Wafer based aberration metrology for lithographic systems using overlay measurements on targets imaged from phase-shift gratings

Sven van Haver,[1,2,*] Wim M.J. Coene,[1] Koen D'havé,[3] Niels Geypen,[1,4] Paul van Adrichem,[1] Laurens de Winter,[1] Augustus J.E.M. Janssen,[5] and Shaunee Cheng[3]

[1]*ASML (Research SMC), Flight Forum 1900, 5657 EZ Eindhoven, the Netherlands*

[2]*S[&]T Experts Pool (STEP), Olof Palmestraat 14, 2616 LR Delft, the Netherlands*

[3]*IMEC, Kapeldreef 75, B-3001 Leuven, Belgium*

[4]*Sioux, LIME BV, Esp 405, 5633 AJ Eindhoven, The Netherlands*

[5]*Department of Mathematics and Computer Science, Eindhoven University of Technology,*

P. O. Box 513, 5600 MB Eindhoven, The Netherlands

In this paper a new methodology is presented to derive the aberration state of a lithographic projection system from wafer metrology data. For this, new types of phase-shift gratings (PSG) are introduced, with special features that give rise to a simple linear relation between the PSG image displacement and the phase aberration function of the imaging system. By using the PSGs as the top-grating in a diffraction based overlay (DBO) stack, their displacement can be measured as an overlay error using a standard wafer metrology tool. In this way, the overlay error can be used as a measurand based on which the phase aberration function in the exit pupil of the lithographic system can be reconstructed. In practice, the overlay error is measured for a set of different PSG targets after which this information serves as input to a least-squares optimization problem that, upon solving, provides estimates for the Zernike coefficients describing the aberration state of the lithographic system. In addition to a detailed method description, this paper also deals with the additional complications that arise when the method is implemented experimentally and this leads to a number of model refinements and a required calibration step. Finally, the overall performance of the method is assessed through a number of experiments in which the aberration state of the lithographic system is intentionally de-tuned and subsequently estimated by the new method. These experiments show a remarkably good agreement, with an error smaller than $5m\lambda$, between the requested aberrations, the aberrations measured by the on-tool aberration sensor and the results of the new wafer-based method.

OCIS codes: 120.0120, 220.3740, 100.5070, 220.1010, 050.0050

http://dx.doi.org/10.1364/XX.99.099999

1. Introduction

In recent years, the specifications and tolerances for aberration control in the lithographic industry have reached a level at which also the contributions from dynamic effects form a significant fraction of the overall aberrations of a lithographic system. In this context, an aberration is considered dynamic when its contribution may change during the complete exposure of a wafer. This includes, for example, lens heating, vibrational modes of lens elements due to the scanning motion of the reticle stage or even refractive index variations of the immersion fluid caused by flow or bubbles. Although, modern day lithographic systems are all equipped with on-tool sensors to measure their aberra-

tions, these sensors commonly operate in between wafer exposures and consequently are unable to pick-up the dynamics of the mentioned effects. In the present work we aim at developing an alternative wafer based aberration metrology method, with a desired accuracy better than $2.5m\lambda$, that can pick-up and study these effects.

In the semiconductor industry, lithography is applied to transfer a desired pattern from a reticle onto a light-sensitive layer on a wafer. If aberrations are present, these give rise to deformations of the transferred pattern, like lateral feature displacements and blurring effects, which in turn may lead to malfunctioning electronic circuits or chips. Therefore, aberration control in the semiconductor industry is mainly driven by the tight error tolerances for the features being printed. This also means that aberrations are only considered relevant if they significantly alter the printed pattern and, looking at the problem from this perspective, it seems log-

———————

* Corresponding author: svenvanhaver@gmail.com

ical to monitor the aberration state of a lithographic system based on information extracted from the wafer. However, quantitative wafer based aberration analysis proves to be very challenging. Not only are the aberration induced changes very small, they are also strongly correlated with other feature shape changing effects like focussing and wafer processing. Because of these and other complications, wafer based aberration metrology methods have never become an industry standard that is widely used. Nevertheless, we will show in this paper that the above mentioned issues can be overcome and that a wafer based method can be devised that is an excellent candidate to study dynamic aberration effects and also has the potential to act as an independent reference for the established on-tool aberration sensors (based on shearing-interferometry [1]).

The wafer based aberration metrology method presented here is inspired by a paper of Nomura published in 2001 [2]. In this paper, Nomura introduces a new type of aberration sensitive phase-shift gratings (PSG), based on which he could devise a method to distinguish between odd and even aberrations, simply by measuring the aberration induced image shift for a few PSG images. In the current work, we build upon his pioneering work, to devise a quantitative method that can retrieve the complete aberration function of a lithographic system.

The remainder of this paper is organized as follows. In Sec. 2 the basic concepts and measurement principles of the new method are discussed. This includes the design and systematic analysis of the PSG targets, a derivation of the model relating PSG overlay errors to Zernike aberrations and also a simulation based proof-of-principle. Sec. 3 deals with the additional complications that arise when the proposed method is implemented experimentally. There it will be shown that, although some of the distinguished experimental issues have the potential to significantly degrade the performance of the method, they can all be accounted for, or compensated, through a number of model refinements and a calibration operation. In Sec. 4 the implementation of the method is discussed and experimental results are given for a number of aberration reconstruction examples. This will also illustrate the impact of the various model refinements and the calibration operation introduced in Sec. 3. Sec. 4 is concluded with an experiment in which the aberrations measured by the new wafer based method are compared with the on-tool aberration sensor results, showing that both methods agree to a level better than 5mλ. We will conclude the overview of our work in Sec. 5, where we shall summarize and comment on the results obtained and where we shall give recommendations for further development of the method.

2. PSG-OVL Aberration Metrology

In this section, the basic concepts and measurement principles exploited in this paper, are discussed. Based on a simplified image formation model and new types of phase-shift gratings (PSG), it is shown that a straightforward relation can be realized between the displacement of a PSG image and the aberrations of the imaging system. In addition, it is explained how standard wafer metrology tools can be exploited to measure this aberration induced displacement for a tailored set of different PSGs, in this way collecting enough information to allow estimation of the aberration function of the optical system.

2.A. Aerial Image and 2-Beam Interferences

The simplified image formation model is based on the following assumptions. For the illumination conditions in the lithographic system, the effective illumination is assumed to consist of a circular centered monopole with a radius σ that is very small compared to the numerical aperture (NA) of the lithographic projection lens and thus effectively generates a plane wave towards the reticle. Upon transmission of the incident plane wave through a periodic structure on the reticle, a discrete number w of diffracted beams are generated that lie within the NA of the projection lens. The aerial image of such a diffraction grating is then a superposition of a number W of 2-beam interferences, with $W = \binom{w}{2}$. Each of these 2-beam interferences can be represented as a cosine-fringe in the aerial image, characterized by two parameters that are its phase-offset and its amplitude. These two beams are characterized by the complex-valued diffraction amplitudes given by $C_1 \exp i\varphi_1$ and $C_2 \exp i\varphi_2$, and by their reciprocal space coordinates denoted by (h_1, g_1) and (h_2, g_2), respectively. The interference of these two beams leads to a spatial frequency component of the aerial image, at the coordinates $(h = h_1 - h_2, g = g_1 - g_2)$, given by

$$\tilde{I}_{h,g} = C_1^2 + C_2^2 + 2C_1 C_2 \cos \left[2\pi(h_1 - h_2)x + 2\pi(g_1 - g_2)y + \varphi_1 - \varphi_2 + \Delta\Phi_{1,2} \right] , \quad (1)$$

where we will assume for now that $\varphi_1 = \varphi_2$ (the $\varphi_1 \neq \varphi_2$ case is discussed at the end of Subsec. 3.D). In this case, the overall phase-offset of the cosine-fringe, denoted by $\Delta\Phi_{1,2}$, is given by the difference of the wave-aberration function taken at the two spatial frequencies or diffraction angles in the exit pupil of the lens; the amplitude of the cosine-fringe is determined by the diffraction amplitudes of the grating at mask level. The pattern-shift of the aerial image is a complicated function of the phase-shift and amplitudes for all the 2-beam interferences that give rise to the aerial image. A directly interpretable and unambiguous relation between the pattern-shift and the aberrations of the projection lens can be achieved when the aerial image is generated by only a single 2-beam interference (see Fig. 1). In this case the observed pattern-shift of the aerial image, ΔX_{AI}, is simply given by

$$\Delta X_{\text{AI}} = \frac{\Delta\Phi_{1,2} P}{2\pi} , \quad (2)$$

with P the period or pitch of the generated fringe pattern.

