The impact of the reticle and wafer alignment mark placement accuracy on the intra-field mask-to-mask overlay

Richard van Haren, Steffen Steinert, Orion Mouraille, Koen D’havé, Leon van Dijk, et al.
The impact of the reticle and wafer alignment mark placement accuracy on the intra-field mask-to-mask overlay

Richard van Haren\textsuperscript{a}, Steffen Steinert\textsuperscript{b}, Orion Mouraille\textsuperscript{a}, Koen D’havé\textsuperscript{c}, Leon van Dijk\textsuperscript{a}, Jan Hermans\textsuperscript{c}, Dirk Beyer\textsuperscript{b}

\textsuperscript{a}ASML, Flight Forum 1900 (no. 5846), 5657 EZ Eindhoven, The Netherlands
\textsuperscript{b}Carl Zeiss SMT GmbH, Carl-Zeiss-Promenade 10, 07745 Jena, Germany
\textsuperscript{c}IMEC, Kapeldreef 75, B-30001, Leuven, Belgium

\textbf{ABSTRACT}

The mask-to-mask writing error contribution as part of the on-wafer intra-field overlay performance has been extensively studied over the past few years. An excellent correlation ($R^2 > 0.96$) was found between the off-line registration measurements by the PROVE\textsuperscript{®} tool and the on-wafer intra-field overlay results. The residual mismatch between the off-line registration measurements and the on-wafer intra-field overlay was around 0.58-nm. This value is approximately 30\% of the dedicated chuck overlay performance of the scanner that was used.

A careful analysis was performed to understand and quantify the two dominant underlying contributors that are responsible for the 0.58-nm mismatch. The first contributor could be attributed to the reproducibility of the reticle alignment of the scanner (~0.43-nm after 10 wafers averaging). The second contributor was assigned to the sampling difference between the PROVE\textsuperscript{®} registration measurement and that of the alignment sensor inside the scanner (~0.39-nm). The sampling difference is a direct result of the relatively large metrology feature (alignment mark diffraction grating) in combination with older generation e-beam mask writing tools that were used in the experiments. Local grating placement variations are averaged out when the scanner alignment sensor is used for an overlay measurement. This is due to the large spot size and the scanning principle to obtain a position. This is fundamentally different for a mask registration tool since it has been designed to perform dedicated measurements on single features (globally or in-die) across the entire mask. Previous investigations used only two sampling points for each individual alignment mark diffraction grating in order to keep the total number of measurements and time under control. It is expected that the sampling difference will significantly decrease if state-of-the-art mask e-beam writers are used and/or if the number of sampling points as measured by the PROVE\textsuperscript{®} will be increased.

It might be obvious that the ability to perform dense off-line local registration measurements has large value to reveal local mask writing errors. The new local registration map (LRM) mode of PROVE\textsuperscript{®} can be used to average out local reticle writing errors enabling a more accurate placement determination of large metrology features like reticle and/or wafer alignment marks. The application of LRM can be used to further improve the accuracy between the scanner and the PROVE\textsuperscript{®} mask registration tool if required.

So far, all published correlation studies between off-line mask registration measurements and on-wafer overlay measurements were based on TIS (Transmission Image Sensor) reticle alignment marks. In this paper, we have applied LRM to improve the placement accuracy of more advanced PARIS (Parallel ILLIAS) reticle alignment marks. A comparison with on-wafer measurements is made. In addition, the placement accuracy of a wafer alignment mark is considered as well. The impact of a wafer alignment mark placement error due to reticle writing errors on the intra-field overlay is experimentally determined and discussed. This includes the effect of an applied intra-field scanner (reticle alignment) correction on the wafer alignment mark placement.

\textbf{Keywords:} Registration Error, Overlay, Computational Overlay, Reticle, Mask, Local Registration Map, Wafer alignment mark, APC, Feed-Forward

*Richard.van.Haren@asml.com; Mobile: +31-6-11786083; fax +31-402685430; www.asml.com
1. INTRODUCTION

