An Empirical Approach Addressing the Transfer of Mask Placement Errors During Exposure

B. Alles* (balles@ma.tum.de), B. Simeon (simeon@ma.tum.de)
University of Technology Munich, Boltzmannstr.3, 85748 Garching, Germany

E. Cotte (Eric.Cotte@amtc-dresden.com), T. Wandel (Timo.Wandel@amtc-dresden.com)
Advanced Mask Technology Center Dresden, Raehnitzer Allee 9, 01109 Dresden, Germany

B. Schulz (Bernd.Schulz@amd.com), R. Seltmann (Rolf.Seltmann@amd.com)
AMD Saxony LLD, Wilschdorfer Landstrasse 101, 01109 Dresden, Germany

ABSTRACT

Today's semiconductors consist of up to forty structured layers which make up the electric circuit. Since the market demands more powerful chips at minimal cost, the structure size is decreased with every technology node. The smaller the features become, the more sensitive is the functional efficiency of the chip with respect to placement errors. One crucial component for placement errors is the mask which can be viewed as a blueprint of the layer's structures. Hence, placement accuracy requirements for masks are also tightening rapidly. These days, mask shops strive for improving their positioning performance. However, more and more effort is required which will increase the costs for masks. Therefore, the transfer of mask placement errors onto the wafer is analyzed in order to check the guidelines which are used for deriving placement error specifications.

In the first section of this paper the basic concepts for measuring placement errors are provided. Then, a method is proposed which is able to characterize the transfer of placement errors from mask to wafer. This is followed by two sections giving a thorough statistical analysis of this method. In the fifth section, the connection to placement accuracy specifications on mask and wafer is established. Finally, the method is applied to a set of test masks provided by AMTC and printed by AMD.

Keywords: Registration, overlay, correlation analysis, specifications

1. REGISTRATION AND OVERLAY

There are basically two concepts for measuring the placement accuracy of structures on a mask or a wafer[1]:

- **Registration:**
  Describes the difference between real position and the design position. Mathematically speaking, registration is a function reg, which maps to each position (x, y) of the design area $\mathcal{M}$ the deviation $(\mathrm{dx}, \mathrm{dy})$ at this point:
  $$\begin{array}{rrcl} \mathrm{reg}: & \mathcal{M} & \rightarrow & \mathbb{R}^2 \\ & (x,y) & \rightarrow & (\mathrm{dx}(x,y), \mathrm{dy}(x,y)) \end{array} \tag{1}$$
- **Overlay:**
  Represents the relative placement errors between two registration fields reg_1 and reg_2:
  $$\begin{array}{rrcl} \mathrm{ovl}: & \mathcal{M} & \rightarrow & \mathbb{R}^2 \\ & \mathrm{ovl} & = & \mathrm{reg}_1 - \mathrm{reg}_2 \end{array} \tag{2}$$

Overlay is often used for specifying the relative placement error between two consecutive layers on a wafer, since it is crucial for the functional efficiency of the electric circuit. Wafer fabs want to be able to combine masks from several vendors. Hence, placement accuracy of masks is usually discussed in terms of registration. However, in rare cases the relative placement error between two masks is discussed and referred to as mask-to-mask overlay. The mask-to-mask overlay between two masks is given by the difference of their registration fields.

A detailed discussion of mask registration and its sources is given in.[2]

---

*Presenting & Corresponding Author, phone: +49/+89/289-17465, fax: +49/+89/289-17483

2. MODELLING THE PLACEMENT ERROR TRANSFER DURING EXPOSURE

In this section, a model connecting mask registration $(\mathrm{dx}_{\mathrm{mask}}, \mathrm{dy}_{\mathrm{mask}})$ and wafer registration $(\mathrm{dx}_{\mathrm{wafer}}, \mathrm{dy}_{\mathrm{wafer}})$ is proposed. If one were capable of printing wafers and measuring wafers and masks perfectly, the only source of wafer registration would be mask registration. In this case, the wafer registration would be given by the formula:

$$m \cdot \begin{pmatrix} \mathrm{dx}_{\mathrm{wafer}} \\ \mathrm{dy}_{\mathrm{wafer}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathrm{dx}_{\mathrm{mask}} \\ \mathrm{dy}_{\mathrm{mask}} \end{pmatrix} , \tag{3}$$

with m denoting the magnification factor of the exposure tool, typically $m = 4$.

However, the assumptions for this model are unrealistic: Nonlinear effects during exposure, like mask clamping, lens aberrations and limited accuracy of mask and wafer stage-speed during exposure, also contribute to the wafer overlay. Accordingly, the model given by equation (3) can be extended to:

$$m \cdot \begin{pmatrix} \mathrm{dx}_{\mathrm{wafer}} \\ \mathrm{dy}_{\mathrm{wafer}} \end{pmatrix} = \begin{pmatrix} o_x \\ o_y \end{pmatrix} + \begin{pmatrix} m_{xx} & m_{xy} \\ m_{yx} & m_{yy} \end{pmatrix} \begin{pmatrix} \mathrm{dx}_{\mathrm{mask}} \\ \mathrm{dy}_{\mathrm{mask}} \end{pmatrix} + \begin{pmatrix} \mathrm{res}_x \\ \mathrm{res}_y \end{pmatrix} . \tag{4}$$

The parameters (o_x, o_y) denote a translation, which might be caused by bad relative alignment of wafer and mask. The magnification errors during exposure are represented in the parameters m_{xx} and m_{yy}. Rotation of the image during exposure and the loss of orthogonality are described by m_{xy} and m_{yx}. All nonlinear effects are summarised into res_x and res_y.

