Development of ultra-precision tools for metrology and lithography of large area photomasks and high definition displays

Peter Ekberg

Doctoral Thesis

School of Industrial Engineering and Management
Department of Production Engineering
KTH Royal Institute of Technology
Stockholm June 2013
Large area flat displays are nowadays considered being a commodity. After the era of bulky CRT TV technology, LCD and OLED have taken over as the most prevalent technologies for high quality image display devices. An important factor underlying the success of these technologies has been the development of high performance photomask writers in combination with a precise photomask process. Photomask manufacturing can be regarded as an art, highly dependent on qualified and skilled workers in a few companies located in Asia. The manufacturing yield in the photomask process depends to a great extent on several steps of measurements and inspections. Metrology, which is the focus of this thesis, is the science of measurement and is a prerequisite for maintaining high quality in all manufacturing processes. The details and challenges of performing critical measurements over large area photomasks of square meter sizes will be discussed. In particular the development of methods and algorithms related to the metrology system MMS15000, the world standard for large area photomask metrology today, will be presented.

The most important quality of a metrology system is repeatability. Achieving good repeatability requires a stable environment, carefully selected materials, sophisticated mechanical solutions, precise optics and capable software. Attributes of the air including humidity, CO$_2$ level, pressure and turbulence are other factors that can impact repeatability and accuracy if not handled properly. Besides the former qualities, there is also the behavior of the photomask itself that needs to be carefully handled in order to achieve a good correspondence to the Cartesian coordinate system. An uncertainty specification below 100 nm (3$\sigma$) over an area measured in square meters cannot be fulfilled unless special care is taken to compensate for gravity-induced errors from the photomask itself when it is resting on the metrology tool stage. Calibration is therefore a considerable challenge over these large areas. A novel method for self-calibration will be presented and discussed in the thesis. This is a general method that has proven to be highly robust even in cases when the self-calibration problem is close to being underdetermined.

A random sampling method based on massive averaging in the time domain will be presented as the solution for achieving precise spatial measurements
of the photomask patterns. This method has been used for detection of the position of chrome or glass edges on the photomask with a repeatability of 1.5 nm (3σ), using a measurement time of 250 ms. The method has also been used for verification of large area measurement repeatability of approximately 10 nm (3σ) when measuring several hundred measurement marks covering an area of 0.8 x 0.8 m².

The measurement of linewidths, referred to in the photomask industry as critical dimension (CD) measurements, is another important task for the MMS15000 system. A threshold-based inverse convolution method will be presented that enhances resolution down to 0.5 µm without requiring a change to the numerical aperture of the system.

As already mentioned, metrology is very important for maintaining high quality in a manufacturing environment. In the mask manufacturing industry in particular, the cost of poor quality (CoPQ) is extremely high. Besides the high materials cost, there are also the stringent requirements placed on CD and mask overlay, along with the need for zero defects that make the photomask industry unique. This topic is discussed further, and is shown to be a strong motivation for the development of the ultra-precision metrology built into the MMS15000 system.

**Keywords:** Ultra precision 2D metrology, LCD-display, OLED-display, nm-resolution, random phase measurement, large area, photomask, acousto-optic deflection, self-calibration, Z-correction, absolute accuracy, uncertainty.
The history of Micronic Laser Systems AB started at KTH Royal Institute of Technology in Stockholm, in the beginning of the 1970s by Gerhard Westerberg. At that time his first ideas of a laser-based pattern generator started to bear fruits. The first machine (1975-76) was rather small and could easily be placed on an office desk. An interesting detail of this machine was the servo system for the X-movement that was built from loudspeaker coils and springs. Also at this time, the first continuous-wave Helium Cadmium lasers became commercially available. This enabled Westerberg to use a wavelength of 448 nm, well-suited for the acousto-optical deflector that was used for deflecting the writing beam in the machine. The purpose of this first laser writer was to write the photomasks for the Motorola 6800 microprocessor chip. At the end of the 1970s he sold the whole concept to the company David Mann that manufactured aperture writers and steppers for the semiconductor industry. In the contract the intention was that Westerberg would receive royalties when David Mann had commercialized the machine and started to sell it on the pattern generator market. However, David Mann never reached this stage, and due to a special clause in the contract the rights were restored in 1984 to Westerberg who could then resume the development on his own. By that time he had started a company named Micronic developing handheld terminals for inventory purposes so work on the laser writer was proceeding as a hobby project. After some time Westerberg sold Micronic and started a new company Ventronic. The purpose of that company was to explore and finance new creative and technical ideas from students having received a master's degree at KTH. However his engagement in this company did not last long. In his next venture, he started the company Micronic Laser Systems and moved to a new building in Danderyd in 1985 together with five employees. A bit later this company delivered the first laser-based mask writer to “Svenska Grindmatriser” in Linköping. This system was more of a prototype than a proven product, and much of the mechanics, electronics, optics and software had limited performance.

In 1988 Gerhard passed away and his old friend, Nils Björk took over the company which by that time had about 10 enthusiastic employees. At that time the financial situation was extremely tough and Micronic Laser Systems was very close to the point of having to shut down.
After Gerhard’s death the company moved to Täby, just north of Stockholm. The business situation in the semiconductor industry was not very lively at that time so the company was in dire need to find new applications where the laser-based writing technique could be used. The first breakthrough came after one of the employees read an article about large area photomasks. He managed to convince the rest of his colleagues including Nils Björk that they should design and build a writer capable of patterning photomasks with dimensions of 0.6 x 0.6 m². This was enormous in comparison to the four- and five-inch reticles the current Micronic Laser Systems tool could handle. The first generation large area pattern generators produced by the company were used for writing so-called “shadow masks” that served as a master to the front metal screens in CRT color TVs. The second and most important breakthrough came when the LCD TV manufacturing plants required ultra-precision photomasks for their production. At that time, Japan was the most important market and the first installation of this kind of maskwriter was made in 1996 at a Hoya factory near Tokyo. The stage of this system was 0.6 x 0.6 m² and served as the starting point of the “size race” that soon was to come. Today the largest photomasks are 1.6 x 1.8 m².

In 2005 Micronic Laser Systems launched the metrology system MMS15000 for large area display masks. This system is still the most accurate 2D coordinate measurement machine for large area photomasks in the world. In addition to ultra-precision mechanics, high quality optics and extremely precise environmental control, a number of new measurement principles and algorithms were developed and implemented in the MMS15000. Some of these are presented and discussed in this thesis. In 2010 Micronic Laser systems merged with Mydata Automation AB and became Micronic-Mydata AB with a broadened product portfolio. Today more than 400 people are employed at Micronic Mydata with offices around the world.
You, as a reader, may really question my decision to pursue an academic career after having worked for 28 years at Micronic Laser Systems/ Micronic Mydata. There are many reasons for this and I will mention some of the most important. Firstly, I am very proud of what we have achieved in the company. We have managed to become the only supplier in the world, for the most demanding applications, of ultra-precision large area photomask writers and measurement systems. We have developed the MMS15000 into a world standard metrology tool used for measuring large area photomasks. This in itself, I think is well worth writing a thesis about. Secondly I must say that my doctoral work would probably not have happened if I had not met Professor Lars Mattsson at KTH Industrial metrology and optics. I will never forget the evening when Lars proposed to me that I start a PhD career at his department. At that time I remembered that I thought, “Wow, what a challenge”, but at that time I was of course completely unaware of all the work that would be involved. On many occasions throughout this work it has been really tough but Lars Mattsson has been the person that always supported me.

During my years at Micronic I have worked in the fields of electronics, optics, image processing and metrology. Based on this, you may think that you have gotten quite some experience in these fields. But after having struggled with all of the courses and projects at KTH, I have realized that there are many more challenges and much more to learn in the fields of optics, image processing and metrology than I ever could have imagined. I therefore want to thank all my friends at the department of Production Engineering for a lot of inspiring discussions.

My second supervisor is Techn. Dr. Lars Stiblert. He is the person who has engaged in lengthy technical discussions about details in this thesis and also other tricky technical questions raised by other KTH projects I am getting involved with. Without his support I do not think I would have had all the motivation needed to finish this thesis.

The work presented in this thesis is a result of many years of R&D in areas such as mechanics, optics, software and metrology at Micronic Laser Systems. The development of the MMS15000 was done by a small group of highly dedicated people. I believe that a major reason for our success was the
fact that we were only 8-10 people working closely together with one clear goal. I especially want to thank John-Oscar Larsson, Klas Edgren, Peter Henriksson, Mikael Wahlsten, and Anders Svensson for all their efforts in making the MMS15000 fulfill the measurement accuracy specification of 90 nm (3σ) over an area of 1.4 x 1.5 m². I also want to acknowledge all my other co-workers at Micronic for all their support and in particular Tom Newman for correcting my “Swenglish.”

From the start of my PhD thesis work four years ago, the company has supported me by letting me do some of the work on this thesis during working hours. For this reason I really want to thank my managers at the company, and especially Sven Löfquist.

Without an understanding wife it would not be possible to do such a thing as PhD-thesis work, while at the same time work in a company. Gunilla has really let me work with this thesis without too many complaints. It is difficult to find words to express how much I have appreciated her understanding when I have been sitting late nights and weekends at the computer.

Another person that really has inspired me is my son Joakim. When I started this journey he was a student at KTH. I remembered the strange feeling of being a student at the same university as my son. Luckily we never bumped into each other in the same lecture.

April 2013 Peter Ekberg
“Do not worry about your difficulties in Mathematics. I can assure you mine are still greater.”

Albert Einstein
List of papers

**Paper A**  
Johan Åman, Peter Ekberg, Lars Leonardsson, Klas Edgren, Torbjörn Sandström, Lars Stiblert, "Recent developments in large-area photomasks for display applications”, Journal of the Society for Information Display  
Volume 9, Issue 1, pages 3–8, March 2001

**Paper B**  

**Paper C**  

**Paper D**  

**Paper E**  
Other publications


4. Peter Ekberg, “Zcorrection, a method for achieving ultra-high absolute pattern placement accuracy of large area photomasks“, Submitted to the proceedings of the 13th euspen International Conference, Accepted as an oral presentation, 2013

5. Peter Ekberg, “Metrology – a forgotten added value maker that eliminates cost of poor quality and supports a sustainable zero defect production“, The 5th International Swedish Production Symposium, 2012
# Table of Contents

1 INTRODUCTION
   1.1 BACKGROUND .................................................. 1
   1.2 PROBLEM STATEMENT ......................................... 2
   1.3 GOAL OF THIS THESIS ......................................... 3
   1.4 WHY METROLOGY? ................................................ 3
   1.5 ECONOMIC BENEFIT BY METROLOGY - A REAL EXAMPLE .......... 3

2 CURRENT DISPLAY TECHNOLOGIES REQUIRING MASK METROLOGY ........ 9
   2.1 ELECTROPHORETIC DISPLAY ..................................... 9
   2.2 PLASMA DISPLAY ................................................ 10
   2.3 LCD .................................................................. 11
   2.4 OLED .................................................................. 13
   2.5 BACKPLANES ..................................................... 16
   2.6 MANUFACTURING TECHNOLOGIES ................................. 16

3 MASK WRITER AND MEASUREMENT TOOL ................................. 20
   3.1 MASK WRITER ..................................................... 20
   3.2 THE MMS15000 METROLOGY TOOL ............................... 22

4 MASK QUALITY MEASURES .................................................. 24
   4.1 TERMINOLOGY FOR PHOTOMASK QUALITY ASSESSMENT ....... 25
   4.2 TRACEABILITY OF THE MMS15000 ............................... 28
   4.3 MURA ............................................................... 29

5 METROLOGY ................................................................... 34
   5.1 BASIC SYSTEM ANALYSIS ........................................ 34
      5.1.1 One-dimensional corrections ................................. 37
      5.1.2 Two-dimensional corrections ................................ 40
   5.2 RANDOM PHASE MEASUREMENT TECHNIQUE .................... 42
      5.2.1 One-dimensional measurements ............................. 42
      5.2.2 Two-dimensional measurements ............................. 46
      5.2.3 General two-dimensional measurements .................. 48
   5.3 COMPETING METHODS, RANDOM PHASE AND ADC IMAGING .... 53
1 Introduction

1.1 Background

What is one of the most important technology developments in the last fifty years? Certainly we cannot give a straight answer to this question without providing some clues. If we then say that the answer has to do with communication, entertainment and information spreading in general it is much easier to answer the question. The Internet with its enormous possibilities of sharing information as images and text is certainly the answer of the digital revolution, starting with radio and TV broadcasting many years ago. The combination of the immense possibilities of communication using the Internet, along with high quality images is so natural today that hardly anyone reflects on the tremendous developments we have seen over the recent decades. One result of the digital revolution that most take for granted is that a mobile phone should contain a camera, and not just any camera, but a camera containing several Megapixels. The captured images can be communicated through Internet and presented on large image devices such as computer screens, monitors and TVs. An interesting fact is that the number of pixels per square-inch continues to increase. Is it not enough when we cannot resolve the individual pixels with our eye? In principle the answer is yes but let us take the high definition television standard HDTV, using 1920 x 1080 pixels$^2$ as an example. One would think that everyone prefers much better image quality just by comparing the older standard 720p using 1280 x 720 pixels$^2$ with the HDTV. If we watch such a TV (720p) with, for example, a screen size of 42 inches from a viewing distance of 4.5 meters it is not possible to resolve each individual pixel [1]. Even so we will see a continued increase of pixels/inch in the future. One of the reasons is that we want to use desktop computers, monitors and TVs for watching images captured by our cameras with all details included, and to do this at a much closer distance to the display. This has also led to controlling the contents of the display using the surface of the display itself by an integrated so-called touch panel, without the need for a keyboard.

Thus, we have now seen only the beginning of the development of display technologies. The prevalent liquid crystal display (LCD) will most likely be displaced by technologies based on organic light emitting diodes (OLED), that provide a better performing and environmentally friendlier alternative.
OLED is already a common display in mobile phones and will soon be as common for the larger displays used in tablet computers, desktop computers and monitors.

What has made it possible to make this extremely advanced flat glass combined with electronics that comprise a display today? Essential factors are photomask lithography and high precision metrology. Several photomasks are needed to manufacture displays, and these photomasks must be extremely accurate from the geometrical point of view. As an example, the placement precision of a feature on a display mask must be in the range of 150-200 nm (3σ) over an area measured in square meters. This also leads to the requirement that a metrology tool used for the verification of such precision on the mask must have a capability better than 150 nm. The state of the art metrology tool MMS15000 presented in this thesis is currently capable of verifying photomasks to a level of accuracy, called registration in photomask applications, of less than 90 nm (3σ) over an area of 1.4 x 1.5 m² [2].

Fused silica (amorphous quartz glass) is the material used for the advanced photomasks defining the patterns of the backplane of a LCD display. The physical handling of a finished large area photomask is difficult, and for transport protection a pellicle is mounted over the patterned area. Pellicles are special coverage frames that protect the mask from contamination, humidity and gases that could degrade the mask. The cost of the mask material and the pellicle amounts to 60-80% of the average production cost of a photomask. The purpose of this thesis is to present the technique we have developed and that is used worldwide in the large area photomask industry to achieve the goal of zero-defects of the exceedingly critical photomasks.

### 1.2 Problem statement

The problem that this thesis is trying to solve can be formulated as:

1) Is it possible to develop and calibrate a mechanical/optical measurement system and then verify the X,Y placement of chromium patterns on a glass surface to an uncertainty < 100 nm (3σ) over a square meter sized area?

2) Can a pattern dimension be measured with repeatability in the range of 10 nm (3σ)?
1.3 Goal of this thesis

In this thesis we will explain the fundamentals of the process for making highly accurate photomasks with a size of square meters that are used in large area exposure systems. Because of this large size, new technical challenges appear that are on the verge of what is physically possible. This is true for the photomask writer but it is actually more so for the metrology tool used to verify the quality of the mask. So the main goal is to explain the details of the metrology solutions in the MMS15000 [2]. How is uncertainty in registration and overlay defined and verified? We will present methods for measurements with repeatability expressed in nanometers over large areas. The last and probably the most important goal is to discuss the methods that we use for achieving an accuracy below 100 nm (3σ) and the potential to extend it to the largest stages of over 1.4 x 1.5 m².

1.4 Why metrology?

As in many segments in the manufacturing industry, metrology is a vital part of the manufacturing process in the display industry. Several companies mostly in Asia are involved in manufacturing products such as TVs, tablet computers, mobile phones and other devices containing a display. Photomasks for the display industry are produced either by companies specialized in producing photomasks (merchant mask shops) or in-house (captive mask shops) owned and operated by the panel manufacturing company.

1.5 Economic benefit by metrology - a real example

Because of the extremely high material costs involved in photomask production, metrology and various inspection processes are required in between the processing steps. The only acceptable final product is a completely defect-free photomask, since the photomask serves as the template in the panel production process. In figure 1 the typical manufacturing process from the glass blank to the final photomask is presented, although this process flow may differ slightly between mask shops and products.
Figure 1: The photomask process.

The material used for photomasks is either soda-lime or amorphous quartz glass, with the latter being used for the most critical layers. The size of the glass blank may range from 390 x 610 mm$^2$ up to 1620 x 1780 mm$^2$ depending of the so-called generation of the panel maker’s manufacturing line. Today the only company using the largest size called G10 is Sharp, located in Japan. Most companies are G8 factories using masks with sizes around 1220 x 1400 mm$^2$. The thickness of the glass blank depends on its lateral size and varies in the range of 5-16 mm. On the top surface of the glass, a uniform chromium layer is sputtered with a thickness in the range of 80-150 nm. The mask shop buys the mask blanks from companies specializing in polishing the glass or quartz and applying the chromium. On top of the chromium layer a photo sensitive resist is applied with a thickness of 800-1200 nm. The resist layer is most often applied by the mask shop using spin coating, spray coating or a roll coating process[3][4]. After that the glass blank is inspected in the immediate vicinity of the exposure machine to
minimize the risk of contamination. In this step blanks containing any pinholes, i.e. defects in the resist, are re-coated with a new resist layer.