The on-product overlay performance consists of both scanner and process-related contributors. Both groups need to be addressed and optimized to minimize the overlay in order to keep up with Moore’s law. The scanner-related overlay contributors are use-case specific. Of course, the scanner baseline performance itself plays an important role. The choice whether two subsequent layers are exposed on the same scanner and chuck (Dedicated Chuck Overlay) or on different scanners and different chucks (Matched Machine Overlay) has an impact on the lowest overlay performance that can be achieved. Thermal effects inside the scanner may play a role as well. For a low transmission mask, reticle heating related penalties may show up while for a high transmission mask lens heating may be more prominent. The best overlay performance can be achieved if the individual contributors are the same or very similar per exposed layer. For instance, reticle heating has no impact on overlay if the two masks used have similar transmission and absorption properties, exposed with the same dose, and aligned back to the same common wafer grid. Although the reticle heating contribution may be present in each individual layer, it drops out if the delta is considered. This is for example the case for the multiple Litho-Etch (LE) application. Generally speaking, the higher the commonality between the two exposed layers, the better the overlay. This is why the Dedicated Chuck Overlay performance (same chuck and lens) is generally better than the Matched Machine Overlay (different chuck and lens) [1,2]. The commonality between the exposed layers can be improved and maintained over time by using reference wafers [1].

Examples of process-related overlay contributors are wafer distortion due to patterned stressed thin films and/or etch. Etch induced pattern shifts may occur if the etch direction is not perpendicular to the wafer surface. The pattern that is defined after the photo resist development step will experience an additional displacement during the etching process. Also, for the process-related overlay contributors commonality is key. In case the etch-induced pattern shift is the same for both layers, it drops out if the delta is considered. The same is true for wafer distortion. Two subsequent layers exposed on a distorted wafer can still have a state-of-the-art overlay performance with respect to each other. The commonality is high since they share the same common (but distorted) reference layer. Obviously, the commonality is low if the exposed layers are compared to the common (but distorted) reference layer. Scanner exposure corrections are required to achieve good overlay performance. Consequently, the second exposed layer needs to follow the applied corrections of the first layer to maintain the state-of-the-art overlay performance between the two exposed layers.

Masks can never be made identical since they represent different layers of the device. Some layers only contain contact holes while others have line/space patterns. The pattern density is not the same from mask-to-mask and within the mask. Consequently, each mask will have its own reticle writing error (RWE) signature. Higher commonality between similar masks may be achieved by writing the masks successively on the same e-beam writer. However, in practice it might be impractical, too costly, or even impossible due to the difference in pattern density and/or mask specifications. Masks with different requirements may not be produced all by the same mask shop at the same time. Critical overlay layers co-exist next to non-critical overlay layers.

Commonality can be improved by fully characterizing the masks on off-line mask registration tools and using this information to enable Feed-Forward corrections to the scanner. This makes the masks virtually the same during exposure [3]. This functionality can only be enabled if we have a perfect correlation between the off-line mask-to-mask overlay and the on-wafer measured overlay. In previous work, we managed to isolate the RWE contribution from other intra-field overlay contributors by a carefully setup experiment [4]. An excellent correlation was found ($R^2 > 0.96$) between the off-line determined mask-to-mask overlay and the on-wafer measured overlay. The accuracy was approximately 0.58-nm and all underlying contributors were identified.

All work we published so far was based on Transmission Image Sensor (TIS) reticle alignment (RA). The NXT:1970Ci we used for the exposures is equipped with both a TIS and a Parallel ILIAS (PARIS) reticle alignment sensor. All NXT systems from 1970 and higher are equipped with a PARIS sensor. The PARIS sensor has been designed to enable reticle alignment as well as capturing thermal effects like lens heating and reticle heating [2]. We decided to extend the correlation study and include PARIS reticle alignment for completeness. We further consider the effect of the RWE on the wafer alignment (WA) marks and the impact it has on the measured overlay. It should be realized that the RWE is not the only contributor affecting the position of the wafer alignment mark, as we will see in the next section.
1.1 Reticle writing error impact on the wafer alignment mark position and overlay

In this section, we consider the role of the wafer alignment mark and the impact it has on the measured on-wafer overlay. We basically discriminate three different scenarios when a wafer gets exposed:

- No wafer alignment. The very first layer is always exposed without any wafer alignment. Only reticle alignment is performed. Either alignment marks are exposed as separate images (zero-layer marks) or the alignment marks are exposed together with the first product layer.

- Direct wafer alignment. Wafer alignment is performed on alignment marks that were exposed in one of the previous layers. Overlay is controlled to the layer in which the alignment marks are defined. In this scenario, the RWE of the alignment mark has an impact on the overlay between the exposed layer and the previous layer. The designed position of the alignment mark is different than the actual measured position of the alignment mark. The delta shows up as a translation penalty in the measured overlay between the exposed layer and the previous layer.