Given that, for the aimed aberration metrology, the entrance pupil consists of a single monopole centered at
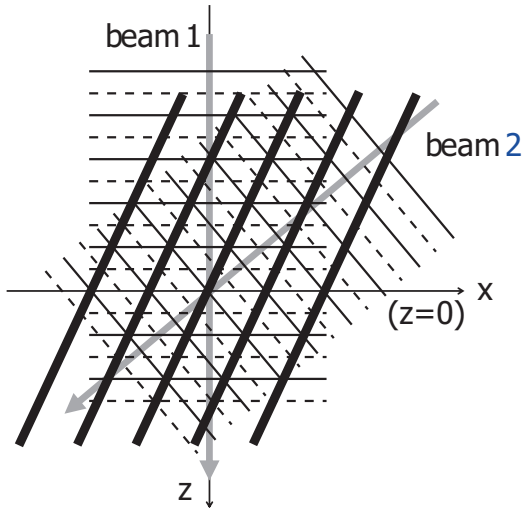
Fig. 1. Two beam interference image formation. Beams 1 and 2 have equal wavelengths with dashed and solid lines representing tops and valleys of the plane wave, respectively. Where solid (or dashed) lines intersect, both beams are in phase and constructive interference takes place to form the image (interference fringes indicated by the thick black lines). A shift of the image plane in the z-direction (defocus) and/or a phase change of either beam will only shift the complete interference fringe in the image plane (z=0) in the lateral (x) direction.

zero, a 2-beam interference can only be realized by a diffraction grating at mask level that has at least one diffraction order that is "forbidden", i.e. has zero amplitude.

2.B. Gratings with Forbidden Diffraction Orders

The analysis of forbidden orders of a 1-dimensional diffraction grating with N equidistant subdivisions or taps in its unit cell, and with for each j-th tap a respective scattering factor denoted by f_j , with $j = 0, 1, \cdots, N-1$, is typically carried out in terms of the unit cell's structure factor, which is, for the k-th diffraction order, given by (see e.g. Ch. 2, Eq. (46) in [3], with $i^2 = -1$)

$$F_N(k) = \sum_{j=0}^{N-1} f_j \exp\left\{2\pi i \frac{jk}{N}\right\} . \qquad (3)$$

First, it should be noted that the tap-based scattering factors f_j have to be complex-valued in order to realize a forbidden order only at the $+k$ diffraction order while the $-k$ diffraction order has non-zero amplitude. Indeed, for real-valued scattering factors f_j, the symmetry relation $|F_N(-k)| = |F_N(+k)|$ is always satisfied (also known as Friedel's law, a well-known property of the Fourier transform of a real-valued function). Next, the question should be addressed how many taps N are to be distinguished within the unit cell. For $N = 2$, only the trivial solution for the set f_j is obtained, equal to $\{1, 1\}$, which actually implies a pitch halving. For $N = 3$, only

k	0	1	2	3	4
1	O	X	O	O	O
2	O	X	X	O	O
3	O	X	X	X	O
4	O	X	O	X	O
5	O	O	O	X	O
6	O	O	X	X	O
7	O	O	X	O	O

Table 1. Possible sequences of diffraction orders for $N = 4$, with "O" and "X" indicating non-forbidden and forbidden diffraction orders, respectively, and the top-row giving the indices k of the diffraction order (note that the 0-th order relates to the centered monopole illumination).

a real-valued solution for the set f_j is obtained given by $\{1, 1, 1\}$, which does not qualify. For $N = 4$, non-trivial complex-valued solutions are obtained for the set f_j. Note that without loss of generality, $f_{(j=0)}$ can be set equal to 1. Possible sequences of diffraction orders in the reciprocal space unit cell $k = 0, \cdots, N$ are given in Table 1. The cases 1 and 2 in Table 1 will be shown to be the (independent) relevant cases.

Case-1 requires that $|F_{(N=4)}(1)| = 0$, which leads to the condition

$$1 + if_1 - f_2 - if_3 = 0 . \qquad (4)$$

This single equation on $\{f_1, f_2, f_3\}$ leaves two degrees of freedom. By making the choice $f_2 = 0 = f_3$, the relation reduces to $1 + if_1 = 0$, which yields the set f_j given by $\{1, i, 0, 0\}$, which is identified as Nomura's phase-shift grating (PSG) [2].

Case-2 requires that both $|F_{(N=4)}(1)| = 0$ and $|F_{(N=4)}(2)| = 0$, which leads to the simultaneous conditions

$$1 + if_1 - f_2 - if_3 = 0 , \qquad (5)$$
$$1 - f_1 + f_2 - f_3 = 0 . \qquad (6)$$

The latter equations still leave one degree of freedom on $\{f_1, f_2, f_3\}$. By making the choice $f_3 = 0$, a non-trivial solution for the set f_j is found to be $\{1, 1+i, i, 0\}$. Clearly, by having two consecutive diffraction orders ($k = 1$ and $k = 2$) that are forbidden, a larger range of pitches can be accommodated before a higher diffraction order is allowed within the NA of the projection lens. In that sense, case-3 would hypothetically be even more preferable, but its simultaneous conditions given by

$$1 + if_1 - f_2 - if_3 = 0 , \qquad (7)$$
$$1 - f_1 + f_2 - f_3 = 0 , \qquad (8)$$
$$1 - if_1 - f_2 + if_3 = 0 , \qquad (9)$$

just allows for the trivial solution for the set f_j given by $\{1, 1, 1, 1\}$, which corresponds to an effective pitch that is a quarter of the original one. Similarly, case-4 corresponds to a pitch halving. Case-5 and 6 are further
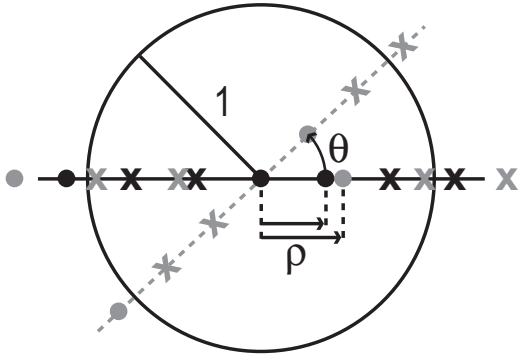
Fig. 2. PSG diffraction pattern control. Dots and crosses represent allowed and forbidden diffraction orders, respectively. Variation of the PSG pitch scales the diffraction pattern and changes the radial sampling position of the +1-st diffraction order. A rotation, θ, of the PSG grating in the xy-plane gives rise to an identical rotation of its diffraction pattern in pupil space.

equivalent to case-1 and 2, respectively, while case-7 obviously cannot lead to a situation with just 2 non-zero diffraction orders within the NA of the exit pupil. Concluding, for $N = 4$, there are two solutions for the set f_j that yield a diffraction pattern that can be used to generate a 2-beam interference process within the NA of the optical system, given by:

$$\{1, i, 0, 0\} \qquad \text{OXOOO} , \qquad (10)$$
$$\{1, 1 + i, i, 0\} \qquad \text{OXXOO} . \qquad (11)$$

Note that a complex scattering factor implies that light transmitted by the corresponding tap is phase-shifted with respect to other taps. This is why a grating involving complex valued taps is commonly referred to as a phase-shift grating (PSG) and we will use this nomenclature to address this type of gratings throughout the remainder of this paper.