In the exposure step the mask pattern is written into the resist by the photomask writer. This step may take up to several days depending on the size of the glass used and the quality requirements. In some cases the same pattern may be re-written up to three times for the purpose of reducing random mechanical errors and laser noise in the exposed resist. After development of the resist the glass blanks already have significant value due to the high material cost and the long exposure time.

In the “Inspection 2” box in figure 1, several metrology steps are involved. First, pattern placement accuracy is checked by measuring special alignment patterns spread out over the exposed resist surface. In some cases the line width, referred to as critical dimension (CD), is also measured at certain points in the pattern. The purpose of this measurement is to verify the process parameters. If the mask passes these two measurements it is then inspected by a dedicated system used for finding defects in the written pattern in the resist. The located defects are repaired either by local deposition of new resist, or by removing resist, depending on the type of defect. In cases when the number of defects is too high or some defects cannot be repaired, the mask needs to be reworked. A manual inspection of the so called “Mura” quality is also done in this step. Mura is a Japanese expression meaning “not-perfect” and is a visually perceived measure of the quality of the mask. Mura defects are extremely difficult to measure quantitatively by optical detection and for this reason the inspection is done manually. The details of Mura and why it is vital for photomask quality will be discussed later.

Up to this point in the mask process it is possible to save the glass if any fatal or un-repairable resist defects have been found. This is done by removing all resist and applying a new layer.

In the etching step the exposed chromium areas are removed using a chemical process. The remaining un-exposed resist is then stripped off. The result after these steps is the final chromium pattern on the top surface of the glass substrate. A third inspection and repair loop of the chromium pattern is now performed. Special repair machines are used to remove any short circuits in the pattern, i.e. chrome bridges. Missing parts of the chromium pattern are repaired by local deposition of the metal. This is a time consuming process including several additional re-inspections. After this step, rigorous CD measurements are done and placement accuracy is more thoroughly verified using the MMS15000 metrology tool. At this point the mask must fulfill the CD and placement accuracy specifications and has to be accepted by specialists checking the Mura quality. If the mask fails any of these tests, it is
rejected with a significant loss of money and time as a consequence. Today it is common that scrapped glass is re-cycled by the glass supplier. That saves about 50% of the blank cost since the other 50% are costs for grinding, polishing and chrome sputtering. After mounting the pellicle (the special coverage frame used for protecting the mask) the photomask is finally manually inspected for potential contamination. Special machines are in some cases used for this procedure. In figure 2 two examples of typical photomask patterns are shown.

![Figure 2: Typical chromium patterns (dark) as viewed in transmitted light on a photomask. The horizontal pitch of the pattern is in the range of 150 µm.

a) A black matrix pattern i.e. one of the layers of the color filter.
b) A pattern of one of the layers of the backplane.](image)

It is clear from the description above that metrology is heavily involved in the mask making process. Since a photomask is used as a master in the display panel production process, no defects are accepted. Because of the increasing complexity of the masks, more metrology is needed to keep up a high yield in the mask shop. By using a real example we will show how the costs are distributed in the production of a G8 1220 x 1400 mm² array photomask. An LCD display requires 4-5 color filter (CF) masks and 4-6 masks for producing the transistor array back plane. In table 1 we have summarized the cost distribution for an array mask relative to the total production cost. Depending of the quality demand the output price of this type of mask will vary between 82 and 141 kUSD.
Table 1: The cost distribution of the blank and the different process steps in a photomask process relative to the total production cost. The remaining percentages are costs for administration, transportation and sales.

Because of the different implementations of the metrology processes among different mask shops we have chosen to present “Inspection and repair” as one line in the table. It is very difficult to accurately determine this kind of information from the mask shops since they consider it to be confidential. For this reason we can only present approximate numbers in table 1. The cost for the blank itself depends of what kind of layer and what process technology the panel maker is using for a particular mask. The mask can either be a binary mask where the pattern contains only two levels in transmitted light, black (chromium) and white (glass) or a so called half tone (HT) mask. In the latter case the pattern has an additional layer giving a half tone with the transparency of 50%. This layer is sputtered on to the glass surface. The processing of such masks is more complicated and involves at least two exposures and extra processing steps. The relative cost for a HT mask blank can be as high as 80% of the total production cost.

In 2012 there was extreme price competition between the large area mask shops that are located in Japan, Korea, Taiwan and China. This led to a profit margin of nearly zero. Only mask shops that have optimized their process and can deliver the highest quality for the most demanding applications currently show some profit. As seen in table 1 metrology stands for only 1-2% of the total production cost. To this amount some percent of the cost for inspection should be added. We include inspection, since this together with metrology assures the quality of the mask, and a failure in metrology and inspection can lead to losses of up to 80%. Therefore the total cost of poor quality (CoPQ) is extremely high in the photomask business [5][6]. Despite the high quality production, 3-4% of the photomasks produced have to be scrapped as a result of the verification by metrology and inspection. There are several reasons for this, e.g. process errors that could not be repaired, CD and/or placement accuracy that did not fulfill specifications, Mura and others.

A really bad situation would occur if a mask shop delivers a poor quality mask to their customer. The effect of this will show up as low or zero yield late in the panel manufacturing process. The reason for this delay is that the

<table>
<thead>
<tr>
<th>Blank cost</th>
<th>60-80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process and etch</td>
<td>1-2%</td>
</tr>
<tr>
<td>Metrology</td>
<td>1-2%</td>
</tr>
<tr>
<td>Inspection and repair</td>
<td>2-3%</td>
</tr>
</tbody>
</table>
panel maker relies on the measurement protocol and the pre-inspected quality check delivered by the mask shop since they do not make their own detailed inspections or measurements of the photomasks before they are used in the panel process. The cost for re-making the mask after several weeks and the “loss of face” to their customer will certainly impact the confidence between the mask shop and the panel maker. In the Asian culture such a failure can be fatal in the sense that the mask shop might be punished by not getting any new orders for a long time.
Today the most important display technologies are LCD and OLED [7][8] and both technologies require ultra-precision photomasks for their lithography processes. But before discussing LCD and OLED let us present two other important technologies, the electrophoretic and plasma display technologies.

2.1 Electrophoretic display

In contrast to glass-supported LCD displays, commercially available electrophoretic displays can be built on soft material like plastic sheets. The electrophoretic display is also called “Electronic paper” [9]. The display is bistable so power is only needed to update the display. Electrically charged microcapsules are controlled by an external electrical field as shown in figure 3.

**Figure 3**: The principles of the Electrophoretic display.
Source: IDTechEx [10].

The power consumption is low in steady state but quite a large power surge is needed for changing the state of the pixels. The display is “paper like” with
very good readability, especially outdoors. The disadvantage today is that such displays are only available in two colors. Other disadvantages are slow response times, high voltages, low resolution and moving parts. These are the most common displays used in e-book readers (EBR). They can be made mechanically flexible and the structures are printed using inkjet or screen-printing technologies. The transistor backplane can use inorganic printable transistors.

2.2 Plasma display

The first commercial flat display technology entering the market at the end of the 1990s was the plasma display. At the beginning, the quality in brightness and color saturation of these displays was poor but they had a market anyway because they were large in size and flat. It is still possible to buy plasma displays and the quality is much improved. In figure 4 the design of the plasma display is shown.

Figure 4: The principle buildup of a plasma display.
Source: Wikimedia commons.

The working principle of a plasma cell is similar to a fluorescent lamp. A plasma, i.e. a collection of charged ions, activates the light emitting phosphors on the chamber walls by the voltage applied between the electrodes. An advantage of the plasma display is that it is flat and delivers a high quality image even when compared with the old CRT technology. Its disadvantages include power consumption, weight and limited resolution.
2.3 LCD

There is no doubt that the reason for the real breakthrough for both small and large size displays was the introduction of the liquid crystal display (LCD) technology [11][12][13]. The LCD display is the successor to the Plasma display. In the beginning, LCD displays could only be made to the size of a laptop screen with good manufacturing yield. The first displays were based on twisted nematic (TN) and super twisted nematic (STN) crystal types [14]. They were driven by a passive matrix backplane and showed rather poor image quality. These displays were also slow in response time and suffered from a narrow viewing angle. The advantage was very low power consumption. These kinds of displays were mostly used in inexpensive mobile phones, information screens and other digital devices. The LCD technology of today has developed and has a much better resolution and image quality. It uses an active backplane of two or more transistors driving each pixel. The refinement of the technology has gone on for a long time by several giants such as Samsung, LG and Sharp in Asia. The different units in a typical LCD module are presented in figure 5.

Figure 5: The different units of a typical LCD module. Today the cold cathode fluorescent tube (CCFL) backlight has been replaced in most cases by an array of light emitting diodes LEDs placed either behind the vertical polarizer or along the edges of the diffuser plate of the backlight unit. Source: Display Search [15].
Two different LCD technologies dominate today, the in-plane switching (IPS) and pattern vertical alignment (PVA) developed by Samsung. IPS technology is used today (2013) in Apple products including iMacs, iPads and iPhones. PVA is a development of multi domain vertical alignment (MVA) technology that was initially introduced by Fujitsu in 1998. In figure 6 the principles of the IPS and PVA technologies are shown.

**Figure 6:** The principle of the IPS and PVA technology. The liquid crystal (LC) is aligned differently in the two cases because of the direction of the electrical field \( E \). In the PVA the Indium tin oxide (ITO) electrodes are patterned both on the back plane and at the color filter substrate. Source: Display Search [15].

There is a fundamental difference in how the liquid crystal is controlled in the two different modes. In the IPS case the crystal is rotated by the electric field in the horizontal plane when a voltage is applied between the two electrodes located on the back plane substrate.

In the PVA case the crystal is vertically tilted in so-called different domains. On the right hand side in figure 6 an example is shown of how the four domains are aligned when the pixel is in its on-state. The direction of the E-field is set by the patterned ITO layer. The PVA technology initially suffered from color degradation when viewing the display at an angle. In order to minimize this effect, Samsung developed super PVA (S-PVA) where each pixel is subdivided into two parts resulting in eight domains.

The light source used today consists of LEDs placed behind or along the edges of the panel in the display module. Polarizers are used to set the incident light polarization angle so that it is aligned with the liquid crystal. Another polarizer is used to block the light as it exits the panel. This means that when the pixel is in off state no light will pass through the panel. When the pixel is turned on the crystal is twisted a certain angle resulting in a change of the polarization angle. This means that some light will find its way through the panel. A built-in problem with LCD technology is its very low
light efficiency, which is only 8-10% for state-of-the-art units. Methods to lower the power consumption have been developed. An example of such a method is the so called dynamic back lighting. When using this method it is possible to locally adjust the back light depending on the contents of the scene. To make this possible, the LEDs are configured in a matrix behind the active area of the panel. Then by using advanced real time image processing algorithms the local brightness can be adjusted. An obvious disadvantage using this method is that the contrast of small objects in dark regions will be reduced.

2.4 OLED

In most applications today the display is built on glass substrates. For this reason these displays are heavy and fragile. In the future we will see soft displays built on plastic substrates especially for hand held devices. OLED is a technology that can be implemented on soft substrates [16]. But many problems remain to be solved, as will be discussed later, before such devices are available on the market.

OLED will also very likely take over as the main technology also for large area displays used in TVs and digital signage applications. Today, mobile phones are available with OLED displays. The reason for this shift in technology is that OLED can deliver a much better image quality, is much faster and does not consume as much power as an LCD panel with a similar size. The fundamental difference between LCD and OLED is how the light is generated. In a LCD pixel the liquid crystal between the electrodes works as a switch for the backlight. So the pixel itself does not generate the light. The applied electrical field controlled by a voltage between the electrodes, mechanically rotates the quite heavy liquid crystal. This results in a slow response time in the range of milliseconds. An OLED pixel is actually a light emitting diode where the light emission is controlled by a current. So the light is emitted by the pixel itself. The switching time is much faster and it is in the range of microseconds. When no current is applied the pixel is truly black, which is not the case for a black LCD pixel. A black LCD pixel is not perfectly black because the liquid crystal (LC) switch and the polarizers cannot be made to have no light leakage. In figure 7 the design principle of an OLED pixel is shown.
Figure 7: The OLED pixel consists of a stack of several organic layers. Onto the glass substrate the anode is deposited as a translucent ITO layer and works as the bottom electrode. The rest of the layers are the hole injection layer, hole transport layer, emissive layer shown as R, G, B and the electron transport layer. At the top the cathode (made of metal or ITO) is finally deposited. Several photomasks are needed for the definition of the different layers. Source: Display Search [15].

Electrons and holes are recombined in the emissive layer and create light emission. Different wavelengths of the light can be generated by using different types of light emitting molecules [16]. There are three different designs of OLED displays currently used today: Top emission RGB OLEDs, bottom emission RGB OLEDs and designs using white OLEDs (WOLED) together with a color filter for generating the colors red, green and blue. In figure 8 the principles of top and bottom emission are shown.
The top emission design clearly has its benefits. Since the light is emitted close to the top surface there are no apertures blocking the light. This results in a very good viewing angle and high light efficiency. In the bottom emission design some of the active emission area is blocked by the thin film transistors (TFTs) indicated by the black rectangles in the figure. Also, since the light is emitted from the backside of the substrate the viewing angle will not be as good as in the top emission design.

White OLEDs are more stable than combinations of red, green and blue OLEDs. The small molecules used for the red, green and blue emission have different aging effects and will therefore generate color shifts if these drifts are not compensated for. For white OLEDs, the degradation will be the same for all pixels and will therefore not cause any significant color shifts. An advantage in the processing of white OLEDs is that no patterning of the OLED stack is necessary, simplifying the process significantly. In the RGB OLEDs case the colors are patterned using a shadow mask, i.e. a metal mask filled with holes defining the color layer in an evaporation process. The density of holes in this mask is limited because of the difficulty of keeping precise alignment between the different layers in the process. The definition of a pixel in white OLED displays is done by the color filter. A different lithography process based on photomasks is used in the manufacturing of the color filter, which does not have the same density limitations of the holes as in the former case.

Figure 8: The principles of bottom and top emission OLEDs respectively. Source: Changhee Lee [17].
Another great advantage of OLED compared to the LCD technology is that OLEDs can be made on soft materials such as plastics substrates or thin metal sheets. Demonstrations of this technology has already been done, but there is still a long way to go before we will see these types of OLED displays having the same quality as OLEDs built on glass substrates.

2.5 Backplanes

The control of the pixels is done by the backplane. It is also called the array, and is made of a matrix of thin film transistors (TFTs). The array can either be passive or active where each pixel is controlled by one or several TFTs. Today, high quality LCD or OLED displays use active arrays. The backplane for LCD displays is made of amorphous silicon. OLEDs are driven by a low-temperature polycrystalline silicon (LTPS) backplanes [18]. Because of the relatively large current needed for driving an OLED pixel, the electron mobility is about 100 times larger for the transistors in an OLED backplane compared to the transistors driving the LC. The majority of LTPS backplanes currently in production rely on excimer laser annealing (ELA) for crystallization of the silicon [19].

Several photomasks are used to define the pattern in the lithography steps in the array process. Depending on what process the panel maker uses, 4-6 masks are required for processing an amorphous backplane and up to 8 masks are required for processing a LTPS backplane. The quality demands regarding critical dimension (CD) and Registration are similar for LCD and OLED backplanes. Halftone technology is used to reduce the number of masks in the array process. Even if a half tone mask is more expensive to manufacture, the gain in costs by reducing one lithography step in the panel process is significant. Optical proximity correction methods (OPC) that have been used in the semiconductor industry for years to enhance pattern fidelity are now introduced and used in large area panel production [20]. This greatly increases the pattern complexity of the photomasks.

2.6 Manufacturing technologies

In the previous sections we have described the technology behind different display types. We will now complete this subject by describing the manufacturing process of LCD modules and bringing in the importance of photomasks. The manufacturing process of OLED modules is very similar. Mass production of these modules is done solely in Asia. The most important companies are Samsung, LG display in Korea, Sharp in Japan and AOU and CMO in Taiwan. In figure 9 a typical TFT module process is shown.
As seen in figure 9 the process is divided into four different processes: the Array, Color filter (CF), Cell and the Module process. In the array process several photomasks are used to define the pattern of the different layers of the transistor, data and gate metal lines. In the sputtering/CVD step different materials are deposited onto the substrate. After the substrate has been covered with resist, a photomask is used in the exposure step to define the pattern of this particular layer. After development, etching and stripping off the remaining resist this whole process is repeated 4-6 times to build up structures with different materials.

The color filter process is different from the array process in the sense that no CVD coating is involved. Photomasks are used for defining the patterns of the different layers. Special colored resists are used for the three different layers - red, green and blue (RGB) as seen in figure 9. In the last step the common ITO electrode is deposited.

In the cell process a polyimide alignment film is applied both on the array and color filter substrates. This film is required to control the pre-tilt angle and orientation of the liquid crystal. After baking in an oven at around 325 ºC, the alignment films hardens and is transformed to a polyimide that is often called the PI material. Next follows a rubbing step that forms small grooves in the alignment film, which make the liquid crystal alignment easier. The rubbing is done using a metal roller covered with cloth made of rayon or cotton.
Rubbing is only necessary for LCDs working in TN or IPS mode. LCDs working in vertical alignment (VA) mode use protrusions or patterned ITO to achieve the alignment. There are many issues involved in the rubbing process so a variety of different methods have been developed over the years.

After rubbing the panel is sealed. The purpose of this process step is to fix the alignment of the array with the color filter. After sealing, spacers, i.e. spheres or rods made of silica, resin or glass fibers are applied. The spacers are used to keep a constant distance between the array and color filter substrate. Different methods are used for the application of spacers. Previously, spherical spacers were applied using spraying. Today column spacers are used. These spacers are applied between the sub-pixels and will therefore not be visible or degrade the contrast and uniformity of the display which was previously a problem. The liquid crystal is then injected, the panel is sealed and the polarizers are applied.

In the module process the driver integrated circuits are bonded to the glass module. Today some of the driver electronics are integrated directly on the array substrate. This reduces costs and at the same time allows the bezel (i.e. the edge) of the display to be made thinner. However this will also lead to higher demands on the photomask used for defining the pattern. Good CD uniformity is much more difficult to achieve because the pattern defining the drivers is completely different from the pattern defining the TFTs in the active pixel array. After this step the backlight module is mounted. Before, Cold Cathode Fluorescent lamps (CCFL) backlights were used, but today backlight units of LEDs are much more common. Finally it should be mentioned that the description of the LCD process given above might differ in details among different panel makers.

