EINDHOVEN UNIVERSITY OF TECHNOLOGY

Eindhoven University of Technology

MASTER

A basis for the order-to-order method
theory and validation

van der Aa, N.P.

Award date:
2003

Link to publication

Disclaimer
This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain

TECHNISCHE UNIVERSITEIT EINDHOVEN
Department of Mathematics and Computing Science

MASTER'S THESIS

A basis for the Order-to-Order method

Theory and validation

by
N.P. van der Aa

Supervisor: Dr. H.G. ter Morsche

Eindhoven, April 2003

ASML

File name : Masterthesis.pdf	
Status : final	**Doc ID** : MasterPIR
Info type : PIR	**Last update** : 27 - 3 - 2003

Program / Machine type:	**Subsystem / Project:**
PAS5500 / TwinScan	Order-to-Order

Title:

PIR A basis for the Order-to-Order method

Summary: **Confidential**

This master's thesis discusses the Order-to-Order method, which is a method meant to optimize the alignment process during a photolithographic process step. The goal of the Order-to-Order method is to find the diffraction order with smallest wafer-to-wafer variance of the process-induced shift. This order is recommended during production.

In the current implementation of the Order-to-Order method it is assumed that the process induced errors are uncorrelated for different orders. This thesis shows that this assumption is not correct by means of experimental results and simulations.

It is shown that the presence of covariance between diffraction orders, ensures that the Order-to-Order method can not compute the exact process induced variance, since the true values of the covariances can not be determined with the available data.

This document may form the basis of a future patent application. For this reason the document must be kept strictly confidential!

For change control		Reviewers		For info	
Frank van Bilsen	ASML	Arie den Boef	ASML	mea_al	
		Sicco Schets	ASML	mea_po	
		Jeroen Huijbregtse	ASML	Jan Willem Martens	ASML
		Frank van Bilsen	ASML	Chris de Mol	ASML
		Henny ter Morsche	TU/e		
		Alessandro Di Bucchianico	TU/e		
		Robert Mattheij	TU/e		

Maintainer : Nico van der Aa	**Keywords:**
Pages incl. cover : 50	Order-to-Order
Word processor : LaTeX asml-0.2	

© ASML 2003. All rights reserved.
Reproduction in whole or in part is prohibited without the prior written consent of ASML.

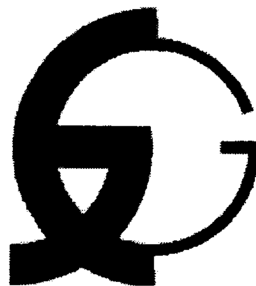
Preface

This master's thesis project was a cooperation between ASML and the Scientific Computing Group. ASML is one of world's leading providers of photolithographic systems for the semiconductor industry. Since photolithography is a process step in the production of computer chips, which requires a high accuracy, much research is carried out to improve the measurement robustness of a so-called alignment process. During such a process the position of a wafer is determined, which is a round slice of silicon where transistors, resistors and capacitors are fabricated on. The Order-to-Order method is a way to optimize the stability of the result of an alignment process. It identifies the most stable measurement from a set of measurements by determining the process induced effect on every measured value. From experimental results it follows that the Order-to-Order method works quite well. The question of ASML is why it works well and a theoretical basis for the Order-to-Order method is desirable.

The Scientific Computing Group is part of the Department of Mathematics and Computing Science at the Technical University of Eindhoven. This group is specialized in numerical methods to solve problems in solid and fluid mechanics, but also in electro-magnetics. In the past another project is carried out in cooperation between ASML and the Scientific Computing Group. Although the Order-to-Order method is mainly statistical, also numerical issues arise and an electro-magnetic approach could be used to enforce the basis of the Order-to-Order method.

I would like to thank Arie den Boef, Sicco Schets, Jeroen Huijbregtse and Frank van Bilsen from ASML for their inspiring dedication to this subject and steer me with questions such as "*why do you do this and why is that necessary?*". From the TU/e I would like to thank Henny ter Morsche for his general support, Robert Mattheij for enabling this joint project with ASML and Alessandro Di Bucchianico for his guidance concerning the statistics involved. Unfortunately, the statistics used in the Order-to-Order method is still lean, but that is due to my inexperience with statistics. Furthermore, I would like to thank everybody from the process overlay group at ASML for creating a friendly and supportive atmosphere, which was pleasant to work in.

TU/e technische universiteit eindhoven



Contents

1 Introduction **5**
1.1 Context of the Order-to-Order method 5
1.2 Definition of the Order-to-Order method 7
1.3 Overview of master's thesis 8

2 Description of experiment **9**
2.1 General description of the processes 9
2.1.1 Tungsten Chemical Mechanical Polishing (W-CMP) 9
2.1.2 Resist Spinning (RS) 10
2.2 Multiple Mark Measurement method (3M) 11
2.3 Derivation of the data sets 11
2.3.1 W-CMP process 12
2.3.2 RS process 13

3 Theoretical model for the Order-to-Order method **16**
3.1 Basic assumption 16
3.2 Statistical model 17
3.3 Deterministic model 19
3.4 3M data vs production data 20

4 Prior knowledge based on assumed correlation between orders **23**
4.1 Reduce the number of unknowns 23
4.1.1 Set covariances to a certain value 23
4.1.2 Set a variance to zero 31
4.2 Add extra equations 32

5 Using the mark shape as prior knowledge **34**
5.1 Vertical Propagation Model for the RS process 35
5.1.1 Assumptions of the Vertical Propagation Model 35
5.1.2 Derivation of the near-field for one mark segment 35
5.1.3 Near-field for an infinitely large grating 37
5.1.4 Derivation of the shift a diffraction order 38
5.2 Determine the shape of the resist layer 39
5.3 Wafer quality (WQ) and contrast 40
5.4 Simulations 41
5.4.1 Illustrative example: flat resist surface 41
5.4.2 Simulations for separate variables 42
5.4.3 Verification of the assumptions made to obtain a solution 45

6 Conclusions **48**

7 Open issues **49**

1 Introduction

1.1 Context of the Order-to-Order method

An integrated circuit (IC) is a collection of components connected together in a complete configuration to perform some useful electronic function [1]. The term "integrated" refers to the manufacturing process that combines separate electronic components into an integral whole. An IC is the result of a multistep photochemical manufacturing process. For producing a printed circuit the emphasis is on interconnecting components, such as transistors, resistors and capacitors. For IC's the primary emphasis shifts to physical and chemical processes. The production of such an IC starts with a round slice with a diameter of typically 300 mm, made of silicon, a so-called wafer. Next, circuit elements are produced simultaneously layer after layer on this wafer.

To build circuit elements like transistors, resistors and capacitors in layers on a silicon wafer, the wafer undergoes several processes like photolithography, material deposition, Chemical Mechanical Polishing (CMP), etching, etc [2]. To illustrate where these processes are used, consider Figure 1. This figure shows the fabrication process of an IC. The process starts with a cylinder of silicon that is cut into slices (1). Next, the slice is polished to obtain an ultra-flat wafer (2). In (3) a layer of material is deposited on the wafer, followed by a deposition of a thin layer of photoresist (4). This process is referred to as the Resist Spin (RS) process. A circuit pattern (on a reticle) is projected onto a section of the wafer using UV light. This light reacts with the photoresist and transfers the circuit image onto the wafer. This section of the wafer will eventually become an IC. The process is repeated until the wafer is covered with many patterns, each of which will become an IC (5). This part of the process uses photolithographic machines, which ASML develops and manufactures. ASML is market leader in these kind of machines. The wafers are then baked to dry, evaporate remaining solvents, and hardens the photoresist. The exposed resist is washed away (6). Etching and ion implantation are done to create vertical or horizontal paths between layers on the wafer (7). The last process is to remove the remaining pattern of photoresist (8). Now the wafer is ready (9) and cut into individual IC's (10). Finally, the IC's are packed, and connector pins are added to produce the final chip (11). Steps (3)

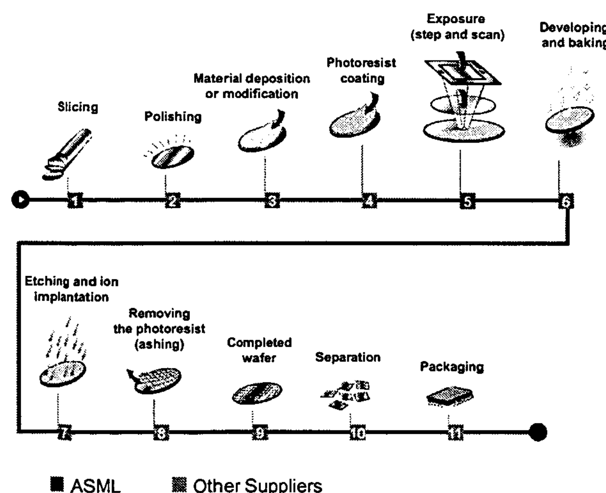


Figure 1: *The production process of a chip.*

to (8) are repeated 20-30 times and each time it is crucial that the patterns are placed on top of each other as accurately as possible.

In this master's thesis, two processes are considered. One of them is Resist Spinning (RS), which is a part of the photolithographic process, the other is tungsten CMP (W-CMP). Both processes are not selective, which means that the process cannot be influenced on a specific position of the wafer, when it is implemented. The RS process ensures the entire wafer is covered with a photoresist layer, while the W-CMP process step planarizes a surface covered with tungsten.

RS is a part of the photolithographic process. During the RS process a thin layer of light sensitive material is deposited on the wafer. With an exposure step a pattern is illuminated. The part of the photoresist layer which is exposed, is removed. The remainder of the photoresist prohibits processes like etching to affect the unexposed parts of the wafer. Note that the exposure process is selective, in contrast to other processes like RS and W-CMP.

Because the pattern in one layer has to connect with the pattern of another layer, it is very important to know the exact position of the wafer during the exposure process. This introduces the term overlay, which is defined as the precision by which a pattern is aligned and printed onto a previous printed layer of an IC. Overlay errors are caused by both the machine and the process itself. Currently, the overlay accuracy is approaching < 20 nm (mean plus 3 sigma).

To locate an object in a three-dimensional space, one needs six degrees of freedom: x-, y- and z-position and rotation around the x-, y- and z-axis. For the Order-to-Order method, the ATHENA[1] or Alignment Sensor is considered. The goal of this sensor is to measure the x- and y-position and the rotation in z-direction of a wafer. To align a wafer, every layer must contain several alignment marks. These marks are periodic structures, so-called gratings, that are etched into a specific layer. Illumination of the alignment marks causes diffraction of light. By measuring the intensities of the inference between diffraction orders $(n, -n)$ and comparing the intensity pattern with a reference grating, the phase difference between orders $(n, -n)$ can be determined (see Figure 2). From the ob-

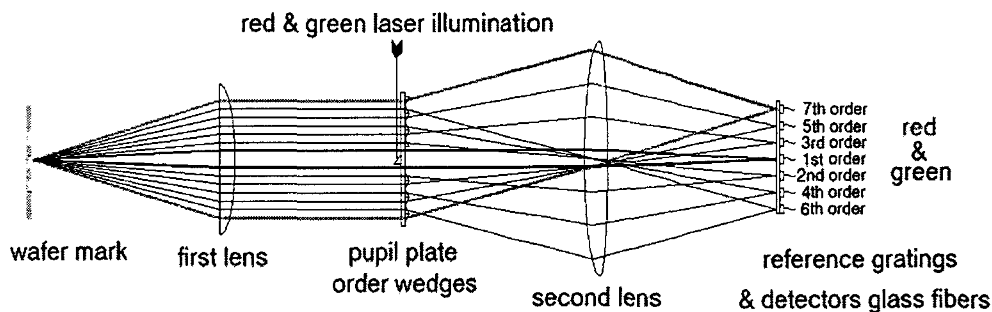


Figure 2: *A schematic representation of the measurement principle of the ATHENA sensor. Diffraction order are captured by the first lens. Between the first and second lens the orders are split in such a way that the orders are separated behind the second lens and can be detected separately.*

[1]ATHENA = Advanced Technology using High-order ENhanced of Alignment

tained phases, the ATHENA sensor computes the x- or y-position of the grating direction of the marks.

When a wafer is processed (e.g. with W-CMP), the alignment marks on this wafer are also processed. In case of the RS process, a thin layer of photoresist covers the alignment mark. During the W-CMP process asymmetric roundings of the phase grating lines could occur. Both effects cause the alignment mark to become asymmetric. Symmetric marks have a uniquely defined position, namely all diffraction orders measure the center of mark symmetry. Asymmetric marks do not have a uniquely defined position. In that case each diffraction order points to an apparent position which is not necessarily equal for all diffraction orders.

In addition, each measured diffraction order contains a diffraction order independent part, which causes the measured position of the alignment mark to deviate from the expected position. This term is equal for all diffraction orders and not introduced by the process. The term "diffraction order independent part" is used for all deviations from the exact positions, which are not caused by changes in shape of the alignment marks due to processing, but that can be contributed to the intrinsic system performance.

1.2 Definition of the Order-to-Order method

The idea behind the so-called Order-to-Order method is that each measured diffraction order contains information about the changes in shape of an alignment mark caused by processing. A basic assumption for applying the Order-to-Order method is that the diffraction order dependent part is caused by processing the wafers only. This assumption is justified because symmetric marks ensure all diffraction orders give the same position, while for asymmetric marks the measured positions of each diffraction order are different. As a consequence of this assumption the remainder part is order independent. This is the reason why this part is called the diffraction order independent part. However, this error can be much larger than the process induced part.

If there is a certain diffraction order of which the process part indicates approximately the same position for almost all marks for all wafers, this diffraction order yields the most stable alignment. Note that all marks over the wafer must be comparable. The description of the problem concerning the Order-to-Order method is given in the problem definition.

Problem definition
Identify that diffraction order which results in the most stable position when alignment marks are influenced by process effects.

As noted before, the most stable order may have a large offset from the actual position. However, because this offset is approximately equal for all marks, one can correct for this offset. As indicated above, the marks must be comparable. When only wafer-to-wafer variability is considered, the marks are comparable, since a mark on a certain position undergoes a similar process effect on all wafers. In case of the RS process one might expect that with increasing radius the resist spin effects become more perceptible. This is also the case for the W-CMP process. Therefore, process effects on two different positions on a wafer can be very different. This thesis is limited to marks located at roughly the same radius. Future extensions of the Order-to-Order method might account for mark location variations.

To find the desired diffraction order, the differences between two diffraction orders, the so-called Shift-between-Orders (SbO's), are investigated. The advantages of considering SbO's instead of APD's is that the diffraction order independent error drops out. The wafer-to-wafer variability of a SbO can be computed easily from the measured positions. However, this is still not the desired variability, since a SbO contains two diffraction orders. According to the problem definition, the variability of the position for a single order due to processing must be computed. Apparently, the problem has changed to find a good decomposition algorithm to identify the desired variability from SbO variability. A difficulty in this algorithm is that correlation between diffraction orders must be taken into account.

Currently, the Order-to-Order method is implemented by assuming there is no correlation between the measured positions of different diffraction orders. Unfortunately, it can be made plausible with Fourier optics that diffraction orders are correlated. Note that [3] shows quite good results when the no covariance is assumed. The main question is why? This question will be the subject of this thesis.

1.3 Overview of master's thesis

In this master's thesis a theoretical basis is given for the Order-to-Order method. To validate this basis, experiments are performed with two processes: the W-CMP process and the RS process. The general use of these processes is described in Chapter 2. In this chapter also the background of the experiments is given which resulted in the data sets that are used to verify the Order-to-Order method. In Chapter 3 the theory of the Order-to-Order is derived and also the way of treating the measured data is discussed. The result of this chapter is an underdetermined system of equations, which results in infinitely many solutions. In Chapters 4 and 5 an attempt is made to reduce the number of variables or to add extra equations. In Chapter 4 an ad-hoc way method is followed, while Chapter 5 uses prior knowledge on the mark shape to get insight in the behaviour of the variables. Finally, this master's thesis ends with conclusions and recommendations in Chapter 6 and a description of open issues in Chapter 7.

2 Description of experiment

To verify the Order-to-Order method experimentally, two data sets of different processes are used. The general implementation of these processes is discussed in Section 2.1. Since both experiments use the 3M method to determine the process effects, this method is pointed out in Section 2.2. The setup of the experiment used to gather the data sets, are described in Section 2.3.

2.1 General description of the processes

In this section the two considered processes W-CMP and RS are described. For each process the actual process is described.

2.1.1 Tungsten Chemical Mechanical Polishing (W-CMP)

The first data set is retrieved from a process called tungsten CMP (W-CMP). W-CMP is used to interconnect two vertically separated metal layers to each other. The name tungsten refers to the material that is used. The process of W-CMP is shown schematically in Figure 3. The insulating material in-between the metal layers is oxide (2). To achieve good imaging, the oxide layer is planarized to obtain a smooth and flat surface (3). At the positions where the connection between the metal layers must be made, small holes are etched into the oxide (4) and afterwards filled with tungsten (5). The superfluous tungsten over the bulk oxide can be removed through a W-CMP removal step (6). Finally, metal[2] is deposited (7). The so-called non-zero layer marks will look like the right side of (7) and this is the mark that has process induced effects.

During W-CMP, deformations of alignment marks such as the asymmetric rounding of the phase

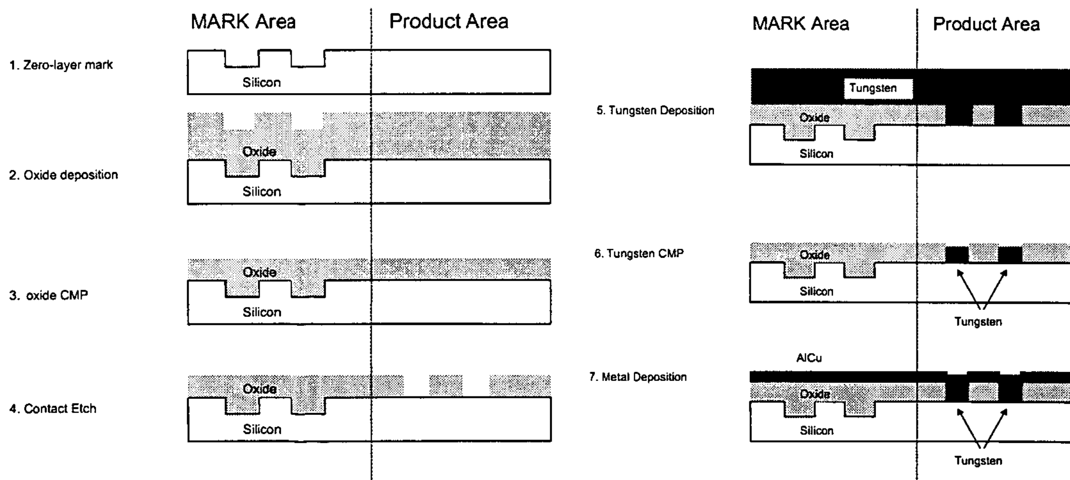


Figure 3: *Schematic representation of the W-CMP process. In each step the left side shows the zero mark area whereas the right side shows a non-zero layer mark.*

[2]This metal is often aluminium.

grating or changes in optical depth of the mark occur (see Figure 4). The latter can cause destructive interference when the optical depth is approximately equal to half the wavelength. Destructive interference causes a severe reduction in signal strength, which strongly amplifies the effect of mark asymmetry on the detected mark position.