2.C. Sampling the phase aberration function

Using the grating designs given in (10) and (11), it is possible to generate a diffraction pattern having only two non-zero (allowed) diffraction orders within the NA of the optical system, while all other diffraction orders are either zero (forbidden) or lie completely outside the NA. In this case, the image of the grating will be generated from a single 2-beam interference and the simple relation given in Eq. (2) applies. Consequently, the phase difference of the aberration function evaluated at the position of the two allowed diffraction orders can be determined by measuring the displacement of the aerial image at the wafer level.

As the aberration function is probed at the location of the allowed diffraction orders, we can control the probing locations by varying the grating parameters. Assuming normal incident plane wave illumination, the 0-th diffraction order of the grating will be generated at the center of the exit pupil, while the higher orders can be

generated anywhere in the pupil by changing the pitch and rotation of the grating, accordingly (see Fig. 2). Consequently, the gratings defined in (10) and (11) can be used to sample the phase difference of the aberration function between the center position and another position elsewhere in the pupil, by controlling the 1-st diffraction order position via the pitch and orientation of the grating.

However, the grating designs given in (10) and (11) cannot be used to sample the entire pupil. This is because Eq. (2) is only valid in the case of 2-beam imaging. This means that, for larger pitches, when the diffraction pattern is compressed relative to the size of the pupil, the condition should be satisfied that no non-zero higher diffraction orders enter the pupil. As a results, the following constraint applies on the maximum allowed grating pitch

$$P \le \frac{k_{min} \lambda}{NA} , \qquad (12)$$

where k_{min} specifies the lowest diffraction orders that should be kept outside the NA to satisfy the 2-beam requirement. Note that k_{min} is different for both designs given in (10) and (11); the first (Nomura) design has one allowed 2-nd diffraction order implying $k_{min} = 2$, while the second design has both 2-nd diffraction orders forbidden thus allowing $k_{min} = 3$. As a result of the constraint given in Eq. (12), the aberration function cannot be sampled close to the center of the pupil (see also Fig.4), with the excluded area being larger for the design in (10) than in (11).

2.D. Aberration Information from Overlay Measurements

In the previous subsections it was explained that the image displacement of a PSG directly provides the phase difference between two discrete points of the aberration function. However, since the aberrations we intend to measure are expected to be very small, typically in the order of $5m\lambda$ for well corrected (lithographic) systems, it can prove quite challenging to measure the aberration induced image displacement with a high enough accuracy to pick-up these effects. Fortunately, the measurement of image displacements is a common task in the lithographic industry, known as overlay metrology, and we can exploit existing commercially available tools to measure our grating displacements accurate enough for the aimed aberration metrology described in this paper.

In the lithographic industry several different overlay metrology methods, which can be either diffraction-based or image-based, are available to determine the shift between layers, known as the overlay error. They all have in common that both layers, for which the displacement should be compared, should contain standard grating features from which it is possible to deduce the relative layer placement error. In the remainder of this paper we shall use an overlay measurement methodology known as diffraction based overlay (DBO) to measure
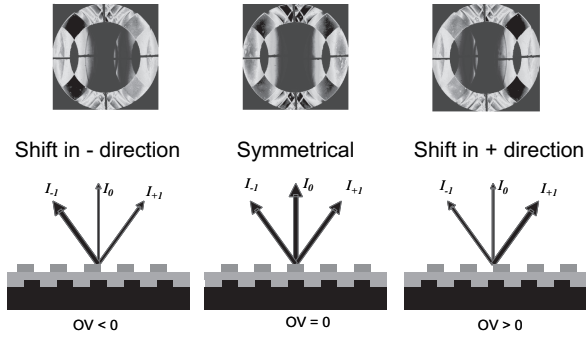
Fig. 3. Basic principle of diffraction based overlay (DBO). A focused spot, with an annular pupil distribution, is used to illuminate a stack with two gratings. The back reflected light, containing diffraction orders, is collected and imaged on a CCD camera (top row images). In the case that both gratings are perfectly aligned, the image on the CCD will be symmetrical (center column), while if a displacement is present, the recorded CCD image is asymmetrical (left and right columns). Then using two such grating stacks, with known biasses in overlay, the relative displacement between the top and bottom grating layers can be determined from the measured asymmetries in free first diffraction orders reflected by both stacks.

the PSG image displacements. We choose DBO as it is the method that currently provides the highest measurement accuracy and reproducibility based on a commercially available system (ASML YieldStar S/T-200). Nonetheless, the basic ideas and principles enabling the aberration metrology method presented in this paper can also be used to devise a similar aberration metrology method based on another overlay measurement technique.

In DBO, the displacement between two layers is assessed by including identical line-space (LS) gratings in each layer which are then effectively printed on top of each other. What results is a stack with two gratings and possibly some other unstructured layers in between (see Fig. 3). The relative shift between both gratings in the stack is determined via an asymmetry in the intensity of the +1st and −1st diffraction orders; the stack is illuminated by a spot after which the back reflected light, containing the first diffraction orders coming from the gratings, is collected on a CCD camera. Combining pupil images from two such grating stacks, both with a different yet known displacement between top and bottom grating, it is then possible to accurately compute the overlay error between both layers (for DBO details see [4, 5]).

The above described standard DBO measurement can be adapted for our application as follows. We generate the top resist grating in the stacks from a PSG. As bottom grating we use a standard line-space (LS) pattern etched in silicon having the same pitch as the PSG and which we assume manufactured and placed perfectly so that it may act as an absolute reference. In

this case the OVL error measured between both gratings in the stacks, directly gives us the PSG image shift due to aberrations present during the exposure of the PSG layer (potential issues originating from a non-perfect bottom reference grating and/or additional OVL contributions coming from other sources than aberrations are discussed in Sec. 3).

A restrictive consequence of using DBO is that it imposes a constraint on the minimum pitch of the PSG gratings that can be used. If the pitch of the gratings in the stack is too small, their first diffraction orders will not lie within the NA of the wafer metrology sensor and consequently no overlay error can be determined. This leads to the following constraint on the minimum pitch

$$P > \frac{\lambda_{\mathrm{DBO}}}{D\,NA_{\mathrm{DBO}}}\;,\tag{13}$$

where NA_{DBO} and λ_{DBO} are the numerical aperture and radiation wavelength of the DBO-tool and D is a parameter controlling the amount of the first diffraction order present within the NA of the DBO-tool. For $D \leq 2$, first diffraction order information is available, however to allow accurate displacement measurements a typical value of $D = 1.8$ is required. Consequently, the minimum grating pitch allowed by the DBO overlay measurement (with a YieldStar S/T-200; $NA = 0.95$, $\lambda = 425$nm) is $P \geq 249$nm. In pupil space, this constraint translates to a sampling area in the pupil bounded by the radius $\rho \leq 0.57$ thus preventing us to sample close to the edge of the pupil (see Fig. 4).

