INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

PHYSICS- AND ALGORITHM-BASED TECHNIQUES TO IMPROVE ALIGNMENT AND OVERLAY FOR LITHOGRAPHY

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING AND
THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

XUN CHEN
DECEMBER 1998

UMI Number: 9924411

UMI

300 North Zeeb Road
Ann Arbor, MI 48103
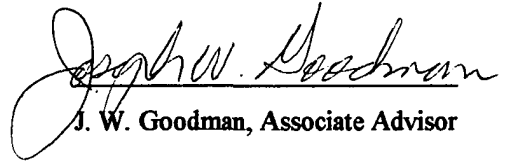
I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

_____

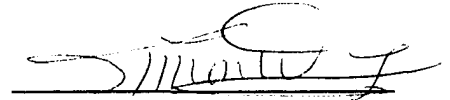R. F. W. Pease, Principal Advisor

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
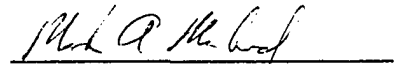
_____

J. W. Goodman, Associate Advisor

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

_____

Simon Wong, Professor

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

_____

M. A. McCord, Consulting Professor

Approved for the University Committee on Graduate Studies:

_____

iii

TO MY WIFE, PING

FOR HER LOVE AND WISDOM

iv

Abstract

The need to overlay patterns in semiconductor circuit manufacturing to better than 20nm is now well recognized as a difficult and important challenge. Two broad strategies for meeting this challenge are described in this dissertation, one is physics-based and the other algorithm-based.

A physics-based approach is used to minimize the alignment errors resulting from asymmetric resist coating, the most ubiquitous cause of alignment errors. The technique employs transmagnetically polarized light at the Brewster angle of the resist. Experimental results show that resist-induced alignment errors as large as 25nm can be mitigated to less than 6nm by using this method.

We provide a general framework for designing advanced algorithms that are capable of correct alignment on marks that are damaged or distorted by various processing. One example, a subspace decomposition based algorithm, is applied to many types of mark damage and distortion. Experimental results show that for both artificially damaged marks and CMP-corrupted marks from a manufacturing process for which conventional algorithms failed (overlay errors of 200nm), the new method achieved a mean plus three sigma alignment error less than 50nm.

It is demonstrated that the combination of the physics- and algorithm-based techniques will be able to satisfy the overlay requirement of, and possibly beyond, the 0.13um generation of lithography.

v

Statement of Original Contribution

My original contribution to this project includes:

1. The physics-based analysis of the resist-induced alignment error presented in Chapter 3

2. The theoretical model and simulation results in Chapter 4

3. The general framework of the learning approach and the preliminary algorithmic implementation presented in Chapter 6

4. The CMP experiment described in Chapter 8

The following are joint contribution by Amir G. Ghazanfarian and myself:

1. The Brewster angle illumination experiments described in Chapter 5

2. The experiments with programmed asymmetry in Chapter 7

As noted in this thesis, the final implementation of the subspace decomposition algorithm that was applied to the experiments in Chapters 7 and 8 was largely due to Amir G. Ghazanfarian.

vi

Acknowledgments

I would like to express my sincere gratitude to my principal advisor, Professor Fabian Pease. Dr. Pease introduced me to the challenging problem of alignment, offered me with invaluable guidance throughout my graduate education at Stanford, and provided me with numerous connections with the industry. Even more importantly, Dr. Pease has been an invaluable mentor for many other aspects of my life. In the later years when Dr. Pease took a leave of absence to join DARPA, I had the opportunity to work with Professor Mark McCord who acted as my advisor. Dr. McCord's encouragement was instrumental in motivating us towards a practical solution of the alignment problem. Without their help and supervision, none of my accomplishments would have been possible.

Throughout this research, I worked very closely with Amir Ghazanfarian, a fellow student in the Pease Group. Amir's friendship, understanding and optimism have not only made this project possible, but also instilled a sense of great enjoyment in this experience.

Many engineers and managers in the semiconductor industry has contributed to this research. In particular, I would like to thank Mr. Dave Markle of Ultratech Stepper, who offered me a summer internship in 1995. Mr. Kent Green of Motorola has provided me the opportunity to carry out the pivotal experiment in a manufacturing environment. In addition, I have had many invigorating discussions and fruitful collaborations with Chiaki Sato of Cannon; Alex Chen and Mike Takac of IBM; Alexander Starikov of Ultratech Stepper; Lloyd Litt, Mike Acosta and Dana Dupere of Motorola; and Mircea Dusa of National Semiconductor.

Special thanks to Dr. Jun Ye of KLA-Tencor for the many stimulating and insightful discussions we had. Also, Sandra Eisensee's administrative help is highly appreciated.

I would like to thank my parents and parents-in-law. I am deeply indebted to my wife, Ping Yang, to whom this dissertation is dedicated. Her patience, understanding and encouragement are tremendous in helping me to finish this research.

Last but not least, I would like to acknowledge Professor Max Lagally of University of Wisconsin at Madison, under whose guidance I earned my MS degree in solid state physics. His deep understanding of science, technology and life in general has never failed to enlighten me. His influence has profoundly altered my personal growth and development.

Contents

PART III: ALGORITHM-BASED TECHNIQUES

List of Tables

List of Figures

xiv

PART I: CHALLENGES IN PATTERN PLACEMENT

Chapter 1. Introduction

In this chapter, we give a brief introduction to the problem of pattern placement in lithography, and discuss the associated challenges that are facing current and future lithographic technologies.

1.1 Planar processing and lithography

The spectacular advances of the development and application of integrated circuit technology are unparalleled in our history. As gauged by many critical measures, the technology of making integrated circuits has been progressing following an exponential trend. Widely known as the Moore's Law [1], the number of transistors that can be packed on a certain area of silicon substrate approximately doubles every eighteen months. Due to the now ubiquitous applications of integrated circuits, this exponential law does not simply characterize the evolution of the semiconductor industry, it is also dictating the rate of progress for many other aspects of information technology. For instance, it has enabled the proliferation of personal computing. It has also catalyzed the revolution of telecommunication. Many of the sweeping technology and life-style changes we are experiencing today can be ultimately attributed to the availability of computing devices that have been made ever so inexpensive, small and powerful by the Moore's Law.

It seems, then, Moore's Law is an important enough trend to justify a rigorous study of the underlying driving forces that made it possible. From this understanding, we can then derive the possibility and probability of it being sustainable over the years to come. This is, of course, an area of active research that encompasses many aspects of semiconductor technology. In this thesis, we would like to examine closely one important aspect of semiconductor technology, namely, the accuracy and precision of pattern placement during the lithography processes.

I
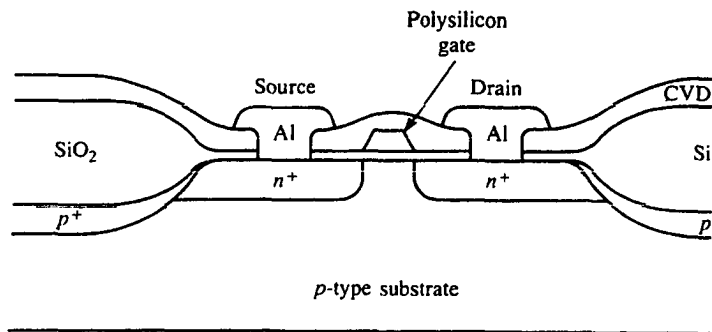
Figure 1-1. Illustration of an NMOS transistor in cross-section view.

We argue that a critical attribute of today's integrated circuit technology is planar processing [2]. A schematic cross-section of an n-type MOS transistor is shown in Figure 1-1. It can be seen that a simple transistor is actually a three dimensional structure of different materials built on top of the wafer surface. The correct functioning of the transistor depends on the precise control of the topographical dimensions and the electrical properties of these materials.

This complicated structure is built in a layer-by-layer fashion, the so-called planar processing, by repeatedly applying a number of basic processing steps that may include oxidation, etching, diffusion, sputtering, chemical vapor deposition, ion implantation, epitaxy and lithography [2]. Among these basic, or elemental processing steps, all but one are blanket processing, i.e., they apply the same action at every location on the surface of the wafer. Any selectivity of these actions must be introduced through the pivotal technology of lithography, which defines the patterns that limit where the subsequent processing steps take place on the substrate [3].

2

Light Source

Condenser Lens

Mask

Chrome

Projection Optics

Resist

Wafer

Stage

After development
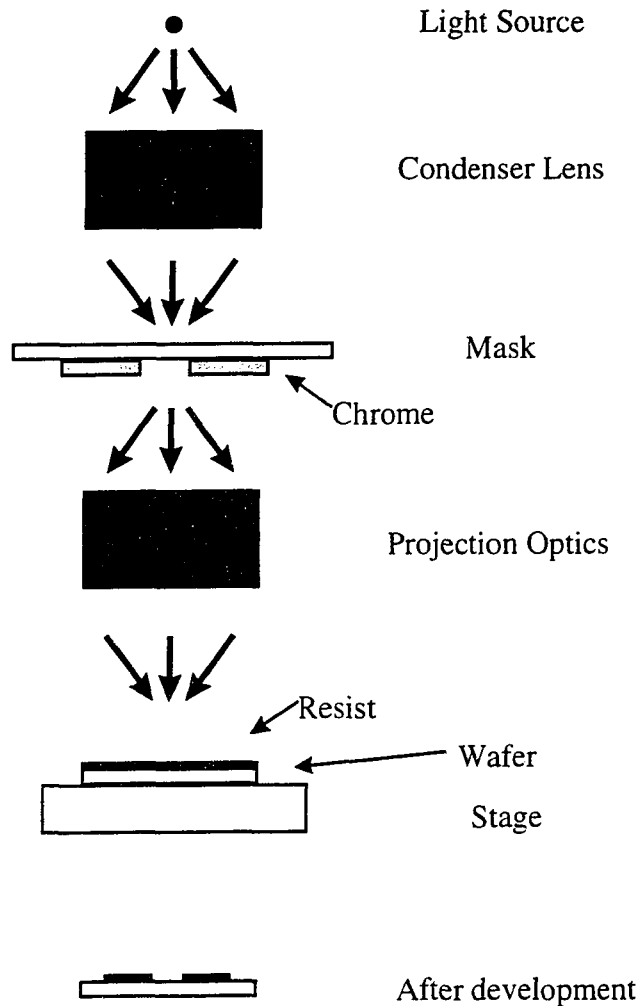
Figure 1-2. Schematics of projection photolithography. The light source and condenser lens form the illumination system. The circuit pattern is first defined in the chrome on the photomask. The projection optics forms an image of the mask pattern in the photoresist coated on the wafer. After development, resist is selectively removed. The remaining resist forms a barrier for subsequent processing.

3

Lithography itself is a decidedly two-dimensional technology. As seen in Figure 1-2, a two-dimensional image of the mask patterns is projected onto the wafer surface. This image is then transferred to the substrate via photoresist. To build a three-dimensional structure, such as the transistor in Figure 1-1, many of these lithography steps have to be used in order to stack up the two-dimensional patterns to build the third dimension.

The planar processing technique is essentially a divide-and-conquer approach that has made the task of building complex three-dimensional structures more amenable to process integration. The multitude of lithographic steps necessitated by planar processing, however, requires the patterns of a particular layer be placed at the desired positions relative to what have already been made on the wafer in previous layers. This means although planar processing has great advantages, it necessitates one important parameter that the lithography technologies must learn to control: pattern placement.

Most of the attention of lithography research has been on resolution, or how small a feature can be reliably made on the wafer [3]. What is as important, but has so far received inadequate research attention, is the issue of pattern placement, or how well the lithographic technology can control the positioning of the patterns.

Adequate pattern placement control is a necessary condition for practical semiconductor technologies. For Moore's Law to be sustained in future years, methods must be found to assure that this necessary condition is satisfied.

1.2 Pattern placement in lithography

Pattern placement is achieved by the procedure of alignment; and measured and monitored by overlay metrology. Without losing any generality, we limit our discussion of pattern placement to the case of optical lithography [3]. See Figure 1-2.

The mask is made by some other lithographic processes, typically electron beam lithography. To facilitate discussion, we will assume that there is only one die per field, i.e., the entire mask contains the circuit patterns of a single chip.

The wafer is held on a stage that is capable of multi-axis motion and adjustment. The position of the stage is monitored by a laser interferometer system that measures position to sub-nanometer precision. Essentially, the laser interferometer readings establish a coordinate system against which positions of features on the wafer can be measured.

1.2.1 The case of rigid wafer and geometric imaging

Let's look at the exposure of one particular die on the wafer. To simplify the discussion, we will assume for now that patterns made on the wafer are perfect replicas of the mask patterns. In other words, we are assuming that the optical image in the lithography step is a perfect geometric image; and secondly, that the pattern transfer process through the resist does not introduce any distortion or smearing effects.

In addition, we assume that the wafer substrate is mechanically rigid, i.e., it is not subject to elastic deformations or thermal expansions. The effects of relaxing these assumptions will be discussed later. For the moment, we also assume that in-plane rotation of the wafer can be corrected using the global alignment procedure outlined in section 1.2.4.

With the above assumptions, the position of every feature in a particular die can be readily deduced if we know the position of any other feature inside the die. The fiduciary can be the center of the die, a corner of the die, or any other feature within the die boundary. It is a common practice to design some structures inside the "scribe lane" between die and use them as the fiduciaries, which are commonly known as alignment marks.

The task of the lithography tool is to first find the fiduciary, or alignment mark, and accurately determine its position. Typically, an optical sensor [4] is used to detect and locate the position of the alignment mark. Suppose that this is found to be (x_m, y_m) in our coordinate system.

According to the rigid-body and geometric imaging assumptions, the position of the alignment mark relative to the die center is known to be exactly as designed. Suppose

5
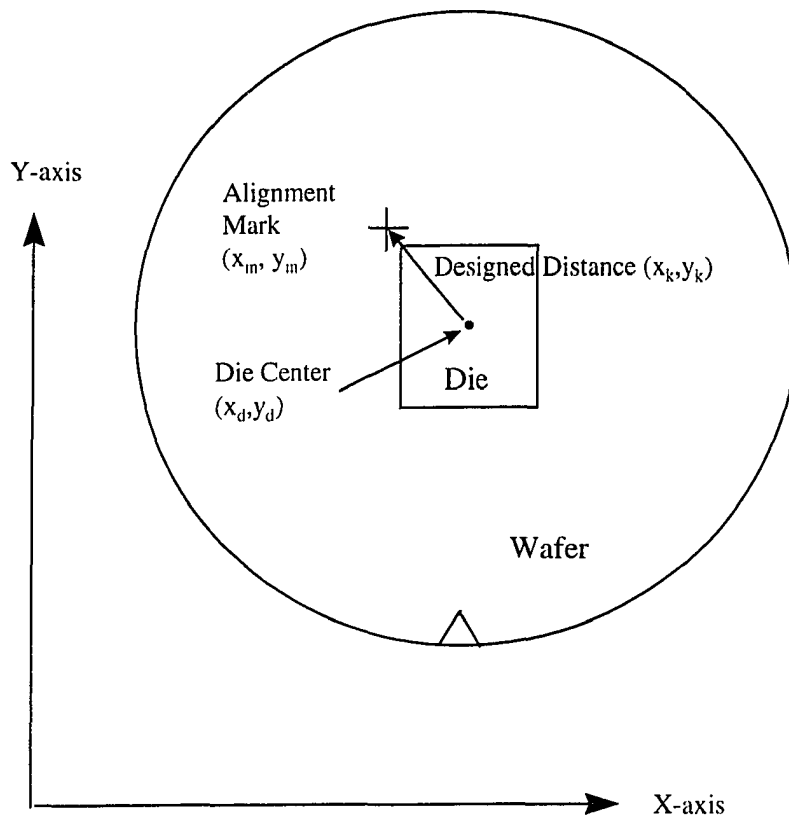
Figure 1-3. The basic alignment procedure. First the lithography tool finds the alignment mark, and determines its position (x_m, y_m). The alignment mark is designed to be located at (x_k, y_k) with respect to the die center. The position of the die center can then be determined. Next, the die center is moved to the optical axis and the die is exposed.

this is (x_k, y_k). We can then find the position of the die center to be $(x_d, y_d) = (x_m, y_m) - (x_k, y_k)$. Next, the lithography tool instructs the wafer stage to move so that the die center is placed at the axis of the exposure optics, or "aligning" the optical axis with the die center. The die is then exposed with the wafer held at this position. The geometry of this procedure is shown in Figure 1-3.

6

Let (x_c, y_c) denote the position where a certain feature in the current layer is imaged. Assume it is supposed to lie exactly on top of a feature in the previous layer which was placed at position (x_p, y_p). The overlay error is then defined, for this particular feature, as $\mathbf{O} = (O_x, O_y) = (x_c, y_c) - (x_p, y_p)$. From this definition, it is clear that overlay error is defined everywhere inside the die. Overlay error, $\mathbf{O}(x, y)$, is therefore a two-dimensional vector function that is defined everywhere inside the die.

The rigid-body and the geometric imaging assumptions greatly simplify the form of this function. In fact, under these assumptions, $\mathbf{O}(x, y)$ is constant across the die. There can be only two reasons why $\mathbf{O}(x, y)$ is not identically zero. First, there could be an error in determining the position of the alignment mark. This is called alignment error. Second, there could be an error in positioning the die center at the axis of the exposure optics, and this is due to the stage positioning error.

1.2.2 Wafer distortion

The situation becomes more complicated if we relax the rigid wafer assumption. Suppose that the wafer deforms when subject to mechanical and thermal stresses. In this case, even if the previous mask layer was perfectly imaged on the wafer, the processing between the previous and current lithographic steps could have deformed the wafer, i.e., the relative positions of circuit patterns with respect to that of the fiduciary are changed. In other words, (x_k, y_k) as defined in the last section is no longer identical to its designed value. The difference can be termed wafer distortion.

Furthermore, the amount of wafer distortion can vary from one feature in the die to another. This makes the overlay $O(x, y)$ a true two-dimensional function: the sum of a constant term across the die, given by the alignment error and the stage positioning error; and a varying term determined by the amount of wafer distortion.

1.2.3 Field distortion

The problem of overlay control is further complicated by the issue of field distortion. The projected image in the resist is never a geometric replica of the mask

7

pattern. Aberrations and distortions in the optics can cause the actual image to differ from the mask pattern. This is known as optical <u>field distortion</u>, a well understood phenomenon in classical optics [5].

Field distortion has the same effects as wafer distortion. It varies across the optical field, and it causes (x_k, y_k) to deviate from its known, designed value. Now, one has to take into account the additional effects of field distortions in order to fully characterize the overlay function $O(x, y)$.

1.2.4 Local and Global Alignment

We have described placement control in lithography as a three-step procedure:

1. Determine the position of the alignment mark associated with the die
2. Align the die center to the optical axis
3. Expose current layer.

This is known as the local alignment procedure, because it achieves placement control by detecting the alignment marks that are associated with the die right before it is exposed. Although used extensively in research laboratories for difficult-to-align situations, it is not usually practiced in high-volume manufacturing environment. The reason is that detecting and locating the alignment marks for each die is too costly – it slows down the lithography process and reduces throughput, and high throughput is needed for obvious economic reasons.

There can be as many as hundreds of die on a wafer, and the locations of these die on the wafer surface are typically laid out in a lattice, as shown in Figure 1-4. This lattice structure is known as the die layout. Because of the regularity of the die layout, it would suffice to find the position of the whole lattice, rather than each individual die. This leads us to global alignment.

In global alignment, only a few alignment marks across the wafer are accurately located. The exact number of marks being used by global alignment can vary from two to ten for one wafer. After the positions of these marks are determined, the shape and the location of the whole lattice is calculated by fitting the mark positions to the regular lattice structure. Often, it is assumed that the actual lattice can only deviate from the

8

designed lattice by an in-plane rotation of each axis, a magnification along each axis, and a translation along each axis: a total of six parameters. Figure 1-5 shows the visual effects of these parameters. It is obvious that such a lattice is only a linear distortion of the designed rectangular lattice.
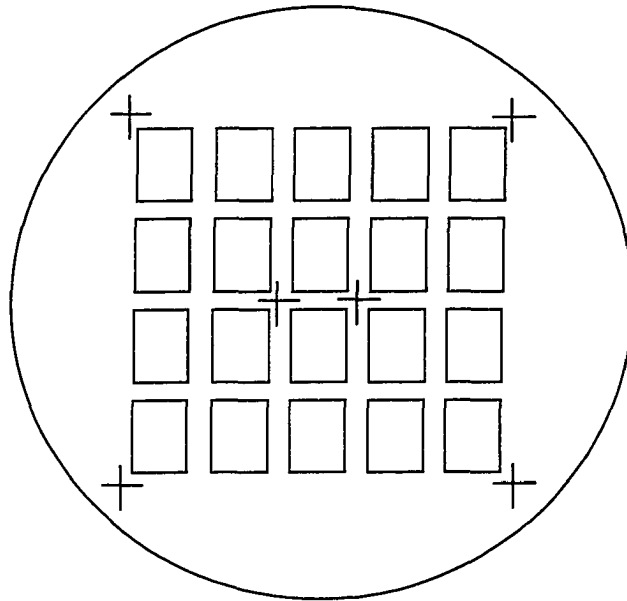
Figure 1-4. Global alignment. Die on a wafer are typically laid out in a regular pattern, similar to the rectangular lattice shown here. In the procedure of global alignmment, positions of several alignment marks are first determined. We are showing six marks in this figure. The position of the lattice can be determined from the positions of the alignment marks.

In this framework, global alignment finds the six parameters that characterize the linear lattice, from which the position of each die is determined. Similarly to local alignment, in the next step of lithography, the tool aligns the die centers to optical axis, and then expose them. A detailed description of global alignment can be found in [6].

9
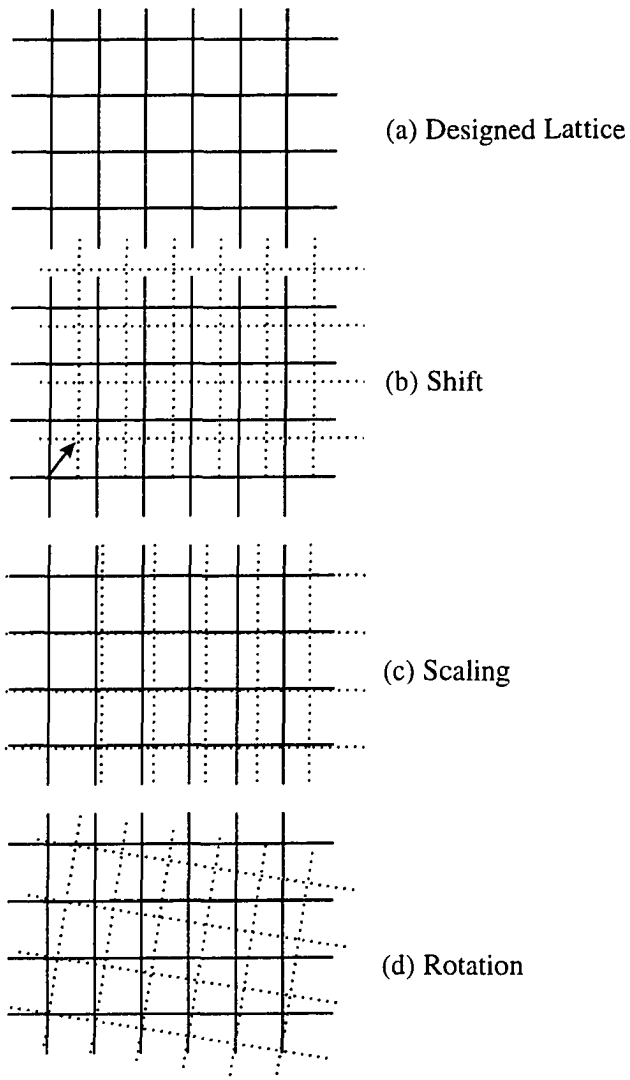
Figure 1-5. Parameters to be determined by global alignment.
Global alignment attempts to determine the position of a die layout
lattice. It assume that there only three kinds of deviations from the
designed lattice, as illustrated in this figure. (a) The designed
lattice. (b) A shift in both the x- and y-directions. (c) A scaling in
the x-direction. (d) An in-plane rotation of the lattice.

Clearly, using global alignment can improve the throughput a great deal if the
process of locating the alignment marks is time-consuming compared to exposure, and
this does seem to be the case. But the throughput improvement does not come without

10

cost, because the global alignment procedure jeopardizes overlay performance. The difficulty lies in the fact that, due to wafer distortions, the die layout is not necessarily a linear lattice. In fact, wafer distortion behaves in totally undetermined ways, and it can undoubtedly make the die layout a nonlinear lattice. This leads us to the conclusion that if global alignment is used, there is yet another contributing factor to overlay error – the nonlinearity of die layout lattices brought about by wafer distortion.

1.2.5 Overlay Metrology

Although overlay is a two-dimensional vector function, we cannot measure it everywhere in the die, because much of the die area is occupied by device patterns. Typically, there are overlay marks designed in the scribe lanes between die to allow the measurement and monitoring of overlay errors. An overlay mark consists of a feature in the current layer interlocking another feature in the previous layer. If perfect overlay is achieved, the center of the current layer pattern and that of the previous layer pattern should coincide with each other. If there is any difference between these two, it
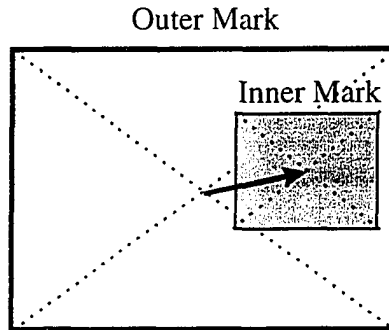
Figure 1-6. Overlay Metrology. Overlay metrology measures overlay error using metrology marks. In the example shown here, the outer mark is an open rectangle made in the precious layer. The solid box is the inner mark and it is in the current layer. The difference between their center positions measures the overlay error between the two layers.

11

represents the overlay error between the two layers. See Figure 1-6.

There is a fundamental difficulty of determining the real overlay performance from metrology results. Overlay metrology can only be performed on several discrete locations in the die. The values of the overlay function $O(x,y)$ at other positions in the die have to be interpolated or extrapolated from the limited number of metrology results. To the extent that both wafer and field distortions exist, such an interpolation or extrapolation procedure can be very unreliable.

Furthermore, for the same reasons that cause alignment errors, overlay metrology also suffers from errors. Overlay metrology has to determine the positions of both the previous and current layer patterns, and errors can happen in either of these measurements.

The performance of overlay metrology is a critical component in pattern placement [7]. It is of vital importance that overlay metrology is obtained using reliable methods. In later chapters, we will see direct impacts of overlay metrology on the design of alignment methodology, because metrology results are used as the basis of analyzing and learning the characteristics of an alignment system.

1.3 Challenges in pattern placement

1.3.1 Overlay requirements of current and future technologies

The required accuracy of pattern placement depends on the size of the features being made in the lithography process: the finer the structure we are making, the

Year	1995	1998	2001	2004	2007	2010
CD (μm)	0.35	0.25	0.18	0.13	0.10	0.07
Gate Overlay (nm)	100	75	50	40	30	20
Metal Overlay (nm)	300	200	150	100	70	50
Current Performance on Metal (nm)		199.5	?	?	?	?

12

Table 1.1. The overlay budgets for the gate and early metal layers for various generations of lithographic technologies [8]. The overlay budget is defined as the mean plus three sigma value of the overlay errors. The quoted current stepper performance is based on an experimental study, which will be presented in Chapter 9.

more accurately we need to place the patterns. Therefore, even for a single chip, the required placement accuracy varies from layer to layer. The critical layer, or the layer that makes the transistor gates, usually has the smallest features and, consequently, the most stringent requirements on placement accuracy. For back-end processing, such as interconnections in the metal layers, the placement requirement is usually much less strict.

Table 1-1 outlines the required overlay performance for different generations of lithographic technologies, both for critical layers and for some representative metal layers [8]. In this table, the quantity called overlay budget, or overlay in short, is a statistical measure of pattern placement. By definition, $overlay = |m| + 3\sigma$, where m is the mean and σ is the standard deviation of a statistical sample of overlay errors taken from a particular process. A rule of thumb that has been used by the industry is that overlay should be less than one third of the feature size in the critical layer in order for the lithography process to be usable.

What is striking about pattern placement control is that, although the overlay requirement has become ten times more stringent since the 1970s, the overlay control capability has improved much less quickly, at best by a factor of two. This imbalance of technology improvement has resulted in a serious danger that future generations of lithography may not deliver according to the schedule in Table 1-1. For instance, the current overlay performance quoted in Table 1-1 is only marginally satisfactory for the 0.25μm generation. Apparently, today's alignment technology cannot satisfy the strict overlay requirements of future technologies.

Let's break down the overlay error into its components and review their historical progress, current status and future outlooks.

