STUDY ON REAL-TIME DEVELOP RATE MONITORS FOR CONTROL OF DEEP-ULTRAVIOLET LITHOGRAPHY

by

Mareike Claassen

Memorandum No. UCB/ERL M98/59

8 October 1998

STUDY ON REAL-TIME DEVELOP RATE MONITORS FOR CONTROL OF DEEP-ULTRAVIOLET LITHOGRAPHY

by

Mareike Claassen

Memorandum No. UCB/ERL M98/59

8 October 1998

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Masters Report

Study on Real-Time Develop Rate Monitors for Control of Deep-Ultraviolet Lithography

Mareike Claassen
Department of Mechanical Engineering
University of California at Berkeley

October 8, 1998

Abstract

The goal of this project is to assess the utility of a Develop Rate Monitor (DRM) as a in-situ sensor for deep-ultraviolet (DUV) lithography control. The DRM collects normal incidence interferometry signals at several wavelengths. The interferometry signals contain information about the resist thickness, the optical properties of the resist and the underlying geometry, and can also be used to calculate the dissolution rate.

This study analyzes different filtering methods for the extraction of the resist thickness from the raw interferometry data. It also examines the sensitivity of the develop rate to processing parameters such as exposure dose and characterizes the quality of measurements needed to accurately and consistently infer thickness. In the context of DUV lithography control, the DRM is evaluated as an in-situ sensor.

Acknowledgements

I would like to thank my research advisors, Professor Kameshwar Poolla and Professor Costas J. Spanos, for their guidance and support. All I have learned is based on their combined knowledge of system identification and semiconductor processing.

Special thanks also to my fellow students in the BCCI and BCAM labs. In particular to Nickhil Jakatdar and Xinhui Niu, who answered all my questions about processing, lithography and reflectometry patiently, to John Musacchio for his help with the control model and to Darin Fisher and Mason Freed for answering mostly random computer questions.

I also want to acknowledge Joe Bendik, Matt Hankinson and Andy Romano for their support during the experiments at National Semiconductor Corporation. I would also like to thank Rick Dill of IBM Corporation and Piotr Zalicki of SC Technology for their valuable comments throughout the years.

Last but not least, warm thanks go out to my family and friends. Their love and support keeps me sane on my journey through graduate school.

Contents

Chapter 1

Introduction and Background

1.1 Introduction

As feature sizes and line widths in integrated circuit production shrink, the lithography sequence becomes increasingly critical in the operation of cost effective, competitive production lines.

It is essential to produce small critical dimensions (CDs) with little variability. The traditional method of reducing variability is to fine-tune machines and design stable processes. This is reaching its limits and the cost of equipment is now so high, that their lifetime has to be extended while they are being pushed to higher performance. This compels us to introduce automatic control methods onto the manufacturing floor.

A key step to successfully applying automatic control is the availability of reliable, accurate sensors. One opportunity for control in the lithography sequence is to engineer a develop rate sensor. This sensor can potentially be used for run-to-run and real-time control. A real-time control application is that of an end-point detector. In combination with other sensors, the develop rate measured on the current wafer can be used in run-to-run control to adjust exposure dose, PEB bake or other process inputs for the next wafer in the lot to achieve less CD variability. This application requires a model of the dependence of develop rate on these process inputs.

In this study we evaluate the performance of a DRM and we asses its capabilities. The current limitations are described. Finally, we develop guidelines to asses the accuracy required for DRMs to be effective for run-to-run control.

1.2 Application of DRMs to I-Line Processes

Develop rate monitors have been in existence for many years and several applications for I-line lithography have been explored.

The classical DRM application is as an end-point detector. A single wavelength reflectometer is used to measure the reflected light intensity off the wafer. This signal reaches a steady state when the resist is completely developed. The flattening of the signal indicates the end of the develop. This can be detected using various signal processing methods. End-point detection is much simpler than measuring thickness or develop rate. Algorithms for the thickness/develop rate extraction (described later in this study) need an initial condition for thickness, models for the dynamic behavior of thickness and normalizing measurements for reflectance. This is in contrast to algorithms for end-point detection which only use the shape of the signal.

The difficulty in end-point detection lies in having to focus the reflector beam onto a spot of resist that is developing. This spot may not represent the average behavior across the wafer. In addition, the noise level of the signal has to be known in order to optimize detection while minimizing the delay in end-pointing.

For I-line resists there was a need to detect the end-point accurately to avoid over or under developing of the resist. These conditions lead to large variations in the CD. It has been shown in SAUTTER ET AL[1] that the develop end-time is highly sensitive to changes in process variables such as exposure dose, and the actual CD is also very sensitive. This has two implications. First it means that by adjusting develop time one can compensate for deviations in earlier processing steps and reduce variability in CD. Second, the end-point can be used to monitor variations in processing variables and faults in previous steps. This knowledge can then be applied to appropriately adjust the settings of the exposure and PEB steps for the following wafers.

Another use of end-pointing is early detection of faulty wafers. It is costly and time consuming to measure CD off-line on every wafer to ensure that they meet specifications. The end-point of every wafer can be measured in line without interrupting or slowing of the process flow. A wafer whose end-point is out-of-spec can be automatically removed from the process flow, measured off-line and reworked before other processing steps occur. This can minimize the amount of scrap and re-processing required.

Despite good solutions to these problems, DRMs did not become hugely popular in pro-

duction facilities, although they were successful in research laboratories. This is principally because the integration of the DRM into the production line was too complicated. For patterned wafers the end-point algorithms needed to be tuned to each specific process, which is difficult on production lines that run a variety of designs and processes.

The introduction of chemically amplified resists (CARs) has shifted the focus from end-point detection to more complex variables such as develop rate. CA resists introduce a high ratio of selectivity between exposed and unexposed areas so that even very long over-develop times do not significantly affect CD. This indicates a decreased need to control the develop time accurately using end-point detectors. Potentially, end-point detection could be used to increase throughput, but the develop station is inexpensive and not a resource bottleneck. Another possible application is to use the end-time as a modeling variable and model its dependence on exposure dose and other process inputs. Given a good model, the end-time could be used to monitor process variations and adapt inputs using run-to-run control algorithms.

An altogether different role of the DRM is to measure thickness and develop-rate in real time. This measurement is much more complex than end-point and could be used for more accurate models. In this study we examine this problem and associated difficulties. We describe the theory that connects resist thickness and reflectivity measurements and explore various algorithms to extract thickness from measured intensities.

Chapter 2

The Deep-Ultra-Violet Lithography Sequence

This chapter provides an overview of the sequence of processing steps involved in DUV lithography and of the characteristics of chemically amplified resists. For each process step the main variables are described, and their influence on the final resist profile is examined. This information allows us to evaluate the necessity of measuring develop rate in the context of various control strategies.

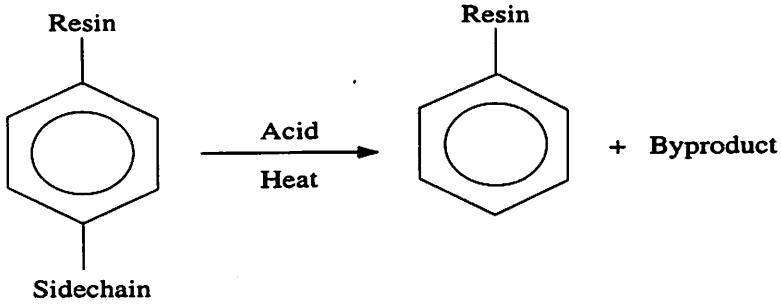
2.1 Chemically Amplified Resists

The term chemically amplified resist (CAR) describes a family of photo sensitive substances that use a catalytic reaction to increase the sensitivity of the resist to exposure. A chemically amplified resist usually consists of multiple components including the polymer resin and a photo-acid generator (PAG). For positive resists the polymer has a side-group attached that renders it unsoluable in the developer, i.e. the side group blocks dissolution. Polymers with the side groups are called blocked sites. During the processing this side group is broken and the resist gets more soluble. This deblocking is done in a sequence of chemical reactions. In the first step, the photon energy during exposure creates acid and byproducts from the photo-acid generator.

$$\mathrm{PAG} \xrightarrow{\mathrm{Light}} \mathrm{Acid} + \mathrm{Byproduct}$$

During the post exposure bake, heat fuels a thermally induced catalytic reaction between

the acid and the side group of the polymer. The side groups disconnect and acid is regenerated. This process is self-catalyzed and self-quenching. Overall the acid is not consumed, and to the first order acid concentration remains constant.



The kinetics of this reaction are more complex, since the reactive sites are consumed during the process and the catalytic rate decreases.

This type of catalytic reaction creates a high ratio between the dissolution rate of exposed and unexposed areas. This high selectivity leads to short develop times and very sharp features. Conversely, it creates high sensitivity to process drifts and need for tight process control. Usually the chosen operating region, as shown in Figure 2.1, can tolerate small deviations in the process inputs without large changes in selectivity. Changes in the process inputs will cause changes in the develop rate for the exposed areas, but this new develop rate is still much larger than that of unexposed areas.

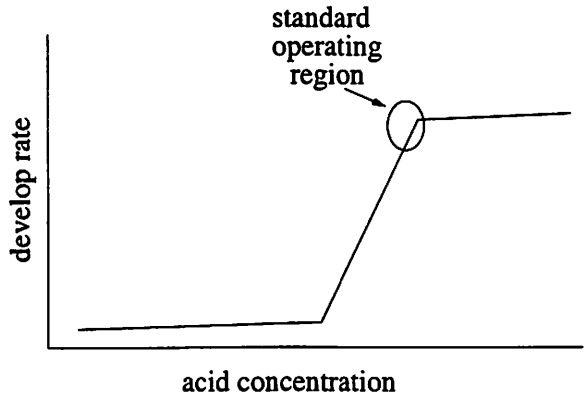


Figure 2.1: Sensitivity of CA resists

2.2 The Processing Sequence and its Variables

The outcome of a lithography processing sequence is a complicated function of many parameters. Variables that influence the final resist profile enter at every step during the process.

The DUV lithography sequence consists of the following main steps: resist coating, pre-bake, formation of an aerial image, exposure, post-exposure-bake (PEB) and develop. The following sections describe the variables entering the system at each process step and show how they influence the final CD.

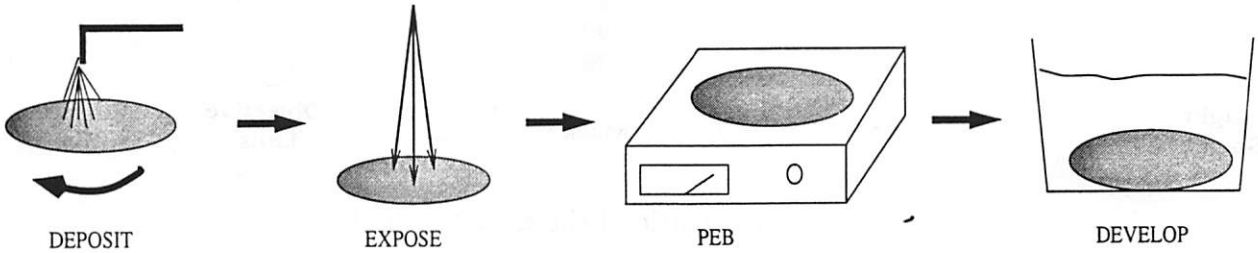


Figure 2.2: The Lithography Sequence

2.2.1 Resist Coating and Pre-bake

The spin-coat process is the most popular way to apply photoresist to a wafer. The thickness of the applied resist depends on the spin speed and spin time and on environmental properties such as humidity, vapor pressure and temperature. The uniformity of the film is influenced by the geometry of the underlying layers, dirt particles and the quality of the spin-coat tool. Aside from thickness and uniformity, the optical properties of the resist influence the final CD. These properties can vary from batch to batch and also as the resist ages. The reflectivity of underlying layers influences the final resist profile by changing the intensity within the resist during exposure even though anti-reflective coating can minimize this effect. The pre-bake step hardens the resist by evaporating the solvent and reduces its thickness (5-10%) which can also influence uniformity if the bake plate is non-uniform.

2.2.2 Aerial Image Formation

The formation of the aerial image, i.e. the distribution of light intensity in the volume of the resist is a combination of many effects, including the intensity of the light source at the exposing wavelength, the optical design, i.e. lenses and mirrors and the mask pattern.

Next to the intensity of the light source, the dominant effect is the diffraction of light when it passes through the mask. The electromagnetic field behind the mask can be described by the Fraunhofer diffraction integral if the slit size is small enough compared to the distance

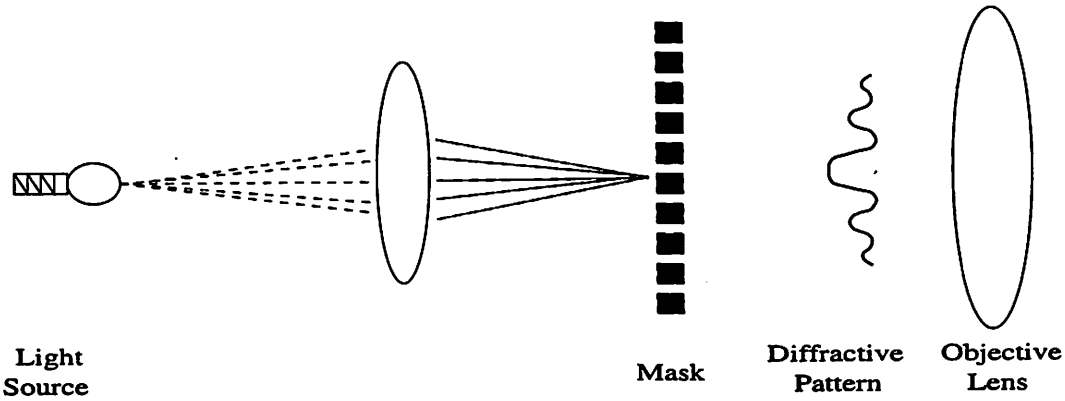


Figure 2.3: Schematic of the exposure tool

from the lens to the mask. The diffraction pattern is then the Fourier transform of the mask pattern. The objective lens only captures part of the light of the diffractive pattern dependent on its numerical aperture. The lens must capture at least one higher diffraction order in addition to the zeroth-order to yield an acceptable image. This defines the resolution of the system. The theoretical resolution of the projected image depends on the type of pattern, e.g. isolated lines, dense equally spaced lines, or others. An ideal lens would produce the inverse Fourier transform of a diffraction pattern and therefore would produce a perfect image. However, a real lens is imperfect and the image will be fuzzy. A optical system is called diffraction limited if the image differs from the pattern primarily because of uncaptured diffraction orders and not because of lens geometry.