In practice, the registration functions are not known. One only knows the deviations for certain points - the measurement points or sample points. Hence, one needs to discretize equation (4):

$$m \cdot \begin{pmatrix} \mathbf{dx}^T_{\mathrm{wafer}} \\ \mathbf{dy}^T_{\mathrm{wafer}} \end{pmatrix} = \begin{pmatrix} \hat{o}_x \mathbf{E}^T \\ \hat{o}_y \mathbf{E}^T \end{pmatrix} + \begin{pmatrix} \hat{m}_{xx} & \hat{m}_{xy} \\ \hat{m}_{yx} & \hat{m}_{yy} \end{pmatrix} \begin{pmatrix} \mathbf{dx}^T_{\mathrm{mask}} \\ \mathbf{dy}^T_{\mathrm{mask}} \end{pmatrix} + \begin{pmatrix} \mathbf{res}^T_x \\ \mathbf{res}^T_y \end{pmatrix} \tag{5}$$

with

$$\mathbf{E} = (1\,1\,\ldots\,1)^T \in \mathbb{R}^n . \tag{6}$$

The deviations on the wafer and mask in x- and y-direction at the measurement points are denoted by $\mathbf{dx}_{\mathrm{wafer}} \in \mathbb{R}^n$, $\mathbf{dy}_{\mathrm{wafer}} \in \mathbb{R}^n$ and $\mathbf{dx}_{\mathrm{mask}} \in \mathbb{R}^n$, $\mathbf{dy}_{\mathrm{mask}} \in \mathbb{R}^n$, respectively. Furthermore, $\mathbf{res}_x \in \mathbb{R}^n$, $\mathbf{res}_y \in \mathbb{R}^n$ represent the residuals at the measurement points. Equation (5) can be reformulated into two matrix-vector-equations:

$$\begin{aligned} m \cdot \mathbf{dx}_{\mathrm{wafer}} &= \underbrace{\begin{pmatrix} \mathbf{E} & \mathbf{dx}_{mask} & \mathbf{dy}_{mask} \end{pmatrix}}_{:=\mathbf{A} \in \mathbb{R}^{nx3}} \underbrace{\begin{pmatrix} \hat{o}_x \\ \hat{m}_{xx} \\ \hat{m}_{xy} \end{pmatrix}}_{:=\mathbf{p}_x \in \mathbb{R}^3} + \mathbf{res}_x \\ m \cdot \mathbf{dy}_{\mathrm{wafer}} &= \underbrace{\begin{pmatrix} \mathbf{E} & \mathbf{dx}_{mask} & \mathbf{dy}_{mask} \end{pmatrix}}_{:=\mathbf{A} \in \mathbb{R}^{nx3}} \underbrace{\begin{pmatrix} \hat{o}_x \\ \hat{m}_{xx} \\ \hat{m}_{xy} \end{pmatrix}}_{:=\mathbf{p}_y \in \mathbb{R}^3} + \mathbf{res}_y \; . \end{aligned} \tag{7}$$

This equation should model the transfer of registration or overlay from mask to wafer during exposure. This is equal to finding parameters vectors $\mathbf{p}_x$ and $\mathbf{p}_y$ for which the residuals are minimal:

$$\left\| \begin{pmatrix} \mathbf{res}_x \\ \mathbf{res}_y \end{pmatrix} \right\|_2 \to \min! \, . \tag{8}$$

This approach is known as *least square fit approach* and it fits in with specifying the registration or overlay in terms of variance or standard deviation[†]. With equation (5), one is able to reformulate the least square fit condition (8):

$$\left\| m \cdot \begin{pmatrix} \mathbf{dx}_{\mathrm{wafer}} \\ \mathbf{dy}_{\mathrm{wafer}} \end{pmatrix} - \begin{pmatrix} (\mathbf{A}\mathbf{p}_x) \\ (\mathbf{A}\mathbf{p}_y) \end{pmatrix} \right\|_2 \to \min! \tag{9}$$

[†] When dealing with specifications based on the range, the corresponding formulation is $\left\| \begin{pmatrix} \mathbf{res}_x \\ \mathbf{res}_y \end{pmatrix} \right\|_\infty \to \min!$. Algorithms for these problems are also known in nonlinear optimization.