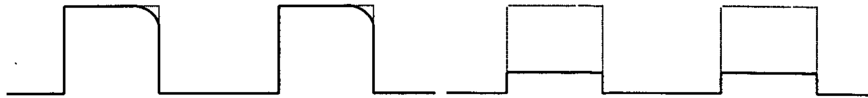


Figure 4: *Schematic representation of the process effects. The left picture shows the asymmetric corner rounding, while the right picture shows the alteration in optical depth.*

Compared to other process steps, the process effects on alignment marks are large for W-CMP. The marks can deform significantly and are therefore suited to validate the Order-to-Order method. For more information on W-CMP see [4].

2.1.2 Resist Spinning (RS)

The second process that is investigated is called Resist Spinning (RS). This process is a preparatory step to expose a wafer. During the RS process, the wafer is rotating and in its center of rotation a few drops of photoresist are released, which spread out to cover the silicon wafer with a thin uniform layer of resist. During the exposure and develop step, parts of the resist layer are removed and the remainders of the photoresist prevent other processes like etching to affect the original layer.

At mark positions a so-called pile-up effect occurs due to the surface irregularity (see Figure 5). The magnitude of this pile-up effect depends upon the topology of the mark. For instance, if the mark depth is more than 25 % of the resist layer thickness, the effect is clearly apparent, while for topographies less then 25 % of the resist layer thickness the effect is negligible.

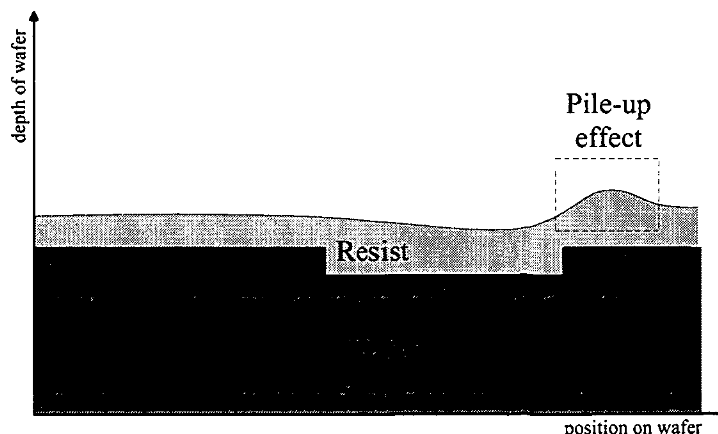


Figure 5: *Schematic representation of the pile-up effect due to the RS process.*

Due to the pile-up effect, the shape of an alignment mark together with the resist layer becomes asym-

metric as seen by the alignment sensor. This creates a process induced effect on alignment positions and, consequently, this process is suitable for experimental validation of the Order-to-Order method (just as the W-CMP process). For more information on the RS process see [5].

2.2 Multiple Mark Measurement method (3M)

An accurate measurement method to determine processing effects, and therefore a way to verify the outcomes of the Order-to-Order method, is the Multiple Mark Measurement (3M) method. This method is capable of determining the process contribution of the total deviation between measured and actual position. As described above, the process effects cause alignment marks to become asymmetric which causes positional differences between diffraction orders. In fact, the 3M method is designed to identify what the Order-to-Order method tries to compute: the Process Induced Alignment Shift (PIAS). This is the reason why this measurement method is used.

The 3M method uses two pairs of two marks to determine the process effect. Each two marks are printed on the same distance from each other on the wafer. Note that both sets must be very close to each other. After processing, all marks except one are cleared, which means that the layers on top of the marks are removed (see Figure 6). The process effect can now be separated from other effects by taking the difference within each pair. The difference between the two sets is equal to the process effect. The 3M method is the most accurate measurement method to determine process effects on alignment available at this moment. However, to implement the 3M method in production is impossible due to the price of additional steps and the complexity. For more information see [6].

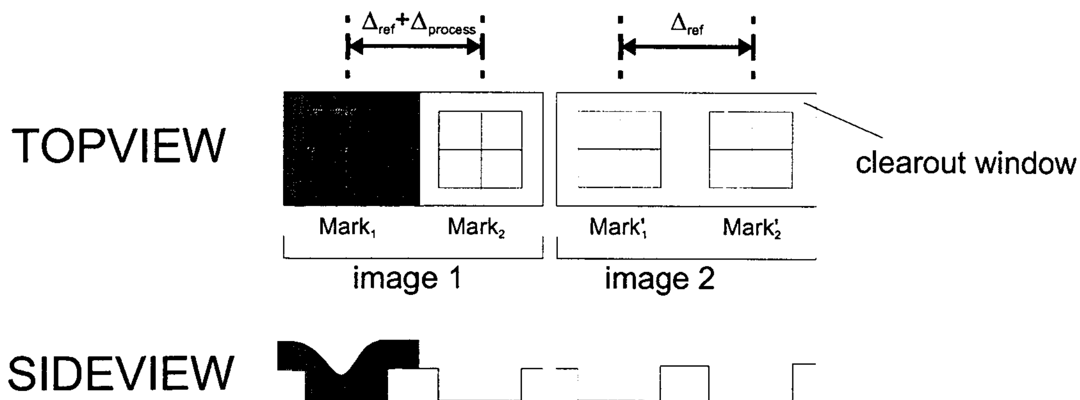


Figure 6: *Schematic representation of the two pairs of marks. Only one mark contains process effects, the other marks are cleared.*

2.3 Derivation of the data sets

In this section the set-up of the experiments for both the W-CMP process as the RS process is discussed as a background for following chapters.

2.3.1 W-CMP process

The W-CMP experiment is carried out with a batch of 10 wafers (300 nm). Since the effect of processing depends also on the kind of alignment mark used, three types of marks are considered. These are the XPA-AH32, XPA-AH53 and XPA-AH74 alignment mark. XPA stands for eXtanded Pattern Area and each mark has a X variant (XPA-X) and a Y variant (XPA-Y) to determine the x- or y-position of the wafer. The schematic representation of the three types of marks is given in Figure 7. These mark types are developed to enhance the intensity of a certain diffraction order as can be seen from Fourier series. The XPA-AH32 mark enhances the third diffraction order, while the XPA-AH53 and XPA-AH74 enhances the fifth and seventh order, respectively.

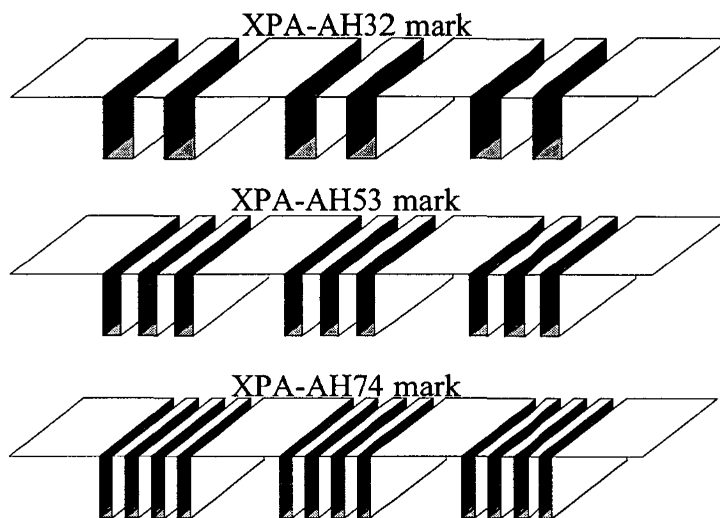


Figure 7: *Schematic representation of the XPA-AH32, XPA-AH53 and XPA-AH74 mark.*

Eight marks of every type are considered: four X- and four Y-marks. The X- and Y-marks are located on the same position on the wafer. The locations of each mark used in this experiment are given in Figure 8.

Since this experiment has not been carried out to verify the Order-to-Order method, each mark has only been measured once per measurement method. In Chapter 3 it is shown that this is statistically not sufficient and this is the reason the RS experiment has been carried out. The measurement methods used in this experiment are the production like measurement method, which serves as input for the Order-to-Order method and the 3M method to validate the results.

The measurements are executed in the following way. Each mark is illuminated with two laser beams knowing a red laser beam (wavelength is 633 nm) and a green laser beam (wavelength is 532 nm). The main reason to use two different wavelength is the chance of destructive interference (see Chapter 5). When the laser beam hits the (processed) marks, diffraction occurs and the intensity and phase of the first seven diffraction orders are measured by the ATHENA sensor. From the phases the positions are derived for every diffraction order. Also the Wafer Quality (a quantity for the signal strength), WQ for short, is measured for each diffraction order. As indicated before, the XPA-AH32 enhanced the 3th diffraction order, so the WQ of this order will be large. Similar the 5th and the 7th diffraction

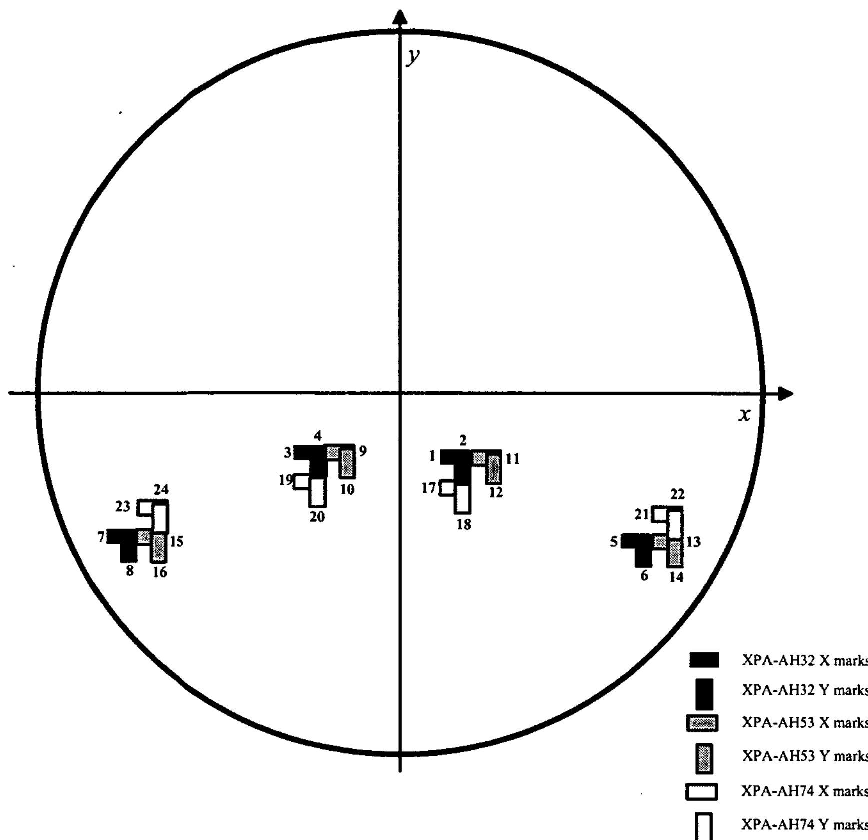


Figure 8: *A schematic plot of the mark positions for the XPA-AH32, XPA-AH53 and XPA-AH74 marks. In reality, the mark do not overlap. In this figure the marks are much larger than in reality.*

order have a large WQ for the XPA-AH53 and XPA-AH74 mark.

Summarizing, the W-CMP experiment results in 2 data sets, one for the 3M method and one for the production like measurement method. Each data set consists of 14 simultaneous positions (7 positions for each color) and WQ's for all 24 marks.

2.3.2 RS process

The experiment involves a batch of six wafers (300 mm) which has undergone the RS process. Because the shape of the resist layer depends on the shape of the alignment mark underneath, two types of marks are considered. The first type is the XPA 8.0 μm mark and the XPA-AH74 mark (see Figure 9). The resist layer is less uniform on a XPA 8.0 μm mark then on a XPA-AH74 mark.

On each of the six wafers 4 mark pairs are considered for both mark types. The x-positions are determined by marks whose grating vectors runs in the x-direction (X-marks). Analogously, the y-positions are determined by marks which gratings lie along the y-direction (Y-marks). Summarizing, there are 4 X-marks and 4 Y-marks per type of mark (2 types), i.e. in total there are 16 marks per wafer. The

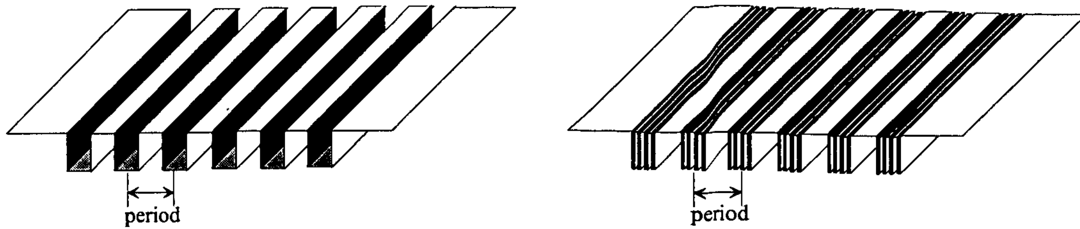


Figure 9: *A schematic representation of XPA 8.0 μm mark (left) and XPA-AH74 mark (right).*

positions of the marks on the wafer are given in Figure 10.

The changes in shape of a mark depend on the position of the mark on a wafer. Therefore, analogous to the W-CMP experiment, two mark pairs are located at one radius and the remaining two at another radius. The diffraction orders are again produced with a red and a green laser. After the execution of the experiment, it appeared that, for example, the fifth diffraction order for the red laser has a very low WQ for both X- and Y-marks of the type XPA-AH74.

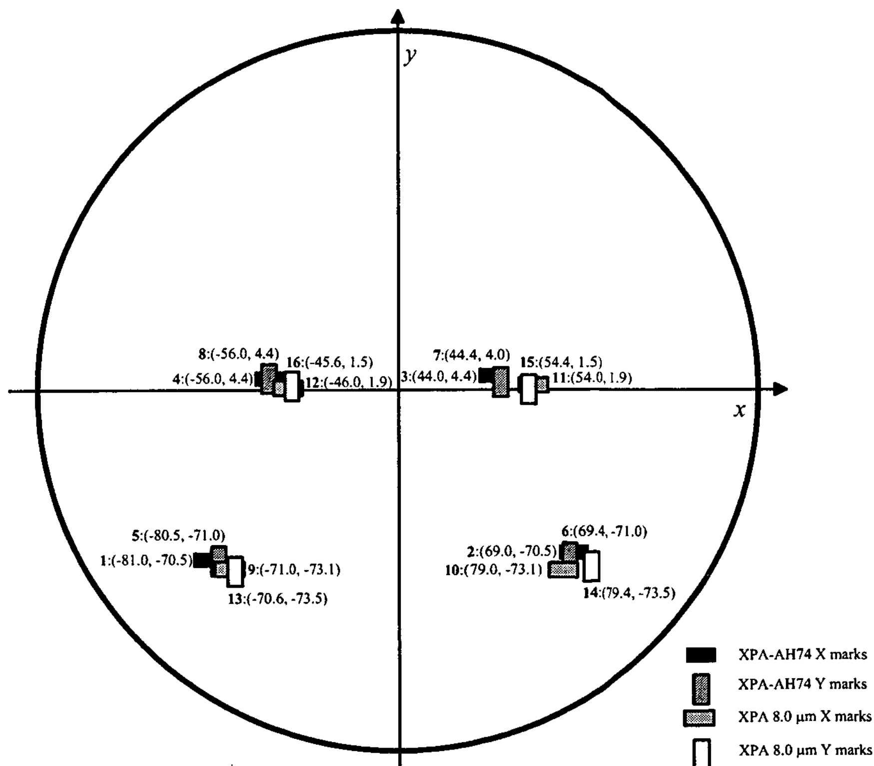


Figure 10: *Positions of the alignment marks for both XPA-AH74 and XPA 8.0 μm marks. Each mark has a label which indicates the mark number and mark x- and y-position. In reality, the mark do not overlap. In this figure the marks are much larger than in reality.*

To verify the Order-to-Order method, two types of measurement methods are used. The first one is the 3M method, the other measurement method is a production like measurement method. In IC fabrication, only the latter is available.

Since noise could ruin the outcomes of the Order-to-Order method, each mark is measured several times. In production, there is a limit for the number of marks that can be measured. Therefore, each mark is measured 10 times consecutively without the wafer leaving the machine. When all wafers are measured, the measurement sequence is repeated, so each mark is scanned 20 times. This number of measurements is not an exceeding of the limit, but those limits should be kept in mind since the number of measurements could not be much larger.

The RS experiment gives two data sets, for each measurement method one. Each data set contains information for the diffraction orders of all 16 marks (4 marks for 2 types (XPA-AH74 and XPA 8.0 μm) and both X- and Y-marks) measured with the red and the green laser. Also the WQ's are measured.

3 Theoretical model for the Order-to-Order method

In this chapter the theoretical basis of the Order-to-Order method is presented. The development of this basis is described step by step, starting with the basic assumption in Section 3.1. The second section provides a statistical approach and in Section 3.3 the model that will be used in the following two chapters of this master's thesis is completed. This chapter is concluded with a remark about the usage of two different data sets, namely the 3M and the production like data sets. Most of the statistical definitions and formulas presented in this chapter are taken from [7], unless stated otherwise.

3.1 Basic assumption

The Order-to-Order method is a technique to separate process induced variations from other variations that combine into the measured aligned position deviation (APD) for a certain diffraction order. An APD is the difference between measured and expected position of an alignment mark. The APD can be split into three parts. The first part is the process induced error, also called Process Induced Alignment Shift (PIAS). Processing an alignment mark causes alterations in the shape of the mark, which can become asymmetric. Diffraction orders react differently on asymmetry as will be shown in Chapter 5. Therefore, the observed PIAS's are order dependent. Note, that this is the only part of an APD which is of interest in this thesis. The second part is the diffraction order independent part, which for example include translation, rotation and thermal expansion of a wafer. Diffraction orders do not react differently on e.g. rotation of a wafer, since these kinds of effects do not alter the shape of an alignment mark. Although the name already indicates the diffraction order independent part is order independent, this part can be different from mark-to-mark and from wafer-to-wafer since, for example, rotation is radius dependent (mark-to-mark) and translation is the same for all mark positions on one wafer, but could be different for another wafer (wafer-to-wafer). Finally, the measurement noise is taken into account, which is a random error influencing each measurement. Clearly, this noise error contributes to the differences between the observed diffraction orders. However, there are techniques to dispose of them, as will be shown in the following section.