In addition to these first order effects that cause variations in the projected image, second order effects result from partially coherent and non-normal incident light. IN addition to the theoretically known effects there are deviations of the real imaging system from its ideal theoretical behavior which are called aberrations. They result from defects in the imaging system, changes in the environmental conditions and intentional design deviations to minimize other effects. One of the most important types of aberration is defocus, which can be described by the optical path difference between the focus plane and the wafer. Defocus adds a phase difference to higher diffraction orders which changes the intensity of the superposition of different orders. The focal plane is an important processing parameter and its optimal position is determined experimentally.

Most of the inputs that influence the aerial image depend on the equipment and are assumed to be fixed. The inputs controlled in the process are the energy delivered, the light source, the mask pattern and the focal plane.

2.2.3 Exposure Kinetics and Standing Wave Effects

In chemically amplified resists the exposure energy generates acid that during the post exposure bake (PEB) changes the solubility of the resist. A positive resist consists of a photo-acid generator (PAG) and a polymer resin, which in part is blocked by side chains to inhibit dissolution. The photo-acid generator forms an acid upon exposure to DUV light. The amount of generated acid depends locally on the intensity and length of exposure and the concentration of photo acid generator.

During exposure, standing waves are produced in the resist layer. Standing waves are the result of interference of the incoming and the reflected light. Assuming normal incidence light the intensity at each point through the depth of the resist depends on the resist thickness, the refractive indices of the resist and the underlying layer and the wavelength. Standing waves lead to variations in develop rate through the resist, which cause the final features to have fringed edges. The use of an anti-reflective coating can minimize standing wave effects at the expense of additional processing steps.

After exposure a 3-dimensional latent image of acid remains in the resist. For the calculation of this image the spatial distribution of intensity (including standing wave effects and variations in intensity during exposure) have to be taken into account. Differential equation models that describe the rate of production of acid during exposure are available.

2.2.4 PEB Chemistry and Diffusion

During post exposure bake two main processes take place: diffusion, and a change in the chemical composition of the resist. The post exposure bake thermally induces a chemical reaction that is catalyzed by the acid generated during the exposure. This reaction, called deblocking of the polymer, changes its solubility. Initially a hydroxyl group is attached to the longer chain polymers, which cause the solubility of the polymer to be greatly reduced. The acid and heat induce a reaction that detaches the hydroxyl groups and renders the remaining polymer soluble. During this reaction acid is consumed but is automatically replenished. Some of the reaction byproducts are also volatile, leading to thickness loss in the resist. The sensitivity of the resist can be changed by varying the ratio of blocked to unblocked polymer sites.

The second major effect of PEB on the resist is diffusion. The original acid distribution changes due to diffusion, quenching and evaporation of the acid. Since the acid concentration

magnifies the deblocking of the resist, accurate knowledge of the acid profile at all times is essential. The acid diffusion reduces the effect of standing waves but exact models of the diffusion process are not well known. This is because secondary effects such as the change of diffusivity and reactivity of the blocked sites contribute to complicating the equations. First order diffusion models combined with differential equations describing the chemical reactions provide the models currently used.

2.2.5 Develop

The main variable related to develop rate is the 3-dimensional distribution of deprotected sites in the resist. Surface inhibition effects and the developer diffusion between the solution and the resist must also be considered in modeling the develop process. Secondary effects are the enhancement of the solubility by reaction byproducts. Develop is a surface limited process, i.e. the develop rate of the top part of the resist is more constraining than the diffusion of the developer into the resist and dissolution from within.

2.3 The Dominant Variables

Even though there are dozens of variables influencing the develop rate of the photoresist, we can only discuss a few in detail in this study. We will introduce two parameters, effective exposure dose and effective PEB time, which can account for the effects of variations in other variables.

In general, the optics of an exposure tool are such that we can consider most deviations in the aerial image formation secondary. There is an intensity loss factor from the light source to the focal plane, that is assumed to be uniform across the image and is lumped into the effective exposure dose parameter. Aberrations are taken into account by adjusting the focal plane whenever the process is changed.

Variations in the absorption of the resist and in the exposure intensity both lead to variations in the acid generation. Therefore, they can be lumped into the effective dose parameter which can also accommodate changes in the optical properties of the underlying layers and variations of the resist composition.

Similarly, one can lump many other variables during the PEB step into an effective bake time parameter. For example, an increase in the initial percentage of protected sites can

be described as an increase in bake time needed to achieve an equal final percentage of deprotected sites.

The choice of these particular two parameters, namely effective dose and effective bake time is reasonable because they can account for many other variables and because one can easily actuate them to control the process.

Chapter 3

The Develop Process

This section examines the develop step in more detail. It describes how a dissolution rate monitor works, the theory behind reflectometry and the difficulties that arise when taking real measurements.

3.1 Typical Develop Rate Curves

Since chemically amplified resists are very sensitive and are often tuned to specific processing conditions, the develop rate varies greatly with the processing parameters. A given process does not have a single develop rate. Rather it varies both spatially across the wafer and into the depth of the resist. Locally, a typical thickness vs. time plot during develop for typical processing conditions is shown in Figure 3.1.

Even though typical develop steps last about 50 seconds, most of the resist is removed during the first 10 seconds. The good selectivity of chemically amplified resists makes this long over-develop time possible without causing damage.

To compare different process settings, we must agree on a measure of quality that can be computed from the develop rate time signal. There are many choices for this measure, such as the average rate until the thickness is less then a threshold, the coefficient of an exponential fit to the develop rate vs. time curve, or the slope of a linear fit to the first few seconds of the thickness vs. time curve. Although these choices will lead to different numerical values for the quality measure, they exhibit the same qualitative behavior when varying the process settings, such as exposure dose. In the remainder of this report we refer to the time signal and the chosen quality measure by "develop rate".

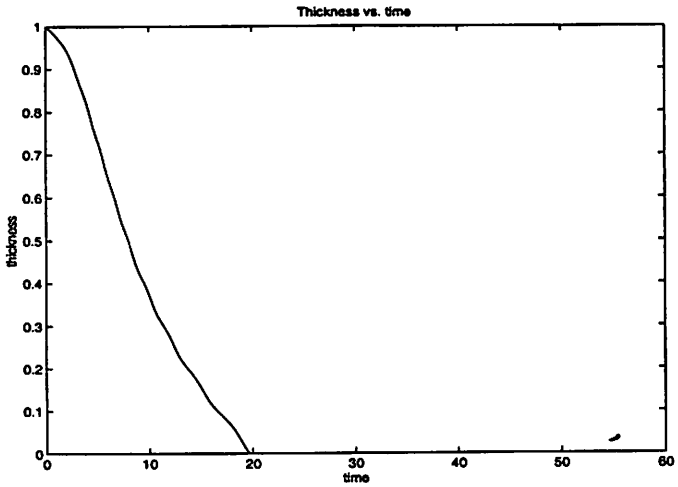


Figure 3.1: Prolith simulation of thickness vs. time

A typical profile of develop rate vs. exposure dose is shown in Figure 3.2

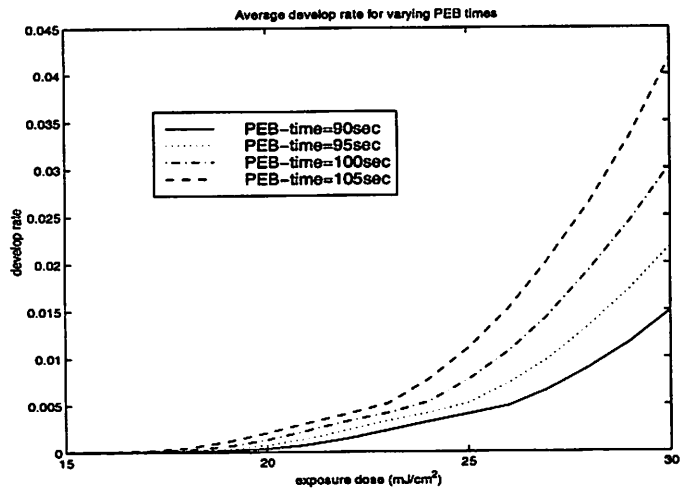


Figure 3.2: Prolith simulation of develop rate vs. exposure dose

One commonly used method to obtain profiles of develop rate vs. exposure dose is called Poor Man's DRM (described in detail in [9]). A Poor Man's DRM experiment uses multiple wafers with the same processing conditions and develops each wafer for increasing lengths of time. The remaining thickness is measured on each wafer. The difference in thickness and develop time determines the develop rate. This is a relatively simple experiment that can be used to corroborate DRM results.

3.2 The Dissolution Rate Sensor

A develop rate monitor (DRM) is basically a reflectometer. It measures the intensity of light reflected of a thin film which then is used to infer thickness of the thin film. A DRM records light intensity at a single wavelength or multiple wavelengths.

To understand the measurements of a DRM one has to study the behavior of electromagnetic waves as they pass through thin films and reflect on film boundaries. A good source for the relevant theory is [7].

The measured light intensity is a function of the reflected and absorbed portions of the electromagnetic wave as shown in Figure 3.3.

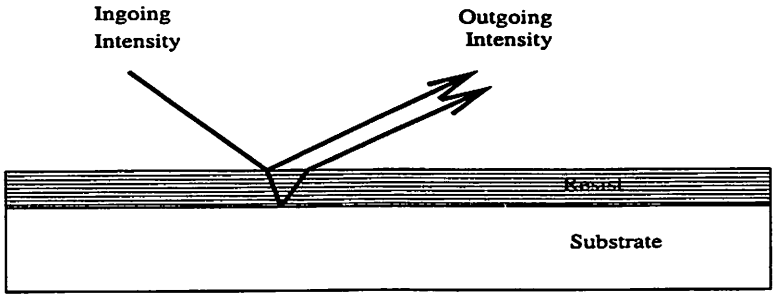


Figure 3.3: Light reflected of a thin film

Consider a thin film of thickness d and complex refractive index $n+ik$ on a thick substrate with index n_s. At the measured wavelength between $700-900nm$ the absorption coefficient k is negligible for materials such as air, silicon and photoresist, and the characteristic matrix M simplifies to

$$M = \begin{pmatrix} \cos(\frac{2\pi}{\lambda} n\, d) & -\frac{i}{n}\sin(\frac{2\pi}{\lambda} n\, d) \\ -i\, n \sin(\frac{2\pi}{\lambda} n\, d) & \cos(\frac{2\pi}{\lambda} n\, d) \end{pmatrix} \tag{3.2.1}$$

The characteristic matrix M describes the behavior of the electro-magnetic wave as it passes through the film. The characteristic matrices of a stack of thin films can be multiplied (respecting the order) to obtain the combined characteristic matrix of the stack.

For transparent layers the reflectance coefficient is

$$r = \frac{(m'_{11} + m'_{12}\, n_s) n_{air} - (m'_{21} + m'_{22}\, n_s)}{(m'_{11} + m'_{12}\, n_s) n_{air} + (m'_{21} + m'_{22}\, n_s)} \tag{3.2.2}$$

where m_{ij} denotes the ij-th element of the characteristic matrix M and $()'$ denotes complex conjugation.

The reflectivity $\mathcal{R} = |r|^2$ is the ratio of incident to reflected light intensity. Since the DRM measures only the reflected intensity, this measurement has to be calibrated with a measurement for the incident intensity. We use a multiplicative parameter p to account for absorption not captured in this measurement. We do not include the dependence of this parameter on wavelength although in reality absorption of the developer depends on wavelength. Modeling this dependence would unnecessarily increase the number of unknown parameters. As a consequence,

$$\mathcal{R} = p\,|r|^2 \tag{3.2.3}$$

It is easy to see that $\mathcal{R}$ is a complicated non-linear function of thickness d, refractive index n and wavelength λ. The refractive index itself also depends on the wavelength. For wavelengths above $300nm$ the dependence is straightforward, and it is possible to analytically compute first derivatives of $\mathcal{R}$ with respect to thickness d and other parameters, which can be used for computation and optimization.

The shape of the reflectivity data depends on the dynamic behavior of the thickness during the develop step. If the thickness decreases linearly, i.e. develop rate is constant, the reflected intensity has the shape of Figure 3.4

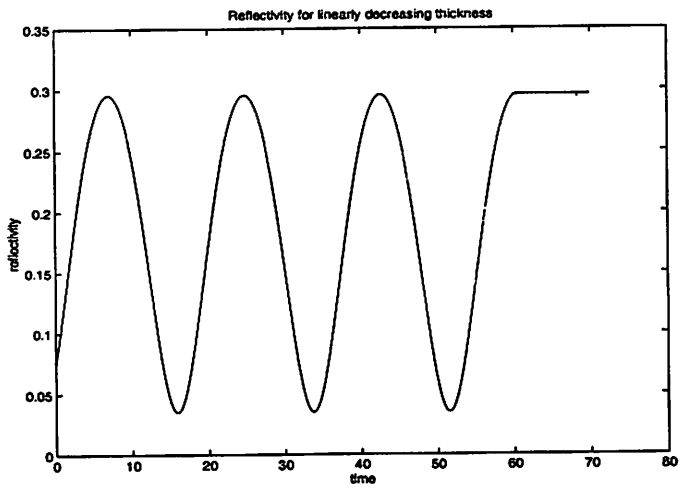


Figure 3.4: Simulated reflectivity for linearly decreasing thickness

The simplest way to infer develop rate is to record the time between maxima and minima of this curve which are $\frac{1}{2}\frac{\lambda}{na}$ apart, where a is the develop rate. This method is not adequate for

real-time application and does not yield thickness. On the other hand the signal processing methods to detect a minimum are very robust and the results are accurate. In reality the develop rate decreases with depth into the resist and the distance between maxima and minima changes with time as can be seen from Figure 3.5.