As one can easily see, equation (9) can be decoupled into two separate problems with half the dimension:

$$\|m \cdot \mathbf{dx}_{\text{wafer}} - \mathbf{A}\mathbf{p}_x\| \quad \rightarrow \quad min! \tag{10}$$

$$\|m \cdot \mathbf{dy}_{\text{wafer}} - \mathbf{A}\mathbf{p}_y\| \quad \rightarrow \quad min! \tag{11}$$

This allows to reduce the discussion of all results in just one direction, the other one follows analogously. There are many algorithms for solving least square problems like equations (9), (10) and (11) in a numerically stable manner. Golub's algorithm was used in the present implementation, which is based on a QR-decomposition.[3]

3. STATISTICAL ANALYSIS OF THE MODEL

In the previous section, an approach was developed which allows to describe the transfer of mask registration onto the wafer. There are certain parameters which can be determined by fitting to a set of mask and wafer data. Still, there are three topics that need to be discussed when the model is applied:

- How well does the model match reality?
- Can the result for the sample points be extended to the whole (pattern) area?
- What are the confidence intervals for the regression parameters $\mathbf{p}_x$ and $\mathbf{p}_y$?

Before these questions are addressed, some statistical terms need to be introduced:

- **Coefficient of Correlation:**
  The coefficient of correlation for two scores X and Y is defined by

  $$\rho_{XY} := \frac{cov(XY)}{\sqrt{\operatorname{var}(X)\operatorname{var}(Y)}} \in [-1;1] \tag{12}$$

  This measure is usually unknown, but it can be estimated by the empirical coefficient of correlation, which is given for the samples x and y by

  $$r_{xy} := \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{13}$$

  with

  $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \,. \tag{14}$$

- **Multiple Correlation:**
  The multiple correlation $\rho(Z,(X,Y))$ is defined as the largest correlation (absolute value) of Z with all linear combinations of (X,Y). The corresponding estimator — called empirical multiple correlation — to the samples x, y and z is given by:

  $$r_{z,(x,y)} := \sqrt{\frac{r_{zx}^2 + r_{zy}^2 - 2r_{xy}r_{xz}r_{yz}}{1 - r_{xy}^2}} \tag{15}$$

- **Shared Variance / Coefficient of Determination:**
  Besides the coefficient of correlation, there is another measure for the quality of a regression: The shared variance or coefficient of determination, which is defined by

  $$R_x^2 = 1 - \frac{s^2(\mathbf{res}_x)}{s^2(\mathbf{dx}_{\text{wafer}})}\,, \tag{16}$$

where the least square fit problem of equation (10) and $s^2(x) = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2$ was used. This measure relates the variance of the residuals with the variance of the input data. If the fit matches the data well, the residual is small and hence its variance. In this case, R_x^2 is approximately 1. If the fit does not explain the data, the variance of the residual is about as big as the variance of the input data and therefore R_x^2 is around 0.
Note, that one can show, that $r^2_{\mathbf{dx}_{\text{wafer}},(\mathbf{dx}_{\text{mask}},\mathbf{dy}_{\text{mask}})} = R_x^2$, please see[4] for more details.

Furthermore, one non-standard distributions is needed:

- **Centered Fisher F-Distribution:**
  The density to the centered F-distribution is defined by

$$f_{m,n}(x) = \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}}\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}, \tag{17}$$

  where n denotes the number of observations, m is the number of parameters and $\Gamma(x) = \int\limits_0^\infty e^{-t}t^{x-1}dt$ represents the Gamma-function.[5] The corresponding distribution is defined by:

$$F_{m,n}(x) = \int_{-\infty}^{x} f_{m,n}(s)ds \tag{18}$$

Moreover, the concept of *statistical hypotheses tests* is required[6]:
First, one needs to state a *null hypothesis* H_0. Here, one states an exact value or values of one or more population parameters. It is temporarily assumed for rejection. The opposite of H_0 is the alternative hypothesis H_1. One hopes to support H_1 with the test. (This way of thinking is quite common in jurisdiction: The defendant is assumed as innocent until the prosecution demonstrates that this assumption is highly unlikely and concludes the defendant to be guilty.)
Furthermore, one needs to choose a *significance level* or *niveau* $\alpha \in [0, 1]$. It represents the threshold level for which H_0 is rejected in favour of H_1. It also gives the α-error, which is the probability of the test to assume H_0 when H_1 is true.
Moreover, one has to find a suitable *decision rule.* It states the value of a *test statistic* at which H_0 is to be rejected or not rejected.
Finally, a sketch of a statistical test can be given:

1. State H_0 and H_1
2. Choose appropriate test statistic
3. Set niveau α
4. Determine decision rule
5. Calculate test statistic
6. Reject or accept H_0
7. If H_0 rejected, assume H_1

For more details, please see[5] or.[6]

Finally, everything is assembled which is needed for a proper investigation of the three questions:

- **How well does the model match reality?**
  This topic can be reformulated: How well is the data explained by the fit? Hence, the multiple coefficient of correlation provides a good measure. Its estimator for a sample is given by equation (15). Another good measure is the shared variance, due to the straight forward interpretation. However, the shared variance does not supply any information about positive or negative correlation. The model applies well if R_x^2 and R_y^2 (or $|r_{\mathbf{dx}_{\text{wafer}},(\mathbf{dx}_{\text{mask}},\mathbf{dy}_{\text{mask}})}|$ and $|r_{\mathbf{dy}_{\text{wafer}},(\mathbf{dx}_{\text{mask}},\mathbf{dy}_{\text{mask}})}|$) is larger than an a priori chosen threshold value. it does not apply if it is below that threshold.