Since only the PIAS and the noise are supposed to be diffraction order dependent and the noise contribution could be disposed of, the basic assumption that is made in the Order-to-Order method is:

Basic assumption
The PIAS is the only part of an APD that is diffraction order dependent.

The mathematical model for a measured APD is given by

$$\boldsymbol{Y}_{ijkl} = \boldsymbol{\tau}_{ijk} + \boldsymbol{\beta}_{ijk} + \boldsymbol{\varepsilon}_{ijkl}, \qquad i = 1, ..., p \quad j = 1, ..., q \quad k = 1, ..., r \quad l = 1, ..., m. \tag{1}$$

Here, $\boldsymbol{Y}_{ijkl} \in \mathbb{R}^n$ are the simultaneous measurements of the n diffraction orders coming from mark i on wafer j measured with the kth colour during the lth measurement. For every diffraction order, Y_{ijkl} can be split in the diffraction order independent error $\boldsymbol{\beta}_{ijk}$, the PIAS $\boldsymbol{\tau}_{ijk}$ and the noise error $\boldsymbol{\varepsilon}_{ijkl}$. For given i, j and k, system (1) has n equations and (disregarding the noise errors) $2n$ unknowns. However, the diffraction order independent error is the same for every diffraction order. Implementing

© ASML 2003

this property gives:

$$\beta_{ijk} = \beta_{ijk} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^n. \tag{2}$$

For given i, j and k, system (1) is reduced to n equations with $n+1$ unknowns. This means that system (1) is still an underdetermined system of equations and therefore has infinitely many solutions. Fortunately, the actual values of the variables $\tau_{ijk}(\in \mathbb{R}^n)$ and $\beta_{ijk}(\in \mathbb{R})$ are not the quantities of interest. The Order-to-Order method searches for the diffraction order that varies least due to process variation or, in other words, that diffraction order for which the PIAS varies least. This implies that the diffraction order independent part is not of interest. Actually, this is the fundamental reasoning behind the Order-to-Order method [8]: for any given i, j and k, subtracting the measured APD's (each of a different diffraction order) ensures that the diffraction order independent part is removed. This subtraction is defined as a Shift-between-Order (SbO) and is expressed mathematically by:

$$Y_a - Y_b = \tau_a - \tau_b + \varepsilon_a - \varepsilon_b, \quad a, b = 1, ..., n \quad a \neq b. \tag{3}$$

For readability reasons and without loss of generality the indices i, j, k and l are disregarded. This equation is valid for any given combination of i, j and k. The index l is removed when the noise error has been erased. For n diffraction orders, $\binom{n}{2} = \frac{1}{2}n(n-1)$ possible SbO's can be computed.

A SbO contains only information about the difference between two PIAS's of two different diffraction orders. This implies that if the PIAS for a certain order of a mark is known, a reference point is available for all other PIAS's and they can all be determined. The problem is reduced to find the PIAS of a single diffraction order. However, this is not possible so therefore the variance of a PIAS is considered.

The key problem of converting system (1) consisting of APD's to system (3) consisting of SbO's, is that the goal is to find the variance of PIAS's over wafers of single orders instead of the variance of SbO's. In theory, if the variance of a certain Shift-between-Order is small, it is possible that the two involved diffraction orders behave quite the same, but the actual variance of the PIAS's for the separate orders can be large. The main difficulty in decomposing variances of SbO's lies in the fact that the PIAS's of single diffraction orders can be correlated and therefore also covariances between PIAS's must be taken into account.

The following section deals with the noise contribution of an APD and provides a statistical approach to obtain the wafer-to-wafer PIAS variance for each diffraction order from the measured SbO's.

3.2 Statistical model

Because the data set consists of measured data, the obtained data is polluted with measurement noise. To implement a method to obtain information from measured data, a model must be developed to interpret this data. There are two kinds of errors involved with the model, knowing noise errors and the so-called "fit" errors. By measuring every mark several times and averaging those values, the obtained values are more accurate and the noise is reduced to such a level, that it can be neglected. However, the magnitude of the fit errors depend on the underlying model.

Averaging over all measurements gives a more reliable estimate of the true value, which can be expressed by confidence intervals. To compute these intervals, the sample standard deviation s must be

determined. Note that noise is assumed to be normally distributed. So, if x_1, x_2, ..., x_m is a sample of m observations, the sample variance is defined as

$$s^2 = \frac{\sum_{i=1}^{m} (x_i - \overline{x})^2}{m-1} \tag{4}$$

where $\overline{x}$ denotes the mean of the sample. The sample standard deviation s is the positive square root of the sample variance. Now, a confidence interval can be determined for the sample. If $\overline{x}$ and s are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2, a $100(1-\alpha)$ % confidence interval on μ (the actual value) is given by

$$\overline{x} - t_{\alpha/2,m-1}\frac{s}{\sqrt{m}} \leq \mu \leq \overline{x} + t_{\alpha/2,m-1}\frac{s}{\sqrt{m}} \tag{5}$$

where $t_{\alpha/2,m-1}$ is the upper $100(\alpha/2)$ % point of the t distribution with $m-1$ degrees of freedom. A 100 % confidence interval contains no information, since this interval equals $\mathbb{R}$.

To illustrate the length of the confidence intervals, consider the RS experiment. Each mark was measured 10 times consecutively. For the XPA-AH74 marks each diffraction order has a 99 % confidence interval of approximately 1 à 2 nm and 1 à 3 nm for a 99.9 % confidence interval, depending of the measured mark, wafer and color.

For the W-CMP data only one measurement per mark is performed. Hence, the sample variance could not be computed and no confidence interval can be determined. This is the reason that an additional experiment was executed: the RS experiment. In this experiment, each mark on every wafer and for every color is measured several times in order to compute this sample variance and determine the confidence intervals.

however, if the standard deviation of the noise error is known (e.g. from previous experiments), it is possible to set up a confidence interval, even for one measurement. If $\overline{x}$ is the sample mean of a random sample of size m from a normal population with known variance σ^2, a $100\,(1-\alpha)$ % confidence on μ is given by

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{m}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{m}} \tag{6}$$

where $z_{\alpha/2}$ is the upper $100\,(\alpha/2)$ % point of the standard normal distribution. For the W-CMP data this σ^2 has to be estimated.

The data set can be divided into so-called factors such as measurements per wafer, per position on the wafer and per color. Existing statistical techniques like (Multivariate) ANalysis Of VAriance ((M)ANOVA) can be used to estimate the variance of any of these factors from the different levels of the specific factor. A difficulty in such a technique is that a single level of a factor can not be used (a single value has no variance). The Order-to-Order method is developed for production like environments, which implies that the wafers are processed wafers. This can be considered as one level of the factor "processing". If data was available of the same, but not processed wafers, the "processing" factor would have two levels and the process effects could be separated from other effects. As a result, (M)ANOVA techniques cannot distinguish the PIAS variance from the APD variance from the processed wafers alone, since there is no second level that can make the distinction between an APD and a PIAS. Therefore, in case of (M)ANOVA's more information is required about the relation between the PIAS variance and the APD variance in order to use such a standard technique. Therefore, standard statistical methods can not be used. For more information on (M)ANOVA techniques see [7].

3.3 Deterministic model

In this section, these measurement values are considered to be deterministic. The SbO's and the variances of these SbO's are computed directly from the deterministic data. When n diffraction orders are measured, there are $\binom{n}{2}$ possible combinations to form SbO's.

Decomposition of the SbO variances into variances and covariances of the single order PIAS's is done according to the following formula [9]:

$$\operatorname{var}(X - Y) = \operatorname{var}(X) + \operatorname{var}(Y) - 2\operatorname{cov}(X, Y). \tag{7}$$

When this decomposition equation is applied on the variances of SbO's, the following system is obtained:

$$\boldsymbol{b} = A\boldsymbol{x}, \tag{8}$$

with $\boldsymbol{b} \in \mathbb{R}^{\binom{n}{2}\times 1}$ the known vector consisting of the computed SbO's, $\boldsymbol{x} \in \mathbb{R}^{(\binom{n}{2}+n)\times 1}$ the solution vector consisting of variances and covariances of single order PIAS's, and $A \in \mathbb{R}^{\binom{n}{2}\times(\binom{n}{2}+n)}$ the coefficient matrix given by

$$A = \left[\begin{array}{cccccc|ccc}
1 & 1 & & & & & -2 & & \\
\vdots & & \ddots & & \emptyset & & & \ddots & \emptyset \\
\vdots & \emptyset & & & \ddots & & & & \\
1 & & & & & 1 & & & \\
0 & 1 & 1 & & & & & & \\
\vdots & \vdots & & \ddots & \emptyset & & & & \\
\vdots & \vdots & \emptyset & & \ddots & & & & \\
0 & 1 & & & & 1 & & & \\
 & & & \vdots & & & & & \\
 & & 1 & 1 & 0 & 0 & & & \\
 & & 1 & 0 & 1 & 0 & & & \\
 & & 1 & 0 & 0 & 1 & \emptyset & & \\
 & \emptyset & 0 & 1 & 1 & 0 & & & \\
 & & 0 & 1 & 0 & 1 & & & \ddots \\
 & & 0 & 0 & 1 & 1 & & & -2
\end{array}\right]. \tag{9}$$

Equation (8) is a system of $\binom{n}{2}$ equations with $\binom{n}{2} + n$ unknowns (n variances and $\binom{n}{2}$ covariances). Note that the right-hand side of system (8) consists of only PIAS's since the diffraction order independent part and the noise error are eliminated. Because there are more unknowns than equations, this system is an underdetermined system of equations. Without additional knowledge, it is impossible to compute the single order PIAS variances.

In previous research [3] and [8], it has been assumed that all covariances are equal to zero or, in other words, the diffraction orders behave completely independent. The following example shows that this assumption can lead to erroneous results. To keep the example simple, a situation is considered with only three diffraction orders. Consider the diffraction orders a, b and c for a certain mark over m wafers. It is assumed that the wafer-to-wafer variance of the SbO's is given by:

$$\begin{cases} \operatorname{var}(r_a - r_b) = 0 \\ \operatorname{var}(r_a - r_c) = A \\ \operatorname{var}(r_b - r_c) = A \end{cases}. \tag{10}$$

© ASML 2003

The Order-to-Order method wants to solve system (8) with only three diffraction orders and with the assumption that the diffraction orders behave independently (all covariances are zero). The solution is given by

$$\begin{cases} \text{var}(r_a) = 0 & \text{cov}(r_a, r_b) = 0 \\ \text{var}(r_b) = 0 & \text{cov}(r_a, r_c) = 0 \\ \text{var}(r_c) = A & \text{cov}(r_b, r_c) = 0 \end{cases} . \tag{11}$$

However, when complete correlation between diffraction order a and b is assumed and no correlation is assumed between orders a and c and between order b and c, the solution is given by

$$\begin{cases} \text{var}(r_a) = A & \text{cov}(r_a, r_b) = A \\ \text{var}(r_b) = A & \text{cov}(r_a, r_c) = 0 \\ \text{var}(r_c) = 0 & \text{cov}(r_b, r_c) = 0 \end{cases} . \tag{12}$$

Both solutions are illustrated in Figure 11. The impact of the additional knowledge is severe since solution (11) indicates both points a and b are most stable under processing, while solution (12) indicates point c varies the least.

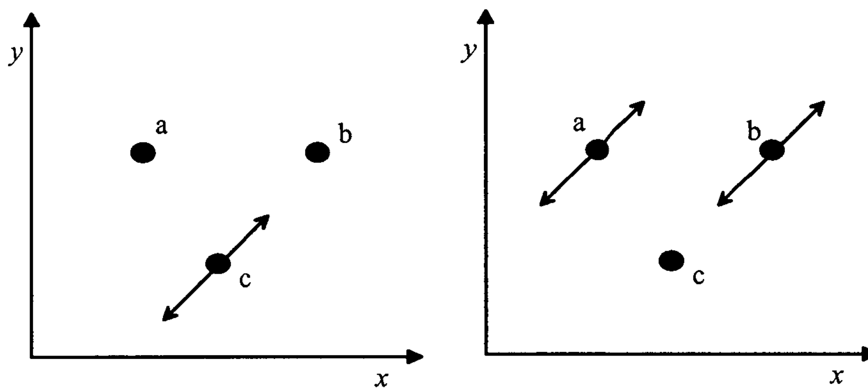


Figure 11: *The left plot shows points a and b which do not vary and point c which does vary. The right plot shows the case where one point is not varying (point c) and the two others (a and b) vary such that both points keep the same distance to each other. The arrows show the wafer-to-wafer variation.*

3.4 3M data vs production data

In theory, the 3M data must result in exactly the same SbO's as the production data. However, both data sets are derived from measured data. Although the measurement errors are reduced by measuring each mark several times, as is done in the RS experiment, infinitely many measurements have to be executed to ensure all measurement errors are filtered out.

Figure 12 (left) shows the computed SbO's of a certain mark on a specific wafer for both data sets. As can be seen, the order of magnitude of the SbO's in the production environment data is larger than in the 3M data. This can be explained by the fact that the 3M measurement method gets rid of off-sets between APD's of the same diffraction order. Large constant offsets are eliminated during the subtraction. In production like environments, the even diffraction orders are not calibrated, which means that these orders may still contain large constant offsets. As can be seen in Figure 12 the SbO's

of two odd or two even diffraction orders result in a small SbO, while SbO's of combinations of an odd and an even order are very large.

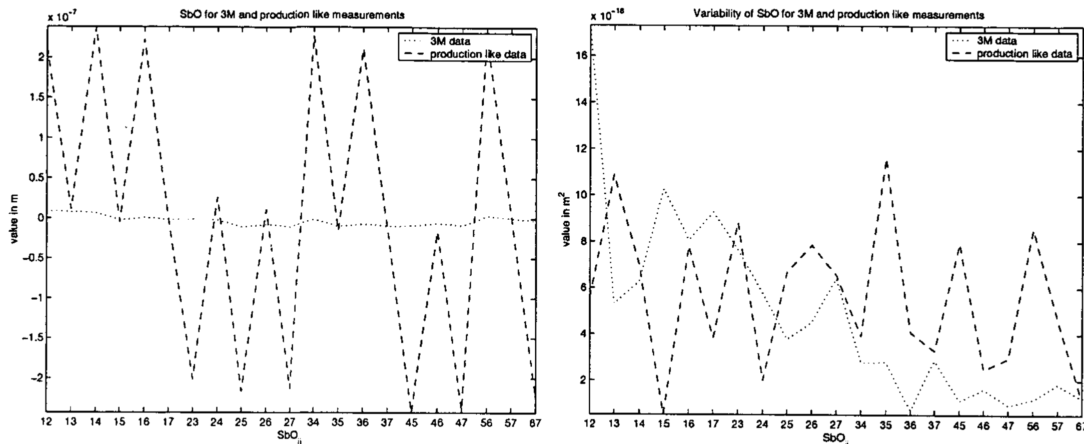


Figure 12: *Left the SbO plot is given for a certain mark on a specific wafer. The right figure gives the wafer-to-wafer variance of this mark over all six wafers. The horizontal axis denotes which SbO pair is considered. For example, 24 means that the following SbO is considered:* $SbO_{24} = Y_2 - Y_4$.

The variance of SbO's computed for both measurement methods should be identical. The reason is that SbO's from other wafers have the same offsets due to calibration. A variance is a measurement of spread from the mean of these SbO's and therefore each SbO is automatically corrected for this offset when a variance is computed. However, the measurement noise again plays a role, since the variance is estimated from measured data. Only six wafers are measured in case of the RS experiment. An error in one of the six measured wafers has large consequences for the final estimate of the variance. To quantify this phenomenon, a confidence interval is given for the actual variance σ^2. If s^2 is the sample variance from a random sample of m observations from a Gaussian distribution with unknown variance σ^2, the a $100\,(1-\alpha)$ % confidence interval on σ^2 is

$$\frac{(m-1)s^2}{\chi^2_{\alpha/2,m-1}} \leq \sigma^2 \leq \frac{(m-1)s^2}{\chi^2_{1-\alpha/2,m-1}} \tag{13}$$

where $\chi^2_{\alpha/2,m-1}$ and $\chi^2_{1-\alpha/2,m-1}$ are the upper and lower $100\,(\alpha/2)$ % point of the chi-squared distribution with $m-1$ degrees of freedom, respectively. To indicate how large the effect is of the number of wafers, the 99 % confidence intervals are given for 6 c.q. 25 wafers:

$$\begin{array}{ll} 0.2985s^2 \leq \sigma^2 \leq 12.1951s^2 & \text{for } m = 6 \\ 0.5268s^2 \leq \sigma^2 \leq 2.4267s^2 & \text{for } m = 25 \end{array} \tag{14}$$

This explains why in Figure 12 (right) the variances of both data sets differ significantly from each other in the RS experiment.

If the variances of the SbO's differ significantly from each other, it is clear that the solutions of the Order-to-Order method are also different. This implies that, although the 3M data set is the only way to verify if the Order-to-Order method indicates the right diffraction order, both data sets can not be compared. In future experiments, more wafers should be taken into account to have a more

accurate estimate of the SbO variance. Since there were no batches available of more than 10 wafers suitable for 3M experiments, the 3M data can not be compared to production like data to verify the Order-to-Order method. However, the 3M data set is the most accurate measurement method to establish the single order PIAS's. Therefore, the variance of PIAS's could be computed directly from the data. The Order-to-Order method only provides another method to determine these PIAS's. This implies that in Chapter 4 the 3M data is used to verify the Order-to-Order method.

The W-CMP data show a remarkable resemblance between 3M data and production like data. In Figure 13, a certain mark is chosen which is representable for all marks. One can see that the SbO's are approximately the same and the wafer-to-wafer variance for this mark are also similar for 3M and production like data. Therefore, in Chapter 4 the 3M data could be used as a comparison for the outcome of the Order-to-Order method.

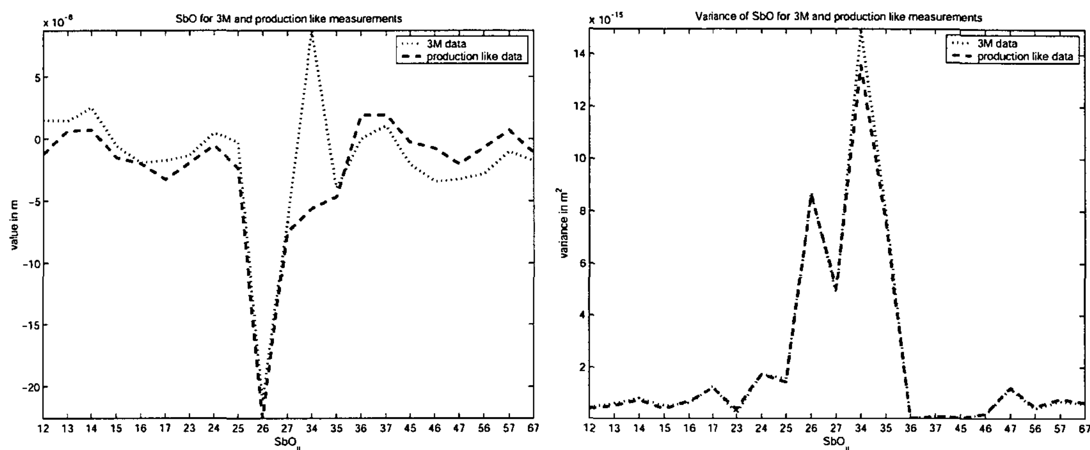


Figure 13: *Left the SbO plot is given for a certain mark on a specific wafer. The right figure gives the wafer-to-wafer variance of this mark over all six wafers. The horizontal axis denotes which SbO pair is considered.*

In summary, the data sets of the RS process are not suitable for mutual comparisons, while for the W-CMP process the 3M data could be used to verify the outcome of the Order-to-Order method. Even if the data sets are comparable, the WQ's of some diffraction orders could be small, and then these orders should be excluded from comparison.

