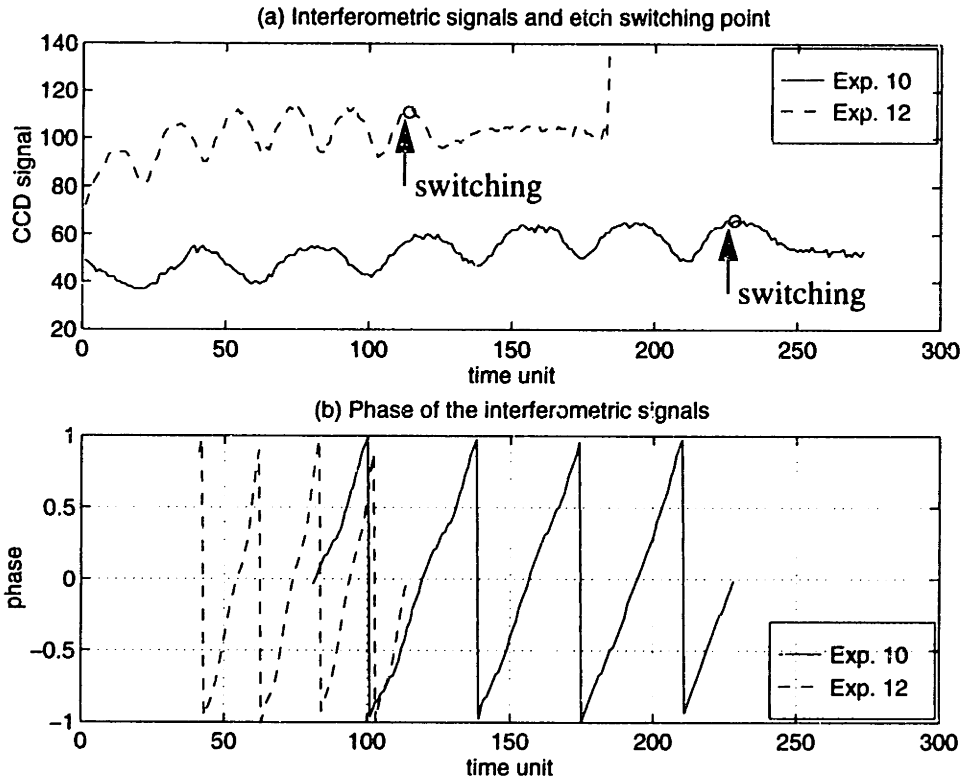


Figure 5-44: Predicted etch switching time for bare poly areas II

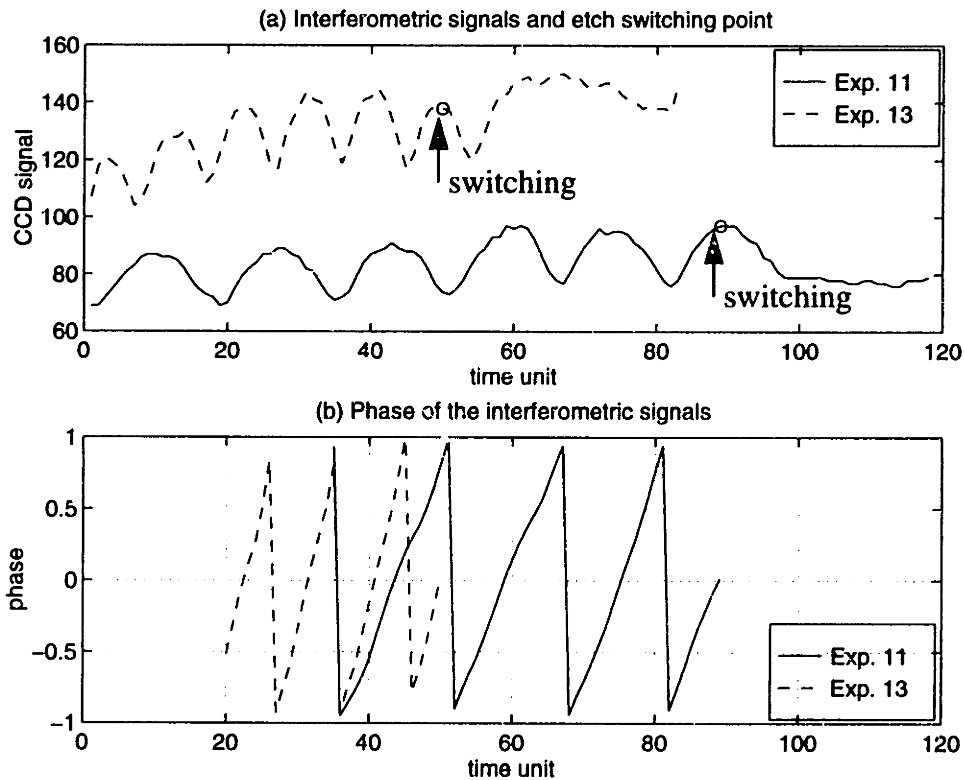


Figure 5-45: Predicted etch switching time for normalized signals

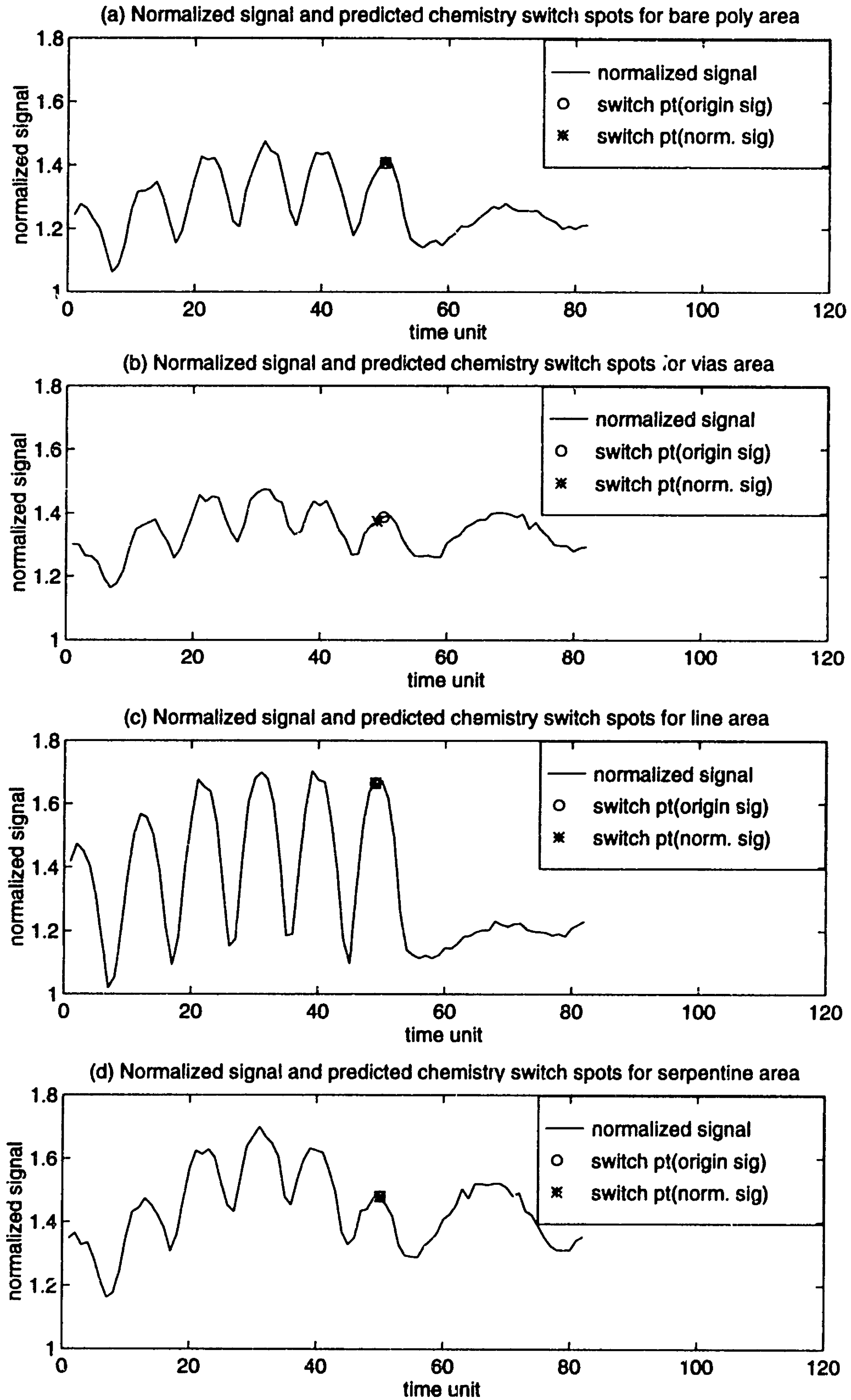
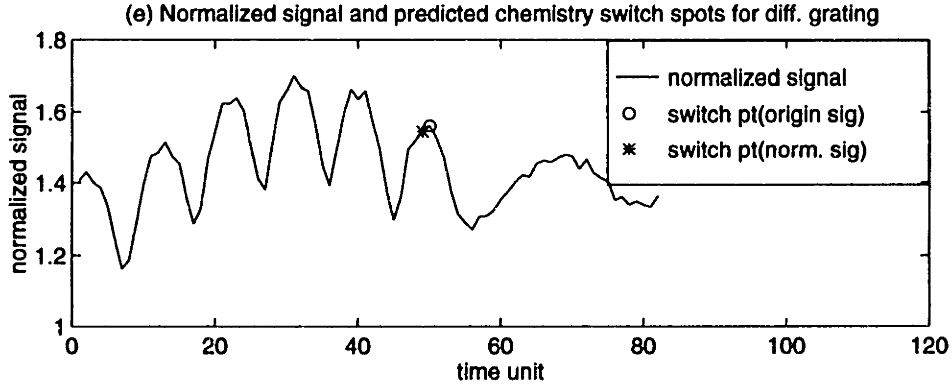


Figure 5-45: Predicted etch sw itching time for normalized signals (cont.)



5.7.4 Film thickness estimation and time-to-endpoint

By knowing the phase at endpoint, the current phase, the total number of signal minima and the signal minima that are received, the film thickness left can estimated as:

$$L_{film} = \left(cycle_{total} - cycle_{now} + \frac{(\theta_{endpoint} - \theta_{now})}{2}\right) \times \Delta \tag{5-8}$$

where L_{film} is the film thickness left, $cycle_{total}$ is the total number signal minima, $cycle_{now}$ is the number of cycle minima detected, $\theta_{endpoint}$ is the endpoint phase, θ_{now} is the current calculated phase, and Δ is the thickness of film etched per cycle of signals. In addition, given a constant etch rate, the calculated time-to-endpoint ($T_{endpoint}$) can be determined by dividing Equation 5-8 by the etch rate, which is then reduced to:

$$T_{film} = \left(cycle_{total} - cycle_{now} + \frac{(\theta_{endpoint} - \theta_{now})}{2}\right) + \omega_{etch} \tag{5-9}$$

where ω_{etch} is the angular frequency of the interferometric signals. Figure 5-46*a* shows the film thicknesses determination based on several selected signals, and Figure 5-46*b* shows the corresponding time-to-endpoint.