Over the years the size of the display panel has grown and we refer to different size generations. Today we see generation 10 (G10) being manufactured with mother glass sizes of 2.6 x 3.1 meter. In the lithography steps mask aligners are used for positioning the masks in relation to previously exposed patterns. In the array process, projection aligners made by Nikon or Canon expose the pattern of the photomask onto the resist-covered mother glass. In the color filter process, proximity aligners or projection aligners are used. The mask size is approximately a quarter of the mother glass. Therefore an array of four to six mask exposures is required to expose the entire substrate area. A complete exposure of a mother glass with one mask takes approximately 70 seconds. The systems used for the lithography step are enormous in size. Shown in figure 10 is a G10 aligner developed by Nikon.
Figure 10: The FX-101S mask aligner used for 2.6 x 3.1 m$^2$ generation 10 mother glass sizes. Source: Nikon.
3 Mask writer and measurement tool

3.1 Mask writer

The photomasks discussed in the previous chapter are produced by dedicated systems specially designed for writing patterns on quartz or soda lime glass plates. The backplane made in the array process has the most demanding pattern. The Micronic Mydata mask writer systems LRS15000 and Prexision are used both for the array and for the Color filter (CF) mask manufacturing [21]. The Prexision8 and Prexision10 are the most advanced mask writers used today for LCD and OLED array patterns (figure 11). These systems use several scanning laser beams of 413 nm wavelength to expose the pattern into the resist and chrome coated glass plate by using a so-called microsweep with a scan length of approximately 200 µm. A very high degree of pattern accuracy is achieved.

Figure 11: The principle of the Prexision mask writer. Source: Micronic Mydata.
A diffractive optical element (DOE) is used to split the incoming laser beam into several sub-beams separated by small angular deviations [22][23]. In the Prexision system, 11 sub-beams are used. The sub-beams enter an Acousto-Optic Modulator (AOM) that is used as an analog switch for each individual beam. The AOM switching is controlled by data from the data path, which has information defining the pattern being written. Optics are used to merge the beams from the DOE to be overlapping (but still separated in angle) in the Acousto-Optic Deflector (AOD), which is the most critical component in the optical path. In figure 12 the principle of the AOD is shown.

![Figure 12: The principle of the deflection of the laser beam using an AOD in the optical head. By applying an ultrasound wave in the frequency range 150-250 MHz through a transducer to the crystal the sub-beams are deflected in the Y-direction. The incoming beam is Bragg diffracted by the incoming modulating acoustical wave. Source: Micronic Mydata.](image)

The AOD is a crystal made of tellurium dioxide (TeO$_2$). When an ultra-sound wave with a frequency span of 100 MHz is applied it is possible to generate a linear change in the angle of the output laser beam [24]. By applying this frequency span an effective deflection angle of 4 degrees is obtained. In Figure 12 the principle for one beam is shown. In practice, several beams are deflected simultaneously and separated in the X-direction. Depending on the focal length of the system, microsweeps with different lengths are generated. For the systems with the currently highest resolution the length of the microsweep is 200 µm. In the writing process the X-bridge is moved one stroke in the X-direction while at the same time the AOD is scanning several microsweeps in the Y-direction, creating an exposed “scan-strip” in the photoresist. Subsequently the Y-stage is moved slightly less than one scan-
strip width in the Y-direction so that there is a small overlap between scan-strips. This scheme is repeated until the entire mask pattern has been written. The stage position is controlled by two separate interferometer systems using one HeNe laser source with a wavelength $\lambda$ of 632 nm. To achieve maximum stability against temperature drifts the Y-stage and some other parts of the system are made of Zerodur, which is a sophisticated glass composite material that has been optimized to have essentially zero temperature expansion in the temperature range 20 – 25 ºC.

3.2 The MMS15000 metrology tool

The 2D ultra-precision metrology tool MMS15000 is built on the same principles as the Prexision mask writer [2]. In figure 13 the layout of the MMS15000 is shown.

Figure 13: The principle layout of the MMS15000 metrology tool
Source: Micronic Mydata.

As seen in the figure the machine uses a single laser beam with a wavelength of $\lambda$=532 nm for the measurement. The laser beam power is adjusted using an
AOM in a similar way as in the writer. The measurement beam scans the photomask using a microsweep, and the reflected beam is picked up by a detector in the back focal point of the optical system. In figure 14 a more detailed description of the involved hardware is shown.

**Figure 14:** The principle of the optics in the metrology tool MMS15000. Source: Micronic Mydata.

We do not use a CCD or CMOS sensor to grab images from the plate. Instead, recordings are created from information in the time domain. The technology used here is based on the principle of random sampling which is thoroughly described in Paper B. This subject will later be described in more detail in this thesis summary. The incoming beam is deflected by the AOD and back-reflected by the chromium pattern on the photomask. A beam splitter transfers the reflected beam to the detector which converts the beam intensity to an analog electrical signal. This signal is then further amplified and filtered in the electronic hardware. The back focal point of the optics is located in the AOM so the signal, separated by the beam splitter placed close to the AOM, will therefore not contain any spatial information. This is because the reflected beam from the photomask will follow the same optical path as the incoming beam from the modulator. Thus, the only information the signal carries is intensity variations in the time domain.
Practically all modern image displays are built up by pixels. A pixel is further divided into three or more sub-pixel elements. The most common pixel configuration is three sub-pixels of red, green and blue (RGB) color. There are also other configurations using more sub-pixels in order to enhance the quality of the display. As an example a pixel is sometimes configured using four sub-pixels RGBW where an extra white pixel is added. The size and location of a pixel on the display is of primary importance for obtaining an image with high quality and with no defects. To achieve the high quality, the lithography process based on photomasks has to be performed in a nearly perfect way. In practice a mask writer with high performance and a production staff with deep process knowledge are needed to achieve that end. A common rule of thumb regarding quality in this area is that about 50% of the outcome depends on the mask writer and 50% depends on the processing of the photomask. Both the X,Y pattern placement and the CD variation must be maintained within very tight specifications. Today the placement accuracy of features over the whole mask area that can be as large as 1.6 x 1.8 m² must be better than 150-200 nm (3σ), with CD variation not exceeding 50 nm (3σ).

Besides these dimensional specifications that can be measured using the MMS15000, an even more critical visual quality measure of photomasks has to be passed. This is called the Mura test. Mura is a Japanese word meaning not perfect and here it refers to the non-perfect diffraction of reflected light from the mask. This subject has many influences on the methods used for writing the pattern. At least until now, quantitative measurements of Mura have failed, so this test has to be done with visual inspection by skilled experts. Mura is indeed a very interesting subject of its own, and for this reason it will be discussed in a chapter later in this thesis.

Metrology of the large area photomasks discussed earlier is a significant challenge. The MMS15000 is a specially designed 2D metrology tool used for verification of geometrical placement and CD of the patterns on the photomask. Besides geometric placement uncertainty, called registration in the display industry, the overlay between different photomasks is also measured. In the next section the different properties of the photomask and how it is measured will be defined.
4.1 Terminology for photomask quality assessment

Large area mask manufacturing has been developed within a few companies. The terminology for quality assessment of these masks has therefore evolved from discussions within this small group. Already well-known notions like “registration” in graphical printing has been adopted but gotten a different meaning in mask quality evaluation. To clarify the terminology used in large area photomask industry we introduce the specific definitions here.

Registration: Registration is a special term used in the display industry and is a measure of the position uncertainty of a feature on a mask. It is defined according to eq. (1) and eq. (2) as three times the root mean square average distance, in the X-direction and in the Y-direction respectively, between the nominal positions and the measured positions of a set of reference points covering a certain area in the X,Y plane of the writer or the measurement tool. In practice registration is measured as the difference in X and Y between measured and nominal Cartesian coordinates for a set of calibration marks on a reference plate called the Golden Plate (GP). The calibration marks are placed in a \( n_x \cdot n_y \) matrix covering the specified area of the stage. Before registration is calculated, the translation and rotation of the measured data are removed. Registration \((R_X\) and \(R_Y) \) can be measured through a traceable metrology chain based on a one-dimensional reference that has been measured by the metrology institute MIKES [25].

\[
R_X = 3 \cdot \sqrt{\frac{1}{n_x \cdot n_y} \sum_{i=0}^{n_x-1} \sum_{j=0}^{n_y-1} (M_x(i_x, j_y) - Ca_x(i_x, j_y))^2} \quad (1)
\]

\[
R_Y = 3 \cdot \sqrt{\frac{1}{n_x \cdot n_y} \sum_{i=0}^{n_x-1} \sum_{j=0}^{n_y-1} (M_y(i_x, j_y) - Ca_y(i_x, j_y))^2} \quad (2)
\]

where

\(M_x(i_x, j_y)\) is the measured X location of the mark at the matrix location \(i_x, j_y\).

\(Ca_x(i_x, j_y)\) is the corresponding Cartesian X location of the mark on the GP.

\(M_y(i_x, j_y)\) is the measured Y location of the mark at the matrix location \(i_x, j_y\).

\(Ca_y(i_x, j_y)\) is the corresponding Cartesian Y location of the mark on the GP.

Overlay is a measure of the how well identical mask patterns can be written to exact locations on different plates and therefore represents a measure of the reproducibility in the writing process. Overlay is measured as the 3\(\sigma\) deviations of the differences obtained between the location measures of a
single mask relative to the average locations of the same mark on three or more masks. The overlay number (Ovl) is reported separately in the X and Y directions. Overlay can be measured in the mask writer itself by its built-in measuring unit but can also be independently measured using e.g. the MMS15000 metrology system [2]. In the specification of the mask writers or the MMS15000, overlay is defined as described above using three measurements. For each point the following properties are calculated.

\[ \text{Ovl}_a(i_x, j_y) = a(i_x, j_y) - \frac{(a(i_x, j_y) + b(i_x, j_y) + c(i_x, j_y))}{3} \]  
\[ \text{Ovl}_b(i_x, j_y) = b(i_x, j_y) - \frac{(a(i_x, j_y) + b(i_x, j_y) + c(i_x, j_y))}{3} \]  
\[ \text{Ovl}_c(i_x, j_y) = c(i_x, j_y) - \frac{(a(i_x, j_y) + b(i_x, j_y) + c(i_x, j_y))}{3} \]  

Where \( a(i_x, j_y) \), \( b(i_x, j_y) \) and \( c(i_x, j_y) \) are measurements of a mark’s X or Y coordinates in the matrix location \( i_x, j_y \).

We now assume that the measurements are samples from a normal distribution \( N(\mu, \sigma) \) with the mean \( \mu \) and the standard deviation \( \sigma \). We can express \( \text{Ovl}_a(i_x, j_y) \) as three independent measures as:

\[ \text{Ovl}_a(i_x, j_y) = \frac{2}{3} \cdot a(i_x, j_y) - \frac{1}{3} \cdot b(i_x, j_y) - \frac{1}{3} \cdot c(i_x, j_y) \]  

Since the measurements of \( a, b \) and \( c \) are independent, this leads to the fact that the variance \( \sigma^2_{\text{Ovl}} \) can be expressed as:

\[ \sigma^2_{\text{Ovl}} = \frac{4}{9} \cdot \sigma^2 + \frac{1}{9} \cdot \sigma^2 + \frac{1}{9} \cdot \sigma^2 \]

Thus, the standard deviation for any of the overlays can be expressed as:

\[ \sigma_{\text{Ovl}} = \sqrt{\frac{2}{3} \cdot \sigma} \]  

**CD**: is the abbreviation for critical dimension and it is a measure of the size of a written pattern on the photomask. This specification is normally split into Critical Dimension Uniformity (CDU) and Critical Dimension Linearity (CD linearity). CDU is expressed as the (max – min) range or as 3\( \sigma \), where \( \sigma \) is the standard deviation of the differences between the measured line width and its average width for a certain width of the line. CD linearity is a measurement of the difference between a set of a range of line widths and their nominal widths. CD is measured in a dedicated CD measurement tool [26].

**Measurement tool**: A coordinate measurement machine like the MMS15000 specially designed for measurements of flat artifacts e.g. photomasks. The Cartesian coordinate system of the machine is traceable to a measurement
standard. In our case the absolute measurements are verified by interferometry but they are also traceable to an artifact developed and verified by MIKES, a metrology institute in Finland [25][27]. Details regarding this will be discussed later in this thesis.

**Golden Plate:** A Golden Plate (GP) is a reference chromium patterned plate, made of quartz glass that is used as the registration standard in a mask writer and/or the measurement tool (see figure 15). To qualify a plate to become such a reference it must be measured in a traceable measurement tool with accuracy better than the system it is to be used for in a calibration. On the GP a set of measurement marks are arrayed in a matrix pattern. The X and Y pitches, i.e. repeated distances between the marks, are typically 20-50 mm. These pitches are usually the same in the X and Y direction.

![Figure 15: A typical GP with strongly exaggerated deviations between measurement mark positions and the Cartesian grid with crosses in a matrix pattern. The line width of the crosses is typically 15 µm.](image)

The deviations in X and Y to the Cartesian grid for each cross, the so called GPdata is measured in a traceable measurement tool. When these deviations are known the plate is qualified to be a Golden Plate. The cross marks are made of highly reflective chromium surrounded by low reflectance quartz glass. Typically \( n_x \) and \( n_y \) are in the range of 20-50.

**Target system:** The system of interest for the calibration of overlay, registration and CD. This is normally the mask writer but can also be a measurement tool. The writer referred to in this thesis also has a built in measurement capability, but it is more cost-effective in mask manufacturing to use a dedicated metrology tool such as the MMS15000.
4.2 Traceability of the MMS15000

As will be discussed later we have developed a robust self-calibration procedure that allows compensation for all systematic deviations of the MMS15000 measurement stage except for the absolute scale. Operating the laser interferometers at optimal conditions yield a repeatability of the order of ten nanometers over the entire stage. However, we cannot claim absolute traceability without measuring a calibrated artifact, measured at a National Metrology Institute. Unfortunately none of these institutes can deliver a large area 2D artifact with a specified uncertainty considerably less than 100 nm. As will be shown in the conclusion, the uncertainty obtained in our registration measurements is far better than the length-scale standard that we have adopted for showing that the MMS15000 is traceable. In other words the uncertainty of traceable measurements is considerably larger because of the relatively large uncertainty of the artifact itself. We therefore claim “registration” as our measure of photomask X,Y uncertainty instead of traditional and traceable length-scale uncertainty.

The length-scale artifact used for the traceability calibration of the MMS15000 is a quartz reference provided by MIKES in Finland [25]. In this calibration process we also make an absolute calibration of our temperature sensors. The atmospheric pressure, CO$_2$ and humidity sensors are all absolutely calibrated at one point in the calibration process. From this point on we rely on the instruments’ calibration. We use the Edlén equation for the correction of the refraction index of air [28][29]. In this way we achieve traceability of the interferometers. The laser interferometer we use has an uncertainty in frequency of 0.1872 MHz (σ), corresponding to an uncertainty of 1.186 ppb (3σ) of the length scale. The temperature stability of the laser wavelength is specified to be 0.5 MHz/ºC which corresponds to 3.2 ppb/ºC. The latter temperature stability specification can be neglected since the chamber temperature of the system is controlled to better than ±0.01 ºC.

In a measurement of one of our quartz references we have a certified expanded uncertainty from MIKES of 116 nm (2σ). In figure 16 one of the quartz references is shown after it has been placed on the MMS15000 measurement stage for verification of the calibration. When transferring this uncertainty to the calibration of the machine we factor in the measured repeatability of 10 nm (3σ). So the uncertainty will then be: \( \sqrt{174^2 + 10^2} \approx 174 \text{ nm (3σ)} \) i.e. entirely determined by the uncertainty of the length-scale artifact.
Figure 16: The quartz reference artifact we use for absolute calibration of the MMS15000. The length of the reference scale is 750 mm. We use three integrated temperature sensors and one extra free hanging sensor (not shown in the image) for the chamber temperature measurement in the calibration process. We use four calibration plates for microsweep and CD calibration. One of the plates has a traceable CD reference pattern.
Source: Micronic Mydata.

In addition to the quartz reference we use a traceable certified reference standard for CD calibration. This reference is used for calibration of an in-house CD measurement tool that is used for transferring the traceability to the calibration plates mounted in the machine.

4.3 Mura
The high complexity of the photomasks is further augmented by requirements from the human visual perception [30]. If we watch a display from some distance it is not possible to resolve each individual pixel. A small systematical change in the grayscale of a display in operation is however easier to observe. Defects causing grayscale changes, commonly called Mura, may have different causes. Besides the driver electronics, the process or the photomasks used for the manufacturing of the display may also generate Mura. In the photomask case extremely small variations in linewidth, registration or overlay between different photomasks may cause thin lines,
spots or regular repeatable patterns in the finished display. As an example, systematic CD errors in the range of 10-20 nm in a pixel pattern with a pitch of 80-120 µm may cause Mura. Human visual perception is very sensitive to such variations in gray scale. About a 0.2% intensity change in grayscale can be seen at a distance of 50 cm if the spatial period of the pattern has a period of 2.4 mm. A graph illustrating the visual sensitivity for Mura is shown in figure 17 [Paper A].

**Figure 17**: Measured human visual perception of contrast sensitivity of gray levels as a function of spatial period at a viewing distance of 50 cm of a 17 inch display. The luminance level is 100 cd/m². The peak sensitivity for a periodic grayscale variation is 0.18% at a 2.4 mm period [31].

In figure 18 an image of a typical “line Mura” on a photomask is shown. As seen in this example thin horizontal lines can easily be observed in the image. The image is obtained by illuminating the mask at an oblique angle with a collimated white light beam.
Figure 18: An example of so called line Mura. The Mura can be observed in diffracted light by using a collimated light source. Source: Micronic Mydata.

