

Calibration strategies for precision stages in state-of-the-art registration metrology

Alexander Huebel^a, Uwe Schellhorn^a, Michael Arnz^a, Gerd Klose^a, and Dirk Beyer^b

^aCarl Zeiss SMT AG, Lithography Optics Division (Germany)

Rudolf-Eber-Strasse 2, 73447 Oberkochen

^bCarl Zeiss SMS GmbH (Germany)

Carl-Zeiss-Promenade 10, 07745 Jena

ABSTRACT

State-of-the-art pattern registration tools have to fulfill stringent requirements both in terms of reproducibility and accuracy, as photomask sets are often fabricated in a distributed fashion worldwide and yet have to perfectly match in their respective overlay properties. One option to calibrate the various metrology tools with respect to each other is to utilize a standard, a “golden mask.” Alternatively, “self-calibration” strategies can be employed, which offer certain distinct advantages. This concept of calibration is illustrated by several examples. Merit functions are defined to compare the quality of the calibration procedures, and it is shown how they can be used to optimize the calibration with respect to its efficiency in filtering measurement noise. A symmetry based analysis is introduced to reveal systematic weaknesses of potential calibration sequences, also indicating the necessary steps to overcome these problems. An extension of the theory is given that allows to suppress a certain class of systematic errors, and it is studied under which conditions it can be applied.

Keywords: registration, pattern placement, metrology, self-calibration

1. INTRODUCTION

Future lithography trends pose new challenges to the metrology equipment used in the semiconductor industry. For example, photomasks have to fulfill even tighter specifications as pattern placement errors can have direct effects on the final device yield. In order to keep pace with this development, measurement systems like PROVE, the state-of-the-art registration tool currently under development at Carl Zeiss [1][2][3][4], must be able to measure the position of photomask patterns with an unprecedented precision – not only in terms of reproducibility, but also accuracy. As the production of mask sets can be distributed across several sites, different registration tools have to match each other in an absolute sense. Furthermore, even with the latest manufacturing capabilities, assembled stage mirrors will not be perfectly flat to the required sub-nm level. In order to both correct for the individual stage inaccuracies and to match different tools to a given standard, a calibration strategy has to be employed.

One approach is to use a so-called “golden mask,” which could be a photomask with a precisely known pattern. When such a mask is taken as a standard, it is indeed possible to match various registration tools by measuring this mask and storing the observed deviations as calibration data for later corrections. This simple strategy, however, suffers from various drawbacks. The most obvious one stems from the fact that no “golden mask” has a ‘precise enough known geometry.’ There simply does not exist any reference tool which can be employed to determine the pattern geometry any better than the most advanced registration tools – exactly the tools that should be calibrated with this mask. In terms of absolute accuracy, a “golden mask” cannot resolve this dilemma: the accuracy of the calibration is limited by the accuracy of the mask! Only by arbitrarily defining a mask as ‘perfect’ without errors one can circumvent the issue. Besides this accuracy problem, there are other, very practical disadvantages of this “golden mask” approach. In order to match separated registration tools, one has to transfer the standard mask from one tool to the other with the associated risk of damage and the limited availability upon a calibration demand. One could also attempt to duplicate the standard mask, but this cannot be done precisely enough for the latest registration requirements. Instead, one typically measures the “golden mask” in a dedicated registration tool, hence transferring the standard to a “golden tool.” With this tool

multiple photomasks can be measured, again transferring the standard to these other masks, which can be sent around then. The transfer of standards, the so-called traceability, however, suffers from the measurement errors and the calibration inaccuracies of the tool used. One last disadvantage to be mentioned is the required stability of the standard masks. If such a photomask is degrading over time, the calibration will vary, which can only be detected by comparing the mask regularly with other standards.

Alternatively to the “golden mask” approach, a so-called “self-calibration” technique can be used. In that case, the metrology tool itself performs several measurements on a sample mask in various possible ways. Under the assumption that the sample has not changed over the short time of the measurements, the results should be identical within the limits of the measurement precision. Any significant deviation constitutes a signature of the tool inaccuracy. By designing a clever sequence of ‘redundant’ sample measurements, most of the tool’s accuracy errors can be detected and further used as correction data. This self-calibration approach provides several advantages and circumvents potential problems with a “golden mask.” The accuracy of the method does not depend on the accuracy of a specific mask, but is based only on the sequence of redundant measurements and the limitations on the precision of a single measurement (as in the “golden mask” approach). This has the welcome benefit that a more precise measurement automatically results in a more accurate tool. The requirement on the sample mask is merely to be stable during the time of the calibration. This is also a demand for the tool, which has to be fulfilled for practical purposes anyhow. Furthermore, it is not necessary to match one calibration mask in one tool to other masks in other tools, hence a calibration can be performed autonomously and anytime on the individual tool. One potential drawback of this approach is that a suitable calibration sequence can entail many measurements, hence require more tool time to perform. Depending on what property has to be calibrated, it is viable to tailor the sequence to the specific needs and thus potentially reduce the overall effort.

Comparing the two approaches, it is apparent that the self-calibration method offers distinct advantages over the use of “golden masks.” Hence it is of no surprise that this type of calibration becomes more and more widespread among advanced metrology tools. The following chapters will explain the principles of self-calibration in more depth, and address specifically the task of how to decide on and qualify a well-suited calibration sequence. One issue, however, still has to be dealt with a “golden mask” approach, and this is scale. The length measurement on an absolute level is a separate topic where many institutions have worked extensively to come up with solutions. Any registration tool deals with this by requiring a traceable length standard, which is commercially available, or might even be customer specific.

2. PRINCIPLES OF SELF-CALIBRATION

The idea behind self-calibration can be intuitively understood with the following simple, but characteristic example: Assume a 10-by-10 grid of markers on a photomask to be measured with an uncalibrated registration tool. By design, the marker grid should be strictly Cartesian. The result is supposed to look like in Fig. 1a, where the measured grid is no longer square, but sheared by a certain angle. Without additional information, it is not possible to decide whether the observed pattern indicates a pattern placement error on the mask, or whether it is an artefact of the registration tool whose axes are not orthogonal to the required accuracy. A possible way to distinguish these two cases is to perform a consistency check with a second measurement. This time the photomask is rotated in the tool by 90° (Fig. 1b). If the measured pattern also turns by 90 degrees as illustrated in Fig. 1b, it is correct to assign the registration error as a writing error of the mask. However, if the measurement result remains unchanged, one must conclude that the observed error is due to a limitation in the accuracy of the metrology tool. In a more general case, the measurement might result in a linear combination of both patterns. Nonetheless, the procedure with the two measurements would still serve to separate the device contribution from the writing error on the mask.