- Indirect wafer alignment. Wafer alignment is performed on alignment marks that were exposed in one of the previous layers. However, in this case overlay is not necessarily controlled to the layer in which the alignment marks were defined. If two layers share the same common wafer alignment reference layer, the impact of the RWE on the alignment marks is the same for both layers and drops out if the overlay between the layers is considered.

A common wafer alignment reference layer has many more advantages. In principle, the alignment mark position is not only affected by the RWE. Many other contributors like reticle heating, lens heating, reticle clamping, and pellicle usage may play an important role as well. Some of these contributors are constant while others vary from wafer-to-wafer. Moreover, a mismatch between the exposure and measurement grid may be present as well. However, if two exposed layers share the same wafer alignment reference layer, the commonality is high, and all the contributions mentioned above are exactly the same for both exposed layers. As a direct consequence, they drop out and a state-of-the-art overlay performance is achieved.

The direct wafer alignment scenario is the most challenging one. All contributors mentioned above do play a role if the overlay between the exposed layer and the previous wafer alignment layer is considered. Luckily, not all contributors are present at the same time. Moreover, many applications have been developed to strongly suppress or even eliminate these contributors. One of the goals of the current publication is to explore the impact of the RWE on the alignment mark position and its impact on the measured overlay.

![Figure 1: Illustration of the RWE impact ($\Delta x$, $\Delta y$) on the alignment mark defined in the origin $\mathcal{O}$ at (0,0). Layer 1 is exposed in the nominal grid. The alignment mark defined at (0,0) is used for wafer alignment to expose Layer 2. In the new exposure grid the new origin is shifted by ($\Delta x$, $\Delta y$). Layer 2 is exposed in this new exposure grid.](image)

In Figure 1 we show the impact of an RWE on the alignment mark that is defined in the origin $\mathcal{O}$ (in blue). The wafer alignment mark at (0,0) is on the same mask as all other marks represented by the blue dots and exposed as Layer 1 in the blue reference grid. Suppose the RWE on the alignment mark at (0,0) is ($\Delta x$, $\Delta y$) and the only error present on the mask. All other marks are located on their designed position. In case wafer alignment is defined on the mark at (0,0), the...
alignment system of the scanner actually finds the mark at \((\Delta x, \Delta y)\). This will define the new origin \(O\) and grid (in black) in which the second layer is exposed. In case the same mask is exposed as second layer all points on the mask will experience a shift of \((\Delta x, \Delta y)\). The measured overlay between Layer 2 and Layer 1 can be described by a translation with a magnitude of \((\Delta x, \Delta y)\). This overlay penalty can be avoided if the actual alignment mark location is known. In case the alignment mark position is defined at \((\Delta x, \Delta y)\) in Layer 1, the black origin and grid coincides with the blue origin and grid and the overlay penalty will be zero. To enable this functionality, we need off-line mask registration measurements.

2. EXPERIMENTAL DETAILS

2.1 Mask-to-mask overlay measurements

In this work, we make use of two different methods to determine the mask-to-mask overlay due to reticle writing errors:

1. First of all, the mask-to-mask overlay can be fully characterized by the PROVE® registration tool. The characteristics of this tool are the DUV actinic wavelength (193-nm) in combination with litho-grade optics, high NA, superior stage concept with tight environmental control, and sophisticated 2D correlation methods [5]. Although in this study the focus is on DUV masks, the PROVE® registration tool is capable to perform registration measurements on EUV masks as well [6].

2. The mask-to-mask overlay can also be characterized on the scanner provided that all other overlay contributors are negligible or drop out in the overlay measurements. This can only be achieved by carefully setting up the experiment. We extensively reported on how this can be achieved in an earlier publication [4]. It basically boils down to increasing the commonality between the layer 1 and layer 2 exposures as much as possible. Since we physically have two different masks, the RWE contribution in the on-wafer measured overlay remains.

Excellent correlation results \((R^2 > 0.96)\) were shown between the mask-to-mask overlay as measured on the PROVE® registration tool and the on-wafer overlay as measured by the scanner. The accuracy was determined to be 0.58-nm which is ~30% of the scanner baseline overlay performance that was used in the experiments. A careful analysis was done to understand the 0.58-nm mismatch and the largest underlying contributors were identified to enable further reduction [4].