In this example the Mura is generated by a small systematic shift in registration, i.e. the position of parallel horizontal lines in the pattern. The measured Mura pitch was 1 mm and it is caused by a pitch error of the pattern of 50 nm. The Mura effect in figure 18 is just one example of several kinds of Mura that may occur because of imperfections in the mask writing process. Over the years a number of different Mura effects have been classified and they have also been traced back to different aspects of the writer hardware and software. So each type of Mura has its own signature and can therefore be analyzed and tracked to different error sources in the mask writer. The most common type of Mura is generated when the writer pitch and pattern pitch differs. In such a case power variations in the microsweep, not visible because of the small spatial distribution, may be “expanded” to a Mura pitch much more evident to the human eye. To exemplify this effect we will perform the following simulation:

In this simulation, the distance between identical features of the pattern, the so-called pattern pitch, is 80 µm in the X and Y directions. The writer has another natural pitch called the writer grid. In the Y-direction this grid corresponds to the scan strip width of 200 µm. When the pattern pitch does not coincide with the writer grid, the difference in spatial frequency will
enhance the risk for Mura. With a known writing grid ($wg$) and a known pattern pitch ($pp$) the Mura pitch ($mp$) can be calculated as:

$$mp = \frac{1}{\frac{1}{wg} - \frac{1}{pp}}$$

(8)

As can be seen in the expression above the Mura pitch $mp$ will expand towards infinity when $wg$ approaches $pp$. One way to suppress Mura is therefore to adjust the writer grid to be the same as or an integer multiple of the pattern pitch. We exemplify this effect with the help of figure 19. Here we illustrate what happens in the Y-direction when the pattern pitch differs from the writer grid and we simultaneously have a small intensity error in the writing laser at the end of the microsweep.

**Figure 19:** The effect of a non-linear laser intensity profile of a maskwriter may be revealed as the illustrated visual appearance in the figure, when the writing grid differs from the pattern pitch. The effect is highly exaggerated in the figure.
When a pattern edge coincides with the segment of the microsweep having different intensity (shown as the *Intensity error* in figure 19) the edge of the pattern will be exposed with a slightly different power. The effect of this will be to move the edge relative the surrounding edges. Another way to express this is that all features along the microsweep will have the same CD except one feature that has been affected by the intensity error. The writing grid in the Y-direction is the scan strip width $wg$. So in this example the small difference between $wg$ and $pp$ will have different impact on the location of the pattern feature edges within two microsweeps. In this example, due to the difference between $pp$ and $wg$ the features will be located in the same position within the microsweep after twelve pattern pitches, and so $mp = 12 \cdot pp$. The Intensity error causing the CD variation along the microsweep will thus be expanded to a much more critical spatial period for the human visual system. When the spatial period of $mp$ is in the range of 2.5-20 mm the sensitivity for this error is high enough to cause visual Mura.

Mura is extremely difficult to measure in a quantifiable way [32]. Even when using advanced metrology based on a huge amount of images from the pattern it is still not an acceptable measure of Mura. Instead special experts in the mask shop do a qualified final visual inspection of the Mura quality. Dedicated tools are only used for pre-selections of good and rejected photomasks in the process but before a photomask is shipped it will always be manually inspected.

Besides the Mura generated by the writer, Mura can also be caused by the processing of the photomask. This happens after the pattern has been exposed into the photoresist when the plate is developed and etched. Instabilities in this process may cause visual defects called process Mura.
5 Metrology

5.1 Basic system analysis

When calibrating the target system that may be a maskwriter or a MMS15000 the following principal assumptions are used:

\[
\text{Registration} = \text{Systematical errors} + \text{Random errors} \quad (9)
\]
\[
\text{Overlay} = \text{Random errors} \quad (10)
\]

As already mentioned we use the word “Registration” for the deviation of a written pattern to the Cartesian grid. When the pattern is measured in a metrology tool the same word is used as a measure of the geometrical placement quality of the photomask. In other areas in the industry the word “uncertainty” is used for the same purpose [paper A].

Our definition of overlay, i.e. a measure of differences of several photomasks, has the same meaning as reproducibility. In a calibration of the MMS15000, repeatability is the most important property. Repeatability is a measure of several measurements over the entire surface of a GP or a photomask without moving it from the stage between the measurements. The 3σ of the variation between the measurements is defined as the repeatability.

As seen in equations (9) and (10) above Overlay always yields better values than Registration. As we reduce the systematic errors, the fundamental limit of Registration is the Overlay. The systematic errors are divided into different categories that describe the roots of the systematic errors in the target system. The systematic errors can be corrected by using a scalar, a correction in one dimension or two-dimensional correction maps as shown in table 2.
<table>
<thead>
<tr>
<th>Category of systematic error</th>
<th>Root cause</th>
<th>Correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global scale</td>
<td>Incorrectly compensated wavelength used in the interferometer system.</td>
<td>Scalar</td>
</tr>
<tr>
<td></td>
<td>Error in temperature of the mask blank during exposure or measurement.</td>
<td></td>
</tr>
<tr>
<td>Global orthogonality</td>
<td>Misalignment of the angle between X and Y axis in the target system.</td>
<td>Scalar</td>
</tr>
<tr>
<td>Global stage bow</td>
<td>An error in straightness of the mechanical X and Y axis due to limitations in the manufacturing of different stage materials.</td>
<td>One-dimensional in each direction X and Y</td>
</tr>
<tr>
<td>Higher order local errors</td>
<td>Clamping of the photomask on the stage.</td>
<td>Two-dimensional</td>
</tr>
<tr>
<td></td>
<td>Second and higher order flatness variations of the stage.</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2**: The table shows the different systematic error categories, root causes and how corrections are performed.
Random errors may also be divided into different categories and causes as presented in table 3.

<table>
<thead>
<tr>
<th>Category of random error</th>
<th>Cause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interferometer</td>
<td>Fluctuations of the refractive index when the wavelength is measured caused by temperature, humidity, CO₂ and pressure variations.</td>
</tr>
<tr>
<td>Mechanical</td>
<td>Residual and hysteresis errors in the mechanical system. Temperature variations of mechanical parts.</td>
</tr>
<tr>
<td>Electrical</td>
<td>Electrical noise affecting the hardware control servos in the mechanical system. Also electrical noise has an impact on the writing and the measurement laser control system.</td>
</tr>
<tr>
<td>Writing and Measurement laser</td>
<td>High frequency variations in power and laser beam angle due to temperature and pressure variations.</td>
</tr>
</tbody>
</table>

Table 3: The different random error categories and their causes.

A summary of the requirements of the photomasks for different display types is presented in table 4. As mentioned Mura cannot be measured objectively and for this reason we give a subjective indication of how critical this demand is.

<table>
<thead>
<tr>
<th></th>
<th>Registration (nm 3σ)</th>
<th>Overlay (nm 3σ)</th>
<th>CD (nm 3σ)</th>
<th>Mura demand</th>
<th>Resolution (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasma</td>
<td>250</td>
<td>200</td>
<td>120</td>
<td>Medium</td>
<td>3.0</td>
</tr>
<tr>
<td>LCD</td>
<td>150</td>
<td>120</td>
<td>60</td>
<td>Severe</td>
<td>0.75</td>
</tr>
<tr>
<td>OLED</td>
<td>150</td>
<td>120</td>
<td>50</td>
<td>Severe</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Table 4: The requirements of the photomasks for the most common display types.
Random errors cannot be compensated for. The only way to keep these errors to a low level is by proper design of mechanical, electrical and optical parts and in addition by keeping the environment in a very stable condition. In our case the system is kept in a temperature controlled chamber. The maximum allowed variation in temperature is ± 0.01 °C.

In principle registration should be traceable to a measurement standard calibrated at a national metrology lab. However, in the mask making business absolute registration is less important due to commercial reasons. It is better for a mask maker to deliver a complete set of several masks of a design to their customer. An efficient way to do this is to use a “company” registration standard that differs to some extent from their competitors. In this way their customer cannot mix critical masks from different suppliers. Another reason for the mask users to buy a set of masks from one and the same supplier is that they can expect much better overlay in such a case, especially if all the masks in the set are written by the same writer.

5.1.1 One-dimensional corrections

Systematical errors over the stage can be divided into four principal deviations [33].

- $dX(Y)$: Y-bow
- $dX(X)$: X-local scale
- $dY(X)$: X-Bow
- $dY(Y)$: Y-local scale

In figure 20 X and Y bow curves are shown. These curves illustrate deviations from straightness of the mechanical axes in the system.

Figure 20: Illustration of X and Y bow curves and how these can be measured from a GP. The linearly fitted tilt is shown for the Y-bow. The X and Y bow curves and the misalignment angle between the axes are highly exaggerated.
The deviations illustrated in figure 20 are the most important to eliminate by calibrations and they can easily be obtained from a measurement of a GP. The corrections of these deviations can be done using vectors or one-dimensional functions. The measurement marks of a GP are configured as a matrix with the nominal pitches \( xPitch \), \( yPitch \). In the next section we present the equations used for obtaining the derivations in local scale, bow and orthogonality based on a measurement of a GP containing \( n_x \cdot n_y \) measurement marks (refer to figure 15).

\[
m_x(i_x,j_y) = M_x(i_x,j_y) - i_x \cdot xPitch \quad (11)
\]

\[
m_y(i_x,j_y) = M_y(i_x,j_y) - i_y \cdot yPitch \quad (12)
\]

where \( m_x \) and \( m_y \) represent the local deviations at a measurement mark at row \((i_x)\) and column \((j_y)\) in figure 15. \( M_x(i_x,j_y) \) and \( M_y(i_x,j_y) \) are the absolute measurements of the mark \( i_xj_y \).

From equation (11) and (12) we calculate the deviation vectors according to equations (13)-(16) below.

\[
dX(j_y) = 1/n_x \cdot \sum_{i_x=1}^{i_x=n_x} m_x(i_x, j_y) \quad (Y\text{-bow})
\]

\[
dY(j_y) = 1/n_x \cdot \sum_{i_x=1}^{i_x=n_x} m_y(i_x, j_y) \quad (Y\text{-local scale})
\]

\[
dY(i_x) = 1/n_y \cdot \sum_{j_y=1}^{j_y=n_y} m_y(i_x, j_y) \quad (X\text{-bow})
\]

\[
dX(i_x) = 1/n_y \cdot \sum_{j_y=1}^{j_y=n_y} m_x(i_x, j_y) \quad (X\text{-local scale})
\]

From the derivations, linear regression lines are calculated using the equation:

\[
u = y \cdot m + c \quad (17)
\]

Where \( y \) is a tilt and \( m \) is an index \( i_x \) or \( j_y \) depending of what deviation vector we use for the linear fitting.

From the derivations expressed in equation (13) and (15) linear regression lines are calculated having slope angles:

\[
X_{angle} = \tan(y_x / xPitch) \quad (18)
\]

\[
Y_{angle} = \tan(y_x / yPitch) \quad (19)
\]
where $\gamma_x, \gamma_y$ are the linear tilts of the deviation vectors $dX(j_y)$ and $dY(i_x)$ respectively.

From the equations (18) and (19) the orthogonality (i.e. the difference in angle between the X-axis and Y-axis) is calculated as:

$$\text{Orthogonality} = X_{\text{angle}} + Y_{\text{angle}}$$  \hspace{1cm} (20)

according to our definition of the angles described above. As seen the orthogonality is defined as the difference in angle between the X-axis and Y-axis when the default angle of $\pi/2$ has been subtracted.

The Mechanical X-axis is the reference axis in both the writers and the MMS15000 measurement system. This means that the orthogonality error is adjusted by the Y-axis angle only. The start position of successive scan stripes is therefore adjusted in the X-direction along the mechanical Y-axis as shown in Figure 21.

**Figure 21:** The principle used for adjusting the orthogonality in the writer by moving the start position for successive scan stripes away from the mechanical Y-axis. The angle in the figure is highly exaggerated.

The X-bow and Y-bow shapes is extracted from the deviation vectors $dY(i_x)$ and $dX(j_y)$ after the linear tilt angle has been removed (see dashed line in figure 20). Also any constant offset is removed. After this, the remaining deviations are saved in lookup tables and are used as corrections for the two mechanical axes. In the Y-direction, the start of the scan strip is moved in addition to the already mentioned orthogonality adjustment, for correction of
the Y-bow. The X-bow correction is done by adjusting the Y-start position of successive microsweeps in real-time during the scan of a stripe.

The scale deviations in the X and Y directions correspond to the slopes of the linear regression lines to the vectors \( dX(i_x) \) and \( dY(j_y) \). The unit we use for scale deviations is expressed in parts per million (PPM) and is defined as follows:

\[
X_{\text{scale}} = \frac{d_x}{x\text{Pitch}} \cdot 10^6 \\
Y_{\text{scale}} = \frac{d_y}{y\text{Pitch}} \cdot 10^6
\]  
(21) (22)

where \( d_x, d_y \) is the linear tilt of the vectors \( dX(i_x) \) and \( dY(j_y) \) respectively. In practice the scale deviation is compensated by adjusting the nominal wavelength of the interferometer system in each direction separately. The adjusted scale factors are calculated as:

\[
\lambda_{cx} = \lambda_{cc} \cdot (1 - \frac{X_{\text{scale}}}{10^6}) \\
\lambda_{cy} = \lambda_{cc} \cdot (1 - \frac{X_{\text{scale}}}{10^6})
\]  
(23) (24)

where \( \lambda_{cc} \) is the temperature-, pressure-, humidity- and CO2- compensated wavelength of the interferometer laser.

5.1.2 Two-dimensional corrections

After the one-dimensional corrections have been performed using the deviation vectors described in the previous section, the remaining deviations are corrected for by using a two-dimensional correction map. The higher order deviations are caused by both mechanical and optical imperfections. The systematic contributions of these errors are extracted from the measured 2D data after the X-bow, Y-Bow and orthogonality have been removed. The correction map contains deviations separated in the X and Y direction as illustrated in figure 22.
Higher order errors are defined as all deviations that cannot be corrected for by linear functions. Since these deviations are local they also need to be corrected for locally. The correction map is obtained by the real measurements on a matrix of measurement marks. This map is organized as a matrix with a smaller pitch in X and Y in comparison to the pitches of the GP measurement marks. Each entry in this matrix defines an X and Y correction vector at a certain position on the stage that has been calculated from data in the GP measurement. When positioning to a certain X and Y coordinate, bilinear interpolation is done in this correction map.

The calibration methods presented above have been used for calibrating the mask writers for many years. We needed to rely on our own registration performance since there were no large area metrology tools in the market or at metrology labs that were good enough to be used as a reference system. For this reason a strong demand from the mask manufacturing market motivated us to take on the challenging task of developing a metrology tool with even better performance than the writers. The result from this demand is the metrology tool MMS15000. To enhance the calibration we developed and implemented self-calibration and Z-correction for the purpose of establishing a highly accurate reference system. This system could then be used for finding the geometrical shape of the GP used as reference in the writer. Self-calibration and Z-correction will be discussed later in this thesis.

So far we have been analyzing large area writer and measuring system performance regarding requirements, and how we can control linear and non-linear deviations in the mechanical positioning system. In the following section we will present the new measurement technique we have implemented to achieve the extreme demands discussed above.
5.2 Random phase measurement technique

Instead of using ordinary CCD or CMOS images of patterns on the photomask we introduced a more efficient method in the writer and the MMS15000 for grabbing one- or two-dimensional patterns. The method is based on random sampling of events in the time domain. This method was a natural choice from the beginning since we already used this technique for calibration of the microsweep. Before discussing how this method is used for grabbing 2D structures we start with the one-dimensional case, which we thoroughly describe in paper B.

5.2.1 One-dimensional measurements

The maximum accepted non-linearity error of the microsweep is <10 nm. As an example we use the actual parameters from the Prexision platform [21]. The spatial length of the microsweep is 200 µm and the total time to perform one sweep is 25 µs. About 64% of this time (16 µs) is efficiently used for the scanning across the 200 µm scan-strip with the laser beam. The maximum non-linearity error of 10 nm will therefore correspond to a maximum error in time of $\frac{10}{200000} \cdot 16000 = 0.8$ ns. To be able to measure such a short time using ordinary frequency measurement techniques one needs a clock frequency of 1.25 GHz. To handle this high frequency in the electronic hardware years ago was a real challenge. For this reason we developed another method that did not need such high clock frequency. The method is instead based on the principle of a statistical random phase clock combined with acousto-optical deflection. In figure 23 the electrical signals involved in a time measurement are shown.

![Figure 23](image.png)

**Figure 23:** The signals involved in a time measurement.
The microsweep is triggered by the Start Of Sweep (SOS) signal. During a scan a detector records the reflected signal from the mask. The output from the detector is an analog signal revealing the reflectance signature of the pattern. The glass area generates a low reflectance compared to the chromium areas of the pattern. A threshold of approximately 50% is used in the hardware to generate the digitized signal. The time \(\text{event}^{\text{ns}}\) measured is referred to the SOS pulse as shown in the figure 23. Each positive or negative flank of the digitized signal generates an event. The elapsed time to an event can easily be scaled to spatial locations since the length of the microsweep (the ruler) is known. Thus, we make use of a statistical approach with random sampling for the detection of edge locations. By using a measurement clock with a phase that is random relative to the time interval we want to measure, it is possible to estimate the location of \(\text{event}^{\text{ns}}\) to any preferred uncertainty. The method makes use of a massive averaging of counts of measurement clock pulses having different phases in the time interval \(\text{event}^{\text{ns}}\). The principle of the random sampling technique is shown in figure 24 and described in detail in paper B.

The repetition rate of the microsweeps is 40 kHz. The SOS signal that triggers the microsweep also resets a measurement clock counter. Four
measurement clock phases are shown in figure 24. Since the phase of the measurement clock is random relative to the SOS pulse (that is synchronized with the microsweep) it is only possible to count $w$ or $w + 1$ counts in the $event_{ns}$ interval. No other counts are possible. If this procedure is repeated $N$ times, the distribution of the number of measurements clock pulse flanks, $w$ and $w+1$ will give the position of where in the microsweep the event takes place, and that is a measure of the length of the interval. In the example shown in figure 24 we count 2 flanks in the first measurement, 1 flank in the second, 2 flanks in the third and 2 flanks in the fourth measurement. The average count here $(2+1+2+2)/4 = 1.75$ clock flanks is an estimation of the interval time $event_{ns}$. It can be shown that the uncertainty of the estimation ($u_{Yave}$) of $event_{ns}$ can be expressed as:

$$u_{Yave} = \frac{1}{\sqrt{N}} \cdot 0.5 \cdot tm$$

(25)

where $N$ is the number of repetitions and $tm$ is the measurement clock period [34].