2.E. Sampling close to the pupil edge: 2D-PSGs

The combination of the constraint on the minimum PSG pitch which was given in Eq. (13), together with the constraint on the maximum PSG pitch in Eq. (12), implies that the resulting pupil area that can be addressed (see Fig. 4) is clearly insufficient to collect enough phase information to successfully retrieve the aberration function. Also note that the pupil area that can be addressed by the design in (10) is completely contained by the area accessible with the design in (11), we will therefore exclusively use the latter design in the remainder of this paper and shall refer to it as 1D-PSG. To enlarge the fraction of the pupil that can be sampled using PSG targets in combination with DBO, we shall now introduce a new type of two-dimensional phase-shift gratings (2D-PSG) on top of the 1D-PSGs introduced in subsection 2.B and will show how the additional design freedom for these 2D-PSGs enables one to sample the aberration function near the pupil edge.

A 2D-PSG is defined in this context as a two-dimensional phase-shift-grating with a typically oblique unit cell, that allows for 2 (truly) diffracted beams within the NA of the projection lens, on top of the zeroth order beam. This implies that there are 3 independent 2-beam interferences that make up the aerial image at the wafer level; two of these three 2-beam interferences comprise the zeroth order beam, and the third 2-beam in-
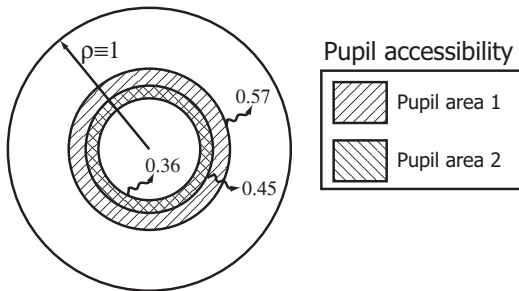
Fig. 4. Pupil accessibility of 1D-PSGs in combination with DBO. The concentric circle at pupil radius 0.57 represents the maximum radial sampling position allowed by the constraint in Eq. (13) (for DBO with a YieldStar S-200; $NA = 0.95$, $\lambda = 425$nm). The circles at pupil radius 0.45 and 0.36 represent the minimum radial pupil position allowed by the constraint in Eq. (12) for the PSG design given in (10) and (11), respectively. Consequently, phase information from the hatched *Pupil area* 1 can be obtained using PSG design (11) while design (10) can only be used to sample the smaller pupil subset labeled *Pupil area* 2.

terference occurs between the two truly diffracted beams (with non-zero reciprocal space coordinates (h_1, k_1) and (h_2, k_2), respectively). The parameters of the 2D unit cell, i.e. lattice parameters a and b and the angle ϕ between them, are chosen such that the DBO metrology tool is only sensitive to the latter 2-beam interference. This will be explained in more detail in the sequel.

The structure factor of the 2D unit cell, comprising N taps per direction and $M = N^2$ objects, is given by

$$F_N(h,k) = \sum_{j=0}^{M-1} f_j \exp\left\{2\pi i(x_j h + y_j k)\right\} , \qquad (14)$$

with (h, k) the indices of the diffraction order, and f_j and x_j, y_j the scattering factor and the (fractional) 2D-coordinates of the j-th object, respectively. A 2D-PSG must then satisfy the following conditions (indicating a number of so-called *forbidden* diffraction orders)

$$F_N(-1,0) = 0$$
$$F_N(0,1) = 0$$
$$F_N(-1,-1) = 0$$
$$F_N(1,1) = 0 \qquad (15)$$

These conditions yield a diffraction pattern with forbidden orders as shown in Table 2.

Two structures that satisfy the conditions in (14) are shown in Fig. 5, both having 4 non-trivial (non-zero) scattering objects in the unit cell. Note that outside of the scattering object, the unit cell has no transmission. It should further be noted that some of the objects need to have a complex-valued scattering factor to enable the generation of forbidden diffraction orders, following a similar argumentation as in Subsec. 2.B for the 1D-PSG case. The first type unit cell has three different phases, 0 , +90 and −90 degrees, while the second type has only

h	-1	0	1
k			
1	O	O	X
0	O	O	X
-1	X	X	O

Table 2. Schematic representation of the diffraction pattern defined by the conditions in (15), with "O" and "X" indicating non-forbidden and forbidden orders respectively.

Fig. 5. Two 2D-PSG unit cell designs satisfying the conditions in (15). Both designs contain 4 scattering objects and their relative scattering factors are color coded according to the legend on the right; i and -i indicate a relative phase offset of +90 and −90 degrees respectively. Note that the correct positioning and relative scattering factor of the objects in the unit cell are essential in achieving destructive interference for certain diffraction orders, the actual shape of the scattering objects is, in this respect, of minor importance.

2 different phases, 0 and −90 degrees (parameters for both types are given in Table 3). The 2nd type 2D-PSG has the benefit that it uses only two phases, thus requiring less process steps in reticle manufacturing, and has better diffraction efficiency due to its larger scattering objects. In the remainder of this paper we shall exclusively use the second type and we shall simply refer to it as 2D-PSG.

Next, the 2D unit cell can be deformed from a square into an oblique one, hereby retaining the conditions of the forbidden orders. At this stage it proves more convenient to use an alternative, yet equivalent, representation of the 2D-PSG unit cell as given in the upper-left of Fig. 6. The obliquity operation for the original unit cell in Fig. 5 is equivalent to a ratio change between the width, A, and height, B, of the alternative design (see second row of Fig. 6). In addition to this, we can also apply an obliquity operation to the new unit cell definition as shown in the bottom-left of Fig. 6. The impact

j	f_j	x_j	y_j
0	1	$\frac{1}{8}$	$\frac{1}{8}$
1	$+i$	$\frac{7}{8}$	$\frac{1}{8}$
2	$-i$	$\frac{1}{8}$	$\frac{7}{8}$
3	1	$\frac{7}{8}$	$\frac{7}{8}$

j	f_j	x_j	y_j
0	$-i$	$\frac{1}{2}$	$\frac{1}{4}$
1	$-i$	$\frac{3}{4}$	$\frac{1}{2}$
2	1	$\frac{1}{4}$	$\frac{1}{2}$
3	1	$\frac{1}{2}$	$\frac{3}{4}$

Table 3. Parameters defining the left and right unit cell shown in Fig. 5, respectively.
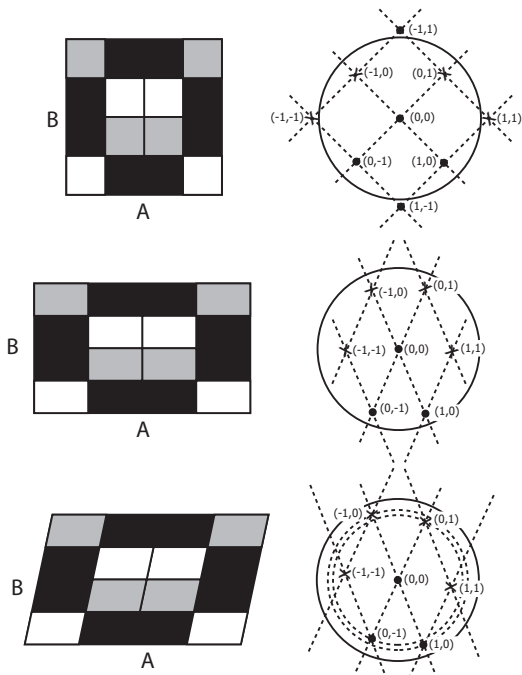
Fig. 6. 2D-PSG unit cells and their corresponding diffraction patterns (dots and crosses represent allowed and forbidden orders, respectively). The first row pertains to the nominal 2D-PSG design which is equivalent to the minimum unit cell given on the righthand side of Fig 5. In the second row, the unit cell is transformed such that $A \neq B$, resulting in a compression of the diffraction pattern in one direction. In the third row, the obliquity transformation is applied, resulting in the generation of the orders $(1,0)$ and $(0,-1)$ at different distances from the pupil center. Note that the diffraction orders $(1,0)$ and $(0,-1)$ are the two truly diffracted beams. The orders that are forbidden because of the 2D-PSG are: $(-1,0)$, $(0,1)$, $(-1,-1)$ and $(1,1)$. Some higher orders, for example $(-2,-2)$ and $(2,2)$, are not forbidden, but remain outside of the NA of the optical system. Note that the nomenclature used here, and throughout this paper, to indicate a diffraction order, pertains to the minimum unit cell as defined in the right-hand side of Fig.5.