13

1.3.2 Progress in each error component of overlay

The first component is alignment error. As we will see in much more detail later, alignment error depends on two factors: the optical systems used to detect the alignment marks; and the processing steps that have been applied to the wafer which can damage or distort the topographic structure of the alignment mark. [4]

The processing that an alignment mark has to go through has undoubtedly become much more complex over the years. In particular, the recent introduction of Chemical Mechanical Polishing [9] and the use of copper and damascene processes [10] pose serious challenges to alignment because the alignment mark structure can be seriously damaged during the polishing process. In contrast, the alignment systems on most commercial lithography tools have remained the same designs as twenty years ago. Very little fundamental progress, if any, has been made in how the alignment system determines the position of an alignment mark. The increase of processing complexity and the idling of alignment technologies lead us to believe that alignment errors have become worse in recent years.

Stage positioning error is one area where some progress has been made recently. The three sigma repeatability of stage positioning has been gradually decreased to its current levels around 10nm [11]. Much of this progress was made possible by the improvements in the mechanical and electrical designs of the stage. Further improvement can be made if stage distortion can be calibrated out. This effect can be as large as 20nm. But methods exist to exactly calibrate and eliminate stage distortion effects, at least for the low-frequency components. [12].

As described in Section 1.2, the high-frequency components of wafer distortion cause intra-field overlay errors, and the low-frequency parts contribute to alignment errors if global alignment is used. Wafer distortion as large as 40nm has been measured , but typically wafer distortion values are around a three sigma of 20nm [13]. It is very likely that wafer distortion will increase as larger wafers are used in manufacturing. Some efforts have been spent on how the wafers should be held on the stage, which was found to be an important cause of wafer distortion.

14

Field distortion is characterized by a three sigma of around 40nm [14]. Since this is due to optical imperfections, we do not believe field distortion can be reduced to a level substantially below this. A useful approach that has been devised to combat the field distortion problem is called "stepper matching" [15], which means that subsequent lithographic layers are made either on the same stepper, or on steppers that have similar distortion patterns. Stepper matching is very effective: field distortion effects less than 15nm were reported after stepper matching was performed. [15]

1.3.3 Relative importance of various error components

Let's put aside the issue of alignment errors, and assume that all other error components are statistically independent of each other. This gives us an overlay of approximately 30nm in the absence of alignment errors. This number is close to the overlay budget of $0.10\mu m$ generation. Obviously, to achieve the necessary pattern placement control for technologies beyond the $0.10\mu m$ generation, improvements must be made not only in alignment, but also in all the other contributing factors of overlay errors.

Local and global alignment errors, however, remain the wildcards of overlay control. For difficult-to-align situations such as chemical-mechanically polished wafers, overlay error in excess of $0.1\mu m$ is not uncommon [16]. Assuming we have a 30nm contribution from all other error components, $0.1\mu m$ overlay error means alignment errors alone can be as large as 95nm. Hence, an overwhelmingly large portion of the overlay budget is consumed by alignment errors.

This analysis motivates us to seek advanced methods of improving alignment performance.

15

Chapter 2. Generic approaches to improve alignment

In this chapter, we describe the basic components of an alignment system. Two broad categories of approaches to improve alignment will be introduced.

2.1 Components of an alignment system

2.1.1 The alignment mark and its optical description

As described in Chapter 1, any topographic structure can be used as an alignment mark. It can be a trench, a mesa, a combination or even an array of both. Typically, alignment mark structures are translationally invariant along a certain direction. For instance, a long, straight trench etched into the substrate can be described as being translationally invariant in the direction along the long axis of the trench. This is shown in Figure 2-1. This type of mark is in fact two-dimensional and is used to do the alignment in the direction perpendicular to its long axis. To facilitate discussion, let's define the coordinate system so that the x axis is perpendicular to the long axis of the mark; y axis is the long axis of the structure; and z axis is normal to the wafer surface. See Figure 2-1.

To give a concrete example, a cross-section, i.e., the x-z plane, microscope picture of an alignment mark is shown in Figure 2-2. The original mark is an etched trench into the substrate. Tungsten is then deposited on top and excess materials are removed in a polishing process. However, some Tungsten is left inside the trench. Aluminum is then deposited over the structure. The composite structure that comprises the trench, the leftover Tungsten and the Aluminum overcoat is our alignment mark in this particular case. In fact, during alignment, there is also a resist coating on top of what is shown in the figure, further complicating the alignment mark structure. As illustrated by this example, an alignment mark is almost never a clean and nicely defined trench or mesa. Because of the processing steps that the wafer has to go through, a resist coating, a

16

deposited metal layer, a grown polysilicon layer, etc., can all become parts of the structure that we are aligning to.

(a)

(b)
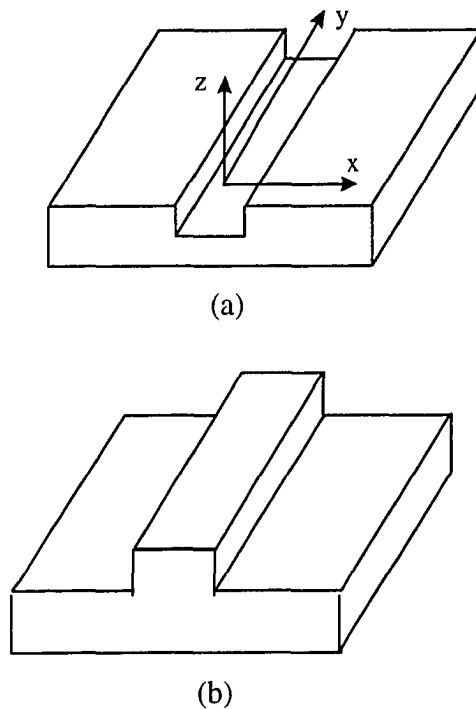
Figure 2-1. Garden variety alignment marks. (a) A
trench etched in the substrate. (b) A mesa. Both
marks can be considered two dimensional structures.
The Cartesian coordinate system used in this thesis is
also shown.

As described in Chapter 1, the position of an alignment mark is known relative to the center of the die and this information is used in the procedure of alignment. Since the alignment mark itself has a structure, when we speak of the "position" of an alignment mark, we must have referred to the position of a certain feature in the mark structure, for instance, the symmetry axis of the trench in dielectric as shown in Figure 2-2. Hereafter, we will refer to this part of an alignment mark as the "reference position", and we will always define the x axis so that the origin $x = 0$ is at the reference position of an alignment mark.

17
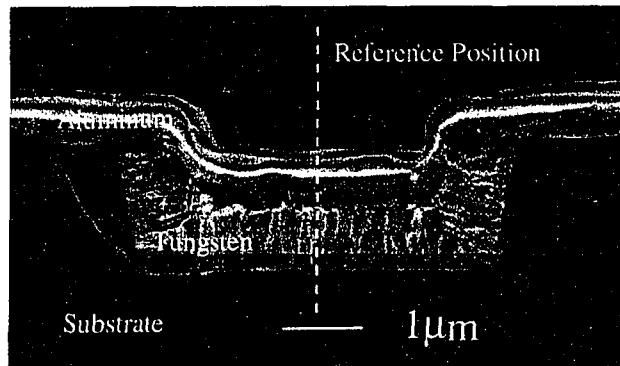
Figure 2-2. Cross-sectional view (SEM) of an
alignment mark covered with subsequent coatings.

The sensors that we use to detect the alignment marks are typically optics-based [4]. The response of an alignment mark in any kind of optical detection system can be predicted if its scattering matrix is known. The scattering matrix $S(p,q)$ is defined as the following with the reference position placed at $x = 0$: if an incident plane wave comes in at an angle p, the light scattered by the alignment mark has a plane wave (or Fourier) component that is outgoing at angle q, with a magnitude and phase represented by $S(p,q)$. The scattering matrix can be calculated using vector theory of electromagnetism based on the mark's three-dimensional function of the index of refraction, $R(x,y,z)$. For the translationally invariant marks, this function can be reduced to two dimensions: $R(x,y,z) = R(x,z)$.

For reasons soon to become clear, symmetry properties of the scattering matrix are very important. Suppose the distribution of refractive index R(x,y,z) is symmetric about the reference position, i.e., $R(x,y,z) = R(-x,y,z)$. Because of the space symmetry of electromagnetism [17], this leads to a symmetric scattering matrix, i.e., $S(p,q) = S(-p,-q)$. We say the alignment mark is symmetric if the above equality holds.

18

We can draw two conclusions from the above description. First, all parts of the alignment mark contribute to the index of refraction function, $R(x,y,z)$, and hence ultimately to the scattering matrix. Therefore all coatings, polishing and other processing that have been applied to the alignment mark will likely change the mark's appearance in an optical detection system. Second, to know the exact extent and nature of these effects, we have to obtain detailed knowledge of the $R(x,y,z)$ function for every alignment mark, and that is impossible to accomplish.

We now turn to the optical systems that are used as alignment tools.

2.1.2 Imaging and scanning optics

A classical optical microscope, or a scanning optical microscope as shown in Figure 2-3 can be used as an alignment system. Working principles of these instruments are well-documented [18]. We provide a few simple examples to illustrate how the optics interacts with the alignment mark. We assume that the reference position of the alignment mark is on the optical axis of the microscope.

Let $F(p)$ represent the phase and amplitude of the plane wave component of the illumination field that comes in at an angle p, where p spans from $-p_A$ to p_A, representing the numerical aperture of illumination, as shown in Figure 2-3. These plane waves are then scattered by the alignment mark, resulting in a scattered optical field given by

$$A(q) \propto \int_{-p_A}^{p_A} F(p)S(p,q)dp, \qquad \text{Eq. 2-1}$$

where $A(q)$ gives the amplitude and phase of the plane wave outgoing at an angle q. Suppose now that all light within the aperture between $-q_A$ and q_A are collected, the image is then given by [19]

$$I(x) = \left| \int_{-q_A}^{q_A} A(q)e^{i2\pi\sin(q)x/\lambda}dq \right|^2, \qquad \text{Eq. 2-2}$$

where I is the intensity image as a function of position x.

19

As is clear from the above example of a simple optical microscope, the alignment signal obtained from a scanning or imaging microscope is the outcome of the alignment mark's scattering matrix interacting with the specifics of the optics. Although the above equations are quite simple, this interaction can be nonlinear and very complicated if effects such as partial coherence are included [20]. The complexity prevents a clear and
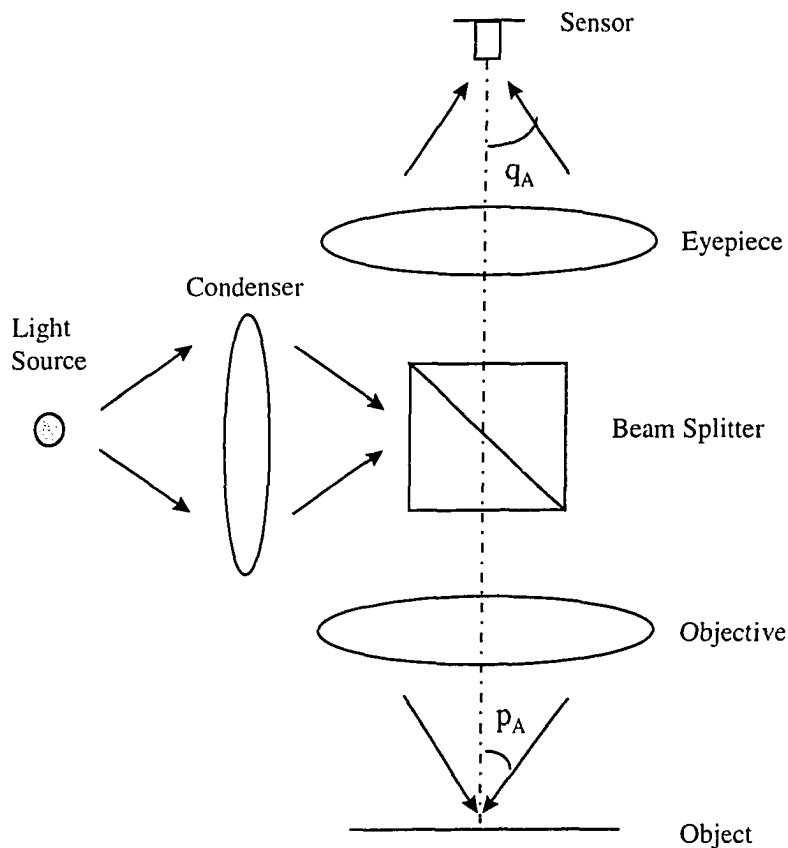
Figure 2-3. Schematic of an optical microscope used as an alignment tool. The critical parameters include the type of illumination, illumination aperture p_a and the aperture of the collection optics q_a.

intuitive, let alone quantitative, understanding of how various processing of the alignment mark structure translates into changes of the alignment signal.

The symmetry property of the image $I(x)$ can be easily deduced. For the simple system described by Equations 2-1 and 2-2, the microscope is symmetric about its optical axis because we have specified symmetric illumination and collection apertures and we have ignored all aberration effects. In this case, it is easy to show that if $S(p,q) = S(-p,-q)$, then $I(x) = I(-x)$, i.e., the image will be symmetric about the reference position. The fact that a symmetric alignment mark imaged by a symmetric optics gives symmetric alignment signal is expected to hold in general, as long as we include in the definition of symmetric optics both the apertures and the aberration effects.

2.1.3 Interferometric Optics

Another optical system that has been successfully employed as a commercial alignment tools is based on coherent interferometry [21, 22]. A schematic of such a system is shown in Figure 2-4.

A periodic line-and-space pattern is made on the mask, or it could be on a permanent fiduciary somewhere inside the lithography tool. A beam of laser light passes through the periodic pattern and is diffracted into multiple orders. A spatial filter is used to block all but the +1 and -1 orders. After passing through the imaging optics, we can write the optical field of the +1 and -1 order diffraction as, respectively,

$$g_{+1} = e^{i(k_x x + k_z z)} \text{ and } g_{-1} = e^{i(-k_x x + k_z z)}. \qquad \text{Eq. 2-3}$$

where the coordinate system is defined as in the last section; in particular, we note that $x = 0$ corresponds to the reference position of the alignment marks. k_x and k_z are the wave vector components in the x and z directions respectively, where

$$k_x^2 + k_y^2 = (\frac{2\pi}{\lambda})^2 \text{ and } \tan(p) = \frac{k_x}{k_z}, \qquad \text{Eq. 2-4}$$

λ is the wavelength of light, p is the incident angle as shown in Figure 2-4.

21
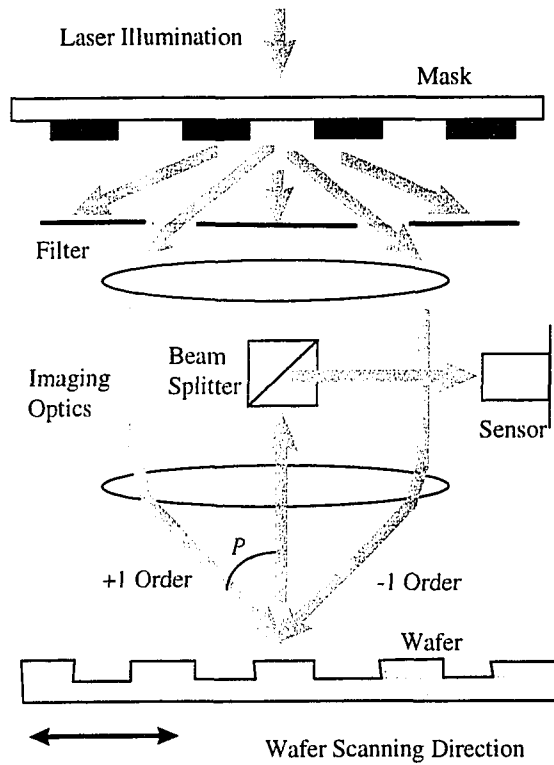
Figure 2-4. Interferometric alignement system. Two plane wave incoming at angle P illuminate the mark, which is a periodic pattern. The illumination is diffracted by the periodic mark. The alignment signal is formed by measuring the intensity of the outgoing plane wave along the optical axis, while the wafer is scanned.

The alignment mark is also a periodic line-and-space structure made on the wafer, with a spatial frequency f given by

$$\sin(p) = \lambda f .$$

Eq. 2-5

Eq. 2-5 specifies that the periodic structure of the wafer alignment mark diffracts both the +1 and -1 orders into the $-z$ direction. See Figure 2-4. The light that's outgoing at an angle $q = 0$ has two components in it, the diffraction of the original +1 order and that of the -1 order. Utilizing the scattering matrix $S(p,q)$, we can write the outgoing light at zero angle as

22

$$g_0 = [S(p,0) + S(-p,0)]e^{-ikz} .$$ Eq. 2-6

Now let's suppose that we move the wafer by a small shift s. Then it is easy to see that Eq. 2-6 must be modified as,

$$g_0(s) = [S(p,0)e^{ik_x s} + S(-p,0)e^{-ik_x s}]e^{-ikz} .$$ Eq. 2-7

If the alignment mark is symmetric, we have $S(p,0) = S(-p,0)$. Then Equation 2-7 can be written as

$$g_0(s) = 2S(p,0)\cos(k_x s)e^{-ikz} .$$ Eq. 2-8

The intensity of this signal is

$$I(s) = 4|S(p,0)|^2 \cos^2[\frac{2\pi \sin(p)}{\lambda}s] .$$ Eq. 2-9

With Equation 2-5 plugged in, this can be simplified as

$$I(s) = 4|S(p,0)|^2 \cos^2(2\pi f s) .$$ Eq. 2-10

Equation 2-10 shows that when the wafer is scanned in the x direction, the outgoing signal at angle $q = 0$ has an intensity that depends on the position of the wafer in an sinusoidal manner. This sinusoidal function can be used as the alignment signal. The phase of this sinusoidal signal is proportional to the position of the wafer, with zero-phase corresponding to the reference position of the mark.

This system has some advantages over the imaging type. First, the alignment is based on the collective location of a periodic feature on the wafer; therefore it is more robust against localized damage to the alignment mark. Second, because the optics essentially uses only two plane waves of light, which we labeled +1 and -1 diffraction orders, they can be differentially modulated in the time domain. Such a technique is called heterodyne [22]. We will not go into the details of heterodyne other than stating that it is a very effective method to improve signal-to-noise ratio if used in combination with a phase-locked amplification system.

23

2.1.4 Alignment algorithm

The signal from an optical alignment system is given in Equation 2-2 for an imaging system and in Equation 2-10 for an interferometric system. Although the signal may contain thousands of data points, we need only one number from it – the reference position of the alignment mark.

For imaging optics, many of the algorithms are based on the expectation that alignment marks are made to be symmetric about their reference positions. Therefore, algorithms rely on finding the symmetry axis and use it as a proxy of the reference position.

For instance, a simple algorithm to find the symmetry axis of a symmetric signal using the concept of center-of-mass can be defined as,

$$x_0 = \frac{\int xI(x)dx}{\int I(x)dx}$$

Eq. 2-11

where x_0 is the estimated reference position found from the signal. Recall that we have defined our coordinate system so that the true reference position is at the x axis origin, $\hat{x}_0 = 0$. If the signal is symmetric, i.e., $I(x) = I(-x)$, the estimate of the reference position given by Equation 2-11 is error-free, because $x_0 = 0 = \hat{x}_0$.

The disadvantage of this particular algorithm is that is uses all parts of the signal, even including where the signal strength is low and therefore photon-counting noise can be dominant. A better approach would be to use only part of the signal where the intensity is strong so that the noise effect is relatively small.

Another possible variation on the theme can be the following: if there is a peak in the signal, use only the part that is close to its peak and fit it to an inverted parabola. The center of the parabola should give the reference position.

As shown by the examples above, there can be a variety of different algorithms, all of which are designed to look for the symmetry axis of a symmetric signal, but in slightly different ways [4]. The main differences between them are their responses to

24

noise. Finding data in a noisy environment is a very well-studied and documented topic. Since this is not the emphasis of this thesis, we will note in passing that most of the classical results such as matched filtering can readily find their application in designing alignment algorithms [4, 23].

If the signal is indeed symmetric, these algorithms are, on average, equivalent to each other. In other words, given a statistical sample of alignment signals, these algorithms may have different variances in their findings of the reference positions, but they all have the same mean value. For the sake of conciseness, we will refer to this group of algorithms as "symmetric algorithms". Interestingly, almost all alignment algorithms reported in the literature, including those used on commercial tools and those in research projects, belong to this category of symmetric algorithms [4]. The only exception being the correlation based algorithm implemented on one of the commercial metrology tools, as will be detailed in Chapter 7.

2.1.5 Constructing an alignment system

We are now ready to build a complete picture of an alignment system. The alignment mark is specified by its scattering matrix. An optical system, either imaging-, scanning- or interferometry-based, can be used as an alignment system to detect the alignment mark and form an alignment signal. The signal is then processed by an algorithm to find the reference position. This is shown in Figure 2-5.
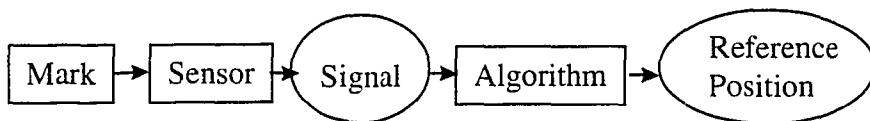
Figure 2-5. Block diagram of a generic alignment systems. The alignment mark is fully characterized by its scattering matrix. The sensor is usually an optical imaging or scanning system. The signal is processed by the algorithm to find the reference position of the mark.

25

Of particular importance are the following symmetry properties of the alignment problem:

- A symmetric mark has a symmetric scattering matrix.
- A symmetric scattering matrix interacting with a symmetric optics gives, on average, a symmetric signal.
- These signals yield, on average, correct reference positions when analyzed by symmetric algorithms.

Equipped with this understanding, we next discuss the possible causes of alignment errors in the following sections.

2.2 Causes of alignment errors

The generic picture of an alignment system shown in Figure 2-5 provides a convenient framework to categorize and analyze different causes of alignment errors.

Suppose that the true reference position of an alignment mark is $\hat{x}_0$, and the reference position found by the alignment system is x_0, the alignment error is then $\Delta x_0 = x_0 - \hat{x}_0$.

First of all, alignment error can be caused by noise, resulting from the detection electronics and the photon counting process. Since we are not particularly concerned with noise effects, let's examine the mean value of alignment error,

$$E(\Delta x_0) = E(x_0) - \hat{x}_0. \qquad \text{Eq. 2-12}$$

where $E(.)$ stands for the expectation value over a statistical ensemble formed by repeatedly sampling the alignment signals from the same alignment mark and optics.

If the above mean value of alignment error is not zero, assuming a symmetric algorithm, we must have a situation where the signal is, on average, not symmetric. According to the symmetry properties that are outlined in section 2.1.5, this leads us to the conclusion that either the alignment mark or the optics must be asymmetric. We emphasize that what is important is the mark's symmetry property <u>relative</u> to the reference position. The mark itself can be symmetric about some axis, but as long as the

26

symmetry axis is different from the reference position, the mark is still asymmetric for the purpose of alignment. In the following sections, each of these causes of signal symmetry is discussed.

2.2.1 Wafer-induced signal asymmetry

If the cause of signal asymmetry is an asymmetric mark on the wafer, it is called a wafer-induced effect.

As we have shown earlier, the scattering matrix depends on the details of the mark structure, which is given by $R(x, y, z)$, the function of index of refraction. If any of the coating, polishing or processing on the mark results in an asymmetric $R(x, y, z)$, we will ultimately have an asymmetric signal even when the detection optics is symmetric. Unfortunately, a plethora of different processing on the alignment mark can make it asymmetric.

Resist coating is one example. During the spin-on process, resist flows in the radial direction. If the relief features of the alignment mark are not parallel to the radial direction, the resist has to climb up or slide down in order to cover the mark structure. This often results in the so-called snow-fencing effect where the coating of the resist is asymmetric about the reference position of the underlying mark. This is illustrated in Figure 2-6.
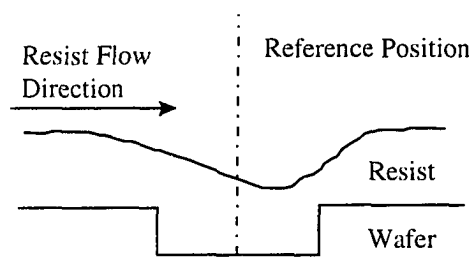
Figure 2-6. Asymmetric resist-coating over alignment mark. When the resist flow direction is not parallel to the mark structure, the coating can become asymmetric about the mark reference position.

27

Another example is metal coating. Since metal layers are coated in a sputtering process, if there is any non-uniformity in the angular flux of atoms when they impinge upon the wafer, the snow-fencing effect will also happen. Although the wafers are often rotated in a deposition chamber in order to even out the deposition flux from different angles, statistical fluctuations of the flux can easily cause the metal coating to become asymmetric. See Figure 2-7.
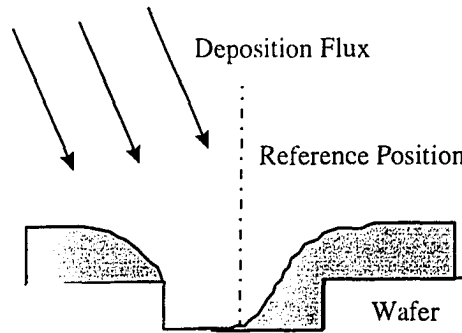
Figure 2-7. Asymmetric metal coating over alignment mark. When the deposition flux is not angularly uniform, the coating can become asymmetric about the reference position.

The most serious cause of mark asymmetry is undoubtedly Chemical Mechanical Polishing (CMP). During a CMP process, a pad is pressed against the wafer surface, while both the wafer and the pad are rotating about their own axes. A chemical solution, known as a slurry, is applied between the wafer and the pad. The action of mechanical polishing is of course abrasive, which causes patterns on the wafer to "bend" towards the direction of the polishing. Perhaps what is more subtle is that the chemical reaction catalyzed by the slurry can also lead to asymmetry. It is found that the reaction speed varies depending on the topography of the neighboring relief patterns, because they can either trap inside or leak away the chemical solutions and thus change the chemical's concentration [24]. This is can be called a proximity effect of polishing. Thus, an

28

alignment mark can be made asymmetric if it does not have identical patterns in its neighborhood on two sides of the reference position. See Figure 2-8 for an illustration.
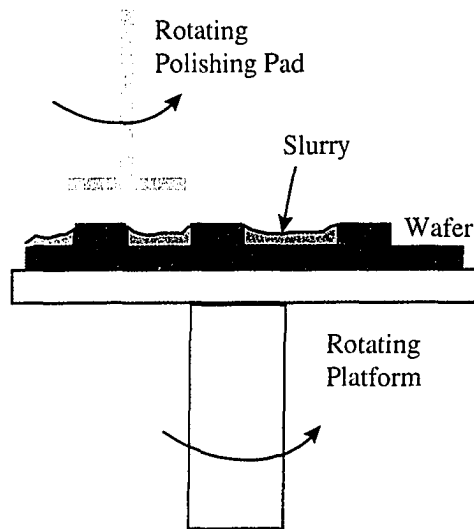
Figure 2-8. Illustration of Chemical-Mechanical Polishing. The wafer is held on a rotating platform. The polishing pad rotates and is pressed against the wafer surface. A chemical solution known as slurry is applied between the pad and the wafer.

These are but a few examples of identifiable causes that can make an alignment mark asymmetric about its reference position. These problems tend to be very difficult to model or understand quantitatively, but they obviously exist, and contribute to signal asymmetry and alignment errors to varying degrees.

2.2.2 Tool-induced signal asymmetry

Even if the alignment mark is symmetric, the signal can still have asymmetry if the optics is not perfectly symmetric.

Optics-caused asymmetry can be convincingly illustrated with the following example. Suppose our alignment optics has a considerable amount of coma in it. To make the argument simple, let's also make the assumption that our alignment mark consists of two point sources of light and that the reference position is the mid-point

29

between them. Furthermore, assume that the reference point is not placed on the optical axis. Both point sources will be imaged into comet-like clouds because of the coma. However, since the two point sources are not placed symmetrically about the optical axis, they are imaged differently. If we take the composite image of the two point sources as the alignment signal, it is asymmetric. This signal asymmetry is caused by coma alone because the idealized alignment mark is perfectly symmetric about the reference position. See Figure 2-9 for a pictorial of this analysis.
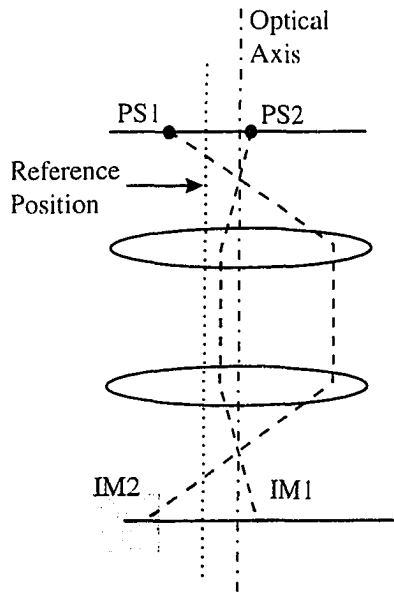
Figure 2-9. Asymmetry in image caused by coma.
Suppose two point sources PS1 and PS2 together form an alignment mark, which is symmetric about the reference position. An optics with coma is the alignment tool, therefore the images IM1 and IM2 are comet-like features. Because PS1 and PS2 have different distances from the optical axis, IM1 and IM2 have different shapes. The alignment signal, or the composite image made up by IM1 and IM2, is asymmetric about the reference position.