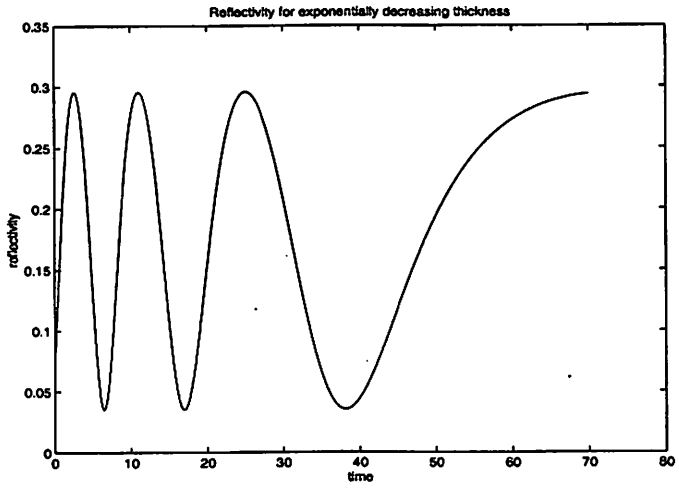


Figure 3.5: Simulated reflectivity for decreasing develop rate

Real data from a DRM is much more complex and also contains standing waves and noise such as (Figure 3.6). This figure shows all the complications that arise in real measurements, which are described in the following section.

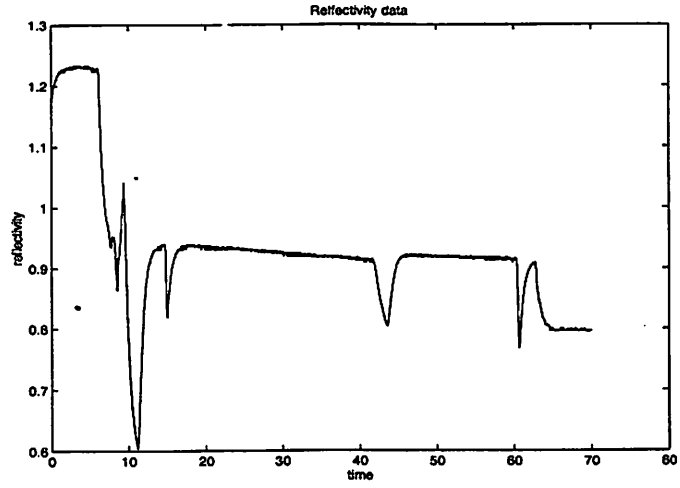


Figure 3.6: Measured reflectivity data from a DRM

3.3 Measurement Difficulties

There are two main problems associated with recording DRM data. First the DRM actually measures only the intensity of the reflected light, but the theoretical model of reflectance $\mathcal{R}$ versus thickness d depends also on the intensity of the light entering the film which therefore needs to be measured as well. In practice, we measure the incident light intensity needed to calibrate reflectance using a pure silicon wafer, which we assume to completely reflect the light and not absorb any intensity at long wavelengths. During this measurement the developer is present, so that it also captures the absorption of the developer when the light passed through it, while entering and exiting the film. The developer is a thick film and does not contribute to the interference pattern. We use the data taken on a pure silicon wafer not only to normalize reflectance but also to assess the noise levels (see Section 4.2)

The second problem is the combination of develop speed and the mechanics of the develop station. In a typical industrial process the main part of the develop takes place during the first 5-10 seconds. This is problematic, especially for a puddle develop tool, since the dispensing of developer happens during that period and this adds large amounts of noise to the measurements. It is difficult to track a rapidly changing variable such as thickness using very noisy data. Starting the measurement after this noisy initial period is also difficult, because of the lack of a good initial condition, to which most of the filtering methods are sensitive.

Chapter 4

Experimental Study

This section describes the experiments that we conducted at National Semiconductor Corporation in order to obtain DRM data. In addition, the amount of noise in the experimental data is estimated in order to evaluate algorithm performance. The discrepancies between the experimental data and the model are described.

4.1 Data Acquisition

We obtained the data for this section using a Site Services DRM-Lithacon. We conducted the experiment at National Semiconductor Corporation. Eight silicon wafers were coated with a 750nm thick resist film and exposed between $1.3 - 3.25 mJ/cm^2$. They were all baked at the same temperature for the same amount of time. We did not use anti-reflective coating and we used blanket exposures as opposed to a patterned mask. After the PEB step, we took the wafers out of the track and measured the thickness using a Tencor 1250SE ellipsometer. Thickness variations in the wafer are due primarily to the different exposure doses. Next, we returned the wafers into the track and developed them for 60 seconds. The DRM was mounted on top of the develop station as shown in Figure 4.1, and uses a broadband light source and records the reflected light at eight wavelengths, namely $704, 767, 800, 830, 840, 890, 930, 960 nm$.

Since the automatic trigger was not installed on this track, we initiated the measurements manually, which results in small delays that vary from wafer to wafer.

After the develop step, we measured the film thickness again using the 1250SE. The wafers with low exposure dose exhibited little loss of resist, whereas the ones with high

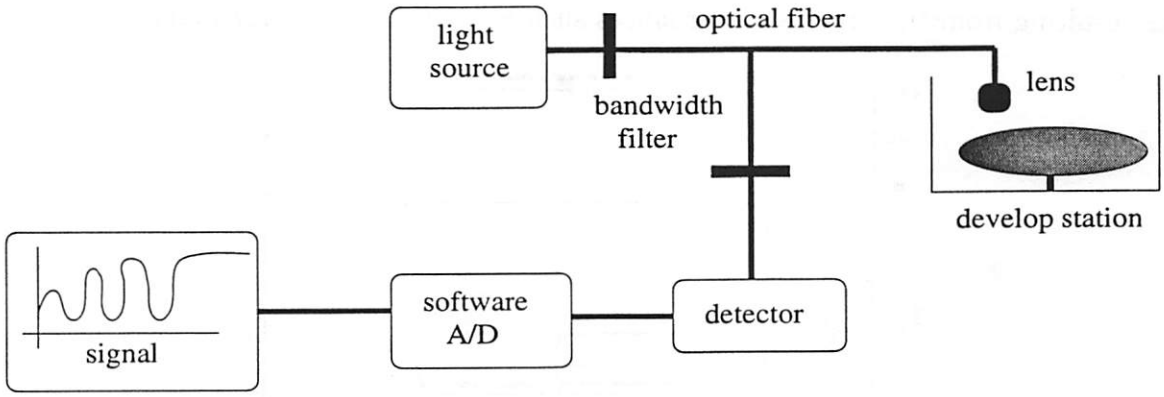


Figure 4.1: Set-up of the Site-Service Sensor

exposure dose were completely stripped of resist. For the wafer with medium exposure dose the remaining thickness can be used to confirm the results from the various thickness estimation schemes.

In addition to these wafers, we used a blank silicon wafer to record DRM data to provide a reference for the incident light intensity. This measurement was also used to evaluate measurement noise. The data can also be used to estimate the correlation of the noise between the eight measured wavelengths.

4.2 Noise Model

The Site Services reflectometer records reflected intensity data at eight wavelengths. Initially, it appears that the measurement noise in each channel should be independent from the others. However, the influence of the motion of the developer adds some correlation. To investigate this, we compare the signals at different wavelengths from a pure silicon wafer with each other. There are two noticeable effects shown in Figure 4.2. First there seems to be significant correlation between the noise channels that needs to be taken into account in the models and algorithms. Secondly there are spikes at similar times in each channel and for each wafer, independent of the resist thickness. This can be seen by comparing Figure 4.2 and Figure 4.3. In both figures there are spikes at $10sec$, $15sec$, $42sec$ and $60sec$. Some of theses spikes are caused by the wafer motion, e.g. at $45sec$ the wafer is rotated and at $60sec$ it is removed from the develop station. The spikes in the beginning of the develop are due to the dispense of the developer. Spikes resulting from the process should be disregarded, but it is difficult to automatically differentiate between maxima due to thickness change and

spikes resulting from the process disturbances such as motion of the developer.

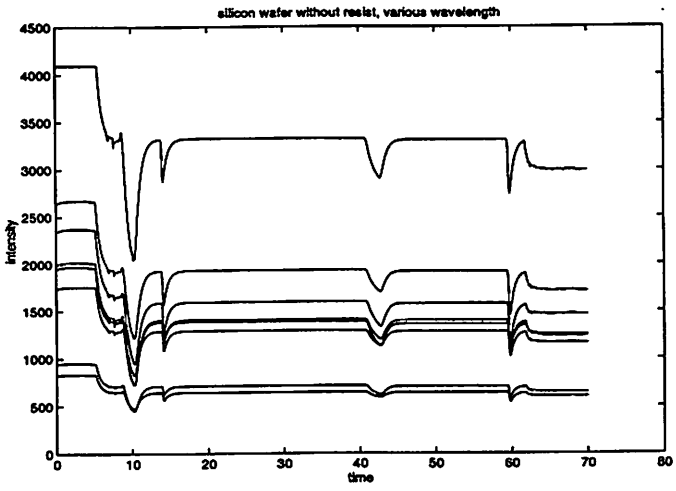


Figure 4.2: DRM measurement of a pure silicon wafer. The steady state section of this measurement is used as a reference.

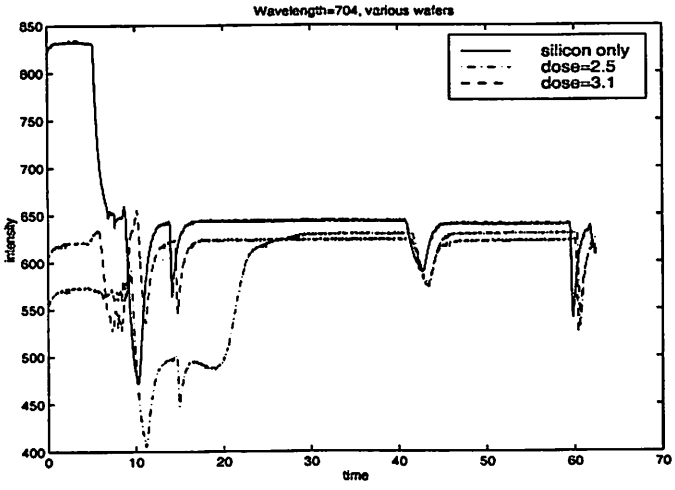


Figure 4.3: DRM measurement of wafers with different exposure doses

We computed the covariance matrix over the stationary portion of the eight signals from the pure silicon wafer shown in Figure 4.2 to estimate the size of the noise and the correlation between the channels.

Let σ_v be the diagonal elements of the noise covariance matrix and Λ_v the normalized cross covariance, so that

$$C = diag(\vec{\sigma}_v)\, \Lambda_v\, diag(\vec{\sigma}_v) \tag{4.2.1}$$

where C is the covariance matrix of the noise signal and $diag(\vec{\sigma}_v)$ is a matrix whose diagonal is the vector $\vec{\sigma}_v$ and all other elements are zero. In this experiment

$$\vec{\sigma}_v^T = [0.56\ 0.72\ 1.18\ 0.50\ 4.11\ 1.32\ 0.13\ 1.62] \tag{4.2.2}$$

and

$$\Lambda_v = \begin{pmatrix} 1 & -0.33 & -0.01 & 0.25 & \cdots \\ -0.33 & 1 & 0.24 & -0.52 & \cdots \\ -0.018 & 0.25 & 1 & -0.026 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \cdots \end{pmatrix} \tag{4.2.3}$$

The largest off-diagonal element of Λ_v is 0.52 and this indicates that there is quite a bit of correlation in the channels. However, no pattern emerges to explain the source of this correlation.

4.3 Modeling Errors

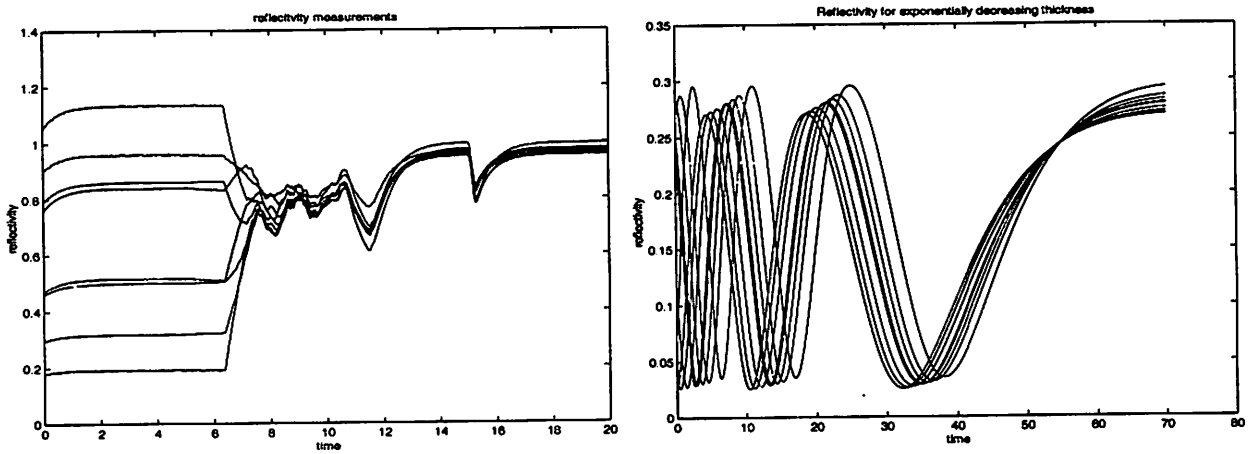


Figure 4.4: Measured and simulated reflectivity data

This section describes the differences between the experimental data and the available model. During this study we assume that the model for reflectance derived in section 3.2 describes the true system yielding the experimental data. Comparing the experimental data with a set of data simulated using this reflectance model (Figure 4.4) raises questions about this assumption. The differences cannot only be attributed to using the wrong parameters for the simulation. The shape of the experimental data does not resemble the simulated data, even after discounting the noise spikes described in the previous section. For these reasons, we believe that a different model for reflectivity or directly for the measured intensity has to

be found in order to succeed in extracting thickness or develop rate. In order to find a better model for reflectivity, many experiments need to be conducted that use methods similar to the poor mens DRM to measure thickness.