4 Prior knowledge based on assumed correlation between orders

Because the system obtained in Chapter 3 has n more unknowns than equations, two approaches could be used to solve the system. One is to add n extra independent equations (4.1). Another approach is to reduce the number of unknowns (4.2). Obviously, a combination of both is also possible. For both methods, additional knowledge is necessary, for example, including knowledge about the shape of a mark topology (see Chapter 5). This chapter discusses the effects of eliminating unknowns or adding additional equations to system (8).

4.1 Reduce the number of unknowns

As already mentioned, current research of the Order-to-Order method assumes that there is no correlation between APD's of different diffraction orders. This is an example of setting unknowns to a certain value. In total $\binom{n}{2}$ covariances are given a value, leaving only n variables to be solved. Actually, in this approach more unknowns are set to zero than is required to obtain a square coefficient matrix in system (8).

There are two kinds of variables: variances and covariances. In (4.1.1) covariances are given a value, while in (4.1.2) also variances are pinpointed to a certain value.

4.1.1 Set covariances to a certain value

- **Set covariances between all orders or between even and odd orders to zero**
  To solve system (9) at least n unknowns must be given a certain value. It would intuitively make more sense to set those variables to zero which influence the solution the least. In general, these variables are the covariances, since variances are often larger. This property is a direct consequence of the Cauchy-Schwarz inequality:

$$|\mathrm{cov}\,(X,Y)| \le \sqrt{\mathrm{var}(X)} \cdot \sqrt{\mathrm{var}(Y)} \le \max\{\mathrm{var}(X), \mathrm{var}(Y)\}. \tag{15}$$

  Another example of setting covariances to zero is to assume that even and odd diffraction orders are not correlated. The reasoning behind this assumption is that for a perfectly symmetric unsegmented XPA 8.0 μm mark (which is used in the RS experiment), there are no even diffraction orders. When the marks are nearly symmetric, the even diffraction orders have a low signal strength, which indicates that the signals are significantly affected by noise (see Chapter 5). Setting all covariances to zero is possible for all $n \ge 3$. Setting all covariances between odd and even orders to zero is valid only for $n \ge 4$. The assumption of no correlation between shifts of diffraction orders and the assumption of no correlation between shifts of even and odd diffraction orders still result in setting more covariances to zero than required when $n \ge 4$ and $n \ge 5$, respectively.

  In Figure 14 the Order-to-Order outcomes of setting all covariances to zero and setting only the covariances to zero which involve an even and an odd diffraction order, are presented for the RS experiment in the case there are 7 diffraction orders. As the figure shows, both solutions roughly follow the exact variance, which is computed directly from the 3M data, but the computed variance and the exact variance do not result in the same least varying diffraction order due to process effects.

  As indicated in Chapter 3 the production like data set could be used in case of the W-CMP

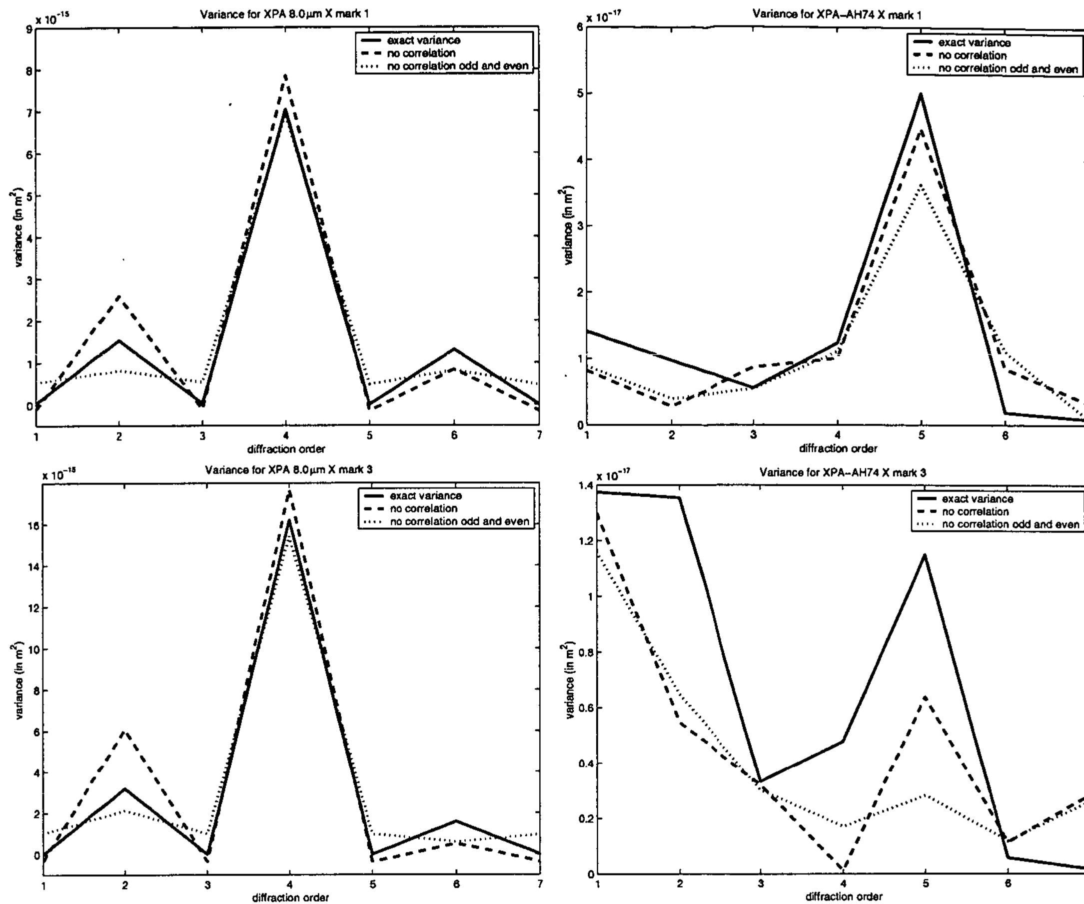


Figure 14: *Results of setting all covariances to zero and setting only covariances between odd and even orders to zero. All results are compared to the exact variance. The left column gives the results for the XPA 8.0 μm marks on two different positions, while the right column does the same, but for the XPA-AH74 mark.*

experiment. Figure 15 shows some results for different types of marks. As can be seen, the exact 3M variance and the computed variance with both 3M data and production like data are not the same. This is an indication that the covariances are absolutely not equal to zero in the W-CMP experiment. However, the two computed solutions are approximately the same as they should be, since the wafer-to-wafer variances of the SbO's are approximately the same.

In the remainder of this chapter only the RS experiment is considered, since the assumptions made in this chapter are theoretically verified in Chapter 5. Note, that for the RS experiment only 3M data could be used (see Chapter 3).

- **Set only n covariances to zero**
  A difficulty arises when exactly n covariances are set to zero. Consider the coefficient matrix A of system (8) of which the rank is equal to the number of rows, namely $\binom{n}{2}$. If n covariances

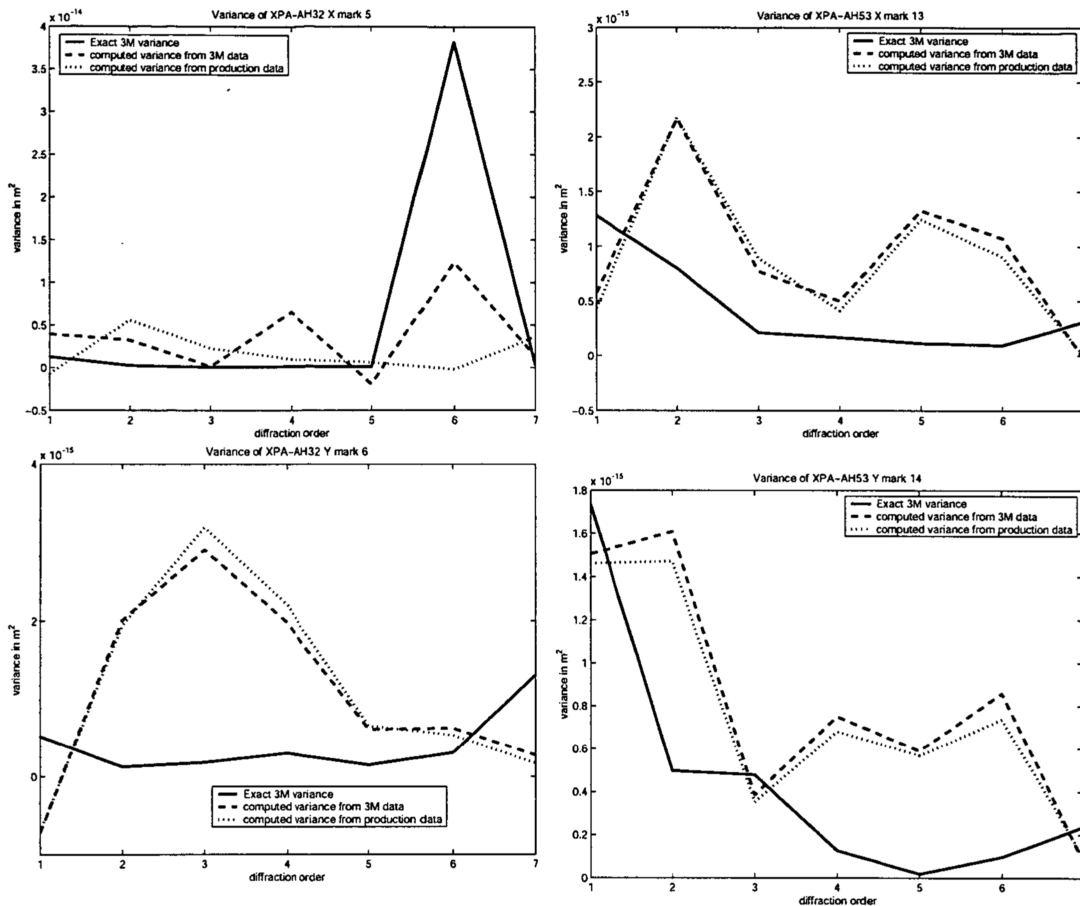


Figure 15: *Presentation of the results of the assumption of no covariances for the W-CMP process. The left column represents the XPA-AH32 marks, while the right column represents the XPA-AH53 marks. The top row are X-marks and the bottom row are Y-marks. For each mark, the exact 3M solution is given and for both data sets the variance is computed with the Order-to-Order method.*

are set to zero, matrix A has n rows with only variances as unknowns. These n rows can make matrix A dependent. This is very important to keep in mind, since a dependent system results in infinitely many solutions. A question is how to find a coefficient matrix A with full rank (all columns/rows are independent). This seems trivial, but when one has n diffraction orders, there are $\binom{\binom{n}{2}}{n}$ possibilities to set n covariances to zero. To give an indication of how large these numbers are and how many of these matrices are of full rank, consider Table 1.

- **Select a combination of covariances to set to zero**
  The question remains which covariances should be set to zero. With a brute force technique (try all 116280 different matrices in case of 7 diffraction orders) a solution is obtained which satisfies the following requirements:
  - the resulting coefficient matrix has full rank;

♯ diffraction orders	♯ possible matrices	♯ full rank matrices
4	15	12
5	252	162
6	5005	1525
7	116280	31406

Table 1: *Indication of the number of possible SbO combinations to obtain a full rank coefficient matrix, compared to the total number of SbO combinations.*

* the solution is closest to the exact solution.

"Closest" is defined by means of the Euclidean norm $\|\hat{x} - x\|$ of the difference between the computed solution $\hat{x}$ and the exact solution x, which is only available when 3M data can be obtained. If this norm is minimal than the solution is said to be closest to the exact

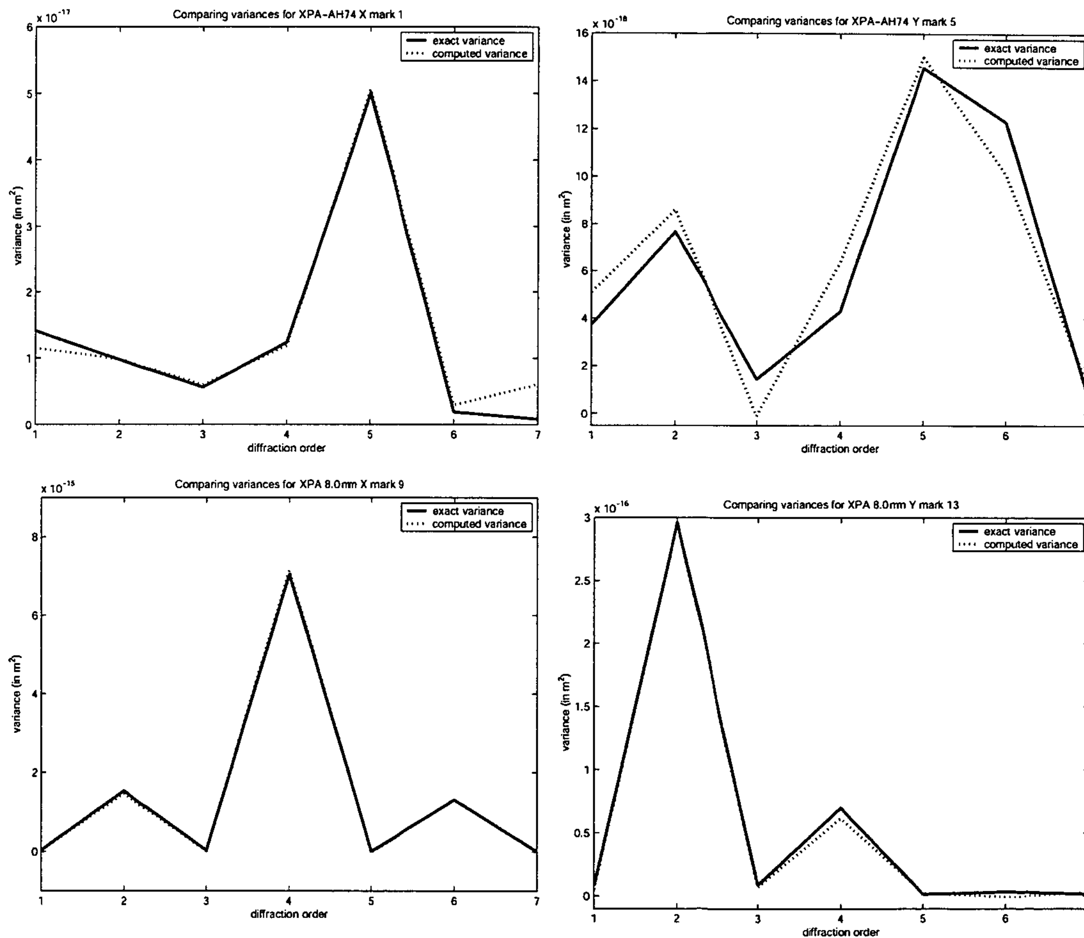


Figure 16: *The computed variances compared to the exact variances for four different marks. The top row are XPA-AH74 marks and the bottom row are XPA 8.0 μm marks. The left column are X marks, while the right column are Y marks.*

mark no	covariances set to zero						
1	cov (τ_1, τ_4)	cov (τ_2, τ_3)	cov (τ_2, τ_6)	cov (τ_3, τ_4)	cov (τ_3, τ_6)	cov (τ_4, τ_7)	cov (τ_5, τ_7)
2	cov (τ_1, τ_2)	cov (τ_1, τ_4)	cov (τ_2, τ_3)	cov (τ_3, τ_6)	cov (τ_4, τ_6)	cov (τ_5, τ_6)	cov (τ_6, τ_7)
3	cov (τ_1, τ_5)	cov (τ_2, τ_7)	cov (τ_3, τ_5)	cov (τ_3, τ_7)	cov (τ_4, τ_7)	cov (τ_5, τ_7)	cov (τ_6, τ_7)
4	cov (τ_1, τ_7)	cov (τ_2, τ_3)	cov (τ_3, τ_4)	cov (τ_3, τ_5)	cov (τ_3, τ_6)	cov (τ_3, τ_7)	cov (τ_4, τ_6)

Table 2: *Table of covariances that are set to zero to obtain the closest variance compared to the exact variance.*

solution. In Figure 16 this solution is given for four marks of the RS experiment, knowing a XPA-AH74 X- and Y-mark and a XPA 8.0 μm X- and Y-mark. Note that the scale on the vertical axis is much larger for the XPA 8.0 μm marks than for the XPA-AH74 marks. It can be concluded that the diffraction orders do not behave independently, since the computed variance differs from the exact variance. Although the computed variance approximates the exact variance much better than in Figure 14, the best order is not identified. This indicates that the assumption that there are n covariances which are negligibly small, does not hold. However, the differences between the computed variance and the exact variance are small. If two diffraction orders both vary in the same way, it is likely that the assumptions made to obtain a solution, determines which of the two diffraction orders is the least varying, although both orders are good orders to do the alignment with.

– **Position dependency of the combination of covariances**
In Figure 17 four solutions are plotted for four marks of the same type, namely XPA-AH74 X marks, but on different positions on the wafer to illustrate if the combinations of covariances set to zero is position dependent. Table 2 gives a summary of which covariances are set to zero for each mark of Figure 17 to obtain the closest variance. It can be seen that, although all four marks are of the same type of mark (XPA-AH74 X mark), different covariances are set to zero.

As can be seen, the set of covariances differs from mark to mark, which indicates that either the covariances lie very close to each other or the diffraction orders react very differently on other positions on the wafer. Figure 18 shows the variance obtained for mark 1. The combination of covariances which are set to zero to compute this variance, is used to compute the solutions for mark 2, 3 and 4. As can be seen from Figure 18 the covariances are not approximately the same, since the combination of covariances for mark 1 does not result in a good approximation of the exact variances for mark 2, 3 and 4. This figure gives the results when setting the optimal combination of covariances derived for mark 1 to zero, but analogously, this procedure could be repeated for all four marks and the results are similar. Apparently, process effects are different on other positions. To explain this, note that the mark is etched in the x- or y-direction and since the resist spin process could be considered to be axi-symmetric but dependent on the radius, the effect of such a process is different for each mark position.