Although this is a thought experiment rather than a realistic example, it vividly demonstrates that a simple aberration such as coma can cause serious problems in alignment. Other optical imperfections such as aberration, distortion and misalignment,

30

can all cause signal asymmetry [25, 26]. Although these tool-induced effects cannot be completely avoided even when the optics is designed with great care, they can be minimized by optical design and calibration. In some sense, wafer-induced effects, including resist-coating and CMP, are more fundamental, as they are much more difficult to avoid.

2.2.3 Asymmetry problem exacerbated by signal variations

So far we have concerned ourselves with the mean alignment error defined in Equation 2-12. The combination of wafer- and tool-induced effects can make the alignment signal seriously asymmetric, which, when processed by symmetric algorithms, inevitably result in non-vanishing mean alignment errors.

The statistical ensemble that the expectation in Equation 2-12 is defined on represents the repeated sampling of the signal coming from the same alignment mark and the same alignment tool. In other words, they describe the noise effect only. The statistical ensemble can be enlarged to include samplings of signals from alignment marks located on different part of a single wafer, from different wafers, or maybe signals collected by different alignment tools. As is clear from the discussion of both the wafer- and tool-induced signal asymmetry, there is no reason for us to believe that each of these signals contains the same amount of asymmetry.

In fact, if all signals in our ensemble contain the same amount and type of asymmetry, the problem is very easy to solve, because a fixed amount and type of asymmetry gives a fixed amount of alignment error. We can experimentally determine how much we <u>always</u> err. The alignment algorithm can then be instructed to compensate for this amount and error-free alignment can be achieved.

A varying amount or type of asymmetry in the alignment signals, however, leads to <u>imprecise</u> alignment. To the extent signal asymmetry varies, alignment errors can have a large variance. Therefore, the problem of signal asymmetry is exacerbated by signal variation.

31

2.3 Generic approaches to improve alignment

To look for ways to minimize alignment errors, we again turn to Figure 2-5, which shows the flow chart of a generic alignment system.

If we remain within the confinement of symmetric algorithms, we must reduce the amount of asymmetry and variation in the alignment signals in order to achieve accurate and precise alignment. Since the signal is the result of physical interactions between the mark and the optics, any methods that reduce signal asymmetry must be physics-based. Typically, these methods study the physics of a specific cause of signal asymmetry and try to eliminate it either by changing the alignment mark structure or the sensor design. A very simple example of a physics-based improvement is minimizing the coma in the optics. More sophisticated examples will be provided in subsequent chapters.

The other possibility is to focus the attention on the algorithms. This would require us to devise an algorithm that is capable of finding the correct reference positions even when the signals are asymmetric. Evidently, this algorithm cannot be based on finding the symmetry axis, like all reported algorithms do.

Figure 2-10 summarizes the different approaches of physics- and algorithm-based techniques.

In subsequent chapters, we will provide the details of both physics- and algorithm-based approaches. Although both of them will be shown to significantly improve alignment performances compared to existing methods, there exist major differences in their approaches to the alignment problem. We will compare these two categories of methods and analyze their relative strengths and weaknesses. In the end, we hope to arrive at an ideal solution.

2.4 Summary

In this chapter we detailed the building blocks of an alignment system. We emphasized the importance of symmetry, the lack of which results in alignment errors. The asymmetry problem is further exacerbated by signal variation. When alignment signals are both asymmetric and varying, alignment will be inaccurate and imprecise.

32

Methods to improve alignment performance can be either physics- or algorithm-based. Physics-based algorithms aim at reducing the amount of asymmetry and variation in the signal. Algorithmic approaches attempt to minimize the impact of signal asymmetry and variation on alignment results.
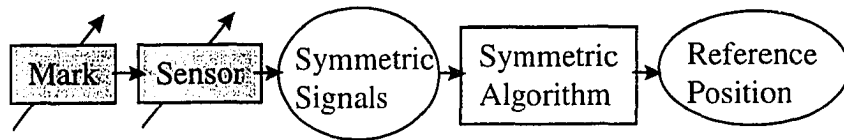
(a)
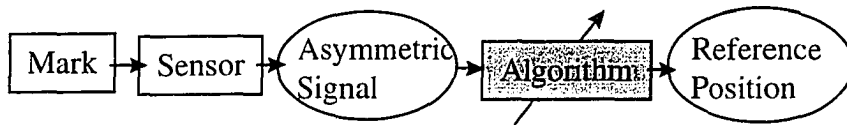
(b)

Figure 2-10. Generic approaches to improve alignment. (a) Physics-based solution. It adjusts mark and sensor designs until the signals become symmetric. (b) Algorithm-based solution. It adjusts alignment algorithm so that correct results can be obtained even from asymmetric signals.

33

PART II: PHYSICS-BASED TECHNIQUES

Chapter 3. Physics of resist-induced signal asymmetry

In this chapter, we analyze the physics of resist-induced alignment errors. The analysis leads to the alignment technique of Brewster Angle Illumination.

3.1 Resist-induced alignment errors

Among the many factors that give rise to signal asymmetry, asymmetric resist coating is the most unavoidable one, because each time alignment is performed, the wafer must be coated with resist and ready for lithography.

Resist coating is applied in a spin-on process. Drops of resist material are placed at the center of the wafer, which is then accelerated to a high rotation speed of one to two thousand revolutions per minute. The spinning motion creates a force field that causes the resist material to flow in the radial direction. This force field is strong enough to evenly distribute the resist material over the surface of the wafer. The combined effect of flowing, drying and the wetting of the wafer surface gives rise to a thin, uniform coating over the wafer [3].

The relief structure of an alignment mark behaves as a disturbance to the radial flow of resist. If the relief structure is not parallel to the local radial direction, resist has to "climb up" or "slide down" in order to cover the structure. And all these happen when resist is flowing in high speed in the radial direction. The outcome is often a coating that is asymmetric about the reference position of the alignment mark. This is illustrated in Figure 2-6, and the above snow-fencing effect is briefly discussed in Section 2.2.1.

The asymmetric resist coating becomes part of the alignment mark, which may otherwise be symmetric about its reference position. Because of the asymmetry in resist coating, the index of refraction function, $R(x,y,z)$, no longer satisfies the symmetry

34

condition, i.e., $R(x,y,z) \neq R(-x,y,z)$. We therefore have an asymmetric mark with an asymmetric scattering matrix, which can eventually results in asymmetric signals.

In some situations, resist-induced alignment errors can become very serious. [27] reports an alignment error as large as 35nm that is believed to be caused by resist coating. In research environments where devices of critical dimensions less than 0.1μm are routinely fabricated, such a magnitude of alignment errors can be devastating.

3.2 Physics of the resist-induced signal asymmetry

We can gain some insight into the physics of the problem by asking, qualitatively, how an asymmetric resist coating results in an asymmetric alignment signal.

Reflected light from the alignment mark forms the alignment signal. When the alignment mark is covered by resist, there are two components in the reflection. One component is the reflection from the resist-air interface, and the other one is the reflection from the resist-wafer interface. The latter is the one that contains the information regarding the relief structure of the alignment mark. Note that resists are usually transparent at the wavelength of the alignment light. See Figure 3-1.

Typical resist thickness is around 1μm, shorter than most alignment illumination's coherence length. Hence an interference effect exists between the two components described above. The net intensity of the reflection varies depending on whether the two reflections interfere constructively or destructively. The path difference, or relative phase, between the two reflections is approximately proportional to the local resist thickness.

When we have an asymmetric resist coating as illustrated in Figure 3-1, the resist thickness varies across the alignment mark structure. Most importantly, the thickness variation is <u>asymmetric</u> about the reference position of the alignment mark. Therefore, the phase difference between the two reflections, hence the net intensity of interference, varies asymmetrically across the alignment mark.

This analysis qualitatively explains the physics of how an asymmetric resist coating gives rise to asymmetric alignment signal. It is found that the important link is

35

the thin-film interference between the reflection from the resist-air interface and that from the resist-wafer interface.
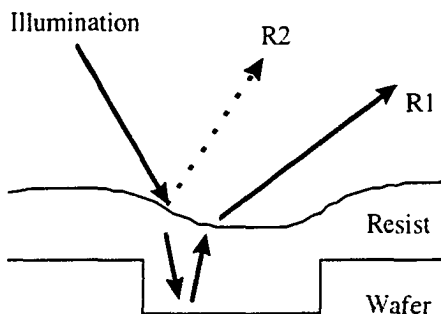
Figure 3.1. Reflected light from a resist-covered alignment mark. R1 represents the reflection from the resist-wafer interface. R2 is the reflection from the resist-air interface. Both contribute to the alignment signal.

The thin-film interference effect causes other problems in alignment as well. As reported in [28], interference effects are found to be responsible for ringing, changes in signal contrast and signal shape, and may even invert the tone of a signal, i.e., an expected peak in the signal becomes a trough. Because many alignment algorithms rely on a particular shape, contrast and tone, these uncertainties brought about by thin-film interference can cause the alignment algorithm to fail all together.

At the heart of the problem is the sensitivity of the alignment signal to resist thickness variation: the intensity of the reflection varies from minimum to maximum if the thickness changes by a quarter of the wavelength. Solutions aimed at mitigating these effects should seek ways to desensitize the signal. One approach is to use broad band illumination [28], which helps to reduce interference, but it has the shortcoming of lowered resolution [18]. Options of different coherence factors in the alignment light are often implemented on commercial steppers to provide different combinations of sensitivity and resolution.

36

3.3 Brewster angle illumination technique

In Figure 3-1, only the light reflected from the resist-wafer interface contains the positional information about the alignment mark. The reflection from the resist-air interface is in fact a noise or disturbance to the alignment problem. Ideally, we would like to reduce or even eliminate the reflection from the resist-air interface, and form the alignment signal with the resist-wafer reflection only. If this can be achieved, the alignment signal should become much less sensitive to the local variation of resist thickness. This should result in much less asymmetry in the signal even if the resist coating remains asymmetric.

The theory of electromagnetism [17] shows that the light reflection from an interface can be made to vanish if certain conditions are satisfied. First, the light must be transmagnetically polarized, i.e., TM polarized. Secondly, the light must be a plane wave with an incident angle of

$$\theta_B = \tan^{-1}(n) \, , \qquad \qquad \text{Eq. 3-1}$$

where n is the relative index of refraction of the two media that make up the interface. In the case of resist-air interface, n is simply the index of refraction of the resist. θ_B as defined in Equation 3-1 is known as the Brewster angle of the interface.

The question remains whether we can build an alignment system under such constraints. Interestingly, the interferometric system shown in Figure 2-4 and treated at great length in section 2.1.3 can be adapted to satisfy these conditions. First, we can polarize the illumination light before it enters the system. Second, we can find the correct value of the alignment mark's spatial frequency, f, so that the impinging angle $p = \theta_B$. According to Equation 2-5, we simply need

$$f = \frac{\sin(\theta_B)}{\lambda} \, . \qquad \qquad \text{Eq. 3-2}$$

For a typical photo resist, the index of refraction is about 1.6, which determines a Brewster angle of about 58°. If we assume that the alignment light is He-Ne laser with a wavelength of 633nm, we find a mark periodicity of 0.75μm. Although this mark

37

periodicity is much smaller than what is commonly used in interferometric alignment systems [21], it can be fabricated using today's technology without many difficulties.

We therefore have found a physically feasible alignment system that promises to significantly reduce, even possibly eliminate resist coating-induced signal asymmetry. The interferometry-based system employs linearly TM-polarized light, and the alignment mark should have a periodicity of around 0.75μm. We will name this approach the Brewster Angle Illumination Technique, or BAIT in short.

3.4 Summary

By analyzing the physics of how an asymmetric resist coating gives rise to alignment signal asymmetry, we reached the conclusion that the thin film interference effect is the culprit. To eliminate the interference effect, we propose the technique of Brewster Angle Illumination, which minimizes the reflection from the resist-air interface.

BAIT is believed to be very effective in suppressing the amount of signal asymmetry even when resist coating is seriously asymmetric. In subsequent chapters, we will demonstrate, both theoretically and experimentally, the effectiveness of BAIT in reducing resist-induced alignment errors.

Chapter 4. Brewster angle illumination – Theory

In this chapter, we provide some theoretical results to demonstrate the advantages of the Brewster Angle Illumination Technique.

4.1 Introduction

In order to assess the effectiveness of the Brewster Angle Illumination Technique in reducing the resist-induced alignment errors, we need to quantify the amount of alignment errors caused by asymmetric resist coating, both with and without employing the BAIT.

A reliable quantitative method should follow the scattering matrix approach outlined in Chapter 2. Such an approach, however, requires the availability of reliable methods to model the resist coating process. Unfortunately, such modeling techniques are not yet available in the commercial or research communities. Modeling of resist coating can be very complicated. First, the spin-coating process depends on many different parameters, including the viscosity of the photoresist, spin speed, acceleration, post-baking conditions, stress in the resist, and possibly even how the resist is initially applied onto the wafer. Secondly, the spin-coating process involves the dynamics of viscous fluid under violent motion conditions. The physics of such topics is still not very well understood, nor are reliable numerical methods available. Due to these difficulties, we elected to use a simple model of resist coating. Although simple, this model is believed to capture the general behavior of resist coating.

In addition, solving three-dimensional Maxwell equations to obtain the scattering matrix is largely a black box that is not amenable to analysis. It can hardly provide any insights into the problem. We therefore adopt a much simpler method using basic scalar diffraction theories [29].

As a consequence of the simplified resist model and the use of scalar diffraction theory, we do not expect the modeling results to be directly comparable to experimental data. But our model should be capable of at least revealing major qualitative trends.

39

As we will see later, this simpler model also suggests an optimal design of the wafer mark that is very intuitively appealing. The difficulty of obtaining any kind of optimality using the scattering matrix approach attests to the power of our semi-quantitative but simpler methods.

4.2 Scalar diffraction model

4.2.1 Model of resist profile

As explained above, it is difficult to model the detailed shapes of resist coating in any quantitative manner. There are, however, several properties of the resist coating process that hold in general.

First, the peak-to-valley variation in the resist profile, labeled as *2d* in Figure 4-1, is usually much smaller than the peak-to-valley height variation in the underlying relief structure, labeled as D in the same Figure. This is especially true when the resist is relatively thick [30].

Second, the high spatial frequency components of the relief feature of the alignment marks will not be followed exactly by the resist profile, i.e., the resist profile behaves as a low-pass filtered version of the mark relief feature. The attenuation of high frequency components of relief structures is more pronounced when the spatial periodicity of the mark structure is small.

A possible explanation for these two observations can be found on the grounds of surface tension. Because of the existence of a strong surface tension, the resist profile tries to shrink its own surface area in order to lower the surface energy. Of course, it is highly unlikely that the resist can reach energy equilibrium during the relatively short and violent spinning and drying processes, but surface energy considerations must still have their signatures on the resist profile.

Let's examine the coating of an isolated step, as shown in Figure 4-2(a). Assume the ideal situation where the plateaus to the left and right of the step extend to infinity. In
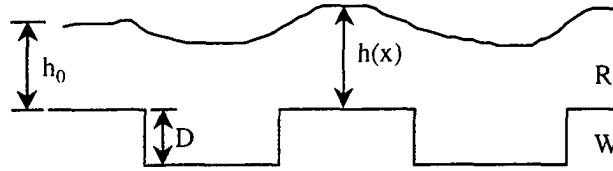
Figure 4-1. General resist coating behavior. The peak amplitude of resist profile variation is denoted by 2d. I the depth of etched trenches in the alignment mark. Ge is much smaller than D. Also, resist profile does not fo high-frequency components of the underlying relief sti shown are the nominal resist thickness, h_0; and resist p function of position, $h(x)$.

the areas that are far away from the step, resist profiles must be flat and parallel to the substrate. Around the step, however, the resist profile cannot follow exactly the sharp corners of the substrate. because such corners contain extremely large amount of surface energy due to their vanishing radii of curvature. Thus, the resist surface follows the familiar rounded shape to minimize its own surface energy, as shown schematically in Figure 4-2(a). This idealized example of an isolated step establishes a length scale Λ: the existence of the step is felt only within the distance Λ around the step; beyond Λ, the resist ignores the existence of the step and becomes flat. This length scale is of course determined by the surface tension, and the extent the resist is allowed to relax into energy equilibrium during the coating and drying processes.

Now assume that the isolated step in Figure 4-2(a) is part of a periodic structure of lines and spaces with a periodicity W, shown in Figure 4-2(b). In the case of $W \gg \Lambda$, we expect the resist to follow faithfully the underlying structure except for within the distance Λ of each step. Therefore, we have $2d \to D$. Since the resist profile is a smoothed version of a square wave, its Fourier Transform should have some energy in the

41

higher harmonics of the base frequency $1/W$, although not as much as the Fourier
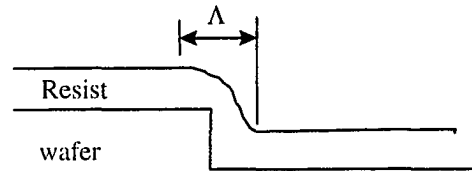
(a)
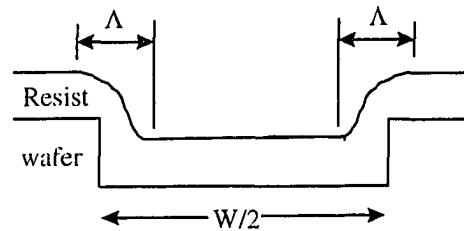
(b)
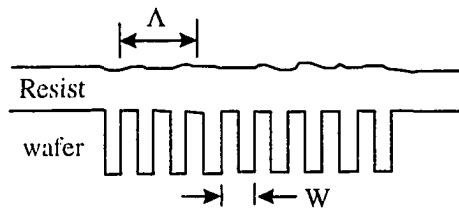
(c)

Figure 4-2. Lengths scales of resist coating. (a) An isolated step. The presence of the step is felt within the length scale Λ. (b) A periodic mark with periodicity $W \gg \Lambda$. (c) A periodic mark with $W \ll \Lambda$.

Transform of the underlying square wave.

In the other extreme case, where $W \lll \Lambda$, the resist profile ignores the ups and downs of the high frequency structure and tends towards the limiting case of a completely flat profile, i.e., $2d \to D$. The only Fourier component substantially different from vanishing is the DC. This is schematically shown in Figure 4-2(c).
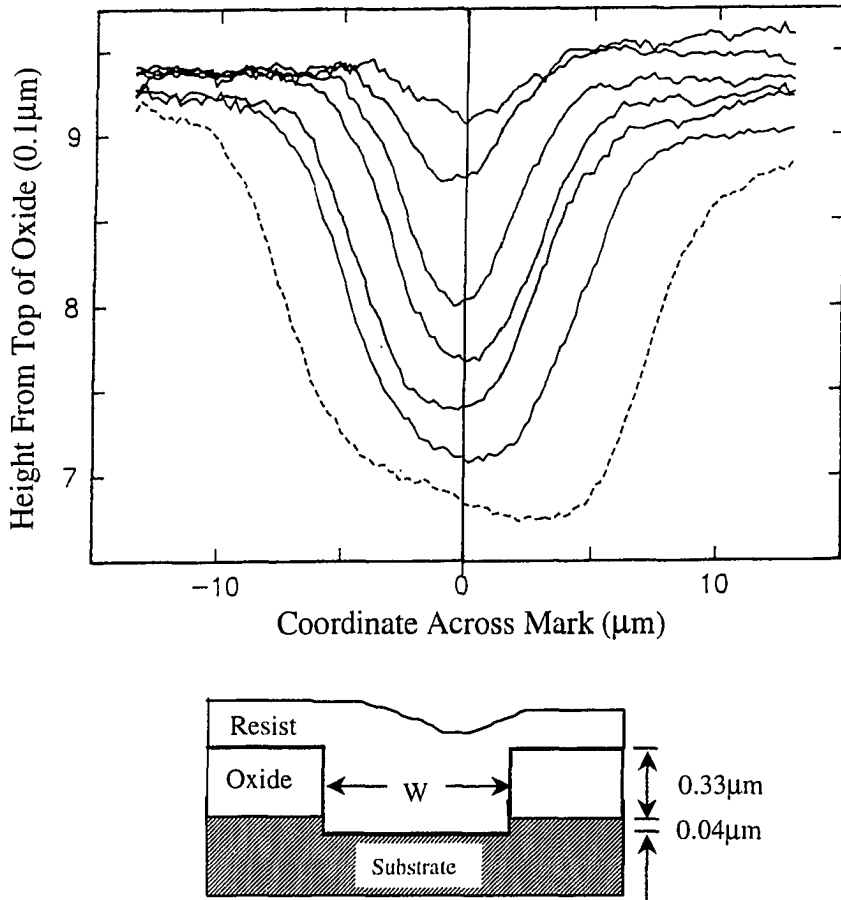
Figure 4-3. Measured resist profile as a function of mark width. The top panel shows several resist profiles measured on 1μm thick Novolak® resist coated on alignment marks. The bottom panel shows the underlying mark structure in cross-section. The oxide thickness is 0.33μm and the mark is etched 0.04μm into the substrate, thus depth of the mark is 0.37μm. From the top curve to bottom, the width of the mark W is, respectively, 1, 2, 4, 6, 8, 10 and 15μm.

43

When $W \sim \Lambda$, the situation lies somewhere between the two extremes, but the low-pass and attenuating properties are still expected to hold true.

Let's look at some experimental measurements to support the above qualitative analysis. Figure 4-3 shows profilometer measurements of the resist profiles covering trench type alignment marks with varying width [31]. The structure is shown in cross-section in the same figure. Complications of the measuring mechanism of a profilometer prevent us from using this image as a quantitative measurement of the resist profile, but several conclusions can be drawn from it.

First, the measurements can be used to estimate the length scale Λ. From this figure, we estimate $\Lambda \succ 10\mu m$. Second, since the mark etch depth is kept constant at 0.37μm, the changes in the modulation depth of resist profiles are caused by the variation of mark width. On one hand, when the mark width is 15μm, $2d \rightarrow D$, and on the other hand, when the mark is 1μm, $2d \rightarrow 0$. Both of these behaviors are predicted by a simple analysis using the length scale and mark width, as we have described earlier. Third, the resist profiles are much smoother than the underlying structure etched in the oxide, indicating the low-pass property. And finally, there are visible amounts of asymmetry in the measured profiles.

For the case of BAIT, the required mark periodicity W is 0.75μm, using the value of Λ estimated from Figure 4-3, we conclude that, $2d \prec\prec D$. Furthermore, since the resist profile is a low-passed replica of the underlying structure, we assume that it contains only the base frequency of the wafer mark structure.

The resist profile is thus modeled as,

$$h(x) = h_0 + d\cos(2\pi fx) , \hspace{3cm} \text{Eq. 4-1}$$

where $h(x)$ is the resist profile measured from the wafer surface, h_0 is the nominal thickness of the resist and f is the spatial frequency of the mark. See Figure 4-1 for illustrations of these quantities. As usual, the x-axis is defined so that its origin lies at the reference position. To be consistent, we also assume that the reference position is a symmetry axis of the underlying mark structure.

44

Since the $h(x)$ function contains only one frequency, the only way that it can be asymmetric about $x = 0$ is to have a lateral shift, i.e.,

$$h(x) = h_0 + d \cos[2\pi f (x + \delta)] , \qquad \text{Eq. 4-2}$$

where δ is the lateral shift of the resist profile with respect to the underlying wafer mark structure.

4.2.2 Optical description of resist-covered mark

When the alignment light comes in, part of it is reflected from, and the other part transmits through the resist-air interface; the local reflection and transmission coefficients as functions of position are given by

$$r^{ar}(x) = r[\theta_0 \pm \tan^{-1}(\frac{dh}{dx})] \qquad \text{Eq. 4-3}$$

and

$$t^{ar}(x) = t[\theta_0 \pm \tan^{-1}(\frac{dh}{dx})] , \qquad \text{Eq. 4-4}$$

where the superscript "ar" stands for "from air to resist". The functions $r(.)$ and $t(.)$ are the reflection and transmission coefficients of the resist-air interface as functions of the local impinging angle. They also depend on the polarization. These functions are given in [17]. θ_0 is the illumination angle measured from the optical axis. $\theta_0 = \theta_B$ if BAIT is used. The $\pm$ signs correspond to the left and right beams in Figure 4-4.
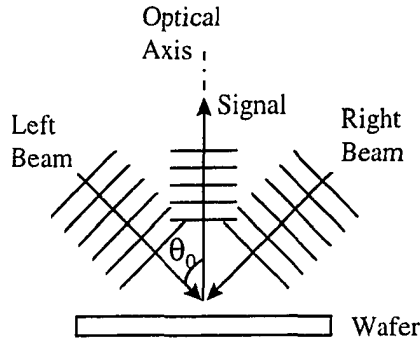
Figure 4-4. Interferometric alignment system.
This figure shows the definition of the left beam,
right beam, incident angle θ_0 and the optical
axis in an interferometric alignment system.

The reflected and transmitted light also pick up some phase modulation from the resist relief profile. Putting these phase terms into the generalized reflection and transmission coefficients in Equations 4-3 and 4-4 yields,

$$r^{ar}(x) = r[\theta_0 \pm \tan^{-1}(\frac{dh}{dx})]\exp[-i2kh(x)]$$

and   Eq. 4-5

$$t^{ar}(x) = t[\theta_0 \pm \tan^{-1}(\frac{dh}{dx})]\exp[ik(n-1)h(x)],$$

where k is the wave vector.

Note that both r^{ar} and t^{ar} are periodic functions of position x. They can be decomposed into Fourier series, or diffraction orders. Since only the light outgoing at angle $q = 0$ is collected to form the alignment signal, we need only the first orders for the reflection coefficient r^{ar}: r_1^{ar} and r_{-1}^{ar}.

The situation of the transmitted light is much more complicated. It is first diffracted by the resist-air interface, then propagates to the wafer surface and gets diffracted there. The reflected light propagates back through the resist and gets diffracted again at the resist-air interface. Part of it transmits to air and those orders emitting along

46

the optical axis contribute to the signal. The other part of it will be reflected back into the resist and undergo multiple reflections inside the photoresist. The hope is that we can make some approximations and truncate this process to analytically obtain the alignment signal as a function of the relative position of wafer and mask s.

It is interesting to note that [32].

$$FT\{\exp[ik(n-1)h(x)]\}$$

$$= e^{i2\pi s} \sum_{m=-\infty}^{\infty} (-i)^m J_m[kd(n-1)]\delta(f_x - mf) \qquad \text{Eq. 4-6}$$

where $FT(.)$ stands for Fourier transform and $J_m(.)$ is the m-th order Bessel function. The last term in Equation 4-6 is the Dirac δ-function. The Bessel functions are uniformly bounded from above [32].

$$|J_m(c)| = o(c^n), m \neq 0, c \neq 0. \qquad \text{Eq. 4-7}$$

If c is small, i.e., $c \prec\prec 1$, we can truncate the series in Eq. 4-6. Since $c = kd(n-1)$, $c \prec\prec 1$ is equivalent to

$$2d \prec\prec \frac{\lambda}{\pi(n-1)} \sim 0.32\mu m. \qquad \text{Eq. 4-8}$$

Therefore, if the peak-to-valley variation of the resist profile is small compared to 0.32μm, we can probably truncate the diffraction orders to keep the 0 and ±1 orders only. The above approximation is $o(c^2)$. We also obtained $FT(f'')$ numerically. The result shows that, for $2d \leq 0.4\mu m$, we can indeed safely ignore diffraction orders higher than ±1, because, collectively they account for less than 0.5% of the total incident energy. As we will see later, the quantity $2d$ is never allowed to become larger than 0.4μm in our calculation. Thus, the expansion in Equation 4-6 can be truncated.

Another interesting phenomenon that also works in our favor is the fact that when the spatial frequency f of the wafer mark structure is high, the higher diffraction orders become evanescent, and hence do not contribute to the alignment signal. In fact, for the

47

case of BAIT, $f = \dfrac{1}{W} = \dfrac{1}{0.75\mu m}$, and all diffraction higher than the ± 2 orders is evanescent and can be safely ignored when calculating the alignment signal.

We conclude that the transmission coefficient, t^{ar}, can be accurately characterized by the three diffraction coefficients t_0^{ar} and $t_{\pm 1}^{ar}$. The same arguments apply to the transmission of light from the resist into air; it can also be represented by three diffraction coefficients, t_0^{ra} and $t_{\pm 1}^{ra}$. Note the change of superscript from "ar" to "ra", the latter reads "from resist to air".

To concentrate our attention on the asymmetrical resist coating, we assume an idealized wafer mark. The underlying structure is modeled as a square wave of amplitude D and spatial frequency f. This is a classical phase grating and its diffraction efficiencies can be analytically obtained. Denote these by E_m, where m is the diffraction order. We have,

$$E_m = \frac{1}{W} \left(\int_0^{W/2} e^{j2\pi mfx} dx + e^{j2kD} \int_{W/2}^{W} e^{j2\pi mfx} dx \right).$$
  Eq. 4-9

The optical property of the resist-covered alignment mark is fully characterized by the amplitude diffraction coefficients given above, $r_{\pm 1}^{ar}$, t_0^{ar}, $t_{\pm 1}^{ar}$, t_0^{ra}, $t_{\pm 1}^{ra}$ and E_m.