This example illustrates well how consistency checks on multiple measurements of a single mask under different tool conditions can serve to differentiate real writing errors on the mask from systematic measurement errors of the tool. However, if the observed pattern looks like in Fig. 2a or 2b, a rotation of the mask does not help since the pattern is invariant with respect to mask rotation – the rotated result looks alike. Instead, one could attempt to measure the pattern a second time after shifting the mask sideways, because the observed pattern then looks noticeably different. Not every shift of the mask is always helpful. A shift-invariant error-pattern like in Fig. 2c can not be resolved by a shift in the vertical direction. Already these few thoughts reveal that the success of a self-calibration depends on the choice of the measurement sequence, and that symmetry arguments play an important role in deciding on a well-suited sequence.

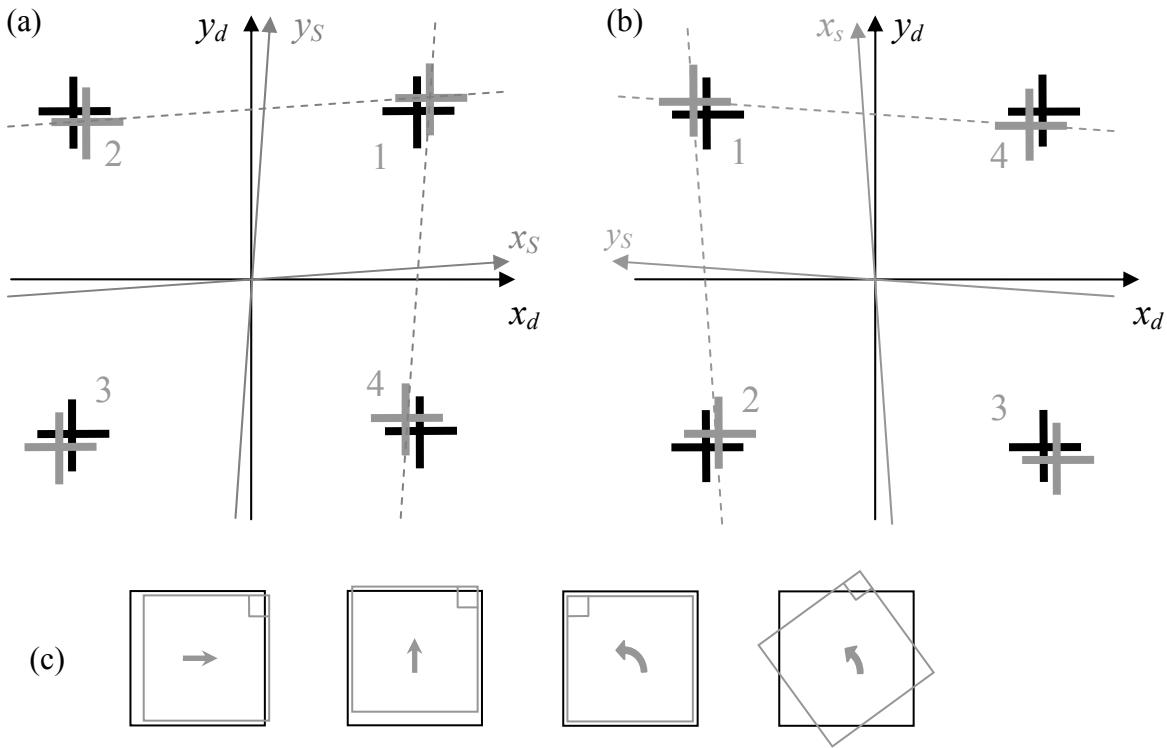


Fig. 1: Consistency check on a measurement with an observed orthogonality error: (a), the hypothetical measurement result (black) in comparison with the expected pattern positions (grey) for the unrotated specimen and the coordinate systems (x_d, y_d) for the device (stage) positions, and (x_s, y_s) on the surface of the mask. Part (b) shows a possible outcome after rotating the mask by 90° . Since the placement errors have rotated along with the mask, a writing error on the mask has been detected. (c), pictograms for different possible positions of the mask (grey, with marked upper right corner) on the stage (black): x-translation, y-translation, grid-symmetrical rotation by 90° , and rotation by an arbitrary angle.

Not worrying about a possible implementation in a real metrology tool, one can think of many variations on how to measure one and the same mask in different ways. Fig. 1c illustrates just a few. Besides rotating the mask by 90° or multiples thereof, one could also shift the mask into any direction, potentially flip it around or upside down [5], or even rotate it by an arbitrary angle. In addition one could use more than one mask, and e.g. rotate a smaller mask off-axis [6]. Since the accuracy error of the tool is typically not known in advance – this is why the calibration is needed in the first place – a combination of these measurements has to be chosen in such a way that any imaginable pattern error can be detected and separated in its mask and tool contributions. It is therefore necessary to combine the information from both rotations and translations to get a full understanding of the measurement device, and look at all possible error patterns, i.e. a complete vector field description. Fortunately, there is a systematic mathematical approach to this calibration problem. While it is possible to analyse calibration sequences even on irregular grids and with non-commensurable positions upon translation or rotation, this paper limits itself to commensurable calibrations on a Cartesian grid.