In practice the microsweep is repeated with a frequency of approximately 40 kHz, i.e. a period time of 25 µs. Normally 10 000 microsweeps are used for scanning one or several edges inside the microsweep length of 200 µm. So the total measurement time for the events will be $(0.025 \cdot 10000) = 250$ ms with an uncertainty ($u_{Yave}$) of 0.125 ns according to equation (25). With a scanning speed equivalent to 12.5 nm/ns the corresponding uncertainty in the spatial domain is $0.125 \cdot 12.5 \approx 1.6$ nm.

We will now exemplify the result of a pitch measurement using the random sampling method (see figure 25 and 26). In this example some lines of a calibration raster with 4 µm pitch have been measured using a MMS15000 (using the same parameters as for a Prexision). The cursor length, i.e. the length used for sampling along the raster line, is approximately 30 µm.
Figure 25: A schematic of the calibration raster used in the MMS15000. The non-linearity of the microsweep is tuned by using such a calibrated raster.

**Pitch repeatability test**

Figure 26: The result of a pitch repeatability test. 100 measurements were made of the pitch between two of the bars in figure 25. The standard deviation (σ) of the measurement is 0.53 nm. Each measurement takes 250 ms.

In the measurement results from figure 26 we see that the average distance is 4.003 µm instead of 4.000 µm. This 3 nm offset error was caused by the length scale setting of the microsweep, which was not perfectly tuned to 200 µm prior to the measurement.

The main purpose for the development of the one-dimensional random phase technique was to be able to calibrate the microsweep in the writer. Due to the outstanding performance of the random phase technique we have developed it a bit further to also be capable of measuring special calibration patterns with extension in two dimensions. The purpose of this was to use a GP to calibrate the writer stage in both dimensions.
5.2.2 Two-dimensional measurements

By using a specially designed measurement mark on the GP, shown in figure 27, it is possible to retrieve both X and Y information for registration calibrations using a scanning microsweep in the Y-direction.

**Figure 27:** The special four-fold rotationally symmetric measurement mark is shown at the left, and to the right are the corresponding signals obtained by scanning the “V” horizontally. The central cross is used by measurement tools capable of measuring horizontal and vertical lines like the MMS15000. The “V” mark is used for registration calibrations using a scanning microsweep in the Y-direction with a slight oscillation in X.

At the right hand side of figure 27 the principle of retrieving X and Y information from a “V” mark is shown. In a measurement of the “V” the optical head is oscillating in a sinusoidal motion with an amplitude $A$ of approximately 15 $\mu$m (half of the cursor length) around the center of the interval $\Delta x$. Inside this interval the microsweep randomly samples the four edges of the mark. Since the angle of the tilted leg is 45 degrees, the coordinate $x$ is the measured distance from the center of the mark to the intersection point of the legs marked as a circle in the lower right in figure 27. This distance is calculated as the measured difference in center of gravity of the right and left legs, $X_m-Y_m$. The X coordinate $X$ is calculated as the sum of $X_m-Y_m$ and the center of the interval $(x_0+x_1)/2$. The Y-coordinate $Y$ is simply measured as the distance from the local origin $(0,0)$ to the center of the vertical leg as shown in figure 27. The absolute coordinate of the mark is then calculated as the sum of the interferometer-controlled absolute position $X_{abs}, Y_{abs}$ corresponding to the local origin $(0,0)$ in figure 27 and the measured $X, Y$ position of the mark.
The repeatability of the method is presented in figure 28. In this test a full GP has been measured in a MMS15000 and the repeatability was verified. Each measurement takes approximately 45 minutes.

**Figure 28**: Visualization of the measurement repeatability in absolute position of 22 x 19 "V" marks in a Golden Plate matrix having the marks separated by 42,000 mm in X and Y. The pattern covers a total area of 0.882 x 0.756 m$^2$. The local scale at each intersection is 50 nm.

In the right hand side of figure 28 an example intersection is expanded. The separate measurements are presented as red, green and blue lines and crosses in different gray levels. The average of the measurements of a cross (an intersection in the plot) is marked as an encircled cross and functions as a local origin. The offset for each measurement relative to the local origin is presented in a local scale (here 50 nm). This scale is indicated by the black arrows at the right of figure 28. The deviating points are then connected by lines to their neighboring intersection deviating points for each measurement separately. This type of grid graph is a convenient way of presenting the deviations in a much finer scale compared to the absolute pitch of the pattern that is in mm. The purpose of the connecting lines is to visualize long wave systematic deviations and make it easier to interpret the long range deformations. In this test three measurements were made and the repeatability in X and Y was calculated as standard deviations by comparing each measurement with the average of all three measurements. The result is shown in table 5.
<table>
<thead>
<tr>
<th>X σ (nm)</th>
<th>Y σ (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6</td>
<td>3.0</td>
</tr>
<tr>
<td>3.4</td>
<td>2.8</td>
</tr>
<tr>
<td>3.0</td>
<td>2.7</td>
</tr>
</tbody>
</table>

**Table 5:** The obtained standard deviations of three measurements in the X and in the Y-direction respectively of the Golden Plate results shown in figure 28.

### 5.2.3 General two-dimensional measurements

The random sampling technique using specially designed marks is a very robust method for registration calibrations both in the writer and the MMS15000. However, a general metrology tool must be capable of measuring any kind of pattern. For this reason, the method was further developed to achieve this goal [Paper C].

In the one-dimensional case, we have presented that we use a measurement clock with random phase relative to the time interval we are measuring. In the general two-dimensional case, we now introduce the SOS trig pulses as a random clock relative to the λ/2 X-interferometer clock. The latter clock generates pulses from the interferometer fringes of the X-interferometer, when the stage is moving in the ± X-direction. Thus, the λ/2 clock is traceable to the absolute coordinate in the X-positioning system. Also, in the Y-direction, we use a similar λ/2 clock that is traceable to the absolute coordinate in the Y-direction. In the initiation phase, two λ/2 counters are reset, one in each direction. This reset is done using a Hall sensor detecting a mechanical zero reference position for both the X and Y interferometer system. In figure 29, the principle of two-dimensional random sampling is shown.
Figure 29: Principle of two-dimensional random sampling for edge detection of the shaded area. The microsweep in the Y-direction is identical with the one-dimensional case, using SOS as trigger and measurement clock pulses for flank detection of reflectance changes within the microsweep. In the X-direction the same principle is applied, but with the X-clock pulses being the trigger and the SOS pulses replacing the measurement pulses. The period of the measurement clock pulses and the SOS-pulses are shown as spatial distances in the figure. The size of the rectangle is not drawn to scale.

As seen in figure 29 the problem in the X-direction is similar to the case in the Y-direction but rotated 90 degrees. In the Y-direction the SOS is the reference and the measurement clock is the random clock used for the measurement of a Y-coordinate. In the X-direction the X-clock (λ/2) is the reference and the SOS is the random clock used for the measurement of an X-coordinate. In a measurement the X-car speed is set to be as low as possible in the X region of interest (ROI) for a measurement. The spatial distance between two X-clock pulses is 316 nm. At the low-speed, about 7-8 SOS clock pulses appear inside the X-clock pulse interval. The travel distance between two SOS ticks, \( X_{inc} \), is therefore approximately \( (316/8) \approx 40 \) nm. In
other words the spatial resolution of the random SOS clock is approximately 40 nm. In the Y-direction the beam moves at 11.7 nm/ns and therefore one measurement clock period corresponds to \((25 \cdot 11.7) = 292\) nm.

In each microsweep, X-events and Y-events are recorded simultaneously. In figure 30 we illustrate a situation where seven microsweeps have passed over a pattern of two rectangles.

![Diagram](image)

**Figure 30:** The resulting events after one X-scan over the pattern. In microsweep 2-5 we record "yes" for X events. In scan 2 and 3 we receive four Y events (one for each edge) and in scan 4 and 5 we receive two Y events. In the rest of the scans no events are recorded. The size of the rectangles is not to scale.

An X-event is recorded by the fact that we have Y-events within a microsweep. These occasions are indicated as “yes” in figure 30. Inside a \(\lambda/2\) interval one scan in the X-direction generates X-events with the positioning uncertainty of \(\pm 20\) nm. The position of the Y-events is obtained with an uncertainty of \(\pm 146\) nm. As the SOS phase is random relative to the X-clock we can use the same theoretical treatment as in the one-dimensional case and express the uncertainty \(u_{X_{\text{ave}}}\) of an edge measurement as:

\[
u_{X_{\text{ave}}} = \frac{1}{\sqrt{n}} \cdot 0.5 \cdot X_{\text{inc}}\]

(26)

where \(n\) is the number of samples in the same \(X_{\text{inc}}\) interval of the edge, i.e. the number of X-scans.

With \(X_{\text{inc}} = 40\) nm we can express the maximum uncertainty caused by the measurement principle as:

\[
u_{X_{\text{ave}}} = \frac{1}{\sqrt{n}} \cdot 20 \ [\text{nm}]\]

(27)
As mentioned above we can receive approximately 7-8 possible X-events within an X-clock interval. The number of events depends on the actual speed of the X-car when it traverses the interval and the reflectance changes. To be able to calculate the spatial location of an event in a certain interval we need to know the local speed \(v_x\) to be used for the interval. The position of an X-event can be expressed as:

\[
X_{\text{pos}} = v_x \cdot X_{\text{event}} \ [nm]
\]  

(28)

where \(v_x\) is the interval length divided by the number of SOS-pulses.

To reduce measurement time of a mask pattern it is measured in both forward and backward X-direction. As the servo system is not designed to keep a constant speed we measure the actual speed of the X-car during the X-scan. In figure 31 the number of SOS pulses \(n_i\) per X-clock interval is shown from a measurement when the X-car has traversed a \(\sim 100 \mu m\) distance in one forward and one backward scan.

![Figure 31: The graph shows the obtained number of microsweeps \((n_i)\) (7,8 or 9) per X-clock interval after one forward and one backward scan over a distance in the X-direction of \(320 \cdot 316 \text{ nm} = 101.1 \mu m\). The local speed \(v_x\) in a certain \(\lambda/2\) interval is calculated based on \(n_i\). Note the difference in speed between the forward and the backward direction.](image)

The physical weight of the X-car is several kilograms so an abrupt change of speed between two successive X-clock intervals is not possible due to inertia. For this reason a low pass polynomial filter is used to filter the speed data. After this filtering, a very accurate local speed \(v_x\) can be estimated for any interval within the measured distance in the X-direction. Figure 32 shows the result of a measurement of a measurement mark as obtained using the two-dimensional random sampling technique.
Figure 32: A measurement mark recorded by random sampling in both X and Y. The obtained pattern represents the X,Y coordinates of a change in reflectance between the highly reflecting chromium and the low-reflective glass area. White and black pixels represent steps from low-to-high and high-to-low reflectance respectively. The measured area is approximately 100 x 200 µm.

The event caused by reflectance changes, at the edges of the mask pattern, appears as black or white dots, depending on the derivative of the signal. A great advantage of this technique is the limited number of “pixels” to be handled, compared to a camera image with a vast amount of pixels containing useless information.
5.3 Competing methods, Random phase and ADC imaging

During the development of the MMS15000, two edge detection techniques were developed in parallel. Besides the random phase technique described in the previous chapter, a more conventional method was also implemented. Already available in the system we had a pixel clock generator that is used in the writer to control the AOM during the exposure process (cf. figure 11). This hardware is also available in the MMS15000. The more traditional way of recording image information is simply to record the analog signal pixel-by-pixel using an analog to digital converter (ADC) and then generate gray level intensity images of the pattern on the photomasks. The pixel resolution using this method is 250 x 250 nm². For a microsweep length of 200 µm the image width is 800 pixels in the Y-direction. In the X-direction the number of pixel rows is limited by how much memory is available. A simplified view of the electronics used for the recording of the analog signal and the digitizer is shown in figure 33.

![Diagram of the alternative pixel imaging systems and the random sampling control.](image)

**Figure 33:** Principle of the alternative pixel imaging systems and the random sampling control. After amplification the analog signal enters the *Sample and Hold*, *ADC* and the control electronics used for the real time recording of the analog signal. Simultaneously the analog signal is digitized and delivered directly to the control logic block where it serves as input for the random phase electronics.

An advantage of using the analog signal for generating the image instead of adding a CCD or CMOS detector with additional optics is that the image is free from geometrical distortion and power gradients. This is because the microsweep generating the image is calibrated to have a maximum linearity error of less than 10 nm and maximum power error of less than 1%.
disadvantage of using this latter described technique is a much more complex data channel for handling huge data volumes, in contrast to just handling the events generated by edge transitions as is the case when the random phase technique is used. For critical dimension (CD) measurements the pixelated images are more practical to use since more complex image processing algorithms can be used on images containing gray level data. Another purpose for implementing the two different approaches is that it opens up possibilities for optimizing the two methods for different measurement cases. An example of an ADC recording of a line raster is shown in figure 34.

![Figure 34](image)

**Figure 34**: A recording of the reflected image of a chromium raster with a pitch of 9.0 μm. The dark areas are glass and the bright areas are chromium lines. In the region of interest (ROI) marked as a red rectangle all edges have been measured.

In this example, the sub-pixel locations of 42 edges have been calculated based on the ADC data. From this data the CD and pitch may be calculated (table 6). Besides the pitch, the non-linearity inside the ROI has also been calculated by subtracting the first order linear term from the accumulated pitch data.

<table>
<thead>
<tr>
<th></th>
<th>μm</th>
<th>(σ) nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD chrome lines</td>
<td>3.929</td>
<td>± 7</td>
</tr>
<tr>
<td>CD glass lines</td>
<td>5.071</td>
<td>± 7</td>
</tr>
<tr>
<td>Pitch</td>
<td>8.999</td>
<td>± 6</td>
</tr>
</tbody>
</table>

**Table 6**: CD, pitch and standard deviation of the data shown in figure 34. The data is based on 21 vertical lines covering 180 μm in the Y-direction and approximately 80 μm in the X-direction.
In figure 35 the non-linearity is shown in the case that the line raster has been calibrated to a level of single nanometers. This graph will be a measure of the non-linearity of the microsweep. A way to produce a calibrated raster is to write the lines in the X-direction in the writer and then measure them in the perpendicular direction. In such a case each location of an edge is traceable to the X-interferometer system.

![Non-linearity graph](image)

**Figure 35:** An example of how non-linearity of the microsweep can be verified by using a raster.

### 5.4 Inverse convolution

We will now make a minor step aside from the thread in this thesis and discuss an important method used in the MMS15000 for enhancing the resolution in CD measurements of thin lines.

All optical systems have a limited resolution due to the wavelength ($\lambda$) and the numerical aperture (NA) used. The effective NA used in the MMS15000 is approximately 0.5. The resolution is proportional to the inverse of the NA, and so a higher NA will enhance the resolution. The resolution $r$ can be expressed as:

$$r = \frac{\lambda}{2\cdot NA}$$  \hspace{1cm} (29)

where $r$ is the minimum distance between resolvable points.
There is however a price for a higher NA, in that the effective microsweep length will be shorter. The impact of the NA for the same deflection angle ($b$) is illustrated in figure 36.

**Figure 36:** For a given angle $b$ the efficient sweep length is inversely proportional to the $NA$ as illustrated in the figure.

A simple description of what happens when we are approaching the resolution limit is that the measured width at a constant threshold level of a line is reduced as the line gets narrower. This effect is illustrated in figure 37 where three different line widths and their corresponding signals are shown.

**Figure 37:** Due to the limited resolution the width of the line in the scanned image at a 50% threshold level will not be linearly proportional to the actual size of the line on the plate.
This effect is quantified as the CD linearity and causes small objects to appear with a reduced size when using the same threshold as used for larger objects. In the photomask industry the resolution limit is normally described as the smallest linewidth that fulfills the CD linearity specification. When using this definition, the resolution of the MMS15000 is approximately 1.0 µm for a linearity specification of 30 nm. This means that a true line width of 1.0 µm will give a measured width of 970 nm.

Instead of using a higher NA, that as a side effect gives both more complicated optics and reduced measurement time due to the shorter microsweep length, we instead developed a software algorithm that artificially increases the resolution. An advantage of using software to enhance the resolution is that this option is implemented only when objects narrower than 2 µm are to be measured. Thus, we will not sacrifice performance when measuring larger objects.

5.4.1 Enhancing optical resolution

We will now describe the filtering method, i.e. the Wiener de-convolution used in the MMS15000 for resolution enhancement [35]. The algorithm is based on a mathematical model of the optical system. This model is then used for designing an inverse filter directly applied to the intensity image before the conversion to the 2D image. It is mainly λ and NA (disregarding optical aberrations) that define the size of the point spread function (PSF) as shown in figure 38.

![PSF](image)

**Figure 38:** The PSF is the system response in the image plane to a point source in the object plane.

We now assume linear conditions, i.e. that the intensity in the object plane gives a linear response in the image plane and that the system is space invariant. That means that a translation in the object plane corresponds to a linear translation in the image plane. We can then describe the image with the convolution:
image(x, y) = \int \int Object(x - dx, y - dy) \cdot PSF(dx, dy) = Object \ast PSF \quad (30)

It is not obvious that we can use this simple model for the case we have in the MMS15000. This is because no spatial information is stored in an “image” on the detector as is the case described by the model (30).