of these transformations on the gratings diffraction pattern is shown on the right of Fig. 6. Going from the first to the second row, it can be observed that a ratio change between A and B compresses the diffraction pattern in one direction, effectively bringing the allowed diffraction orders (-1,0) and (0,1) closer together. On the other hand, the obliquity transformation makes that the two allowed orders are generated at a different radial distance from the pupil center and this will proof to be an essential feature measuring spherical aberration components later on. Full details on the relation between the 2D-PSG unit cell parameters and the resulting position of its allowed diffraction orders are provided in Appendix B.

Since the diffraction pattern of a 2D-PSG contains three allowed orders within the pupil, its aerial image is comprised of a superposition of three 2-beam interfer-
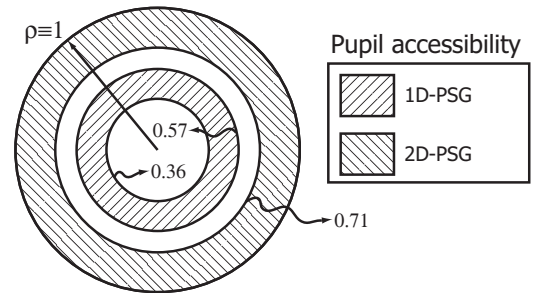

Fig. 7. Pupil accessibility of 1D- and 2D-PSGs in combination with DBO.

ence fringes (with their respective reciprocal space coordinates (h,k)):

$$(0,0) - (1,0) = (-1,0) \, ,$$
$$(0,0) - (0,-1) = (0,1) \, ,$$
$$(1,0) - (0,-1) = (1,1) \, . \qquad (16)$$

Now the important observation to be made here is that the first two fringes in (16), involving the zeroth order, will have a small period, as the distance between the orders is relatively large, and will therefore not satisfy the constraint in Eq. (13). On the other hand, the third fringe in (16) stems from interference between orders $(0,-1)$ and $(1,0)$ and, since these orders are relatively close together, generates a fringe ($h = h_1 - h_2, k = k_1 - k_2$) with a large enough period to be observed by the DBO tool. As a result, the complicated 2D resist grating generated from a 2D-PSG will be observed by the DBO tool as a simple 1D pattern as if it was generated exclusively from interference between orders $(0,-1)$ and $(1,0)$. Moreover, it turns out that even the simple relation given in Eq.(2) remains valid for the fringe generated by the orders $(0,-1)$ and $(1,0)$. Then, by applying a resist grating imaged from the 2D-PSG as the top grating in an overlay stack, and using a LS grating etched in silicon matched to the DBO observable 1D component, the phase difference between two points close to the pupil edge can again be measured as an overlay value.

Similar as in the 1D-case, a multitude of 2D-PSG gratings at different orientations at mask-level can be used so that the aberration function is sampled at different azimuths in the pupil. Moreover, some freedom in the allowed unit cells parameters can be used to achieve additional measurement diversity (for more details see Appendix B). Consequently, combining 1D-PSGs and 2D-PSGs, it is possible to collect differential phase information on the aberration function for the larger part of the pupil via DBO measurements (see Fig. 7). This, in principle, opens up the route towards a method capable of estimating the aberrational state of a lithographic system based on a large number of overlay measurements.

2.F. PSG-OVL aberration model

In the previous subsections it was explained that the phase difference, measured as an overlay value, between

different points in the pupil can be obtained by varying the type, unit cell parameters and orientation of the PSG targets used in the overlay stack. In this way, using both 1D- and 2D-PSGs with varied parameters, it is in principle possible to collect enough information on the aberration function to retrieve it. However, to estimate the aberration function based on differential phase information we need to define an appropriate model and aberration representation and for this purpose we proceed as follows.

Let $\Phi(\rho, \theta)$ be the aberration function defined on the pupil in normalized polar coordinates ($0 \leq \rho \leq 1$ and $0 \leq \theta \leq 2\pi$). The classical Zernike expansion of the aberration function is then given by

$$\Phi(\rho, \theta) = \sum_{n,m} \alpha_n^m Z_n^m(\rho, \theta) \tag{17}$$

where α_n^m denotes a Zernike coefficient for integer n, m such that $n - |m|$ is even and non-negative and $Z_n^m(\rho, \theta)$ denotes a classical Zernike circle polynomial defined as

$$Z_n^m(\rho, \theta) = R_n^{|m|}(\rho) \begin{cases} \cos(|m|\theta), & \text{for } m \geq 0 \\ \sin(|m|\theta), & \text{for } m < 0 \end{cases} \tag{18}$$

Note that we choose to use the double index Zernike convention here as it will prove mathematically more convenient to construct an appropriate model. On the other hand we will use the single index Zernike convention in the text and figures to refer to specific aberration terms. A conversion rule, to go from the single index Zernikes (FRINGE convention) to the double index, and vice versa, is provided in Appendix A.

Next, let (ρ_1, θ_1) and (ρ_2, θ_2) be the polar pupil coordinates of the two non-forbidden diffraction orders of a given PSG and let $\Delta\Phi_{1,2}$ be the phase difference of the aberration function evaluated at these positions

$$\Delta\Phi_{1,2} = \Phi(\rho_2, \theta_2) - \Phi(\rho_1, \theta_1) . \tag{19}$$

Then, using Eqs. (2), (17) and (19), we can write the aberration induced overlay error, OVL, for the current PSG as

$$OVL = \Delta X_{AI} = \frac{\Delta\Phi_{1,2} P}{2\pi}$$
$$= \frac{P}{2\pi} \sum_{n,m} \alpha_n^m [Z_n^m(\rho_2, \theta_2) - Z_n^m(\rho_1, \theta_1)] , \tag{20}$$

with the Zernike coefficients α_n^m describing the aberration state of the imaging system.