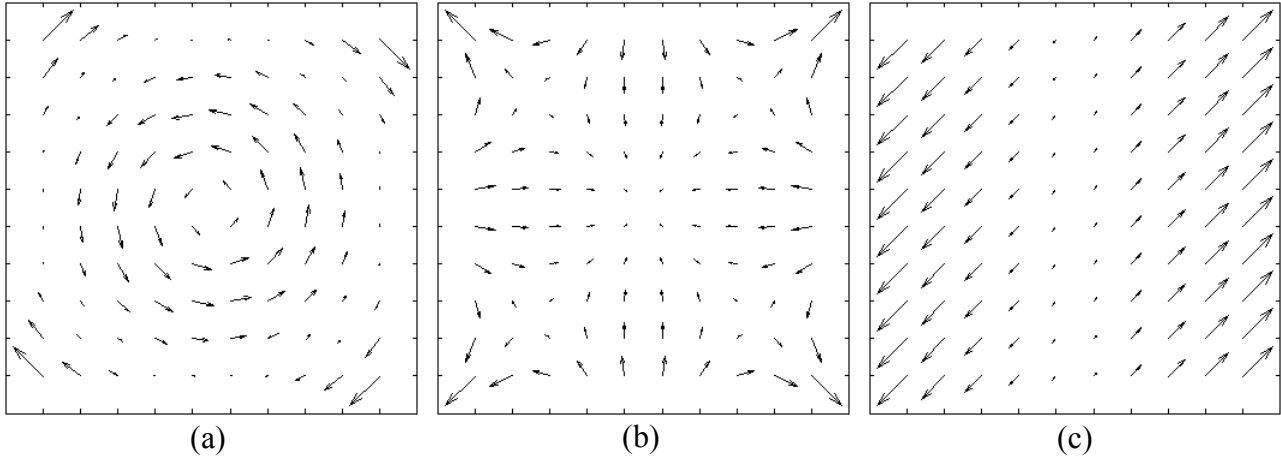


Fig.2. Three examples of symmetrical vector fields on a 10-by-10 grid, corresponding to possible real pattern placement errors or artificial measurement errors of the tool. (a) shows a rotationally symmetrical field, (b) a field that is invariant under the operations of the symmetry group C_{4v} of the square, in particular under rotations by multiples of 90 degrees. In (c), a vector field is shown that has translational symmetry with respect to the vertical direction.

3. MATHEMATICAL BACKGROUND

A rigorous mathematical treatment of the calibration problem starts from the idealized assumption that the measurement results in all possible positions of the mask can be written as a linear superposition of two vector fields, one describing the device/tool error and the other one the pattern positions on the mask. The field of the tool error is constant in all positions, whereas the mask field undergoes rigid transformations whose parameters (rotation angles and translation vectors) are approximately known. For the purpose of this paper, only grid-symmetrical calibration problems are described, where the two vector fields are defined on a Cartesian n -by- n grid, and the rigid transformations remain grid-symmetrical to within the required accuracy [7][8]. Under these circumstances, each possible transformation induces a permutation of the grid points on the mask with respect to the fixed tool grid. For the $4 \cdot n^2$ unknown vector components of the two fields of the calibration problem, a corresponding linear equation system can be written down:

$$\begin{pmatrix} \underline{m}_x \\ \underline{m}_y \end{pmatrix} = \begin{pmatrix} \underline{1} & \underline{0} & \underline{T}_{xx} & \underline{T}_{xy} \\ \underline{0} & \underline{1} & \underline{T}_{yx} & \underline{T}_{yy} \end{pmatrix} \cdot \begin{pmatrix} \underline{d}_x \\ \underline{d}_y \\ \underline{s}_x \\ \underline{s}_y \end{pmatrix} \quad (1)$$

The n^2 -vectors \underline{d}_x and \underline{d}_y are the two components of the tool error on the n -by- n grid points, likewise \underline{s}_x and \underline{s}_y for the pattern positions on the sample mask. The field $(\underline{d}_x, \underline{d}_y)$ is given in a coordinate system (x_d, y_d) fixed to the tool, whereas $(\underline{s}_x, \underline{s}_y)$ relate to a coordinate system (x_s, y_s) fixed to the mask (Fig. 1). On the left-hand side of Eq. 1, the x - and y -components of all measured data points in all positions of the mask are contained in the column vectors \underline{m}_x and \underline{m}_y , respectively. In the matrix that maps the unknowns onto the measurement values, the blocks denoted by T_{ij} contain permutation matrices describing the rigid transformations of the specimen. For translations and the 180° rotation, these amount to a pure (signed) permutation of the grid points, whereas the off-diagonal blocks T_{xy} and T_{yx} are needed to describe rotations by 90° or 270° .

The matrix in Eq. 1 is generally not square, but rather represents an overdetermined system of equations, reflecting the redundancy of the measurement sequence. The mathematical problem can be solved by a Gaussian least-squares

optimization: for every set of measurement data given on the left-hand side, there exists a set of parameters ($\underline{d}_x, \underline{d}_y, \underline{s}_x, \underline{s}_y$) that minimizes the sum of squared deviations between the right- and left-hand side. The remaining deviations, i.e. the residual error, are a measure of the consistency of the assumptions with the measurement result, and therefore automatically constitute an error estimate for the calibration map ($\underline{d}_x, \underline{d}_y$).

In reality, Eq. 1 is not yet sufficient since the mask positions are not known well enough to sub-nm accuracy. Moreover, as long as there is no valid calibration, the device itself cannot be used to measure the positions of the specimen. The only way out of this dilemma consists in disregarding parts of the measured information: in each measurement, the three vector fields corresponding to x-translation, y-translation, and rotation (the *primary disturbances*, Fig. 3) must be allowed to take on arbitrary values. This restriction can be mathematically included by introducing additional Lagrange variables into the solution vector of Eq. 1 [9]:

$$\begin{pmatrix} \underline{m}_x \\ \underline{m}_y \end{pmatrix} = \begin{pmatrix} \underline{1} & \underline{0} & \underline{T} & \underline{T} & \underline{C} \\ \underline{0} & \underline{1} & \underline{T} & \underline{T} & \underline{C} \end{pmatrix} \cdot \begin{pmatrix} \underline{d}_x \\ \underline{d}_y \\ \underline{s}_x \\ \underline{s}_y \\ \underline{c} \end{pmatrix} \quad (2)$$

The parameter vector \underline{c} contains three additional parameters for each position, and the corresponding x- and y-components of the primary disturbances are abbreviated by C_x and C_y , respectively.

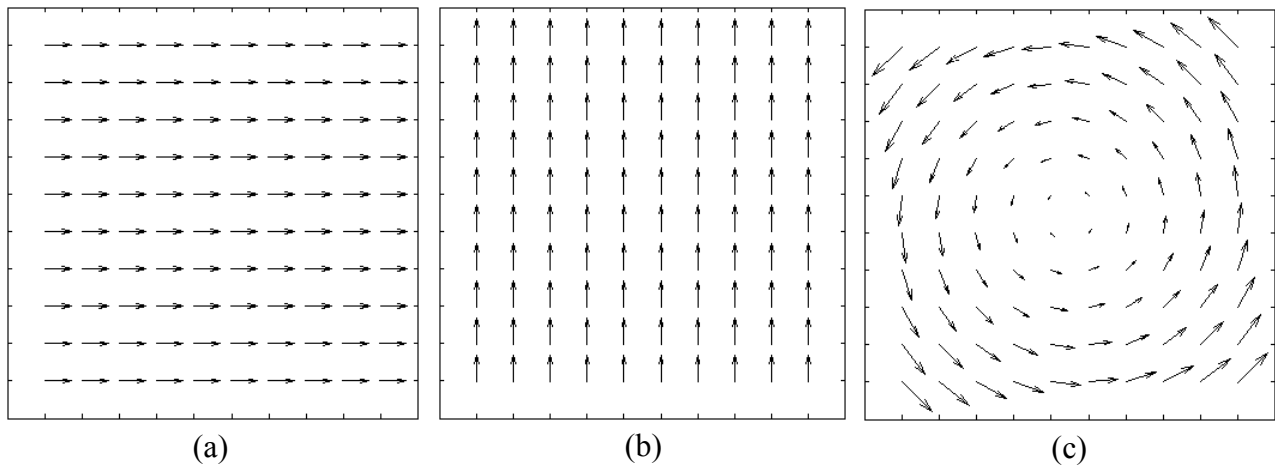


Fig. 3: The three vector fields corresponding to x-translation (a), y-translation (b), and rotation (c).