- **Can the result for the sample points be extended to the whole (pattern) area?**
  Equation (5) allows to adjust 6 parameters. If a measurement set consists of only three data points, the residual is equal to zero and hence $R_x^2 = R_y^2 = 1$ no matter what the points are (as long as they do not coincide). However, the information of only three data points is meaningless for the whole area. Therefore, a confidence interval for R_x^2 and R_y^2 is needed.[7] Another, simpler idea is to analyze R^2 when a certain number of points is removed from the original set. This will be discussed further with the example.

- **What are the confidence intervals for the regression parameters $\mathbf{p}_x$ and $\mathbf{p}_y$?**
  One needs to assume that the residuals, e.g. res_x are gaussian distributed with a variance σ^2. An estimator for σ^2 for a regression with three parameters is given for

  $$\bar{s}_x^2 = \frac{1}{n-3}\sum_{i=1}^{n}\left((\mathbf{dx}_{wafer})_i - \hat{o}_x - \hat{m}_{xx}(\mathbf{dx}_{\text{mask}})_i - \hat{m}_{xy}(\mathbf{dy}_{\text{mask}})_i\right)^2 , \tag{19}$$

  where n is the number of sample points. Note, this is a re-scaled empirical variance: Usually the variance has the coefficient $\frac{1}{n-1}$. Based on this assumption and estimator, one can find a hypothesis test for the parameters and furthermore one can derive corresponding confidence intervals[5]:

  $$\begin{array}{ccc} \left[\hat{o}_x - \sqrt{3\mathbf{c}_1\bar{s}_x^2 F_{3,n-3}(1-\alpha)}\right. & ; & \left.\hat{o}_x + \sqrt{3\mathbf{c}_1\bar{s}_x^2 F_{3,n-3}(1-\alpha)}\right] \\ \left[\hat{m}_{xx} - \sqrt{3\mathbf{c}_2\bar{s}_x^2 F_{3,n-3}(1-\alpha)}\right. & ; & \left.\hat{m}_{xx} + \sqrt{3\mathbf{c}_2\bar{s}_x^2 F_{3,n-3}(1-\alpha)}\right] \\ \left[\hat{m}_{xy} - \sqrt{3\mathbf{c}_3\bar{s}_x^2 F_{3,n-3}(1-\alpha)}\right. & ; & \left.\hat{m}_{xy} + \sqrt{3\mathbf{c}_3\bar{s}_x^2 F_{3,n-3}(1-\alpha)}\right] . \end{array} \tag{20}$$

  Here, the vector $\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_3)$ is defined by

  $$\mathbf{c} = \left(\left(\mathbf{A}^T\mathbf{A}\right)^{-1}_{1,1} \quad \left(\mathbf{A}^T\mathbf{A}\right)^{-1}_{2,2} \quad \left(\mathbf{A}^T\mathbf{A}\right)^{-1}_{3,3}\right)^T \tag{21}$$

4. MEASUREMENT ACCURACY

So far, the analysis neglected the fact that the measurement at the sample points could contain some noise:

$$\underbrace{\begin{matrix}\tilde{\mathbf{dx}}_{\text{wafer}}\\ \tilde{\mathbf{dy}}_{\text{wafer}}\\ \tilde{\mathbf{dx}}_{\text{mask}}\\ \tilde{\mathbf{dy}}_{\text{mask}}\end{matrix}}_{\text{measured data}} \quad \begin{matrix}=\\=\\=\\=\end{matrix} \quad \underbrace{\begin{matrix}\mathbf{dx}_{\text{wafer}}\\ \mathbf{dy}_{\text{wafer}}\\ \mathbf{dx}_{\text{mask}}\\ \mathbf{dy}_{\text{mask}}\end{matrix}}_{\text{actual registration or overlay}} \quad \begin{matrix}+\\+\\+\\+\end{matrix} \quad \underbrace{\begin{matrix}\boldsymbol{\delta}_x\\ \boldsymbol{\delta}_y\\ \boldsymbol{\eta}_x\\ \boldsymbol{\eta}_y\end{matrix}}_{\text{measurement error}} \tag{22}$$

where $\boldsymbol{\delta}_x$ and $\boldsymbol{\delta}_y$ denote the measurement errors of the registration measurements on the mask, and $\boldsymbol{\eta}_x$ and $\boldsymbol{\eta}_y$ denote the measurement errors of the registration or overlay measurements on the wafer. Furthermore one can assume that all those errors are gaussian distributed:

$$\begin{array}{lclclcl} \boldsymbol{\delta}_x & \sim & N(0,\sigma_1^2), & \qquad & \boldsymbol{\eta}_x & \sim & N(0,\sigma_2^2) \\ \boldsymbol{\delta}_y & \sim & N(0,\sigma_1^2), & \qquad & \boldsymbol{\eta}_y & \sim & N(0,\sigma_2^2)\,, \end{array} \tag{23}$$

with σ_1^2 denoting the variance of the mask's measurement noise and σ_2^2 representing the variance of the wafer's measurement noise. Note, one can switch to mask-to-mask overlay by multiplying the measurement noise with $\sqrt{2}$ if both masks were measured with the same tool.