– **Select the wrong combination of covariances**
Since it is not known which covariances should be set to zero, an example is presented to give an idea of how erroneous the computed variance can be. In Figure 19 the variance is given of which the norm of the difference between the obtained variance and the exact

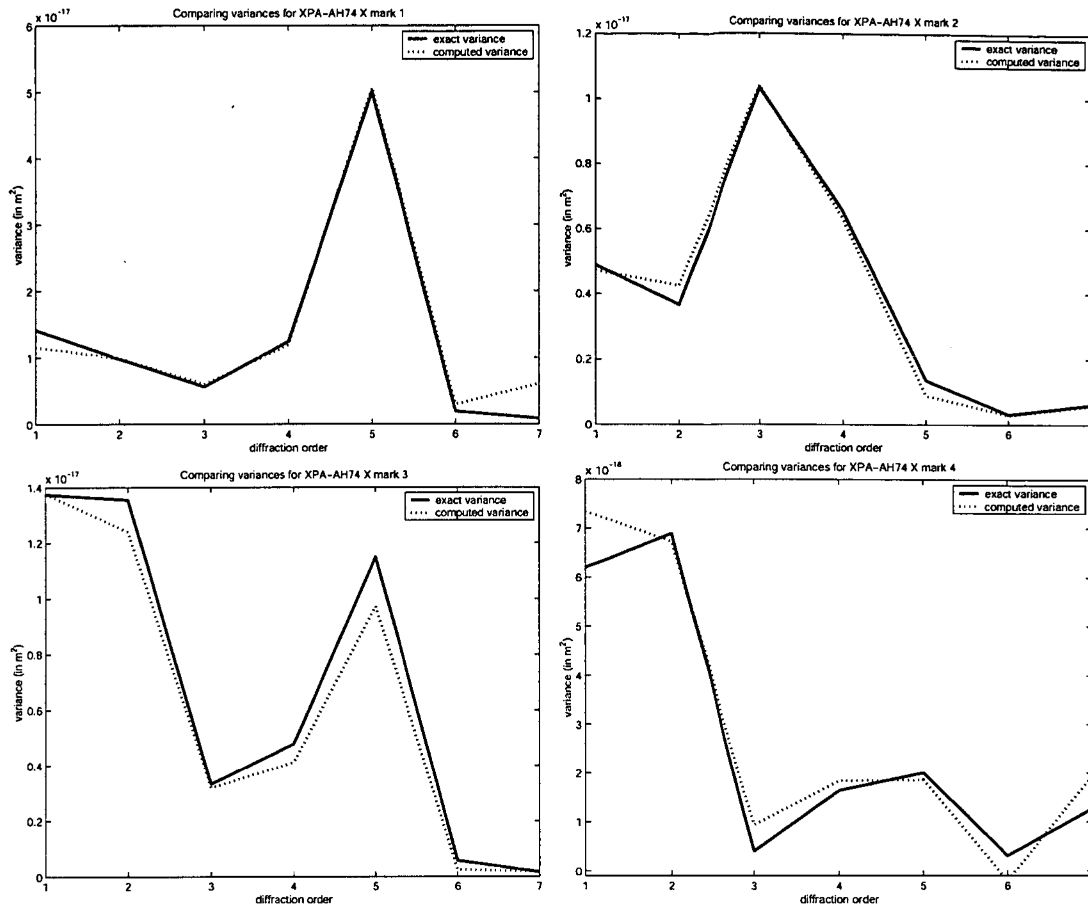


Figure 17: *The computed variance for the mark type XPA-AH74 X marks for four different mark positions compared to the exact variance.*

variance is maximal, and this variance is compared to the exact variance. Some of the resulting variances are even negative, which is a contradiction to the definition of a variance. According to the definition a variance is always nonnegative.

- **Condition of the resulting system**
  Even when the correct n covariances are set to zero, the input vector $\boldsymbol{b}$ from system (8) (the vector consisting of SbO's) determines the solution. The resulting matrix has full rank and is a square matrix. Therefore, the inverse of this matrix exists. In that case, it becomes important to know how the system is conditioned to determine how sensitive the system is to perturbations in vector $\boldsymbol{b}$. A system is called ill conditioned or badly conditioned if a small relative error in data causes a large relative error in the computed solution, regardless of the solution method [10]. To indicate whether the problem is ill or well conditioned, the condition number is introduced. This number gives an indication of the relative error in the solution when a small permutation is applied on the input data. The condition number of a matrix A is defined as:

$$\mathrm{cond}(A) = \|A\| \left\|A^{-1}\right\|, \tag{16}$$

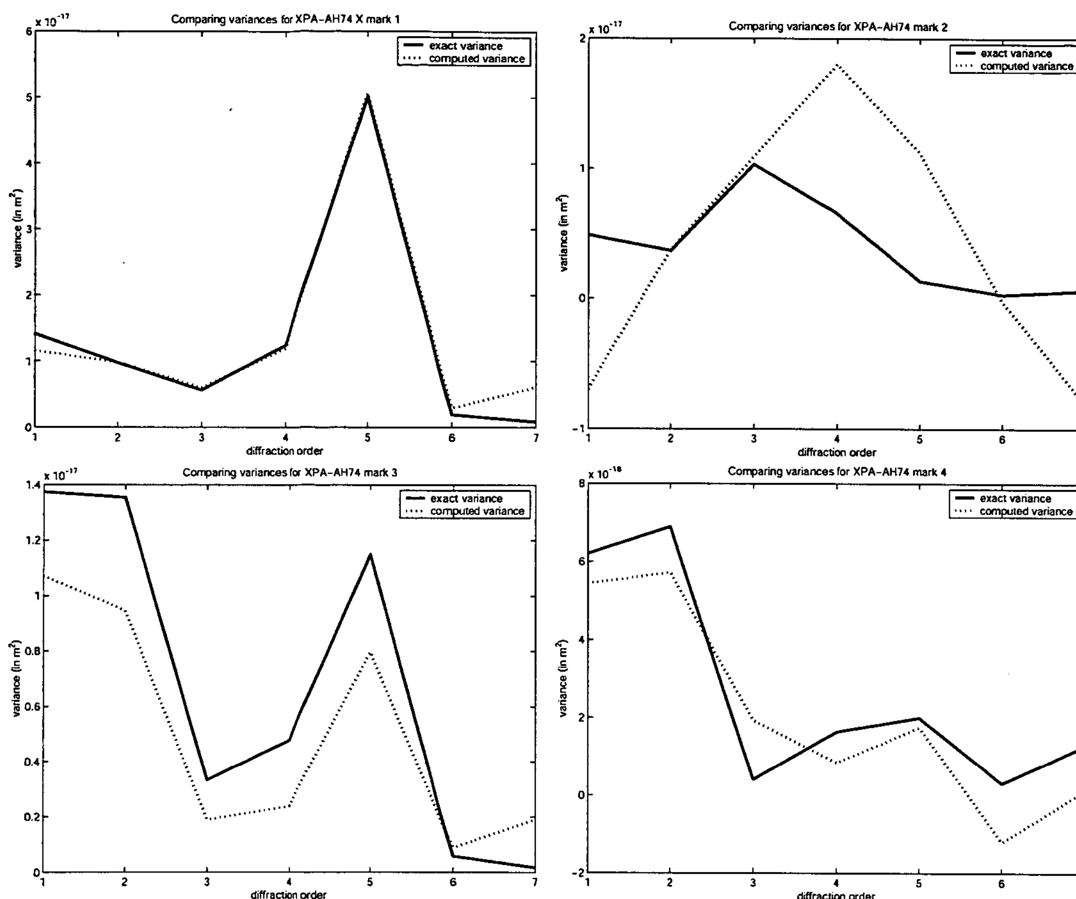


Figure 18: *The computed variance for the XPA-AH74 X marks obtained by using the optimal combination of covariances of mark 1 for all four marks, compared to the exact variance for all four marks.*

where $\|.\|$ is the Euclidean norm or 2-norm. Note, that the condition number is a property of the system, and is independent of the input vector $\boldsymbol{b}$.

Since matrix A is a coefficient matrix which consists of constant entries, it is useless to investigate perturbations in this matrix. Instead, what is of importance are the perturbations in the vector consisting of the computed SbO's, which is the input of the Order-to-Order method. Consider system (9), but now with a full rank square matrix A. If, in the linear system $A\boldsymbol{x} = \boldsymbol{b}$, $\boldsymbol{\delta b}$ and $\boldsymbol{\delta x}$ are the perturbations of $\boldsymbol{b}$ and $\boldsymbol{x}$, respectively, and $\boldsymbol{b} \neq \boldsymbol{0}$, then

$$\frac{\|\boldsymbol{\delta b}\|}{\operatorname{cond}(A)\,\|\boldsymbol{b}\|} \leq \frac{\|\boldsymbol{\delta x}\|}{\|\boldsymbol{x}\|} \leq \operatorname{cond}(A)\frac{\|\boldsymbol{\delta b}\|}{\|\boldsymbol{b}\|}. \tag{17}$$

For any nonsingular matrix A it holds that [11]:

$$1 = \|I\| = \left\|A \cdot A^{-1}\right\| \leq \|A\| \cdot \left\|A^{-1}\right\| = \operatorname{cond}(A). \tag{18}$$

When in system (9) n covariances are set to zero, the upper and lower boundary of the condition number are given for each number of diffraction orders in Table 3.

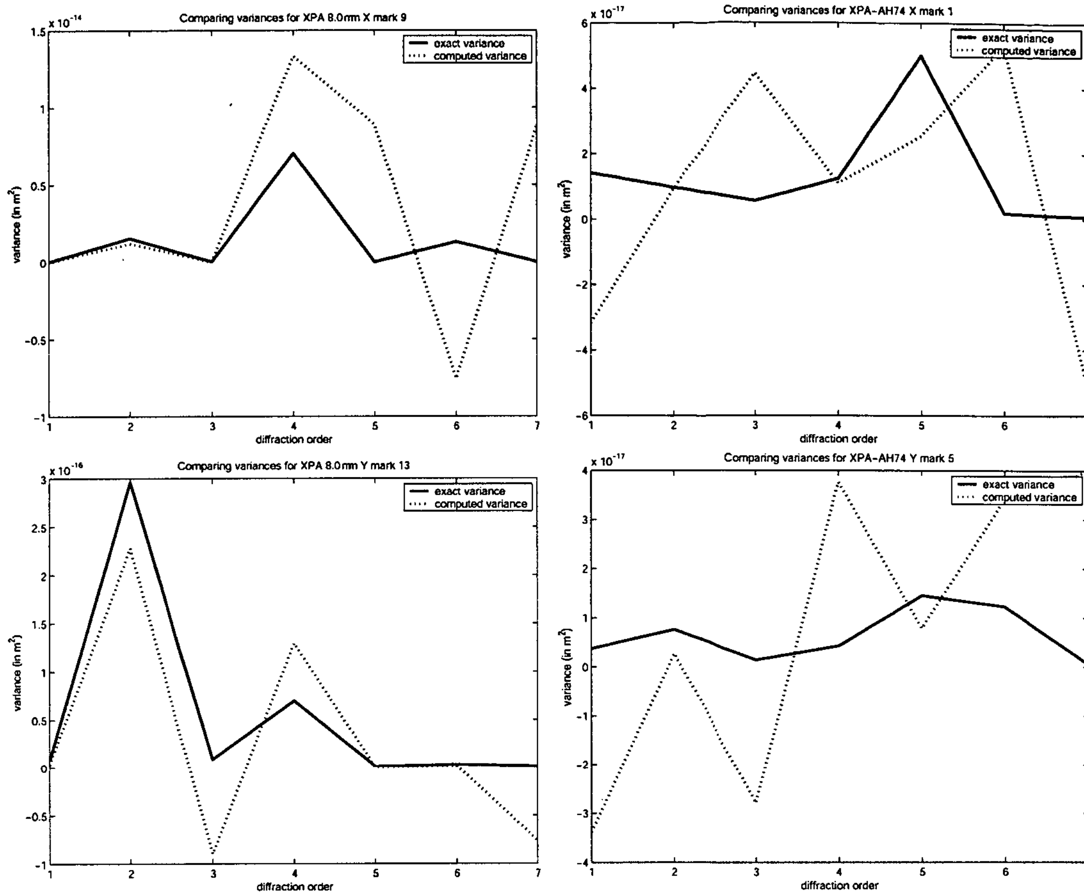


Figure 19: *An example of how bad the computed variance can be compared to the exact variance for two types of marks (XPA-AH74 X and XPA 8.0 μm X marks) one two different positions. In the left column the XPA 8.0 μm marks are given, while in the right column the XPA-AH74 marks are presented.*

♯ diffraction orders	minimum condition number	maximum condition number
4	5.4371	5.4371
5	6.8142	9.0243
6	4.8766	13.9349
7	8.9513	19.5205

Table 3: *Lower and upper boundaries of the condition number for different number of diffraction orders.*

Four observations could be made from Table 3. The first observation is that the obtained system is well conditioned, since condition numbers of order 10 are normally considered to be small. The second observation is that when the number of diffraction orders grows the condition number also becomes larger. The third observation is that the larger the number

of diffraction orders is, the larger the range in condition numbers is. The last observation is that the dependency between rows always involves an even number of rows. Therefore, if 6 or 7 diffraction orders are considered, the condition numbers are somewhat related. The most important conclusion of these computed condition numbers is that the problem is well conditioned and deviations in the vector consisting of variances of SbO's results in small deviations in the solution vector.

As a final remark on condition numbers, note that the closest solution in Figure 17 does not necessarily have the smallest condition number. This could be concluded since all four marks of this figure have a different optimal set of covariances and also a different condition number.

– **Final remarks**
Note that in this master's thesis the exact variance is available, but in practice, this variance is unknown. It is nice to know that if only n covariances are set to zero the solution can approximate the exact variance, but it is also shown in Figure 19 that the computed solution could be very wrong, even compared to the computed variance when assuming there is no correlation between all diffraction orders. Therefore, it is very important to know which covariances you must choose to set to zero before conclusions are made about the outcomes of such an assumption. In Chapter 5 simulations are carried out for the RS case to show how diffraction orders react to asymmetries due to this process to have a better insight how shifts of diffraction orders correlate.

4.1.2 Set a variance to zero

Another method to solve the problem is to assume that the variance of one diffraction order is equal to zero. The idea behind this approach is that the diffraction order which varies the least due to processing has the lowest variance. In comparison to other diffraction orders, this order yields the lowest Euclidian norm for the difference between vectors of the computed and the exact (co)variances. If it is assumed a variance of a diffraction order is equal to zero, this physically means that the involved diffraction order is constant. In turn, this implies that all covariances which include this diffraction order also have to be zero, since a constant cannot correlate with a variable. Thus, setting one variance to zero results in setting $n-1$ covariances to zero, giving us the n unknown. The resulting coefficient matrix A is of full rank. The results of this experiment are given in Figure 20. One can see from this figure that when the right diffraction order is chosen to set the variance to zero (the diffraction order for which the exact variance is minimal), the computed solution follows the exact variance best. However, the problem is again to find out which variance vary least due to processing, because without special experiments no exact variance is available.

To conclude section 4.1, it can be stated that the Order-to-Order method does not identify the same diffraction order as the exact variance does, when n unknowns are pinpointed to zero. However, when the correct covariances are set to zero, the differences between exact and computed variances become insignificant. Still, choosing a set of n variables requires a sufficient insight in the process effects on alignment marks. Anyhow, assumptions must be made about process effects on diffraction orders. This indicates that when the assumptions do not hold for a specific mark, the Order-to-Order method is likely to find the wrong diffraction order.

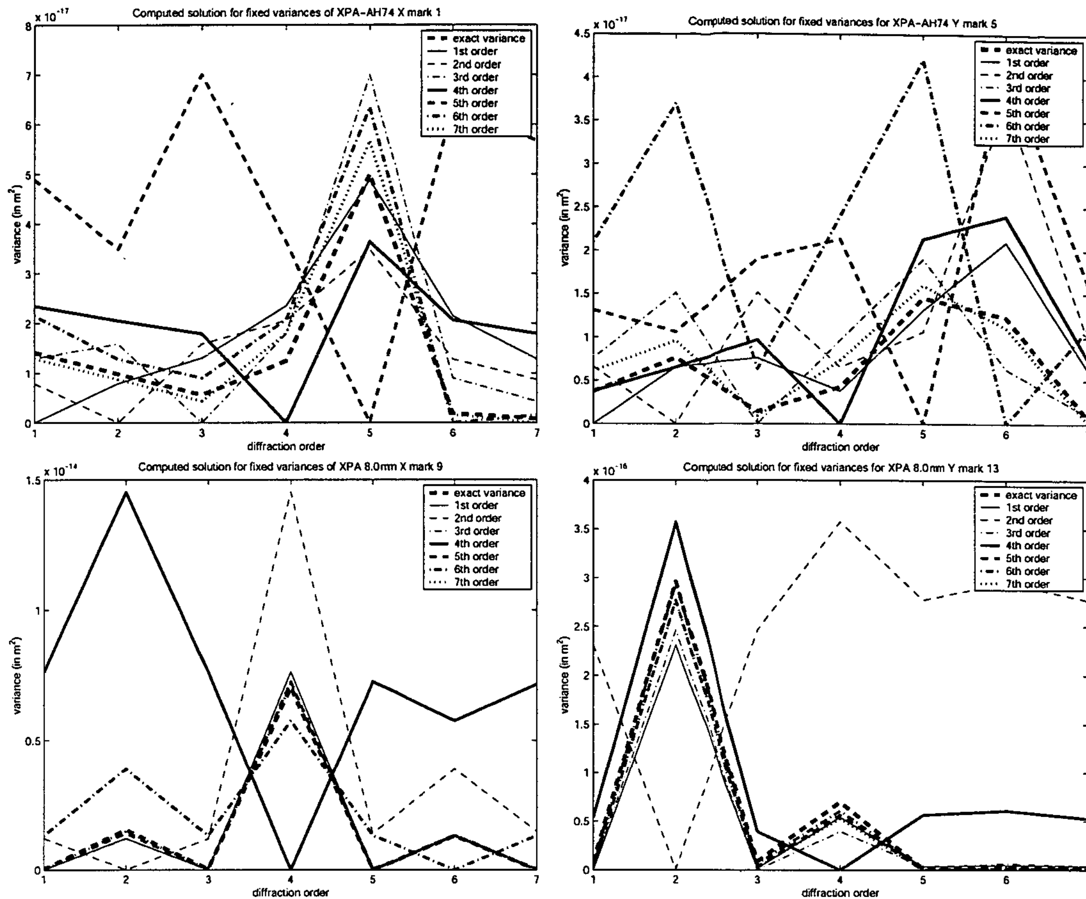


Figure 20: *Results of the simulations where one variance of a certain diffraction order is set to zero.*

4.2 Add extra equations

It is important to know that only APD's can be used as input for the Order-to-Order method. WQ's are used as an indication of the signal strength of a measured APD and in the Order-to-Order method they are only used to select signals with sufficient signal strength. The choice of defining a SbO is logical, since it removes the diffraction order independent contribution. A method to include additional equations is to keep the diffraction order independent part and sum two APD's or use direct variances and covariances of APD's. The basic formula is given by

$$\mathrm{var}\left(\alpha X + \gamma Y\right) = \alpha^2 \,\mathrm{var}\left(X\right) + \gamma^2 \,\mathrm{var}\left(Y\right) + 2\alpha\gamma \,\mathrm{cov}\left(X, Y\right). \tag{19}$$

Implementation gives the following set of equations:

$$\begin{aligned}
\mathrm{var}\left(Y^a\right) &= \mathrm{var}\left(\tau_a+\beta\right) \\
&= \mathrm{var}\left(\tau_a\right)+\mathrm{var}\left(\beta\right)+2\mathrm{cov}\left(\tau_a,\beta\right), \qquad (20) \\
\mathrm{cov}\left(Y^a,Y^b\right) &= \mathrm{cov}\left(\tau_a+\beta,\tau_b+\beta\right) \\
&= \mathrm{cov}\left(\tau_a,\tau_b\right)+\mathrm{cov}\left(\tau_a,\beta\right)+\mathrm{cov}\left(\tau_b,\beta\right)+\mathrm{var}\left(\beta\right), \qquad (21) \\
\mathrm{var}\left(Y^a+Y^b\right) &= \mathrm{var}\left(\tau_a+\tau_b+2\beta\right) \\
&= \mathrm{var}\left(\tau_a\right)+\mathrm{var}\left(\tau_b\right)+2\mathrm{cov}\left(\tau_a,\tau_b\right)+4\mathrm{var}\left(\beta\right) \\
&\quad +2\mathrm{cov}\left(\tau_a,\beta\right)+2\mathrm{cov}\left(\tau_b,\beta\right). \qquad (22)
\end{aligned}$$

By adding these equations, $n+1$ variables are added, knowing the variance of the diffraction order independent part of an APD and n covariances between the PIAS part and the order independent part of an APD. In principle, the equations above provide enough equations comparing to the number of variables, but unfortunately, the system is dependent and the rank equals $\binom{n}{2}+n$, which is less than the number of unknowns and the number of equations. This means that the system still has infinitely many solutions.