In a follow-up publication we demonstrated that the commonality between the layer 1 and layer 2 exposures can be broken if one of the masks is equipped with a pellicle [7]. Additional overlay contributors come into play and need to be addressed separately. After isolating and removing these additional overlay contributors from the measured overlay, the excellent correlation results between the mask-to-mask overlay as measured on the PROVE® registration tool and the on-wafer overlay as measured by the scanner is restored. The conclusion that off-line mask measurements can be used to improve the on-wafer overlay remained therefore unchanged.

All experiments in the present work are executed on the ASML NXT:1970Ci (≤ 2-nm single machine overlay, dedicated chuck, full wafer coverage). On this scanner reticle alignment can be done using TIS reticle alignment or PARIS reticle alignment. All results we presented so far were based on TIS reticle alignment. In this work, we include PARIS reticle alignment to see if the conclusions remain the same.

2.2 Mask layout

The same ASML BMMO (Baseline Matched Machine Overlay) masks that were used previously are also used in the current investigation. The mask identification labels we use throughout this paper are:

- N004 \(\rightarrow\) EBM-6000
- N005 \(\rightarrow\) EBM-5000

The masks were manufactured on two different (older generation) e-beam writing tools, the EBM-6000 [8] and the EBM-5000 [9]. All scanner experiments and PROVE® measurements were performed without pellicles.
Figure 2: Layout of the two Baseliner masks that were used in this experiment. Both masks are identical by design and contain metrology modules in a 13×19 layout. PROVE® measurements were done on the locations indicated by the black dots. The XPA grating that can be read by the scanner alignment system as well as the PROVE® enabling a direct comparison. PARIS reticle alignment marks were used this time.

Figure 2 shows the layout of the N004 and N005 masks used. Both masks are identical by design and contain metrology modules in a (13 × 19) layout within the image field. One module is shown in an expanded view. It contains a wafer alignment mark (XPA) that can be read out on the scanner. The overlay between two masks can be determined by exposing the second mask with a shift of +520-µm in the y-direction with respect to the first mask exposure. By measuring the distance between the two XPA’s and subtracting 520-µm, the overlay is obtained. Note that this method can be used to determine the overlay between two masks that are physically different (like N005 and N004) or for a single mask to itself. In the first case the RWE difference between the masks will contribute to the on-wafer measured overlay while in the latter case the RWE contribution drops out. Both methods are used in the current work.

PARIS reticle alignment marks (2×7) are located outside the image field on both ends of the mask. The PARIS marks are indicated by the small red rectangles in Figure 2. An expanded view is added to show more details of a single PARIS mark. It consists of a center cross in between two diagonal (U and V) gratings. The U and V gratings are used by the scanner to determine lens aberrations at 7 locations across the width of the field as well as to extract the positional information for reticle alignment. Apart from the regular linear 6 parameters like for TIS reticle alignment, 6 additional parameters that are linked to intra-field distortions can be corrected as well. We assumed that the center cross represents a PARIS mark position on the mask very well since it is located in between the U and V gratings. The distance between the cross and the center of the U or V grating is 100-µm (at 4x). This is why we started off to measure the center-cross positions to be representative for the PARIS reticle alignment mark locations on the PROVE® registration tool.

3. RESULTS

3.1 PARIS cross measurements: mask-to-mask overlay compared to on-wafer scanner overlay measurements

Figure 3 shows the N004 to N005 mask-to-mask overlay. The on-wafer overlay measurements are compared to the PROVE® measurements that are based on the center cross positions of the PARIS marks. The results were not in line with our expectation, a large offset of approximately (3.0, -1.3) nm was observed in the difference between the mask-to-mask measurements as measured by the PROVE® tool and to the scanner on-wafer measurements. Initially, the scanner PARIS reticle alignment was suspected. The main reason was that the mask-to-mask overlay fingerprint measured on the PROVE® looked very similar to the case where the TIS reticle alignment (RA) gratings were measured [4]. The only difference is that the field locations now get a 12 parameter PARIS reticle alignment correction instead of the linear 6
parameter correction when TIS reticle alignment was used. The magnitude of the field fingerprint was quite comparable, ~2.2-nm (mean+3σ) for TIS RA and ~1.9-nm (mean+3σ) for PARIS RA, see left-hand side field plot in Figure 3.

Figure 3: Mask-to-mask overlay based on the PARIS cross measurements as measured by the PROVE®. The expected intra-field overlay performance is shown in the plot on the left-hand-side. The on-wafer intra-field overlay after PARIS RA shows a large translational offset as shown in the center plot. As a consequence, a large mismatch is observed in the delta plot.