Since the periodicities of the interferometric signal calculated using Fast Fourier Transform method are quantized, the limited amount of data will lead to fluctuations in the frequency (or the etch rate) estimates, which will affect the endpoint time accuracy. In Figure 5-47*a*, the quantization in etch rate estimation is shown. However, by averaging the

Figure 5-46: Film thickness and time-to-endpoint estimates for line area in Site #5

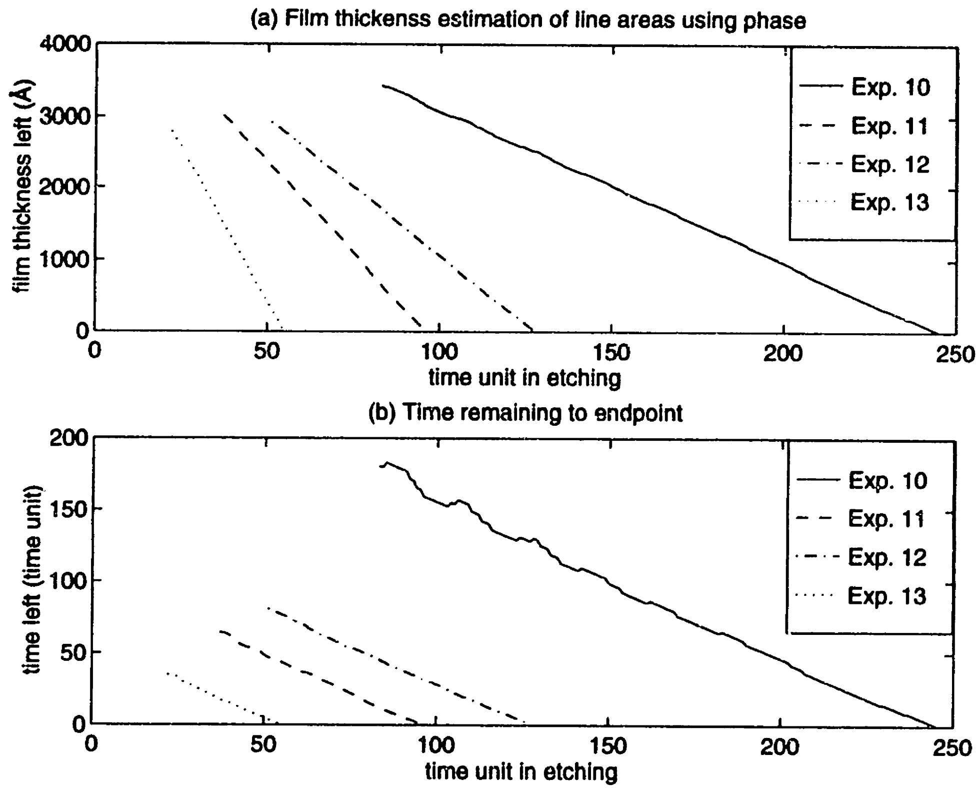
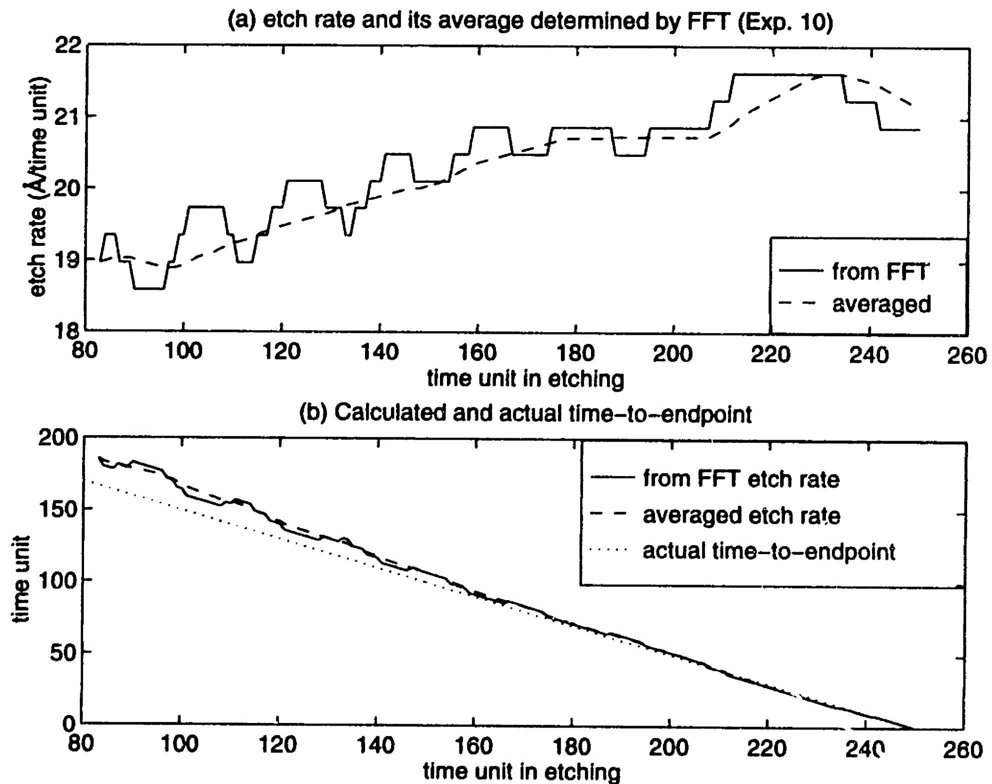


Figure 5-47: Etch rate and time-to-endpoint estimates for a drifting etch



FFT calculated frequencies over the most recent half cycle of the interferometric signals, fluctuations in etch rate estimation and time-to-endpoint can be reduced, as shown in Figures 5-47*a*,*b*. In Figure 5-47*b*, since the actual time at endpoint is known, the actual time-to-endpoint line is also plotted. The calculated time-to-endpoint assumes that the etch rate is constant. Since the etch rate is increasing, the calculated time-to-endpoint is always shown to be larger than the actual one. If the etch rate for a site is constant, than the calculated and the actual time-to-endpoint will coincide. Figure 5-48*a* shows the etch rate for a line area for site #8 from experiment #1. The etch rate is very steady despite the quantization effect due to the Fast Fourier Transform method. In Figure 5-48*b*, the calculated and actual time-to-endpoints are plotted, and are shown to be nearly the same. In Figure 5-49, the differences between the calculated and actual time-to-endpoints are found to be within the ±1 time unit bound in the constant etch rate region. Again, it can also be observed that the averaged etch rate method can provide more accurate results.

Figure 5-48: Etch rate and time-to-endpoint estimates for a "more" steady etch

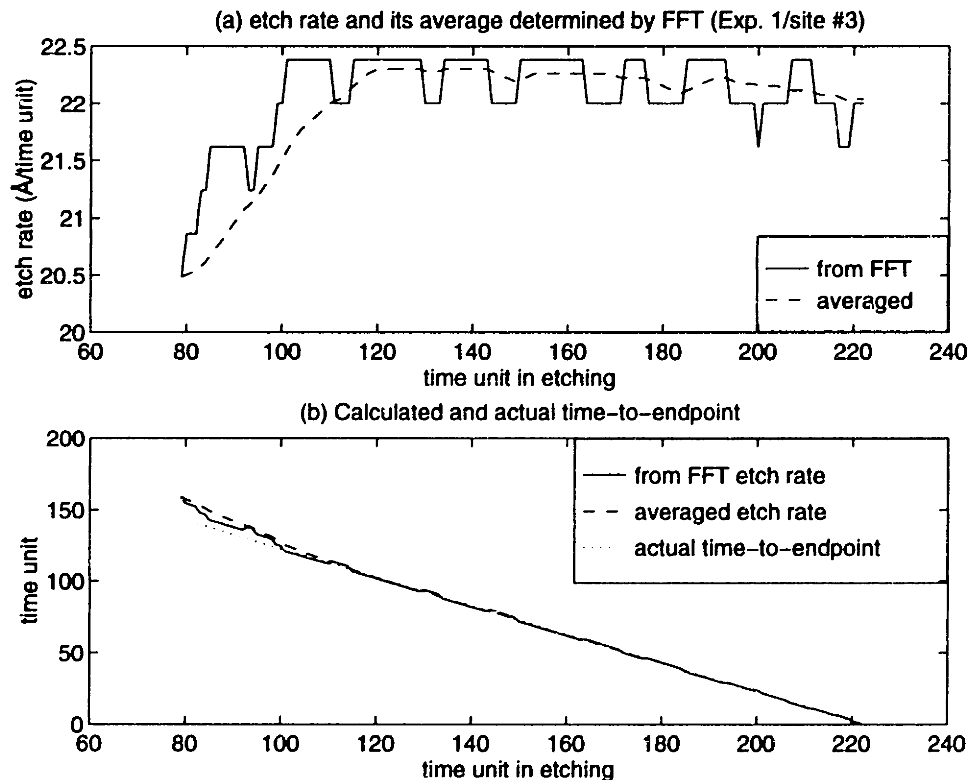
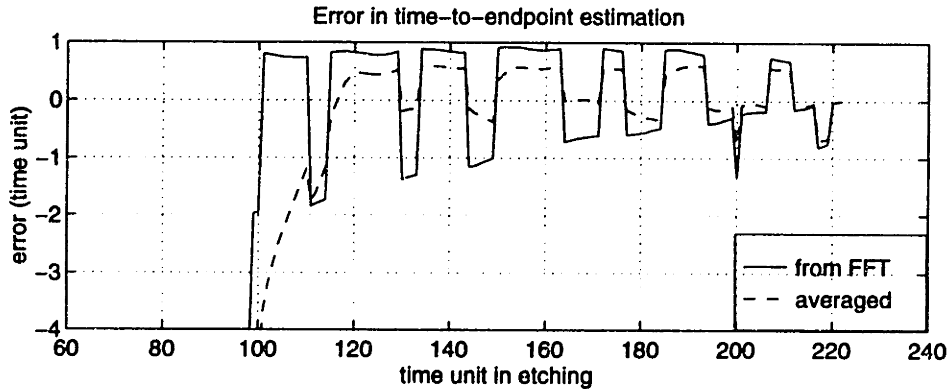