Equation (5) is formulated for the actual registration or overlay of mask and wafer. Combined with equation (22), one obtains:

$$\begin{pmatrix} \tilde{\mathrm{dx}}_{\mathrm{wafer}}^T \\ \tilde{\mathrm{dy}}_{\mathrm{wafer}}^T \end{pmatrix} = \frac{1}{m} \begin{pmatrix} \hat{o}_x \mathbf{E}^T \\ \hat{o}_y \mathbf{E}^T \end{pmatrix} + \frac{1}{m} \begin{pmatrix} \hat{m}_{xx} & \hat{m}_{xy} \\ \hat{m}_{yx} & \hat{m}_{yy} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{dx}}_{\mathrm{mask}}^T \\ \tilde{\mathbf{dy}}_{\mathrm{mask}}^T \end{pmatrix} + \underbrace{\left[\begin{pmatrix} \hat{m}_{xx} & \hat{m}_{xy} \\ \hat{m}_{yx} & \hat{m}_{yy} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_x^T \\ \boldsymbol{\eta}_y^T \end{pmatrix} - \begin{pmatrix} \boldsymbol{\delta}_x^T \\ \boldsymbol{\delta}_y^T \end{pmatrix} + \begin{pmatrix} \mathbf{res}_x^T \\ \mathbf{res}_y^T \end{pmatrix} \right]}_{= \begin{pmatrix} \tilde{\mathbf{res}}_x^T \\ \tilde{\mathbf{res}}_y^T \end{pmatrix}} \tag{24}$$

Moreover, one can assume that the measurement history can be neglected. Hence, the errors $\boldsymbol{\delta}_x$, $\boldsymbol{\delta}_y$, $\boldsymbol{\eta}_x$ and $\boldsymbol{\eta}_y$ are independent. Furthermore, the registration or overlay values do not affect the measurement noise. Statistically speaking, $\mathbf{dx}_{\mathrm{wafer}}$, $\mathbf{dy}_{\mathrm{wafer}}$, $\mathbf{dx}_{\mathrm{mask}}$ and $\mathbf{dy}_{\mathrm{mask}}$ are independent from the error terms. In this case, the additional terms do not change the transfer matrix or the offset vector. Among statisticians this model is known as Berkson model and it provides a justification of a least square fit approach for data with measurement noise.[5]

The residuals are no longer caused solely by the model error, as the measurement noise is also part of the residuals in Berkson's model.

5. IMPACT ON DERIVING SPECIFICATIONS

A guideline for deriving wafer overlay and mask registration requirements is given by the ITRS-roadmap,[8] see table 1. As one can see easily, decreasing structure sizes, e.g. MPU/ASIC Metal 1 (M1) halfpitch, lead to stricter overlay requirements. In the past, overlay limits were given by approximately 35% of the MPU/ASIC Metal 1 (M1) halfpitch. Today, this guideline no longer holds: Overlay specifications are tightened to 25%. . . 30% of the MPU/ASIC Metal 1 (M1) halfpitch.

Year	2006	2007	2008	2009	2010
MPU/ASIC Metal 1 (M1) halfpitch (nm)	78	68	59	52	45
Wafer Overlay (3σ,nm)	13	11	10	9	8
Mask Registration (3σ,nm)	8	7	6.1	5.4	4.8

Table 1. Placement accuracy requirements for the future technology nodes according to the ITRS-roadmap.[8]

Furthermore, three basic propositions are needed for deriving the specification guideline:

1. The contribution of the mask to wafer overlay is the mask-to-mask overlay divided by the exposure tool's magnification factor which is typically $m = 4$.
2. The mask is assigned to maximal 30% of the wafer overlay.
3. The overlay error is bounded by twice the allowed registration error. This can easily be proven by applying the absolute value to equation (2) and using the triangular inequality:

   $$|\,\mathrm{ovl}\,| \leq |\,\mathrm{reg}_1\,| + |\,\mathrm{reg}_2\,| \tag{25}$$

   Suppose, the registration limit is the same for all layers, $\mathrm{reg}_1 = \mathrm{reg}_2 = \mathrm{reg}$, one obtains:

   $$|\,\mathrm{ovl}\,| \leq 2|\,\mathrm{reg}\,| \tag{26}$$

Assumptions 1 and 2 allow to derive a guideline for the mask-to-mask overlay limit ovl_m for a given wafer overlay limit ovl_w:

$$\mathrm{ovl}_m = 4 \cdot 0.3 \cdot \mathrm{ovl}_w = 1.2\,\mathrm{ovl}_w \tag{27}$$

Finally, one obtains the guideline by combining equations (26) and (27):

$$\mathrm{reg} = 0.6\,\mathrm{ovl}_w \tag{28}$$

6. APPLICATION TO SET OF MASKS

A set of two product-like masks was printed at AMD Fab30. The mask registration and the wafer overlay measurements were made on the same structures. Figure 1 shows the mask-to-mask overlay which was measured at the AMTC and the wafer overlay which was measured at AMD Fab30. The mask-to-mask overlay is given after the usual six-parameter-model correcting shift, rotation, scale (also known as asymmetric magnification) and orthogonality. The wafer overlay given is corrected by the standard ten-parameter-model[1] and it is averaged over all exposure fields of the wafer. Furthermore, the orientations of the measurements for mask and wafer are usually not the same and they need to be adjusted. Here, the orientation of the mask was used.