However, no specific reason could be found for the large translation overlay penalty that was observed in the on-wafer overlay measurements. For that reason, the assumption that the center cross of the PARIS grating represents the position as measured by the scanner was reconsidered. Reticle alignment inside the scanner is performed on the PARIS gratings. The crosses are not used. For the TIS reticle alignment marks the gratings are far away (~8-mm) from the crosses and since the scanner is measuring the TIS gratings only, a grating measurement on the PROVE® was a natural thing to do. For the XPA wafer alignment marks (see Figure 2) both cross measurements and grating measurements were done. Although the XPA grating is much closer (~750-µm) to the cross, differences in the nanometer regime were observed. This triggered us to start measuring the PARIS gratings as well. In the next section, we elaborate more on the PARIS grating measurements on the PROVE® tool.

3.2 PARIS grating measurements: dense local registration measurements

In this section, we aim to determine the PARIS grating position on the mask more accurately. In Figure 4 on the left-hand-side, the cross in the center of the PARIS mark is shown. The PROVE® measurement yields an (x, y) position. The mask-to-mask overlay shown in Figure 3 was based on these PARIS cross registration measurements. For the PARIS grating measurements, we made use of the new Local Registration Map (LRM) functionality.

Figure 4: Center cross inside the PARIS reticle alignment mark. For the PARIS gratings, 1D and 2D measurement modes are selected. Only at the grating edges, a 2D measurement mode can be applied since the features to be measured contain both x and y information. Since the scanner only measures the gratings away from the edges, we use the 1D measurement mode results only. Measurements were done every 5-µm with a (6×6) µm² ROI setting.

A user-defined 2D array measurement layout is superimposed on the PARIS mark enabling dense registration measurements of the grating lines. The measurement mode is either 1D or 2D as indicated by the blue dots and red dots in Figure 4. The PARIS mark design consists of two gratings under -45 and +45 degrees with respect to the x-axis, see Figure 2. Most of the measurements are done in 1D measurement mode. Only at the outer edges, 2D registration
measurements can be done since the features contain both $x$ and $y$ position information. Since the scanner only measures the inner parts of the $U$ and $V$ gratings, we restrict ourselves to the 1D measurements. An illustration is shown on the right-hand-side in Figure 4. The measurement pitch (at mask level) is 5-$\mu$m in both the $x$ and $y$ direction and the Region of Interest (ROI) is $(6\times6)$ $\mu$m$^2$ for each measurement.

In Figure 5, the PROVE® LRM results are shown for both the $U$ and $V$ gratings of the PARIS mark. The scale is 5-nm (1×) as shown at the bottom. The PARIS gratings and the center crosses positions were measured during the same run and in the same reference grid. If we consider the cross of one of the PARIS marks, an $x$ and $y$ value of (-0.04, -3.40) nm represented by the red vector was measured. The grating position can be determined by combining the average LRM vectors $\langle u \rangle$ and $\langle v \rangle$ from the $U$ and $V$ gratings, respectively. This provides us the PARIS grating position of (1.46, -3.57) nm represented by the blue vector. The delta between the PARIS cross measurement and the grating measurement is (-1.50, 0.17) nm as shown by the black vector. All the vector values are at wafer level (1×) and are not negligible.

![PARIS U and V grating Local Registration Measurements](image)

Figure 5: PROVE® LRM results for the $U$ and $V$ gratings of the PARIS mark. The PARIS center cross is measured as well and shown in red. By combining the average of the local measurements for the $U$ and $V$ gratings, the PARIS grating position is obtained as shown by the blue vector. The delta between the cross and the grating positions is shown by the black vector (-1.50, 0.17) nm.

Interestingly, the delta between the PARIS cross and grating positions is observed for all 14 PARIS reticle alignment marks and for both the N004 and N005 masks, see Figure 6. The PARIS mark that was considered above is located on the N004 mask at the bottom left corner.

![PARIS Cross Grating delta PROVE, N004](image)

Figure 6: PARIS cross grating delta measurements for N004 and N005. An offset is observed between the cross and the grating positions on the mask. It is not the same for N004 and N005. The plot on the right-hand-side shows the delta between the two masks. It reflects the difference in the offline determined mask-to-mask overlay if the PARIS gratings are used instead of the crosses.