Because of this discrepancy in the available model we only tested some of the filtering methods on the experimental data. Most of the conclusions are drawn, based on our experience with simulated data.

4.4 The State Space Model

The measured intensity is modeled by the reflectivity described in Section 3.2 normalized by the measured incident intensity.

Any unknown parameter can be treated as a state of the system. In the simplest case the reflective index, or the Cauchy parameters are known and the only parameter to be estimated is the thickness d at each time.

There are many possible models for the dynamic behavior of the thickness d. The simplest is assuming a constant develop rate r

$$\vec{x}_{k+1} = \begin{pmatrix} d_{k+1} \\ r_{k+1} \end{pmatrix} = \begin{pmatrix} d_k - r_k \\ r_k \end{pmatrix} = \vec{f}(x_k) \tag{4.4.4}$$

where k denotes the time index and r is the develop rate.

A more realistic model assumes exponential decay in the thickness

$$\vec{x}_{k+1} = \begin{pmatrix} d_{k+1} \\ m_{k+1} \\ r_{k+1} \\ p_{k+1} \end{pmatrix} = \begin{pmatrix} m_k\, d_k - r_k \\ m_k \\ r_k \\ p_k \end{pmatrix} = \vec{f}(x_k) \tag{4.4.5}$$

where m is the exponential decay coefficient and p is the extra absorption coefficient introduced in Section 3.2.

On can conceive of increasingly complex models that include other parameters such as the Cauchy coefficients as states. The tradeoff between estimating more parameters and errors due to noise or the lack of the data limit the benefits of more complex models. The output of this state-space model is reflectivity R

$$y_k^i = \mathcal{R}(\lambda^i, d_k) + \sum_j C_{ij} e_k^j \tag{4.4.6}$$

where e is white noise and C is the covariance matrix of the noise channels. For simplicity we assume only additive output error. We are neglecting modeling error which would appear as a colored noise term adding to the state equation. The models described in this section, are used in the following chapter as the basis for the filtering methods that estimate the "state" thickness.

Chapter 5

Filtering methods

In this chapter we describe the filtering methods used to extract thickness. They include Extended Kalman filtering, non-linear Least Squares and filtering for output non-linearities.

5.1 Extended Kalman Filtering

The Extended Kalman Filter (EKF) is an estimation method for the state of a non-linear dynamic system. Unknown parameters in systems governed by non-linear ordinary differential equations can usually be estimated as artificial states.

The basic principle of the EKF is to use, at each sampling point, a linear Kalman Filter computed for the linearization of the non-linear system model at that time. Equivalently this algorithm:

- takes the non-linear system model at the current time
- linearizes around current state estimate
- computes Kalman Gain for linearized system
- computes new state estimate with linear KF

5.1.1 The Kalman Filter

The standard Kalman Filter is an optimal observer for the states of a linear systems driven by white noise. This type of system can be expressed in the following form:

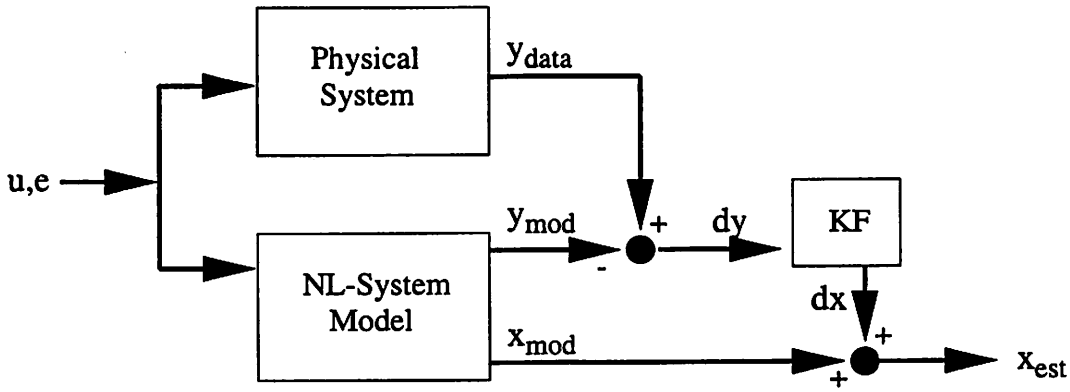


Figure 5.1: Schematic of the Extended Kalman Filter

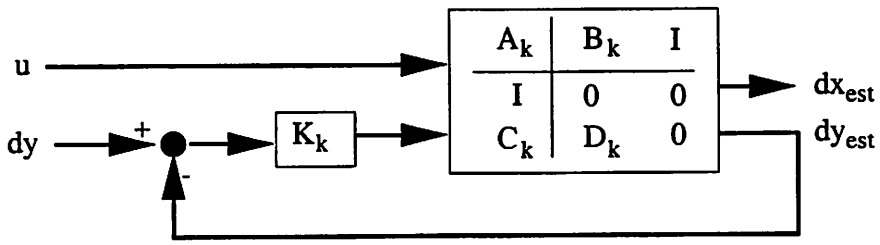


Figure 5.2: Schematic of a Kalman Filter

$$x_{k+1} = A_k x_k + B_k^u u_k + B_k^e e_k \tag{5.1.1}$$

$$y_k = C_k x_k + D_k^u u_k + D_k^e e_k \tag{5.1.2}$$

Here x_k denotes the system state at time k, y_k is the measured output, u_k is the known input and e_k is random noise. A, B, C and D are the matrices describing the dynamic system.

The general problem of estimating the system state x given data for u and y and a model A, B, C, D consist of solving the following minimization problem

$$\hat{x}_k = \arg\min_m \mathbf{E}[(x_k - m)(x_k - m)^*] \tag{5.1.3}$$

where $\mathbf{E}$ is the expected value.

The solution of this problem is the best estimate for x given the information y in a Maximum-Likelihood sense.

To solve this problem, we make the following standard assumptions

1. $\{A_k, B_k, C_k, D_k, u_k, y_k\}_0^\infty$ are known.

2. $x_0 \sim \mathcal{N}(0, P_0)$

3. $e = \{e_k\}_0^\infty$ is zero-mean, Gaussian, white noise with $\mathbf{E}[ee^*] = I$

4. x_0 and $\{e_k\}_0^\infty$ are uncorrelated.

Here, the crucial assumption is that e and x_0 are jointly Gaussian. The zero mean and unit variance assumptions can be met by scaling, and whiteness can be met by adding a coloring filter. It is eminently reasonable to assume that x_0 and e are uncorrelated in the DRM application. The computation relies on the correctness of the model and problems will arise in complex systems that include large amounts of modeling error.

The solution to the minimization problem in equation (5.1.3) is

$$\hat{x}_{k+1|k} = A_k\hat{x}_{k|k} + B_k^u u_k + K_k(y_k - C_k\hat{x}_{k|k} - D_k^u u_k) \tag{5.1.4}$$

where K is the Kalman gain matrix that is computed by solving Riccati equations (see Appendix). $\hat{x}_{k+1|k}$ is the optimal state estimate in the sense that it has the smallest covariance. For more information on Kalman Filtering see [5] and [6].

5.2 Non-linear Least Squares

There are many methods that can be applied to estimate thickness from the intensity measurement. The simplest method is to use non-linear least squares, which minimizes the Euclidean norm between the measured data and a non-linear model function depending on the minimization parameter. In this case

$$\min_d \frac{1}{2}\|y^{data} - \mathcal{R}(d)\|^2 \tag{5.2.5}$$

where y^{data} is the measured intensity vector normalized by the measured incident intensity.

This minimization is done at every time point. Again, there are many methods to solve the minimization itself. The method of choice for this problem is non-linear programming using a Gauss-Newton approximation as the search direction for a line search.

Non-linear programming uses iterative algorithms to solve $\min_\theta J(\theta)$. The basic steps are:

- Choose θ_0, set $k = 0$

- Choose search direction δ^k
- Solve $\hat{\alpha} = arg\ \min_\alpha J(\theta^k - \alpha\delta^k)$ and set $\theta^{k+1} = \theta^k - \hat{\alpha}\delta^k$
- Check stopping rule
- stop or $k = k + 1$ and repeat step 2

Two common methods are the Steepest Descent Method which uses the gradient $g = \frac{\partial J}{\partial \theta}$ of J as a search direction and secondly the Newton Method which uses the inverse Hessian times the gradient $H^{-1}g$. To be a valid choice the search direction has to be a descent direction, i.e. the cost has to decrease in that direction. This is always true for the Steepest Descent method but not necessarily for the Newton Method.

Since one searches in a different direction at the next iteration, the line search itself in each step does not need to be solved exactly.

In addition to the choice of search direction, there are many variables in these algorithms that have to be determined by the user. These include the initial guess for θ, the stopping criteria for the line search and the stopping accuracy for the whole algorithm.

In this particular application, we choose the Gauss-Newton approximation to the Hessian as a search direction δ, which is computed by

$$J(d) = \frac{1}{2}[\mathcal{R}(d) - y^{data}]^*[\mathcal{R}(d) - y^{data}] \tag{5.2.6}$$

$$Q = \frac{\partial \mathcal{R}}{\partial d} \tag{5.2.7}$$

$$g = \frac{\partial J}{\partial d} = Q^*[\mathcal{R}(d) - y^{data}] \tag{5.2.8}$$

$$H = \frac{\partial^2 J}{\partial d^2} \approx Q^*Q \tag{5.2.9}$$

$$\delta = (Q^*Q)^{-1}Q^*[\mathcal{R}(d) - y^{data}] \approx H^{-1}g \tag{5.2.10}$$

The advantage of this method is that the search direction is easily computed, and that it is always a descent direction. Convergence analysis is difficult, because the line search is often ad-hoc. We chose the initial guess of thickness at each time step to be the computed estimate of thickness of the previous step. This assumes that thickness is changing slowly compared to the data acquisition rate which, with 40 samples per second, appears reasonable.

The disadvantage of this method is that it does not take the dynamic behavior of the thickness into account, i.e. there is no model on how thickness changes with time, and consecutive time steps are only connected by the choice of initial thickness. Also the method does not account for noise and is not computationally suitable for full spectrum DRMs because the involved matrices become too large.

5.3 Filtering for Output Non-linearities

This section describes an ad hoc filtering method that can be used for systems governed by a linear state equation, but with a non-linear output equation. To use this method we assume that the thickness changes linearly with time. The model here is:

$$d_{k+1} = d_k - r_k \tag{5.3.11}$$

$$r_{k+1} = r_0 + w_k \tag{5.3.12}$$

$$y_k^i = g^i(d_k) + v_k^i \tag{5.3.13}$$

where d is the thickness, r is the develop rate, y^i is the reflectivity indexed by the wavelength λ^i and w and v are white noise sequences.

The filter used is based on a second order approximation of the error, and assumes knowledge of the gradient of the output function g. The filtering equations are:

$$\hat{d}_{k+1} = \hat{d}_k - \hat{r}_k - L(\hat{y}_k^i - y_k^i) \tag{5.3.14}$$

$$\hat{r}_{k+1} = r_0 \tag{5.3.15}$$

$$\hat{y}_k^i = g^i(\hat{d}_k) \tag{5.3.16}$$

where $\hat{(\,)}$ denotes estimated variables and L is the filtering weight.

Let the errors be denoted with $\tilde{(\,)}$, then

$$\tilde{d}_{k+1} = \hat{d}_{k+1} - d_{k+1} \approx (1 - L\nabla g)\tilde{d}_k - \tilde{r}_k + L\,v_k^i \tag{5.3.17}$$

$$\tilde{r}_{k+1} = \hat{r}_{k+1} - r_{k+1} = -w_k \tag{5.3.18}$$

$$\tilde{y}_k^i = \hat{y}_k^i - y_k^i \approx \nabla g\ \tilde{d}_k - v_k^i \tag{5.3.19}$$

where high order terms are assumed to be negligible. In choosing the weight L appropriately, the errors decay to zero. The appropriate choice is for the eigenvalues of the dynamic matrix of the errors to be inside the unit circle.

The eigenvalues μ are determined by the equation

$$\mu^2 - (1 - L\nabla g)\mu = 0 \tag{5.3.20}$$

The optimal choice is

$$L\nabla g = 1 \qquad or \qquad L^i = \frac{1}{n}\frac{1}{\frac{\partial g^i}{\partial d}} \tag{5.3.21}$$

where n is the number of wavelengths that are being measured. This choice of solution gives equal weight to all noise channels. If the data came from the exact model, the estimate would converge extremely fast, unless the noise takes the estimate out of the region where the first order approximation is valid.

Chapter 6

Results

In this chapter the various filtering methods described previously are tested on simulated and experimental data. Their advantages and disadvantages are discussed.