At first glance it might seem that the loss of only four pieces of information per position could be compensated by the redundancy of the measurement sequence. However, the exclusion of these three basic error fields always reduces the rank of the matrix in Eq. 2 regardless of the chosen calibration sequence. As a consequence, translations, rotation and scale of the calibration map ($\underline{d}_x, \underline{d}_y$) and of the pattern positions ($\underline{s}_x, \underline{s}_y$) will remain undetermined, confirming our physical intuition that the absolute position and orientation of the coordinate systems (x_d, y_d) and (x_s, y_s) are not fixed, and the absolute scale can only be measured with an external reference (“golden mask”).

4. STATISTICS AND STABILITY OF SQUARE-GRID CALIBRATIONS

The theoretical considerations above have shown that in the case of ideal measurements, the solution of Eq. 2 is unique for the $2 \cdot (2n^2 - 4)$ relevant unknowns and independent of the chosen calibration sequence - as long as the sequence contains enough information (condition of completeness). In reality however, one always has to deal with measurement errors. Since not every calibration sequence responds to these errors in the same way, this section concentrates on the design of a well-suited measurement sequence in the presence of stochastic noise; systematic errors will be discussed in section 5.

As a test case, we study the calibration of a 20-by-20 grid. A simple calibration sequence with two mask shifts and one additional rotation is illustrated in Fig. 4a. The rotation by 90° or 270° has to be included in order to fulfill the requirement of completeness (see above) - it is the only possibility to compare the parameters d_x with the parameters d_y . The two short translations by one grid constant ensure that the Nyquist criterion is not violated, so the full spatial resolution of the grid can be used. This calibration sequence was tested using a Monte-Carlo simulation with 10000 runs, where the measurement of each grid point suffered from white noise, i.e. a Gaussian error distribution without systematic bias was assumed for all grid points and vector components. In order to quantify the quality of the calibration, a merit function f_{loc} is defined which describes the local noise transfer at each grid point - the ratio of the standard deviation of the calibration result d_x (or d_y) at this point, divided by the standard deviation of the measurement noise.

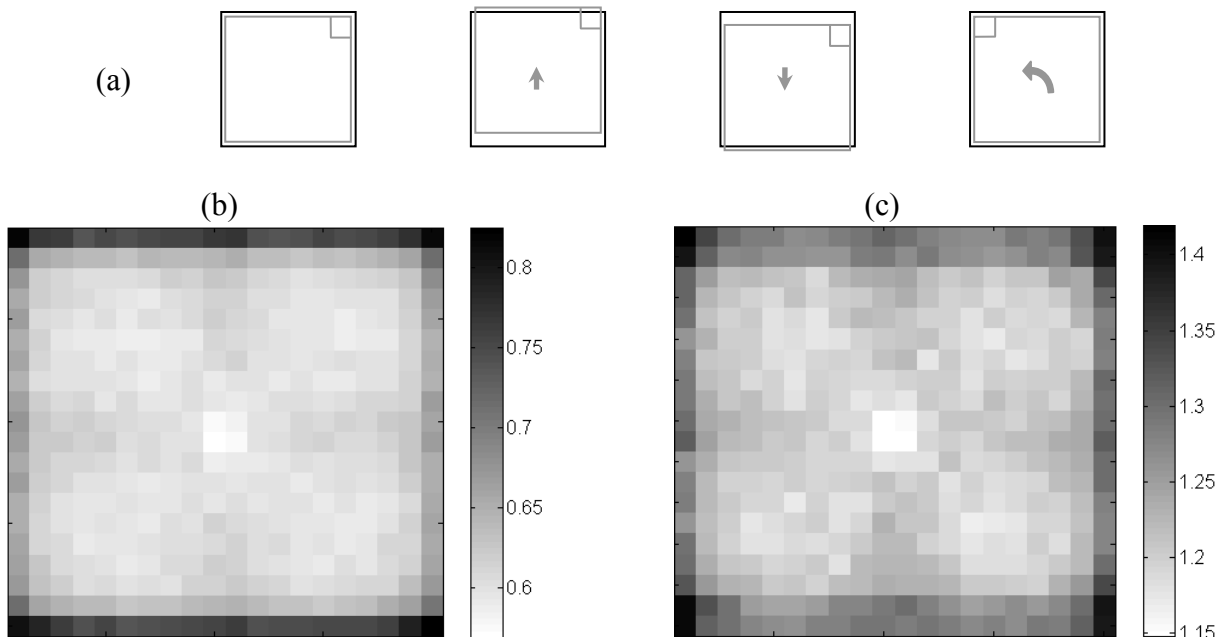


Fig. 4: A simple calibration sequence consisting of two short translations by one grid constant along opposite directions and a rotation by 90° (a). Part (b) shows a greyscale plot of the white noise transfer function f_{loc} for this sequence, calculated from 10.000 runs. The normalized noise transfer function f_{norm} is plotted in (c).

Fig. 4b shows this *spatial noise transfer function* f_{loc} for the sample sequence of Fig. 4a, obtained by the Monte-Carlo simulation. Only the noise transfer for d_x is plotted since d_y exhibits the identical amount of noise by symmetry. A weakness of this sequence becomes apparent on the upper and lower edges, where the error propagation turns out to be less favourable than in the centre. Theoretically, the standard deviation at each grid point (i, j) could reach a value of $1/\sqrt{n_{ij}}$, where n_{ij} is the number of individual measurements on this point. At the upper and lower edges, the grid points participate in only three individual measurements (instead of four in the centre), so their standard deviation could reach

at best $1/\sqrt{3} \approx 0.58$ compared to 0.5 in the centre. The absolute numbers could be further reduced by simply measuring the sequence more often. In order to compare the systematic quality of various calibration sequences, it is beneficial to take out these purely statistical effects. This suggests to introduce an additional *normalized noise transfer function* of the calibration sequence (Fig. 4c), defined as

$$f_{norm}(i, j) = \frac{f_{loc}(i, j)}{\sqrt{n_{ij}}} \quad (3)$$

This function measures locally how efficient a certain calibration sequence is in statistically averaging the measured data. For the sample sequence of Fig. 4a, the optimum efficiency of 1 is approached only in the very centre, being between 20 % and 30 % worse for the rest, and deviations reach beyond 40 % near the corners.