Although the main difference can be described by a translation, variations between the crosses and gratings of the different PARIS marks can be observed as well. This is quite clear by considering the cross-grating differences for the top-row PARIS marks on mask N005. Note that the average offset between the cross and the grating is not the same for the N004 and N005 masks. In the present case, the delta measurements even point in different directions. This implies...
that if the offline mask measurements are based on the center crosses and not on the grating positions of the PARIS marks, large translation overlay penalties are expected when the offline measurements are compared with the on-wafer measurements. This is exactly what we see in Figure 3.

The LRM functionality offers the opportunity to quantify the local placement variations inside large metrology structures like the PARIS and XPA gratings. In Figure 7, we show dense registration measurements for the PARIS and XPA marks. The PARIS mark size is 400-μm×200-μm at 4x. This is 200-μm×200-μm for each U and V grating. The XPA mark consists of 4 quadrants and only the indicated x and y-gratings are relevant for the scanner readings. These gratings are even larger, the x-grating is 736-μm × 512-μm and the y-grating is 512-μm × 736-μm, all at reticle level (4x). The XPA mark at (0,0) on both N004 and N005 mask was selected for the dense registration measurements.

![Figure 7: PARIS U and V grating local registration measurements for two different reticle alignment marks.](image)

Two PARIS marks were selected to demonstrate the LRM capability. Apart from an average grating position, local variations within a ~4-nm range are observed as well. Especially for PARIS mark 2, areas inside the V grating can be observed with elevated reticle writing errors. They may find their origin in the main deflection field size of 512-μm and the sub-deflection field sizes of 64-μm and 32-μm of the EBM systems used in this work [8,9].

For the XPA gratings, we removed the average grating position. The remaining intra-grating variation varies approximately from -2-nm to +2-nm at mask level.

In the next section we consider the mask-to-mask overlay as determined on the PROVE® in case the PARIS gratings measurements are used. We will again compare the results with the on-wafer overlay measurements.

### 3.3 PARIS grating measurements: mask-to-mask overlay compared to on-wafer scanner overlay measurements

Before comparing the offline mask-to-mask overlay based on PROVE® measurements on the PARIS gratings with the on-wafer scanner overlay results, we will first consider the impact on the N004 and N005 masks separately. Figure 8 shows the XPA grating positions in the reference grid that is based on the 14 PARIS marks. To be more specific, all XPA and PARIS grating positions are first measured with respect to their design positions in the measurement grid of the PROVE® tool. After that, the 14 PARIS mark measured positions are used as input for the 12-parameter PARIS reticle alignment model. The modelled parameters are used to calculate the impact on the XPA intra-field positions and added to the measured position of the XPA at the same location. The new positions of the XPA marks for both the N004 and N005 masks after PARIS grating reticle alignment are shown in Figure 8. All intra-field positions are affected by the 12-parameter reticle alignment. The same position offsets are expected on-wafer. This implies that for N004 or N005 the difference between the actual measured position by the scanner alignment system and the designed position is the same as shown by the N004 and N005 field plots in Figure 8.
Figure 8: The expected intra-field XPA offsets after 12-parameter PARIS grating reticle alignment for N004 and N005. The vector plots represent the offsets of the XPA marks with respect to their designed positions. The arrows contain both the RWE and the RA impact. All plots are based on PROVE® measurements. The offline determined mask-to-mask overlay between N004 and N005 is shown on the right-hand-side. A large translation in the on-wafer overlay is expected.

The expected on-wafer overlay is obtained by subtracting the N005 field plot from the N004 field plot. The mask-to-mask overlay based on PROVE® measurements and the PARIS reticle alignment model is shown on the right-hand-side in Figure 8. A dominant intra-field translation penalty is expected in the overlay.

As a last step, we compare the mask-to-mask overlay as shown on the right-hand-side in Figure 8 with the on-wafer measured overlay. The results are shown in Figure 9. PARIS reticle alignment is used and the PROVE® measurements are now all based on the PARIS gratings as opposed to the crosses in Figure 3. The mismatch between the off-line determined mask-to-mask overlay and the on-wafer measured overlay can only be judged by considering the delta field plot on the right-hand-side in Figure 9. The residual mismatch is extremely small as represented by the 99.7% and max metrics of approximately 0.45-nm. This means we have a perfect correlation between the PROVE® and the scanner overlay measurements.