If the aberration state of a system is unknown, the expression in (20) allows one to devise a linear system of equations relating the unknown Zernike coefficients, α_n^m, to measured overlay values. Estimates for the α_n^m are then obtained by solving the standard least-square minimization problem defined by [6]

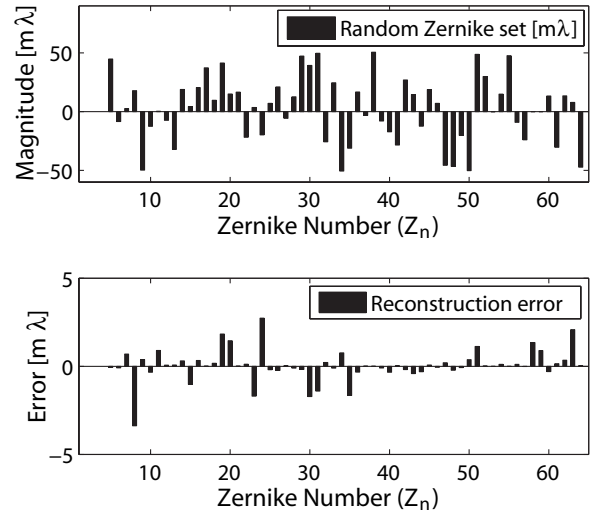$$\alpha_{\min} = \arg\min_{\alpha} \|E\alpha - OVL_{\text{meas}}\|^2 , \tag{21}$$

Fig. 8. Numerical proof-of-principle. A set of 60 Zernikes (Z_5 to Z_{64}) is defined with random values between -50 and $+50\text{m}\lambda$ (top row). For this Zernike set, the aberration induced shift of all 420 PSG targets is computed and the resulting simulated values serve as input to the minimization problem defined in Eq. (21). The bottom row presents the error in the estimated Zernike coefficients, which is better than $5\text{m}\lambda$.

with $\alpha \in \mathbb{R}^G$ the vector of unknown Zernike coefficients, $OVL_{\text{meas}} \in \mathbb{R}^H$ the vector of measured overlay values per PSG target and $E \in \mathbb{R}^{H \times G}$ a 2D matrix containing the model predicted overlay per Zernike term per PSG as constructed using Eq. (20). Note that $H > G$.

2.G. Proof-of-principle in simulation

In Subsec. 2.F it was shown how to relate the measured PSG overlay values to the aberrations of the lithographic system. Based on this model, a linear system of equations can be constructed that, upon solving, provides estimates for the Zernike coefficients representing the aberration function. To evaluate the potential of our method we have executed the following numerical experiment.

1. A set of 420 PSG targets (with a set of matching 1D line-space reference gratings at the bottom of the grating stack) is defined, aimed at sampling throughout the pupil while satisfying the constraints given in Eqs. (12), (13) and Appendix B.

2. Using a commercial lithographic simulator (Prolith (64-bit) Version 14.2.0.29) the aerial image of all PSG targets is simulated under realistic exposure conditions, given an arbitrary aberration state of the lithographic system.

3. The images of the reference gratings are simulated in the same way. However, since these gratings are on a different layer, the aberration conditions

during their exposure are possibly different, which is accounted for in the simulation by using another (unknown) aberration state.

4. The aberration induced overlay error per PSG target is then obtained as the shift between the aberrated PSG and reference grating aerial image.

5. Finally the simulated overlay values serve as input to the least-square minimization problem given in Eq.(21), which, upon solving, provides estimates for the α_n^m describing the aberrational state as defined during the simulation of the PSG images.

The results of the above described numerical experiment are presented in Fig. 8. In the top figure, the arbitrary set of 64 Zernikes used for this simulation, is shown. In the bottom figure the error per estimated Zernike is displayed. These simulation results clearly show that the Zernike coefficients describing the aberrational state of a lithographic system can in principle be estimated based on measured relative image displacements. Note that, an error is observed in the estimated coefficients due to the fact that the various constraints do not allow the pupil to be sampled optimally and because the simulation assumed that the reference grating was produced under different aberration conditions (random set with same order of magnitude as shown in top of Fig. 8). Nevertheless, the presented simulations shows that the method set out in this paper is capable of estimating the aberrational state of a lithographic system under these conditions, with an accuracy better than $5m\lambda$.

3. Experimental challenges and Refinements

In the previous section, the basic concepts and a numerical proof-of-principle were presented for our wafer based aberration metrology method. In the present section, we discuss the complications that arise when the proposed method is implemented experimentally and we will evaluate the impact of these potential issues on the overall performance of the method. Where necessary, refinements of the PSG-OVL model and measurement procedure are proposed in order to minimize their negative effect impact.

3.A. Finite illumination source

During the discussion of the basic measurement principles in Sec. 2 it was assumed that the PSG targets at the reticle are illuminated coherently by a plane wave at normal incidence. However, in reality all lithographic systems have a finite source, commonly combined with a Köhler illumination system. In such an illumination configuration, every source point generates a plane wave at a different angle towards the reticle and, consequently, the reticle image can be considered as the sum of many coherent contributions. In addition to this, the source is imaged in the pupil planes of the projection optics. This implies that light incident on a PSG, originating from different points on the source, will generate diffraction orders at slightly different spatial frequencies in the
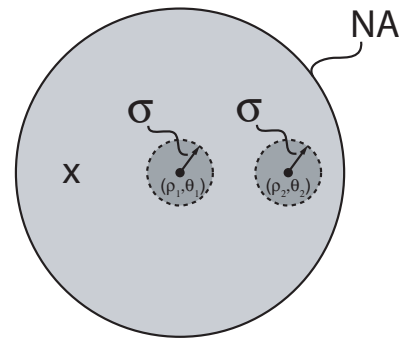

Fig. 9. Schematic representation of the projection optics exit pupil for the case of a 1D-PSG in combination with an on-axis finite monopole effective source. The monopole source, the size of which is defined by the radius σ relative to the NA of the projection system, is convolved with the PSG diffraction orders at locations (ρ_1, θ_1) and (ρ_2, θ_2). As the source is assumed spatially incoherent, each point on the resulting diffraction order discs can only interfere with its corresponding point on the other disc and each of these point-pairs effectively generate a single coherent contribution.

pupil and consequently will accumulate a different relative phase on propagation through the projection optics if aberrations are present (see also Fig. 9). Under these conditions, Eq. (2) is still valid to predict the aberration induced fringe displacement for a single coherent contribution, but does not give a good prediction for the overall displacement of the total PSG image, which consist of the superposition of all differently displaced coherent contributions. Consequently, the method presented in Sec. 2 will also not provide optimal aberration estimates for realistic lithographic systems due to the impact of their finite sources.

A straightforward way to reduce the impact of the finite source is to make it smaller so that it better approximates the coherent (single point) illumination case. In modern day lithographic systems this is possible via a user-defined effective source in the entrance pupil, for example a small centered disk. However, the minimum area of such an effective source is constrained, because making it too small either increases the exposure time too much or requires the energy density to be so high that the optics can get damaged. Consequently, the source representation in the pupil cannot be small enough to allow for an accurate coherent treatment of the problem and we need to accommodate for the finiteness of the effective source through the PSG-OVL model, for which we proceed as follows.