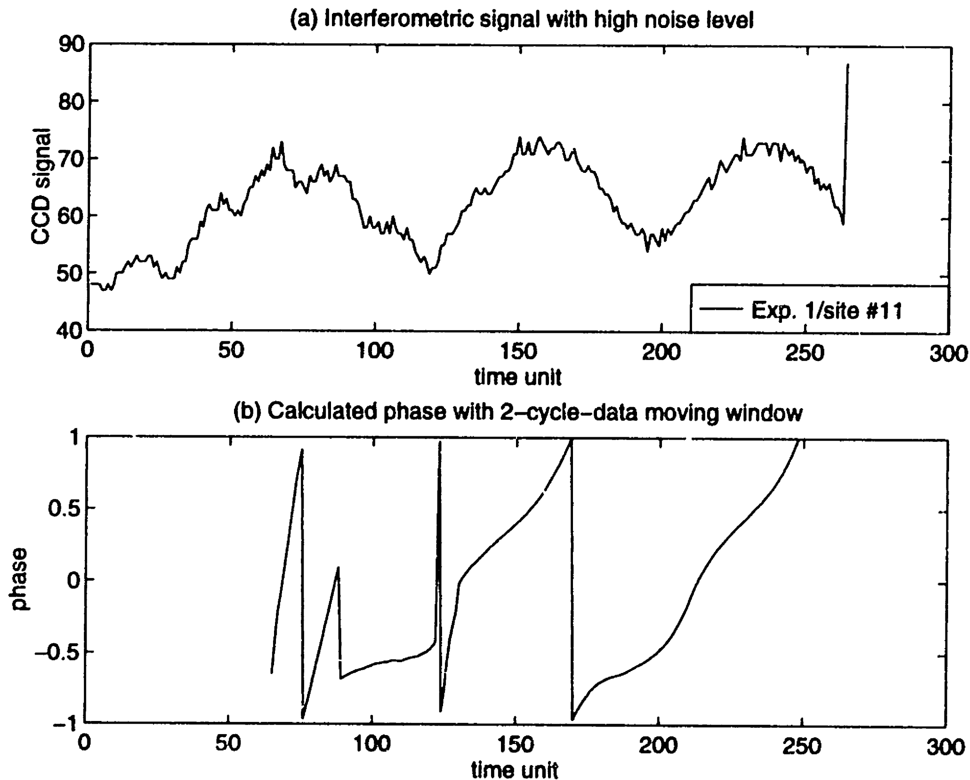
Figure 5-49: Error in time-to-endpoint estimates for a "more" steady etch



5.7.5 Modified algorithm for noisy data

In general, the signals could be contaminated with noise and the phase cannot be obtained easily. Figure 5-50*a* shows the signal from a bare polysilicon area in dice #31 from experiment #1. The signal to noise ratio is very low. In addition, there is an extra low frequency component to the signal. It is suspected the extra frequency is due to incorrect alignment of the CCD pixel with the selected sites and the photoresist effects become dominant. As a result, the phases as shown in Figure 5-50*b* are unreliable for predicting the endpoint.

Figure 5-50: Example for incorrect phase estimation using 2-cycle window



One way to improve the successful rate of phase estimation at time t is to use all the signals available up to time t instead of a sliding window of 2-cycle-signal. A coarse estimate of the signal frequency can first be obtained from the initial two cycles of data. In the subsequent estimations, the frequency for etch rate and phase calculation are then determined within a bounded range of the initial coarse estimate, which will reduce the chance of picking up wrong frequencies due to noise contamination or the photoresist effect. Since more data are used, more signal distortions can be tolerated. In Figures 5-51*a*,*b*, the phase estimations using the all-data window are plotted, which give more accurate results. However, the accuracy of the phase estimate may also be compromised if the etch drifting rate is large, as mentioned in Section 5.3. Table 5-12 gives statistics for the failed endpoint prediction rate for designed experiments listed in Table 5-9 using the all-data window. For each of the 27 experiments, 36 dice are considered, and five pattern features are examined per die; hence, there are altogether 36×27=972 signal traces for each feature. A few results using a sliding window of 2-cycle-signal are also shown in Table 5-13. By comparing these results, it can be seen that the all-data window method produces more robust endpoint prediction.

Figure 5-51: Example for correct phase estimation using all-data window

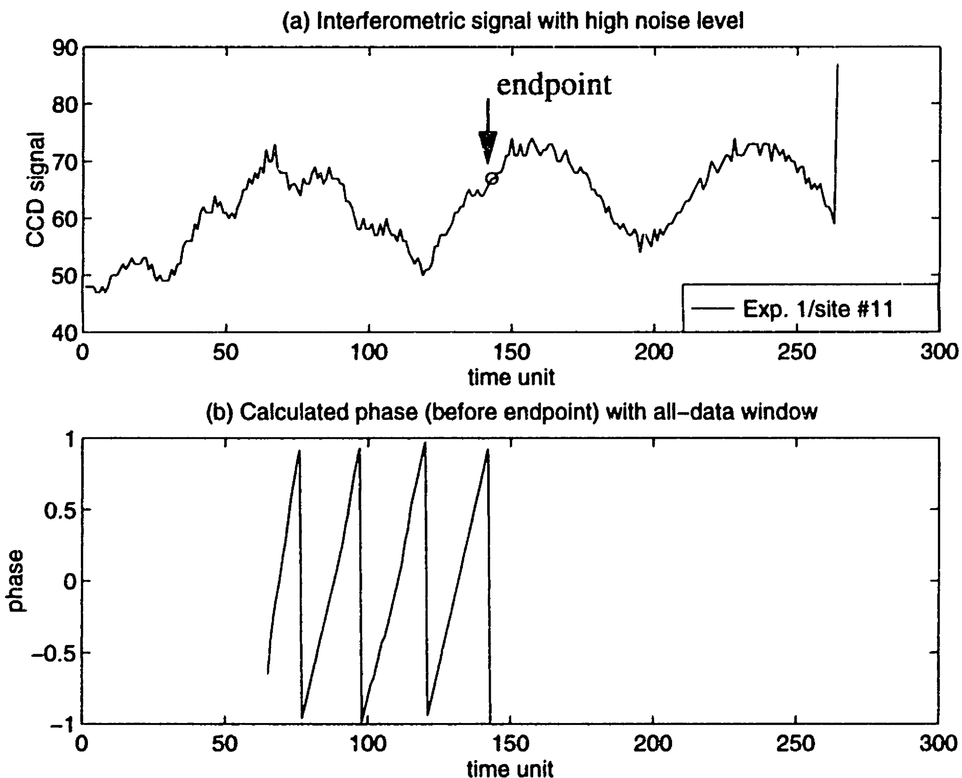


Table 5-12: Number of endpoint prediction failure using all-data window

Exp. #	bare poly	vias	lines	serpentines	dif. gratings
1	0	5	0	3	3
2	0	0	0	4	0
3	1	1	0	0	0
4	0	0	0	0	0
5	5	6	3	5	8
6	0	2	0	0	4
7	0	0	0	0	0
8	3	14	6	8	9
9	0	0	0	0	1
10	0	1	0	0	0
11	0	8	0	0	0
12	0	0	0	0	1
13	0	9	0	0	0
14	0	0	0	0	1
15	0	2	0	0	0
16	0	0	0	0	1
17	0	2	0	0	0
18	0	4	0	0	0
19	4	12	4	12	15
20	1	10	0	1	0
21	1	1	0	2	2
22	0	1	0	0	1
23	1	4	0	0	5
24	2	3	0	0	3
25	0	5	0	1	4
26	0	0	0	1	7
27	0	4	0	0	1

As an example, for the bare polysilicon area, there are 18 failed cases out of a total of 972 cases, i.e., 2%. Percent failure for other features are shown in Table 5-14. It should be noted that the selected CCD pixels could have been mis-aligned in experiments #6 and #19, and that the vias give more failed cases, which is due to the smaller ratio of etching-to-unetching film area.

Table 5-13: Number of endpoint prediction failure using 2-cycle-signal sliding window

Exp. #	bare poly	vias	lines	serpentines	dif. gratings
3	2	2	1	0	0
5	5	7	3	7	11
7	2	9	4	8	15
9	11	15	9	9	12

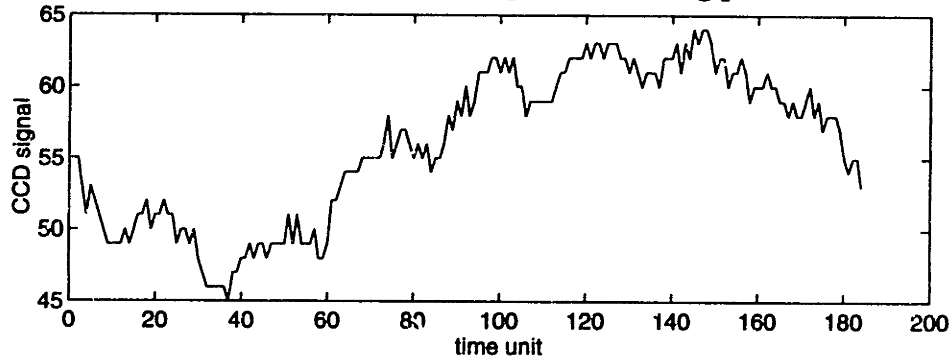
Table 5-14: Percent failure of endpoint prediction using all-data window method

	bare poly	vias	lines	serpentines	dif. gratings
total # of signals	972	972	972	972	972
# of failed cases	18	92	13	37	66
% failure	1.85	9.47	1.34	3.81	6.79

The causes for incorrect endpoint prediction even for the all-data-window method can be classified into three categories:

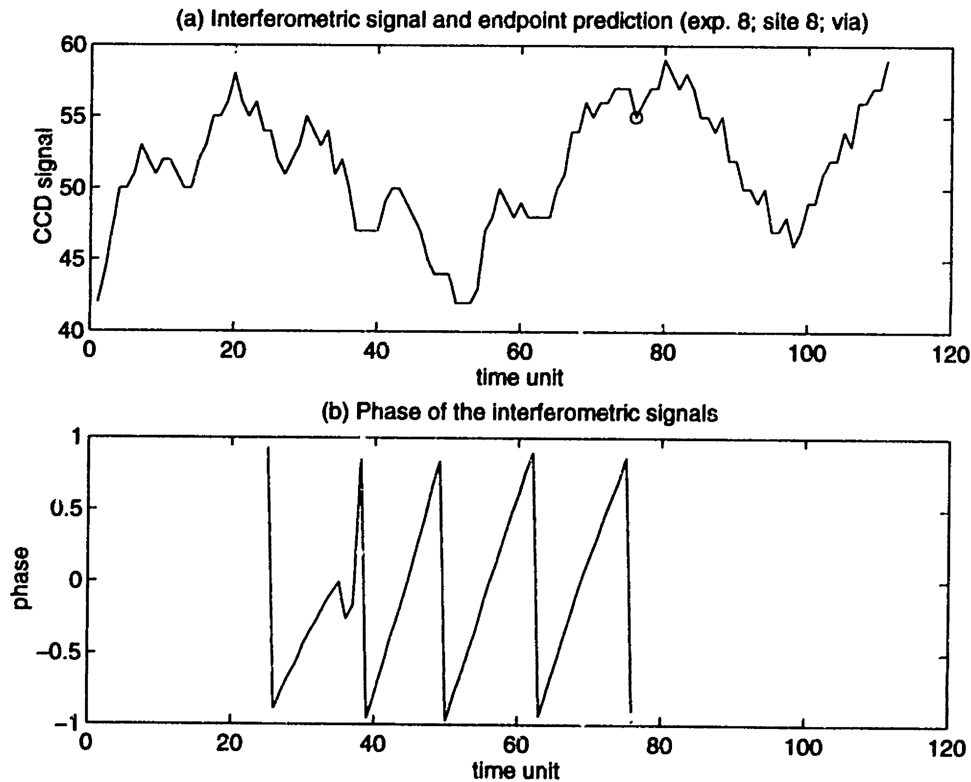
a) The photoresist interference is very large at the beginning of the etch, and its amplitude is even larger than that of the polysilicon etching signal. As a result, the number of data corresponding to the initial two cycles of etching signal, as explained in Section 5.5, cannot be found (using Fast Fourier Transform). An example is shown in Figure 5-52 for the via area signal from site #7 in experiment #5. If the effect due to the photoresist if similar to just adding a linear trend to the etching signals, endpoint prediction will not be affected.