The mean+3σ value in x is slightly higher than expected and above the 99.7% and max values. We attribute this to the intra-grating variation [4], which makes the distribution non-Gaussian. Some horizontal rows have a small offset in the x-direction and are most likely linked to the mask e-beam writing process. These effects will be picked up differently by the PROVE® measurement tool compared to the scanner. The alignment sensor inside the scanner is less sensitive to these penalties due to large spot size and the scanning principle. In general, it looks like these effects start to dominate the mismatch between the off-line determined mask-to-mask overlay and the on-wafer measured overlay. They can only be reduced by applying the LRM mode on the XPA gratings instead of the two measurements per grating we did so far or by writing the masks on more state-of-the-art e-beam writers.

Figure 9: The mask-to-mask overlay as measured by PROVE® compared to the on-wafer results. The mismatch is below 0.45-nm. Although the PARIS gratings are measured more densely (36x36) after applying the LRM functionality, the 2-point XPA grating measurements were still used for the PROVE® mask-to-mask and the scanner on-wafer comparison.
3.4 The RWE impact on the wafer alignment mark placement and the consequence on overlay

One aspect we didn’t address so far is the RWE impact on the alignment mark position inside an image field on the intra-field overlay performance. It should be noted that not only the RWE is relevant but also the reticle alignment impact on the field location of the wafer alignment mark we use for wafer alignment. This means that the vector plots shown for N004 and N005 in Figure 8 become relevant. We selected the alignment mark at field location (0, -5.406) mm on mask N005 for which the reticle heating impact is relatively low. It actually corresponds with the XPA location shown in the expanded view in Figure 2. The deviation from the designed position is (1.75, -0.87)-nm as represented by the vector at the (0, -5.406) mm field position of the N005 mask in Figure 8.

In this paper, we start relatively simple and first consider the overlay of the N005 mask to itself. PARIS reticle alignment is used and N005 is exposed in photo resist covering the full wafer. Subsequently, the latent image of the XPA alignment mark at field position (0, -5.406) mm is selected for 16 different fields. Wafer alignment is performed on these wafer alignment marks and the second layer is exposed. This corresponds to the direct alignment scenario as already was described in the introduction. The same N005 mask is used for the second layer and exposed with a +520-µm shift in the y-direction relative to the first layer. Since the same mask is used for layer 1 and layer 2, all XPA offsets with respect to their designed positions, as shown in the center field plot of Figure 8, drop out. The only overlay penalty that remains is the offset on the XPA mark at field location (0, -5.406) mm that is used for wafer alignment. The photo resist is developed and the XPA marks of layer 1 and 2 were read out and corrected for the +520-µm shift to obtain the overlay.

![Measured overlay and Overlay: RWE induced translation removed](image)

Figure 10: Measured overlay in case an alignment mark is used in the previous layer. The RWE effect on the alignment mark is directly observed in the measured overlay. The overlay plot on the right-hand-side shows the expected improvement if the offline determined RWE contribution on the alignment mark is corrected for.

The overlay results are shown in Figure 10. The second to first layer overlay is largely determined by a translation penalty. A large improvement in the measured overlay is observed if the RWE induced translation penalty of (1.75, -0.87) nm is removed from the measured overlay. The resulting overlay is shown on the right-hand-side in Figure 10. The overlay improvement is significant. Usually, the designed alignment mark positions are used in the exposure job. In case these designed positions are corrected for the reticle writing contribution, the resulting translation penalties in the measured overlay can be eliminated and an automated process control (APC) loop correction for these kinds of overlay errors is not required.

4. DISCUSSION

An accurate off-line mask-to-mask overlay prediction can only be done if the reticle alignment mark positions are measured the same way as it is done inside the scanner. A center cross registration measurement to represent the PARIS alignment mark position is not recommended. We have shown that it may result in a large mismatch between mask-to-mask overlay as measured on the PROVE® registration tool and the on-wafer overlay measurements. Only after measuring the PARIS grating positions, the mismatch reduces to mean+3σ numbers below 0.63-nm. Initially this was not expected since the cross inside the PARIS mark is exactly located in between the two diagonal gratings and only 100-µm (4×) away from the center of the gratings. This in contrast to the TIS reticle alignment mark for which the center cross is...
8,000-µm (4×) away from the center of the x and y gratings. For that reason, we directly started measuring the grating positions of the TIS RA marks which resulted in a perfect correlation between the mask-to-mask overlay as measured on the PROVE® compared to the scanner on-wafer measurements [4]. In this work, we have shown that if the PARIS grating positions are measured on the PROVE®, similar results are obtained.