A refinement was to assume that there is no correlation between the process part and the order independent part of an APD. This reduces the number of unknowns by n. The resulting system still has rank $\binom{n}{2}+n$ and has one more unknown, meaning there are still infinitely many solutions. When the unknown var(β) is given a value, the system is not underdetermined anymore and the rank of the matrix is equal to the number of unknowns. But this answer does not give the real PIAS variances, since it is assumed there is no correlation between the diffraction order independent part and the process part, meaning that the diffraction order independent part must be a constant (also over the wafers). This induces that the answers could just as easily be computed from the measured APD's (corrected for measurement errors). In practice, this assumption does not have to hold, since the order independent part is only constant per mark for every diffraction order, but differs from mark to mark and wafer to wafer.

5 Using the mark shape as prior knowledge

This chapter discusses how Fourier optics can be used to obtain prior knowledge of the mark shape. The goal of using mark shape information is not to find the actual values of the covariances of system (8), but to validate the assumption of no correlation used in Chapter 4 and/or to find the n covariances which should be set to zero to obtain the most accurate approximation of the exact variance. Only the RS process is considered, since there is well-defined experimental data available for this process. An extension to the W-CMP process will be part of a follow-up investigation. To have insight in how diffraction orders correlate, the light reflected by the mark topology has been modeled. In general, refraction and reflection should be taken into account (see Figure 21). However, in this chapter the Vertical Propagation Model is used, which is a zero-order approximation of the reflected so-called near-

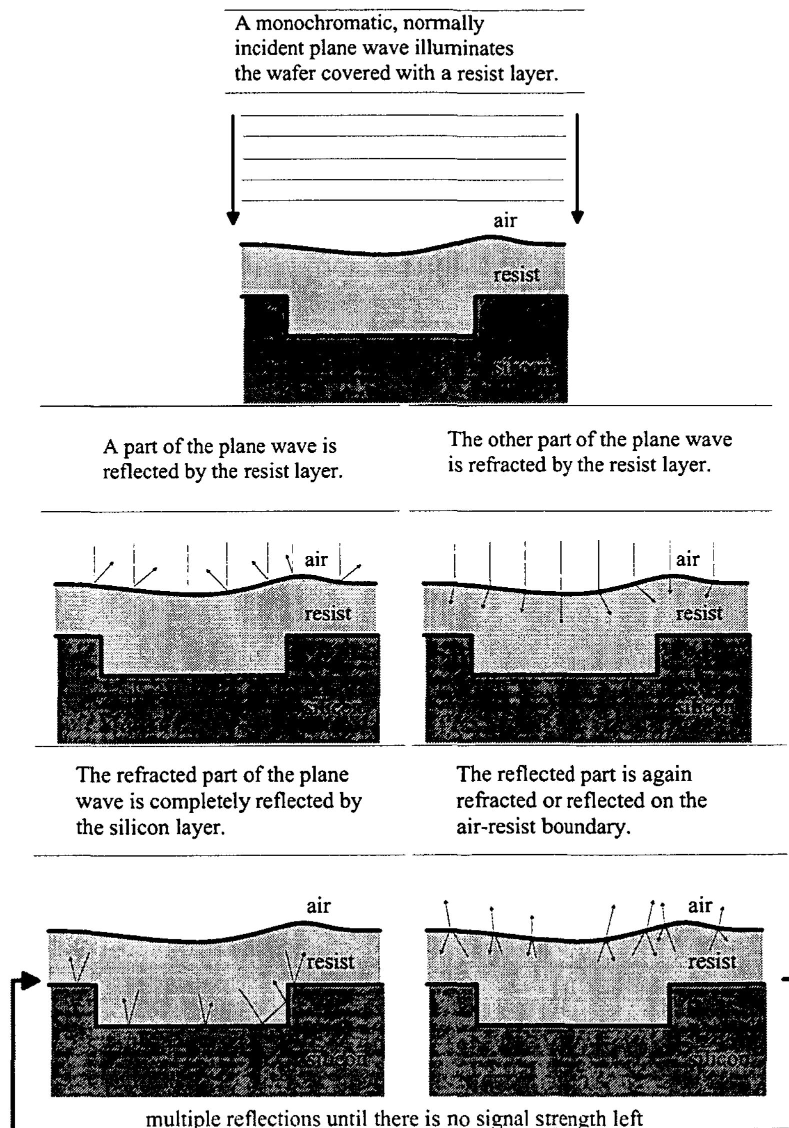


Figure 21: *Schematic representation of the actual illumination process.*

field. Although this model is oversimplified, its accuracy is adequate enough for an initial investigation on the added value of using mark shape information as prior knowledge for the Order-to-Order method.

5.1 Vertical Propagation Model for the RS process

In this section the Vertical Propagation Model is discussed to model the process effects in case of the RS process.

5.1.1 Assumptions of the Vertical Propagation Model

The Vertical Propagation Model assumes that the thickness of a resist layer is small and one period of a mark is large compared to the wavelength of the incoming plane wave (red or green light). Under these limiting cases, the reflected near-field, which is the field just above the mark topology, can be approximated by a simple analytical expression that is derived in the following section.

The thickness of the resist layer used for the experiment discussed in Chapter 2, is 0.7 μm, which is of the same order of magnitude as the wavelength (633 nm for the red light and 532 nm for the green light). Therefore, the first assumption does not hold. Fortunately, the requirement on resist thickness is not very strict. Moreover, one period is much larger than the wavelength and this requirement is often more important.

5.1.2 Derivation of the near-field for one mark segment

Figure 22 shows the definitions for the derivation of the Vertical Propagation Model. Since the thickness of the resist layer is assumed to be small compared to the wavelength, lateral deviation of the propagating beam are small and therefore neglected. Hence, the Vertical Propagation Model assumes that the monochromatic normally-incident plane wave only propagates in the z-direction.

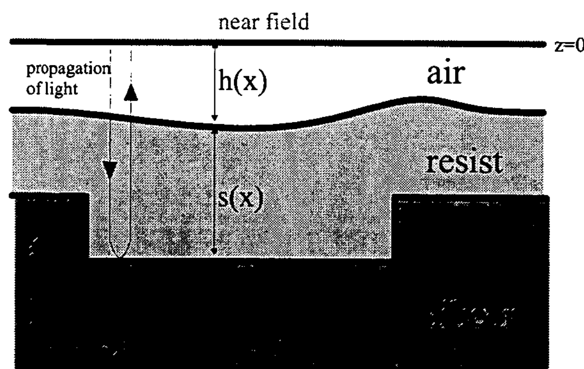


Figure 22: *Definition of parameters for the Vertical Propagation Model. The figure shows one period of an alignment grating covered with resist.*

A monochromatic plane wave can mathematically be written in complex notation ($j^2 = -1$) as

$$\psi(z) = \mathrm{Re}\left\{A\, \exp\left(-jkz + \varphi_0\right)\right\}, \tag{23}$$

where the initial phase φ_0 equals zero and k is the propagation number, defined as

$$k = \frac{2\pi}{\lambda}. \tag{24}$$

The parameter λ is the wavelength. This plane wave travels a distance $h(x)$ through air, before it is partly reflected back at the air-resist interface and partly transmitted into the resist layer. The reflected part travels again the distance $h(x)$ back. The resulting part of the near-field caused by the reflection at the air-resist interface is:

$$\text{reflected field} = -rA\exp[-j\frac{4\pi}{\lambda}n_{\text{air}} \cdot h(x)]. \tag{25}$$

The term r is the reflection coefficient given by the Fresnel equations [12]:

$$r = \frac{n_{\text{resist}} - n_{\text{air}}}{n_{\text{air}} + n_{\text{resist}}}, \tag{26}$$

where n_{air} and n_{resist} are the refractive indices of air and resist, respectively $n_{\text{air}} = 1$ and $n_{\text{resist}} = 1.62$, so $r = 0.24$ [13]. So, 24 % of the incident light is reflected. The refractive index is usually dependent of the wavelength of the light used, but for air and resist the difference between red and green light is small. It is also assumed that resist shows no absorption for the wavelengths used, so the imaginary part of the refraction index of resist is 0.

The remainder of the incident field is transmitted and propagates in the resist towards the silicon substrate. Therefore, the transmission coefficient t is given by

$$t = 1 - r. \tag{27}$$

Substitution of $r = 0.24$ gives $t = 0.76$ which means that 76 % of the original plane wave is transmitted through the resist layer.

The field propagating towards the silicon substrate is partly absorbed and partly reflected at the silicon-resist interface. Since absorption causes a change in the phase of the field, silicon has a complex refraction index n_{silicon}, which is equal to $4.14 + j \cdot 0.045$ in case of the green laser ($\lambda = 532$ nm) and equal to $3.88 + j \cdot 0.019$ in case of the red laser ($\lambda = 633$ nm). Denote the reflection coefficient of the silicon layer by R, then

$$R = \frac{n_{\text{silicon}} - n_{\text{resist}}}{n_{\text{silicon}} + n_{\text{resist}}}. \tag{28}$$

Since R is now a complex number it will also change the phase of the reflected wave. For the green laser $R = 0.4375 + j \cdot 0.0044$ and for the red laser $R = 0.4109 + j \cdot 0.0020$, which means that for green light only 44 % is reflected and for red light only 41 % is reflected.

In principle, multiple reflections can be easily included. However, an exact modeling of multiple reflections makes no sense since the Vertical Propagation Model already assumes that the path traversed through the air and resist layer is short. Even if the 24 % of the wave coming from the silicon substrate, is reflected back into the resist layer is taken into account, the intensity of this reflected wave is approximately 7 à 8% of the original incoming plane wave and this part contributes only 2 à 3 % to the total near field (the wave must be reflected by the silicon-resist interface and transmitted by air-resist layer. Therefore, no multiple reflections are taken into account.

The near-field (see Figure 22) can now be represented as follows:

$$\text{near field} = \text{reflected field} + \text{transmitted field}$$
$$E_{\text{near}}(x) = -rA\exp\left[-j\frac{4\pi}{\lambda}n_{\text{air}}\cdot h(x)\right] - Rt^2A\exp\left[j\frac{4\pi}{\lambda}\left(n_{\text{air}}\cdot h(x) + n_{\text{resist}}\cdot s(x)\right)\right], \quad (29)$$

with $r = 0.24$, $t = 0.76$ and $R = 0.4375 + j\cdot 0.0044$ for green light or $R = 0.4109 + j\cdot 0.0020$ for red light. This near field is only defined on one mark segment, so for $-L/2 \le x < L/2$, where L is the length of the segment.

5.1.3 Near-field for an infinitely large grating

The derived expression for the near-field is the field for only one segment. Outside the mark segment the near-field is assumed zero:

$$E_{\text{near}}(x) = 0, \quad |x| > L/2. \quad (30)$$

Consider the case where the mark segment that is shown in Figure 22 is repeated periodically. To have an expression for the total near field, continue the derived near-field periodically. To take infinitely many mark segments is only an assumption to simplify the computations, since the diffraction orders are delta functions. If a finite number of mark segments are taken into account, the far-field consists of sinc functions. Since only the tops of the sinc functions are searched for, the number of mark segments may be infinite. Mathematically, the periodical repetition is given by:

$$E_{\text{neartot}}(x) = \sum_{m=-\infty}^{\infty} E_{\text{near}}(x - mL), \quad -\infty < x < \infty. \quad (31)$$

To determine the diffraction orders, which are measured with the alignment sensor, the far-field is needed. The far-field can be computed according to the following formula [12]:

$$E_{\text{fartot}}(\omega) = \mathcal{F}\left\{E_{\text{neartot}}(x)\right\} = \int_{-\infty}^{\infty} E_{\text{neartot}}(x)\exp[-j\omega x]dx, \quad (32)$$

where $\mathcal{F}$ is the Fourier transform. Application of the space-shift theorem [14] and the periodicity of the near-field, give:

$$\begin{aligned}
E_{\text{fartot}}(\omega) &= \mathcal{F}\left\{\sum_{m=-\infty}^{\infty} E_{\text{near}}(x - mL)\right\} \\
&= \int_{-\infty}^{\infty}\sum_{m=-\infty}^{\infty} E_{\text{near}}(x - mL)\exp[-j\omega x]dx \\
&= \left(\int_{-L/2}^{L/2} E_{\text{near}}(x')\exp[-j\omega x']dx'\right)\sum_{m=-\infty}^{\infty}\exp[-j\omega mL], \quad \text{where } x' = x - mL \\
&= \mathcal{F}\left\{E_{\text{near}}(x')\right\}\sum_{m=-\infty}^{\infty}\exp[-j\omega mL] \\
&= L\sum_{m=-\infty}^{\infty}\delta(\omega - mL)E_{\text{far}}(\omega) \\
&= L\sum_{m=-\infty}^{\infty}\delta(\omega - mL)E_{\text{far}}(mL). \qquad (33)
\end{aligned}$$

As can be seen, the far-field can be expressed in terms of the Fourier transform of the near-field. In general, it is nearly impossible to find a simple analytical expression for E_{far}. Therefore, a numerical approximation is computed, using the discrete Fourier transform [15]. Suppose that for each segment a set of N samples are chosen:

$$E_{\text{near},k} := E_{\text{near}}(x_k), \quad x_k = k\Delta x, \quad k = 0, 1, ..., N-1, \tag{34}$$

where Δx is the sample period in the x-direction. The discrete Fourier transform then reads as follows

$$E_{\text{far},n} \approx \sum_{k=0}^{N-1} E_{\text{near},k} e^{j\omega_n x_k}, \tag{35}$$

where

$$E_{\text{far},n} = E_{\text{far}}(\omega_n), \quad \omega_n = n\Delta\omega, \quad \Delta\omega = \frac{2\pi}{L} \quad n = 0, 1, ..., N-1. \tag{36}$$

Because there are only N input values to the summation, only N independent values of the transform can be expected.

The discrete Fourier transform is significantly faster when the number of points that are necessary to discretize the near-field, is a power of 2 [16]. In that case, the Fast Fourier Transform can be used.

5.1.4 Derivation of the shift a diffraction order

The position of a mark can be computed from the phase of a diffraction order. The alignment signal that is measured with the ATHENA sensor is given by:

$$\begin{aligned} I(n) &= \left|E_{\text{far}}(n) + E_{\text{far}}(-n)\right|^2 \\ &= \left|E_{\text{far}}(n)\right|^2 + \left|E_{\text{far}}(-n)\right|^2 + 2\left|E_{\text{far}}(n)\right|\left|E_{\text{far}}(-n)\right|\cos\left(\theta(n) - \theta(-n)\right), \end{aligned} \tag{37}$$

where $\theta(n)$ is the measured phase of diffraction order n. To know how much a mark seems to be shifted due to the alteration in shape of the mark, it is assumed that the mark is symmetric. Define the actual phase of the far-field of diffraction order n by $\varphi(n)$. When a phase grating is symmetric ($\varphi(n) = \varphi(-n)$) and the grating is shifted by Δx, the far-field could be derived by using the shift-theorem:

$$\mathcal{F}\left\{E_{\text{near}}(x - \Delta x)\right\} = \mathcal{F}\left\{E_{\text{near}}(x)\right\} \exp[-jk\Delta x]. \tag{38}$$

Now consider diffraction order n. Then

$$E_{\text{far}}(n) = \mathcal{F}\left\{E_{\text{near}}(x)\right\} \exp[-j\frac{2\pi n}{L}\,\Delta x], \tag{39}$$

where L is the length of one grating period.

The measured phase $\theta(n)$ of the far-field is decomposed in

$$\theta(n) = \varphi(n) - \frac{2\pi n}{L}\,\Delta x. \tag{40}$$

The ATHENA sensors determines the position of the mark by searching for the top of the cosinus in equation (37). When this top is found, the cosine has value 1, and therefore $\theta(n) - \theta(-n) = 0$. This gives the desired formula for the shift of the mark.

$$\theta(n) - \theta(-n) = \varphi(n) - \varphi(-n) - \frac{4\pi n}{L}\,\Delta x \quad \Leftrightarrow \Delta x = \frac{L}{4\pi n}\left(\varphi(n) - \varphi(-n)\right). \tag{41}$$

Unfortunately, there are definitely no symmetric marks. Therefore, $\varphi(n) \neq \varphi(-n)$ and Δx indicates the distance which the mark would have been shifted if it had been symmetric. This Δx is now different for each n and therefore denoted by $\Delta x(n)$. The difference in phase between the orders $-n$ and n is enough information to compute the alignment shift $\Delta x(n)$ for diffraction order n:

$$\Delta x(n) = \frac{L}{4n\pi} \left(\varphi(n) - \varphi(-n)\right). \tag{42}$$

Note that $\varphi(n)$ and $\varphi(-n)$ can be determined except for a multiple of 2π. However, these phases do not have to be computed separately, only their difference $\theta(n) - \theta(-n)$ must be determined and this is given by

$$\varphi(n) - \varphi(-n) = \arg\left(E_{\text{far}}(n) \cdot E^*_{\text{far}}(-n)\right). \tag{43}$$

To determine $\Delta x(n)$ the far-field must be known. Without prescribing $h(x)$ and $s(x)$ (see Figure 22) the far-field could not be computed. Therefore, in the following section an example is given to model the air and the resist layer.