Figure 5-52: Incorrect initial cycle counting due to strong photoresist interference



b) The initial two cycles of etching signals are detected, but the phase calculation is distorted by the photoresist etching signals. As a result, the phase is no longer an increasing function (except when it is wrapped around). An example is illustrated in Figure 5-53. Although the film clearing point can still be detected, this case is regarded as a failure because the film thickness (as a function of the phase) has been incorrectly determined during part of the etch.

Figure 5-53: Incorrect phase estimation due to photoresist interference



In general, random sensor noise on the signal will not affect the signal tremendously because they will be filtered in the phase calculation. If better signal processing algorithms are available to obtain the phase of the etching film independent of the photoresist effect, endpoint prediction failure can be reduced. In addition, the signal could be so noisy that the signal periodicity cannot be recognized. This is more common for the dice close to the edge of the wafer, and not for the inner thirty-six dice considered in this section.

5.7.6 Etch rate drifting

Etch rate drifting was shown to affect the accuracy of phase estimation. In addition, the estimation error will increase as the number of data cycles used for the calculation increases. Although reducing the size of the data window will reduce the magnitude of error, it was shown in the previous section that a large data window size will improve the robustness of phase and film thickness estimation. In this section, the amount of etch rate drifting is determined from the experiments. From the 27-experiment design of experiments, dice #15,16,21,22 are considered, because they are at the center of the wafer and have a good signal-to-noise ratio. Using the endpoint prediction algorithm from Section 5.7.3, endpoints for the five patterns in each of the four dice are obtained. Then for each signal, a ratio of the etch rates for the second half of the etching versus the first half of the etching is obtained. Figures 5-54(*a-e*) are histograms for the etch rate ratio for the bare polysilicon area for the twenty-seven experiments. From these histograms, it is shown that the etch rate drifting is about 10%. However, since the ratio is on the averaged etch rates between the first and second half of the etch, the actual drifting bounds should be doubled, i.e., about ±20%. In Table 5-15, the means and standard deviations of the etch rate ratios are given. Since the etch drifting rate is small compared with the cases analyzed in Section 5.3, the use of the all-data-window in the previous section for film thickness estimation and endpoint prediction is more justified.

Figure 5-54: Ratio of the second half etch to initial half etch for different areas

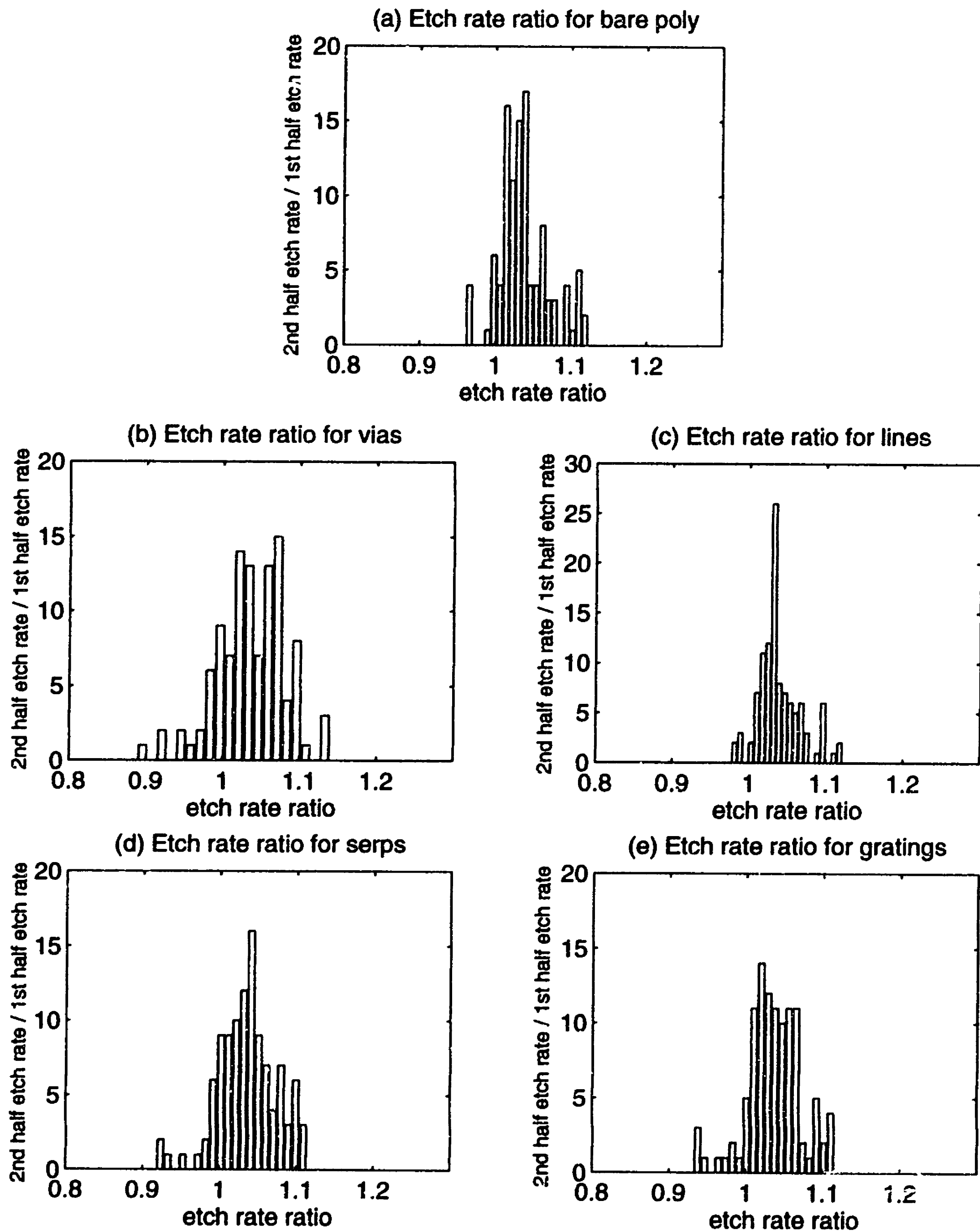


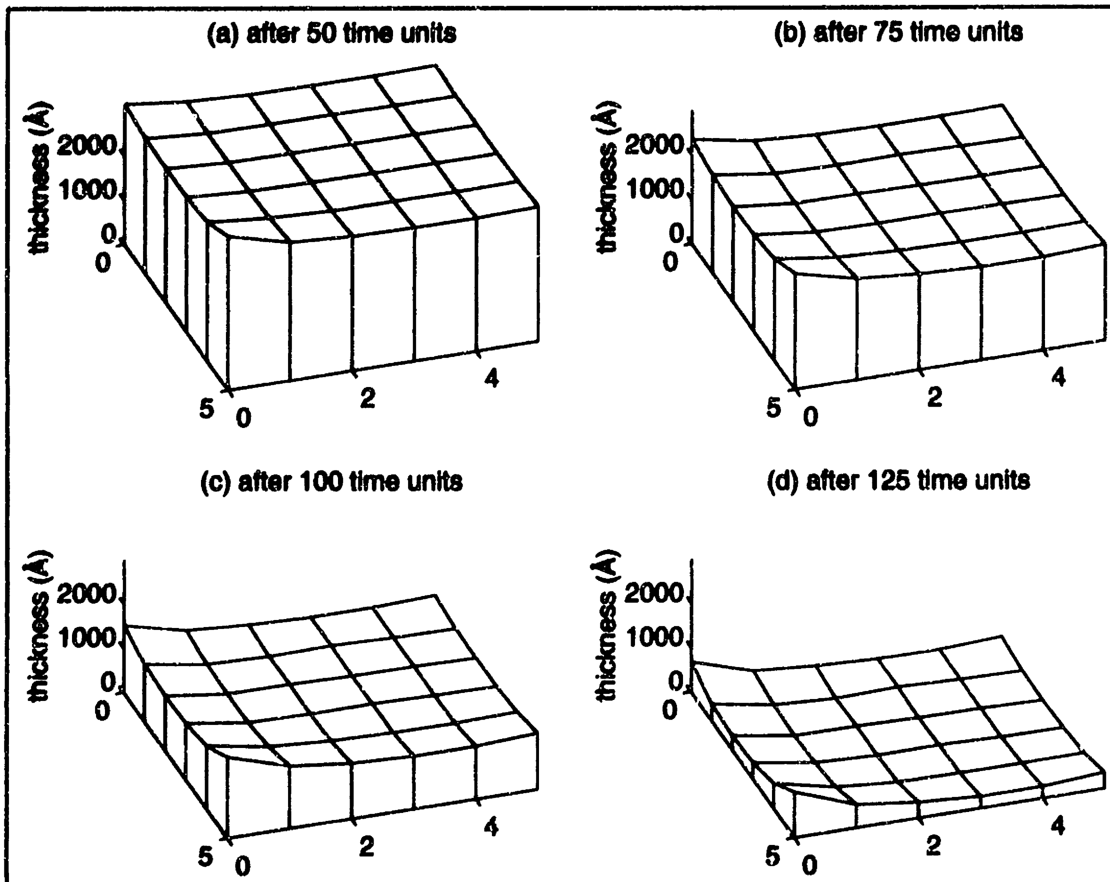
Table 5-15: Statistic of etch rate ratio

	Bare Poly	Vias	Lines	Serpentines	Gratings
mean	1.0378	1.0361	1.0388	1.0359	1.0361
std. dev.	0.0339	0.0441	0.0282	0.0386	0.0359