It is obvious that if a good match between the off-line determined mask-to-mask overlay and the on-wafer scanner overlay is required, PARIS gratings need to be measured and not the crosses. In this study, a sampling layout of two times (36×36) per PARIS mark was used to determine its position. This was done for all 14 PARIS marks for both N004 and N005. Since 14 PARIS marks are used for the reticle alignment, it might be of interest for the end-user to see the effect of a reduced sampling plan. We explored this by a reducing the number of measurements to only one per U and V grating.

The results are shown in Figure 11. On the left-hand-side, the delta between the mask-to-mask overlay as measured on the PROVE® and the on-wafer overlay is shown if the PARIS grating measurements are based on a (36×36) sampling layout. This sampling layout was applied to both the U and V grating, as shown in Figure 5. In order to reduce the measurement time, we evaluated the impact of a reduction in the sampling layout. The two extremes of (36×36) and (1×1) are shown in Figure 11. We definitely see a large improvement in accuracy in case a single PARIS mark is considered. This could already be concluded from the results presented in Figure 7. It is obvious that a (1×1) sampling scheme is more susceptible to local intra-grating variations. However, the intra-field impact is low (<0.1-nm) in case the two sampling schemes are compared. The PARIS mark intra-grating variation impact on the field positions is apparently low. This can be explained by the fact that the RA is done on 14 PARIS marks and thereby averaging out the intra-grating variation. A rough estimation of the residual intra-field penalty would be 0.5-nm/√14 ≈ 0.13-nm. Another explanation could be that the size of the PARIS mark is small compared to the main deflection field size of the e-beam writer. This is not the case for the XPA with its larger dimensions.

Figure 11: The impact of a reduced measurement sampling plan for the PARIS gratings on the mismatch between the mask-to-mask overlay and the on-wafer determined overlay. A reduction in the sampling plan has hardly any impact on the mismatch; even for one measurement point per U and V grating, the results look good.

Apparently, the sampling reduction for the PARIS grating measurements has less impact on the intra-field overlay performance than measuring the gratings instead of the crosses. This is due to the averaging effect of using 14 PARIS marks for reticle alignment. In the next paragraph, we explain that if a single location is considered the LRM mode becomes much more relevant. We will illustrate this based on the XPA and PARIS grating measurements shown in section 3.2.

It might be clear that the intra-grating variation of ~0.5-nm has a stronger impact on the mismatch between a PROVE® measurement and the scanner readout for a single XPA field position. This is a direct consequence of the sampling difference; while the PROVE® is designed to measure the position on the mask locally, the alignment sensor inside the scanner measures a larger part of the grating, averaging out the local intra-grating variations. Only if the LRM mode is applied to the grating area that is scanned by the alignment sensor spot, we expect a reduction in this error budget contribution. For this work, we decided there is no significant added value to measure the XPA marks on both masks for all (13×19) field positions, both x & y gratings, and (36×36) intra-grating positions. It would have resulted in 1,280,448 additional measurements. We concluded that the current accuracy which is 30% of the scanner overlay performance is more than sufficient. Moreover, the solution path is clear in case a better accuracy is required.
However, it is obvious that the PROVE® LRM mode reveals mask details that are worthwhile to explore further. Section 3.2 showed already mask e-beam related reticle writing errors present inside the PARIS and XPA gratings. It is anticipated that these errors might also be present for device patterns even for masks that are produced on more advanced EBM systems. We will continue exploring the mask impact on the edge placement errors (EPE).

5. CONCLUSIONS

In conclusion, the on-wafer mask-to-mask overlay can be fully predicted by performing off-line mask registration measurements on the PROVE® tool. The prediction includes the contribution of reticle alignment and is independent on the type of reticle alignment that is used whether it is TIS or PARIS marks. We have shown that a cross measurement that is supposed to represent the mark location is not recommended. It is crucial to measure the gratings instead. This is important since the gratings are also used by the scanner during reticle alignment. An accurate reticle mark placement measurement is a prerequisite to enable mask-to-mask overlay predictions that are based on off-line mask registration measurements. Since the mask-to-mask overlay including TIS or PARIS reticle alignment can be fully characterized by off-line registration measurements, we would like to conclude this part of the investigation.

One aspect that has also been considered in this paper is the RWE impact on a wafer alignment mark. Provided that the reticle alignment contribution is included in the determination of an accurate wafer alignment mark position, the overlay impact can be quantified as well. This means that the on-wafer overlay prediction capability based on off-line mask measurements can be extended towards the RWE impact on a wafer alignment mark.

REFERENCES