To model the sweep measurement we rewrite (30) as:

\[
image(x, y) = \int \int Object(x + dx, y + dy) \cdot PSF(dx, dy) \quad (31)
\]

The light at “pixel” x,y is the sum of all reflected light when the PSF is centered above the object at position x,y. Thus this is not a normal convolution since \(dx,dy\) have changed sign. However this is a minor difference and can be resolved with the substitution:

\(dx' = -dx\) and \(dy' = -dy\)

Then we can rewrite (31) as:

\[
image(x, y) = \int \int Object(x - dx', y - dy') \cdot PSF(-dx',-dy') \quad (32)
\]

Now we assume that our PSF is circularly symmetric so for this reason we can rewrite (32) so it describes a normal convolution as:

\[
image(x, y) = \int \int Object(x - dx', y - dy') \cdot PSF(dx',dy') = Object \ast PSF \quad (33)
\]

### 5.4.2 Derivation of the filter function

The convolution in the space domain corresponds to a multiplication in the Fourier domain so (33) can therefore be expressed as:

\[
\mathcal{F}(image(x, y)) = \mathcal{F}(Object(x, y)) \cdot \mathcal{F}(PSF(x, y)) \quad (34)
\]

and

\[
\mathcal{F}(Object(x, y)) = \mathcal{F}(image(x, y)) \cdot \frac{1}{\mathcal{F}(PSF(x,y))} \quad (35)
\]

where \(\mathcal{F}\) is the Fourier operator and \(\mathcal{F}(PSF(x,y))\) is called the optical transfer function (OTF).

The Fourier transform of the PSF will have a negative slope. This means that values in the frequency plane will be close to zero for high frequencies. For this reason high frequency noise in the image due to the factor \(\frac{1}{\mathcal{F}(PSF(x,y))}\) in equation (35) will be amplified. In order to control the sensitivity to noise especially for higher frequencies we define the following Wiener filter:
\[ \text{Filter}(f_x, f_y) = \frac{1}{\text{OTF}(f_x, f_y)} \cdot \frac{|\text{OTF}(f_x, f_y)|}{|\text{OTF}(f_x, f_y)|+K(f_x, f_y)} \]  \tag{36}

where \( f_x, f_y \) is the coordinate in the frequency plane.

As seen in (36) we have now introduced a second term \( K(f_x, f_y) \) for the purpose of suppressing high frequency noise. If the term \( K(f_x, f_y)=0 \) for all frequencies, the filter is just the inverse of the system. When the filter is used we can rewrite equation (35) as:

\[ \mathcal{F}(\text{Object}(x, y) = \mathcal{F}(\text{Image}(x, y) \cdot \text{Filter}(f_x, f_y)) \]  \tag{37}

By using a test pattern with lines having known size, the PSF and \( K \) can be found by solving the optimization problem:

\[ \text{Min} \sum |\text{CD}_{\text{ref}}(q) - \text{CD}_{\text{measured}}(q, \text{PSF}(x, y), K(x, y))| \]  \tag{38}

where \( CD \) is the line width and \( q \) is a certain line in the pattern.

The strategy used in the MMS15000 is first to optimize the PSF, which is done by simulating several different PSF functions and then convolving these functions with lines having known widths. The result of these \( q \) number of convolutions is saved in a table \( CD_{\text{ref}} \). By comparing the measured CD with the data in \( CD_{\text{ref}} \), the optimum PSF can then be found by interpolation in the table. After this step the optimal \( K \) can be found by selecting a circularly symmetric function with a certain radius in the frequency domain as shown in figure 39. The optimization criterion used is to choose a radius so that when the filter is applied the measured lines fulfill the CD linearity specification.

**Figure 39:** Illustration of how to select the optimum \( K \) function in the frequency domain. Different radii will give different resolution (i.e. CD linearity). The frequency response for narrow lines and also amplification of noise will be affected by the chosen radius of the \( K \) term. A too strong Wiener filter function will give large CD errors due to low amplification of the total filter for the targeting line widths, and a too weak Wiener filtering will amplify noise and make the CD error large.
5.4.3 Results

After estimation of the PSF and $K$ verification measurements are done on lines with calibrated widths, (see figure 40). A pattern is used that consists of lines with widths of 0.5µm-2.0 µm in steps of 0.1 µm.

![Figure 40](image-url)

**Figure 40:** Measurement signal obtained for a set of chromium lines having widths of 0.5, 0.6, 0.7 …2.0 µm from left to right. The left graph shows the resolution without filtering and the right graph shows the result with filtering applied.

As seen in figure 40 the algorithm introduces some noise. The majority of this noise occurs at low reflectance levels (glass) and high reflectance levels (chromium). At the 50% threshold level this noise will have a minor impact on the estimation of the edge location. Noise will also be suppressed because of the significant number of pixels that are averaged in an estimation of a line edge. In figure 41 the edges of the transitions are shown in the spatial domain.

![Figure 41](image-url)

**Figure 41:** Measured 0.5 µm – 2.0 µm chrome line edges in steps of 0.1 µm using a 50% threshold. The 0.5µm lines are shown as the short lines at the left in the images. The picture at the left shows the unfiltered image and the right one shows the result when the filter has been applied.
6 Self-calibration

In the manufacturing process of large area photomasks one would expect that traceability of geometrical locations of features on the mask would be exceedingly important, and a global Cartesian large area standard to be used as traceable reference. However there are reasons for why this is not the case. As has been mentioned before, mask makers use their own reference standards in order to make it more difficult for competitors to make masks that fit exactly to their geometrical standard. The reason for using the same reference standard with the ability for the end users to mix photomask manufactured by different vendors is therefore of little interest to the mask maker. Up to 11 masks are needed in the process of making a display. Besides the registration requirements, the overlay, i.e. how well each mask fits together with the other masks in a set is also very important. The absolute registration requirement of the pattern on the photomask is today in the range of 150-200 nm (3σ) over an area of square meters. It is obvious that calibration and verification at this ultra-precision level require an extremely stable and accurate coordinate measurement machine (CMM).

The MMS15000 [2] is today the world standard CMM for measurements of photomasks with sizes up to 1.4 x 1.5 m². Besides local measurements of CD of features, the MMS15000 also is capable of absolute registration measurements with an uncertainty better than 90 nm (3σ). To achieve this accuracy, and because traceable reference standards do not exist with this huge size, a novel and more general self-calibration algorithm (see paper D) had to be developed for the MMS15000. Our method differs in some aspects from methods developed by Michael Raugh and others in the 1990s [36][37][38]. Raugh put a lot of effort into deriving the math behind self-calibration using a square lattice. The fundamental principle of self-calibration is to solve an inverse problem and it is the same regardless of what kind of algorithm is used. It is in the details, generality and robustness against error sources that the differences appear between our new method and the old ones.

6.1 Basic principles

By using a fairly accurate measurement tool and a reference plate, referred to as the golden plate (GP), filled with measurement marks not necessarily with known X,Y locations it is possible to achieve an absolute fit to a perfect
Cartesian system of the measurement tool stage and the interferometrically determined location of the measurement marks on the GP down to a precision only limited by the noise level of the measurement tool. To minimize the noise level we use our specially developed measurement technique, described in the previous chapter, which works on the principle of the random sampling of pattern edges by scanning the laser beam in the MMS15000. The measurement repeatability using this technique is around 10 nm (3σ) when measuring a matrix of measurement marks on an artifact of 0.8 x 0.8 m² [Paper B].

The self-calibration method we have developed, to be briefly described here and in detail in paper D, solves the inverse problem with fewer constraints than previously established methods used in small area CMMs in the semiconductor industry. The algorithm we have implemented in the MMS15000 is therefore more general in some sense. We do not require a square grid of measurement marks on the plate or the GP. Display masks are rectangular in shape, so to use the surface of the plate as efficiently as possible the matrix of measurement marks will typically not be a square matrix. Also, the pitch of the measurement marks does not need to be constant. The final MMS15000 stage grid used for the correction (i.e. mapping of the stage deviation from a Cartesian grid) does not depend on the grid of measurement marks on the GP used in the calibration. The GP used is typically made of quartz glass with a size of approximately one meter square and with a thickness of 5-16 mm. Precise control of alignment and rotation of such a large and heavy plate is very difficult. The requirement needed by the traditional methods, that different views (i.e. placements of different rotations and translations of the plate on the stage) must be very precise in alignment, can therefore not be fulfilled. Precise alignment is not required in our case. We can also use different placement schemes for different plates and stage sizes and use weighting of data from certain areas on the stage to obtain tighter tolerances.

A GP of this size cannot be treated as a rigid body, at least not down to the accuracy we are interested in. Therefore the algorithm must be robust not only against ordinary measurement noise but also for the “shape noise” generated by the GP itself when it is placed at different locations on the stage. Several views are used in order to minimize the errors caused by the slight deformation of the plate. Another reason for using more views is that the size of the plate used for the calibration is seldom large enough to cover the entire measurement machine stage.
6.2 The algorithm

The goal of a self-calibration is to find the absolute geometrical shape i.e. deviations from a perfect Cartesian grid of the GP and the measurement machine stage. All deviations can be found in a self-calibration procedure except the absolute scale. To find the absolute scale we need a traceable calibration standard [25], preferably with less uncertainty than what the MMS15000 laser interferometers provide. That has so far not been possible to obtain.

The input data used in a self-calibration are several measurements of the same measurement marks on the GP when it has been placed in different views, i.e. different positions and rotations on the CMM stage. At least three different views of the same plate are needed at a minimum. Normally more views are used for redundancy reasons and to achieve good stability in the calibration along the edges of the plate and the edges of the covered area of the stage. In figure 15 we showed typical errors of a 2D measurement tool that can be assessed and compensated for by a self-calibration process. These include first, second and higher order deviations from a Cartesian grid. An example of a typical first order deviation is the orthogonality, defined as the angle deviation between the X-axis and Y-axis from ninety degrees. Another example is the stage bow, defined as the deviation from a straight line of the X and Y axes respectively.

Normally, in a mask writer, orthogonality, bow and all other shape errors can be determined with a large area GP that has been verified in a self-calibration procedure. But to get such a GP with registration tolerances in the sub-100 nm range we have to verify it in the MMS15000 and simultaneously verify the shape deviations of the machine stage itself. The initial boundary condition for the self-calibration is thus that we put a plate with measurement marks onto the machine stage assuming that both have unknown shape deviations. In figure 42 we illustrate this by a strongly exaggerated shape-distorted GP (plate) with its measurement marks. It is placed on a shape distorted measurement machine stage with its inherent gridpoints.
Figure 42: A strongly exaggerated representation of a plate with measurement marks (+) separated by a pitch of $pp_x$ and $pp_y$ put on a deformed measurement machine stage with a virtual grid having a pitch of $sp_x$ and $sp_y$. The dashed lines refer to the perfect Cartesian grid of the stage. The goal of the self-calibration procedure is to obtain the corrections needed to compensate for the deviations in each stage gridpoint to straighten out the deformed blue solid line grid into the dashed grid.

On the plate we use a set of the $n_x \cdot n_y$ measurement marks distributed in a matrix. As discussed above, the “exact” locations of these marks are unknown but they can still be used to make a two-dimensional sampling of the MMS15000 stage deformation. Note that we will only have a representation of the shape of the plate in each discrete point $i_x,j_y$ of the matrix on the plate. The shape of the measurement machine stage may be found in every accessible X and Y point. This is because the sampling of the stage can be done at any arbitrary point and not necessarily only in the matrix grid $i_{sx},j_{sy}$ representing the shape of the stage after the calibration. It is more a question of what spatial frequencies of errors we are interested in finding. In other words, the method we have developed is not limited to finding deviations on the stage with spatial frequencies lower than the Nyquist frequency set by the pitch of the marks at the plate [39].

In the self-calibration process the placement scheme for the different measurements of the plate on the stage is important. As mentioned before, three different placements of the plate are enough for a calibration provided
that these placements are not redundant. This means that new spatial information of the plate and stage must be retrieved in each measurement. By representing each measurement of the plate as a set of equations, a large number of placements will generate a huge set of equations. For three placements this system of equations will be over-determined. The solution is typically obtained by using a least squares fitting model and so it has been discussed by other authors dealing with self-calibration [40]. But the interesting thing is that the system can actually be degenerated, i.e. the placements do not necessarily add information to generate a unique solution. To avoid this we have developed a dedicated software package to optimize the placement scheme for a certain stage size in relation to the plate size. The details of that are out of the scope of this thesis.

6.2.1 Definitions used in the self-calibration

To describe the self-calibration process we need to introduce the following definitions.

**Grid:** A matrix of points in the X,Y plane separated by well-defined distances, the so-called pitch. We use two different grids, each with its own pitch, in the self-calibration of the MMS15000. One is the grid of physical measurement marks on the plate to be calibrated and the other is the correction grid used for correction of the MMS15000 measurement stage after the calibration.

\( i_x, j_y: \) are integers indexing points in the X and Y direction respectively when addressing plate points (marks) distributed over the plate \( (P) \) and points in a measurement matrix \( (M) \).

\( (i_x = 0, 1, 2 \ldots n_x-1) \) and \( (j_y = 0, 1, 2 \ldots n_y-1) \) where \( n_x, n_y \) is the number of measurement marks on the plate in the X and Y direction respectively. A nominal absolute position on the plate or stage is calculated by multiplying the index with the pitch in the corresponding direction respectively.

\( is_x, js_y: \) is an integer index in the X and Y direction respectively used when addressing points in the matrix describing the stage \( (S) \).

\( (is_x = 0, 1, 2 \ldots m_x-1) \) and \( (js_y = 0, 1, 2 \ldots m_y-1) \) where \( m_x, m_y \) are the number of gridpoints in X and Y directions used on the stage.

**Plate matrix \( (P) \):** Refers to the plate with chromium measurement marks. On the nm-scale it is not perfect and the “shape” of the plate is described as a matrix of vectors, \( P(i_x,j_y) \), with their origin at the center of the plate.
Stage matrix (S): In analogy with the Plate matrix the 2D “shape” of the measurement machine stage is described as a matrix of vectors, \( S(i_x,j_y) \), relative to its origin at the lower left corner.

Measurement matrix (M): Refers to a measurement of the plate measurement marks at a certain stage placement and rotation of the plate in the X, Y plane. A measurement of the measurement mark \( i_x,j_y \) on the plate is expressed as \( M_k(i_x,j_y) \), and represents the vector coordinates relative to the center of the stage. Here, \( k \) is an index representing a specific rotation/placement of the plate on the stage.

Simulated measurement (SM): In the algorithm, simulated measurements are calculated based on the temporary representation of \( S, P \) and the placement and rotation, \( k \), of each measurement \( M \). A simulated measurement in a certain location of the plate is expressed as a vector \( SM(i_x,j_y) \) with its two coordinates.

Placement (plk): Is a vector of two components \( x_k,y_k \) containing the absolute coordinates of the averaged center point of all measured marks \( M_k \) in a measurement view, \( k \), on the stage.

Placement angle \( \alpha(k) \): is obtained by calculating the angle from four corner marks of the measurement \( M_k \) when it has been back-rotated \((0, \pi/2, \pi, 3\pi/2)\). This angle is referred to the X-axis.

6.2.2 The self-calibration procedure

We will now introduce the principle of the algorithm. We start with an unknown plate \( P \) and an unknown stage \( S \). The goal of the exercise is to find the absolute shape of \( P \) and \( S \). When \( S \) is known we can correct the stage movements so it will describe a perfect Cartesian system, limited only by the uncertainty caused by random noise. \( S \) may contain information about scale differences in the X and Y directions and orthogonality. A rotational adjustment of the stage axes is done to make the stage X-axis aligned with the Cartesian X-axis. Then, the stage Y-axis alone carries the information about the orthogonality error [33]. In the algorithm we use two different coordinate systems, one for \( P \) with an origin in the center of the plate and one for \( S \) with its origin in the lower left corner of \( S \) (see figure 42). By rotation and placement of the plate on the stage at \( k \) different views, we get an over-determined measurement problem. This gives us more data points than unknown parameters and while running the self-calibration algorithm as an iteration loop, see figure 43, we make use of a figure of merit.
similar to the least squares criteria to determine when we cannot gain any more information out of the measured data points. We can express the principles of the algorithm as follows.

Let us assume that we have a flat rigid object, in our case a plate with measurement marks, and we put this object on a metrology tool stage for measurement. If this object yields the same position of the measurement marks in the plate coordinate system, independent of where it is placed on the stage, we can say that the stage coordinate system of the tool is consistent. We can also say that if the object does not look the same at the different stage positions, the tool coordinate system is not consistent. The different appearance of the object in different views is therefore a measure of how much the tool coordinate system differs from the perfect Cartesian system in a certain placement of the object. By using this difference in order to compensate the stage motions we have achieved a calibration of the tool stage, but with the important difference that we do not necessarily know the absolute positions of the measurement marks, i.e. the “shape” of the plate. Therefore we approximate the measurement mark pattern on the plate by using the average shape of this pattern from the different views as the best estimation of the plate shape. By virtually replacing this new object on the tool stage in the same locations and rotations as in the first measurements we can update the tool stage coordinate system at these locations. If the plate shape in a second comparison is the same in all placements we are through with the calibration; if not, the procedure is repeated.

In the self-calibration process we introduce a temporary stage correction function \( SCF \) which is a 2D function, describing the correction of the stage \( S \) for the current step in the iteration loop. This function is represented as a matrix with the pitch \( sp_x \) and \( sp_y \) and has the same size as \( S \). This matrix is set to zero at the initiation of the iteration process. \( M, P \) and \( SM \) are all matrices defined earlier and represent different states in the iteration. \( C \) is the merit matrix with the same size as \( S \) and the standard deviation of its elements gives the merit number of the calibration.
Figure 43: The principle of the calibration loop. The above notations are only meant for describing the different steps as principal equations. The R+ and R- refer to forward and backward rotation matrixes of the plate on the stage. The algorithm is described in detail in Paper D.

6.2.3 Ambiguities

To avoid degeneration in the calibration process the placement scheme is important. An example of a degenerated problem is illustrated in figure 44. Here we will try to extract the plate and stage shape using two measurements, 0 and 180 degrees. We imagine two cases. Case 1 with a perfect plate and a deformed stage and case 2 with a rotationally symmetric error in the plate and a deformed stage.
Figure 44: A calibration using two placements, 0 and 180 degrees without any translation in-between placements. The real plate and the stage is presented at the left and the two measurements at the right in the stage coordinate system.