5.8 Full wafer endpoint prediction

Because not all the signals have a high signal-to-noise ratio which allows accurate estimation of the phase, it will be more fruitful to obtain phase and endpoint estimation from several sites from a wafer and rely on those that give accurate phase estimations. As a result, a full wafer endpoint prediction algorithm is developed. This algorithm is very similar to the single interferometric signal endpoint algorithm, except that it checks if the signals are strong enough for accurate cycle counting and phase estimation. As with the single interferometric signal algorithm, the data required for two cycles of data are first extracted from the multiple sites (e.g., a total of M sites). Since it is assumed that not all signals are good enough, it would be more robust to require that only M-1 (or M-2) of this 2-cycle-data information is required. Then, as more data are received, either the 2-cycle-signal sliding window, or all-data window, will be used for determining the phase. During the phase estimation, the phase must be non-decreasing except when it is wrapped around. It is reason-

Figure 5-55: Full wafer film thickness estimation

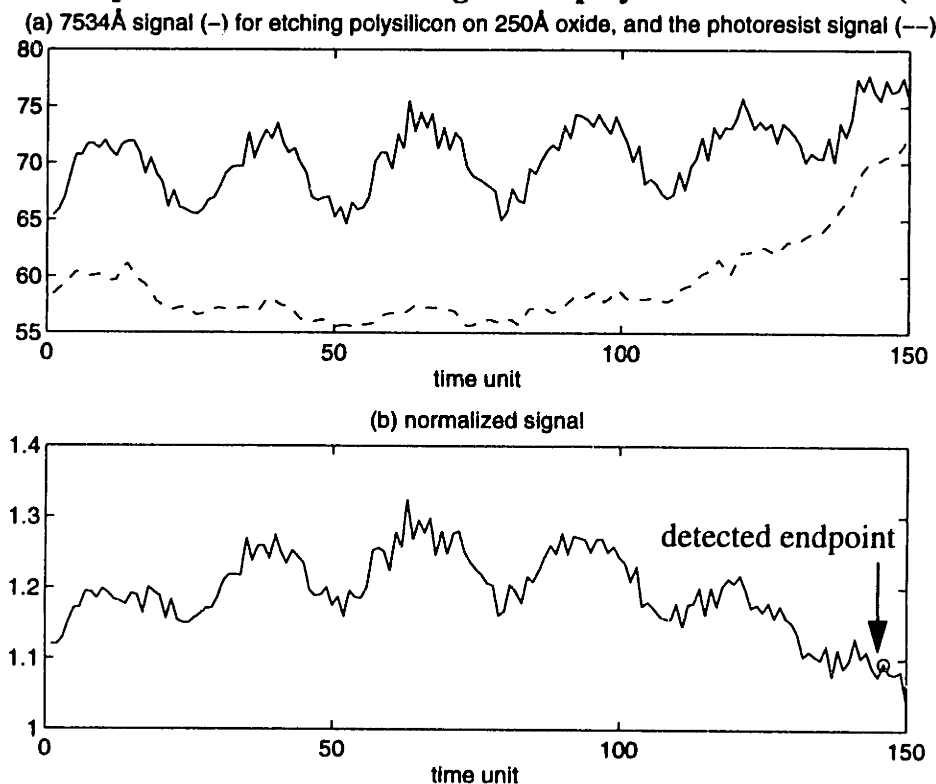


able to assume that if the phase change is negative, the specific interferometric signal is unsuitable for providing valid phase estimation, and should be neglected for reliable thickness estimation or endpoint prediction. Hence, the etch should be stopped or chemistry should be switched based on information provided by the valid interferometric signals. Figures 5-55*a*,*b*,*c*,*d* show the film thickness for the line areas for experiment #11. The film thicknesses are determined using the all-data window. It can be observed that the etching is not very uniform using the current control setting.

5.9 Endpoint prediction versus endpoint detection

In Chapter 4, algorithms for endpoint detection based on the interferometric signals have been examined. It was found that endpointing could be delayed if the sensor signals contain high levels of noise. Given the availability of endpoint prediction, one alternative can be stopping an etch when the predicted film thickness is zero, followed by a short final etch to ensure film clearing. Figure 4-6 is repeated in Figure 5-56 for reference. This signal is very noisy and the detection of endpoint using the model deviation approach was delayed.

Figure 5-56: Endpoint detection of etching 5000Å poly over 1000Å oxide (Model dev.)



In Figure 5-57, the ideal reflectance is shown. The normalized phase at endpoint is found to be at -0.7041 using the Fast Fourier Transform. Figure 5-58 indicates the endpoint predicted using the endpoint phase based on the original data. It can be seen that accurate endpoint is predicted instantaneously. Figure 5-59 indicates the endpoint predicted based on the normalized data. The same endpoint time is concluded based on the original data or the normalized data. Compared to Figure 5-56, the predicted endpoint was found more quickly, indicating that the prediction algorithm is very robust even for very noisy signals, and may be preferred over endpoint detection for use in plasma etch processes.

Figure 5-57: Reflectance of etching 5000Å polysilicon over 250Å oxide

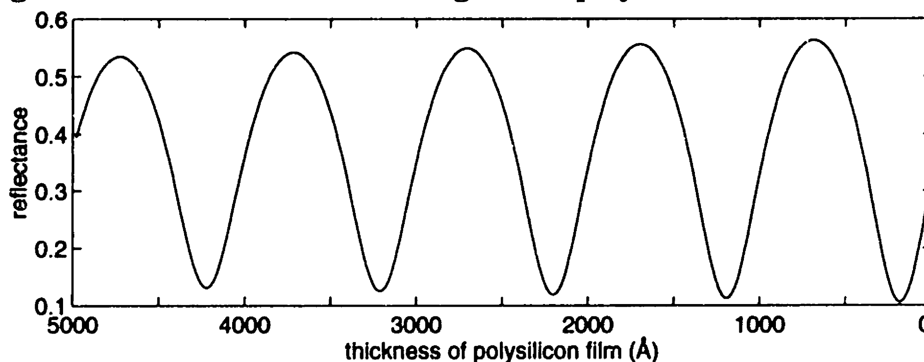


Figure 5-58: Predicted endpoint for etching 5000Å polysilicon over 250Å oxide

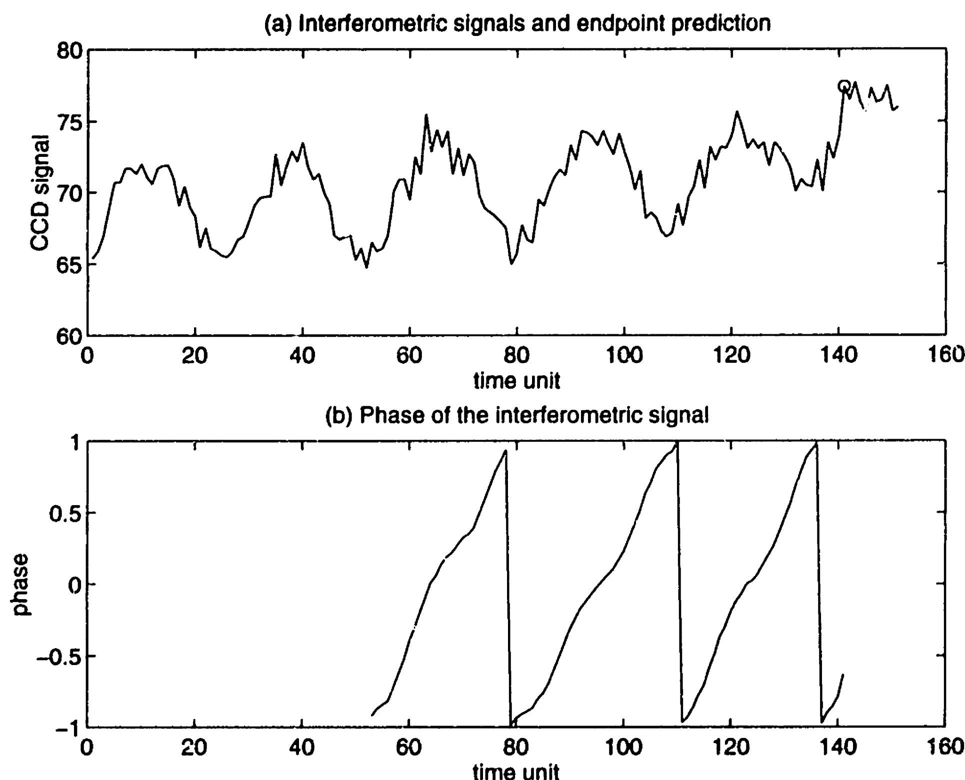
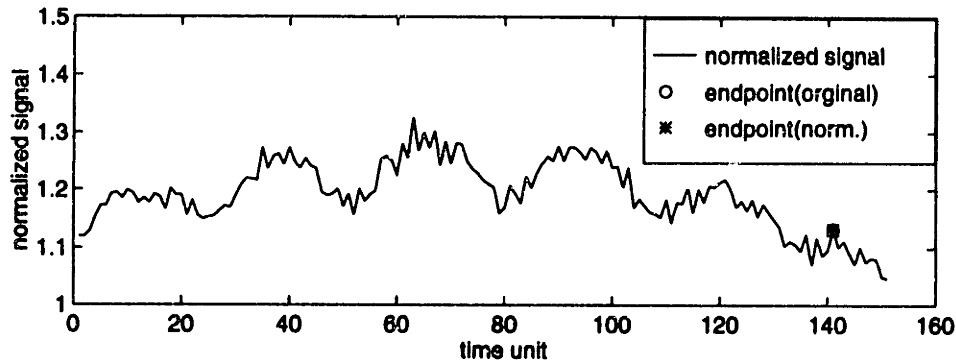


Figure 5-59: Normalized signal and predicted endpoint for etching 5000Å polysilicon over 250Å oxide



5.10 Summary

In this chapter, the use of phase is proposed as a measure of film thicknesses. Given the phase at endpoint and the refractive index of the etching film, the thickness of the film left during etch can be estimated; such endpoint prediction can be used for chemistry switching in an etch process. Sensitivity analysis has been performed with respect to drifting etch rate and small variations in film structure. It has been found that short wavelength signals produce more robust results. The proposed algorithm has been verified with experimental results. It has been found that it can be accurate to about 100Å in terms of polysilicon etch film thickness estimation. The robustness and limitation of using phase to signal etch chemistry switching time or the film clearing time have been examined. Compared to endpoint detection, endpoint prediction using the signal phase may indeed have significantly better robustness accuracy.

Chapter 6

Conclusions and future work

6.1 Conclusions

Plasma etching is an increasingly important step in the fabrication of deep-submicron microelectronic circuits. Motivated by the need to improve the consistency and uniformity of the process, various types of run-by-run and real-time control systems are being developed. Recently, full wafer interferometry has been developed to facilitate spatial monitoring of an etching wafer in real-time. This thesis focuses on obtaining accurate extraction of etch rate, efficient detection of endpoint of the process, and endpoint prediction (or film thickness estimation) based on the interferometric signals from the sensor.