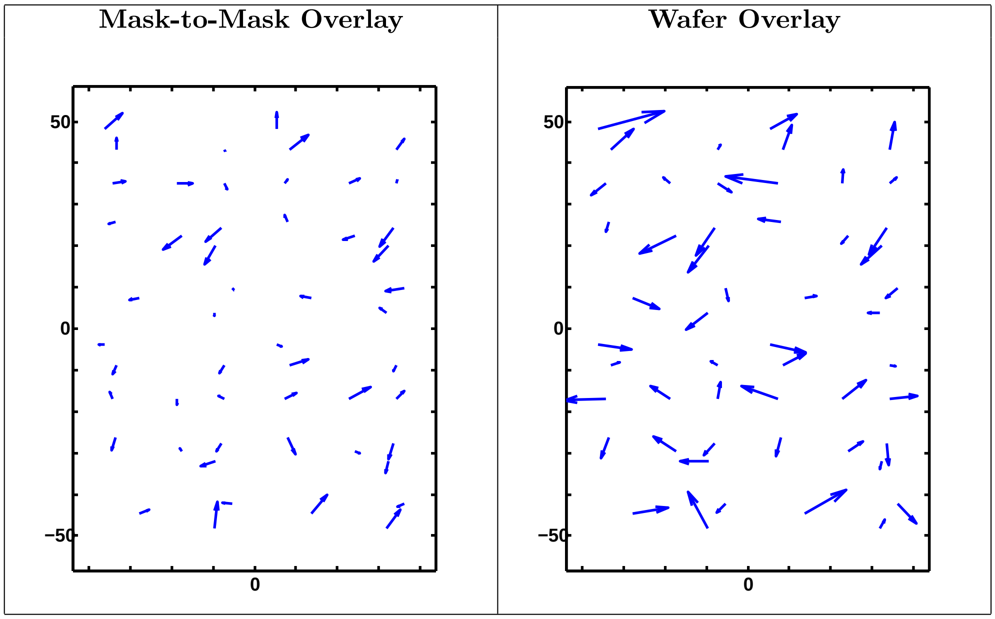

Figure 1. Mask-to-mask overlay of two masks and wafer overlay of their prints. The orientation of the wafer was adjusted to the one of the mask.

Already visually, one can see some correlation between the mask-to-mask overlay and the wafer overlay. If the data is applied to the model given by equation (9), one obtains the parameters given in table 2. The values of the translation vector are virtually zero. Furthermore, the values for $\hat{m}_{xy}$ and $\hat{m}_{yx}$ are significantly smaller than $\hat{m}_{xx}$ and $\hat{m}_{yy}$. The regression according to equation (9) can be visualized easily, see figure 2. This can be interpreted as followed: The wafer overlay in each direction is approximated by a plane depending on the mask-to-mask overlay data.

$\hat{o}_x = -3.52 \cdot 10^{-16}$	$\hat{m}_{xx} = 0.7571$	$\hat{m}_{xy} = 0.4168$
$\hat{o}_y = -1.52 \cdot 10^{-16}$	$\hat{m}_{yx} = 0.2677$	$\hat{m}_{yy} = 1.121$

Table 2. Regression Parameters found for the mask-to-mask overlay and wafer overlay.

It was pointed out that finding a solution to equation (9) is only half the battle. The solution needs to be interpreted properly by the means provided in section 3.

First, the quality of the fit is addressed. It can be measured with the shared variance. For this example, one obtains:

$$R_x^2 = 0.2005 \qquad R_y^2 = 0.6180 \tag{29}$$

These numbers indicate a good correlation of the y-deviations on the wafer with the mask-to-mask overlay data and a poor correlation of the x-deviations on the wafer. One cannot expect R_x^2 or R_y^2 to be very close to one: Nonlinear effects and the measurement noise account for the residuals. In the end, one can state the fit to be of medium quality.