As seen in this simple example we will generate two identical measurements for the two cases. In other words there is no information of the different stages and plates based on the two measurements. The system is therefore degenerated and has an infinite number of solutions. In a calibration we need to know if the solution for the stage and the plate is unique for the involved measurements. We simplify the problem in another example using a square matrix, by placing the GP exactly on the stage grid and making a self-calibration, by using three measurements as shown in figure 45.
Figure 45: A self-calibration using three measurements. The unknown gridpoints of the stage are marked with red dots, and the unknown gridpoints of the plate are marked with crosses.

We make one measurement \( m_1 \) with the rotation 0 degrees and one, \( m_2 \), rotated ninety degrees but placed on the same stage gridpoints. The third measurement, \( m_3 \), is translated to the left on the stage grid with the rotation 0 degrees. We now have \( n_x \cdot n_y \cdot 2 \) unknown parameters of the plate and \( (n_x+1) \cdot n_y \cdot 2 \) unknown parameters in the stage. In three measurements we will have \( n_x \cdot n_y \cdot 2 \cdot 3 \) known measured values. So in total we have \( n_x \cdot n_y \cdot 4 + n_y \cdot 2 \) unknown parameters and \( n_x \cdot n_y \cdot 6 \) known measurements. This overdetermined equation system can be solved using a least square model. But before doing this we need to know if it is possible to find a unique solution at all. As showed in figure 44 the use of a square plate and two rotations does not reveal a unique solution even if we have an equal number of equations as measurements. A way to check if this equation system can be solved is to investigate if it has a sufficient number of independent equations. In our case we can write the equation system to be solved in a matrix form as:

\[
A \cdot ps = B
\]

Where \( A \) is a matrix containing the combinations of the unknown parameters, \( ps \) and \( B \) are vectors containing all the measurements. The rank of a matrix is a calculation of the number of independent equations in \( A \) after that \( A \) has been reduced to its most compact form. For the equation system to have a solution the rank must be equal to or larger than the number of known parameters. If this is the case the solution can be expressed as:

\[
ps = \text{inv}(A) \cdot B
\]

In the case when rectangular GPs together with rectangular areas of the stage are to be calibrated the problem of finding the optimum placement scheme is
quite complicated. For this reason we have developed a special software tool. The input parameters to this software are the pitch \(pp_x\), \(pp_y\), number of measurement marks, \(n_x\), \(n_y\) of the GP and the stage pitch \(sp_x\), \(sp_y\), and number of gridpoints in respective directions on the stage. The output is a suggested placement scheme requiring a minimum number of measurements and a figure of merit as the rank for that placement scheme.

An important difference to solving the equation system as described above and the method we use is that even in a situation when the equation system is slightly underdetermined the solution we achieve for the shape of \(S\) and \(P\) does not differ as much to the correct (and unique) solution as would be the case when inverting the matrix. During the development of the MMS15000 we implemented both methods and found that our method is more robust in such a situation. The reason for this is that we use an initial guess of \(S\), and the method we use when updating the stage correction function in the iteration loop. The latter is a linear operation which does not propagate errors as efficiently as when using a least square model.

The authors of the book *Numerical Recipes in C* have expressed the problem of overdetermined systems as:

“There is a certain mathematical irony in the fact that least-squares problems are both overdetermined (number of data points greater than number of parameters) and underdetermined (ambiguous combinations of parameters exist); but that is how it frequently is. The ambiguities can be extremely hard to notice a priori in complicated problems.” [41].

### 6.3 Self-calibration results

#### 6.3.1 Simulation

The convergence and robustness against noise have been tested using simulations. In this first simulation we demonstrate a simple case of using four placements of a plate with 16 x 17 measurement marks with a pitch of 50 mm in the X and Y directions respectively. The shape of this plate is shown in figure 46a. For the estimation of the two-dimensional stage function we use a grid of 30 x 30 points with a pitch of 40 mm in the X and Y direction respectively. The simulated stage is presented in figure 46b.
Figure 46: The 0.8 x 0.85 m$^2$ plate grid (a) and the 1.2 x 1.2 m$^2$ large stage grid (b) used in the simulation.

The first plate placement (the upper left corner) in this simulation is the absolute position (100,150) mm in the X and Y direction respectively. In table 7 the translation and rotation for each placement relative the first plate position is shown.

<table>
<thead>
<tr>
<th>Placement</th>
<th>Rotation (radians)</th>
<th>Misplacement angle (milliradians)</th>
<th>Translation X (mm)</th>
<th>Translation Y (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>-5</td>
</tr>
<tr>
<td>2</td>
<td>$\pi$</td>
<td>-1.5</td>
<td>10</td>
<td>62</td>
</tr>
<tr>
<td>3</td>
<td>$\pi/2$</td>
<td>-2</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>$3\pi/2$</td>
<td>-1</td>
<td>45</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 7: The different rotations and translations used in the simulation.

Please note that $\alpha$ is the sum of the rotation and the misplacement angle. This is an example where all translations and misplacement angle are far from any gridpoints on the stage.

The different absolute placements and a part of the simulated stage are presented in figure 47.
Figure 47: The four different simulated placements of the 50 mm plate grid shown in red, green, pink and orange on top of the blue 40 mm stage pitch grid. The local scale around a gridpoint is in this case 3 mm. The background grid (dashed in the figure) has a pitch of 50 mm in both X and Y. The covered area of the stage is 0.8 x 0.8 m².

This first simulation will demonstrate the convergence using the placement scheme as shown in table 7. We stop the simulation after 10000 iterations for verification of the results. A comparison of the simulated plate with the output plate is a measure of the precision of the calibration. Another measure is a comparison of the simulated stage and the output stage of the algorithm. In the latter comparison it is only relevant to compare the output stage in the area covered by the placements of the plates. These two comparisons have been done in figure 48a and 48b respectively.
Figure 48: a) The difference between the output plate and the simulated plate. The $3\sigma$ for the difference is 3.7 nm and 2.6 nm in the X and Y direction respectively.

b) The difference between the output stage and the simulated stage in the covered area of the placements. The $3\sigma$ for the difference is 4.5 nm and 3.3 nm in the X and Y direction respectively. The scale in the graphs is 20 nm/div.

In the calibrated area it is very probable that we have a small absolute scale error of the stage. Also the plate may have an absolute scale error. The algorithm will automatically move all scale errors in to the matrix $P$ that is a description of the plate assuming we have no absolute scale errors in the stage. Since absolute scale needs to be calibrated using a reference, the effect from the scale has been removed in the graphs in figure 48. As can be seen, the deviations along the edges are worse compared to in the middle. The reason for this difference is the limited coverage of gridpoints at the edges, i.e. three measurements compared to four different measurements in the middle.

Robustness against noise

In the next simulation we add some random noise to the measurement. The $3\sigma$ variation of this noise is 18 nm in the X and Y directions respectively. In this test we will avoid interpolation errors and therefore we use the same pitch, 50 mm, of the plate and stage grids in the X direction and Y direction respectively. Table 8 shows the placement scheme.
<table>
<thead>
<tr>
<th>Placement</th>
<th>Rotation (radians)</th>
<th>Misplacement angle (milliradians)</th>
<th>Translation X (mm)</th>
<th>Translation Y (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>$\pi$</td>
<td>0</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>$\pi/2$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>$3\pi/2$</td>
<td>0</td>
<td>50</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 8: The different rotations and translations used in the simulation with noise added. In order to make the test relevant we have avoided any interpolation errors by placing everything on grid.

The result of this test is presented in figure 49.

**Figure 49:**

a) The difference between the output plate and the simulated plate when random measurement noise with the $3\sigma$ of 18 nm has been added. The $3\sigma$ for the difference is 18 nm and 17 nm in the X and Y direction respectively.

b) The difference between the output stage and the simulated stage in the covered area of the placements when random measurement noise has been added. The $3\sigma$ for the difference is 16 nm and 15 nm in the X and Y directions respectively. The scale in the graphs is 50 nm/div.
This is an ideal case never occurring in practice. But it is relevant for demonstrating the robustness against noise. As can be noted the simulated result of the noise of the plate and the stage is in the same range as the input noise.

6.3.2 Real measurement data

As a last demonstration of the performance of the algorithm we use real input data. In this example we use the following placement scheme.

<table>
<thead>
<tr>
<th>Placement</th>
<th>Rotation (radians)</th>
<th>Translation X (mm)</th>
<th>Translation Y (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>( \pi )</td>
<td>0</td>
<td>-42</td>
</tr>
<tr>
<td>3</td>
<td>( \pi/2 )</td>
<td>42</td>
<td>-42</td>
</tr>
<tr>
<td>4</td>
<td>3( \pi/2 )</td>
<td>0</td>
<td>-42</td>
</tr>
</tbody>
</table>

Table 9: The four different placements of a reference plate. The translation in X and Y are nominal placement locations.

Since we do not know the exact shape of the GP used, we need another measure of the accuracy of the calibration. So instead of comparing the output plate with the input simulated plate we instead measure the overlay between four different plates generated by the algorithm. Each of these plates corresponds to each of the input measurements. We can express a measurement using the principal equation:

\[
M = P - S
\]  

(41)

where \( P \) is placed in \( S \) as shown in table 9. In each placement we therefore can calculate a plate as:

\[
P(i) = M(i) + S
\]  

(42)

where \( i \) is the index of the placement. If we then back rotate each \( P(i) \) to zero degrees (except the first one when \( i=1 \)) we then have four different versions of the plate corresponding to each measurement \( M(i) \).

The output plate \( P_{\text{ave}} \) is the average of these plates that may be expressed as:
\[ P_{\text{ave}} = \frac{1}{4} \sum_{i=1}^{4} P(i) \]  

(43)

In figure 50 the average plate and the overlay of the four plates is presented. In an overlay plot each plate is compared with the average plate \( P_{\text{ave}} \).

**Figure 50**: The average plate \( P_{\text{ave}} \) and the overlay between the different placements of the plate.

In table 10 the three sigma of the overlay in the X and Y direction is presented.

<table>
<thead>
<tr>
<th>Placement</th>
<th>X 3(\sigma) (nm)</th>
<th>Y 3(\sigma) (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>132</td>
<td>98</td>
</tr>
<tr>
<td>2</td>
<td>149</td>
<td>76</td>
</tr>
<tr>
<td>3</td>
<td>74</td>
<td>84</td>
</tr>
<tr>
<td>4</td>
<td>94</td>
<td>95</td>
</tr>
</tbody>
</table>

**Table 10**: The deviation of each plate compared to the average plate expressed as 3\(\sigma\).

As seen in the table the deviations of the different plates compared to \( P_{\text{ave}} \) are below 150 nm (3\(\sigma\)). These are typical numbers in a self-calibration of the...
MMS15000 over an area of approximately 0.8 x 0.8 m². The registration specification of the MMS15000 is 90 nm (3σ) for a verification measurement of the placements 0, π/2, π and 3π/2 inside the calibrated area. It is not possible to fulfill this specification without correcting for the fact that the plate is not rigid during the calibration. A special function called Z-correction is used for this purpose that will be further discussed later in this thesis. By several simulations and also using real data, we have verified that the loop converges but we cannot prove convergence in the mathematical sense. However, it is possible to prove it mathematically if we put similar constraints on the placements of the plate, pitches of measurement marks and pitches used on the stage in the calibration etc. This has already been done by Raugh [36].

It is possible to calibrate the whole stage using a much smaller plate either by first finding the shape of the plate (in a smaller area of the stage) and then using it as a reference for a calibration of the larger area, or by designing a placement scheme covering the whole stage and in this way finding the shape of the stage and plate simultaneously. A full self-calibration plus verification of a large area stage (1.4 x 1.5 m²) takes several days in time including tens of measurements. In 2005 we received a patent on our method [42].
We will now discuss the function called Z-correction that is used for correcting measurements of a photomask in the MMS15000 that has been distorted due to gravity when it is resting on the stage of the tool. The same function is also used for correction of distortions of the glass when the pattern is exposed on a quartz plate in the writer. Without using this correction it is practically impossible to achieve an uncertainty better than 150-200 nm (3σ) over an area of 0.8 x 0.8 m². When Z-correction is used it is possible to enhance this number to considerably less than 100 nm (3σ) [paper E].

As has been mentioned before, several photomasks are used in display manufacturing for defining the pattern of e.g. transistors on the array substrate. These masks are used for defining the gate and source-drain pattern of the transistor. It is therefore an absolute necessity that the different layers fit together for the transistor to work properly. In the production of the transistor arrays on the substrate an aligner is used (cf figure 10) for positioning and transferring the image of the mask onto the photoresist coated substrate. The physical support of the photomask in the aligner is done by supporting rods as shown in figure 51.

![Figure 51](image)

**Figure 51**: The typical way of how the photomask is supported in an aligner. The bending of the photomask is highly exaggerated.

In the exposure stage the mask pattern resides on the bottom surface of the photomask. Due to gravity, the plate will bend significantly and deform the pattern, and this can be compensated for. However, it is not easy to compensate for the defocus due to local variations in the Z-direction of the mask image along the focal surface. For this reason the patterned surface of the mask has a tighter tolerance regarding flatness than the other surface of the plate [43].
In the MMS15000 the photomask is placed on a flat Zerodur stage prior to the measurement. Distortions will occur for several reasons. Firstly the photomask is not flat on its back side that is resting at the stage. Secondly the stage supporting the mask is not perfectly flat and finally, because contaminating particles and/or air bubbles might be stuck under the plate, it will be deformed when it is exposed or measured on the stage. As will be shown later, particles or air bubbles stuck under the photomask may introduce registration errors of several hundred nanometers in the X,Y directions. A quartz glass with a thickness of 16 mm and a lateral size of square meters cannot be treated as a rigid body in a self-calibration process aiming for sub-100 nm levels in the metrology tool. Correction for gravitation-induced distortions of the plate is therefore important for a correct calibration of the tool stage and for the calibration of the plate.

In the calibration process of the MMS15000, a GP is used as reference. The crosses on the GP are referred to as measurement marks distributed with a certain pitch across the surface. After a successful calibration, the location of the marks are known with certain accuracy in the Cartesian coordinate system of the plate. When this information is known, as mentioned before, the plate is qualified to be a Golden Plate i.e. a GP. Such a plate may then be used as a reference for other metrology tools or writers.

When the Z-correction function is used, we make a calibration of the stage and the plate in a flat plane at the distance T (that corresponds to the thickness of the plate) above the surface of the stage as shown in figure 52.

In a measurement the X and Y coordinate and the Z coordinate of the mark are measured. The Z coordinate is used for compensating the measured X,Y coordinate for the deviation caused by the gravity-induced distortion of the plate when it is resting on the stage. After this compensation the measured X,Y deviations correspond to the deviations in the flat reference plane.
Figure 52: The stage coordinate system and the reference plane used by the Z-correction function. The plate (yellow) is filled by chromium crosses in a matrix. In a measurement the $X, Y, Z$ coordinate is recorded. The origin $(0,0,0)$ coordinate is defined to be at the lower left corner of the stage.

The reference in Z direction is the so called X-bridge. In figure 53 the parts involved in a Z-measurement are shown.

Figure 53: Cross-section of the $xBridge$, $xCar$, plate and the Stage involved in a Z-measurement. A special device measures the distance $mz$ between the $xCar$ and the final lens. A focus servo is keeping the distance from the final lens and the plate constant. The $xCar$ is traveling on air-bearings on the $xBridge$ in the X-direction. The $xBridge$ is floating on air-bearings on the Stage. The Stage is moved in the Y-direction relative the $xBridge$.

The physical deviation from a straight line of the xBridge is automatically taken care of in the self-calibration procedure that will be discussed later in this thesis. The accuracy of the Z-measurements is not at all as critical for mask manufacturing as the X,Y coordinates of the patterns. The reason for this will be further explained below.
7.1 Mechanical deformation model

The measured Z coordinate will vary depending on thickness variations of the plate, Z variations of the stage and possible particles and/or air bubbles under the plate when it is resting on the stage. A very simple mechanical model, illustrated in figure 54, is used for calculating the spatial deviation of a location of a measurement mark on the plate due to the gravity-induced plate bending.

![Plate bending effect](image)

**Figure 54:** A strongly exaggerated illustration in the X,Z plane of the effects caused by a particle residing under the plate. The black dots at the reference plane show the projection of the measurement marks in the case of no particle present under the plate. The open circles at the top of the plate show the location of the measurement mark in the deformed situation. By projecting these onto the reference plane, the \( dx \) deviation is obtained. The tensile stress of the glass caused by the particle will thus enlarge the pitch on the convex top-surface. For the same reason the distance between points on the concave bottom-surface will shrink. On the virtual centerline (dashed in the figure) we assume that the pitch is unaffected. For details see paper E. A measurement of the cross mark \( i_x, j_y \) can be expressed as:

\[
M(i_x, j_y) = P_{abs}(i_x, j_y) - S(i_x, j_y) + Z_{corr}(i_x, j_y)
\] (44)
$P_{abs}(i_xj_y)$ is the absolute vector of a measurement mark on the plate in the Cartesian coordinate system relative the origin of the stage and can be expressed as:

$$P_{abs}(i_xj_y) = P(i_xj_y) + P_{origin}$$

(45)

Where $P_{origin}$ is a vector from the stage origin to the center of the plate.

$S(i_sxs_y)$ represents the vector coordinates in the Cartesian coordinate system, relative the origin of the stage, of the nominal coordinates of the measurement mark that is displaced to $P_{abs} (i_xj_y)$ by the Z-deviation of the plate.

$Z_{corr}(i_xj_y)$ is a correction vector that is representing two components

$$Z_{corr}(i_xj_y) = [dx(i_xj_y),dy(i_xj_y)]$$

In practice the distance difference between $P_{abs} (i_xj_y)$ and $S(i_sxs_y)$ is of the order of 100 nm. Equation (44) is illustrated graphically in figure 55.

**Figure 55:** In the gridpoint $i_sxs_y$ on the stage (the black dot in the figure), chosen as the point closest to the nominal measurement mark position, a measurement $M(i_xj_y)$ is made. With Z-correction the result will be the open red circle (the most correct position) and without Z-correction it will be the filled red dot.
7.2 Different cases

As will be seen, it is not obvious that the correction of a measurement is valid for all shape Z-deviations of the glass plate. To demonstrate that we present three cases where we simulate different kinds of errors in the plate and stage and demonstrate the correction.