Accurate etch rate information is useful for run-by-run improvement of an etch, and for real-time adjustment of the equipment parameters. The etch rate can be determined from the periodicity of the sensor signal readily, and four algorithms (the Fourier Transform, model matching, Kumaresan-Prony, and multiple signal classification methods) are examined for their accuracy, with respect to different levels of noise in the signals, different lengths of data available, and different sampling rates. It is found that despite its simplicity, the Fourier Transform method is able to produce estimates with accuracy comparable to the more sophisticated methods.

Endpoint detection using the signals from full wafer interferometry is also examined. Such information is essential for minimizing the over-etch of the underlying layer. Two endpoint detection algorithms have been developed. The first is applicable for etching non-absorptive films. It takes advantage of the signal deviation from its periodic pattern after endpoint. A model for fitting the periodic etching signal is proposed, the parameters of which can be determined with few computations. The second model is applicable for etching films with high refractive indices. If the etch has a relatively high selectivity, a flat slope in the signal will suggest endpoint. In both algorithms, the signal phase and the total number of signal cycles received are used to reduce false endpoint alarms.

By using the phase information determined from the interferometric signals, and a

prior value of phase at endpoint, the approximate film thickness left during an etch can be determined in real-time. Simulations have been performed to examine the sensitivity of such a measure of film thickness with respect to small drifting in the etch rate and small variations in the film stack structure. It was found that estimates determined from short wavelength signals are more robust against different types of variations. Although simulation results indicate that using long data length for phase computation may produce larger errors in face of large etch rate drift, experimental data show that the etch rate varies less that ±25 %, and the use of all possible interferometric data for phase calculation is more robust than using only a 2-cycle moving data window. Using the interferometric signals from the full wafer interferometer, accurate time-to-endpoint or thickness-to-endpoint can be estimated for appropriate etch chemistry switching or etch stopping. Experimental data show that the accuracy of film thickness estimation is about ±100 Å for the etch of polysilicon.

6.2 Future work

The work of this thesis on etch rate estimation, endpoint detection and endpoint prediction will facilitate the use of full wafer interferometry for real-time assessment and control of single-wafer plasma etching process. However, in order to fully utilize the potential of the sensor, further improvements in the development of the sensor and the control algorithms are required. A critical procedure in obtaining data for analysis is to align the pixels of the CCD camera to the appropriate areas on the wafer. Without correct alignment, a pixel may pick up signal from the photoresist region and lead to incorrect estimates. The re-orientation is currently performed manually by using the scribe lines as references. In the future, a computer pattern recognition algorithm should be developed for full automation of the process.

In Chapter 5, it was found that increasing the data sampling rate can improve the accuracy of etch rate estimation. However, when a fast sampling rate is not synchronized with the rotating magnetic field, the signal will be contaminated with extra high frequency components. Dedicated signal processing algorithms should be developed to filter out the unwanted signal for extracting useful information in real-time.

As the dimension of devices continues to shrink, the percentage of etching area over a wafer surface may decrease, and the full wafer interferometric signals may suffer from low signal-to-noise ratio. Algorithms that map out the etch profile, even if only a few strong signals are available, should be developed. On the other hand, if the computation of signals from a lot of pixels is feasible, efficient algorithms should then be developed to extract the useful wafer level variation.

The various algorithms investigated take advantage of the signals from only a single wavelength. If signals from more wavelengths can be analyzed, their estimates could be combined to produce more accurate results. In addition, for very thin films, the signals from an etch may be very short, and there are not enough cycles of data for accurate phase estimation or curve fitting. However, if the etch is selective and the etching film has a higher refractive index than the underlying layer, simultaneous "flat" signals in different wavelengths may be used for endpoint detection.

Only the transparent film has been examined in this work. However, full wafer interferometry can also be extended to metal etch. For example, as a metal layer is being etched, the distance between the slowly etching photoresist surface and the etching metal surface gradually increases. Light reflection from these surfaces of different heights will result in lateral interference, which can be examined for relative etch rate determination. In addition, when a metal layer is cleared to expose the underlying transparent film layer, the signal content may change suddenly. Such a change in the signal should be explored for endpoint detection.

In terms of control algorithms, the emphasis of the thesis has been on extracting various measures of the performance of an etch for simple real-time control. This information should provide more improvement for an etch if combined with more sophisticated control strategies. For example, given the film thickness measure, the chemistry could be switched for the final etch when the mean thickness of the film is at a predefined threshold value, e.g., 500Å. On the other hand, the threshold value can also be determined either as a function of the spatial uniformity of etch rate or film topology in real-time, or it can be modified according to a run-by-run algorithm. Finally, an effective strategy that integrates real-time control and run-by-run control to reduce process variation at different stages of the control system will significantly improve the quality of process outputs.

Bibliography

[1] H. Alius, and R. Schmidt, "Interference method for monitoring the refractive index and the thickness of transparent films during deposition," *Rev. Sci. Instrum.*, vol. 61, no. 4, pp. 1200-1203, 1990.

[2] R. L. Allen, R. Moore, and M. Whelan, "Application of neural networks to plasma etch end point detection," *J. Vac. Sci. Technol. B*, vol. 14(1) pp. 498-503, 1996.

[3] D. Angell, and G. S. Oehrlein, "Grazing angle optical emission interferometry for end-point detection," Appl. Phys. Lett., vol. 58, no. 3, pp. 240-242, 1991.

[4] P. Banks, W. Pilz, I. Hussla, G. Lorenz, and G. Castrischer, "In-situ diagnostics for plasma processing, in *Proc. SPIE: Monitoring and Control of Plasma-Enhanced Processing of Semiconductors*, vol. 1037, pp. 35-42, 1988.

[5] G. G. Barna, D. W. Buck, S. W. Butler, M. Splichal, H. Anderson, D. White, "Sensor information from a Lam 9600 metal etch process," in *Proc. Electrochem. Soc.: Process Control, Diagnostics, and Modeling in Semiconductor Manufacturing*, vol. 95-2, pp. 306-326, 1995.

[6] D. S. Boning, J. L. Claman, K. S. Wong, T. J. Dalton, and H. H. Sawin, "Plasma etch endpoint via interferometric imaging," *Proc. American Control Conf.: Invited Session on Modeling and Control in Microelectronics Processing*, 1994.

[7] D. S. Boning, S. Ha, and E. M. Sachs, "On-line control of uniformity in single-wafer plasma etch processes," *Extended Abstract, TECHCON 93*, pp. 19-21, Semiconductor Research Corporation, Atlanta, GA, Sept. 1993.

[8] F. Bose, R. Patrick, and H. Baltes, "Measurement of discharge impedance for dry etch process control," in *Proc. SPIE*, vol. 2336, pp. 101-110, 1994.

[9] G. E. P. Box, and N. R. Draper, *Evolutionary Operation: A Statistical Method for Process Improvement*. New York: John Wiley & Sons, 1969.

[10] G. E. P. Box, and N. R. Draper, *Empirical Model-Building and Response Surfaces*. New York: John Wiley & Sons, 1987.

[11] S. Bushman, T. F. Edgar, I. Trachtenberg, and N. Williams, "Process monitoring and control of plasma etch systems," *unpublished note*, 1994.

[12] S. W. Butler, K. J. McLaughlin, T. F. Edgar, and I. Trachtenberg, "Development of techniques for real-time monitoring and control in plasma etching II," *J. Electrochem.* Soc., vol. 138, no. 9, pp. 2727-2735, 1991.

[13] S. W. Butler, and J. Stefani, "Application of predictor corrector control to polysilicon gate etching," in *Proc. American Control Conf.*, pp. 3003-3007, 1993.

[14] J. Cheng, X. Qian, K. Irani, H. Etemad, and M. Elta, "Expert system and process optimization techniques for real-time monitoring and control of plasma processes," in *Proc. SPIE: Advance Techniques for Integrated Circuit Processing*, vol. 1392, pp. 373-384, 1990.

[15] J. L. Claman, *Real time interferometric imaging of the plasma etch process*. M. Eng. Thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 1994.

[16] T. J. Dalton, W. T. Conner, and H. H. Sawin, "Interferometric real-time measurement of uniformity for plasma etching," *J. Electrochem. Soc.*, vol. 141, no. 7, pp. 1893-1900, 1994.

[17] T. J. Dalton, *Pattern dependencies in the plasma etching of polysilicon*. Ph.D. Thesis, Massachusetts Institute of Technology, Department of Chemical Engineering, 1994.

[18] R. Deaton, and A. George, "Quantification of laser interference fringes as applied to plasma etch endpoint detection," in *Proc. SPIE: Lasers in Microlithography*, vol. 774, pp. 162-171, 1987.

[19] V. G. I. Deshmukh, D. A. O. Hope, T. I. Cox, and A. J. Hydes, "Application of adaptive network theory to dry etch monitoring and control," in *Proc. SPIE: Advance Techniques for Integrated Circuit Processing*, vol. 1392, pp. 352-360, 1990.

[20] M. Elta, H. Etemad, J. S. Freudenberg, M. D. Files, J. W. Grizzle, P. T. Kabamba, P. P. Khargonekar, S. Lafortune, S. M. Meerkov, J. R. Moyne, B. A. Rashap, D. Teneketzis, and F. L. Terry, Jr., "Application of control to semiconductor manufacturing: reactive ion etching," in *Proc. American Control Conf.*, pp. 2990-2997, 1993.

[21] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*. California: Academic Press, 1981.

[22] A. E. Gower, *An architecture for flexible distributed experimentation and control with an AME 5000 plasma etcher*. M.S. Thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 1996.

[23] R. S. Guo, and E. M. Sachs, "Modeling, optimization, and control of spatial uniformity in manufacturing processes," *IEEE Trans. Semi. Manufac.*, vol. 6, no. 1, pp. 41-57, 1993.

[24] S. Ha, *On-line control of process uniformity using categorized variabilities*. Ph.D. Thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering, 1993.

[25] M. A. Hanratty, A. P. Paranjpe, S. A. Henck, and B. Jucha, "Interferometric monitoring and control of silicon incorporation in the diffusion enhanced silylated resist process," in *Proc. SPIE: Dry Etch Technology*, vol. 1593, pp. 67-78, 1991.

[26] M. Haverlag, G. M. W. Kroesen, C. J. H. de Zeeuw, Y. Creyghton, T. H. J. Bisschops, T. J. de Hoog, "In situ ellipsometry during plasma etching of SiO_2 films on Si," *J. Vac. Sci. Technol. B*, vol. 7, no. 3, pp. 529-533, 1989.

[27] M. Haverlag, and G. S. Oehrlein, "In situ ellipsometry and reflectometry during etching of patterned surfaces: experiments and simulations," *J. Vac. Sci. Technol. B*, vol. 10, no. 6, pp. 2412-2418, 1992.

[28] S. Haykin, *Neural Networks: A comprehensive Foundation*. New York: Macmillan, 1994.