4.2.3 Model the alignment signal

All diffraction coefficients described above are obtained with the reference position of the mark placed at the origin of the x axis. When the mark, together with the resist coating on top, shifts by an amount s in the x direction, all these coefficients should have an extra phase factor,

$$\Phi = \exp(\mp i 2\pi mfs),$$
  Eq. 4-10

where m is the diffraction order. The upper sign is for the left beam and lower sign for the right beam, because they see different shifts, $+s$ or $-s$, respectively.

48

Similarly, since the asymmetry of the resist profile is modeled as a lateral shift of the resist profile with respect to the mark, the diffraction coefficients associated with the resist-air interface should have an extra phase factor of

$$\Theta = \exp(\mp i2\pi mf\delta).$$ Eq. 4-11

To avoid notation clutter, hereafter we will assume that proper phase factors as given by Equations 4-10 and 4-11 are already included when we write the diffraction coefficients such as $r_{\pm1}^{ar}, t_0^{ar}$, etc.

Using these coefficients we can write down directly both the contribution from the left and the right beams to the alignment signal, denoted by L and R, respectively,

$$L, R = r_{\pm1}^{ar} + \sum t_{m_0}^{ar} (PE_{m_1}P) r_{m_2}^{ra} (PE_{m_3}P) r_{m_4}^{ra} ... t_{m_n}^{ra}$$ Eq. 4-12

Equation 4-12 is a rather complicated formula. We will examine the terms one by one. First of all, all upper signs apply to L, and lower signs to R. Secondly, we emphasize again that the diffraction coefficients can be different for L and R, depending on the phase factors defined in Equations 4-11 and 4-12, and the reflection and transmission coefficients defined in Equation 4-5. The first term on the right, outside of the summation, is the reflection from the resist-air interface. Only the ±1 orders are included because these are the only ones that contribute to the alignment signal. The first term in the summant, $t_{m_0}^{ar}$, means that the light transmits from air to resist, and we are looking at the m_0-th diffraction order. P is an operator that propagates the light through the resist. Assuming no attenuation, t is simply a phase factor that depends on the nominal thickness of the resist h_0 and the direction at which the light is propagating. The term inside the parentheses represents first a propagation through resist, then a reflection from the wafer, followed by another propagation back to the resist-air interface. If it is reflected again, another pair of parentheses will follow to describe the sequence of propagation, reflection and propagation. If it transmits through the resist-air interface back in to the air, the sequence terminates with a transmission coefficients, $t_{m_n}^{ra}$.

Note that the summation is limited to the set

49

$$\left\{ \{m_i\}: m_i \in \{0,1,-1\}, \sum_i m_i = \pm 1 \right\}, \qquad \text{Eq. 4-13}$$

where the upper sign is for L and the lower sign for R. The reason for the requirement $m_i \in \{0,-1,+1\}$ is that we have already justified truncating all reflection and transmission coefficients to include only the 0 and ± 1 orders. The condition

$$\sum_i m_i = \pm 1$$

comes from the fact that only these combinations of diffraction result in light propagating along the optical axis.

We simplify Equation 4-12 by two further approximations. The first one is that multiple reflections can be ignored. The reflection coefficient of the resist-wafer interface is on the order of 0.4, assuming the wafer is silicon covered by native oxide. The reflection coefficient of the resist-air interface from inside the resist is on the order of 0.1. Hence as light is bounced another time inside the resist, its energy is diminished to only 0.16% of its original value. We will therefore keep only those terms that are reflected from the wafer once. The second approximation is that, to be consistent with our earlier $o(c^2)$ approximation, only one of the m_i's can be ± 1 and all others should be zeros. This is equivalent to keeping only those contributions that have undergone ± 1 order diffraction at most once. This approximation can be justified because, according to Equations 4-6 and 4-7, each ± 1 order diffraction involves an $O(c)$ term. If the light goes through such diffraction twice, its amplitude is $O(c^2)$, and we have already justified discarding such terms in the discussion following Equations 4-7. With these approximations, Equation 4-12 has become

$$L, R = r_{\pm 1}^{air} + t_0^{air} \mathbf{P} E_0 \mathbf{P} t_{\pm 1}^{ra} + t_{\pm 1}^{air} \mathbf{P} E_0 \mathbf{P} t_0^{ra} + t_0^{air} \mathbf{P} E_{\pm 1} \mathbf{P} t_0^{ra}. \qquad \text{Eq. 4-13}$$

Finally, the alignment signal is just the interference of L and R, i.e.,

$$I = |L + R|^2. \qquad \text{Eq. 4-14}$$

Some straightforward algebra shows that,

$$I \propto \cos^2[2\pi f(s+\varepsilon)], \qquad\qquad \text{Eq. 4-15}$$

and

$$\varepsilon = \frac{1}{4\pi f} \angle(L^* R), \qquad\qquad \text{Eq. 4-16}$$

where $\angle(.)$ denotes the phase angle of a complex quantity, and $(.)^*$ represents complex conjugate.

Comparing Equations 4-15 and 2-10, we can see that ε is the alignment error induced by the asymmetric resist coating.

4.3 Modeling results and discussion
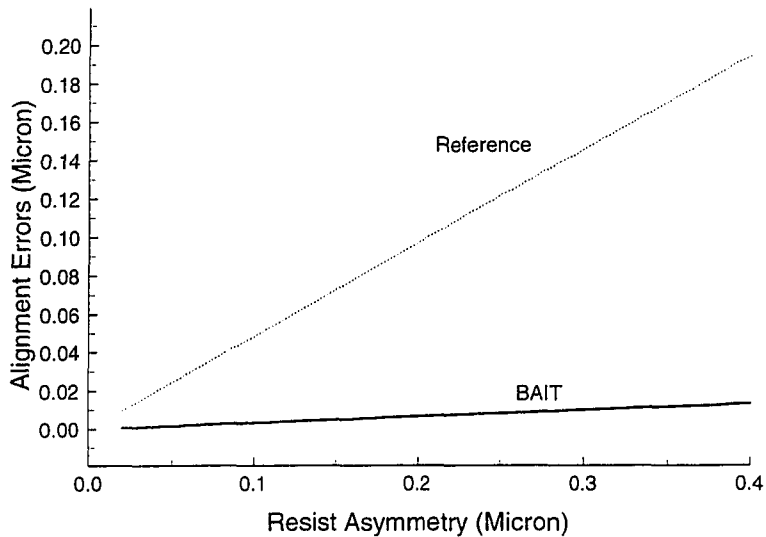
Figure 4-5. Calculated alignment error plotted against resist asymmetry, or the lateral shift δ. The upper curve is obtained under the conditions that emulate a commercial stepper. The bottom curve is calculated using the BAIT conditions.

51

Figure 4-5 plots the alignment error versus the amount of resist asymmetry. For the upper curve labeled "reference", we take numerical values according to the specification of a successful interferometry-based commercial alignment system, in which case, $D = 0.16 \mu m$, $W = 16 \mu m$. [21, 30] We also assume that $d = D/2$. The commercial alignment system uses non-polarized light, which is simulated using an equal-energy mix of TE and TM in our model. This combination of parameters will be referred to as the reference case. If the Brewster Angle Illumination Technique is used, we specify that the polarization is TM and that $W = 0.75 \mu m$. All other parameters such as D and d are kept the same as in the reference case in order to put the comparison on fair grounds.

Both curves are linear in Figure 4-5, indicating a strong linear relationship between the resist asymmetry and its induced alignment error. Because the modeling result in Equation 4-16 is a first-order theory, this linear relationship is not surprising. In fact, for all the numerical situations we explored, we found the alignment error ε to depend linearly on the asymmetry δ. This enables us to eliminate one variable and define an error transfer ratio, η,

$$\eta = \frac{\varepsilon}{\delta}.$$ Eq. 4-17

which measures how much of the resist asymmetry is transferred into the final alignment error, or, equivalently, how sensitive the alignment system is to resist asymmetry.

For the reference case, an error transfer ratio of $\eta_{ref} = 0.45$ is found. But for the BAIT case, we find $\eta_{BAIT} = 0.04$, a better than ten-fold improvement over the performance of the commercial alignment system.

Thus the semi-quantitative scalar diffraction model predicts that BAIT should perform much better than the conventional system. The performance is measured in the error transfer ratio, or how much of the resist asymmetry translates into alignment error.

The resist asymmetry, δ, can be different in the two systems we considered, due to complications in the resist coating process, as explained in Section 4.2.1. The above conclusion regarding the performance of two alignment systems does not specify how

52

much resist asymmetry exists in either of these two system, it simply states that BAIT is much less sensitive to resist coating asymmetry than the configuration of a commercial alignment tool.

Next, let's examine the robustness of the BAIT against process variation. Since the pitch of the alignment mark is fixed at $W = 0.75\mu m$ in order to achieve the Brewster angle effect, the only parameter we can vary is D, the etch depth of the alignment mark.
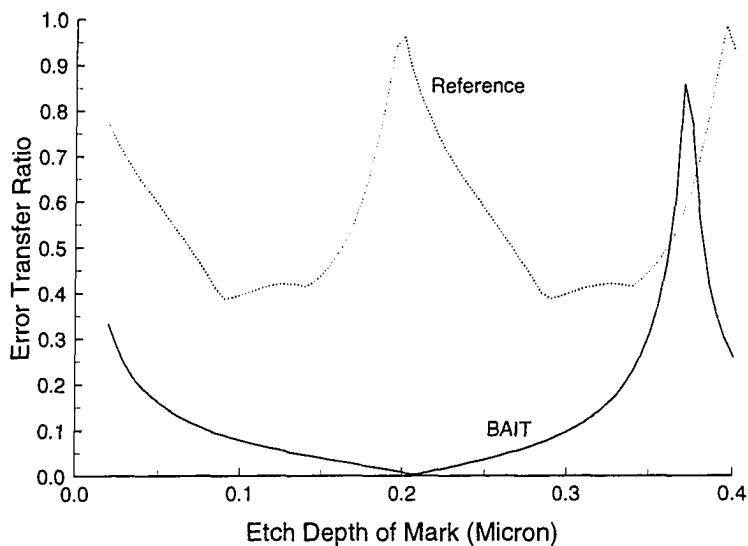
Figure 4-6. Calculated error transfer ratio plotted against mark etch depth. The upper curve is obtained under conditions that emulate a commercial stepper. The lower curve is calculated using the BAIT conditions.

Figure 4-6 plots the error transfer ratio as a function of the etch depth of the alignment mark, which is varied from 0.02μm to 0.4μm. The upper curve is obtained for the reference case and the bottom curve is the predicted behavior if BAIT is used. The first feature to notice in this graph is that the BAIT curve lies below the reference curve,

53

meaning that for most mark configurations, BAIT is less sensitive to resist asymmetry. Also, the BAIT curve changes more gradually than the commercial case, meaning that BAIT is relatively immune to process conditions and works nearly equally well even when the mark configuration varies widely. In addition, there are other very interesting features in this plot.

First, in the reference case, error transfer ratio approaches unity at both $D = 0.19 \mu m$ and $D = 0.39 \mu m$. This again happens in the BAIT case with $D = 0.37 \mu m$. The interesting question is under what condition all resist asymmetry translates into alignment error.

To answer this question, let's turn to Equation 4-13. It states that there are four components in each of the left and right beams' contribution to alignment signal. These four terms are graphically represented in Figure 4-7. The first term is a reflection term, which is minimized if BAIT is employed. Recall that all of the asymmetry of the alignment mark under consideration is contained in the resist profile, namely, the resist-air interface. Also recall that from Equations 4-10 and 4-11, the 0-th order diffraction at an interface does not carry any information about the structures at the interface. Hence, the 0-th order diffraction at the resist-air interface cannot sense the existence of asymmetry. To this light, the mark is actually symmetric. The component shown in Figure 4-7(d) has only 0-th order diffraction at the resist-air interface. It must be error-free.

Similarly, if the light undergoes 0-th order diffraction at the resist-wafer interface, it cannot at all sense the existence of the underlying mark. The components shown in Figure 4-7(b) and 4-7(c) are examples of this situation. Since they cannot "see" the correct location of the mark, which is defined at the resist-wafer interface, these components actually align to the resist-profile. Thus if the components depicted in Figure 4-7(b) and (c) are the only ones present, the error transfer ratio is unity. The same analysis applies to the reflection component shown in Figure 4-7(a), it also fully transfers the error.

54

We therefore speculate that unity transfer ratio occurs when the component in Figure 4-7(d) vanishes and alignment signals are formed by the error-carrying components in Figure 4-6(a), (b) and (c). This can only happen if $E_{\pm 1} = 0$. Recall that Equation 4-9 gives the diffraction efficiency as a function of the etch depth. Inspection of this equation shows that, indeed, each of the unity transfer ratio in Figure 4-5 corresponds to a vanishing first order diffraction efficiency of the underlying square wave mark.
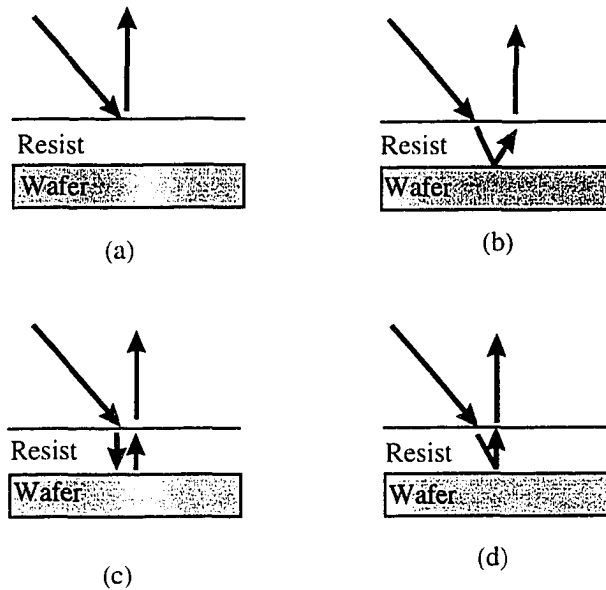
Figure 4-7. The components of alignment signal. (a) The light goes through a +1 order diffraction. (b) The light goes through a 0-th order diffraction at the resist-air interface, then a 0-th order diffraction at the resist-wafer interface, then a +1 order at the resist air interface. The sequence can be denoted 0-0-1. (c) The sequence 1-0-0. (d) The sequence 0-1-0.

Ideally, we would like to eliminate all but the error-free components of the alignment signal. The first error-carrying component, the reflection, can be minimized by BAIT. The two others in Figure 4-7(b) and (c) can be made to vanish if $E_0 = 0$, according to Equation 4-13. A vanishing 0-th order diffraction efficiency is indeed

55

achieved at $D = 0.20\mu m$, explaining the vanishing error transfer ratio in the BAIT curve around this mark configuration.

This is a very encouraging result. It tells us that even when resist coating contains substantial asymmetry, error-free alignment can still be achieved by making the alignment system completely immune to this asymmetry. All that are needed are the Brewster Angle Illumination Technique and properly designed marks. The mark should have vanishing 0-th order diffraction efficiency under resist.

4.4 Summary

A scalar diffraction model reveals many encouraging features of the problem of aligning on resist-coated alignment marks. Calculations show that the Brewster Angle Illumination Technique can suppress resist-induced alignment errors by more than ten-fold compared with commercial alignment systems.

An extremely encouraging result is also indicated by this semi-quantitative model: if BAIT is employed on correctly designed marks, nearly error-free (less than 5nm error) alignment can be achieved even when the resist coating is grotesquely asymmetric, e.g., resist profile shifts from the underlying mark pattern by as much as 0.4 μm.

56

Chapter 5. Brewster angle illumination – Experiments

In this chapter, we present experimental results that convincingly demonstrate the effectiveness of BAIT in eliminating resist-induced alignment errors.

5.1 Motivation

We demonstrated the advantages of BAIT using a scalar diffraction model. It is of course highly desirable to obtain experimental evidence that the BAIT is a viable technique to reduce resist-induced alignment errors.

In fact, an experimental study is also motivated by the confusion over resist induced alignment errors that exists in the literature.

In Chapter 2, we stated that many alignment mark structures can be described as being two-dimensional because they are roughly invariant when translated along some axis. When the mark is coated with resist, the amount of asymmetry in the resist coating depends on the relation between this axis of translational invariance and the local flow direction of the resist. One expects that when the translational axis of the alignment marks is parallel to the flow direction, there should be no asymmetry, because one side of the reference position is no different from the other. One further expects that asymmetry takes the maximum value when the translational axis is perpendicular to resist flow. Between these two extremes, i.e., when the flow direction makes an acute angle with the translation axis, some amount of resist asymmetry is expected to exist.

Explicitly or implicitly, many researchers made the assumptions that, first, the alignment error as a two-dimensional vector always lies in the direction of radial flow, regardless of the orientation of the alignment mark; and second, the magnitude of alignment error depends linearly on the mark's distance to the center of the wafer. Under these two assumptions, the resist induced alignment errors across the wafer behave like the effect of a simple magnification error. Sometimes, this belief is cited in order to justify ignoring the problem of resist-induced error because magnification errors can be easily compensated for in the procedure of global alignment, as is evident from the

57

description in Section 1.2.4. In other cases, this linear and radiating pattern is used as a template to look for resist-induced alignment errors [33].

It is our belief that the above assumptions are very weakly supported. These assumptions can only hold true if the resist behaves linearly. Specifically, we need to break down the radial flow velocity of the resist into two orthogonal components, one in the x direction and one in the y direction. Note that the x flow component is parallel to the y mark and vice versa. Furthermore, we need to assume that the resist asymmetry created in the y mark is proportional to the y velocity, and same applies in the x direction. Only when these are true can we expect the linear and radiating pattern of resist induced alignment errors. However, the resist coating process, which includes the flow and drying of viscous material under very violent motion conditions, is known to be highly nonlinear. The measured resist profiles shown in Figure 4-3 strongly suggests nonlinearity in resist coating. The linear assumptions made above are at best gross idealizations.

We hence speculate that these complications contribute to the inconsistency of reported results regarding resist induced errors [27, 33, 34]. At the same time, we feel compelled to conduct a systematic and well-defined study of the issue of resist-induced alignment errors.

5.2 Experimental setup

To study the resist-induced alignment errors and to evaluate the BAIT, we assembled a relatively straightforward implementation of the interferometric alignment optics. (Figure 4-4)

A sketch of the optics for the experimental study [35] is shown in Figure 5-1. The definition of a Cartesian coordinate system follows the convention used in previous chapters and is shown in the figure.

A 15mW He-Ne laser with wavelength $\lambda = 633nm$ and linear polarization is used as the illumination source. The main optics comprises a pair of microscope objectives (MOs). The first one is an off-the-shelf 10x MO. The second one has a magnification of 100x and a numerical aperture of 0.9. This MO is custom-made so that it can be used in

58
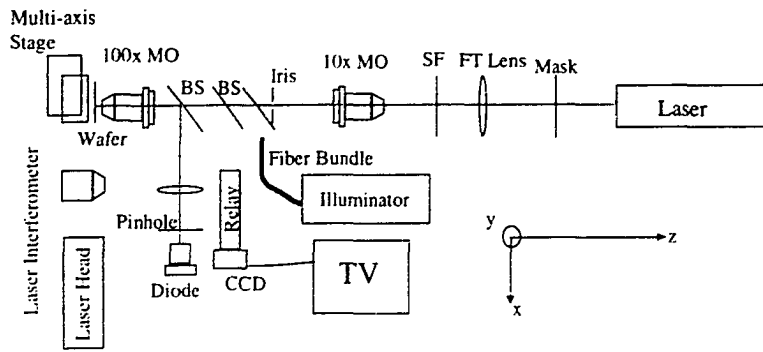
Figure 5-1. Schematic of the experiments. FT stands for
Fourier Transform. SF is spatial filter. BS is beam splitter. MO is
microscope objective. Also shown are the coordinate axes. Please see
text for a detailed description.

either direction. It can be used as a microscope objective which amplifies the object by one hundred times. It can also be used in the reverse direction. in which case it forms a real image that is de-magnified by the same amount. The combination of this pair of MOs images the mask pattern, shrunk by ten times, onto the wafer plane.

The mask is fabricated using electron beam lithography and comprises equally spaced lines and spaces parallel to the y axis. It essentially serves as a transmission grating. On the wafer we also etched some lines-and-spaces patterns parallel to the y axis with a periodicity ten times smaller than that of the mask; therefore the wafer mark and the mask image have the same periodicity. thus satisfying the condition for an interferometry-based alignment system as shown in Figure 4-4.

After passing through the mask. the laser light is diffracted into multiple diffraction orders. The Fourier Transform (FT) lens behind the mask focuses each diffraction order into a spot on the FT plane. where a spatial filter is placed. The spatial filter is essentially a plate with only two openings. coated with absorptive material. The openings pass only the +1 and -1 orders of the diffraction. After this spatial filtering step. the effective mask in our experiment is no longer a square wave transmission grating. it becomes a sinusoidal one. The above are standard optical processing techniques that can be found in classical Fourier Optics textbooks [15].

59

The mask image formed by the two MOs on the wafer surface is the interference pattern of two plane waves, i.e., a sinusoidal intensity pattern. The reflection from the wafer is directed into the diode by the beam splitter. The FT lens and the pin-hole in front of the diode form a spatial filter that collects only the plane waves propagating along the optical axis, i.e., the 0-th order diffraction. It is clear from this description that our experimental setup as shown in Figure 5-1 is an implementation of the principles described in Figure 4-4, i.e., an interferometric alignment system.

There is also an incoherent imaging system that exists in parallel with the coherent interferometric alignment tool. The incoherent light from a bulb is directed on the wafer by the fiber bundle and the beam splitter. The wafer's image is formed on the CCD camera and monitored on a TV set. This image is used for finding the patterns on the wafer, for focus, and for other adjustments.

The wafer is held on a computer-controlled stage, which can move in the x, y and z directions and rotate about the z axis. In each of the x, y and z directions, stage positioning is controlled by the dual motion capability of a lead-screw driven coarse stage and a piezo-electrically driven fine stage. The coarse stage has a positioning precision of around $1\mu m$ but is capable of linear motion within a 2mm range. The fine motion stage has a positioning precision of 3nm but can only travel $100\mu m$, according to factory specification. The combination of these two allows for fine positioning over a rather large area. The stage's x and y motion are used to scan the wafer relative to the mask. The z motion is used to find the correct focus during imaging. Furthermore, the stage's x position is monitored by a laser interferometer system, which has a resolution of 5nm.

The stage vibration has a standard deviation of 5nm when the system is left idle. The whole experimental setup is enclosed in a cloth tent to minimize the effect of varying temperatures and air turbulence. We measured the laser-interferometer readings when the stage was left idle for a long period of time. The stage drift obtained from these experiments varied from 0.003nm/s to 0.12nm/s. Since the data collection for a single experiment takes much less than a second, the drift effect is small compared to the measurement results, as will be shown later.

60

When the wafer stage is driven by the piezos, the diode collects the alignment signal which is then amplified, sampled and digitized, while the laser interferometer system records the position of the stage. Thus an alignment signal as a function of the stage position is formed. A sample signal is shown in Figure 5-2. A typical value for the signal-to-noise ratio is 13dB.
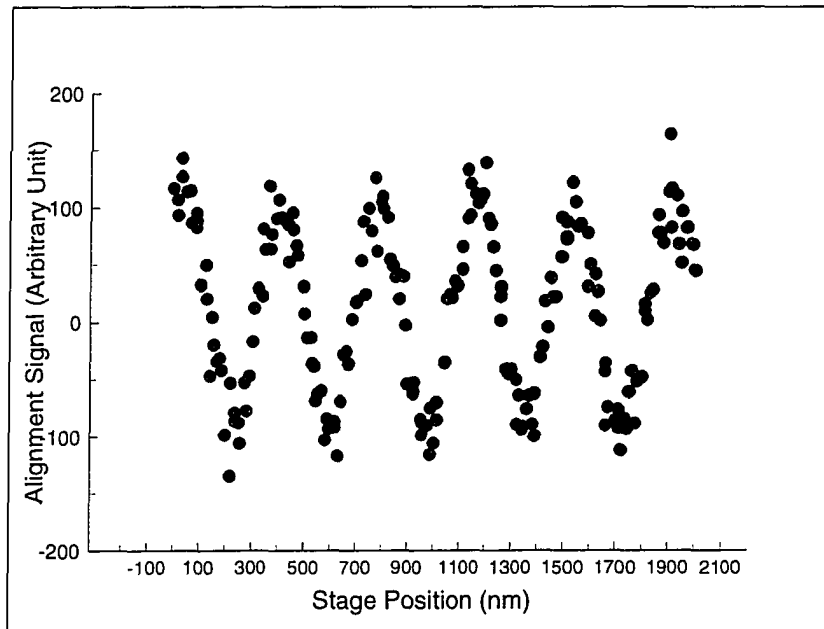
Figure 5-2. Plot of a sample alignment signal obtained from our experimental setup.

To characterize the performance of the instrument, we carried out repeated measurements on the same location of the wafer. These experiments showed that alignment error measurements have a 1σ repeatability of about 6nm. This offers an estimate of the detectable length scales using this instrument.

61

5.3 Detailed measurement procedure
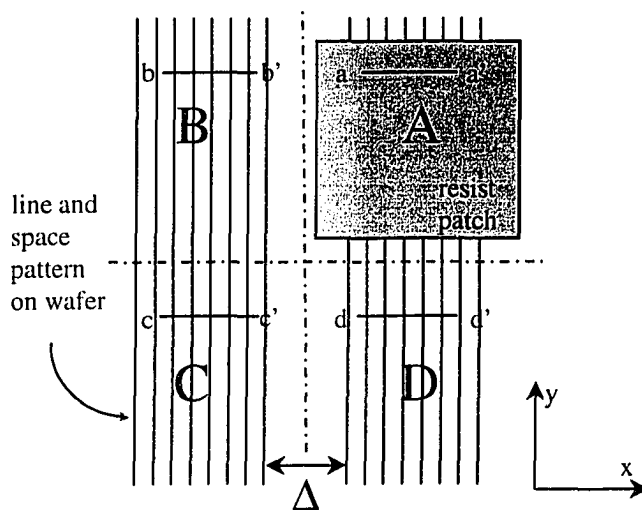
Figure 5-3. Illustration of the experimental procedure. There are four quadrants in this graph, from A to D. Position of the resist-covered mark can be determined from the aa' scan. Position of the uncovered mark can be found by the bb', cc' or dd' scan. The difference between these is the resist-induced error. Many complicated issues, such as rotation, stage cross-talk, etc. can complicate such measurements. Also shown is the distance between the left and right halves of patterns, Δ.

The resist-induced alignment errors are found to range from several nanometers to several tens of nanometers. To reliably measure effects on this scale, the experimental procedure needs to be careful designed. We tested several basic differential methods and they are found to give statistically identical results if careful calibrations and adjustments of the instruments are performed.

In Figure 5-3, we show the patterns made on the wafer. Note there are four quadrants in the figure, labeled as "A" to "D". The A quadrant is covered by resist and the three other quadrants contain uncovered marks.

In the first approach, we first take a scan in the A quadrant, and immediately afterwards, the wafer is moved so that a scan in the B quadrant can be taken. The scans are labeled as aa' and bb' in Figure 5-3. Recall that the alignment signal is sinusoidal in

62

such a system and that the zero phase occurs when the wafer and the mask are aligned. Suppose that fitting the signal from the A quadrant to a sinusoid yields zero-phase locations,

$$x_A = x_1 + jW/2, \qquad\qquad \text{Eq. 5-1}$$

where j is an integer, W is the period of wafer mark structure and $W/2$ is the periodicity of alignment signals. From the B quadrant, we find zero-phase locations,

$$x_B = x_2 + mW/2 - \Delta, \qquad\qquad \text{Eq. 5-2}$$

where m is another integer and Δ is the distance between the left and the right halves of patterns, $\Delta = 40\mu m$. Taking the difference of Equations 5-1 and 5-2, we get

$$\varepsilon = (x_1 - x_2) + kW/2 + \Delta = \varepsilon_0 + kW/2 + \Delta, \qquad\qquad \text{Eq. 5-3}$$

where k is also an integer. We expect that ε_0 represents the resist-induced alignment error. The only complication is that there is an ambiguity due to the periodicity. This ambiguity, however, can be clarified quite easily if we realize that the smallest W we will encounter is $W = 0.75\mu m$ and an resist-induced error as large as $W/2 = 0.375\mu m$ is unimaginable. Thus the resist induced error, ε_0, can be determined by eliminating all multiples of $W/2$ from ε in Equation 5-3. We can call this approach an "A-B comparison".

In the second approach, the comparison is made between two scan lines such as those labeled aa' and dd' in the figure. This can be called an "A-D comparison".

A third possible approach is to obtain alignment on each of the four quadrants and use the results in quadrants B and C to make sure that the lines-and-spaces are lined up with the y axis, then the comparison between quadrant A and quadrant D is much more reliable. This can be called an "four-quadrant measurement".

Both the A-B and A-D comparison methods can be undermined by some subtle problems. In the A-B comparison method, the stage has to move the wafer from A quadrant to B quadrant; this is a relatively large step in the x direction. Suppose that there is a cross-talk between the x and y stages, i.e., what is supposedly a step in purely the x direction involves both x and y movements. Then the A and B quadrant signals are

63

taken from different y-locations on the wafer. If the wafer has any residual-rotation in the x-y plane, y-location difference translates into a difference in the x locations. This unknown amount of x location difference becomes part of our measurements of alignment error.