6.1 Extended Kalman Filtering

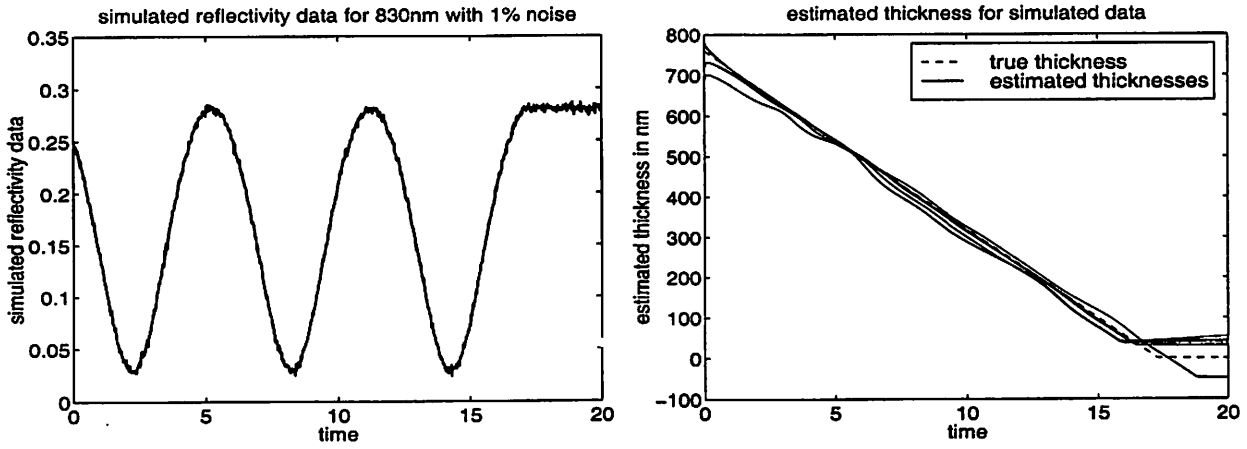


Figure 6.1: Simulated data with 1% noise and estimated thickness using random 5% deviations from true initial conditions .

Because of the difficulties with the measured data described in section 4.3, the filtering methods have to be evaluated using simulated and experimental data. First we use simulated data to evaluate the Extended Kalman Filtering algorithm. White noise is added to the reflectivity, which is simulated using the linear state equations (6.1.1). The three states (thickness, absorption and develop rate) are estimated from this data using initial conditions

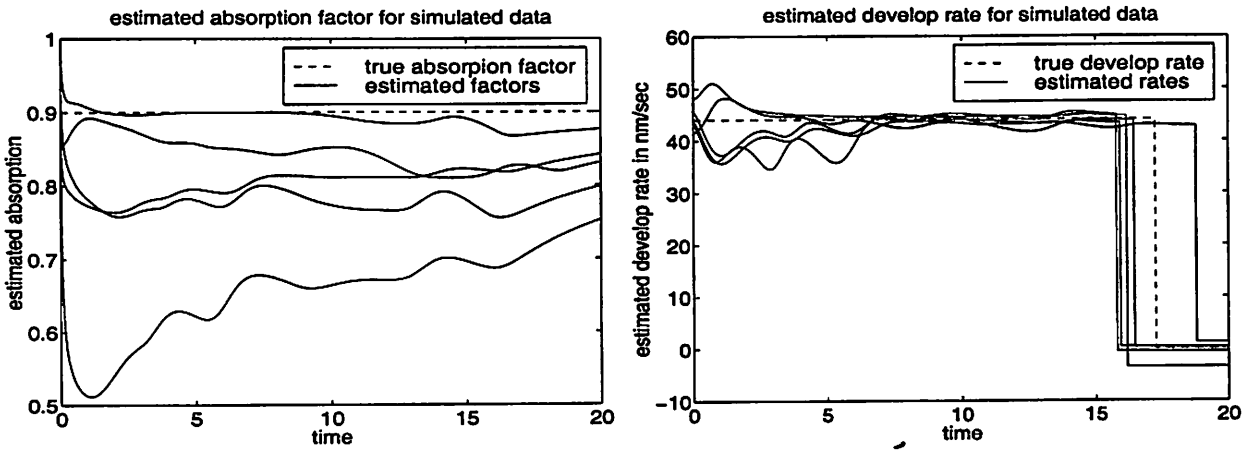


Figure 6.2: Estimated absorption coefficient and develop rate from simulated data with 5% noise using random 5% deviations from true initial conditions.

randomly distributed around the simulation conditions.

$$\vec{x}_{k+1} = \begin{pmatrix} d_{k+1} \\ r_{k+1} \\ p_{k+1} \end{pmatrix} = \begin{pmatrix} d_k - r_k \\ r_k \\ p_k \end{pmatrix} = \vec{f}(x_k) \tag{6.1.1}$$

The first simulation had 1% noise added to the reflectivity data (Figure 6.1). The estimated states shown in Figure 6.2 used initial conditions deviating 5% from the true initial conditions used in the simulations. The thickness and develop rate are estimated well, but the absorption coefficient is not a good estimate of the true value.

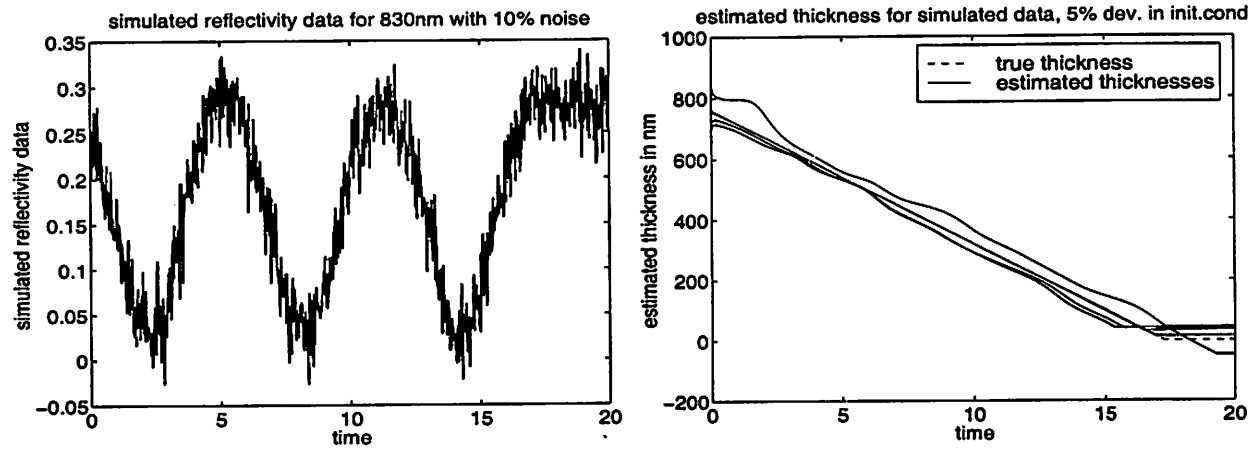


Figure 6.3: Simulated data with 10% noise and estimated thickness using random 5% deviations from true initial conditions.

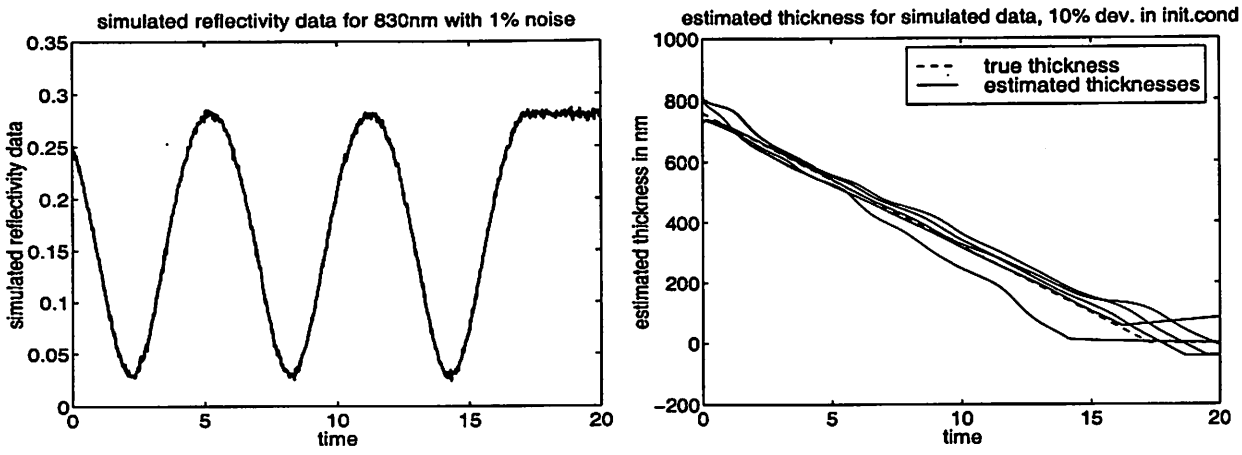


Figure 6.4: Simulated data with 1% noise and estimated thickness using random 10% deviations from true initial conditions.

Figure 6.3 shows the same estimation of thickness using simulated data with 10% noise added. The thickness continues to be estimated accurately. This shows that the EKF algorithm rejects output noise well as predicted. Figure 6.4 shows an estimation using the same reflectivity data, but with initial conditions deviating by 10% from the true values. The estimation is much less accurate, which shows the sensitivity of the EKF to initial conditions.

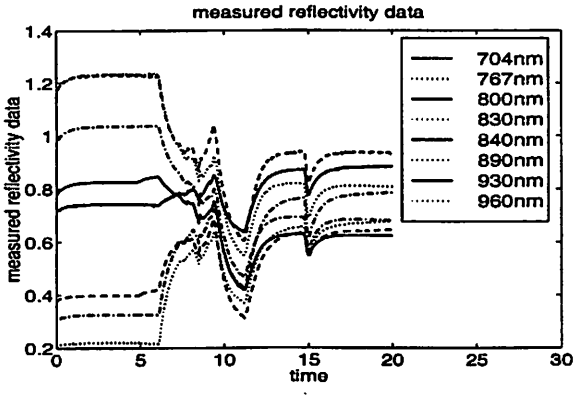


Figure 6.5: Measured reflectivity data with exposure dose $1.9mJ/cm^2$.

We also tested the EKF algorithm on the measured data shown in Figure 6.5. The mean initial condition for the estimation was determined from the thickness measured before develop and the end-point of 15sec that was extracted from the data. These initial conditions were also used to simulate data for the previous section.

To assess the robustness of the algorithm we also estimated the thickness with a different mean initial condition for develop rate using the same measured data set. The two thickness

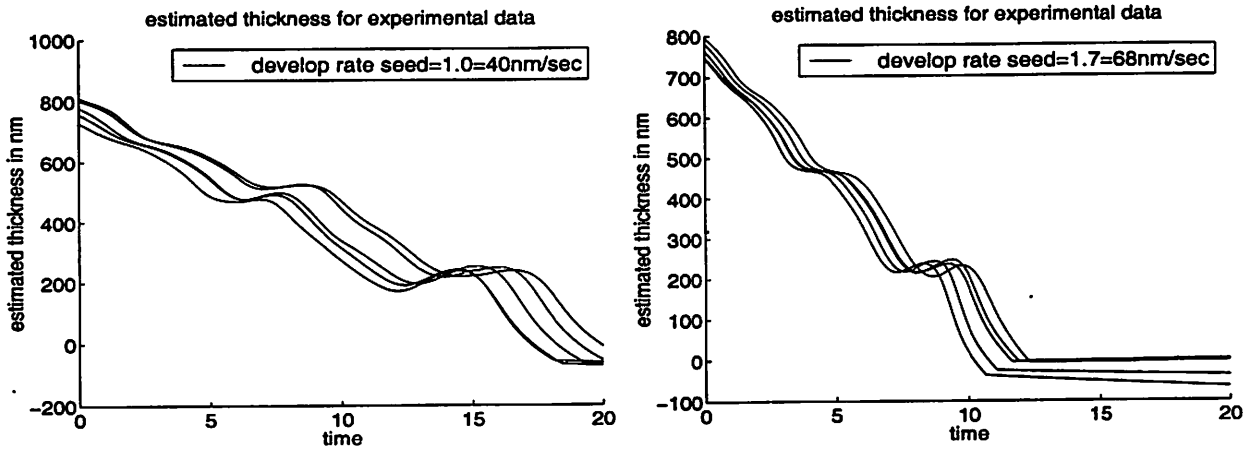


Figure 6.6: Estimated thickness from measured data with two different initial develop rates for five random initial states

estimates are shown in Figure 6.6. Although the same data is used, both estimates follow the path prescribed by the model, instead of both converging to the path prescribed by the data. This shows that the EKF method is not robust to modeling errors. It follows the faulty model and does not compensate for the discrepancy between the data and that model.

The first plot in Figure 6.6 and the associated develop rate and absorption estimates shown in Figure 6.7, suggest that the algorithm works well. However, the estimates follow any prescribed trajectory from the state equations and only adapt for random noise. Figure 6.7 also demonstrates the dynamic behavior of the filter. The overshoot in the estimate of develop rate is damped out by the filter dynamics.