[29] O S. Heavens, *Optical Properties of Thin Solid Films*. New York: Dover Publications, 1995.

[30] E. Hecht, and A. Zajac, *Optics*. Massachusetts: Addison Wesley, 1987.

[31] P. A. Heimann and R. J. Schutz ,"Optical etch-rate monitoring; Computer simulation of reflectance," *J. Electrochem. Soc.*, vol. 131, no. 4, pp. 881-885, 1984.

[32] P. A. Heimann, "Optical etch-rate monitoring using active device areas: lateral interference effects," *J. Electrochem. Soc.*, vol. 132, no. 8, pp. 2003-2006, 1985.

[33] S. A. Henck, "In situ real-time ellipsometry for film thickness measurement and control," *J. Vac. Sci. Technol. A*, vol. 10, no. 4, pp. 934-938, 1992.

[34] S. A. Henck, W. M. Duncan, L. M. Lowenstein, and S. W. Butler, "In situ spectral ellipsometry for real-time thickness measurement: etching multilayer stacks," *J. Vac. Sci. Technol. A*, vol. 11, no. 4, pp. 1179-1185, 1993.

[35] N. Hershkowitz, and H. L. Maynard, "Plasma characterization and process control diagnostics," *J. Vac. Sci. Technol. A*, vol. 11(4) pp. 1172-1178, 1992.

[36] C. D. Himmel, and G. S. May, "Advantages of plasma etching modeling using neural networks over statistical techniques," *IEEE Trans. Semi. Manufac.*, vol. 6, no. 2, pp. 103-112, 1993.

[37] A. K. Hu, *An optimal bayesian controller for flexible manufacturing processes*, Ph.D. Thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 1992.

[38] Y. L. Huang, T. F. Edgar, D. M. Himmelblau, and I. Trachtenberg, "Constructing a reliable neural network model for a plasma etching process using limited experimental data," *IEEE trans. Semi. Manufac.*, vol. 7, no. 3, pp. 333-344, 1994.

[39] B. Kim, and G. S. May, "An optimal neural network process model for plasma etching," *IEEE Trans. Semi. Manufac.*, vol 7, no. 1, pp. 12-21, 1994.

[40] R. Kumaresan and D. W. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zeros modeling in noise," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 30, pp. 833-840, 1982.

[41] S. F. Lee, and C. J. Spanos, "Prediction of wafer state after plasma processing using real-time tool data," *IEEE Trans. Semi. Manufac.*, vol. 8, no. 3, pp. 252-261, 1995.

[42] Low Entropy Systems, Inc., *1000-IS Full Wafer Imaging Interferometer Manual*, 1995.

[43] P. J. Marcoux, and P. D. Foo, "Methods of end point detection for plasma etching," *Solid State Technology*, April, pp. 115-122, 1981.

[44] S. L. Marple, *Digital Spectral Analysis with Applications*, New Jersey: Prentice Hall, 1987.

[45] *MATLAB User's Guide*, The MathWorks, Inc., Natick, MA, 1993.

[46] H. Maynard, E. Rietman, J. T. C. Lee, and D. Ibbotson, "Plasma etching endpointing by monitoring RF power systems with an artificial neural network," in *Proc. Electrochem. Soc.: Process Control, Diagnostics, and Modeling in Semiconductor Manufacturing*, vol. 95-2, pp. 189-207, 1995.

[47] M. McGeown, K. I. Arshak, and E. Murphy, "Automated process control for plasma etch," in *Proc. SPIE: Integrated Circuit Metrology, Inspection, and Process Control VI*, vol. 1673, pp. 629-639, 1992.

[48] H. L. Mishurda, and N. Hershkowitz, "Instantaneous etch rate measurement of thin transparent films by interferometry for use in an algorithm to control a plasma etcher," in *Proc. SPIE: Advanced Techniques for Integrated Circuit Processing*, vol. 1392, pp. 563-569, 1990.

[49] D. C. Montgomery, *Introduction to Statistical Quality Control*. New York: Wiley, 2nd edition, 1990.

[50] D. C. Montgomery, *Design and Analysis of Experiments*. New York: John Wiley, 3rd edition, 1991.

[51] P. K. Mozumder, and L. M. Loewenstein, "Method for semiconductor process optimization using functional representations of spatial variations and selectivity," *IEEE Trans. Components, Hybrids, and Manu. Tech.*, vol. 15, no. 5, pp. 311-316, 1992.

[52] P. K. Mozumder, and G. G. Barna, "Statistical feedback control of a plasma etch process," *IEEE Trans. Semi. Manufac.*, vol. 7, no. 1, pp. 1-11, 1994.

[53] R. Mundt, "Model based training of a neural network endpoint detector for plasma etch applications," in *Proc. Electrochem. Soc.: Process Control, Diagnostics, and Modeling in Semiconductor Manufacturing*, vol. 95-2, pp. 178-188, 1995.

[54] J. Neter, W. Wasserman, and M. H. Kutner, *Applied Linear Regression Models*. Illinois: Irwin, 1983.

[55] A. V. Oppenheim, and R. W. Schafer, *Discrete-Time Signal Processing*. New Jersey: Prentice Hall, 1989.

[56] M. S. Phadke, *Quality Engineering Using Robust Design*. New Jersey: Prentice Hall, 1989.

[57] E. A. Rietman, "Neural networks in plasma process," *J. Vac. Sci. Technol. B*, vol. 14(1), pp. 504-510, 1996.

[58] E. A. Rietman, "A neural network model of a contact plasma etch process for VLSI production," *IEEE Trans. Semi. Manufac.*, vol. 9, no. 1, pp. 95-100, 1996.

[59] E. A. Rietman, R. C. Frye, E. R. Lory, and T. R. Harry, "Active neutral network control of wafer attributes in a plasma etch process," *J. Vac. Sci. Technol. B*, vol. 11(1), pp. 1314-1318, 1992.

[60] E. A. Rietman, and S. H. Patel, "A production demonstration of wafer-to-wafer plasma gate etch control by adaptive real-time computation of the over-etch time from in situ process signals," *IEEE Trans. Semi. Manufac.*, vol. 8, no. 3, pp. 304-308, 1995.

[61] E. M. Sachs, R. S. Guo, S. Ha, and A. Hu, "Process Control System for VLSI Fabrication," *IEEE Trans. Semi. Manufac.*, vol. 4, no. 2, pp. 134-144, 1991.

[62] E. M. Sachs, A. Hu, and A. Ingolfsson, "Run by run control: combining SPC and feedback control," *IEEE Trans. Semi. Manufac.*, vol. 8, no. 1, pp. 26-43, 1995.

[63] S. S. Saxena, P. K. Mozumder, and K. J. Taylor, "Simultaneous control of multiple measures of nonuniformity using site models and monitor wafer control," *IEEE Trans. Semi. Manufac.*, vol. 9, no. 1, pp. 128-1[illegible], 1996.

[64] C. Schaper, M. Moslehi, K. Saraswat, and T. Kailath, "Control of MMST RTP: repeatability, uniformity, and interaction for flexible manufacturing," *IEEE Trans. Semi. Manufac.*, vol. 7, pp. 202-219, 1994.

[65] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 1986, pp. 276-280, 1986.

[66] *SEMI Equipment Communications Standard 2, Message Content*, Book of SEMI Standards 1994, vol. 2B, Semiconductor Equipment and Materials Institute, Inc., Mountain View, CA, 1994.

[67] T. H. Smith, *Novel techniques for the run by run process control of chemical-mechanical polishing*. M.S. Thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 1996.

[68] J. Stefani, and S. W. Butler, "On-line inference of plasma etch uniformity using in-situ ellipsometry," *J. Electrochem. Soc.*, vol. 141, no. 5, pp. 1387-1391, 1994.

[69] M. Sternheim, W. van Gelder, and A. W. Hartman, "A laser interferometer system to monitor dry etching of patterned silicon," *J. Electrochem. Soc.*, vol. 130, no. 3, pp. 655-658, 1983.

[70] P. Stoica, "List of references on spectral line analysis," *Signal Processing*, vol. 31, no. 3, pp. 329-340, 1993.

[71] *Ultramax software program*. Ultramax Corp., Cincinnati, OH, 1984.

[72] K. Wong and D. S. Boning, "On in-situ etch rate extraction from interferometric signals," in *Proc. Electrochem. Soc.: Process Control, Diagnostics, and Modeling in Semiconductor Manufacturing*, proc. vol. 95-2, pp. 360-371, 1995.

[73] K. Wong, D. S. Boning, and H. H. Sawin,"On endpoint detection of plasma etching via optical emission interferometry," in *Proc. Electrochem. Soc.: Plasma Processing XI*, proc. vol. 96-12, 1996.

[74] S. Wolf, and R. N. Tauber, *Silicon Processing for the VLSI Era: Volume 1-Process Technology*, California: Lattice Press, 1986.

[75] N. Yamada, P. L. G. Ventzek, Y. Sakai, and H. Tagashira, "Simulations of a feedback control scheme for an inductively coupled plasma source for etching applications," *J. Electrochem. Soc.*, vol. 143, no. 4, pp. 1375-1383, 1996.

[76] Y. Yao and S. Pandit, *IEEE Trans. Signal Process.*, vol. 43, pp. 878-885, 1995.

[77] G. C. H. Zau, and J W. Butterbaugh, P. Rummel, and H. H. Sawin, "Monitoring and control of real power in RF plasma processing," *J. Electrochem. Soc.*, voi. 138, no. 3, pp. 872-873, 1991.