Similarly, for the A-D comparison method, the large step in the y direction can introduce an unknown amount of x movement if there is a cross-talk between the x and y stages. Or, if there is a residual rotation of the wafer in the x-y plane, the measured alignment errors contain an unknown amount of rotation-induced x location difference between the scan "aa'" and "dd'".

The residual in-plane rotation can be measured in-situ and corrected by the rotation capability of the wafer-stage about the z axis. Cross-talks between axes of the stage can also be measured. They are compensated for by special computer programs which take the cross-talks into considerations when commanding the stage to move.

Such calibration procedures performed at the beginning of the experiment after mounting each chip on the stage are necessary for the experiment to be trustworthy. We performed such calibration and found that, after these effects are taken care of, the aforementioned three methods of obtaining the resist-induced errors become statistically equivalent. Figure 5-5 shows the histograms of measured values of resist-induced errors, obtained using the three different methods on the same chip. Statistical testing shows that they are identical distributions.

Because the stage cross-talk, drift and rotation can all change slightly in the course of time, we chose the method of A-B comparison due to its superior speed. All experimental results presented in later sections are obtained using this method.

The analysis presented here underscores the necessity of a well-designed procedure for the measurement of such minute effects as resist-induced alignment errors. The results shown in Figure 5-4 give us the confidence that our instrument and experimental design are capable of performing such measurements.

64
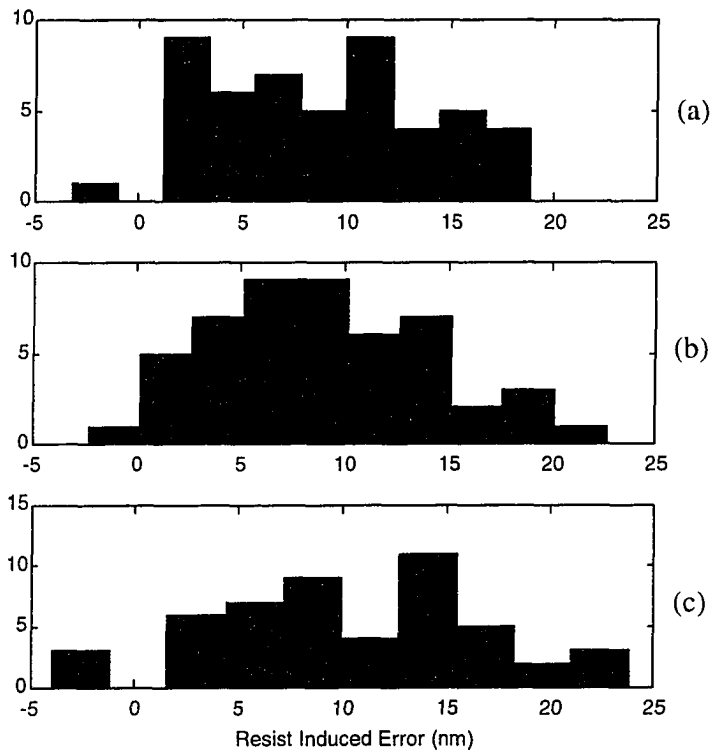
Figure 5-4. Histograms of measured resist-induced errors using three different methods. From top panel to bottom, (a) AB Comparison; (b) AD Comparison and (c) Four Quadrant Measurement. These distributions are shown to be identical by statistical testing.

5.4 Experimental results and discussion

We first measured the resist-induced errors across the wafer. To avoid the complications described in the introduction section, we separated out the cases when the wafer marks are parallel or perpendicular to the radial flow, which will be referred to as the "parallel" and "perpendicular" marks, respectively. The resist used for this study is SAL601®, and resist thickness is $1.0\mu m$. Figure 5-5 shows the results obtained for a reference case. The alignment mark is equally wide lines and spaces with a pitch of $4\mu m$.

65

The impinging angle in this case is 10°, far from the Brewster angle condition. In the figure, the circles and triangles represent perpendicular and parallel marks, respectively. Each data point is the average of twenty repeated measurements, with the error bar representing standard deviation in the sample. It is clearly seen that the perpendicular marks show measurable alignment errors that have a roughly linear dependence on the distance from the wafer center. The parallel marks, however, show alignment errors across the wafer being close or below the instrument detectability, which can be considered vanishing. This result clearly confirms the intuitive expectation of the behavior of resist-induced alignment errors, as outlined in the introduction section.
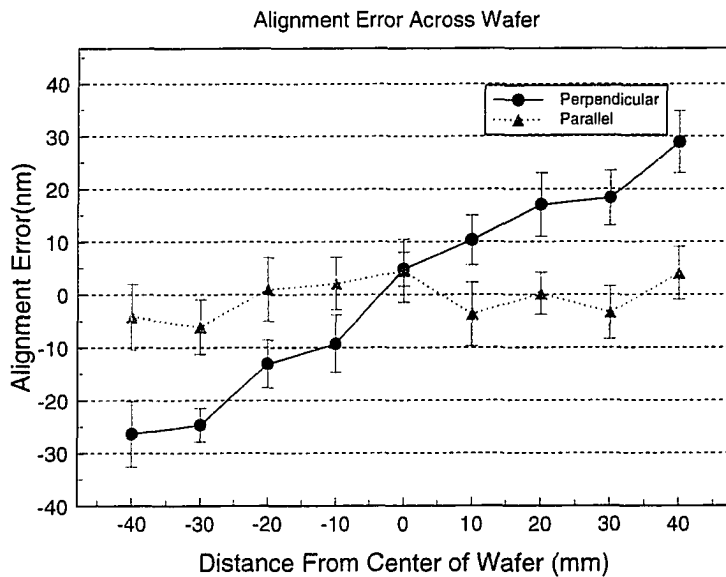
Figure 5-5. Measured resist-induced alignment errors at conditions far away from Brewster angle illumination. When the mark is perpendicular to the radial direction, there exists significant amount of alignment error. For the marks that are parallel to resist flow, there does not exist measurable resist-induced error.

Next, we repeat the same experiment on alignment marks that are optimized to eliminate resist-induced errors, i.e., using the error-free condition established in Chapter 4. These marks are line-and-space patterns with 0.75μm periodicity and an etch depth of

66

0.2μm. (Figure 5-6) Both parallel and perpendicular marks show alignment errors well below 10nm across the wafer, in most cases close to or below the instrument resolution.
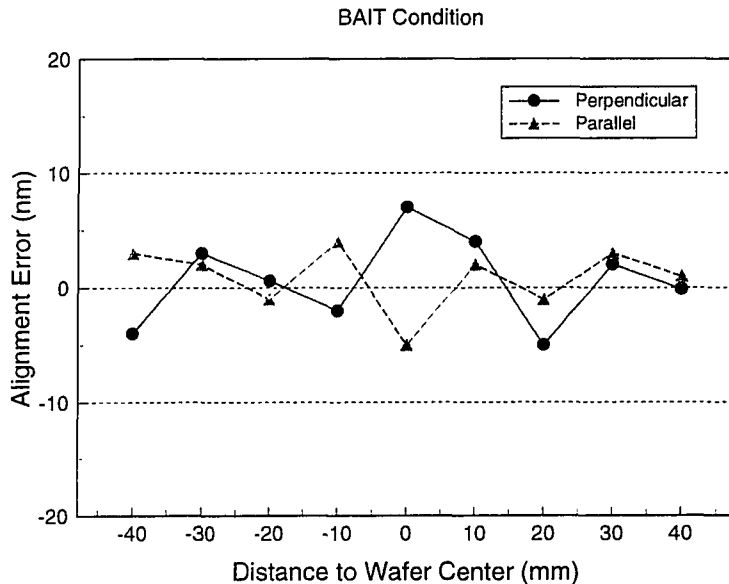
Figure 5-6. Measured resist-induced alignment errors at BAIT condition. Both perpendicular and parallel marks show alignment errors below detectability.

It is not clear from this result alone whether the BAIT has eliminated the resist-induced alignment errors or whether there is simply less asymmetry in the resist coating. These wafers have much finer structures than the one used in Figure 5-5. As discussed in Chapter 4, when coated on very fine structures, resist approaches a flat profile, ignoring the underlying relief. Since a completely flat coating cannot be asymmetric, we expect the resist asymmetry also lessens for finer mark structures.

To distinguish these two possibilities, we rotated the polarization of the illumination from TM to TE by inserting a half-wave plate into the optical path. All other parameters are left unchanged. Since the reflection from the resist-air interface does not vanish at the Brewster angle when the light is TE polarized, we expect that the

67

error-reduction mechanism to break down and hence, if there is any asymmetry in the resist coating, we should be able to measure some alignment errors. The results in Figure 5-7 indeed show measurable amount of errors for perpendicular marks, yet the parallel marks are still error-free.

BAIT but TE Polarization
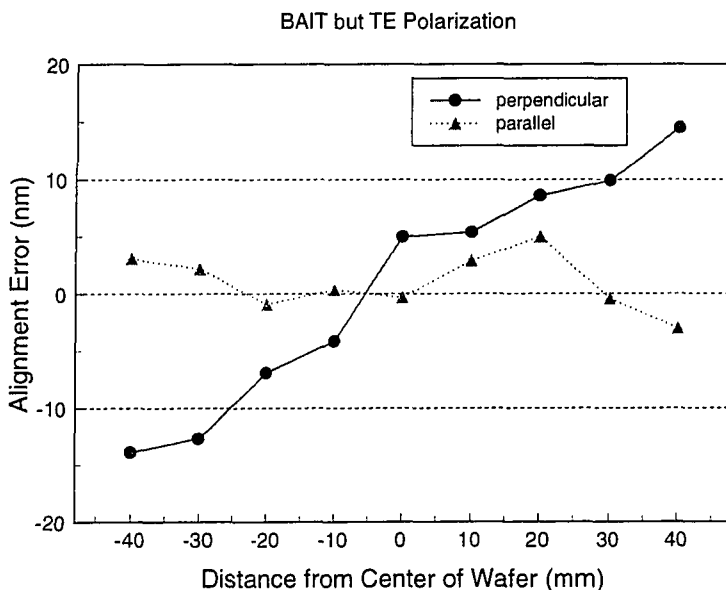
Figure 5-7. Measured resist-induced errors using Brewster angle but with TE polarization. These results proves the existence of resist coating asymmetry on the BAIT-optimized wafers. The asymmetry is detectable using TE polarization.

Comparing Figures 5-6 and 5-7 leads us to the inarguable conclusion that first, there exists resist asymmetry on the perpendicular marks, and second, alignment is error-free on these asymmetrically coated marks if BAIT is employed.

For the perpendicular marks on these BAIT-optimized wafers, the approximate linear relationship between the alignment error and the distance to center of wafer is also apparent. The magnitudes of these errors are smaller than those of the reference wafer, which may be attributed to lesser resist asymmetry as a result of finer structures of the alignment marks.

68

5.4.2 Exploring the behavior of resist coating

Several of the figures above show that, at least for the perpendicular marks, the resist coating asymmetry seems to be proportional to the radial distance to center of wafer. As discussed in Chapter 4, we do not expect generally linear behavior from the process of resist coating.

Define the orientation, ζ, of an alignment mark as the sine of the acute angle ϕ made by the radial direction and the lines-and-spaces of the mark, i.e., $\zeta = \sin\phi$. For instance, the perpendicular mark has an orientation of unity and a parallel mark has zero. If the resist induces linear errors over the wafer, we shall expect the x component of the alignment error to be proportional to $r\zeta$, where r is the radial distance from the mark to the center of the wafer. In Figure 5-8, we show a scatter plot of measured alignment errors across a wafer obtained in the TE mode, against the orientation of the mark. There is very little evidence of a linear dependence in this plot. The absence of linearity confirms our expectation regarding the behavior of resist asymmetry.

In Figure 5-8, even for marks that have the same value of $r\zeta$, the measured alignment errors can vary widely. The variation is definitely beyond the measurement repeatability, which is around 6nm. This indicates that resist asymmetry cannot be completely characterized by radial distance and orientation. The largest possible error for each $r\zeta$ value, however, does show a linear relationship. This occurs when the orientation is unity, as already shown in Figures 5-9. Apparently, when the mark is not perpendicular or parallel to the resist flow, there are other factors than the radial distance and orientation affecting the values of alignment errors. Similar results were already observed in the literature [33].
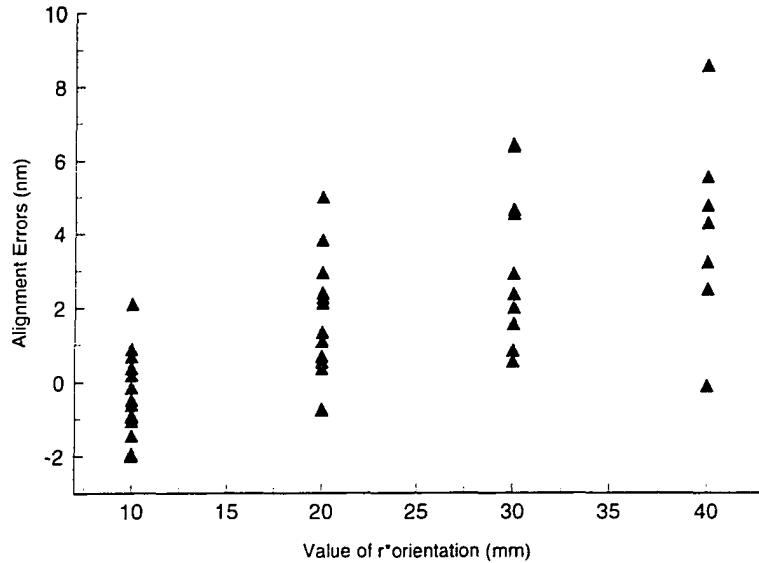
Figure 5-8. Measured alignment error vs the value of rζ. These data were obtained using TE polarization across one wafer. If resist coating behaves linearly, there should exist a linear relationship in this plot. Each point in this graph is the average of 15 repeated measurements.

Next, we explored the cases of different spin speeds and resist thickness. Figure 5-9 shows that in general the resist-induced error decreases as the resist becomes thicker. However, these results are ambiguous because there can be two factors that contributed to the decrease of error. First, because the resist is thicker, it "feels" less of the underlying structure. In the limiting case when resist is so thick that it cannot feel the existence of the mark structure, the resist profile becomes completely flat and no asymmetry exists. Hence a thicker resist tends to show less asymmetry. Second, resist is thicker because we used a slower spin speed and so, the radial flow is also slower. This could also lead to a decreasing amount of resist coating asymmetry.

70
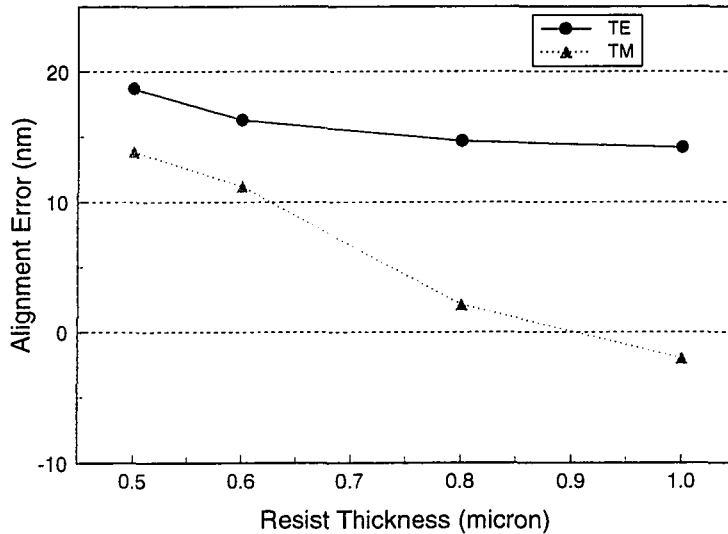
Figure 5-9. Alignment error vs resist thickness. The alignment error used in the graph is the average of the measured errors on the two outermost perpendicular marks. Different resist thickness was obtained by adjusting spin speed.

To isolate these different factors, we diluted the resist with various amount of Shipley Microposit® thinner. In one experiment, we diluted the resist so that a constant spin speed was used but different resist thicknesses resulted. The spin speed was kept at 1800rpm, while the volume ratios of thinner to resist were 0.2, 0.3 and 0.5 for the resulting thickness of $0.91\mu m$, $0.78\mu m$ and $0.61\mu m$, respectively. The alignment errors were measured from the two outermost perpendicular marks and the average value is shown in Figure 5-11. The errors measured using TE polarized light are visible but show a very slow dependence on resist thickness. Another interesting phenomenon is that the effectiveness of BAIT starts to break down when the resist becomes as thin as $0.6\mu m$. This failure is believed to be attributable to the ripples on the surface of the resist, which cause the local impinging angle to deviate from the ideal Brewster angle setting, and therefore the exact BAIT condition can not be achieved. When resist becomes thin, its

71

surface has large ripples due to the underlying structure and therefore BAIT fails to eliminate the error.
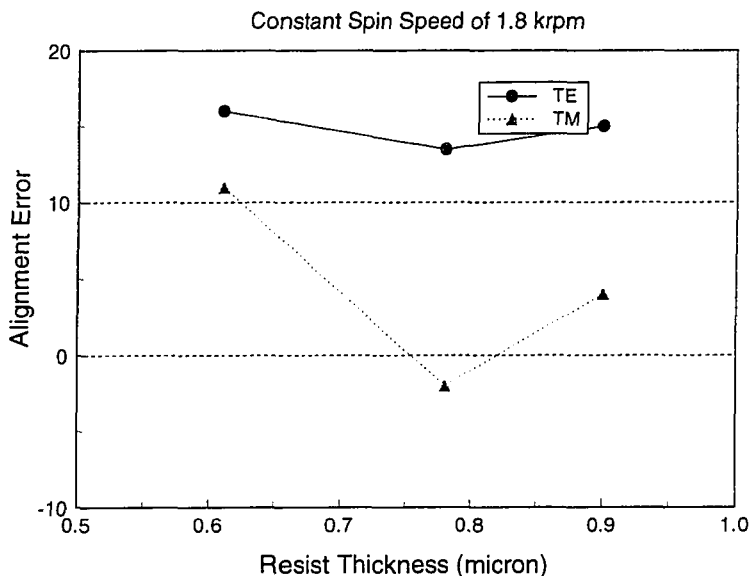
Figure 5-10. Alignment error vs resist thickness. The alignment errros shown here are averages of those measured on two outermost perpendicular marks. A constant spin speed of 1.8krpm was used. The different resist thicknesses were obtained using different dilution ratios.

In another experiment, we use different combinations of spin speed and dilution ratio to obtain a constant thickness. The thickness is held at $0.610\pm0.005\mu m$, while volume ratios of thinner to resist of 0.2, 0.33, 0.4 and 0.5 were used in combination of spin speeds of 4.0, 2.9, 2.3 and 1.8kprm, respectively, to achieve the constant thickness. Figure 5-11 shows the measured errors from the two outermost perpendicular marks. Noticeably, the asymmetry as measured by TE data shows a stronger dependence on the spin speed than on thickness. In this case, resist induced error can no longer be eliminated because the resist thickness is small and the BAIT is showing signs of failure.

72

Even under these adverse conditions, though, BAIT still performs better than non-BAIT conditions such as the TE polarization used for comparison.
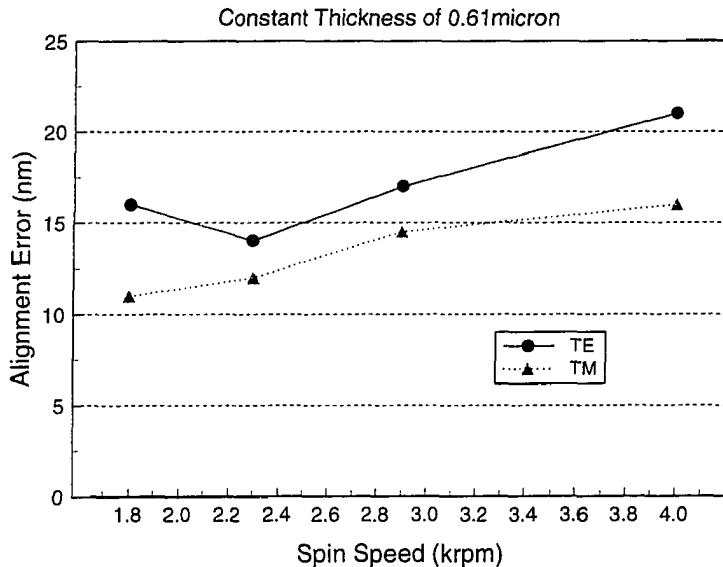
Figure 5-11. Alignment error vs. spin speed. The alignment errors shown here are averages of those measured on two outer-most marks. The resist thickness was kept constant at 0.61μm by combining different spin speeds and dilution ratios.

5.4.3 Conclusions

From the experimental results shown above, we can draw the following conclusions. As detected by using the reference wafers or TE polarization on the BAIT wafers, resist coating is in general asymmetric when the marks are not parallel to the radial flow direction, but no significant amount of asymmetry was detected for marks parallel to radial flow. For marks that are neither parallel or perpendicular to the radial flow direction, we found no evidence of a linearly-behaving resist coating asymmetry, as assumed by some casual observers. The amount of asymmetry on the perpendicular marks is more sensitive to spin speed of the coating process than the resulting thickness.

73

We observed, on the perpendicular marks, an alignment error that depends on the distance to the center of wafer in an approximately linear fashion, and error amplitude can be as large as 25nm. This magnitude depends on the structure of the underlying alignment mark: less asymmetry is observed for finer marks with smaller periodicity. The use of Brewster angle illumination with TM polarized light greatly reduces resist-induced error to below the instrument resolution in most of the parameter spaces we explored.

Comparison of TM and TE polarized light unambiguously shows the improvement in alignment accuracy provided by BAIT. When the resist becomes thinner than $0.6\mu m$, however, BAIT starts to become less effective, most possibly due to the enhanced relief structure on the surface of the resist. Even in these cases, however, BAIT still performs considerably better than a non-BAIT alignment system. Reduction in topography of the alignment marks should improve the effectiveness of BAIT for thin resist.

5.4 Practical limitations of BAIT

In the delineation of experimental results, we already noticed an important limitation of BAIT: its relative ineffectiveness in reducing errors when the resist becomes thin.

There are other rather fundamental limitations to BAIT as well.

First of all, indices of refraction of most photoresist determine that the Brewster angle is around 58°. This angle of incidence translates into an extremely high numerical aperture of 0.85. Since most exposure optics for lithography require a relatively large field-of-view and a reasonable depth-of-focus, the numerical aperture is usually limited to 0.65 or below. This means that the exposure optics cannot be used for alignment, i.e., the alignment system must be implemented in a separate system. In the terminology of stepper technology, this is called an "off-axis, non-TTL (Through-The-Lens)" alignment system. It is known that such non-TTL systems' performances are inferior to those of the TTL type[36], because of the introduction of a new error source for overlay. This new factor is the instability of the distance between the optical axes of the exposure and

74

alignment subsystems. Often called "base-line drift", this instability was found to introduce overlay errors on the order of twenty to forty nanometers [36].

Furthermore, the indices of refraction dictate that the mark periodicity must be as small as 0.75µm, unless the light wavelength is significantly larger than 633nm. Recall that there is a 2π uncertainty associated with determining the phase of a sinusoid, as explicitly expressed in Equations 5-1 and 5-2. An unambiguous alignment using such phase-detecting techniques must rely on a pre-alignment method that is accurate within the signal periodicity. For BAIT, this translates into a requirement of 0.375µm maximum error on the pre-alignment procedure. Since some pre-alignment procedures are based on coarse and inexpensive methods [4], achieving such accuracy can be difficult or expensive in some occasions.

Finally, just like resist-coating, many other processes behave like low-pass filters. When the mark is coated by other materials, its modulation depth D may be seriously attenuated if the mark has a very fine structure, resulting in a relatively flat surface relief and hence a low-contrast signal. Therefore the fine structure of the alignment marks needed by BAIT may lead to loss of signal contrast.

5.5 Summary

The Brewster Angle Illumination Technique was found to be a robust tool in eliminating resist-induced alignment errors; for mark geometry optimized for BAIT (Chapter 4.), resist-induced alignment errors as large as 25nm were reduced to an unobservable level of below 6nm.

While exploring general behaviors of resist coating, a reduction in BAIT's effectiveness was observed for resist thickness thinner than 0.6µm. It is attributed to the ripple on the surface profile of the resist.

Other practical limitations of BAIT were also discussed.

PART III: ALGORITHM-BASED TECHNIQUES

Chapter 6. The learning approach – Simulation

In this chapter, we introduce the learning approach of alignment, which is necessitated by signal asymmetry. Simulation models and results will be presented.

6.1 The necessity of learning

6.1.1 The absence of truth

The difficulties of alignment come from the existence of asymmetry and variations in the signal. The physics-based techniques presented in the first part of this thesis aim at reducing or eliminating signal asymmetry and variation. But there is another approach to solve the alignment problem. We can take the signal asymmetry and variation as given and ask whether or not we can still achieve accurate alignment. In other words, we can try to weaken the link between, or possibly disconnect, signal asymmetry and errors in the final alignment results. As depicted by Figure 2-5, the conduit from signal asymmetry to alignment error is the algorithm. We are therefore looking for algorithm-based approaches.

Almost all reported algorithms are based on the hypothesis that the signal is symmetric about the reference position. For these symmetric algorithms, signal asymmetry inevitably translates into alignment error. To achieve correct alignment on asymmetric signals, we must look beyond the group of symmetric algorithms.

But if the algorithm does not look for the symmetry axis of a signal, what should it look for?

This difficulty is a fundamental one. To clarify the predicament of aligning on an asymmetric signal, let's examine another thought experiment. In Figure 6-1(a), we show

76

an ideal alignment mark, as designed. Note that it is symmetric about the reference position. Suppose what is actually made on the wafer is the one shown in Figure 6-1(b). The task of alignment is to find the correct reference position from this mark. If the mark shown in Figure 6-1(b) is all the information we have, it is absolutely <u>impossible</u> to find the correct reference position, because we have no convincing reason to pick the correct reference position from the many other possible candidates shown in Figure 6-1(b). The information about the reference position is simply absent from the mark structure, or any alignment signals resulting from the mark.
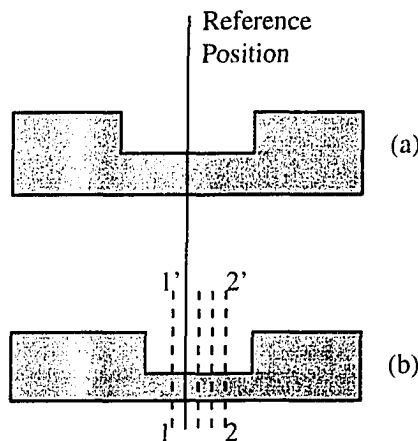
Figure 6-1. Sketch illustrating the difficulty of alignment on asymmetric signals. (a) The designed alignment mark; it is symmetric about the reference position. (b) The fabricated mark on the wafer; is it asymmetric. 11' and 22' are other possible candidates for reference positions. There exists no logical reason to pick any of them and ignore the others.

The case in Figure 6-1 is definitely an overly simplified version of what can possibly happen in reality, but it underscores the general predicament of trying to align on asymmetric signals. When an alignment signal is asymmetric, the signal alone does not provide sufficient information for us to determine the correct reference position. This

77

problem can be named "the absence of truth". It is the fundamental difficulty of aligning on asymmetric signals or marks.

6.1.2 Analogy to a communication problem

In its essence, alignment is very similar to a communication problem.

A general communication system [36] is shown in Figure 6-2(a). The information

Information To Be Transmitted $\hat{x}_0$

Correct Reference Position $\hat{x}_0$

Encoder

Mark Design

Code

Ideal Mark

Communication Channel

Processing, Coating Detection Optics

Noisy, Distorted Version of Code

Asymmetric & Varying Signal

Decoder

Algorithm

Received Information x_0

Determined Position x_0

(a)   (b)

Figure 6-2. Parallelism between the alignment and communication problems. (a) The flow chart of a communication system. (b) The flow chart of an alignment system. A one-to-one correspondence between the components of an alignment system and those of a communication system is apparent.

78

to be transmitted is contained in the code. The code is then transmitted through the communication channel. At the other end of the channel, the receiver senses a possibly distorted and noisy version of the original code. This distorted and noisy version is then used by the decoder which tries to determine the information being transferred.