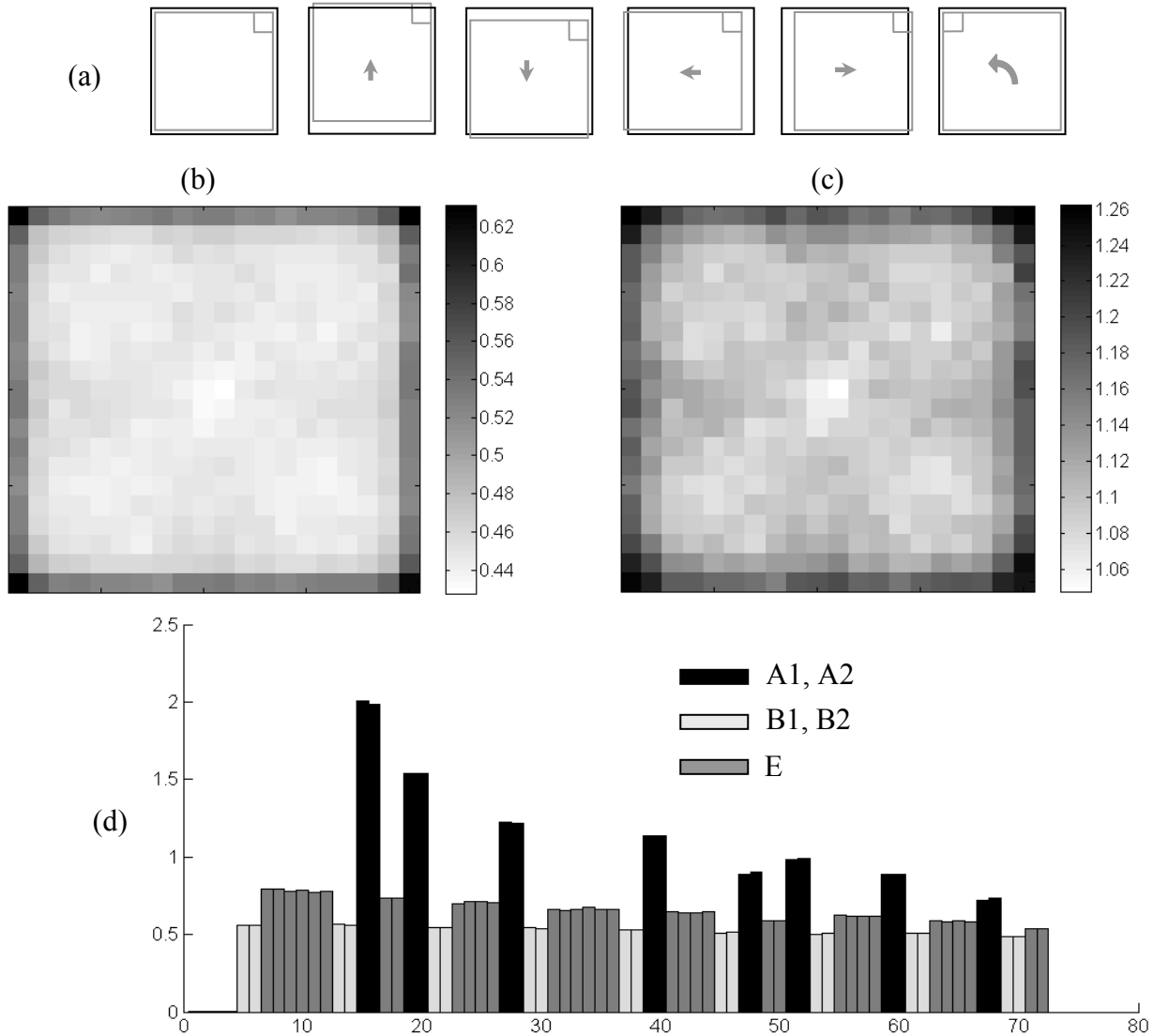


Fig. 5: A symmetrical calibration sequence with short translations by one grid constant in all four directions and a rotation by 90° (a), and its white noise transfer function f_{loc} (b). Part (d) shows the noise transfer in the basis of vectorial Legendre polynomials up to 5th order. The first four vector polynomials correspond to the non-determinable patterns of translation, rotation, and scale and they have been set to zero. A clear dependence of the noise transfer on symmetry can be recognized. The normalized version f_{norm} of the transfer function (c) is relevant for comparison with the sequence of Fig. 4 in terms of statistical averaging efficiency.

In order to improve these results, one could try to employ the more symmetrical calibration sequence illustrated in Fig. 5a. It contains all mask positions of the first sequence, but has been extended by two horizontal translations to equivalently cover all four directions. One would hope that the edges are better conditioned, since each edge is missing in only one out of six mask positions. Because of the sixfold averaging, an optimal noise transfer factor of 0.41 should be reached in the centre. The results of the simulation are shown in Fig. 5b and 5c. The absolute noise transfer has been reduced as expected, but still the edges and moreover the corners reveal weak spots. The normalized plot shows that 10% statistical efficiency is lost at most of the points, and the edges are not much better in this regard compared to the first sequence (Fig. 4b,c). Comparing the two examples, one can conclude that by adding new positions to a calibration sequence, though an improvement of statistical averaging can be accomplished, it is much harder to improve the efficiency of the averaging, i.e. the inherent stability of the calibration.

To understand where the statistically unexpected noise stems from, it helps to reconsider that the noise in the calibration is no longer homogeneously distributed across the grid. Since the mathematical description of the problem is linear, such a behaviour can only be explained with the occurrence of spatial correlations between different points. This can be visualized by describing the Monte-Carlo results in a functional representation. For this purpose, the spatial noise transfer is calculated in an orthonormal basis of vectorial Legendre polynomials. Fig. 5d shows a noise plot for the corresponding coefficients, whose standard deviations are taken over the 10000 runs. Certain functions or spatial patterns clearly suffer from more noise than others, which is a signature of the predicted spatial correlations. The description of the noise transfer in terms of vectorial Legendre polynomials has the additional advantage to indicate which types of patterns are most sensitive. The polynomials can be characterized by the five possible symmetry properties corresponding to the irreducible representations of the group C_{4v} , the symmetry group of the square (greyscale code in Fig. 5d). The 'black' polynomials belonging to the A_1 and A_2 representations of fourfold rotational symmetry exhibit the highest sensitivities to white measurement noise. The transfer function looks fairly uniform for all other functions, those of types B_1 and B_2 (light grey) being slightly better. These differences tend to wash out for the short-ranged polynomials of higher order.