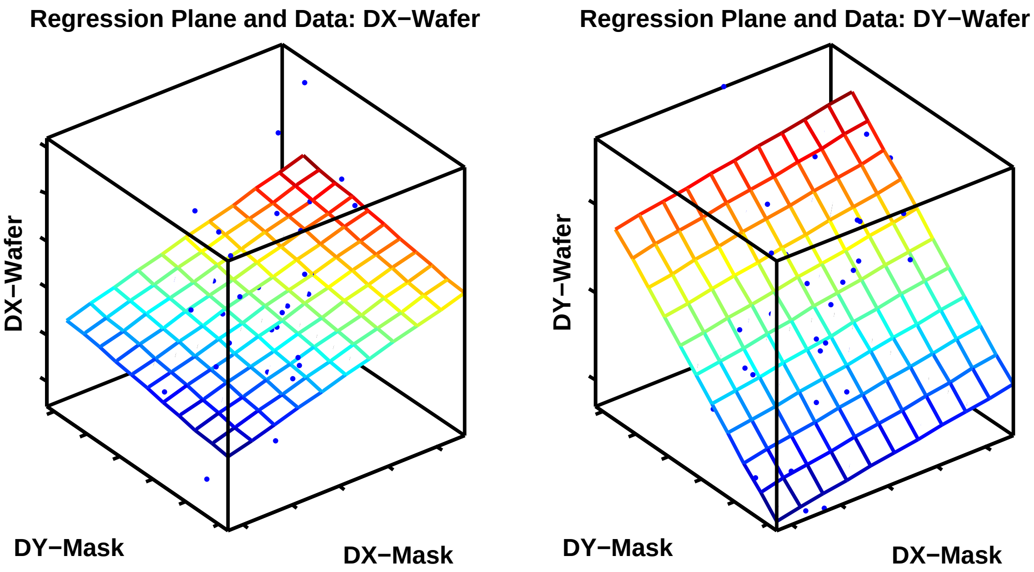

Figure 2. Visualisation of the measured data and the linear regression fit

Subsequently, one needs to check if the residuals are — at least approximately — gaussian distributed. A histogram, as it is given in figure 3, is a first step. Visually, the residuals in the x-direction are closer to a gaussian distribution than the residuals in y-direction.

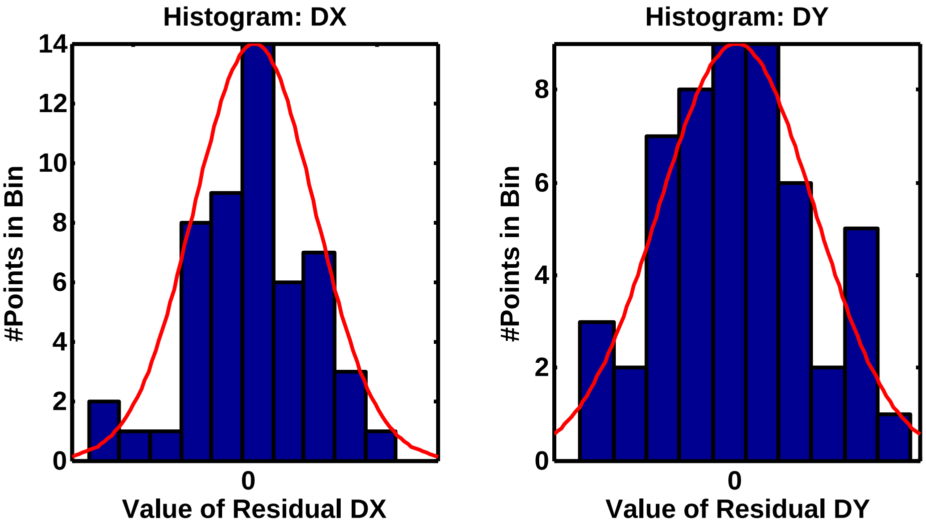

Figure 3. Histograms of the residuals of the fit.

A better way of checking for normality, is given by a QQ-plot, see figure 4. The percentiles of a gaussian distribution are plotted against the percentiles of the input sample. If the sample is gaussian distributed the points make up a straight line.[5] This analysis shows that the residual is approximately gaussian distributed for both directions. Due to additional systematics, one cannot expect a perfect gaussian distribution. However, as a first order approach the normality assumption is valid.

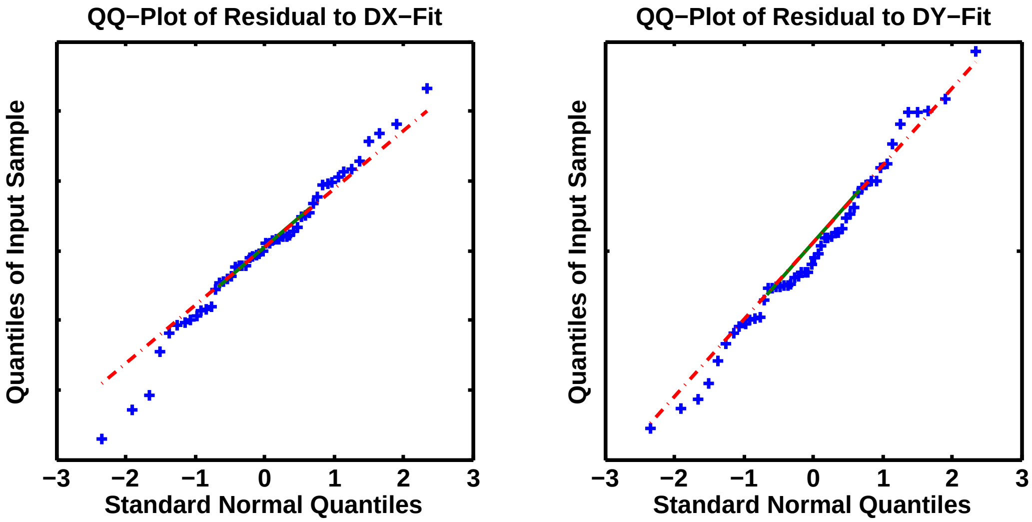

Figure 4. QQ-Plots of the residuals of the fit.

Hence, the confidence intervals of the parameters are given by equation (20). For the niveau $\alpha = 0.05$ one ends up with the confidence intervals given in table 3. Note, the confidence intervals for parameters connected with dx_{wafer} are significantly larger than those for parameters connected with dy_{wafer}. This matches with the values for R_x^2 and R_y^2.