Case 1: The first case has already been presented in figure 54 and is redrawn in figure 56. It assumes a perfectly flat plate put on a particle on a flat stage.

Figure 56: Strongly exaggerated illustration of how a particle will elastically bend the heavy plate by gravity, and the corrections needed for a displaced measurement mark in the X-direction, caused by the deviation of the surface in Z-direction.

In the figure the X-axis has been moved to the reference plane for better clarity. Because of the bending of the plate, the measurement marks will move outwards (gray dots in the figure). The corrections (the black arrows) will bring them back to their original locations along the X-axis of the reference plane (black dots). The particle could just as well be an air bubble under the plate or a bump on the stage. The bending will of course also displace the measurement marks in the Y-direction.

Case 2: In the second case, shown in figure 57, a plate with varying thickness is treated as a bump on the bottom surface put on a flat stage.
Figure 57: A plate with a free form (varying thickness) is placed on a flat stage. By “free form” we mean the shape of the plate when not subjected to any external forces. Gravity will bend the plate (depending of the elastic module) and will deform its top-surface.

The deviations in figure 57 are exaggerated for better clarity. The corrections (the black arrows) will bring the measurement marks back to their original locations in the reference plane (black dots). A bump will also move the marks in the Y-direction, but for clarity only the movement in the X-direction is show in the figure.

Case 3: In the third case, shown in figure 58, a plate with a convex shaped top-surface and a flat bottom-surface is put on a flat stage.

Figure 58: A flat plate bottom on a flat stage will not cause any bending. Thus, no measurement marks will physically move on the top-surface when the plate is resting at the stage. The “correction” will in this case introduce an error. The distances in the figure are exaggerated for better clarity.
The Z-correction function cannot distinguish between the cases shown in figure 57 and figure 58. The same gradient directions will be measured by the system at the top surface. For this reason the correction in case 3 will erroneously “move” the marks as indicated in the figure.

Case 3 shows that the Z-correction can introduce errors rather than correcting them. In practice this could be a serious problem, but normally high quality quartz glass blanks used by the photomask industry have a flatter top-surface than bottom-surface. The reason for this is that the mask pattern is written into the photoresist coated top-surface. When the mask is used in the exposure equipment it is very important that this surface is deformed according to the prescribed form for pure optical reasons (see figure 51). Without going into much detail on this subject we just say that the image surface must stay inside the focal depth of the optics used in the exposure equipment.

7.3 Front surface corrections

The high quality quartz blanks are usually supplied with a measurement record of the shape of the top-surface and the bottom-surface. These height deviation maps are available from the glass supplier [45]. By using data from the top-surface of the blank it is possible to compensate for the gradient calculations and make use of the real top-surface shape instead of using the flat reference plane. By using the top-surface data it is thus possible to separate case 2 from case 3 that otherwise would not be possible by just measuring the Z-coordinate. In figure 59 examples of the shape data of the top and bottom surfaces of a plate are shown. In this case the top surface is slightly flatter with a maximum deviation of 28 µm compared to the 32 µm of the bottom surface.
Figure 59: An example of the shape of the top and bottom shape of a photomask. The size of the plate is 0.5 x 0.75 m²

Which surface to use as reference is in practice determined by how the mask is supported in the exposure equipment. Two ways are possible [46]. In the proximity aligner the mask is put in close contact (100 µm) with the substrate. The distortions of the mask will in this case depend on how well this proximity distance can be controlled over the mask area. In the projection aligner the mask is clamped to two rods at its edges (see figure 51). This clamping introduces distortions of the image, especially in areas close to the edges where the clamping is done. It is possible to compensate for these distortions by mapping the geometrical locations of features on the mask, provided these details of its design and tolerances are known. This is only of interest if we beforehand know which mask aligner will be used for the mask.

To sum up, we can claim that by using the Z-correction the absolute X,Y errors caused by Z-deformations of the plate, the stage, contamination on the stage, uneven plate support or air-bubbles, will be suppressed significantly.

7.4 Results of Z-correction

To prove the advantage of the Z-correction we will now present a result taken from a real case, cf figure 60, of a 0.6 x 0.6 m² plate suffering from a 16 µm particle stuck under the right hand edge. This figure and the following figures have dual scales in the X,Y plane. The local scale in figure 60, shown above at the top right corner, is 500 nm, while the absolute distance between gridpoints is on the order of several centimeters.
Figure 60: Illustration of a real case that may occur in the production of a photomask. A 16 µm particle was stuck under the plate along the right hand edge of a 0.6 x 0.6 m² photomask. Another 6 µm particle was stuck under the plate close to the upper left corner. The Z-data shows the difference of two measurements before and after the plate was cleaned. As seen in the upper right graph the tiny particle causes an X,Y deformation of several hundred nm. The lower left graph shows the corrected results after applying the Z-correction. The Z-scale is 2 µm/div.

7.5 Self-calibration in three dimensions

The purpose of a self-calibration is to find the absolute deviations of the plate and the stage of the metrology tool in their different Cartesian coordinate systems. Z-correction will enhance the calibration and also make the calibration much more robust against deformations of the plate during the calibration process. The bottom surface of the plate is not perfect. For this reason we cannot avoid that the measurement marks will move from their
original locations on the top-surface when the plate is deformed in Z when resting on the stage. Another error source is the non-flatness of the stage. If we imagine a perfectly flat plate and put it on a non-perfect stage, the measurement marks will also move sideways. A big advantage with the Z-correction is that it is possible to separate deviations caused by a non-flat plate from deviations caused by a non-flat stage. This is because in a self-calibration the plate is always translated and rotated inside the calibration area on the stage. In a calibration, deviations of the plate will move with the plate i.e. stay in the plate coordinate system and deviations in the stage will stay in the stage coordinate system. This is true also when Z induced deviations are involved. That fact is important to realize in a self-calibration. Another equally important condition is that the plate used for calibration should be rigid. When using large plates of square meters in size, the latter condition cannot be fulfilled. But when Z-correction is applied in the self-calibration, the non-rigid plate will be “made” rigid and therefore fulfill that condition.

In figure 53 the different parts involved in a Z-measurement were illustrated. It is not possible to separate deviations of the X-bridge to flatness variations on the stage in a Z-measurement. The specification of the X-bridge surface flatness and the Zerodur stage flatness are in the range of ±2µm. One reason for the former tight specification is that the X-car is floating on air-bearings on the X-bridge at an approximate distance of 8 µm. For this reason the flatness demand is high in order to avoid oscillations or that the X-car gets stuck in this narrow gap. When a plate is measured, the X-bridge in combination with the stage will serve as the reference in the Z-direction. Potential deviations of the X-bridge from a straight line can therefore give an erroneous measurement of the plate thickness. In practice only third-order and higher terms will introduce errors as a first order effect (tilt) will only generate an offset and a second order effect will generate a scale error in the X-direction. Scale errors must anyhow be calibrated using a traceable reference, and so this error will be removed by calibration anyway. In the development of the MMS15000, the X-bridge shape was investigated both from the measurement record supplied by the manufacturer and with the help of a theoretical model of the sag of the X-bridge when it is mounted in the machine. The result from this investigation showed that the remaining errors (above second order) could be neglected.

We will now demonstrate how Z induced errors from the plate and the stage can be separated using self-calibration. In this example we simulate a perfect plate in the sense that all measurement marks are placed in a perfect Cartesian grid with a pitch of 40 mm in the X and Y direction respectively when it is not affected by gravity. The top-surface is also perfectly flat. However, the
bottom surface has some variation in shape. We also simulate a perfect stage meaning that no X, Y deviations exist in the calibration plane within the distance T over the stage surface. The stage flatness is however not perfect. The pitch used for the reconstruction of the stage function is 50 mm in the X and Y direction respectively. In table 11 the placements of each measurement are shown.

<table>
<thead>
<tr>
<th>Placement</th>
<th>Rotation (radians)</th>
<th>Misplacement angle (milliradians)</th>
<th>Translation X (mm)</th>
<th>Translation Y (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>-0.25</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>π</td>
<td>0.25</td>
<td>0</td>
<td>41</td>
</tr>
<tr>
<td>3</td>
<td>π/2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>3 π/2</td>
<td>0.3</td>
<td>39</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 11: The different placements of the plates on the stage.

We simulate a long wave Z-shape of with the approximate amplitude of ± 10 µm on the bottom-surface of the plate and a long wave Z-shape with an approximate amplitude of ± 5 µm of the stage. The plate and stage used are shown in figure 61.

Figure 61: The plate and stage used in the simulation.
The goal of this exercise is to reconstruct the Z-shape of the plate and stage after calibration. If this can be done we will have shown that shapes in the different coordinate systems will stay in their coordinate systems and that the self-calibration algorithm will be able to separate them. In figure 62 the result from the calibration is shown.

**Figure 62**: The output X,Y gradients of the plate and stage after calibration and the reconstructed Z shapes. The X,Y data is presented without Z-correction applied to show the physical deviations to the Cartesian grid caused by the non-flat stage and plate. Only the calibrated area of the stage is shown. The scale in the X,Y data graphs are 100 nm/div. The Z-scale in the graphs is 2µm/div.

In the figure the plate and stage X,Y deviations are shown. These deviations are only caused by the non-flat bottom-surface of the plate and the non-flat surface of the stage. The output X,Y data corresponds to the gradients of the surfaces. The reconstructed surfaces are therefore calculated 2D integrals of
the gradients scaled by the reference distance T above the stage surface. Note that only the covered area of the measurements can be reconstructed. Therefore the stage X,Y gradient data and the reconstructed stage Z-data can only be calculated in this area.

The Z-correction is a linear operation. This means that Z-related errors may be added to X,Y related errors in the same coordinate system. It is not possible to separate these errors in a single measurement just by looking at the deviations. But since we measure the Z-height together with the X,Y coordinate we can compensate for effects directly related to the bending of the plate, so that after Z-correction has been done we are back to solving a problem in the reference plane that is a purely 2D problem. As mentioned above, the calibration is only valid in the reference plane T above the stage surface. However, it is possible to perform a calibration using another height reference plane, i.e. a different thickness of the plate. By using these two calibrations it is easy to interpolate between the two calibrated heights and thus obtain the correct compensation.
7.6 Real data 3D self-calibration

To demonstrate the necessity of applying the Z-correction for ultra-precision measurements we will now present the results of a self-calibration procedure of a MMS15000 with and without using Z-correction in the self-calibration process. The full details are discussed in paper E. In a verification of a calibration, four placements of the plate with its measurement marks are measured by rotating it in the X,Y plane to the angles 0, π/2, π and 3π/2 radians. This is done in the center of the calibrated area of the measurement machine stage. In this case we use a quartz plate having a matrix of 18 x 16 measurement marks separated by a pitch of 42 mm in both the X and the Y directions. The thickness of the plate is 8 mm. Figure 63 shows the four placements graphically with the local gridpoint scale of 1 mm.

![Placements](image)

**Figure 63**: The four measured placements used in the verification of the calibration. The local scale is 1 mm in both X and Y directions.

The X,Y data and the measured Z-data for the first placement (0) are shown in figure 64.
Figure 64: The X,Y deviations of the measurement marks, shown in a local scale of 500 nm at each gridpoint, and the measured Z data with the plate at placement angle 0. The Z-range (min to max) of this measurement is 32 µm. The Z-scale is 2 µm/div.

The quality of a self-calibration is verified by four new measurements with the placement angles 0, π/2, π and 3π/2 in the calibrated area. In figure 65 the result of two such verifications are shown, one without using Z-correction and the other when Z-correction was used.
Figure 65: The overlay of the four verification measurements at placement angles 0, $\pi/2$, $\pi$ and $3\pi/2$ without (left) and with Z-correction (right) used in the self-calibration process. The local X,Y-scale at each gridpoint is 200 nm/div.

In table 12 the $3\sigma$ of the different overlays are shown.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>angle</th>
<th>X without Z-correction (3σ) nm</th>
<th>Y without Z-correction (3σ) nm</th>
<th>X with Z-correction (3σ) nm</th>
<th>Y with Z-correction (3σ) nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>273</td>
<td>215</td>
<td>57</td>
<td>67</td>
</tr>
<tr>
<td>2</td>
<td>$\pi/2$</td>
<td>174</td>
<td>210</td>
<td>65</td>
<td>59</td>
</tr>
<tr>
<td>3</td>
<td>$\pi$</td>
<td>228</td>
<td>176</td>
<td>78</td>
<td>83</td>
</tr>
<tr>
<td>4</td>
<td>$3\pi/2$</td>
<td>201</td>
<td>201</td>
<td>57</td>
<td>67</td>
</tr>
</tbody>
</table>

Table 12: The deviations from the Cartesian grid measured as $3\sigma$ after a verification with and without using Z-correction.
The $3\sigma$ is a standard figure of merit used in this industry for verification of overlay and registration (i.e. uncertainty). In this case the overlay $3\sigma$ is calculated as the deviation of each measurement compared to the average of all measurements. Here an $18 \times 16$ matrix covering an area of $0.8 \times 0.8$ m$^2$ was measured. As seen in the table and in figure 65, $Z$-correction enhances the self-calibration significantly.

We have in this thesis shown that $Z$-correction improves the self-calibration of the MMS15000 and it is an absolute prerequisite for the ultra-precision metrology needed in the large area display industry. How much improvement is achieved depends on the flatness of the bottom-surface of the photomask, stage flatness and error sources such as contamination or pockets of air that might be stuck under the plate when it is resting on the stage surface of the metrology tool. In the photomask writer, $Z$-correction is used for compensating the geometrical placement of features in the writing process. In the latter case this compensation is very important since the cost is extremely high of losing a photo mask that might have taken four days to write, because of a particle stuck at the bottom surface.

From our experience an absolute placement accuracy of the mask pattern of less than 100 nm ($3\sigma$) for large area masks is not possible without using $Z$-correction even when using state of the art quartz blanks in the calibration process. By default the $Z$-correction function will adjust feature locations so that the deviation to the Cartesian grid is minimized in a flat reference plane. It is also possible to use the top-surface of the photo mask as the reference if the supplied quartz glass top-surface $Z$-deviation data is available for the gradient calculations. This technique is expected to become very important in the forthcoming manufacturing of photomasks used by the next generation high definition projection aligners. They will be used in the production of large area 4K ($3840 \times 2160$ pixels$^2$) and even higher resolution displays where the flatness demand of the mask pattern side of the photomask is higher than today [47].
The development of the 2D ultra-precision measurement machine MMS15000 was indeed a challenge. The large area combined with extremely challenging geometrical tolerances required new techniques to be developed for the ultra-precise measurements. The most important quality we had to build into this metrology tool was repeatability. Without good repeatability it is not possible to achieve good registration, normally referred to as uncertainty in dimensional metrology. We have in this thesis shown that the developed random phase method has an extremely good repeatability of around 1.5 nm (3σ) in a pitch measurement. We have also shown that the repeatability of hundreds of measurement marks over an area of 0.8 x 0.8 m² is in the range of 10 nm (3σ) despite being measured over time intervals of hours.

When launching the idea of the MMS15000 the error budgets showed that it was not possible to fulfill a registration specification of 90 nm (3σ) over an area of 1.4 x 1.5 m². However, experiments done on older smaller systems did show that the potential was there. Even if the fundamental physics is known beforehand the optics and the mechanical system with its moving parts are very difficult to estimate in all the details. It is practically impossible to separate correlated error sources from random error sources. Another way to express this is that all information about the system behavior is not known beforehand. The selection of materials for parts in the machine is of vital importance. But even if low-expansion materials are chosen it is still important to have extremely good temperature control of the entire system. Air-turbulence in the interferometer legs is another error source that is extremely difficult to model. Therefore we had to run many experiments in order to understand the airflow in the chamber and how this source of error could be minimized. If you have made all possible efforts in the design of the measurement system you still need to consider how the measured object, i.e. the photomask, is affected by the system in a measurement. As discussed in this thesis it is not possible to make a good calibration without compensating for the fact that the mask itself is not rigid during the self-calibration process. To achieve a measured X,Y position accuracy of 100 nm (3σ) on a large photomask is hardly possible without using the Z-correction. This is true even when using state of the art quartz blanks in the calibration process.
The general self-calibration algorithm developed for the MMS15000 has been shown to be very robust even in situations when the layout scheme of a plate with measurement marks is close to being underdetermined. The iterative method we use is more robust than a classical least-squares minimization because we put in an initial stage function as a first guess. The shape of this initial guess will severely influence the final result when the parameter matrix has a low rank.

The registration specification of the MMS15000 is 90 nm (3σ) over the full stage area of 1.4 x 1.5 m². To verify this number is a highly time consuming process, especially since the largest available golden plate (GP) used in the calibration is smaller than the stage.

To conclude this thesis we will therefore show (cf figure 66) a final example of a real verification of a 0.7 x 0.7 m² golden plate showing the outstanding performance of the MMS15000. This verification was performed by four rotation placements after a self-calibration. Each set of measurement took ~45 minutes in time. As seen, the maximum error is 49 nm (3σ). We are proud of having achieved this outstanding result considering all the mechanics, optics, software and handling of the plate involved in the self-calibration and the verification over such a large area.

So far all MMS15000 systems have been shipped to Asia. The future will tell if Europe will acquire a MMS15000 as an ultra-precise reference system, perhaps as a joint venture among European National Metrology Labs to be able to cope with large area measurements at the nanometer-scale. By adding an ultraprecise Z-axis we look forward to extending the performance of the machine to full 3D ultra-precision.
Figure 66: An uncertainty (registration) verification measurement after calibration of an area of approximately 0.7 x 0.7 m² on the center of the stage. The local scale in the graph is 100 nm and the results are given as three times the standard deviation of the measured X and Y positions for each rotation placement of the GP.
9 References


Box 3141,SE-183 03 Täby, Sweden
website: www.micronic.se (MMS15000)


[21] Micronic Mydata AB
   Box 3141, SE-183 03 Täby, Sweden, website: www.micronic.se (LRS and Prex 10)


[25] MIKES, Box 9, Tekniikantie 1, FIN-02151 Espoo, Finland, website: www.mikes.fi, (Certificate of Calibration: M-05L37)


[34] Sam Kash Kachigan, ”Statistical analysis”, Radius Press ,pp 111,1986