A significant improvement of the noise transfer uniformity and therefore of the statistical efficiency of the calibration requires additional mask positions that specifically address the black bars. These mask positions have to break the symmetry of the A_1 and A_2 patterns, i.e. additional rotations do not help, but more translations with larger shifts would be beneficial. These considerations lead to the calibration sequence illustrated in Fig. 6a, which indeed yields improved noise transfer properties (Fig. 6b, c). Especially the functional analysis of Fig. 6d reveals the improvement concerning the previously limiting rotationally or fourfold symmetric patterns (black). The statistically determined theoretical limit is achieved over a much larger central area than with the simpler calibration sequences. The price for this inherently better sequence is an increased calibration effort with ten mask positions compared to the four which were initially used.

While it is also possible to reduce the noise transfer by simply increasing the redundancy of a few mask positions, or even selectively repeating measurements only at critical grid points (e.g. near the edges), this additional effort does not address the systematic limitations of the simpler calibration sequences. Only the deliberate addition of mask positions that help with the specific symmetry patterns (here scale-like and fourfold) can improve the calibration efficiency and empower the procedure to approach its statistical possibilities. A well-balanced behaviour of a calibration sequence becomes even more important when not only statistical noise, but also systematic errors come into play. These cannot be averaged out by multiple measurements, but enter directly into the result. When a component of the solution (a specific pattern) is unstable against white noise, it will in general also be unstable against systematic disturbances.

There is one more option to improve the noise transfer of a calibration sequence, which should be mentioned for the sake of completeness, even though in reality its applicability is often very limited. The previous examples showed how simpler calibration sequences were not suited to detect certain patterns. If on the other hand, one can make reasonable assumptions that the tool to be calibrated does not suffer from certain error patterns, the calibration analysis can intentionally filter out these patterns. One could e.g. utilize only the lowest order Legendre polynomials, thus reducing the degrees of freedom in the solution and correspondingly improving the noise susceptibility. However, if the tool should exhibit an error which cannot be described by the limited Legendre basis, this error will pass undetected and will not be calibrated.

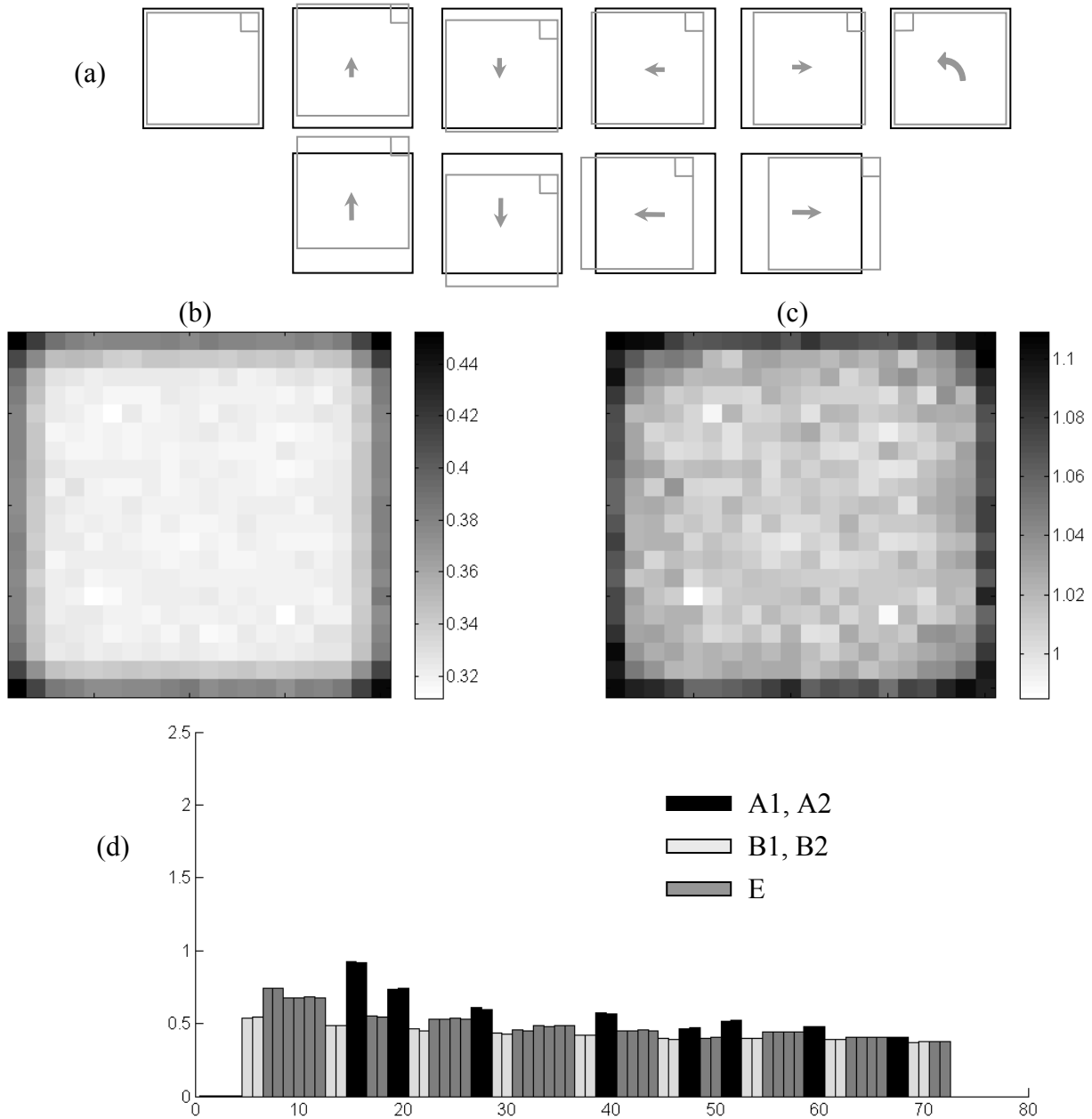


Fig. 6: The calibration sequence of Fig. 5, extended by four longer translations by two grid constants (a), together with its unnormalized (b) and normalized (c) noise transfer functions. The noise transfer uniformity is close to the theoretical optimum, also visible in the vectorial Legendre decomposition (d) where the systematic instabilities have almost disappeared.