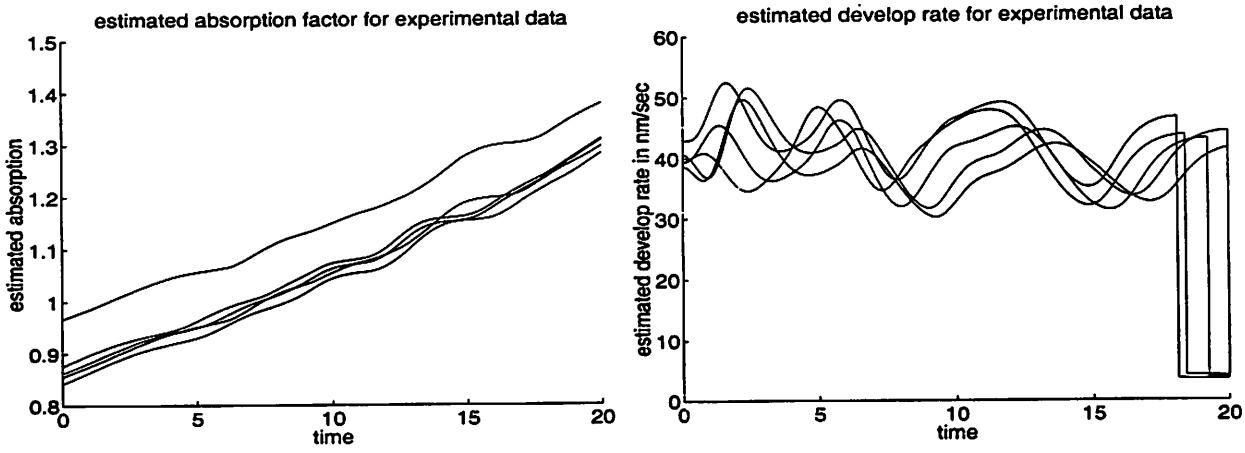


Figure 6.7: Estimated absorption and develop rate using measured data for five random initial states

6.2 Non-linear Least Squares

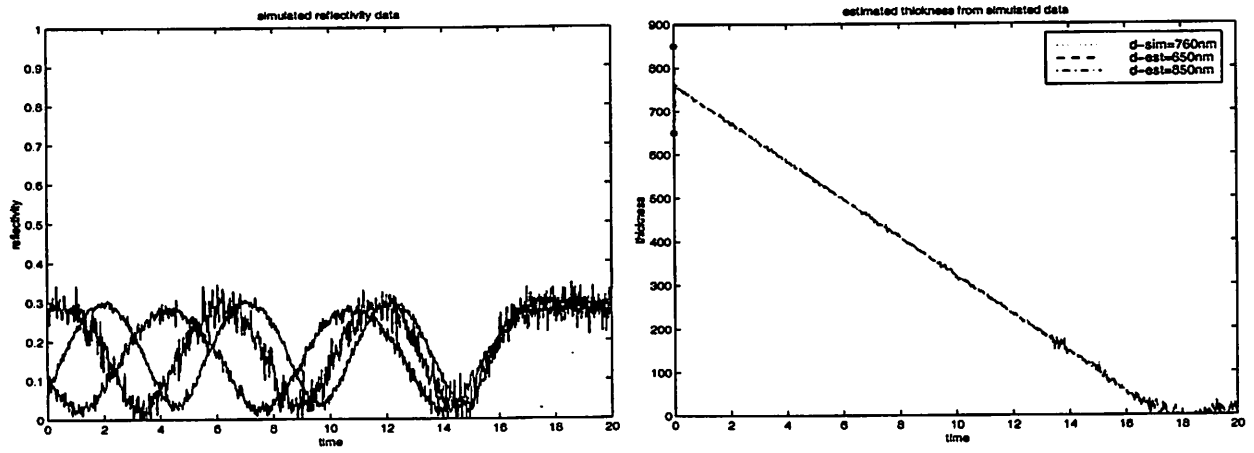


Figure 6.8: Simulated reflectivity data and estimated thickness using initial thickness $d_0^{sim} = 760nm$ and $d_0^{est} = 650nm$.

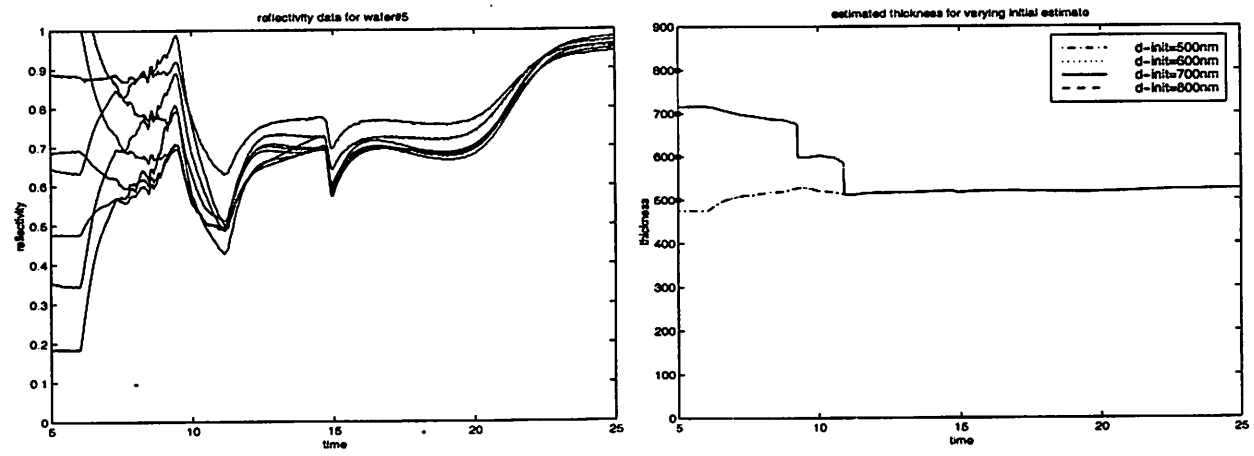


Figure 6.9: Measured reflectivity data with $dose = 2.5mJ/cm^2$ and estimated thickness using varying initial conditions.

First, we tested this method on simulated data. We simulated the reflectivity data using a linear thickness model from equation (4.4.4) and a fixed absorption coefficient. Normally distributed, uncorrelated noise was added to the data. The variance was proportional to the reflectance at the initial condition for each wavelength respectively. As shown in Figure 6.8 the noise added is not insignificant, but the method proves robust for initial estimates over $100nm$ from the true starting value, where the estimate recovers immediately.

The next step was to run the estimation on real data. We chose the data set from wafer

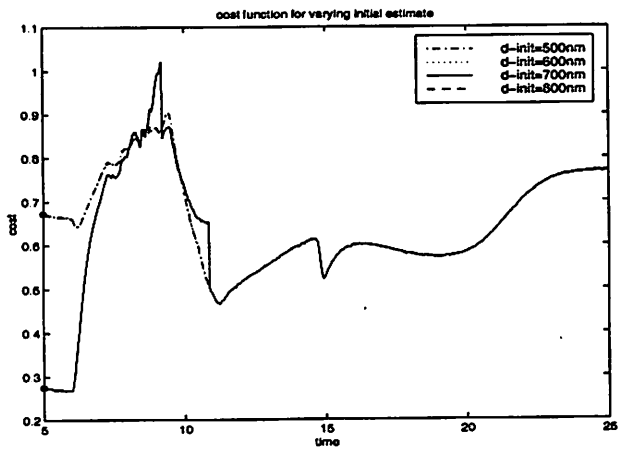


Figure 6.10: Cost function for varying initial guess using experimental data

#5 which had an exposure dose of $2.5mJ/cm^2$ and a measured thickness of $713nm$ before develop, and no resist remaining after develop.

Figure 6.9 shows that the real reflectivity data does not resemble the theoretical data shown in Figure 3.4. It is therefore not surprising that the estimation does not provide good results. One can observe from Figure 6.9 that the robustness of this estimation method still holds. Even though the four estimates are started with initial conditions with a range of $300nm$ they eventually return to the same estimate.

The remaining cost shown in Figure 6.10 is correlated to the spikes in the reflectivity data between $10 - 20sec$. This shows that the method is even robust to these disturbances from the process in the data. It appears that there is another source of error that accounts for the difference between the estimated thickness that plateaus around $500nm$, when it should return to zero according to the final measurement taken on the Tencor. This source might be a fundamental difference in the model for the reflectivity $\mathcal{R}$ and the true system that generates the intensity data. Overall the robustness of this method is promising if a better model can be provided.

6.3 Filtering for Output Non-linearities

Again, we first tested this algorithm on simulated data. In all models the simulation parameters were restricted to be positive, i.e. the develop rate and thickness are set to zero when the simulated thickness becomes negative. This raises the issue of on how to naturally restrict the estimation. There are two restrictions that should be incorporated. First the

estimated thickness should decrease and secondly it should be non-negative. These should not be strictly enforced because of noise in the data.

We chose to not let the estimated thickness increase by more then 5% and to set the estimated develop rate to zero, when the thickness estimate was negative. We added these restrictions to the state estimate equations (5.3.14)-(5.3.16).

We measured the noise levels for w and v in percent relative to the main signal. Let

$$r_k = r_0 + w_k = r_0(1 + \phi e_k) \tag{6.3.2}$$

where $e \sim \mathcal{N}(0,1)$, then x% noise corresponds to $\phi = \frac{x}{300}$ with a choice of 3σ bounds for the percentage. Similarly, we determined the noise level for v in percent of $\mathcal{R}$ and we chose the random initial condition of the estimation relative to the known initial condition of the simulation.

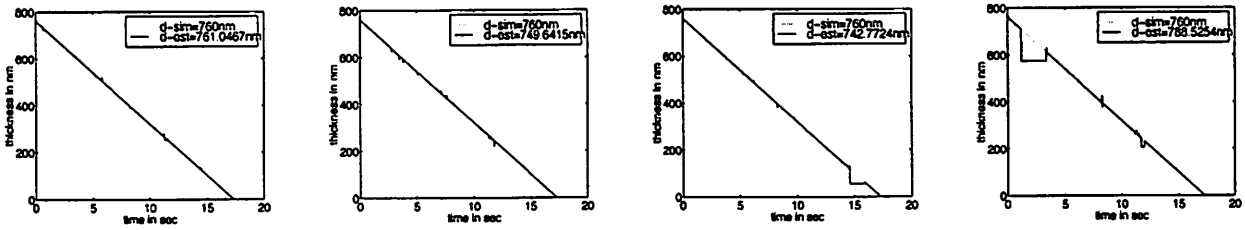


Figure 6.11: Estimated thickness from simulated data with various random initial thicknesses (5% variation) for uncorrelated noise levels $(w \sim 10\%, v \sim 1\%)$.

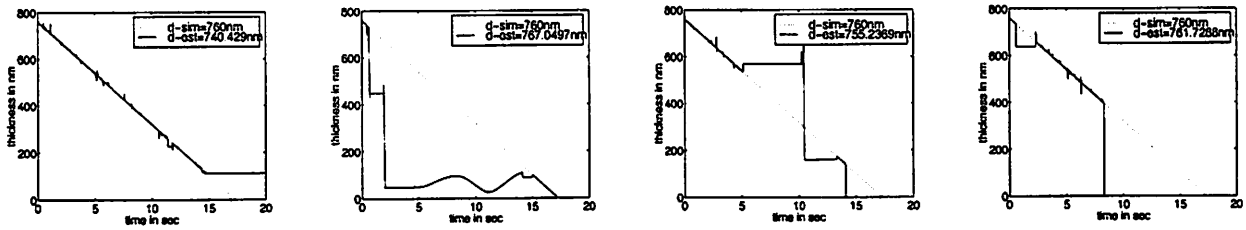


Figure 6.12: Estimated thickness from simulated data with various random initial thicknesses (5% variation) for uncorrelated noise levels $(w \sim 10\%, v \sim 5\%)$.

Figure 6.11 shows the simulated and estimated thickness for four random initial conditions varying within 5% of the real initial condition. The noise level for w and v were chosen to be 10% and 1% respectively.

The algorithm seems to be especially sensitive to output-noise, i.e. the noise level for v. For more than 1% noise in the output the estimate diverged for almost every run (see Figure 6.12).

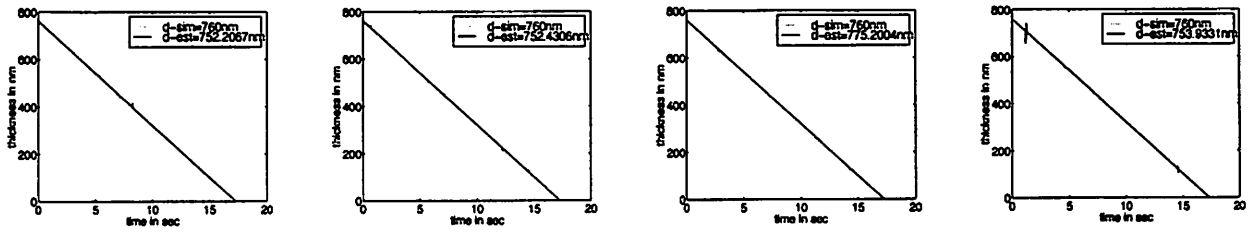


Figure 6.13: Estimated thickness from simulated data with various random initial thicknesses (5% variation) for correlated noise levels ($w \sim 10\%, v \sim 1\%$).

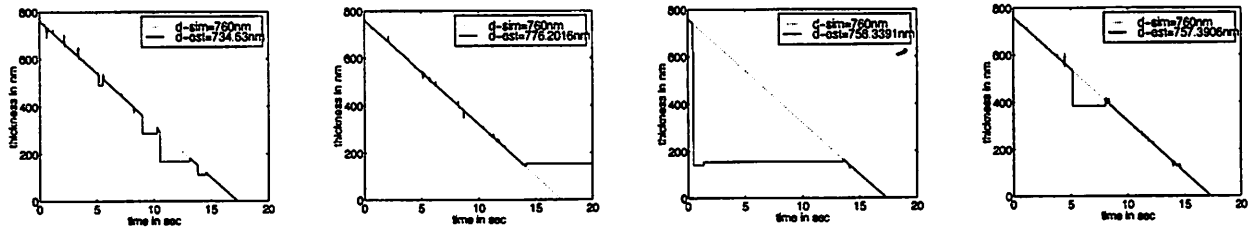


Figure 6.14: Estimated thickness from simulated data with various random initial thicknesses (5% variation) for correlated noise levels ($w \sim 10\%, v \sim 5\%$).

An increase in the modeling noise w, did not result in a significant reduction in the estimation accuracy. The algorithm is also not very sensitive to the initial thickness estimate.

In the first set of simulations the noise channels of the output were not correlated. Since they are correlated in reality, we attempted to see if this makes the estimation easier. For this second set of simulations half of the noise level was uncorrelated and the other half was completely correlated.

We used the same noise levels as in the previous plots for Figure 6.13 and Figure 6.14. There appears to be at most a slight improvement in tolerance.

Chapter 7

Simulation Study Using PROLITH

In this chapter we use PROLITH simulations to explore the sensitivity of the develop rate with respect to process input variables such as exposure dose and PEB time. The PROLITH simulator uses detailed models for all lithography processing steps starting from the aerial image formation to diffusion during the post exposure bake and the develop step. The modeling parameters for the resist are obtained experimentally and supplied by the manufacturer. The PROLITH GUIDE TO LITHOGRAPHY [3] provides an excellent summary of the models used and variables involved in each process step.

We used the simulated data obtained from PROLITH to evaluate the capabilities of a DRM as a sensor for run-to-run control.