A generic alignment system is shown in Figure 6-2(b). The information that we would like to find is the location of the reference position. Assume this is $\hat{x}_0$. The designed alignment mark, the symmetric one without all the coverings and processing, is a code that represents $\hat{x}_0$. The coverings and processing on the mark, together with the alignment optics, represent the communication channel through which the code propagates. The receiver senses the distorted a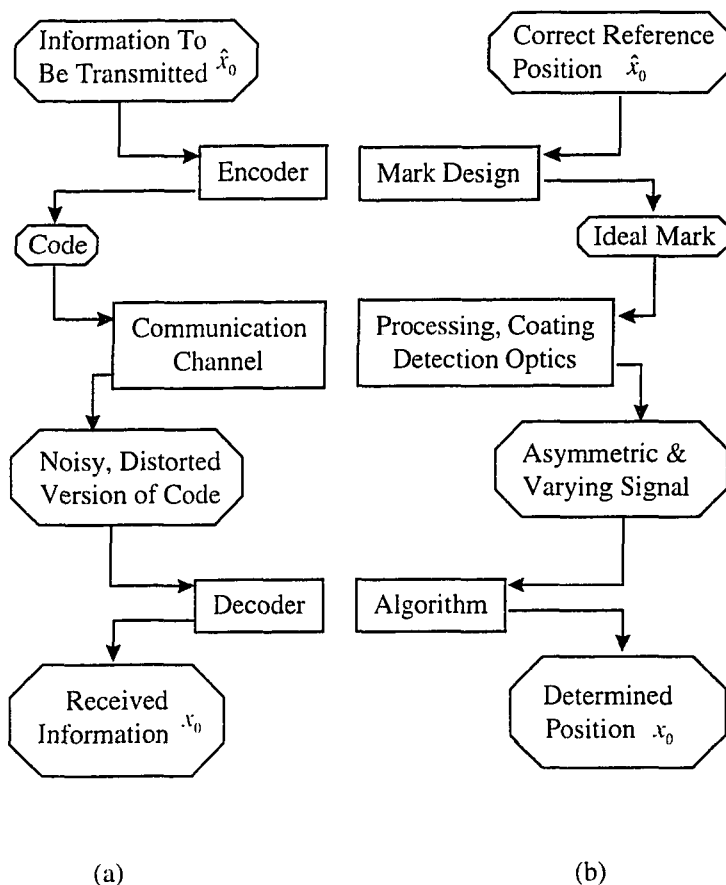nd noisy version of the code, which is the alignment signal $I(x)$. The alignment algorithm functions as the decoder and tries to find the hidden information of $\hat{x}_0$ from the signal $I(x)$. In the communication problem, channel distortion and noise is what makes the decoding problem difficult. Analogously, *coverings and processing of mark and imperfections of the alignment optics are the culprits of alignment difficulties.*

In any kind of decoding problem, <u>external</u> information must be available. By external, we mean "not contained in the code itself". The external information is typically contained in a "code book". The same applies to alignment algorithms. For instance, all the symmetric alignment algorithms makes use of the important external information that the signal is symmetric about the reference position. Asymmetry is a form of channel distortion, in the presence of which the external symmetry assumption breaks down. To correctly align on asymmetric signals, we must find other external information, or hypotheses, to build our alignment algorithm on.

There is a very general method for designing decoding algorithms in the presence of channel distortion and noise [36]. Transmit some known information, denoted by $\hat{x}_0$, through the channel and receive the transmitted code, denoted by $I(x)$. Repeat this N times so that we have a statistical sampling of the channel behavior. Let $\psi(\hat{x}_0) = \{I_i(x), i = 1...N\}$ represent this collection of received codes. Repeat this procedure for all possible choices of $\hat{x}_0$. We then have a one-to-one mapping, given by

79

$$\hat{x}_0 \mapsto \psi(\hat{x}_0), \forall \hat{x}_0.$$ Eq. 6-1

This mapping, which will be called ψ-mapping, should be able to fully characterize the channel distortion and noise, as long as the channel is static in the sense that its properties are not changing with time. The ψ-mapping is a thorough but verbose mathematical representation of the communication channel.

Let's give an example of how the ψ-mapping can be used. From the ψ-mapping, one can calculate the conditional probabilities of $P[I_i(x)|\hat{x}_0]$ and $P[\hat{x}_0|I_i(x)]$. The former represents the probability of receiving a particular signal when the transmitted information is known. Conversely, the latter is the probability of the transmitted information being $\hat{x}_0$ when the received code is $I_i(x)$. A decoding algorithm can be designed as the following if these probabilities are known,

$$x_0 = \arg \max_{all \, \hat{x}_0} P[\hat{x}_0|I_i(x)]$$ Eq. 6-2.

i.e., the chosen value of $\hat{x}_0$ is the one that has the most probability of occurring when the code is known to the received one.

Direct implementation of the decoding algorithm in Equation 6-2 is of course very inefficient. But the above derivation supports one important observation: the much needed external information for any decoding problem must be contained somewhere in the ψ-mapping of Equation 6-1. The task of designing reliable and efficient algorithms is admittedly the one of reducing or compressing the information contained in the ψ-mapping.

The same conclusion can be drawn regarding the alignment problem. In order to find the algorithm that is capable of accurate alignment on asymmetric signals, we must first go through the exercise of building the ψ-mapping, which may then lead us to some useful and concise external information, based upon which the alignment algorithm can be constructed.

The above points us in the direction of a learning approach, because during the process of building up the ψ-mapping, we are indeed learning about the characteristics of the system.

6.2 Constructing and deciphering the ψ-mapping

6.2.1 Constructing the ψ-mapping

The first issue in constructing the ψ-mapping is that the signals in the set $\{I_i(x), i = 1, 2, ..., N\}$ need to be <u>representative</u> of the processes under consideration. For instance, if we use several different tools to polish the wafer, and we have reason to believe that significant difference may exist between these tools, then we should include signals from each tool in order for this information to be present. It may as well be the case that differences between the tools are not observable. In this situation, we hope we can eliminate the "redundancy" when interpreting the information contained in the ψ-mapping. We will see later that this is indeed achieved.

To construct the ψ-mapping, we need to find corresponding pairs of $\hat{x}_0$, the reference position and $\{I_i(x), i = 1, 2, ..., N\}$, the resulting alignment signals. The alignment signals can be readily obtained if we simply record the signals during the process of lithography. There are essentially two methods to find the corresponding reference positions. One relies on overlay metrology, the other on differential alignment.

To utilize overlay metrology, alignment is first performed using any existing algorithms, for instance, one of the symmetric algorithms. The algorithm finds an estimate of the reference position, x_0, which can be different from the correct reference position, $\hat{x}_0$. Exposure is then performed assuming x_0 is the mark position. After development and possibly other processing, overlay metrology is used to measure the alignment error. From the metrology data and the estimate of the reference position, we can reconstruct the correct reference position for each signal. This procedure is illustrated in Figure 6-3.

81
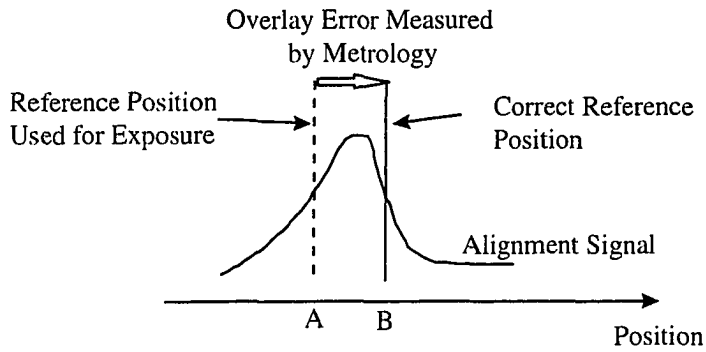
Figure 6-3. Illustration of finding correct reference position from overlay metrology. For a given signal, any existing algorithm is used to find an estimate of the reference position (A). This estimate is then used for exposure. Overlay metrology tells how much error resulted from this procedure. The correct reference position (B) can be found based on this information.
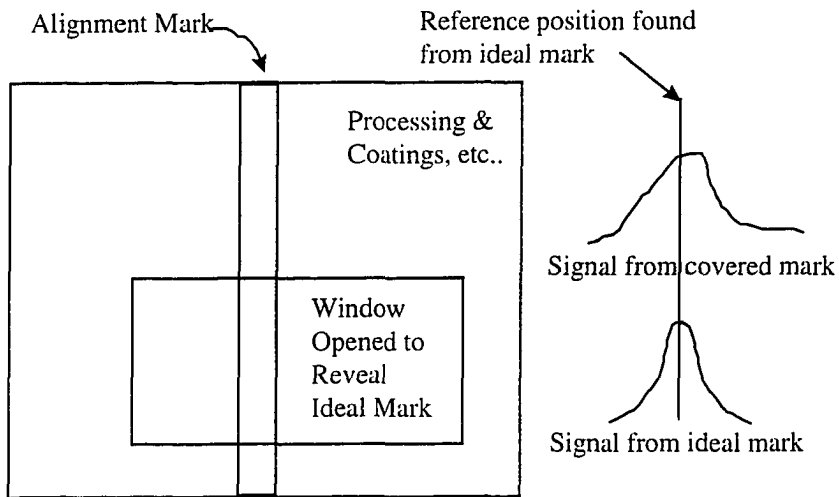
Figure 6-4. Illustration of finding reference position from differential alignment. An extra lithography step is used to remove the coatings on the alignment mark within a window. The correct reference position of the mark can be determined from the ideal mark revealed inside the window.

82

To use differential alignment, an extra step of lithography is used to remove partially the coverings that are affecting alignment. This step requires only coarse alignment. After this step, we have essentially two alignment marks, placed right next to each other, one with and one without the error-causing covering. Alignment can be performed on uncovered part, which is presumably symmetric. The alignment result on the uncovered mark is used as the correct reference position for the covered mark. This is very similar to the experimental procedure presented in Chapter 5. See Figure 6-4 for an illustration.

Neither of these two methods is error-free. As described in Chapter 1, in addition to alignment error, overlay error contains many other effects. The differential method is also error-prone, as discussed in Chapter 5. In particular, the tool-induced effects cannot be detected in the differential methods. Thus there could be errors in the reference positions obtained using these methods. We will see the applications of both these methods in the experimental studies presented in Chapter 7.

6.2.2 Simplifying the ψ-mapping

Each alignment signal can be written as, $I(x; \hat{x}_0)$, with the dependence on $\hat{x}_0$ explicitly expressed. If we change the reference position and shift it by the amount Δ, the signal should also shift by Δ, i.e.,

$$I(x; \hat{x}_0) \rightarrow I(x - \Delta; \hat{x}_0 - \Delta). \qquad \text{Eq. 6-3}$$

In other words, the functional form of the alignment signal should only depend on the difference of x and $\hat{x}_0$, i.e.,

$$I(x; \hat{x}_0) = I(x - \hat{x}_0). \qquad \text{Eq. 6-4}$$

We can therefore simply the ψ-mapping by shifting all the signals until their corresponding reference positions satisfy

$$\hat{x}_0 = 0. \qquad \text{Eq. 6-5}$$

The ψ-mapping now contains only a single map:

83

$$0 \mapsto \psi(0) = \{I_i(x), i = 1, 2, ..., N, \hat{x}_0 = 0\}. \qquad \text{Eq. 6-6}$$

We've reached the conviction that all the information regarding the alignment system is contained in the set of alignment signals taken with the reference position fixed at the origin of the x axis.

6.2.3 Deciphering the ψ-mapping: A physical model

We now build a physical model that characterizes the form and variation of alignment signals [37].

Because the form of the alignment signal depends on which processing layer is being aligned, which alignment sensor is used and where the alignment mark is located on the wafer, etc., the only practical way to model alignment signals is to build a model for a particular parameter set. We build a model for what we call a "controlled environment" — a fixed processing layer, a particular alignment sensor and a fixed mark position on the wafer. When any of these change, we have to build a different model. Possible relaxation of some of the restrictions will be discussed later.

Assuming additive noise, the alignment signal can be written in the most general form as $I(x) = I(x; p_1, p_2, ..., p_k) + \Pi(x)$ where p_k, $k = 1,2,...,n$, are n independent variables that characterize the environment; $\Pi(x)$ is random noise. Assume also we have already simplified the ψ-mapping to the set $\{I_i(x; p_1, p_2, ..., p_k), i = 1, 2, ..., N, \hat{x}_0 = 0\}$. We will refer to this set as the sample data set.

In a perfectly controlled environment, all the p_k's are fixed. Also, all N copies of the alignment signals are identical, except for the noise term. We can take the average signal $\bar{I}(x - \hat{x}_0)$ as a very simple model of the alignment system. We will refer to this simple model a template, for reasons soon to become clear. Whenever we receive a signal $I(x)$, we know that it is simply the template plus some noise. The issue of finding the position of a known signal in the presence of noise is well-studied [38], and a very simple solution is the use of correlation, i.e.,

$$x_0 = \arg \max_t [I(x - t) \otimes \bar{I}(x - \hat{x}_0)] \qquad \text{Eq. 6-7}$$

84

In this overly simplified case of a perfectly controlled environment, all the information regarding the alignment problem is reduced to a single template function, and a straightforward alignment algorithm can be designed based on correlation with the template, as expressed in Equation 6-7.

In reality, however, there are always a few parameters that cannot be perfectly controlled. Suppose that p_j, $j=1,2,...,r<n$, undergo small but random fluctuations around some fixed values $p_j(0)$. With this assumption, an alignment signal can be written as,

$$I(x) = I(x; p_1(0), p_2(0),...,p_n(0)) + \sum_{k=1}^{r} \frac{\partial I}{\partial p_k} \Delta p_k + h.o. + \Pi(x)$$

$$I(x) = I(x;0) + \sum_{k=1}^{r} I'_k(x;0)\Delta p_k + h.o. + \Pi(x)$$

Eq. 6-8

where we have used the shorthand notation $I(x;0)$ and $I'_k(x;0)$ to represent the "average signal" and its derivative evaluated when all parameters are at their fixed value $p_j(0)$; "$h.o.$" stands for higher order terms; $\Pi(x)$ represents for noise. The set $\{I_i(x; p_1, p_2,...,p_k), i = 1, 2, ..., N, \hat{x}_0 = 0\}$, or the sample data set corresponds to random variations in the parameters, i.e.,

$$I_i(x) = I(x;0) + \sum_{k=1}^{r} I'_k(x;0)\Delta p_{ik} + h.o. + \Pi(x) .$$

Eq. 6-9

Let's examine Equation 6-9 in detail. First of all, the alignment signals from a controlled environment can be expressed as linear combinations of several functions. The basis functions of the linear combination include $I(x;0), I'_k(x;0)$ and higher order derivatives. The first one is the alignment signal when all process parameters are held at their fixed-point, or the intended values. The other basis functions represent the response of this signal with respect to the changes in process conditions. Both of these can be said to characterize the controlled environment we have defined. Despite the existence of process variations, these basis functions remain the same as long as we are working in the same controlled environment.

85

The linear combination weights are <u>random</u> fluctuations of process parameters. They characterize imperfections of process control. Although the variations are random in nature, we expect their statistical properties to remain constant if the controlled environment is not changed.

Lastly, the presence of noise has a very interesting effect. Some of the random fluctuation terms in Equation 6-9 may be small compared to the amplitude of noise and cannot be observed reliably. If we assume that the environment is reasonably well-controlled so that fluctuations are small, there should be only a finite number of terms fluctuating above the noise level. Thus the presence of noise truncates the series of expansion in Equation 6-9. Consequently, the linear space spanned by the linear combinations discussed above has a finite dimension.

We can now draw the conclusion that the alignment system as a whole, including the covering and processing on the marks and the alignment optics, can be represented by a finite-dimension linear space, if the process conditions are well-controlled so that fluctuations are small. The dimension is determined by the number of process parameters fluctuating above the noise level. The bases of the linear space are given by the average alignment signal and its responses to various fluctuations of process conditions.

6.3 Subspace-based alignment algorithm

6.3.1 Finding the subspace

With the simplified and clear understanding of the information contained in the ψ-mapping of the alignment communication channel, we are ready to devise an alignment algorithm.

In the first step, the linear space described above needs to be determined. The problem is stated as the following. Given a set of alignment signals $\{I_i(x; p_1, p_2, ..., p_k), i = 1, 2, ..., N, \hat{x}_0 = 0\}$, knowing that these signals lie in a linear space, determine the dimension and basis vectors that span the linear space.

86

This is a well known problem. The problem of determining the basis vectors of a linear space is solved by the procedure called Singular Value Decomposition (SVD). SVD is a very versatile procedure that is well documented in many advanced linear algebra textbooks [39,40]. Methods of determining the dimension are also available from statistical literature, in particular, the ones related to Minimum Description Length (MDL) theory [41, 42]. Similar to the procedures associated with an eigen-system, SVD and MDL are most often used as canned procedures, because their implementations are rather complicated, and very reliable numerical routines already exist in many numerical calculation software packages [41 - 46]. Since we use both SVD and MDL as provided in the references, we will skip the detailed descriptions here.

Let $\Omega = \{B_k, k = 1,2,...,D\}$ represent the linear space found from the SVD and MDL procedures, where D is the dimension and B_k's are the bases vectors. Next, we describe how a new signal with unknown reference position can be aligned using this model space Ω.

6.3.2 Alignment using the subspace

Start with the model Ω and any estimate of the reference position, $x_0 = t$. Shift the basis functions of the model B_k so that their reference positions are at $\hat{x}_0 = t$, and denote these functions as $B_k(t)$. The collection of these shifted basis functions constructs a shifted model space $\Omega(t)$. Define the projection operator $\mathbf{P}_j(t)$ that projects onto the shifted basis functions as,

$$\mathbf{P}_j(t)I(x) = < I(x), B_j(x,t) > B_j(x,t),$$
Eq. 6-10

where $< . >$ represents inner product, which is defined for two real-valued functions $f(x)$ and $g(x)$ as

$$< f(x), g(x) > = \int_{-\infty}^{\infty} f(x)g(x)dx.$$
Eq. 6-11

Also, define the residual energy, as a function of t, as

87

$$E(t) = \left\| (1 - \sum_{j=1}^{D_m} \mathbf{P}_j(t))I \right\| \geq 0. \qquad \text{Eq. 6-12}$$

In this construction, we first project the signal onto each of the shifted bases and sum up the projections; this gives the projection of I into $\Omega(t)$. Subtracting this part from I results in the residual part, which is orthogonal to $\Omega(t)$. If t is the correct alignment position, the alignment signal $I(x)$ should lie in the subspace $\Omega(t)$. This property can be used to devise a search algorithm to find the reference position, as given by

$$x_0 = \arg\min(E(t)) \qquad \text{Eq. 6-13}$$

i.e., the reference position is found when the residual energy is minimized.

Equation 6-13 represents the best effort alignment in a least square sense using the given model Ω. A geometric depiction of the procedure that leads to Equation 6-13 is shown in Figure 6-5.

The value of minimum residual energy, E_m, is a strong indication of how well the model Ω describes the new signal. If E_m is comparable to the noise level in the alignment signals, we should be confident that the model Ω reliably characterizes the alignment system from which the new signal resulted, because almost all energy in the signal can be accounted for by the model.

Otherwise, if E_m is much bigger than the noise magnitude, it is an indication of a misfit between the model Ω and the alignment system. There could have been a dramatic change in the process environment due to maintenance, adjustment or change of recipes, etc. In this case, a new model needs to be found before reliable alignment can be performed.

We will refer to the new alignment algorithm as Sub-Space Decomposition (SSD) algorithm [37, 46-48]. It contains two steps. In the first step, the model space Ω is constructed. In the second step, the model Ω is used to align new signals.
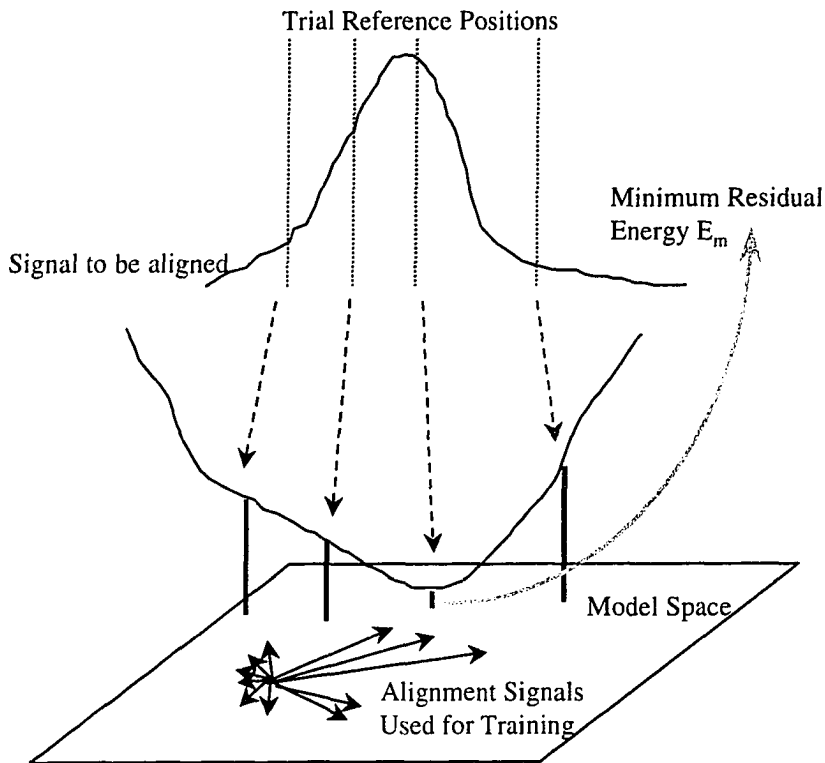
Figure 6-5. Geometric interpretation of the SSD algorithm. Sample
alignment signals are used for training, i.e.., the determination of the
model space. To align a new signal, a minimization procedure is used.
Start with any trial location of the reference position, one can calculate
the residual energy, i.e.., the amount of energy in the signal that cannot
be explained by the model space (See Equation 6-12 for a definition of
residual energy). By minimizing the residual energy, the best location
of the reference position can be found.

We'd like to emphasize that it is of vital importance we explore and evaluate
some other critical issues surrounding the new algorithm, including numerical routines,
computational considerations, noise response and optimality, etc. In-depth and detailed
discussions on these issues of the SSD algorithm form the core of another thesis [46]. In
this thesis, instead of repeating these materials, we elected to emphasize the logical bases,
physical backgrounds, experimental procedures and results.

89

6.4 Simulation results

6.4.1 Signal generation

We have implemented the SSD algorithm using computer simulation [37].

The signals are generated as

$$I_j = S + \alpha A_0 + \beta_j A_1 + \Pi,$$ Eq. 6-14

where S is symmetric about a known position $\hat{x}_0$ and A_0 and A_1 are asymmetrical about the same position. Π is a Gaussian-distributed noise with zero mean. The signal to noise ratio is on the order of 20dB. α is fixed at 0.2 and β is a random variable uniformly distributed between 0 and 0.2. α and β are introduced to simulate both fixed and random amounts of asymmetry in the signals. The signals are sampled at 128 points with a sampling period of 0.25μm. Some sample signals are shown in Figure 6-6.

The model building step is implemented using 20 samples generated by Equation 6-14. The resultant basis functions and the original ones are shown in Figure 6-7. We see that the determined basis functions are linear combinations of the original ones, characteristic of the SVD process.

We then generated 50 data signals, again using Equation 6-8. The minimization procedure is implemented using Brent's method of minimization [45].

6.4.2 Control algorithms

For comparison, we also implemented two algorithms: peak-detection and template - correlation. The peak-detection algorithm belongs to the category of symmetric algorithms. It band-passes the signal, which has the combined effect of smoothing out the noise and estimating the first-order derivative of the signal. A peak in the signal corresponds to a zero-crossing in the derivative. The peak position is found based on the location of the zero-crossing of the band-passed signals. This algorithm is implemented on one of the commercial steppers. To protect the proprietary information, we do not provide the details of the algorithm in this thesis.

90
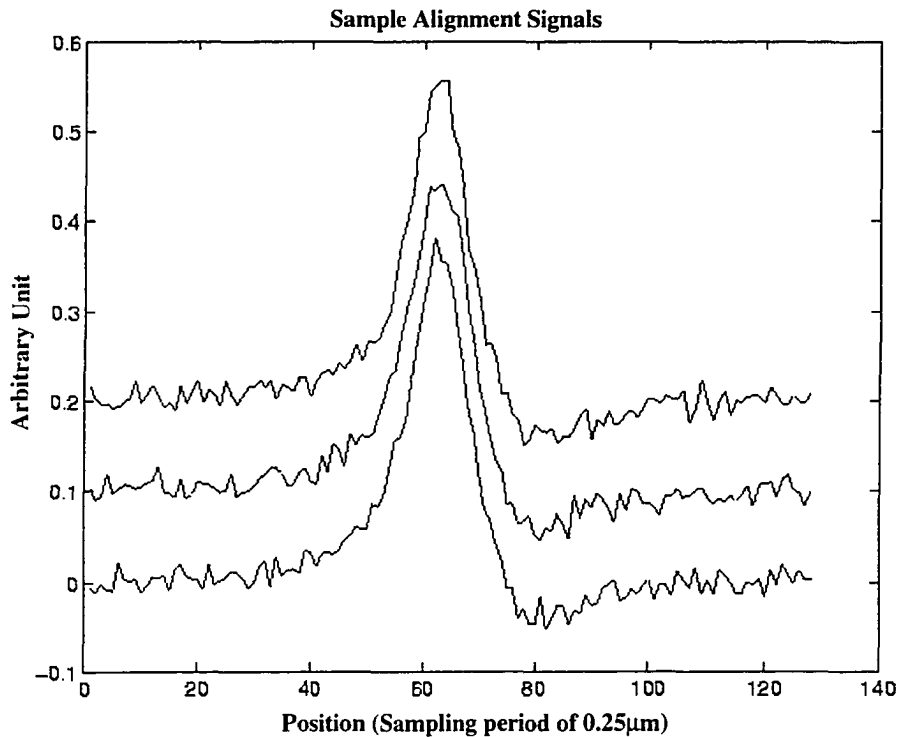
Figure 6-6. Sample alignment signals generated for simulation. The signals are shifted in the vertical direction for clarity. It is clear from this plot that the signals used in the simulation contain asymmetry.

The correlation algorithm is already described in the discussion proceeding to Equation 6-7. We saw earlier that the correlation algorithm is the algorithm of choice in the absence of process variation, i.e., all process parameters are fixed exactly at their desired values. A version of this basic correlation algorithm is implemented on a commercial overlay metrology instrument.

91
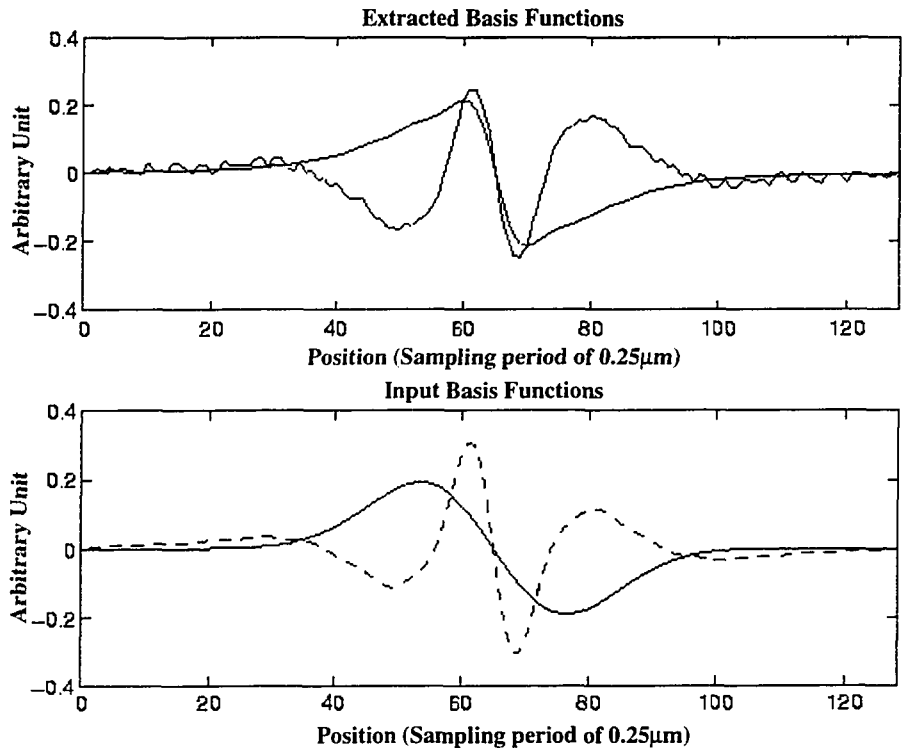
Figure 6-7. Extracted and input basis functions of the model space. The bottom panel plots the two functions used as bases for the linear space to generate alignment signals. The top panel plots the extracted basis function using the SAD process. These two functions are linear combinations of the inputs.

Together, these two algorithms should be representative of the state-of-the-art in alignment algorithms before SSD is available.

6.4.3 Simulation results and discussion

The alignment performances of the control algorithms and the SSD algorithm are summarized in Table 6-1.

92

Alignment Performance (nm)	SSD	Peak Detection	Template - Correlation
Error Mean	-0.43	510.2	50.1
Error Standard Deviation	7.17	76.3	47.3

Table 6-1. Comparison of the alignment performance of SSD, Peak Detection and Correlation algorithms on simulated alignment signals. Compared to the two control algorithms, SSD significantly reduces both the mean and variance of alignment errors, even on these asymmetric and varying signals.

From Table 6-1, we see that not only the mean, but also the standard deviation of the alignment performance is improved by using the SSD approach. The reason for the improvement on error mean is clear, but it is more subtle for the standard deviation. Equations 6-9 and 6-14 clearly state that, because of the fluctuation in the parameters, there is randomness in the signal in addition to the additive random noise. The SSD approach is able to characterize such randomness and eliminate its effects on alignment. For the other algorithms used for comparison, such randomness contributes directly to the variance of the alignment error.