5. FILTERING OF SYSTEMATICAL ERRORS

Since systematic errors usually do not average out, the first and most important strategy must be to avoid them by design. However, since there are technical limitations, the calibration concept has to be able to cope with these remaining systematic effects. Certain systematic tool errors can be described as a linear error contribution with respect to known tool or measurement parameters. This type of problems can be solved with an extension of the mathematics introduced before.

To illustrate the procedure with an example, the tool should suffer from a small rotation of the coordinate axes of the camera image (field of view) relative to the coordinate axes of the stage movement (Fig. 7). This systematic inaccuracy does not introduce an additional error if all marks are measured exactly in the center of the field. However, if the marks are not measured in the center, e.g. because their positions are not known accurately enough beforehand, there will be a measurement error proportional to their distances from the center of the field. In a first order approximation for small rotations the measurement errors of the x- and y-components are linearly given by

$$\begin{aligned}\delta x &= \varphi \cdot b \\ \delta y &= -\varphi \cdot a\end{aligned},$$

where φ is the rotation angle and (a, b) are the image coordinates of the mark in the field of view. When the centering errors (a, b) stem from insufficient knowledge of the pattern positions on the mask, the errors δx and δy will transform along with the mask positions. Such a transformation behavior is consistent with the assumptions of the calibration model, so δx and δy do not augment the *residual* calibration error – instead, it appears as if each mark on the mask has an additional writing error $(\delta x, \delta y)$ in the pattern positions $(\underline{s}_x, \underline{s}_y)$. The actual calibration result $(\underline{d}_x, \underline{d}_y)$, however, is not affected.

The situation changes when the centering problem comes from insufficient knowledge about the positioning of the stage (e.g. an uncalibrated stage): in this case, the error $(\delta x, \delta y)$ depends only on the stage position and not the mask position or orientation, i.e. it is constant for each point of the stage position grid. The error enters then into the calibration map $(\underline{d}_x, \underline{d}_y)$. A device calibrated this way could only measure properly on the image positions used for the calibration, but once the marks are to be measured at different positions a systematic error would appear without noticing. In a general case, the tool inaccuracy in combination with non-centered marks leads to measurement errors which can result in a wrong calibration.

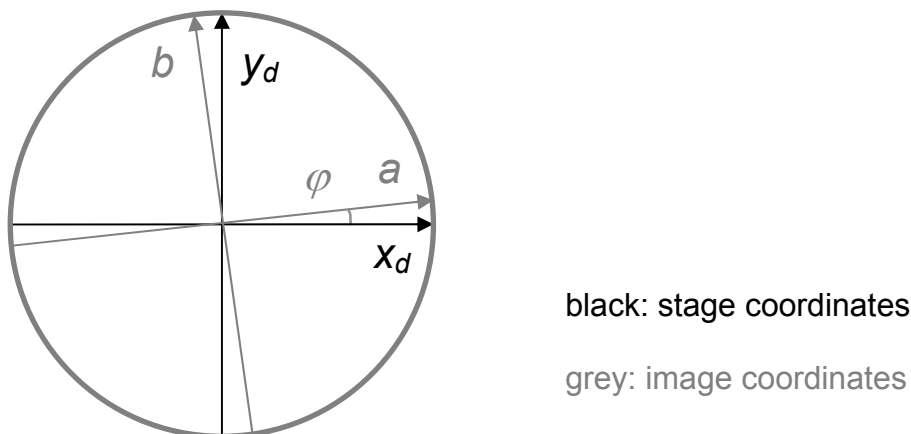


Fig. 7: A rotation of the coordinate axes of the camera image (a, b) relative to the axes of the stage movement (x_d, y_d)

To illustrate the possible effect of a systematic tool error on the calibration results, this example is applied to the calibration sequence shown in Fig. 5. In all the mask positions, the patterns are measured in the center, except for the rotated orientation. For some reason, with that mask orientation the stage should exhibit positioning problems in such a way that the patterns at $y_d > 0$ are assumed to be offset sideways with $a = 5 \mu\text{m}$ and those at $y_d < 0$ with $a = -5 \mu\text{m}$. The systematic image rotation is assumed to be $\varphi = 10 \mu\text{rad}$, but no further errors should be present. The image rotation in combination with the uncentered marks introduces unexpected errors into the measured data. These simulated data are now evaluated by the calibration sequence of Fig. 5. Since no other tool errors or statistical noise are included in the data, a proper calibration should yield vanishing tool error and residuals. Due to the systematic image rotation, however, the calibration leads to the plots in Fig. 8. The calibration map shown on the left side (a) exhibits a faulty non-zero correction field. Furthermore, the residual of the data on the left-hand side of the calibration equation also exhibits a non-zero regular pattern (b). This is a clear signature that the mathematical model used (Eq. 2) is no longer suitable to properly describe the measured data. Even worse, while the residual indicates a problem in modelling the system, the pattern cannot be used to determine the rotation angle φ .