We assume the effective source in the pupil to be a small disk (monopole), with radius σ, that is spatially incoherent and uniform. In addition it is assumed that the diffraction efficiency of the scattering objects at the reticle do not vary with the angle of incidence of the light. The total PSG image is then obtained as the incoherent summation of all fringe intensities while integrating over the source and the effective aberration induced shift
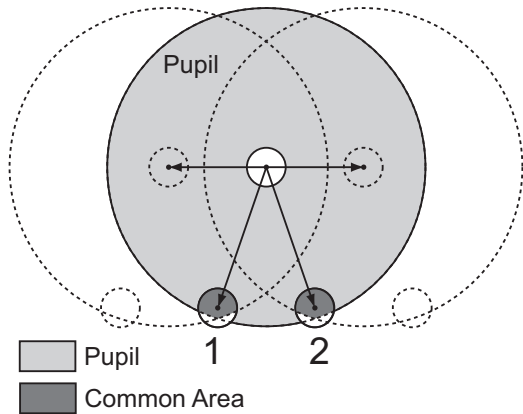
Fig. 10. 2D-PSG diffraction orders partly outside the pupil. A point on one diffraction order disc can only interfere, and thus contribute to the overall image, when its corresponding point on the other disc lies also within the pupil. As a result, the effective diffraction order phase is obtained by integration over the common area of both orders which will be an irregular shaped area when one, or both, diffraction orders lie partly outside the pupil. Systematically computing the average phase for such areas is dealt with in Appendix C.

simply becomes the average of the individual fringe displacements. In this case, a similar expression as Eq. (2) can be obtained, relating the overall PSG-image shift, in the case of a finite source, to the average phase difference between two areas in the pupil

$$\Delta X_{AI} = \frac{P}{2\pi}(\bar{\Phi}_\sigma(\rho_2, \theta_2) - \bar{\Phi}_\sigma(\rho_1, \theta_1)), \qquad (22)$$

where $\bar{\Phi}_\sigma$ denotes the average of the aberration function, taken over a disk with radius σ at the diffraction order positions (ρ, θ) (see Fig. 9). Next, the phase average over a disk-shaped area in the pupil can also be written in terms of Zernikes and we can finally write an equivalent expression for Eq. (20) in which the finiteness of the source is accounted for:

$$OVL = \Delta X_{AI} = \frac{P}{2\pi}(\bar{\Phi}_\sigma(\rho_2, \theta_2) - \bar{\Phi}_\sigma(\rho_1, \theta_1))$$
$$= \frac{P}{2\pi} \sum_{n,m} \alpha_n^m \left[\bar{Z}_{n;\sigma}^m(\rho_2, \theta_2) - \bar{Z}_{n;\sigma}^m(\rho_1, \theta_1) \right], \quad (23)$$

with $\bar{Z}_{n;\sigma}^m(\rho, \theta)$ the average of a Zernike function taken over a disk with radius σ at the pupil position (ρ, θ). Accurate computation of Zernike function averages required in Eq.(23) is discussed in Appendix C. It should be noted that the 2D-PSGs introduced in Subsec. 2.E require some additional attention. As they are meant to sample the aberration function near the pupil edge, we also have to deal with the case that, for either one or both of the relevant interfering diffraction orders, the averaging disc is partly outside the NA of the lithographic system. Since the effective source is considered incoherent, a point on the averaging disc will only contribute
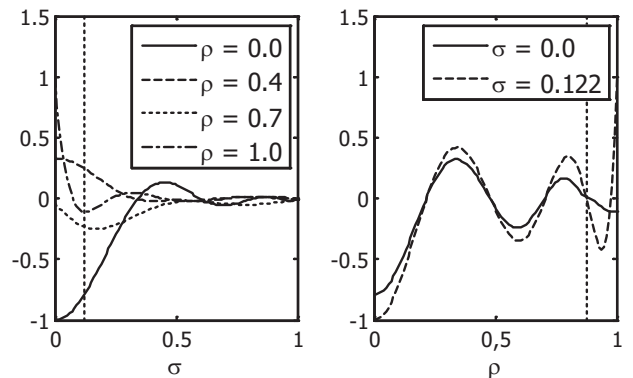

Fig. 11. Finite source impact. Assume a phase distribution in the pupil, defined by Zernike function Z_{36}. On the left, we plot Z_{36} averaged over a disk as a function of the averaging disk radius, σ, for four different radial positions in the pupil. On the right, the disk average is plotted as a function of its radial position to illustrate the impact of a finite monopole source ($\sigma = 0.122$) compared to the coherent case ($\sigma = 0.0$).

if the corresponding points in the other diffraction order also lies within the NA of the system. Source points for which only one diffraction order lies within the NA cannot interfere and consequently only produce a spatially uniform yet irrelevant contribution to the image. In Fig. 10 the common area of both allowed diffraction orders is indicated for a general case that both orders are cut differently by the NA limited pupil. The task of computing the relative phase for the resulting strangely shaped common area, which is indicated in Fig. 10, is also dealt with in Appendix C.

In Fig. 11 the impact of a finite source (on-axis monopole) on the effective phase of a diffraction order area is shown. On the left, the effective diffraction order phase, for Zernike function, Z_{36}, is shown as a function of the source radius, σ, for three radial positions of the averaging disk in the pupil. It can be seen that the effective diffraction order phase deviates substantially from the coherent case ($\sigma = 0$), already for modest illumination σ's, clearly showing the necessity of the finite source correction. The right-hand figure compares the effective diffraction order phase for $\sigma = 0.0$ (coherent case) and $\sigma = 0.122$ (realistic lower bound for the σ in a lithographic system). Especially, for averaging areas near the pupil edge ($\rho \to 1$), as encountered for the 2D-PSGs, the difference between the coherent and finite source models is non-negligible.

3.B. Asymmetric resist gratings

Another experimental complication comes from the fact that the PSGs are imaged by a 2-beam process into a layer of resist with a finite thickness. As a result of this, the resist grating formed from a 1D-PSG will be asymmetric (see left-hand side of Fig. 12). In the 2D-PSG case, where the image is formed by mutual interference between three orders, the resulting resist grating looks
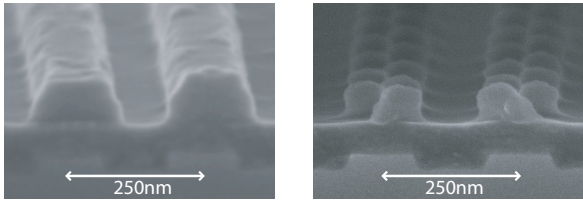
Fig. 12. Cross-section SEM images. On the left a cross-section SEM of a 1D-PSG wafer stack is shown and one can observe a significant difference in side-wall angle (SWA) between both sides of the grating lines. On the right a cross-section SEM of a 2D-PSG wafer stack is shown. It is clearly observed that, although the line appears to be built-up from tilted pillars, the 2D-PSG generates effectively a 1D grating with a pitch matched to the bottom grating. The fine structure in the grating line is sub-resolution for the DBO measurement, but can possibly contribute to the asymmetry signal generated by the whole stack.

even more exotic, with effective grating lines built-up from tilted pillars (see right-hand side of Fig. 12). These irregular grating shapes introduce an offset in the asymmetry signal of a DBO measurement and consequently also give rise to a process and PSG dependent, yet constant, offset in the measured overlay values. When not accounted for, these offsets introduce very large errors on the reconstructed spherical aberration terms (see also Fig. 15). Fortunately, these artifacts are independent of the aberrations that we want to determine, thus making it possible to remove them via a calibration step. For this purpose one can execute the presented aberration metrology method on a very well corrected lithographic system and assume that the overlay values measured on that system for all PSGs are exclusively due to the PSG grating asymmetries. Now by subtracting these measured overlay residuals, as a reference measurement, from any subsequent PSG-OVL measurement, one effectively removes the Zernike reconstruction error coming from the resist grating asymmetry. The price that is paid for this, is that the reconstructed Zernikes are now no longer independent and describe the aberration state of the unknown system relative to the reference system, but this is common practise in the lithographic industry and therefore no problem for our aimed application.

3.C.   Other overlay contributions

In this paper, the aberrations are estimated based on measured overlay values. Therefore, it can be easily understood that, overlay contributions coming from sources other than aberrations, directly propagate into the Zernike reconstruction error. Possible other sources contributing to the overlay are:

1. *Mechanical overlay*
   The actual displacement between different layers of the wafer, caused by wafer and reticle stage positioning errors.