Appendix A

Etch rate data used for neural network model building [23]:

Table A-1: Selected etch rate data from AME-5000 (Å/sec)

mag. field strength (G)	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
20	45.6	46.1	45.8	45.3	46.8	47.7
30	51.2	51.6	51.3	51.4	51.5	52.9
40	59.9	60.1	59.8	60.0	60.1	61.8
50	66.6	66.8	67.4	66.3	67.8	68.6
60	73.1	73.1	73.6	73.1	73.8	74.1
70	75.7	75.5	75.6	75.9	77.3	77.4
80	85.0	84.7	84.5	84.8	85.7	85.9
90	90.0	90.2	90.3	89.9	90.3	91.0
100	95.6	95.8	95.8	95.3	95.8	95.7

Table A-1 (cont.): Selected etch rate data from AME-5000 (Å/sec) (cont.)

mag. field strength (G)	Site 7	Site 8	Site 9	Site 10	Site 11	Site 12
20	47.4	46.8	47.3	48.7	48.7	48.[illegible]
30	52.7	52.2	53.6	54.7	55.2	53.3
40	60.5	59.6	60.8	62.1	60.9	59.6
50	68.2	67.5	68.4	70.0	68.6	66.3
60	73.4	73.5	73.9	74.3	73.4	72.7
70	77.4	76.4	76.7	77.6	76.8	76.5
80	85.3	85.4	84.4	85.6	82.2	82.3
90	90.3	90.0	89.3	90.2	88.3	87.9
100	95.8	95.3	93.6	93.7	92.2	91.9

Appendix B

Neural network etch rate model:

Define the weighting and bias matrices as:

$$W_1 = \begin{bmatrix} 0.121496 \\ 0.109943 \end{bmatrix}; \qquad b_1 = \begin{bmatrix} -11.089601 \\ -4.063199 \end{bmatrix};$$

$$W_2 = \begin{bmatrix} 8.380024 & 2.085767 \\ 5.344423 & 3.020160 \\ 2.804355 & 3.463502 \end{bmatrix}; \qquad b_2 = \begin{bmatrix} -1.418143 \\ 0.123218 \\ -3.971616 \end{bmatrix};$$

$$W_3 = \begin{bmatrix} 24.901555 & 24.643514 & 26.689213 \\ 24.510650 & 24.111219 & 26.823810 \\ 24.798275 & 24.742980 & 26.604534 \\ 24.970044 & 24.712073 & 26.389683 \\ 25.360973 & 24.873498 & 25.721146 \\ 24.956380 & 24.884695 & 24.516553 \\ 24.665267 & 24.410614 & 25.317460 \\ 24.768859 & 24.300637 & 25.715382 \\ 24.562290 & 24.981187 & 22.264304 \\ 24.665568 & 25.207638 & 20.991318 \\ 23.167616 & 23.907994 & 20.248619 \\ 23.154596 & 23.102468 & 22.112291 \end{bmatrix}; \qquad b_3 = \begin{bmatrix} 23.368223 \\ 24.185688 \\ 23.533226 \\ 23.292254 \\ 23.914814 \\ 25.267554 \\ 25.140868 \\ 24.552058 \\ 25.457248 \\ 26.628703 \\ 27.863031 \\ 27.113608 \end{bmatrix}.$$

The etch rates (*ER*) for the twelve selected sites at the magnetic field strength (*B*) are:

$$ER = W_3\Gamma(W_2\Gamma(W_1B + b_1) + b_2) + b_3.$$

Appendix C

Sensitivity analysis of film thickness estimation error at film clearing due to film structure variation:

The film to be etched is polysilicon, which is deposited over silicon oxide on a silicon wafer. If the film structure varies, such as a slight change in the oxide thickness, the phase at endpoint of the interferometric signal will change accordingly. The difference between the prescribed phase (phase from the nominal film structure), and the actual phase will result in error in film thickness estimation. For example, even when the signal reaches the prescribed endpoint phase, a thin polysilicon layer is still present. The follow tables shows the error due to such a difference in endpoint phase with respect to small variations in film structures. The differences have been converted to polysilicon film thickness so that the result from different signal wavelengths can be compared. Table C-1 shows that 9 cases that are considered. For each case, the refractive index of polysilicon, oxide, and silicon substrate are varied at ± 0.05, and the oxide thickness are also varied at ± 0.05. Table C-2 shows the 2-level full factorial design array, and Table C-3 lists the thickness errors for each case.

Table C-1: Film structures for sensitivity analysis

case #	signal wavelength (Å)	poly / silicon refractive index	oxide refractive	oxide thickness (Å)
1	4800	4.42	1.55	80
2	4800	4.42	1.55	220
3	4800	4.42	1.55	1000
4	6100	3.93	1.46	80
5	6100	3.93	1.46	220
6	6100	3.93	1.46	1000
7	7534	3.73	1.45	80
8	7534	3.73	1.45	220
9	7534	3.73	1.45	1000

Table C-2: 4-parameter 2-level full factorial design array (L: low level; H: high level)

Experiment number	Poly refrac. index	oxide thickness	oxide refrac. index	sub. refrac. index
1	L	L	L	L
2	H	L	L	L
3	L	H	L	L
4	H	H	L	L
5	L	L	H	L
6	H	L	H	L
7	L	H	H	L
8	H	H	H	L
9	L	L	L	H
10	H	L	L	H
11	L	H	L	H
12	H	H	L	H
13	L	L	H	H
14	H	L	H	H
15	L	H	H	H
16	H	H	H	H

Table C-3: Film thickness estimation error (Å) due to film structure variation

Experiment number	case 1	case 2	case 3	case 4	case 5
1	1.8	5.6	12.1	2.8	6.8
2	-16.6	-3.2	11.8	-38.3	-11.1
3	-0.2	0.6	0.1	-0.2	0.8
4	-17.6	-7.2	1.8	-37.3	-15.1
5	1.8	4.6	-1.9	1.7	4.8
6	-17.6	-5.2	0.8	-40.3	-13.1
7	-0.2	-1.4	-16.9	-0.2	-1.2
8	-17.6	-9.2	-11.2	-38.8	-17.1
9	20.8	11.6	11.1	44.8	20.8
10	0.4	1.8	11.8	0.2	1.9
11	16.8	5.6	-0.9	38.8	13.8
12	-1.6	-3.2	1.8	-2.3	-3.1
13	20.8	10.6	-2.9	45.8	19.8
14	-0.6	0.8	-0.2	-0.3	0.4
15	16.8	4.6	-18.9	39.3	12.8
16	-2.6	-4.2	-13.2	-2.3	-5.1

Note: "-16.6" means at the prescribed endpoint, 16.6Å of polysilicon is still left.

Table C-3: Film thickness estimation error (Å) due to film structure variation (cont.)

Experiment number	case 6	case 7	case 8	case 9
1	13.5	2.5	6.5	15.5
2	9.7	-67.1	-22.0	8.4
3	1.0	-0.02	0.5	0.5
4	-0.3	-63.1	-25.0	-3.1
5	1.5	1.5	4.5	4.0
6	0.7	-69.1	-25. 0	-1.1
7	-12.5	-0.5	-1.5	-12.5
8	-11.3	-65.1	-28.0	-0.3
9	13.5	76.5	33.5	17.5
10	10.7	-0.1	2.0	9.9
11	1.0	67.5	24.5	2.5
12	0.2	-2.1	-4.0	-2.1
13	1.5	77.5	32.5	5.5
14	0.7	-0.1	0.03	0.4
15	-13.0	68.5	23.5	-11.5
16	-11.3	-2.2	-6.0	13.1

Appendix D

A Matlab script file for endpoint prediction:

```
% This is a simple algorithms for endpoint prediction, (using all data,
% not just a moving window)
%
% Phase at endpoint is pre-determined (theorectically or from experiments)
% which is used for endpointing.
%
% Important variable definitions in the program:
%
% cycle_end: number of minima of signals before endpoint
% an_end: phase at endpoint (pre-determined)
% thick_tot: total film thickness
% thick_per: film thickness per cycle of signal
% rule: phase at the first signal received, or its +/- pi bound
% junk: signals to be analyzed
% alarm: the data number at endpoint
% N: number of data point equivalent to 2 cycles of signals
% an: the (AN)gle or phase for the last signal received; an(1) is associated
%     with the Nth data point.

% Initialize some arrays
clear an
alarm = 0;

% This is the normalized endpoint phase for etching polysilicon on 1000Å oxide
an_end = 3.094836682/pi;

% The initial thickness is less than 5000Å because I need an intial short etch
% to orient the CCD camera to select the right pixels.
thick_tot = 4750;
thick_per = 776.875;

% Calculate the total number of signal minima
cycle_end = floor(thick_tot/thick_per - (an_end+1)/2) + 1;
rule = 1 - rem(thick_tot/thick_per-(an_end+1)/2,1)*2;

% Pick a random number of signal data point to start with, which should be
% between 1 and 2 cycles of signals. The sensor noise may dominate the signal
% if N is too small, and incorrect cycle count will be got if N is too large.
N = 20;
N_new = 10000;
```

```
% Calculating the number of data points equivalent to 2 signal cycle as new
% data are received.
while (N < (fix(N_new*1.999999)))
 sb = junk(1:N);
% Remove a linear trench from the signal before determining the periodicity
 xx = [ones(N,1) [1:N]'];
 sb = sb - xx*inv(xx'*xx)*xx'*sb;
 ft = (fft(hamming(N).*sb,2048));
 [m nn1] = max(abs(ft(1:1024)));
 N_new = 2048/(nn1-1);
 N = N+1;
end

% This is used to determine is the estimated phase wrapped whether it
% crosses over the +/- pi region, or it is just fluctuating about zero.
flip = 1-4/N;

sb = junk(1:N);
xx = [ones(N,1) [1:N]'];
sb = sb - xx*inv(xx'*xx)*xx'*sb;
ft = fft(hamming(N).*sb,2048);
[m nn1] = max(abs(ft(1:1024)));
nn2=nn1;
nn = (nn1-1)/2048;
an(1) = rem(angle(ft(nn1))/pi+1+rem((N-1)*nn,1)*2,2)-1;

% The fundamental frequency from the first 2 signal cycle will
% set up a bound for frequency detection later, to reduce the
% photoresist interference.
nn_old = nn1;
nn_b = ceil(nn_old/3);

% Adjust the number of cycles to be expected before endpoint
% Refer to Section 5.5 of the thesis.
% It took me quite a while to figure this part. If the cycle counting is
% incorrect, it could be due to a scenario that I haven't considered here;
% or the initial N is too big/small.
cycle_now = 2;
if (rule > 0)
  rule = rule - 1;
  if (an(1) < rule)
    cycle_now = cycle_now + 1;
  end
else
  rule = rule + 1;
```

```
 if (an(1) > rule)
  cycle_now = cycle_now - 1;
 end
end

% Analyze new data one after another
for kk=2:(length(junk)-N+1)

% The all-data window is used.
 sb = junk(2:kk+N-1);
 xx = [ones(kk+N-2,1) [1:kk+N-2]'];
 sb = sb - xx*inv(xx'*xx)*xx'*sb;
 ft = fft(hamming(kk+N-2).*sb,2048);

% The last four lines will be replaced by the following lines if a 2-signal-cycle
% sliding window is used for phase calculation:
%  sb = junk(kk:kk+N-1);
%  sb = sb - xx*inv(xx'*xx)*xx'*sb;
%  ft = fft(hamming(N).*sb,2048);

  [m nn1] = max(abs(ft(nn_old-nn_b:nn_old+nn_b)));
 nn1 = nn1+nn_old-nn_b-1;

 nn = (nn1-1)/2048;
% I want to get the phase at the end of the data window, not at the beginning
 an(kk) = rem(angle(ft(nn1))/pi+1+rem((kk+N-3)*nn,1)*2,2)-1;
% Increment cycle count if a signal minimum is detected
 if((an(kk)*an(kk-1) < 0) & (an(kk)<-flip | an(kk-1)>flip))
  cycle_now = cycle_now + 1;
 end

if (((cycle_now>=cycle_end) & (an(kk)>=an_end-0.75*nn) ) | (cycle_now > (cycle_end)))
  alarm = kk;
% Endpoint is reached. Stop the etch!
  break
 end
end
```