It is seen from the simulation above that the SSD approach successfully built a model given some example alignment signals and their correct reference positions. The model can then be used to align signals coming from the same process. When compared with other algorithms currently employed in the industry, the SSD algorithm shows significant reductions in both the mean and the variance of the alignment errors.

The existence of a well-controlled environment so that Equation 6-9 holds true is imperative for the SSD approach to be viable. The definition of the controlled environment can be relaxed in some circumstances. For instance, there may exist such a situation where one can express the alignment signals from different mark positions on the wafer using Equation 6-9, only that the number of the parameters is increased by two

— the x and y positions of the mark. Then the same model can be used for all marks at different positions on the wafer.

The computer simulation presented here does not provide any information regarding the validity of the basic assumption of a well-controlled environment. Exactly to what extent this assumption can be justified can only be established by experimental studies.

6.5 Summary

The fundamental difficulty of alignment on asymmetric signals arises from the fact that an asymmetric signal may not contain sufficient information for us to determine the correct reference position.

An analogy with the generic communications problem led us to a generic approach to fully gather the information of an alignment system, resulting in a system representation by a mathematical mapping.

A physical model of the alignment problem helps to simplify the system representation and condense the information contained in it. It is found that, under certain assumptions, an alignment system can be represented by a linear space, which can then be used to perform alignment.

Computer simulation shows very promising performance for the SSD method. However, the validity of the fundamental assumption, namely, the existence of a well-controlled environment, cannot be established by simulations. Experimental studies are needed to ascertain the foundation of the new approach to alignment.

Chapter 7. Experiments with programmed asymmetry

In this chapter, we describe several experimental applications of the Subspace Decomposition method. The signals in these experiments contain deliberately introduced asymmetry and variations.

7.1 Introduction

In Chapter 6, we developed an algorithm-based solution to the problem of alignment on asymmetric and varying signals. Although the SSD method initiated from some very general considerations such as the analogy of alignment to communication, there were a few central assumptions that the derivation was based upon. In particular, it was assumed that there exists a well-controlled environment, i.e., processes fluctuate around some average conditions and these fluctuations should be small.

It was shown using a straightforward computer simulation that the SSD method is a very reliable algorithm for alignment, even when the signals contain varying amount of asymmetries. The alignment performance of SSD, as measured by the mean and variance of alignment errors, was exceptionally good compared with other algorithms currently employed in the industry.

These results, however, can only hold up if the central assumption is true. We need to provide experimental evidence to substantiate the viability of the new method. In this chapter, we will apply the SSD algorithm to two kinds of alignment problems, both of which contain programmed, i.e., deliberately introduced asymmetry.

7.2 Damaged marks

7.2.1 Mark fabrication

In the first experiment [49], the alignment mark is a 2µm wide and 1µm deep trench etched into a Silicon substrate. This is an ideal alignment mark because it involves the minimum amount of processing. The reference position is defined as the symmetry axis. This is shown in Figure 7-1.
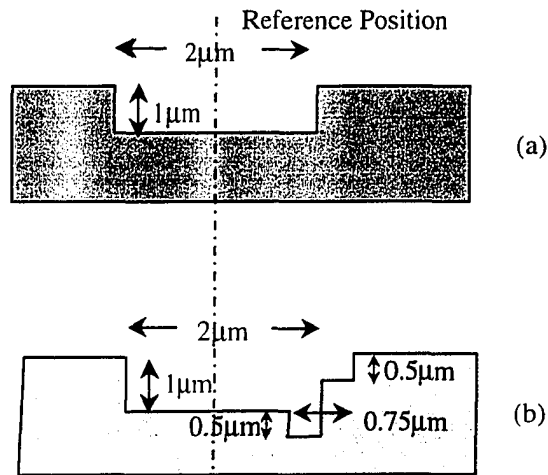
95

Figure 7-1. Cross-section of the alignment mark used in the experiment of damaged Si mark. (a) Ideal mark as designed, it is symmetric about the reference position. (b) The damaged mark using a second lithography step.

After the ideal alignment marks are made on the wafer, a second lithography step is used to introduce some short line segments that straddle one edge of the alignment marks. These line segments are $0.75\mu m$ wide and etched $0.5\mu m$ deep. They serve as deliberately introduced damage to the ideal alignment marks. Since these line segments are short, we have the undamaged and damaged alignment marks placed right next to each other. The process of making these testing structures is shown in cross-section in Figure 7-1.

Interestingly, the relative position of the distorting line segments and the underlying alignment marks varies from die to die because of the overlay error of the two lithographic steps. The stepper tool used for making these structures was an outdated machine that sometimes failed to align even on the ideal marks when making the second layer. Thus it relied on the inferior method of mechanical positioning when exposing the damage. The result is that for different die on a single wafer or on different wafers, the

96

form of damage to the ideal alignment mark can be significantly different. We will see evidence of this later.
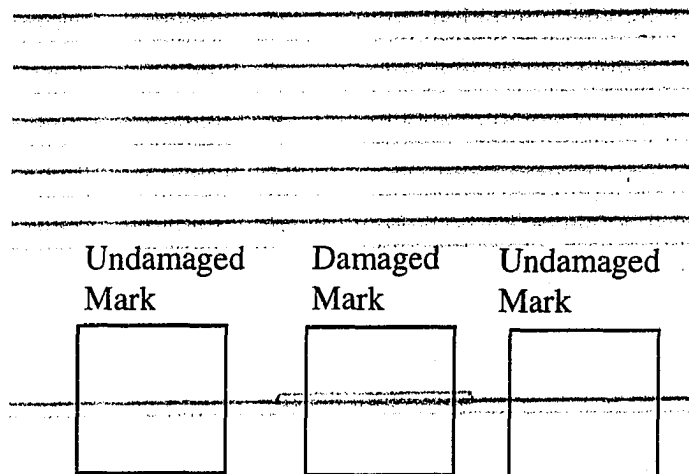

Figure 7-2. Microscope image of the alignment structures. The box marked "Damaged Mark" has a short line segment that straddles one edge of the ideal mark. The boxes marked "Undamaged Mark" show the ideal mark. The ideal mark has width of 2μm.

A 50x microscope was used as the alignment tool. Figure 7-2 shows a typical picture. Note that the image contains both the damaged and undamaged marks.

7.2.2 Image processing

First, residual rotation in each image is compensated. To do this, two sub-images of the undistorted mark, labeled L and R in Figure 7-3, are taken from the image. The gray scale value of each pixel in the two dimensional images of L and R can be written as

$$L = L(i, j) \text{ and } R = R(i, j), \quad\quad \text{Eq. 7-1}$$

where i and j are indices to the pixels, with i indexing the y direction and j the x direction. Note that the definition of these two axes is different from the usual manner in order for the notation to be consistent with previous chapters. See Figure 7-3.

97

One dimensional signals representing the ideal mark are obtained by projecting the images L and R to the x axis,

$$I_L(j) = \sum_i L(i,j) \text{ and } I_R(j) = \sum_R R(i,j) \qquad \text{Eq. 7-2}$$

These one-dimensional signals are processed using the peak-detection algorithm described in Section 7.4.2 and yield reference positions of the ideal marks in the sub-images L and R, x_0^L and x_0^R. If there is any residual rotation in the mark, they should be different from each other. We can define the residual rotation θ using

$$\tan(\theta) = \frac{x_0^R - x_0^L}{y_0^R - y_0^L}, \qquad \text{Eq. 7-3}$$

where y_0^L and y_0^R are the average y coordinates of the L and R sub-images, respectively, as shown in Figure 7-3.

After the residual rotation is determined, we can rotate the axes x and y into x' and y', with the x' axis parallel to, and y' perpendicular to the mark. As shown in Figure 7-3. Now the gray scale of the image pixels in the new coordinate system is given by,

$$I(x', y') = I[i(x', y'), j(x', y')], \qquad \text{Eq. 7-4}$$

where $i(x', y')$ and $j(x', y')$ transform the new x' and y' coordinates of the pixel to the indices (i, j).

Because the image is rotation-free in the new coordinate system, and each of the sub-images D, L and R is translationally invariant in the y' direction, we expect,

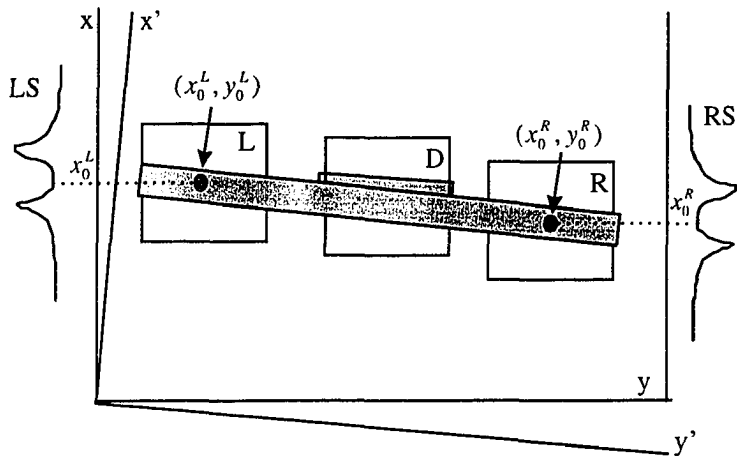$$I(x', y') = I(x'), \qquad \text{Eq. 7-5}$$

98

Figure 7-3. Routine to eliminate residual rotation. The alignment mark is shown as the shaded area. The images of undamaged mark, L and R are projected to the x-axis to form the 1D signals LS and RS. These signals are used to determine the centers of the mark positions, labeled as x_0^L and x_0^R, respectively. These can be used to determine the rotation. The axes are then rotated so that y' is parallel to the mark and x' is perpendicular to it. The mark images can be projected to the x' axis to obtain rotation-free one dimensional signals.

This equation gives us the one-dimensional signal associated with the damaged and undamaged marks.

7.2.3 Alignment results

After we take out the rotation in the picture, the ideal alignment mark is parallel to the y' axis. We can therefore use the undamaged mark in the same picture to determine the reference position of the damaged mark. This is done using the peak-detection algorithm described in Section 7.2.4. In order to evaluate the reliability of the reference position obtained using this method, we took many sub-images of the undamaged mark and calculated their corresponding reference positions. We found a standard deviation of 5.7nm. This small variation of the reference positions not only attests to the stability of the peak-detection algorithm, it also confirms that we have successfully eliminated

99

rotation from the image in the procedure discussed above, because otherwise, the reference position would vary strongly from one side of the image to another.
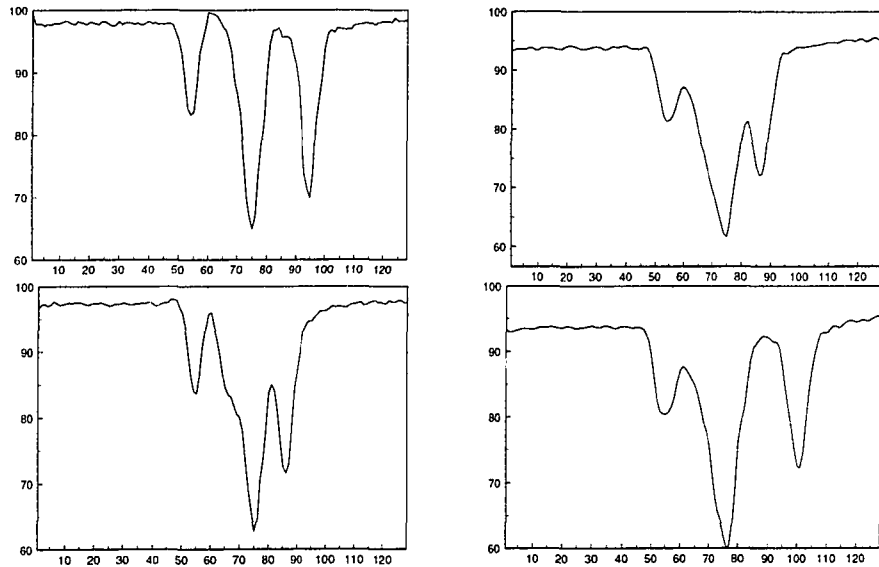

Figure 7-4. Sample alignment signals from the damaged Si marks. Unit for the x-axis is the sampling period of 0.1μm. The y-axis is the measured gray scale intensity; it has arbitrary unit. Apparently, the signals here are all asymmetric. There also exists a significant amount of variation in the signals.

Some typical one dimensional signals of the damaged marks are shown in Figure 7-4. The amount of asymmetry is obvious in each of the signals. Also, there exists strong variation from one signal to another. This variation is due to the overlay errors between the two lithographic steps used to make the mark, as explained earlier. The combination of the asymmetry and variation makes this an extremely difficult situation for alignment.

For comparison, we have again used the two algorithms described in Section 7.2.4. The algorithm of peak-detection was slightly modified because of the existence of two inverse peaks, or troughs, in the signal of the undamaged mark. The algorithm was

100

changed so that it first finds the position of both troughs and then takes the average as the reference position of the signal. The mean-correlation algorithm was unchanged, i.e., the average signal obtained from the training set is used as the correlation template.

Table 7-1 compares the alignment performance of each of the three algorithms [49]. These results are acquired using 143 signals from 19 different die on a single wafer. Among the 143 signals, 113 were split into the training set, which is used to train both the SSD and the correlation algorithm; the remaining signals were used to test alignment.

Alignment Performance	Mean (nm)	Standard Deviation 1σ (nm)	Enabling Technology (μm)
Peak Detection	730.8	224.5	4.5
Mean Correlation	19.7	75.1	0.8
SSD	2.1	13.9	0.13

Table 7-1. Comparison of alignment performances of the SSD, Peak Detection and Correlation algorithms on damaged Si marks. In the last column, the minimum feature size of lithography that is enabled by the alignment performance is shown. This is based on the rule of thumb that the overlay budget, or mean plus three sigma, is roughly one third of the minimum feature size.

The peak-detection algorithm basically fails in this difficult situation. The reason is that the algorithm is designed to look for two troughs in the signal, based on the design of the ideal mark. When the mark is damaged, the signal can contain more than two troughs, as shown in Figure 7-4. This confuses the algorithm. The correlation algorithm performs much better than peak-detection, but the overlay, or the mean plus three sigma value of overlay error is found to be over 200nm. Assume this is the alignment performance we can get on a critical layer, the smallest features we can make using this lithographic technology is 0.8µm. On the other hand, the overlay of SSD algorithm is smaller than 50nm, proving that even if the alignment marks for the critical layer are damaged as badly as Figure 7-2 and 7-4 show, the SSD approach can still enable 0.13µm generation of lithography. The improvement is dramatic.

7.3 Asymmetrically coated marks

7.3.1 Mark fabrication

In this experiment [49], the ideal marks are still 2μm wide and 1μm deep trenches made in a Silicon substrate. After these ideal marks are made, the wafers are then coated with 0.5μm thick Al, using an electron-beam evaporator as the Aluminum source. When depositing the Al layer, the wafers are tilted 40°, so that the metal coverage of the alignment marks is uneven. This is illustrated in Figure 7-5.

A second lithographic step is used to remove the metal on some part of the wafer, revealing the underlying symmetric, ideal alignment marks. Since this step only requires coarse overlay, it can be performed by aligning to the asymmetrically coated marks. See
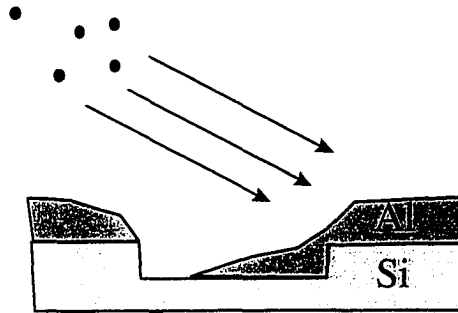
Figure 7-5. Illustration of the tilted metal coating process. When the Al atom flux is tilted with respect to the surface normal, the resultant coating is asymmetric.

Figure 7-6 for an microscope image of the metal-covered and uncovered marks.

In this case, strong variation of mark structure again exists because of the different impinging angles of Al deposition at different locations on the wafer.

7.3.2 Alignment results

Similar to the experiment with damaged marks, the signals are obtained from the images such as the one shown in Figure 7-6. The image processing steps are identical.

102

The reference positions of the covered marks are found by detecting the position of the uncovered marks.
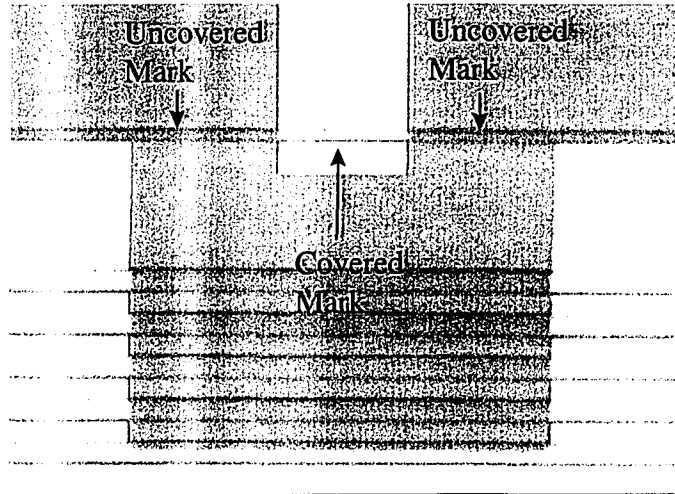
Figure 7-6. Microscope image of the metal coated alignment marks. Al layer is partially removed to reveal the uncovered, ideal alignment mark.

Figure 7-7 shows some sample signals. Evidently, the alignment signals are strongly asymmetric due to the tilted metal coating. The existence of variations in the signals is also apparent.

The alignment results for the control algorithms and the SSD approach are shown in Table 7-2. These are the results of processing 153 signals from twenty die on two different wafers. In this case, 114 signals were used for training and the rest for alignment test.

103
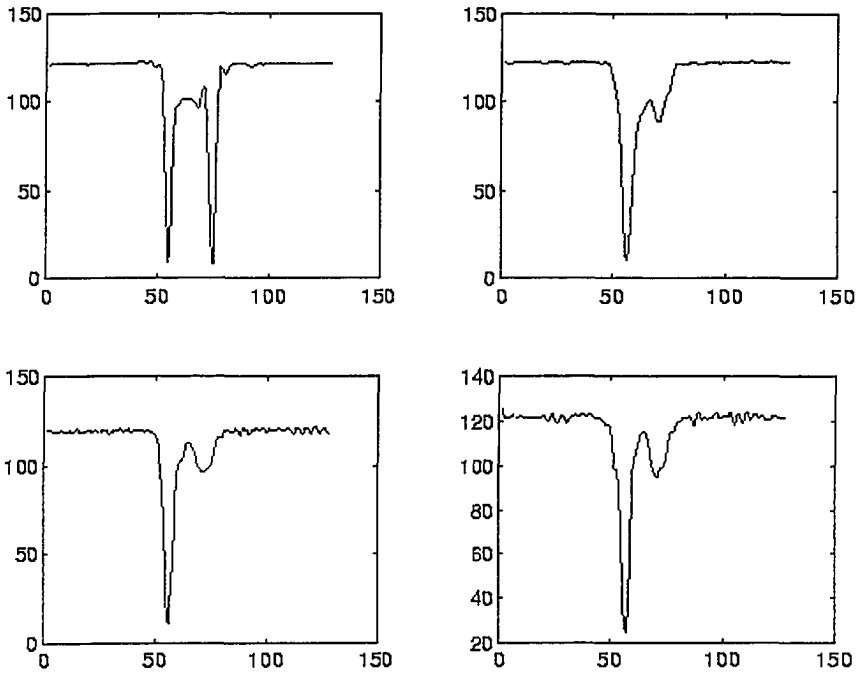
Figure 7-7. Sample alignment signals from the metal-coated marks. Unit for the x-axis is the sampling period of 0.1μm. The y-axis is the gray scale intensity of the signal and it has arbitrary unit. The asymmetry and variations in the signals are apparent.

Alignment Performance	Mean (nm)	Standard Deviation 1σ (nm)	Enabling Technology (μm)
Peak Detection	-28.9	117.7	1.2
Correlation	15.4	45.8	0.5
SSD	-0.9	13.7	0.13

Table 7-2. Comparison of alignment performances of the SSD, Peak Detection and Correlation algorithms on metal coated marks. In the last column, the minimum feature size of lithography that is enabled by the alignment performance is shown. This is based on the rule of thumb that the overlay budget, or mean plus three sigma, is roughly one third of the minimum feature size.

104

7.4 Discussion

The performance of the peak-detection algorithm in the metal-covered mark experiment is somewhat better than the damaged mark case. This is due to the fact that signals from the metal-covered marks do have two troughs, as expected by the peak-detection algorithm. By comparing the results in Tables 7-1 and 7-2, we can conclude that the peak-detection algorithm is overly dependent on signal details, such as the number of peaks or troughs. Such algorithms completely fail if these detailed assumptions on the signal form are not satisfied. This serious flaw is not limited to our version of the peak-detection algorithm, it is quite common among the widely used and discussed algorithms [4].

The correlation algorithm performs better than peak-detection in both experiments with programmed asymmetry. We expect this observation to hold true in general. As explained in Chapter 6, the correlation algorithm does a very simple and primitive form of learning and it is optimal when the alignment system does not have any variations in it. This is the reason why the mean errors obtained using the correlation methods are smaller than those of symmetric algorithms represented by peak-detection. When signal variations are significant, however, the correlation algorithm gives large error variances, as demonstrated by the two examples here.

The performance of SSD is consistently impressive. The most important feature of Tables 7-1 and 7-2 is the simultaneously reduction of mean and variance of alignment errors when SSD is used. This can be understood on the basis of SSD's usage of a sophisticated learning approach that not only characterizes the asymmetry, but also the variations in alignment signals.

Note that in both experiments, we have used the so-called differential method to obtain the correct reference positions. As explained in Section 6.2.1, this method suffers from several problems.

105

The first one is the subtle issues of cross-talks and rotations, as detailed in Section 5.3; this is well accounted for in the process of image processing. See Section 7.2.2. Another problem with the differential method is the possible existence of tool-induced effects. Because we relied on the tool's detection of the undistorted marks to obtain the correct reference positions, even the SSD alignment procedure has no chance of characterizing or eliminating these tool-induced errors, because there is imply no such information in the training set.

In order to minimize this problem, we were careful to place the mark in the same position in the optical field when we took the images. Most of the optical aberrations that can lead to tool-induced errors are functions of the mark's distance to the optical axis. Thus by placing the marks at the same position in the field, the tool-induced errors, if they exist, are made to be approximately constant.

The alignment performances in Tables 7-1 and 7-2 are measured relative to the reference positions of the undistorted marks; if the latter err by a constant amount, the alignment results presented here also err by the same amount. If this constant tool-induced error can be somehow measured, all alignment results can be corrected and the conclusions regarding the performance of different algorithms can remain unchanged.

7.5 Summary

In two experiments, we programmed asymmetry and variation into alignment marks. The differential method was used to find the correct reference positions. The results show that the Subspace Decomposition method dramatically reduces both the mean and variance of alignment errors, compared to a symmetric algorithm and a correlation algorithm.

The convincing performances of the SSD algorithm in these very difficult settings prove that the hypothesis of a well-controlled environment does hold true. Since the asymmetry and variation we introduced in our experiments are substantially larger than what are commonly observed in manufacturing processes, we challenged SSD with two very difficult situations for alignment. It can be argued that if the linear space model is valid in these rather extreme circumstances, it should also be justified in situations where

106

less asymmetry and variation are encountered, such as those in real semiconductor manufacturing.

We will carry out one more experiment to test the SSD algorithm for wafers that have gone through processing steps in a manufacturing environment.

Chapter 8. Experiment with CMP wafers

In this chapter, we apply the Subspace Decomposition method to the alignment of chemical-mechanically polished wafers.

8.1 Introduction

Chemical-Mechanical Polishing (CMP) is needed largely because of the issue of depth of focus in lithography [3]. The usable depth of focus actually shrinks faster than the minimum feature size. Due to various processing steps, the wafer surface can become rough on a microscopic scale, which makes simultaneous focus on different parts of the wafer very difficult. CMP can planarize the wafer surface, and thus increasing the process latitude when trying to focus on the wafer. In addition, the recent breakthrough of Cu interconnect technology also relies on the availability of CMP to complete the special damascene processes [10].

But CMP, in particular Tungsten CMP, is known to wreak havoc with alignment [50].

The damages inflicted on the alignment marks by the CMP process are the combined effects of abrasive mechanical motion and delicate chemical reactions. The mechanics of CMP and how it causes asymmetry in alignment marks are laid out in Section 2.2.1 and Figure 2-8.

To make the problem worse, CMP processes are also unstable. The wearing of the pad, the evaporation and dilution of slurry chemicals, etc., can all contribute to significant changes of CMP conditions over time.

Alignment on polished wafers is a perfect example of how signal variations exacerbate the problem of signal asymmetry, and make it intractable. Because engineers and researchers are accustomed to the notion of a systematic trend in the alignment errors, there have been attempts to compensate such systematic alignment errors caused by CMP. But unfortunately, polished wafers do not show reliable systematic trends in their overlay errors [50].

108

The predicament of randomly varying alignment errors is beyond the alignment capability of most state-of-the-art lithography tools, but it is exactly where the subspace decomposition method can make the most impact. We therefore apply the method of SSD to the problem of aligning polished wafers.

Also, since the experiments in Chapter 7 involve only programmed asymmetry and variation, we would like to use the CMP experiment to test the viability of SSD in a real manufacturing environment.

8.2 Experimental procedures

8.2.1 Wafer fabrication

In this experiment, 19 wafers are processed in a state-of-the-art $0.25\mu m$ manufacturing facility. All processing is done on state-of-the-art commercial tools. Processing conditions are identical to those used for a certain device that is being made in the plant. To protect the proprietary information involve, we only give a qualitative description as to how the wafers are processed.

Aluminum ⟶
W- Plug & CMP ⟶
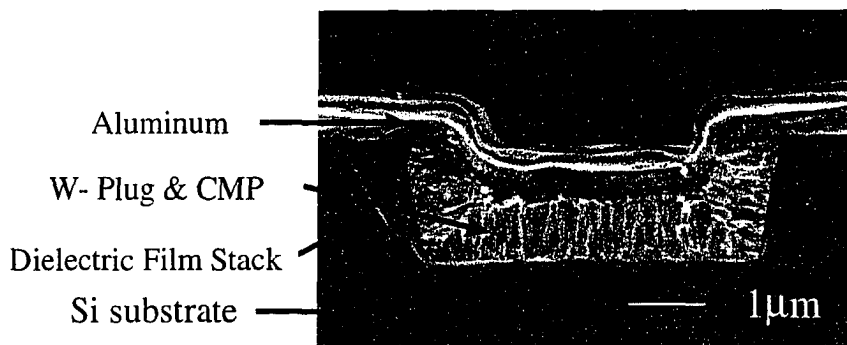Dielectric Film Stack ⟶
Si substrate ⟶

Figure 8-1. Scanning electron microscope image of the cross-section of a typical alignment mark. This picture illustrates the processes that have been applied to the marks. After these marks are etched in dielectric films, W is deposited and then chemical-mechanically polished. Aluminum is then coated on top.

A scanning electron microscope image of a cross-section of the alignment mark is shown in Figure 8-1 [51]. In the first step, a stack of dielectric layers is deposited on the

109

wafer. A lithography step is used to define the alignment marks, together with some overlay metrology marks and other circuit patterns. After exposure and development, the marks are etched into the dielectric film stack. Next, a blanket layer of Tungsten is sputtered on the wafer, similar to the so-called Tungsten-plug process. In this step, etched holes with small dimensions will be completely filled, thus the name "plug". But since the alignment mark is relatively wide compared with the thickness of the W film, the relief structure of the mark is replicated to the surface of Tungsten film. The film is then chemical-mechanically polished, removing the material on top of the wafer surface, but leaving behind some Tungsten <u>inside</u> the mark trench, as can be seen in Figure 8-1. A blanket Al layer is then deposited on top, similar to the first step of making metal connections.

Although it is apparent that the Al coating is asymmetric on the two side walls of the mark, it is not obvious what effects the CMP process has on the mark. The absence of apparent CMP damage on the alignment marks even under close inspections with tools such as the scanning electron microscope is a very common observation, and it contributes to the difficulty of understanding and improving alignment on polished wafers.