5.2 Determine the shape of the resist layer

Because of the shape of an ideal mark, the resist layer shows a dip followed be a pile-up effect [17]. Since all periods on the periodic grating are assumed to be the same, also the resist layer will be the same for each period. Therefore, consider only one period. The shape of the resist is modeled by three cosines. The schematic representation of the modeling is given in Figure 23. The mathematical

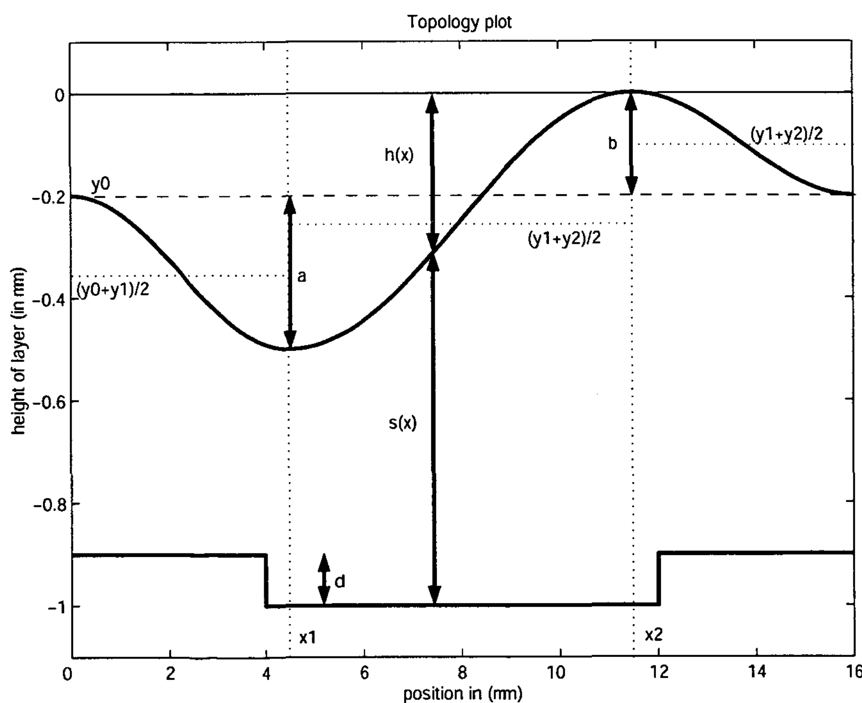


Figure 23: *Model of resist layer.*

formula for the resist shape is given by:

$$f(x) = \begin{cases} \frac{y_0+y_1}{2} + \frac{y_0-y_1}{2}\cos\left(\frac{\pi x}{x_1}\right) & ,x_0 \le x < x_1 \\ \frac{y_1+y_2}{2} + \frac{y_1-y_2}{2}\cos\left(\frac{(x-x_1)\pi}{x_2-x_1}\right) & ,x_1 \le x < x_2 \\ \frac{y_2+y_0}{2} + \frac{y_2-y_0}{2}\cos\left(\frac{(x-x_2)\pi}{x_3-x_2}\right) & ,x_2 \le x \le x_3 \end{cases}. \tag{44}$$

The parameters y_0, y_1 and y_2 determine the amplitude of the cosine, while the parameters x_0, x_1, x_2 and x_3 determine the period of the modeled cosine. The limit the degrees of freedom for the simulation $y0$, x_0 and x_3 are given a certain value. Since f is periodic for the period $x_3 - x_0$, the value of f must be the same for x_0 and x_3. If the problem is redefined with variables a and b instead of y_1 and y_2, defining a as the depth of the dip and b the height of the pile-up effect. The expressions for y_1 and y_2 are as follows:

$$y_1 = y_0 - a \tag{45}$$

$$y_2 = y_0 + b. \tag{46}$$

$$\tag{47}$$

Substitute these two equations in equation (44) results in the following set of equations:

$$f(x) = \begin{cases} y_0 - \frac{a}{2} - \frac{a}{2}\cos\left(\frac{\pi x}{x_1}\right) & ,x_0 \le x < x_1 \\ y_0 + \frac{b-a}{2} - \frac{a+b}{2}\cos\left(\frac{(x-x_1)\pi}{x_2-x_1}\right) & ,x_1 \le x < x_2 \\ y_0 + \frac{b}{2} + \frac{b}{2}\cos\left(\frac{(x-x_2)\pi}{x_3-x_2}\right) & ,x_2 \le x \le x_3 \end{cases}. \tag{48}$$

This concludes the description of the top of the resist layer. Note that only four degrees of freedom are admitted in this model for the resist layer, that is a, b, x_1 and x_2. The shape of the alignment mark is given by a depth d and a width w :

$$g(x) = b - d\ \mathrm{rect}\left(2x/w\right). \tag{49}$$

Now the thickness of the air layer ($h(x)$) and the thickness of the resist layer ($s(x)$) can be computed:

$$\begin{aligned} h(x) &= c - f(x) \\ s(x) &= f(x) - g(x) \end{aligned} \tag{50}$$

The obtained functions for $h(x)$ and $s(x)$ can now be substituted in equation (29).

5.3 Wafer quality (WQ) and contrast

To explain the results of Section 5.4, where the diffraction orders are simulated for several mark topologies, the following two quantities are introduced: the Wafer Quality (WQ) and the contrast of a measured signal. Both quantities indicate how "good" a signal is.

The WQ is proportional to the amplitude of the intensity variation. Mathematically, the WQ of diffraction order n is given by:

$$\mathrm{WQ}(n) = \left|E_{\mathrm{far}}(n)\right| \cdot \left|E_{\mathrm{far}}(-n)\right|. \tag{51}$$

The WQ's that are measured in the experiment of Chapter 2 are standardized with the WQ of an ideal mark with optical depth $\lambda/4$ with λ the wavelength of the light that is used. If WQ(n) is small

then $E_{\text{far}}(n)$ or $E_{\text{far}}(-n)$ is small, which causes the signal-to-noise ratio to deteriorate.

The second quantity, the contrast, is used to identify the difference in intensity between the $(-n)^{\text{th}}$ and the nth diffraction order. Mathematically, the contrast is given by:

$$\text{contrast}(n) = \frac{2\left|E_{\text{far}}(n)\right| \cdot \left|E_{\text{far}}(-n)\right|}{\left|E_{\text{far}}(n)\right|^2 + \left|E_{\text{far}}(-n)\right|^2}. \tag{52}$$

According to this definition, the contrast is a value between 0 and 1. If the contrast is equal to 1, then $\left|E_{\text{far}}(n)\right| = \left|E_{\text{far}}(-n)\right|$, which indicates both signals have the same intensity. If it is small, the intensity of one of the signals is small compared to the intensity of the other.

The reason why these quantities are introduced is that Δx (the PIAS) is determined by both the nth and $-n^{\text{th}}$ diffraction order. If the intensity of one of these signals is small (or both are small), noise is likely to spoil the accuracy of the signal. The WQ and contrast together show what the intensity ratio between the nth and the $(-n)^{\text{th}}$ diffraction order is, and are indications of the accuracy of the position shifts.

5.4 Simulations

In this section the simulations carried out with the Vertical Propagation Model are presented. To have a feeling of the outcome of these simulations, a simple example is given with a flat resist layer (5.4.1). In (5.4.2), the resist layer is modeled by four parameters (two for the x-direction and two for the thickness of the resist). This section is concluded by addressing the problems encountered when combinations of simulation variables are taken.

5.4.1 Illustrative example: flat resist surface

This simple example is only presented here to illustrate the way of simulating. The resist layer which covers the perfectly symmetric mark is flat (see Figure 24). This keeps the topology symmetric, which results in a unique definition of the position of the mark. Therefore, negative diffraction orders have the same phase as positive diffraction orders.

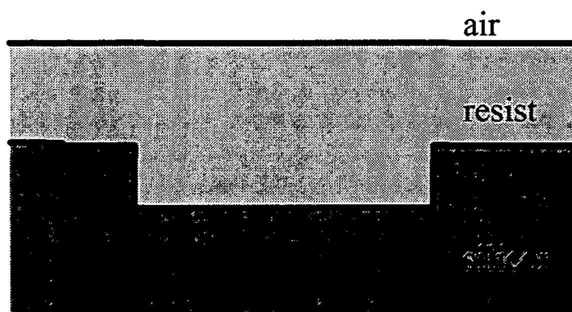


Figure 24: *Schematic representation of the simple case.*

The results of the simulation are given in Figure 25. For this simulation 2^{10} points are used to implement the Fast Fourier Transform. This implies that intervals of approximately 16 nm are used

© ASML 2003

which is small compared to the pitch of a mark segment. The results are as anticipated, since the mark topology is perfectly symmetric and therefore the shift plot shows for each diffraction order that the shift Δx equals 0. The WQ plot shows that even diffraction orders have WQ equal to zero, which should be the case since the mark topology involves a symmetric 50 % duty cycle infinite 8.0 μm mark and such a mark has no even diffraction orders. The odd diffraction orders have decreasing WQ which could be explained from the Fourier series of the shape of the mark. Also the contrast plot gives contrasts equal to one for every diffraction order, which indicates the $+n^{\text{th}}$ and $-n^{\text{th}}$ diffraction order gives the same position.

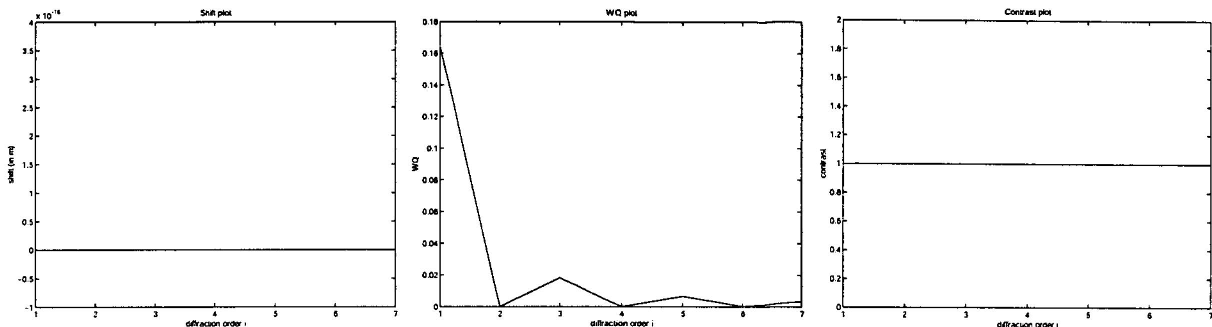


Figure 25: *Results of the test case. The left plot shows the shifts, the middle plot shows the WQ and the right plot gives the contrast for every diffraction order.*

5.4.2 Simulations for separate variables

In this section, four variables are used to simulate RS process effects on the position shift, the WQ and the contrast of the diffraction orders. During the simulations one of the variables is altered and the other three variables are fixed. The four variables are the depth of the dip in the resist layer (a), the height of the pile-up effect (b), the position of the hole (x_1) and the position of the pile-up effect (x_2) (see Figure 23). It is shown that the shifts of all diffraction orders are correlated in the sense that there is a relation between diffraction orders, although it might be a complex one. In the experiments diffraction orders are measured on a mark position over all the wafers of a batch. These marks are assumed to have a similar, but still different, process effect. If these samples are taken out of a set of mark topologies for which the diffraction orders are correlated in a very complex way, the covariance between the measured diffraction orders will become very small. For the mark topologies where the shifts of diffraction orders are nicely correlated (for example by a straight line), the covariances are expected to be large. It is expected that the correlation between diffraction orders is simple as long as the WQ and the contrast of both orders are sufficiently high.

The four simulations are summarized:

- **Variable a**
  As a reference situation, the perfectly symmetric mark topology is taken as given in Figure 24, that is $a = 0$, $b = 0$, $x_1 = 4\,\mu$m and $x_2 = 12\,\mu$m. Variable a is varied from 0 to 100 nm. This means that at position x_1 a dip is created of maximal 100 nm. As soon as $a > 0$ the topology becomes asymmetric. The simulation process and the results for the shifts, WQ's and contrasts are given in Figure 26.

Some trivial characteristics could be noticed from this figure. The reference situation (flat photoresist layer) gives for all diffraction orders a unique position, which is logical because for a symmetric topology, each diffraction order points to the center of gravity. Immediately after altering the variable a (insert asymmetry), the shifts deviate from zero and for every diffraction order this shift is different. The even orders have an abrupt change when the topology becomes asymmetric. This can be explained from the fact that when the topology is symmetric, there are no even diffraction orders since the mark topology can be described with only cosines (odd diffraction orders which follows from Fourier series). But if asymmetry is introduced, even orders start to appear and therefore they are highly sensitive for topology changes. Also the contrast of these orders is very small just after the asymmetry is introduced.

The even orders are alternating in sign, since the shifts of the 2nd and 6th diffraction orders are negative, while the shift of the 4th diffraction order is positive. Also, the abrupt change in shift

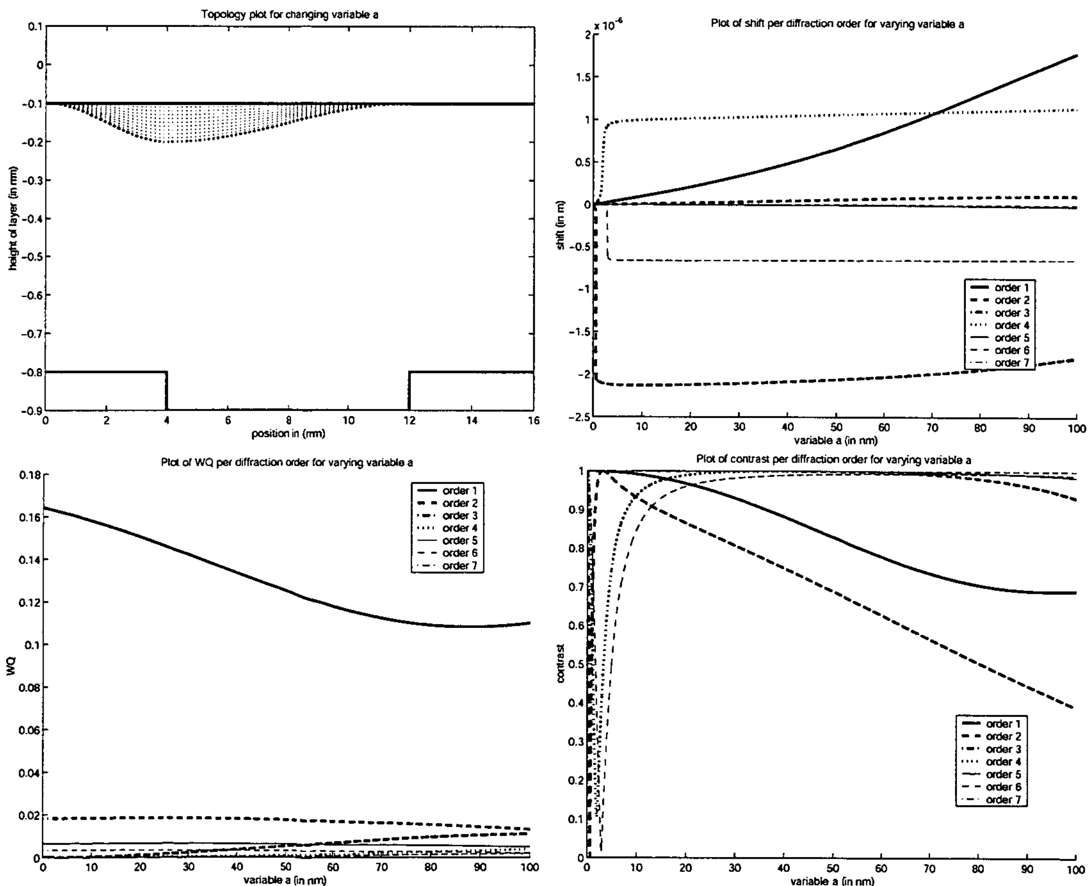


Figure 26: *Simulation results for variable a. Top left the simulation process for varying variable a is given. The top right plot gives the shift plot, while the bottom plots are the WQ and contrast plots. The shift as well as the WQ and the contrast are given as a function of the changes in variable a. Each line represents a single diffraction order.*

and contrast plot is different from order to order: the 2nd diffraction order shows its abrupt change before the 4th order does. These two effects are consequences of the asymmetry that is introduced, since the 2nd order will compensate the first asymmetries, and the 4th and 6th order will only refine the asymmetries and therefore they have a low signal intensity when the asymmetry is added to the topology. Therefore, the abrupt change can be noticed later for higher orders.

The shifts of the odd diffraction orders are small compared to the shifts of the even diffraction orders. This is logical, since the odd diffraction orders are dominant in the symmetric case which could be explained from the Fourier series that describe the shape of the alignment mark. A small change in topology does not change much about this dominance. Also for the shift of the odd diffraction orders a hierarchy is present since the shift of the 1st diffraction order is much higher than 7th diffraction order and this could be explained from Fourier series since the higher orders only refine the mark shape.

To conclude the simulation for variable a, it could be stated that all diffraction orders correlate to each other in a complex way, but when one takes samples it can be supposed that when the contrast shows abrupt changes, the covariance between orders will be small, since the shift shows an abrupt change. This is especially the case between even and odd diffraction orders just after the asymmetry is introduced. If the contrast does not show sudden changes for both orders, the covariance between the orders is certainly not equal to zero.

- **Variable b**

Similar to the simulation concerning variable a, the perfect symmetric mark topology is taken as a reference ($a = 0$ nm, $x_1 = 4\mu$m and $x_2 = 12\mu$m). Variable b is varied from 0 to 100 nm. This means that at position x_2 a pile is created of 100 nm maximal. When b becomes larger than zero, asymmetry is introduced to the topology. The simulation process is pointed out in Figure 27 together with the results for the shifts, WQ's and the contrasts.

By altering the variable b similar features can be noticed from the results as for altering variable a.

- **Variable x_1**

Changing variable x_1 means that the position of the dip is replaced. It is useless to take the reference situation as taken for the simulations with variables a and b, since there is no hole and varying x_1 would not change this topology. The choice of a reference situation is made rather arbitrary: the variable a is set to 100 nm and variable b equals zero. The value of variable x_2 is kept at 12 μm. Now x_1 is varied between 3 and 7 μm. This is illustrated in Figure 28. In this figure also the results are given.