7.1 The Simulations

We used Shipley's UV5 resist to run the simulations on silicon with AR2 anti-reflective coating to reduce standing wave effects.

The output of the simulation was thickness vs. time data for various exposure doses and PEB times, which we converted to develop rate vs. time using Euler differentiation. From this time signal (shown in Figure 7.3) we extracted three quality measures. These are the maximum develop rate and the mean develop rate (excluding regions of zero develop rate) and the initial develop rate at $time = 1.2sec$. While these have different numerical values, their general tendencies are similar. This is evident in Figures 7.1 and 7.2. Thus, either measure can be used as the quality measure in modeling the develop rate. We chose the maximum develop rate to conduct the sensitivity analysis in the following section. The

linear fit to the maximum develop rate is also shown in Figure 7.2.

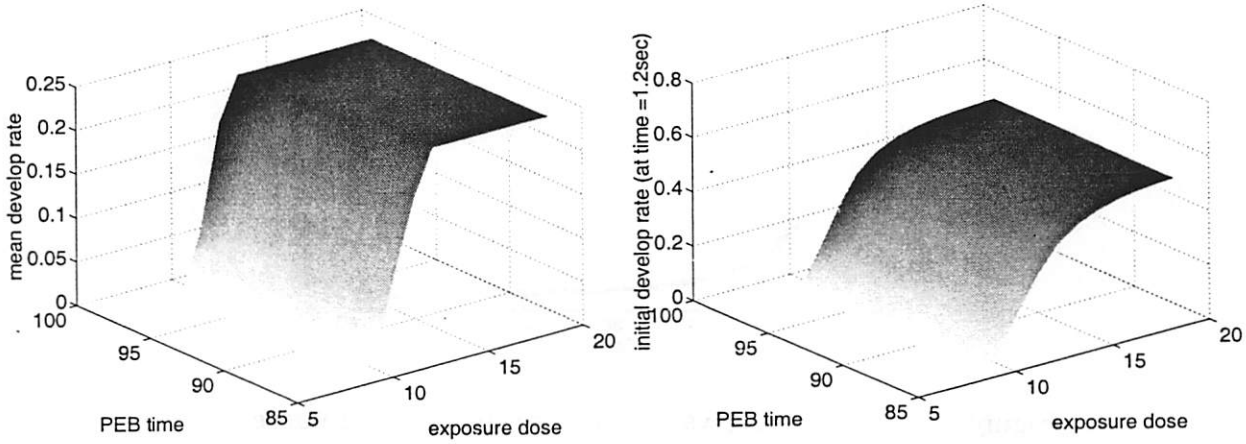


Figure 7.1: Average and initial develop rate vs exposure dose and PEB time.

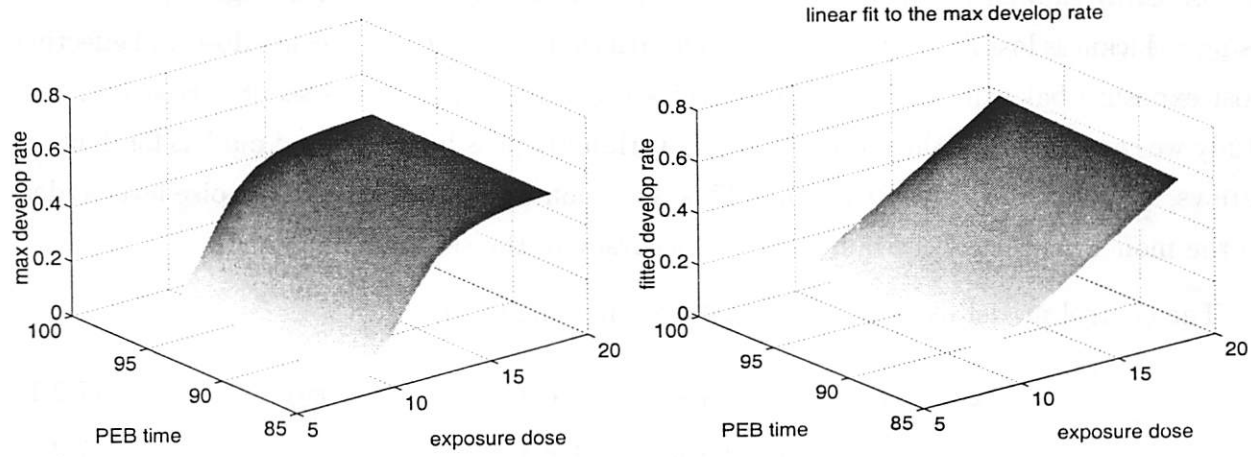


Figure 7.2: Maximum develop rate and its linear fit vs exposure dose and PEB time.

These figures also show the high sensitivity of develop rate to the process inputs. Figure 7.3 demonstrates how the develop rate changes during the develop process and that it is necessary to chose a single valued measurable for the purpose of control models.

7.2 Develop Rate as a Run-to-Run Control Measurable

Important questions remain to be answered. How valuable is a measurement of develop rate for a run-to-run control scheme? How does the difficulty with this measurement trade off with the needed accuracy?

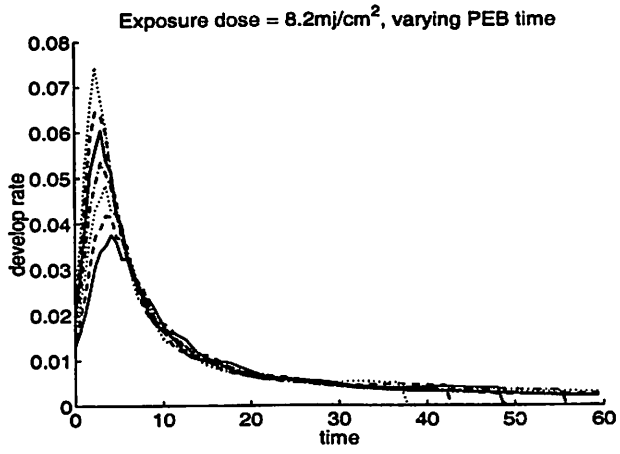


Figure 7.3: Develop rate vs time for varying exposure dose.

To answer these, we examined the use of a develop rate monitor for run-to-run control. MUSACCHIO showed in [4] the benefits of run-to-run control for the lithography sequence using a thickness loss measurement to model variations in effective exposure dose and effective post exposure bake time. For details on thickness loss measurements see [8]. Based on this study we can determine the reduction in CD variability possible with good models for develop rate vs. exposure dose and PEB time. This reduction would depend on the noise level added to the model, which is determined by the accuracy of the sensor.

The control model used by MUSACCHIO [4] has the form

$$z = a_{11}(x_1 + e_1) + \tilde{a}_{12}(x_2 + e_2) + \tilde{a}_{13}\sqrt{(x_2 + e_2)} \tag{7.2.1}$$

$$y_1 = a_{21}(x_1 + e_1) + a_{22}(x_2 + e_2) + a_{23} + e_3 \tag{7.2.2}$$

where z is the CD, x_1 and x_2 are the exposure dose and the PEB time respectively and y_1 is the square root of the measured thickness loss. The noise terms are e_1, e_2 and e_3. The problem is to estimate the z given the measurement of y_1 without knowledge of the dose or time but given the statistics for the noise terms.

The noise e_i is assumed to be normally distributed

$$e_i \sim \mathcal{N}(0, \sigma_i^2) \tag{7.2.3}$$

Adding a measurement for develop rate to equation 7.2.2 would change the variance of the estimated z.

$$y_2 = a_{31}(x_1 + e_1) + a_{32}(x_2 + e_2) + a_{33} + e_4 \tag{7.2.4}$$

To make the computations simpler the square root term in equation 7.2.1 is linearized about the operating point $(\bar{x}_1, \bar{x}_2)$

$$\sqrt{x_2} \approx (\frac{1}{2\sqrt{\bar{x}_2}})x_2 + (\frac{3}{2}\sqrt{\bar{x}_2}) \tag{7.2.5}$$

and the coefficients in equation 7.2.1 are combined to yield

$$\begin{pmatrix} z \\ y_1 \\ y_2 \end{pmatrix} = A \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} + \tilde{A} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix} \tag{7.2.6}$$

where x is known. So

$$\begin{pmatrix} z \\ y_1 \\ y_2 \end{pmatrix} \sim \mathcal{N}(Ax, \tilde{A}\Lambda\tilde{A}^*) \tag{7.2.7}$$

Using Theorem 1 from the appendix we can compute the variance of the conditioned variable $z|y$ which will depend on the variance of the measurement noise in the develop rate σ_4

$$\sigma_{z|y} = K_z - K_{zy}K_y^{-1}K_{yz} \tag{7.2.8}$$

where

$$K = \tilde{A}\Lambda\tilde{A}^* = \begin{pmatrix} K_z & K_{zy} \\ K_{yz} & K_y \end{pmatrix} \tag{7.2.9}$$

$$A = \begin{pmatrix} -0.016 & -0.00050.4321 & \\ 2.5176 & 0.0334 & -23.4615 \\ 0.0483 & 0.0033 & -0.5254 \end{pmatrix}, \quad \tilde{A} = \begin{pmatrix} -0.016 & -0.0005 & 0 & 0 \\ 2.5176 & 0.0334 & 1 & 0 \\ 0.0483 & 0.0033 & 0 & 1 \end{pmatrix} \tag{7.2.10}$$

Figure 7.4 illustrates the relationship between the variance of the estimated CD and the noise level in the develop rate sensor. To get a 10% improvement in the variance of the estimated CD one can tolerate at most 20% measurement noise for develop rate.

Figure 7.5 shows the standard deviation for the estimate of CD with varying measurements. From the model alone the standard deviation is $7.16nm$, but using the thickness loss sensor it can be reduced to $6.89nm$. Using only a measurement of develop rate with 10% noise one can reduce the standard deviation to $6.01nm$. Using both measurements the estimate has a standard deviation of $5.84nm$. The difference of improvement between the

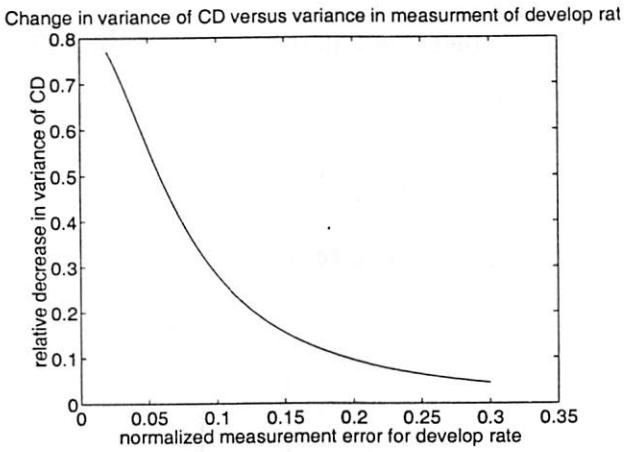


Figure 7.4: Relative change in Variance of estimated CD with various levels of measurement noise for develop rate.

two sensors is due to the assumed measurement accuracy. The model for the square root of thickness loss has about 20% measurement noise.

All of these numbers are dependent on the models for the dependence of develop rate on exposure dose and PEB time and they would need to be recalculated for models derived from experimental data. They do, however, capture the essential trend involved.

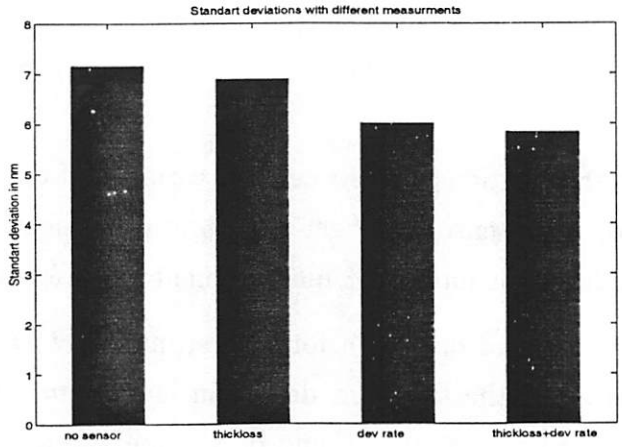


Figure 7.5: Standard Deviation of CD with different measurements.

Chapter 8

Conclusions

This study was intended to explore the potential of a develop rate monitor as an in-situ sensor for the purpose of controlling the DUV lithography sequence.

A number of problems appeared during the investigation. Some of these stem from the filtering methods, others from the physical process, and some are inherent to the DRM or DUV resists.

The principal difficulty with the physical process, which in this case was a puddle develop, is a low signal to noise ratio. During the dispense of the developer the signal is unrecognizable and cannot be used for thickness extraction. A good initial estimate of thickness is available before the develop begins, but lost after the dispense because DUV resists develop rapidly. Some of the filtering methods however, depend on a good initial estimate of thickness. It is this interplay of the problem sources that make this project challenging.

The advantage of using the develop step for an in-situ sensor is that the sensor is easily mounted on the track and does not interfere with the process. The disadvantage is that it is late in the process flow where damage might have already occurred. For spray develop stations, in which the wafer is spinning, the problem of changing location of the measurement becomes an issue. This can be solved by synchronizing the data acquisition rate with the rotational speed of the wafer.

Problems inherent to the DRM are the lack of calibration of the intensity of the light source and its sensitivity to measurement noise. This sensitivity will increase when dealing with patterned wafers, especially when the ratio of exposed to unexposed areas decreases. For rotating develop stations this is compounded by changes in the location of the small developing areas. Using DRMs with additional wavelengths might help to increase the

signal to noise ratio, but it will also increase the cost of the sensor and complexity of the algorithms.

The reason for attempting to use control on the DUV lithography sequence is the high sensitivity of chemically amplified resist to slight variations in the processing conditions. It is this inherent sensitivity that makes printing smaller feature sizes possible, but the same chemistry produces very short develop times, which causes difficulties with the filtering methods.