The wafers are then coated with resist. A commercial stepper is used to align to the alignment marks buried under the Al coating. For eight different die on a wafer, the stepper is instructed to record the alignment signals; and also the estimated corresponding reference positions as given by the current alignment algorithm implemented on the stepper. Without going into the details of the algorithm, we mention that it is similar to the correlation-based approach we discussed earlier, but with a very important difference. The correlation algorithm as we have defined in Equation 7-7 uses training signals to obtain the correlation template, but the algorithm implemented by the stepper uses a fixed template function, which is designed according to the expected signal shape. Also to protect proprietary information, sample alignment signals are not shown here. Assuming these positions found by the existing algorithm are the correct reference positions of the marks, the stepper aligns and exposes the wafers.
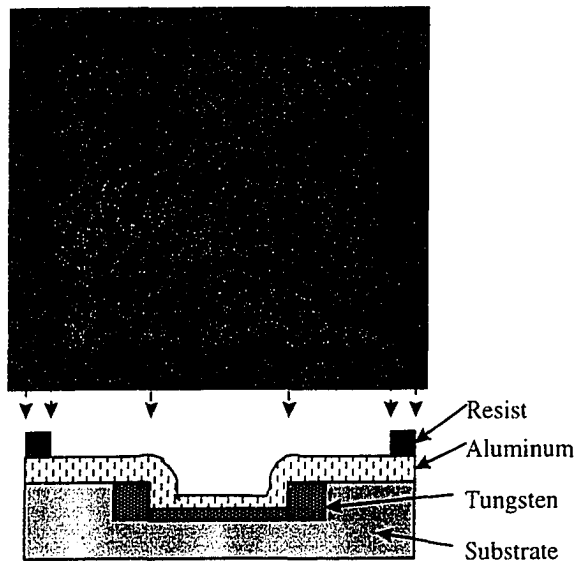
110

Figure 8-2. Image of overlay metrology mark after development. The sketch below the image associates the features in the image with topographic structures in the cross-section. The outer mark is an opening in the resist coating. The inner mark is an Al-covered step in the residual W left inside the mark by the CMP process. The edge definition of the Al-coated step is rugged as indicated by the "cloudiness" in the picture. This is not a very reliable mark for metrology.

8.2.2 Overlay metrology

In this experiment we rely on overlay metrology to determine the correct reference position for each alignment mark. The procedure is detailed in Section 7.2.1. After the exposure, images of the overlay metrology marks are taken and analyzed. One example is shown in Figure 8-2. The outer mark is the opening in the resist and the inner mark is the Al-covered step formed by residual W; see Figure 8-2 for an illustration. For reasons that are not completely clear, the edge definition of the inner mark is rugged and cloudy. Overlay results obtained from these mark images cannot be reliable. Indeed, CMP layers are reported to cause overlay metrology errors [24]. It is also reported that more reliable overlay metrology can be performed after the removal of Al-coating [24]. Following

111

these results, we etched away the Al on the wafers. In the etching step, we deliberately increased the etching time so that the W inside the marks can be at least partially removed.
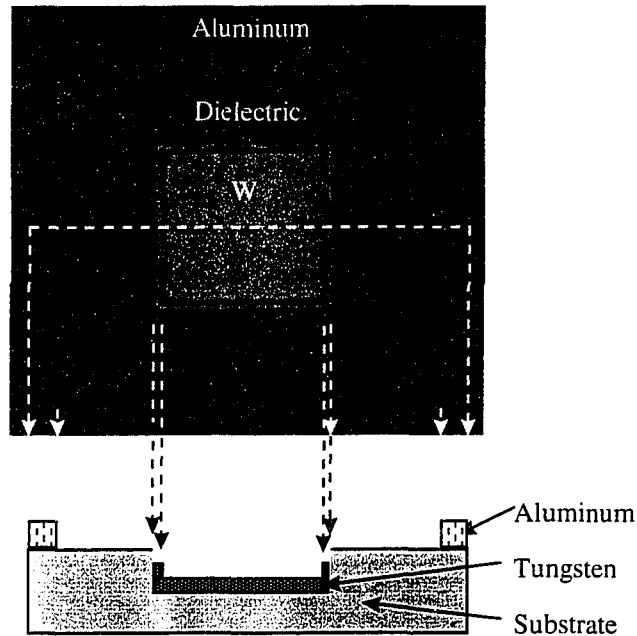

Figure 8-3. Image of overlay metrology mark after metal etch and clean. The sketch below the image identifies the features in the image with topographic structures in the cross-section. The outer mark is a step in Al coating. The inner mark is double edged, with one edge corresponding to the side wall of the trench etched in the substrate; and the other edge a step in the residual W left inside the trench by the CMP process. Note that the W is partially removed in the metal over-etch step. This mark gives more reliable metrology results than that in Figure 8-2.

An optical image of an overlay metrology mark after the metal etch step and the removal of resist is shown in Figure 8-3. Here the outer mark is a step in Al. The inner mark is a double-edged feature. One of the edge is the step down from the dielectric to the remaining W, and the other edges are the steps in W, as shown in Figure 8-3. The overlay mark shown in Figure 8-3 has a much better edge definition. Therefore reliable

112

metrology results can be obtained. We observed a one sigma repeatability of 2nm when measuring the overlay errors.

After the overlay metrology is reliably performed, the procedure outlined in Section 7.2.1 is followed to find the correct reference position corresponding to each alignment signal. Alignment performances presented later are all measured relative to this definition of true reference position. For this experiment, 285 signals were measured from 19 wafers, processed using 4 different polisher/recipe combination.

8.3 Results and Discussions

Alignment Performance	Mean (nm)	Standard Deviation 1σ (nm)	Enabling Technology (μm)
Stepper	36.3	54.4	0.25
SSD	1.4	14.0	0.07

Table 8-1. Comparison of alignment performances of a commercial stepper and the SSD algorithm on chemical-mechanically polished wafers. The last column shows the minimum feature size allowed by the corresponding overlay budget, which is defined as mean plus three sigma. Note that WCMP are used on metal connection layers, the overlay requirement is less stringent than that of the critical layer.

Table 8-1 summarizes the alignment performance of the stepper algorithm, and the expected performance if SSD were used. The overlay, i.e., the mean plus three sigma value of the overlay errors, given by the commercial stepper is nearly 200nm, not too much smaller than the minimum feature size of 0.25μm on the devices that are being made using the same processes. Because this is the alignment on a metal layer, the overlay budget is usually much less stringent than the critical layer. Such an overlay performance is either problematic or marginally satisfactory, depending on the specific overlay requirement for the device.
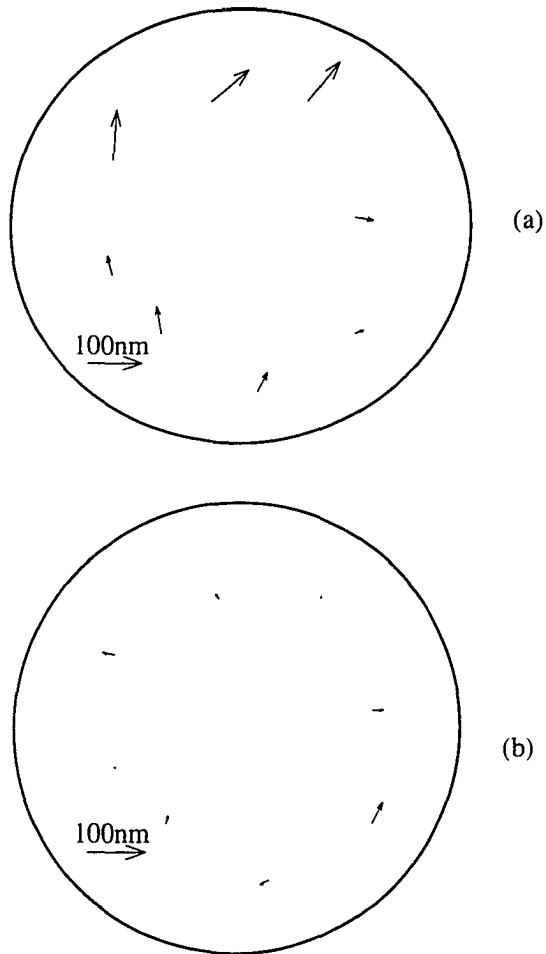
113

Figure 8-4. Overlay vector maps from a commercial stepper and
the SSD method. (a) Overlay vectors across eight die on a wafer,
as measured by metrology. This represents the alignment errors
using current algorithms implemented on the stepper. (b)
Predicted overlay vectors for the same wafer if SSD were used.

On the other hand, the performance of the SSD algorithm can allow the 0.07μm

generation of lithography, considering that we are aligning to a metal layer. The CMP

process studied is known to be notoriously difficult to align, yet the SSD method can still

deliver overlay performance below 50nm. Because the processing used in this

experiment is identical to actual technologies used for real devices, these results

114

convincingly demonstrate the viability of SSD as a practical tool to improve alignment errors.

One interesting effect is how local alignment considered here can lead to problems and improvements for global alignment. Figure 8-4(a) shows the alignment errors across one of the wafers in our samples. There exists a rotational pattern, which results in a rotation error in the procedure of global alignment. This should be clear from the description given in Section 1.2.4. Many other researchers have observed the same phenomenon that, among the six parameters of global alignment, rotation error is most likely the problematic one when CMP is used [24, 52]. Figure 8-4(b) shows the corresponding pattern of alignment errors on the same wafer if SSD were implemented on the stepper. Clearly, the magnitude of the error is much smaller. Also, the rotation pattern is absent. Hence an error-free global alignment can be performed.

As in the case of differential alignment, the overlay metrology method of determining the correct reference position is also prone to errors. In particular, overlay metrology measures overlay error, which is not necessarily equal to the alignment error. Other effects including stage positioning error, wafer and field distortions, etc., can all contribute to overlay errors. These effects are enumerated in Section 1.2,

In our experiment, both layers between which we are measuring overlay errors were made on the same stepper. Thus the effect of field distortion essentially cancels out, i.e., a rather ideal case of the so-called "stepper matching" technique discussed in Section 1.2.3.

Stage positioning inaccuracy should have a minimum impact on our conclusion, because when exposing the experimental die, the stage was essentially sitting idle at the position of exposure. It is widely believed [11] that most of the stage positioning errors occur when the stage is moving relatively fast. To keep the throughput high, exposures are often done before the stage can settle down to the desired position and recover from the fast motion involved in stepping. This impatience results in a considerable amount of stage positioning error. When the stage is held idle at the desired position, stage positioning errors are believed to be less than 5-10nm [53].

115

Lastly, since we performed essentially local alignment, only the intra-field term of the wafer distortion contributes to our alignment results. As explained in Section 1.2.4, inter-field wafer distortion results in overlay error only when global alignment is used. Similar to the discussion of tool-induced effects in Section 7.4, a constant amount of intra-field wafer distortion does not compromise the results shown in Table 8-1. What can be problematic is the variation of intra-field wafer distortions from die to die. This effect can be expected to be substantially smaller than the three sigma value of 20nm often associated with total wafer distortion.

Thus the combined effect of stage positioning, wafer and field distortion is not expected to contribute more than 20nm to the three sigma of reference positions. This is well within the observed three sigma of nearly 50nm in the performance of SSD. Therefore, the conclusions drawn above regarding the performance of SSD on CMP wafers remain unchallenged even in the presence of the above complications in interpreting measured overlay errors.

8.4 Summary

The process of CMP is needed for today's lithography. We expect that the importance of CMP will be increasingly emphasized when the depth of focus of lithography keeps on shrinking. Therefore, it is of extreme importance to solve the problem of alignment on CMP wafers.

The intrinsic asymmetry and variations have made alignment on polished marks a very difficult task, in fact, the most difficult one according to many lithography engineers.

The experimental study described in this chapter confirms that SSD is capable of achieving overlay of less than 50nm on Tungsten CMP wafers, proving the power and versatility of the SSD method.

In addition, results here also support the central assumption for the applicability of SSD, namely, the existence of a linear space model. If this assumption holds true in the case of the notoriously unstable process of CMP, it should be reasonable to extrapolate

116

from these results that SSD can perform equally well for less demanding problems, such as tool-induced effects, etc.

117

Chapter 9. General properties of algorithm-based techniques

Using the Subspace Decomposition method presented in previous chapters as an example, we summarize the general properties of algorithm-based solutions.

9.1 SSD as an algorithm-based solution

In designing the SSD approach, we started with a very general description of the alignment problem by comparing it to a communication channel. In this analogy, all the coatings and processing on the alignment marks, together with the physics of the alignment sensors, are summarily included in the general notion of "channel characteristics". An empirical scenario is used to characterize an alignment system. It treats the alignment system as a black box and learns about its behavior by examining the inputs and outputs. The derivation that led to the SSD solution was basically the one of representing, simplifying, deciphering and utilizing the all-inclusive "channel characteristics" of an alignment system. This procedure does not make use of any detailed information regarding the underlying physical processes.

If we return to Figure 2-5, which shows the block diagram of a generic alignment system, the SSD method is a prototypical algorithm-based solution to the alignment problem.

9.2 General advantages

Algorithm-based solutions, as exemplified by the SSD method, are remedies for many of the shortcomings of physics-based approaches.

First, physics-based solutions usually require costly hardware implementations, but algorithmic solutions can be implemented purely in software. Both the learning and aligning algorithms given in Chapter 6 can be realized using standard digital signal processing techniques. Software implementations have shorter turn-around times and they are easier to test and less costly to implement.

118

Second, physics-based solutions are problem-specific and hence inflexible, but algorithm solutions are generally applicable and versatile. The learning step in the SSD solution finds the type and amount of asymmetry and variation in the signals, and prepares the alignment step for the idiosyncrasies of the targeted alignment system. The use of such a tuning procedure makes the approach of SSD very versatile. Algorithm-based approaches such as SSD should be able to achieve accurate and precise alignment, regardless of the sources of the asymmetry or variation. Hence, they are versatile and flexible. This conclusion is firmly supported by the consistent performances of SSD in the three experiments presented in this thesis. Clearly, the three experiments contain very different causes of alignment difficulties, yet the algorithm was applied without many modifications.

Furthermore, algorithmic methods such as SSD can be said to be adaptive, because, when new processes are introduced and new types of asymmetry difficulties occur, it will automatically fix the problem, and we may never know the existence of the new alignment problem.

Finally, we also find that SSD can help to improve the performance of global alignment in the presence of wafer distortions. [54]

9.3 General limitations

Similarly, in the areas where the physics-based solutions are advantageous, the algorithm-based solutions suffer from some limitations.

While physics-based solutions can help find the underlying causes of alignment problems, algorithmic approaches mask the real culprits of mark damage and distortion. Sometimes, finding the causes of alignment problems can lead to better process and tool design that not only eliminates alignment, but also fixes other associated problems. One example is again related to CMP. The overlay problems caused by CMP have prompted the manufacturers of CMP tools to emphasize stability of the polishing behavior. Apparently, these steps not only alleviate overlay difficulties, but also benefit the uniformity and controllability of the polishing process. Vendors of CMP tools may not have been as motivated if algorithmic approaches such as SSD were used to align the

119

polished wafers. It is important to realize that SSD does learn about the process in the model-building step, but it understands the physics in an abstract and mathematical manner that is largely incomprehensible to us.

Furthermore, physics-based solutions can completely eliminate or at least reduce the amount of asymmetry and variation in the signals. Even though algorithmic techniques can align relatively reliably on imperfect signals, asymmetry- and variation-free signals are still preferred. In the language of the communication problem, although we can have very sophisticated and capable decoding algorithms, we still prefer a distortion-free communication channel.

The advantage of having symmetric and stable signals is underscored by the possibility that algorithmic approaches may require frequent training to function satisfactorily. The procedure of training can be burdensome and expensive in a manufacturing environment, especially when process tuning, maintenance adjustment and tool upgrades are frequent, because these necessitate re-training of the model. Modern and future semiconductor plants will likely emphasize rapid re-tooling and re-configuration to increase technology and business flexibility. This trend will tremendously increase the number of possible combinations, and the frequency of change, of processing parameters, resulting in an alignment system that can only be described by very large models, or very frequent retraining if the model is to remain simple. A possible solution is to find ways to carry out training on a real-time basis. If this can be accomplished, learning based alignment techniques will be more practical and attractive.

9.4 An ideal solution?

Sections 9.2 and 9.3 bring about the impression that physics- and algorithm-based solutions are complementary, in the sense that they fix each other's problems. We therefore speculate that a better solution can be found if we combine both approaches. We use physics-based techniques to find and mitigate the sources of alignment errors. For the residual asymmetry and variation in the signals, we can resort to algorithmic solutions.

In fact, physics and algorithms can be combined in much more profound ways.

120

Recall that, while deriving SSD, we found a very general approach of solving alignment problems, namely, the procedure of characterizing, simplifying and deciphering the channel characteristics of an alignment system. Interestingly, this process can be made much more efficient if it is guided by in-depth physical understandings of the system itself.

For example, the method of subspace decomposition was based on the rather general physical model presented in Section 7.2.3, but we do not have to rely on such general models. If we know the details of the physics of a particular problem, a precise and concrete physical model can be built to replace the generic one in Section 7.2.3. This should lead to more efficient algorithm design. The generic subspace decomposition algorithm is therefore only needed when we do not possess any knowledge of the underlying physics, or when the physics is too complicated to be useful.

A perfect example is CMP. Being a relatively young and immature technology. CMP's detailed physics or chemistry is still not very well understood. Confined by our lack of knowledge, a generic algorithm such as SSD is the method of choice. But as our understanding of the CMP process deepens, we may be able to design tailor-made algorithms that are more efficient in solving CMP related problems.

By more efficient algorithms, we mean that the algorithm contains more *a priori* information. In essence, what we have just described is a method of incorporating more *a priori* knowledge into the algorithm. <u>The more knowledge of the process we build into the algorithm, the less training is necessary</u>.

Let's provide an example of how this can be done. Suppose an alignment tool has pin-cushion distortion caused by aberrated optics (Figure9-1). When the alignment mark is asymmetrically placed in the optical field, its image will be asymmetric. This can be illustrated by imagining the upper left quadrant as the alignment mark. It is a perfectly symmetric square in the object plane. Its image is the irregular shape shown in the upper-left quadrant in the image plane. This image is asymmetric. Suppose that after a careful study, we understand the reason for signal asymmetry is the pin-cushion field distortion of the alignment optics. We can then design an algorithm to first restore the image to a

121

pin-cushion-distortion-free version, which can be processed by subsequent algorithms to find the correct alignment. .

This idealized example demonstrates the relationship between physics and algorithms. Most importantly, we would like to emphasize that, to the extent we understand the physics of the underlying processes, algorithms do not have to be burdened with the task of learning the associated effects on alignment.
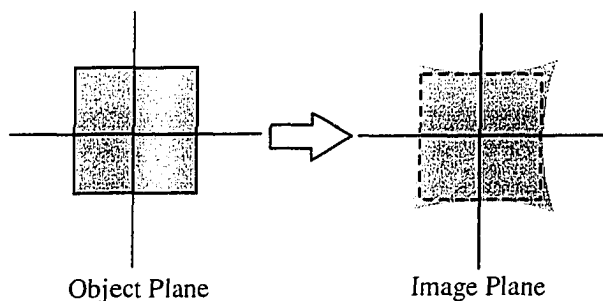

Object Plane   Image Plane

Figure 9-1. Alignment optics with pin-cushion distortion.
The dashed line in the image plane outlines the ideal
geometric image without the distortion. Assume that
the upper-left quadrant is actually an square alignment mark.
It is symmetric about its centers. It is imaged into a shape
shown in the upper left corner in the image plane. The image
is not symmetric about any axis.

An ideal solution to the alignment problem is therefore the following. We carry out an in-depth study of the causes of alignment errors. The knowledge gained during this study is then incorporated in the design of algorithms. Because the knowledge of the process is already built in the algorithm itself, a minimum amount of training is needed. Not only does this approach retain all advantages of algorithmic techniques, it also removes their serious limitations.

122

References

1. G. E. Moore, Proc IEEE 1976, 64(6), p307

2. R. C. Jaeger, Introduction to Microelectronic Fabrication, Addison-Wesley Publishing Company, 1988.

3. L. F. Thompson, C. G. Willson and M. J. Bowden, Introduction to Microlithography, American Chemical Society Professional Reference Book, 1994

4. G. Gallatin, Alignment Signal Generation and Processing, Short Course Notes, SPIE Symposium on Microlithography, 1994.

5. M. Born and E. Wolf, Principles of Optics, Pergamon Press, 1965.

6. A. A. Ghazanfarian, R. F. W. Pease, X. Chen and M. A. McCord; J. Vac. Sci. Technol. B15(6), p2146, 1996

7. A. F. Plambeck, Microlithography World, Vol. 5, No. 1, p17, 1996

8. Semiconductor Industry Association, The National Technology Roadmap for Semiconductors, 1997.

9. J. Mendonca, K.Murella, I. Him, J. Schlueter and C. Karlsrud; Thin Solid Films, Vol. 320, No. 1, p103, 1995

10. M. G. Lee, T. Larson, S. Beilin; Proc. Thirteenth International VLSI Multilevel Interconnection Conferences (VMIC), p395, 1995

11. E. Kouno, Y. Tanaka, J. Iwata and Y. Tasaki; NEC Research and Development, No. 90, p70, 1996

12. J. Ye; Ph.D. Thesis, Department of Electrical Engineering, Stanford University, 1995

13. G. Rivera, P. Canestrari; Proc SPIE, Vol. 1927, p806, 1994

14. C. S. Lee, J. S. Kim, I. B. Hur, Y. M. Ham, S. H. Choi, Y.S. Seo and S. M. Ashkenaz, Proc. SPIE, Vol 2197, p2, 1995

15. C. K. van Peski; Solid Stage Technology, Vol. 25, No. 5, p111, 1994

16. Mircea Dusa of National Semiconductor Inc., private communications.

17. J. D. Jackson, Classical Electrodynamics, John Wiley & Sons, Inc., 1975

18. E. T. Whittaker, The Theory of Optical Instruments, Cambridge University Press, 1915.

19. J. W. Goodman, Introduction to Fourier Optics, The McGraw-Hill Companies, Inc., 1996.

20. J. W. Goodman, Statistical Optics, John Wiley & Sons, 1985

21. G. Bouwhuis and S. Wittekoek, IEEE Trans. Electron Devices, Vol. 26, p715, 1979

22. T. Higashiki, Jpn. J. Appl. Phys., Vol. 29, 2568, 1990

23. A. V. Oppenheim and R. W. Schafer, Discrete-Time Signal Processing, Prentice Hall, 1989.

24. A. Plambeck, N. Knoll and P. Lord; Integrated Circuit metrology, Inspection and Process Control IX, Proc. SPIE, Vol. 2439, p298, 1995

25. T. Sato and H. Nomura, Jpn. J. Appl. Phys., Vol. 37, No. 6A, p3553, 1998

26. T. Saito H. Watanabe and Y. Okuda, Proc. SPIE, Vol. 3051, p686, 1997

27. N. Bobroff and A. Rosenbluth, J. Vac. Sci. Technol. B6, 403, 1988

28. K. Ota, N. Magome and K. Nishi; Proc. SPIE, Vol. 1463, p304, 1991

29. X. Chen and R. F. W. Pease, J. Vac. Sci. Technol. B14, No. 6, p3980, 1996.

30. S. Wittekoek, J. van der Werf and R. A. George, Proc. SPIE, Vol. 538, p24, 1985

31. Courtesy of A. Starikov of Ultratech Stepper.

32. B. D. Gupta, Mathematical Physics, Vikas Publishing House PVT, Ltd., 1978.

33. P. Dirksen, C. Juffermans, A. Leeuwestein, C. Mutsaers, A. Nuijis, R. Pellens, R. Wolters and J. Geman, Proc. SPIE, Microlithography Symposium, 1997.

34. E. E. Moon, P. N. Everett and H. I. Smith, J. Vac. Sci. Technol., B13(6), p2648, 1995.

35. X. Chen, A. A. Ghazanfarian, M. A. McCord, R. F. W. Pease, Proc. SPIE, Vol. 3334, p951, 1998.

36. D. H. Kim, K. H. Lee, Y. H. Oh, J. H. Lee, H. B. Chung and H. J. Yoo; Proc SPIE, Vol. 2440, p928, 1995.

37. X. Chen, A. A. Ghazanfarian, M. A. McCord and R. F. W. Pease, J. Vac. Sci. Technol., B15(6), p2185, 1997.

38. C. W. Helstrom. Statistical Theory of Signal Detection, Pergamon Press, 1960.

39. B. Noble and J. W. Daniel. Applied Linear Algebra, Prentice Hall, 1988.

40. P. R. Halmos; Finite Dimensional Vector Spaces, Springer Verlag, 1974.

41. J. Rissanen, "Modeling by Shortest Data Description", *Automatica*, Vol. 15, No. 5, p465, 1978.

42. J. Rissanen, "A Universal Prior For the Integers and Estimation by Minimum Description Length", *Annals of Statistics*, Vol. 11, p417, 1983.

43. G. H. Golub and C F. van Loan, Matrix Computations, The Johns Hopkins University Press, 1989.

44. S. K. Godunov, A. G. Antonov, O. P. Kiriljuk and V. I. Kostin, Guaranteed Accuracy in Numercial Linear Algebra, The Academic Inc., 1988.

45. W. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, Numerical Recipes in C, Cambrige University Press, 1992

46. A. A. Ghazanfarian, Ph.D. Thesis, Department of Electrical Engineering, Stanford University, 1998

47. A. A. Ghazanfarian, X. Chen, T. Kailath, M. A. McCord and R. F. W. Pease; IEEE Trans. on ASSP, Vol. 4, p1913, 1998

48. A. A. Ghazanfarian, X. Chen, M. A. McCord and R. F. W. Pease; Proc. SPIE, Microlithography Symposium, Santa Clara, California, 1998

49. X. Chen, A. A. Ghazanfarian, M. A. McCord and R. F. W. Pease. The 42nd International Conference on Electron, Ion, Photon Beams and Nanofabraication Technologies, Chicago, 1998.

50. A. F. Scaduto, R. Booth, B. Hwang, J. Mase, J. Miller, J. Steeves and T. Tran-Quinn, IEEE/SEMI, Proc. Advanced Semiconductor Manufacturing Conference and Workshop, 1996.

51. Courtesy of Mike Acosta and Kent Green of Motorola, Inc.

52. Mike Acosta and Kent Green of Motorola Inc., private communication.

53. Chiaki Sato of Cannon U. S. A., private communication.

54. A. A. Ghazanfarian, X. Chen, M. A. McCord and R. F. W. Pease. The 42nd International Conference on Electron, Ion, Photon Beams and Nanofabraication Technologies, Chicago, 1998.

IMAGE EVALUATION
TEST TARGET (QA-3)

150mm

6″
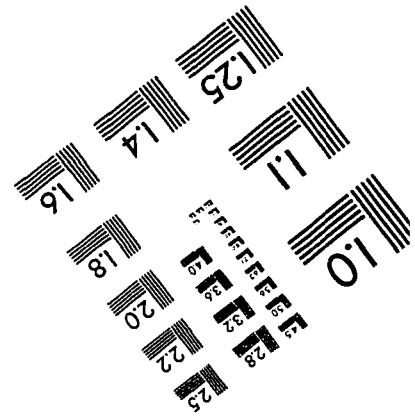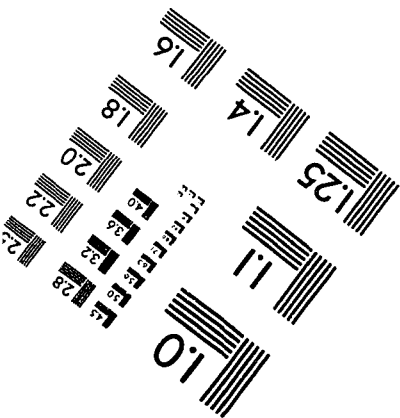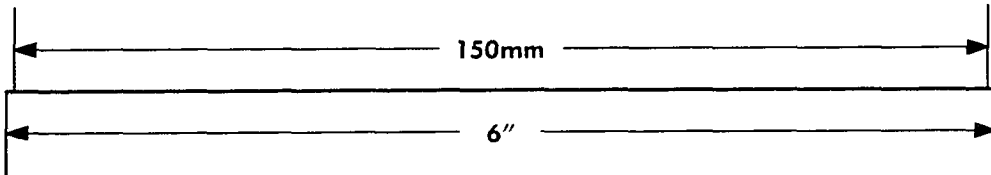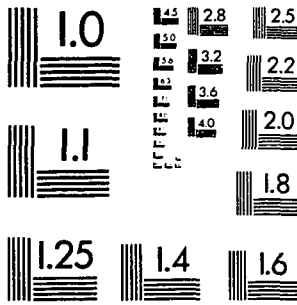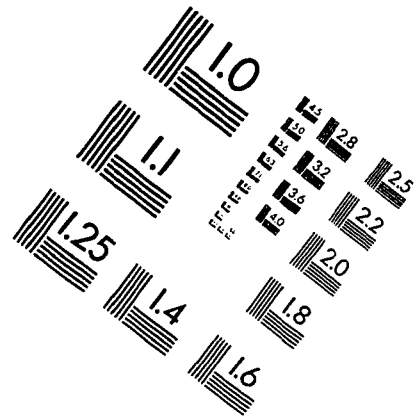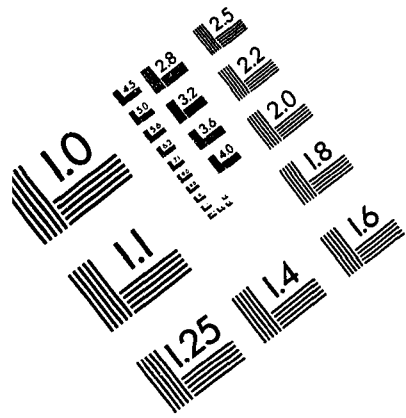
APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© '993, Applied Image, Inc.. All Rights Reserved