It can be seen from the figure that especially the shifts of the lower diffraction orders react significantly to the changes, although the higher orders are certainly not constant. From a Fourier series approach this could easily be explained, since the lower diffraction orders capture the actual form first and this form is refined by the higher diffraction orders. Also the contrast plot shows that the contrast of the lower diffraction orders become low at a certain value of x_1. However, since this simulation shows no symmetric topology, all diffraction orders are present (the WQ never becomes zero) and the abrupt changes in shift and contrast as in the simulation

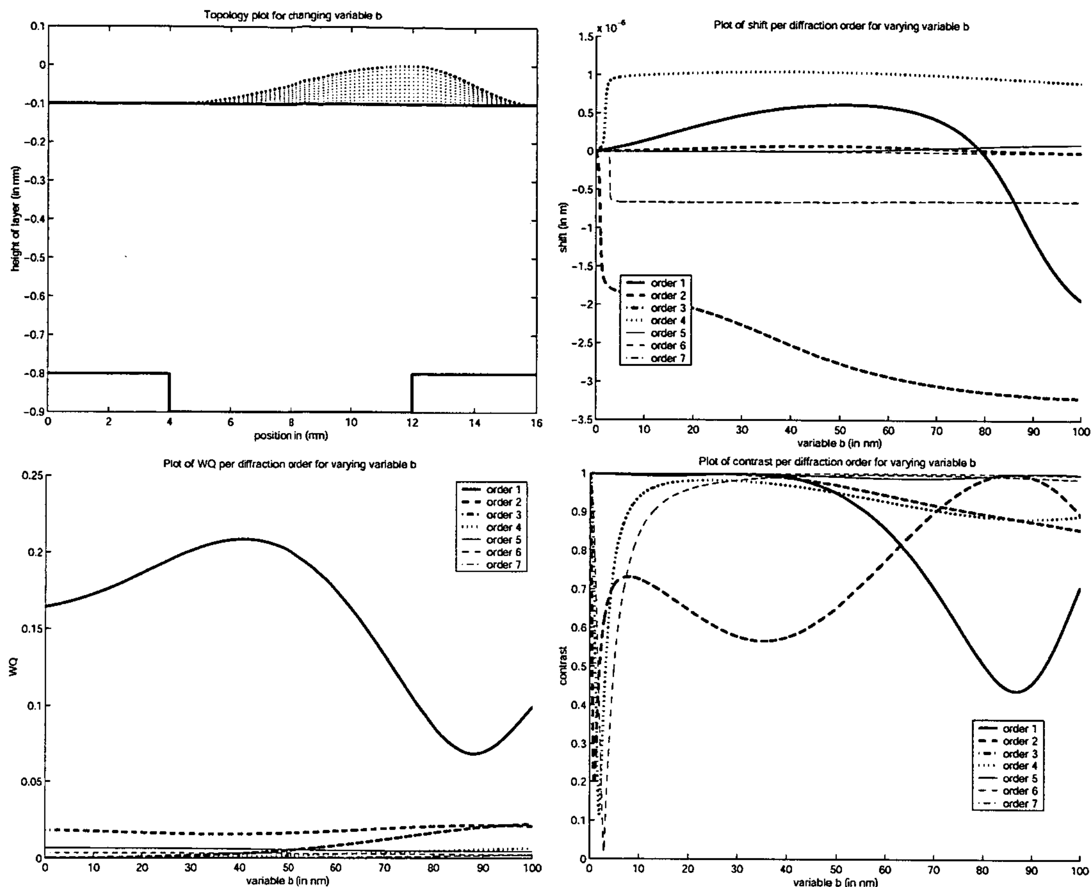


Figure 27: *Simulation results for variable b. Top left the simulation process for varying variable b is given. The top right plot gives the shift plot, while the bottom plots are the WQ and contrast plots. The shift as well as the WQ and the contrast are given as a function of the changes in variable b. Each line represents a single diffraction order.*

for a and b cannot be seen.

- **Variable x_2**

  Similar to the reference situation for x_1, the perfect symmetric topology could not be used here. Variable a is set to zero and x_1 is kept at 4 μm. Arbitrarily, the variable b is set to 100 nm, so the pile is present. Variable x_2 is varied between 9 and 13 μm. The set-up of this simulation and the results are given in Figure 29.

  Similar features could be noticed in this figure as in the simulation of changing variable x_1.

5.4.3 Verification of the assumptions made to obtain a solution

In the preceding section, the effects of changing only one variable were studied. It seems that the diffraction orders are correlated if the contrast is high. In this section, it is made plausible that one

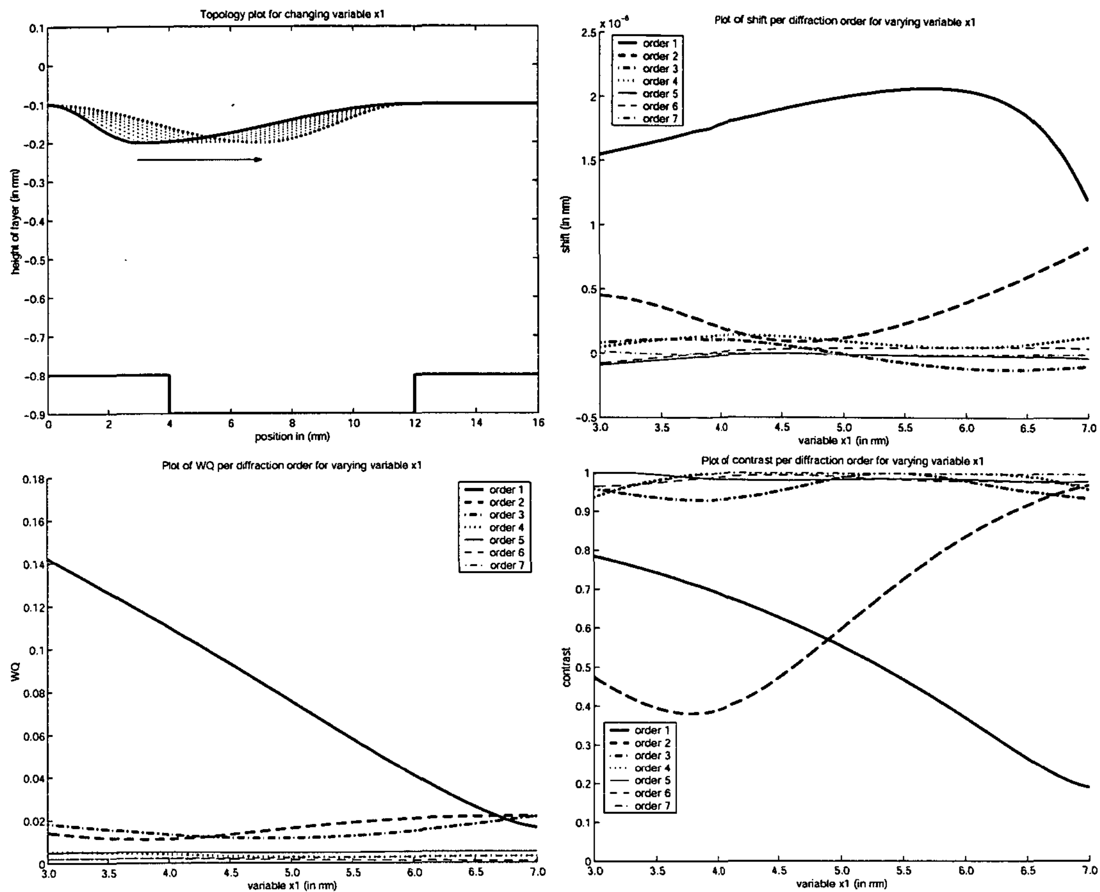


Figure 28: *Simulation results for variable x_1. Top left the simulation process for varying variable x_1 is given. The top right plot gives the shift plot, while the bottom plots are the WQ and contrast plots. The shift, the WQ and the contrast are given as a function of variable x_1. Each line represents a single diffraction order.*

effect coming from a certain variable could be enhanced or reduced by the effects of another variable: the "spaghetti" effect. If such a spaghetti effect takes place, nothing could be said about correlation between diffraction orders since each combination of variables gives some annihilation and/or enforcement effect. This indicates that in some cases covariance could be present and in other cases not. This is an explanation of what in Chapter 4 causes the different combinations of covariances for the 4 different marks, which resulted in the most accurate approximation of the exact solution (see Table 2).

The covariance between diffraction orders are small when the mark topology is almost symmetric. In this case the even diffraction orders appear rapidly but these orders are highly sensitive to the mark topology. In this case it is probable to state that there is no covariance between diffraction orders.

To conclude this chapter, it can be said that the correlation between diffraction orders is difficult to find due to two causes. In the Vertical Propagation Model of this chapter, only four parameters

© ASML 2003

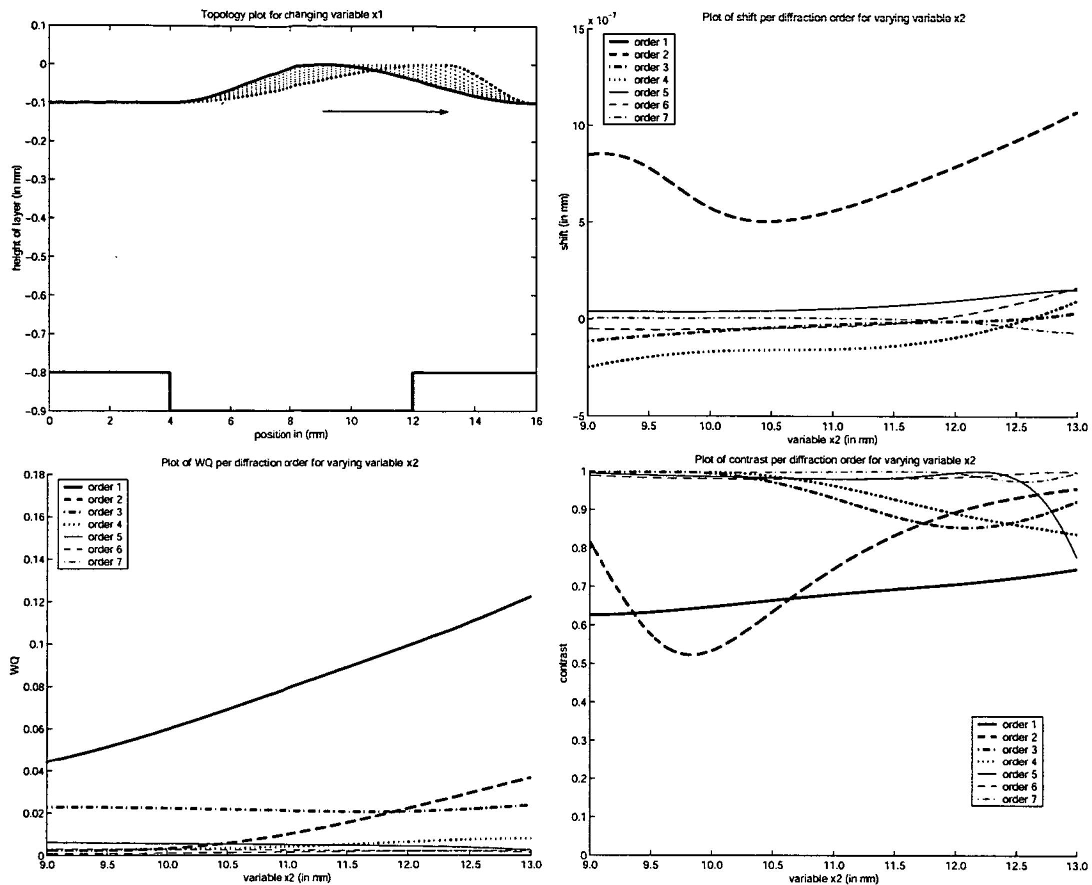


Figure 29: *Simulation results for variable x_2. Top left the simulation process for varying variable x_2 is given. The top right plot gives the shift plot, while the bottom plots are the WQ and contrast plots. The shift, the WQ and the contrast are given as a function of variable x_2. Each line represents a single diffraction order.*

were taken into account, but the shape of the resist layer must be described by much more parameters in practice and each measured mark will have a different set of values for these parameters. Therefore, the spaghetti effect will be enforced. The other cause is that the RS experiment described in Chapter 2, is carried out with only 6 wafers and therefore the resulting covariance is not estimated accurately and large deviations from the actual covariance are possible. When the batch size is much larger, the covariance is expected to be more consistent and less sensitive to coincidences.

6 Conclusions

This master's thesis was intended to provide the Order-to-Order method a mathematical basis and to validate this basis by experiments and simulations. The Order-to-Order method is meant to select that diffraction order that is least sensitive to alterations in the shape of the alignment marks due to processing of the wafers. This information must be derived from the measurements obtained from the processed wafers only.

The following conclusions can be drawn for the Order-to-Order method with the references to the sections they come from:

- Standard statistical techniques like (M)ANOVA's are not usable in the Order-to-Order method since only processed wafers and no reference wafers are available (Chapter 3).
- If shifts of diffraction orders are covariant, the Order-to-Order method results in an underdetermined system since there are more unknowns than equations (Chapter 3).
- To obtain the correct solution of the Order-to-Order method, additional knowledge must be available to reduce the number of unknowns are add more equations (Chapter 3).
- The assumption of no correlation between diffraction orders or no correlation between even and odd diffraction orders do not result in the diffraction order with the minimal variance. However, differences in variance between the obtained order and the optimum order are small (Chapter 4).
- The estimates of the single order PIAS variance can be improved by the Order-to-Order method, when only as much covariances between orders are set to zero as there are diffraction orders. However, if the wrong set of covariances is taken, the computed variances are even worse than when the assumption is used that there is no correlation at all between shifts of diffraction orders (Chapter 4).
- The set of covariances that must be set to zero to obtain an optimal solution is mark position dependent, since for each position this set consists of different covariances (Chapter 4).
- The Vertical Propagation Model shows that the diffraction orders are correlated when mark topologies are similar, except when the contrast and WQ becomes very large as in the transition form a symmetric topology to an asymmetric topology (Chapter 5).
- The more parameters involved in describing the mark topology, the less predictable covariance between shifts of diffraction orders will be (Chapter 5).
- The less wafers are taken into account, the less predictable the covariance between shifts of diffraction orders will be (Chapter 5).

With respect to the question why the Order-to-Order seems to work quite well, the main conclusion from this master's thesis is that, from wafer-to-wafer single order PIAS variances, this question can not be answered. It seems that the shifts of diffraction orders always show some correlation, but this correlation is unpredictable since there was only a small number of wafers available.

7 Open issues

Although this master's thesis answers some questions concerning the Order-to-Order method, there are some open issues. These issues are summarized in this chapter:

- **Find a statistical model to implement (M)ANOVA**
  To determine the PIAS variance from the APD variance, statistical techniques like (M)ANOVA's are the only way to do this in a statistically correct way. However, to compute a variance of a specific factor (in the Order-to-Order method the factor was the effect due to processing), multiple levels of this factor should be present. In the W-CMP and RS experiment, there is only one level for processing, since each mark has undergone only one process. If the batches were measured when they were not processed yet or measured again after another process has been executed, there would be more levels available and the (M)ANOVA could determine the process variance.

- **Compare mark positions on a wafer**
  In this master's thesis only wafer-to-wafer variance is considered, since it is assumed the process effects on alignment marks are similar on one mark position over all wafers. An open issue is to find a way to compare alignment marks on different mark positions, since each position could result in a different diffraction order and the question remains which order should be selected?

- **Verify the Order-to-Order method with more measurements and more wafers**
  Chapter3 also discusses the differences between the 3M data set and the production like data set. In theory, both data sets must result in the same SbO variance. More measurements and more wafers result in better estimates for PIAS variances.

- **Use two colors**
  The alignment marks are illuminated by two lasers: a red and green laser. Since both lasers have a different wavelength, the lasers will diffract differently on a certain mark shape. However, as long as the WQ and the contrast of the signal is sufficiently high, both sets of measured diffraction orders are the result of the same mark shape.

- **Search for a full rank coefficient matrix of system (1)**
  In Chapter 4 a brute force technique is used to determine which n covariances can be set to zero to keep the coefficient matrix of system (1) a full rank matrix. In order to have a much faster method, it is important to know the conditions under which the coefficient matrix has full rank depending on the choice of the n covariances.

- **Improve the reliability of the results when n covariances are set to zero**
  The results obtained from Figure 14 seems to be promising. It should be verified that this works on other processes as well (like W-CMP). To implement this technique, the exact variance is needed to obtain the minimum distance between the computed and the exact variance, but there might be a way to select the right combination of covariances without the need of the exact variance.

- **The availability of the mark shape gives PIAS information**
  If the shape of a mark could be satisfactory estimated by means of the measured diffraction orders, the single order PIAS's of the mark can be reconstructed.

References

[1] Charles F. Wojslaw. *Integrated Circuits - Theory and Applications.* Reston Publishing Company, Inc., 1978.

[2] Dr. Seth P. Bates. Silicon Wafer Processing. *Applied materials*, Summer, 2000.

[3] Jeroen Huijbregtse. Order-to-Order on Wafer Model Parameters for VSPM's. MEMO, ASML, 2001.

[4] Joep Roijers. Process Effects on Alignment - Tungsten-CMP / Aluminium PVD. Applications Note, ASML, January 2003.

[5] Arnout Smit. Resist Pile-up Effect. MEMO, ASML, January 2000.

[6] Sanjay Lalbahadoersing. Measuring with the MMM Method. General Information Document, ASML, September 2001.

[7] Douglas C. Montgomery and George C. Runger. *Applied Statistics and Probability for Engineers.* John Wiley & Sons, Inc., third edition, 2003.

[8] Sicco Schets. A Statistical Method for Order-to-Order Shift Variances Deconvolution into Single Order PIAS Variances. General Information Document, ASML, 2001.

[9] John A. Rice. *Mathematical Statistics and Data Analysis.* Wadsworth, Inc., second edition, 1995.

[10] Biswa Nath Datta. *Numerical Linear Algebra and Applications.* Brooks/Cole Publishing Company, 1995.

[11] J. Douglas Faires and Richard L. Burden. *Numerical Methods.* PWS-KENT Publishing Company, 1993.

[12] Eugene Hecht. *Optics.* Addison Wesley, fourth edition, 2002.

[13] Arie den Boef. The Influence of Resist Thickness on ATHENA Performance. Preliminary Investigation Report, ASML, 2001.

[14] E. Oran Brigham. *The Fast Fourier Transform.* Prentice-Hall, Inc., 1974.

[15] G. J. Borse. *Numerical Methods with Matlab® - A Resource for Scientists and Engineers.* PWS Publishing Company, 1997.

[16] *Matlab® Function Reference*, volume 2: F - O. The Mathworks, Inc., 6 edition, 2002.

[17] Arnout Smit. Process Effects on Alignment. Applications Note, ASML, March 2001.