Finally, one needs to address the question of the number of measurement points. In figure 5, a plot of the dependency of the shared variances on the number of sample points is given. At a certain number there is a plateau for which there are only small changes of the shared variances. Hence, one can conclude that enough data points were used.

One can summarize the results found for the example data set as follows:

- The data set consists of enough data points. Hence, the results for the sample points can be extended to the whole pattern area.
- The values for the shared variances indicate that quality of the fit is good enough for a first order approach.
- The confidence intervals indicate that assumption 1 in section 5 also holds as a first order approximation.

Parameter	$\hat{o}_x$	$\hat{o}_y$	$\hat{m}_{xx}$	$\hat{m}_{xy}$	$\hat{m}_{yx}$	$\hat{m}_{yy}$
Value	$-3.52 \cdot 10^{-16}$	$-1.52 \cdot 10^{-16}$	0.7571	0.4168	0.2677	1.121
min. Value	−0.8827	−0.4253	0.3148	−0.256	0.0546	0.9074
max. Value	0.8827	0.4253	1.199	0.8593	0.4808	1.334

Table 3. Confidence intervals for the regression parameters found for the example

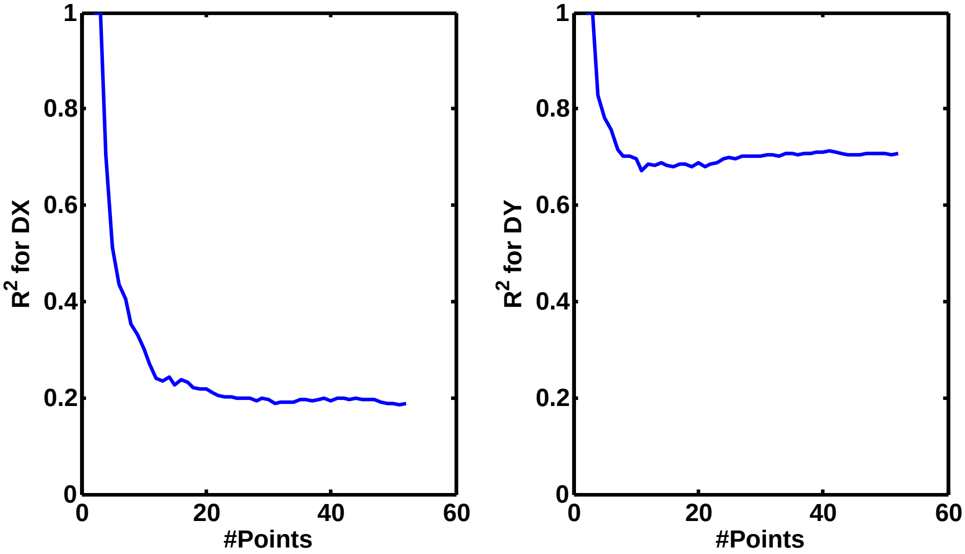

Figure 5. Dependency of the shared variance with the number of measurement points

7. SUMMARY AND CONCLUSIONS

In this paper, a method for analyzing the contribution of mask registration to wafer registration was proposed which can be extended to mask-to-mask overlay and wafer overlay (between to consecutive layers). Furthermore, a thorough statistical analysis including confidence intervals and measuring the quality of the fit was performed. The least square approach can be justified for samples with measurement noise, if one assumes the errors at the points to be independent. This model is known as Berkson's model.

The guidelines used for deriving placement accuracy requirements are based on the fact, that mask-to-mask overlay is transferred with the magnification factor onto the wafer. The example given in this paper shows that this assumption holds for today's lithography processes. Assumptions 1 and 2 of section 5 will be addressed in future work.

Furthermore, mask-to-mask overlay provides a good indicator of wafer overlay performance when the same structures are measured on mask and wafer. This might help in judging the quality of a double-exposure mask set.

Acknowledgements

The authors would like to thank Sven Knoth, Irmingard Eder and Robert Steltzer for fruitful discussions on the statistical topics.

AMTC is a joint venture of AMD, Qimonda/Infineon and Toppan Photomasks and gratefully acknowledges the financial support of the Federal German Ministry of Education and Research (BMBF) under Contract No. 01M3154A ("Abbildungsmethodiken fuer nanoelektrische Bauelemente")

REFERENCES

1. H. J. Levinson, *Principles of Lithography*, 2005.
2. E. Cotte, B. Alles, T. Wandel, G. Antesberger, S. Teuber, M. Vorwerk, A. Frangen, and F. Katzwinkel, "193-nm Immersion Photomask Image Placement in Exposure Tools," in *Optical Microlithography XIX*, D. Flagello, ed., 2006.
3. P. Deuflhard and A. Hohmann, *Numerische Mathematik 1*, 2002.
4. G. Seber, *Linear Regression Analysis*, 1977.
5. J. Hartung, *Statistik*, 1999.
6. J. Wynne, *Learning Statistics*, 1982.
7. S. Chatterjee and B. Price, *Regression Analysis by Example*, 1991.
8. "ITRS-roadmap, 2005 edition." http://www.itrs.net.