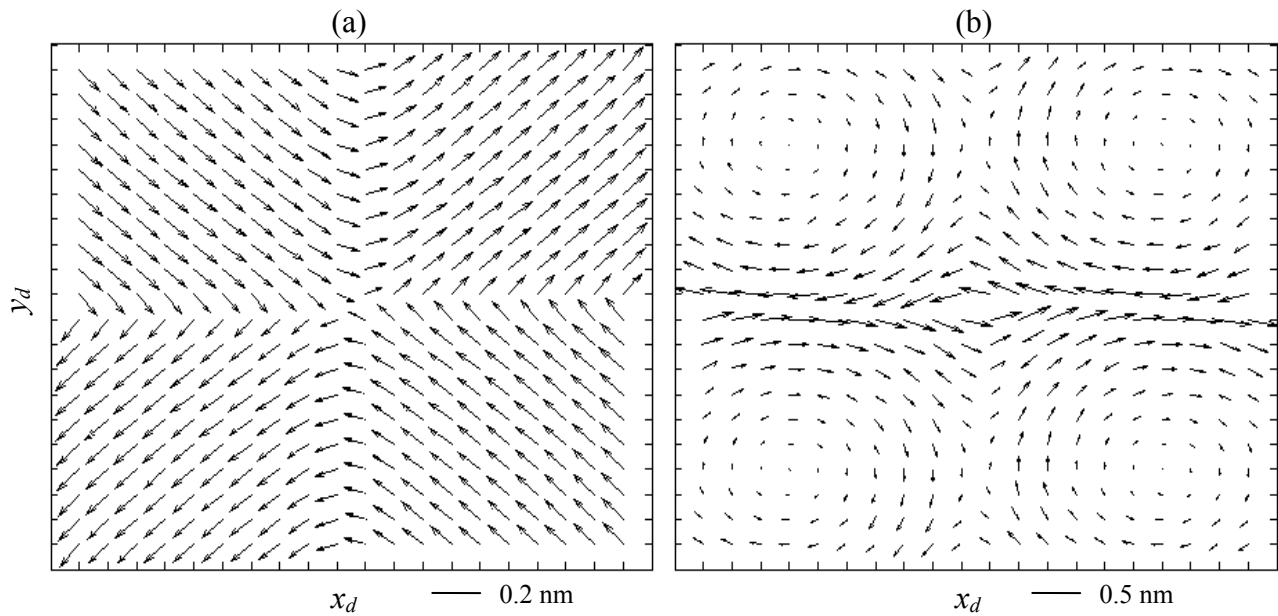


Fig. 8: Influence of an image rotation of $10 \mu\text{rad}$ on the calibration sequence of Fig. 5. The patterns are assumed to be measured at $a = +5 \mu\text{m}$ in the positive half-plane and at $a = -5 \mu\text{m}$ in the negative half-plane in the rotated position. (a) shows the change in the calibration map ($\underline{d}_x, \underline{d}_y$) and (b) the residual error of the rotated position.

In order to successfully deal with the systematic error in this example, the mathematics of Eq. 2 has to be extended - we suspect an image rotation to be present in the tool, but do not know its value φ . The strategy to correct for such a linear error term is to include the unknown (φ) as a new component of the solution vector. The observed image positions (a, b) are included (due to the linear dependency) in the transformation matrix:

$$\begin{pmatrix} \underline{m}_x \\ \underline{m}_y \end{pmatrix} = \begin{pmatrix} \underline{1} & \underline{0} & \underline{T}_{xx} & \underline{T}_{xy} & \underline{C}_x & +\underline{b} \\ \underline{0} & \underline{1} & \underline{T}_{yx} & \underline{T}_{yy} & \underline{C}_y & -\underline{a} \end{pmatrix} \cdot \begin{pmatrix} \underline{d}_x \\ \underline{d}_y \\ \underline{s}_x \\ \underline{s}_y \\ \underline{c} \\ \varphi \end{pmatrix} \quad (4)$$

The unknown φ acts on the additional column containing the image coordinates to produce the disturbance contribution to the measurement values (m_x, m_y). For the discussed example, one is able to retrieve the exact solution both for the calibration map (d_x, d_y) and for the pattern positions (s_x, s_y) by this method – the first being a zero-map with a zero residual indicating that Eq. 4 is a perfect mathematical model to explain the measured calibration data. Moreover, a value for the rotation angle φ can be extracted.

Mathematically, this method works as long as the extra column is linearly independent of all preceding columns. This condition is violated, for example, when the displacement ($b, -a$) depends on the stage position only. One then creates a new symmetry of the calibration problem, reducing the rank of the matrix. Not far from this extreme case, by introducing a near-symmetry, one would still destabilize the calibration. To find out whether this is the case and whether the destabilizing effects can be tolerated, it is necessary to consult the noise transfer functions for the modified equation system. In our example, the increase in noise transfer compared to Fig. 5b,c remains negligibly small.

6. CONCLUSION

Since there are good reasons to prefer self-calibration techniques over “golden mask” approaches in order to meet the required accuracy demands of modern metrology tools, the choice between the possible calibration options remains often illusive. In order to shed light on the topic and provide a quantitative guideline, a few different possibilities for the self-calibration on regular square grids have been compared. It has been shown that measurement redundancy is not only necessary in order to get a complete calibration, but also to average out the white noise present in any measurement process. This averaging process is of critical importance when the calibration must be even better than an individual measurement itself - which is required to reach the accuracy specifications in the photomask registration tools of the next generation. The capability of a calibration sequence to filter white noise has been discussed in terms of three error propagation functions. From the analysis of the results obtained by Monte-Carlo simulations, it has been concluded that the efficiency of noise suppression, measured by the normalized noise transfer function, is not determined by the redundancy of the measurement alone, but can be further optimized by choosing an adequate set of calibration positions. Alternatively, it is possible to use the information contained in the unnormalized transfer function in order to stabilize the calibration by selective averaging, at the cost of a reduced averaging efficiency. Spatial correlations have crucial importance for the efficiency of a calibration and can be minimized by studying the noise transfer in a symmetrical set of basis functions. Symmetries have been shown to be relevant in the correction of systematical errors as well – these can be treated by a modification of the original idea of self-calibration whenever supplementary data are available. The ability to filter such systematical errors without affecting the statistical stability of a calibration can be decisive as the specifications advance into the sub-nm range, where previously unnoticed influences are almost certain to play a role.

REFERENCES

- [1] Klose, G., *et al.*, “High resolution and high precision pattern placement metrology for the 45 nm node and beyond,” EMLC 2008, Dresden, Jan 21 – 24, VDE Verlag, 233 – 238.
- [2] Klose, G., *et al.*, “PROVE: a photomask registration and overlay metrology system for the 45 nm node and beyond,” Photomask and Next-Generation Lithography Mask Technology XV, Proceedings of SPIE Vol. 7028.
- [3] Klose, G., *et al.*, “Photomask Registration and Overlay Metrology by means of 193 nm Optics”, BACUS Photomask Technology Conference, Proceedings of SPIE (to be published)
- [4] Arnz, M., *et al.*, “Monte-Carlo Simulations of Image Analysis for flexible and high-resolution Registration Metrology“, EMLC 2009 (to be published)
- [5] Schellhorn, U., *et al.*, “Method for the determination of residual errors,” int. pat. appl. WO 2008/055589 (2008).
- [6] Rinn, K. *et al.*, “Method for correcting measurement errors in a machine measuring coordinates“, EP0931241 (1999)
- [7] Raugh, M., “Auto calibration method suitable for use in electron beam lithography,” patent US 4583298 (1984).
- [8] Raugh, M., “Self-calibration of Interferometer Stages,” ARITH-TR-02-01, Interconnect Technologies Corp. (2002).
- [9] Schellhorn, U. *et al.*, “Calibrating method, measuring method, optical measuring device and operating method”, int. pat. appl. WO 2005/124274 (2005).