We explored three different filtering techniques in trying to extract thickness from reflectivity data. Each has inherent advantages and disadvantages. Extended Kalman Filtering and Non-linear Output Filtering are faster computationally than Non-linear Least Squares and exploit the dynamic behavior of thickness. Least Squares is less sensitive to initial conditions and output error. Its computational speed could be increased by using a more sophisticated line search algorithm. On the other hand, the speed will decrease when using more wavelengths and might not be practical for real time applications. The Output Filtering method does not allow for a more complex, non-linear model of the state, i.e. in the dynamic behavior of thickness and it is also very sensitive to noise in the measurement signal, which becomes problematic for patterned wafers. The EKF works very well, if the model used is accurate, and it is not sensitive to output noise. Modeling errors will render this method unreliable.

The sensitivity study using PROLITH simulations shows the potential of a good develop rate monitor. A DRM with 10% measurement noise would significantly reduce the variance of the estimate of CD. This can allow reducing the mean value of the produced CD, provided sufficient actuation authority. Still, the difficulties in obtaining such a sensor might favor other sensors that are easier to realize and provide equal improvement.

Overall there are many problems associated with the data acquisition that need to be solved. In spite of the noisy signals the Non-Linear Least Squares Method appears to be most promising because of its robustness. This technique only requires a good model for the source of the data. The challenge is to find a more realistic model for the reflectivity $\mathcal{R}$. Another project for future investigation would be to explore the end-point time as a modeling variable for a run-to-run control scheme. End-point is not nearly as complex to estimate as develop rate and may be sensitive enough to the process inputs, such as exposure dose and PEB time, to yield additional knowledge that can be used in the control algorithm.

Bibliography

[1] SAUTTER, K.M., HA, M. AND BATCHELDER, T.
Photoresist development Process Control And Optimization Utilizing and End Point Monitor
Microelectronic Manufacturing and Testing,11-14 (September 1989)

[2] THOMSON, M.
The Use of In-Situ Develop End Point Control to Eliminate Variance
Xinix Product Literature, Luxtron(1990)

[3] MACK, C.A.
Inside PROLITH^{TM}, A Comprehensive Guide to Optical Lithography Simulation
FINLE Technologies, Inc., (1997)

[4] MUSACCHIO, J.
Run-to-Run Control for Lithography
Masters Report, UC Berkeley, (1998)

[5] LJUNG, L.
System Identification: Theory for the User
Prentice Hall PTR, (1987)

[6] BALAKRISHNAN, A.V.
Kalman Filtering Theory
University Series in Modern Engineering, Springer, (1984)

[7] BORN, M., WOLF, E.
Principles of optics:electro-magnetic theory of propagation, interference and diffraction of light
Pergamon Press, Oxford New York, (1989)

[8] JAKATDAR, N.
In Situ Metrology for Deep Ultraviolet Photolithography Control
Masters Report, UC Berkeley, (1997)

[9] JAKATDAR, N., NIU, X., SPANOS, C.J.
Characterization of a Chemically Amplified Photoresist for Simulation using a Modified "Poor Man's DRM" Methodology
SPIE Metrology, Inspection and Process Control for Microlithography XII, Volume 3332, pg 578-585 (1998)

Appendix A

Appendix

A.1 Derivation of the Kalman Filter

Define the collection of given information as

$$\underline{\overline{I}}_k = \{y_0, y_1, \cdots, y_{k-1}\} \tag{A.1.1}$$

The goal of the Kalman filter is to find

$$\hat{x}_{k|k} = \mathbf{E}[x_k \,|\, \underline{\overline{I}}_k] \tag{A.1.2}$$

and the covariance matrix of this random variable where | denotes conditioning with given data.

Let us introduce some short-hand notation

$$x_{k|j} = x_k|\,\underline{\overline{I}}_j = \text{conditioned random variable} \tag{A.1.3}$$

$$\hat{x}_{k|j} = \mathbf{E}[x_{k|j}] = \text{conditional mean} \tag{A.1.4}$$

$$P_{k|j} = \mathbf{E}[(x_{k|j} - \hat{x}_{k|j})(x_{k|j} - \hat{x}_{k|j})^*] = \text{conditional covariance} \tag{A.1.5}$$

For the derivation of the Kalman Filter we will need the following two results:

Lemma 1

For any random variables X, Y, Z we have

$$X|(Y,Z) = (X|Y)|(Z|Y)$$

Theorem 1

Let X be a random n-vector, Y be a random m-vector, and let

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(m_Z, \Lambda_{zz}) = \mathcal{N}\left(\begin{bmatrix} m_X \\ m_Y \end{bmatrix}, \begin{bmatrix} \Lambda_{XX} & \Lambda_{XY} \\ \Lambda_{YX} & \Lambda_{YY} \end{bmatrix} \right)$$

Then, $X|Y \sim \mathcal{N}(m_{X|Y}, \Lambda_{XX|Y=y})$ where

$$m_{X|Y} = m_X + \Lambda_{XY}\Lambda_{YY}^{-1}(y - m_Y) \tag{A.1.6}$$

$$\Lambda_{XX|Y} = \Lambda_{XX} - \Lambda_{XY}\Lambda_{YY}^{-1}\Lambda_{XY}' \tag{A.1.7}$$

Conditioning the system equations (5.1.1)-(5.1.2) on the data leads to

$$\begin{aligned} x_{k+1|k} &= A_k x_{k|k} + B_k^u u_{k|k} + B_k^e e_{k|k} = A_k x_{k|k} + B_k^u u_k + B_k^e e_k \\ y_{k|k} &= C_k x_{k|k} + D_k^u u_k + D_k^e e_k \end{aligned} \tag{A.1.8}$$

Here we used the fact that u_k is completely known and not random and that e_k is independent from previous output data through y_{k-1}. Now define

$$z = \begin{bmatrix} x_{k+1|k} \\ y_{k|k} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} A_k\hat{x}_{k|k} + B_k^u u_k \\ C_k\hat{x}_{k|k} + D_k^u u_k \end{bmatrix}, \begin{bmatrix} \alpha_k & \beta_k \\ \beta_k^* & \gamma_k \end{bmatrix} \right) \tag{A.1.9}$$

where

$$\begin{bmatrix} \alpha_k & \beta_k \\ \beta_k^* & \gamma_k \end{bmatrix} = \begin{bmatrix} A_k P_{k|k} A_k^* + B_k^e B_k^{e*} & A_k P_{k|k} C_k^* + B_k^e D_k^{e*} \\ C_k P_{k|k} A_k^* + D_k^e B_k^{e*} & C_k P_{k|k} C_k^* + D_k^e D_k^{e*} \end{bmatrix} \tag{A.1.10}$$

$$= \begin{bmatrix} M_k \end{bmatrix} \begin{bmatrix} P_{k|k} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} M_k^* \end{bmatrix} \tag{A.1.11}$$

and

$$\begin{bmatrix} M_k \end{bmatrix} = \begin{bmatrix} A_k & B_k^e \\ C_k & D_k^e \end{bmatrix} \tag{A.1.12}$$

With the use of Lemma 1 we can define an expression for $x_{k+1|k+1}$,

$$x_{k+1|k+1} = x_{k+1}|(y_0, \ldots, y_{k-1}, y_k) \tag{A.1.13}$$

$$= (x_{k+1}|y_0, \ldots, y_{k-1}) \,|\, (y_k|y_0, \ldots, y_{k-1}) \tag{A.1.14}$$

$$= \left(x_{k+1|k}\right) | \left(y_{k|k}\right) \tag{A.1.15}$$

Using Theorem 1 we can compute the mean value and the covariance for $x_{k+1|k+1}$,

$$\hat{x}_{k+1|k+1} = (\hat{x}_{k+1|k}) \,|\, (y_{k|k}) = A_k \hat{x}_{k|k} + B_k^u u_k + \beta_k \gamma_k^{-1} \left(y_k - C_k \hat{x}_{k|k} - D_k^u u_k \right) \quad \text{(A.1.16)}$$

$$P_{k+1|k+1} = \alpha_k - \beta_k \gamma_k^{-1} \beta_k^* \quad \text{(A.1.17)}$$

Comparing equation (A.1.16) with equation (5.1.4) it follows that the Kalman Gain K_k is defined by

$$\begin{aligned} K_k &= \beta_k \gamma_k^{-1} \\ &= (A_k P_{k|k} C_k^* + B_k^e D_k^{e*})(C_k P_{k|k} C_k^* + D_k^e D_k^{e*})^{-1} \end{aligned} \quad \text{(A.1.18)}$$

Substitution of equation (A.1.18) into equation (A.1.17) then leads to the following Riccati equation:

$$\begin{aligned} P_{k+1|k+1} = A_k P_{k|k} A_k^* + B_k^e B_k^{e*} - \\ (A_k P_{k|k} C_k^* + B_k^e D_k^{e*})(C_k P_{k|k} C_k^* + D_k^e D_k^{e*})^{-1}(C_k P_{k|k} A_k^* + D_k^e B_k^{e*}) \end{aligned} \quad \text{(A.1.19)}$$

which iteratively updates $P_{k|k}$. This update is then substituted into equation (A.1.10) to compute α_k, β_k and γ_k for the next time step.

For time invariant systems, often the steady state Riccati equation

$$P = APA^* + B^e B^{e*} - (APC^* + B^e D^{e*})(CPC^* + D^e D^{e*})^{-1}(CPA^* + D^e B^{e*}) \quad \text{(A.1.20)}$$

is solved to reduce computational load.

A.2 Derivatives of the Reflectivity Function

Most optimization algorithms use the derivative of the function that is being optimized, in this case the reflectivity function $\mathcal{R}$ described in Section 3.2 to be

$$M = \begin{pmatrix} \cos(\frac{2\pi}{\lambda} n\, d) & -\frac{i}{n}\sin(\frac{2\pi}{\lambda} n\, d) \\ -i\, n \sin(\frac{2\pi}{\lambda} n\, d) & \cos(\frac{2\pi}{\lambda} n\, d) \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \tag{A.2.21}$$

$$r = \frac{(m'_{11} + m'_{12}\, n_s) n_{air} - (m'_{21} + m'_{22}\, n_s)}{(m'_{11} + m'_{12}\, n_s) n_{air} + (m'_{21} + m'_{22}\, n_s)} \tag{A.2.22}$$

$$\mathcal{R} = p\,|r|^2 \tag{A.2.23}$$

The derivative of this function with respect to a parameter θ such as the thickness d or the absorption coefficient p can be derived analytically instead of numerically. The numerical computation would add computation time and result in less accuracy.

The derivative is developed by repeatedly applying the chain rule. This sequence of equations can easily be implemented to compute $\frac{\partial \mathcal{R}}{\partial \theta}$

Using the shorthand notation

$$c = \cos(\frac{2\pi}{\lambda} nd) \tag{A.2.24}$$

$$s = \sin(\frac{2\pi}{\lambda} nd) \tag{A.2.25}$$

$$r = r_R + i\, r_I \tag{A.2.26}$$

$$n_a = n_{air} \tag{A.2.27}$$

$$n_s = n_{silicon} \tag{A.2.28}$$

We start by expanding equation (A.2.22)

$$\begin{aligned} r &= \frac{(c + i\frac{s}{n} n_s) n_a - (i\, s\, n + c\, n_s)}{(c + i\frac{s}{n} n_s) n_a + (i\, s\, n + c\, n_s)} \\ &= \frac{c^2(n_a^2 - n_s^2) + s^2(\frac{n_s^2 n_a^2}{n^2} - n^2) + i\left[2\, c\, s(\frac{n_s^2 n_a}{n} - n_a n)\right]}{c^2(n_a + n_s)^2 + s^2(\frac{n_s n_a}{n} + n)^2} \\ &= \frac{w_R + i\, w_I}{v} \end{aligned} \tag{A.2.29}$$

So that

$$r_R = \frac{w_R}{v} \tag{A.2.30}$$

$$r_I = \frac{w_I}{v} \tag{A.2.31}$$

$$\tag{A.2.32}$$

where w_R and w_I represent the real and imaginary parts of the numerator and v the denominator of r.

Now we can start applying the chain rule to equation (A.2.23),

$$\frac{\partial \mathcal{R}}{\partial \theta} = \frac{\partial p}{\partial \theta}(r_R^2 + r_I^2) + 2p(r_R\frac{\partial r_R}{\partial \theta} + r_I\frac{\partial r_I}{\partial \theta}) \tag{A.2.33}$$

$$\frac{\partial r_R}{\partial \theta} = \frac{\frac{\partial w_R}{\partial \theta}v - w_R\frac{\partial v}{\partial \theta}}{v^2} \tag{A.2.34}$$

$$\frac{\partial r_I}{\partial \theta} = \frac{\frac{\partial w_I}{\partial \theta}v - w_I\frac{\partial v}{\partial \theta}}{v^2} \tag{A.2.35}$$

$$\frac{\partial w_R}{\partial \theta} = 2c\frac{\partial c}{\partial \theta}(n_a^2 - n_s^2) + 2s\frac{\partial s}{\partial \theta}((\frac{n_a n_s}{n})^2 - n^2) \tag{A.2.36}$$

$$\frac{\partial w_I}{\partial \theta} = 2s\frac{\partial c}{\partial \theta}n_a(\frac{n_s^2}{n} - n) + 2c\frac{\partial s}{\partial \theta}n_a(\frac{n_s^2}{n} - n) \tag{A.2.37}$$

$$\frac{\partial v}{\partial \theta} = 2c\frac{\partial c}{\partial \theta}(n_a + n_s)^2 + 2s\frac{\partial s}{\partial \theta}((\frac{n_a n_s}{n}) + n)^2 \tag{A.2.38}$$

$$\frac{\partial c}{\partial \theta} = -\frac{2\pi}{\lambda}s\,n\frac{\partial d}{\partial \theta} \tag{A.2.39}$$

$$\frac{\partial s}{\partial \theta} = \frac{2\pi}{\lambda}c\,n\frac{\partial d}{\partial \theta} \tag{A.2.40}$$

This can easily be implemented by computing the sub-derivatives from the bottom up where θ is the parameter of interest.