2. *Reticle registration errors*

   The misplacement of features on the reticles with respect to the reticle designs.

3. *Reticle and wafer deformations*
   Temperature gradients, (thermo-)mechanical strain and/or processing can cause the reticle and wafer to deform, introducing position dependent relative displacements of the features they contain.

In its simplest form, mechanical overlay imposes a relative (x, y) position shift between features in two different layers of the wafer, which is identical throughout our small-sized target area in the die. In our method, where we measure the grating displacement in the direction of the grating vector only, the mechanical overlay contribution is given by the projection of the vector (x, y) on the grating vector. As a result, mechanical overlay gives a pitch independent contribution to the PSG grating displacement, but does vary with the PSG orientation. It turns out that the mechanical overlay fingerprint, the fingerprint being the combined response of all PSG targets to a given overlay source, is identical to that of a tilt aberration (linear combination of Z_2 and Z_3). However, in a lithographic system, these aberration components are typically ignored, since they are automatically compensated during alignment of the wafer before exposure, and we may conclude that mechanical overlay does not contribute to the relevant Zernikes (Z_t for $t > 3$).

The second overlay contribution listed is coming from the reticle. If features on the reticle are not placed exactly on the position defined in the reticle design, this displacement contributes directly to the observed overlay for the displaced target. Since the overlay is defined as the displacement between two gratings in two different layers imaged from two different reticles, the combined overlay contribution of both reticles is in fact the target specific relative displacement between its bottom and top grating, also known as the registration error between both reticles. This registration error, which is basically a manufacturing defect that remains static during the lifetime of a reticle, is commonly measured during manufacturing of a reticle set and could therefore be compensated for in the PSG-OVL model. However, this is not even necessary when a calibration as described in Subsec. 3.B is performed, as this calibration step also automatically compensates for the static overlay contribution induced by the reticle registration error. Also other reticle defects contribute to the Zernike reconstruction error of our method, but these are discussed in more detail in the next subsection.

The final contribution listed above is more tricky. The mentioned deformations introduce overlay errors according to the same mechanism as the reticle registration error, but in this case depend on environmental and exposure conditions, implying that they are not static, and are, therefore, not automatically removed by the calibration. The overlay contribution generated by wafer and reticle deformations is position dependent and can in principle result in an arbitrary overlay fingerprint for a
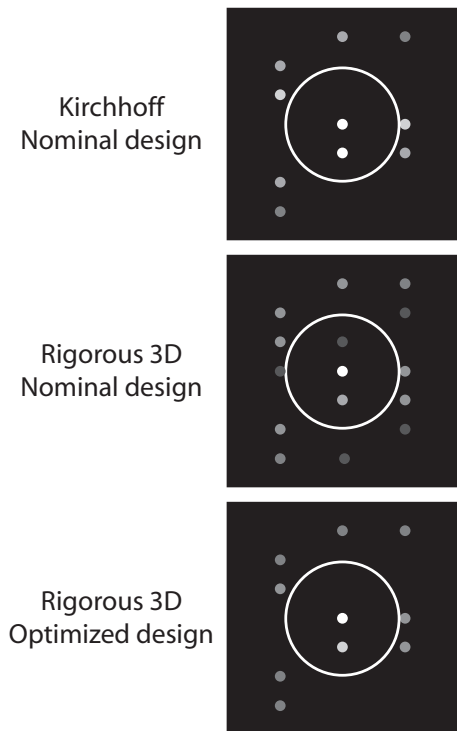
Fig. 13. Computed diffraction patterns for the nominal 1D-PSG defined in (11). The top row shows the diffraction pattern computed according to Kirchhoff, with as expected only two non-zero orders within the NA (indicated by the white circle). The middle row shows the diffraction pattern for the same target when correctly accounting for the reticle 3D-effects. In this case three non-zero orders are predicted within the NA, destroying the pure two-beam process. In the bottom row the rigorously computed diffraction pattern is shown for an optimized 1D-PSG design, showing that the pure two-beam process is restored.

PSG-target set. However, temperature, strain and processing deformations typically give rise to overlay contributions that vary slowly over the wafer. By placing all PSG targets required for our method close together, we can thus minimize the impact of these deformations to an acceptable low level. If this is insufficient, one may resort to complex computational models to predict temperature induced deformations, but this has not been investigated further in the research for this paper.

3.D. Reticle 3D-effects and manufacturing errors

A typical lithographic reticle consist of a quartz substrate on which a thin layer of chromium is deposited to make it opaque. By selectively removing chromium using electron beam lithography, a desired binary pattern can be transferred. In addition to this, our PSG designs require that certain transmissive areas on the reticle are assigned a different phase and this can be achieved by selectively etching those areas into the substrate. Combined, this allows us to manufacture a reticle with PSG targets according to the designs in Sec. 2. However, since
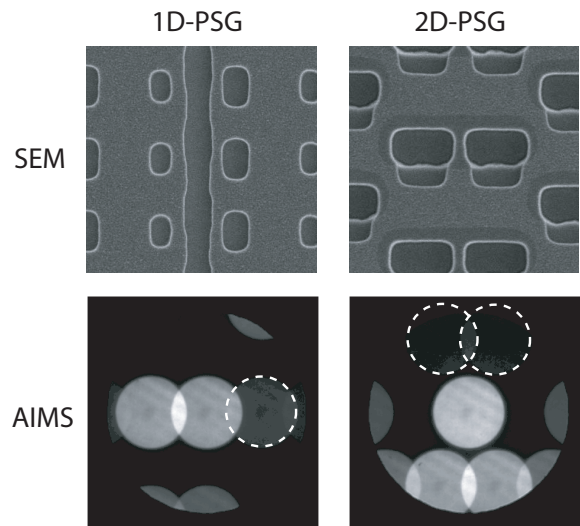

Fig. 14. Illustration and impact of reticle manufacturing defects. Top row shows a SEM image of a typical 1D- and 2D-PSG on the reticle. The 1D SEM image was taken before the phase etch step, while the 2D SEM image was taken afterwards. One can observe significant corner rounding and additional deformations due to the phase etch into the substrate. In the bottom row, a typical diffraction pattern for a 1D- and 2D-PSG is shown as measured by an AIMS microscope. Due to the writing and etching defects, the intended forbidden orders (indicated by dashed circles) are non-zero. Note that the monopole source of the AIMS is relatively large, resulting in overlapping orders and higher diffraction orders showing up at the pupil edge which are of no concern in the actual experiment in the lithographic system where a much smaller source is used.

the resulting PSG feature sizes on the reticle are in the same order of magnitude as the illumination wavelength and the grating can not be considered optically thin, the diffraction efficiency of the PSG targets is not predicted well by the Kirchhoff diffraction model. Instead, a rigorous 3D-diffraction model should be applied to correctly account for the light-matter interactions taking place at the reticle. In the first two rows of Fig. 13, the simulated diffraction pattern is shown for these two cases, clearly illustrating that the forbidden diffraction orders of the nominal PSG design are no longer zero when the 3D structure of the reticle is accounted for. At the bottom row of Fig. 13, however, it is shown that this issue can be solved by optimizing the nominal PSG design so that it compensates the 3D-diffraction effects to again achieve forbidden diffraction orders.

In addition to the 3D-diffraction effects discussed above, also reticle manufacturing errors can contribute to unwanted energy in the forbidden orders. These are for example corner rounding, under-etch and non vertical etch side walls and they can all modify the diffraction efficiency of the features on the reticle (see also Fig. 14). As these imperfections are difficult to predict and to parameterize it is impractical to accurately es-

timate their impact via computational means. Instead, one may choose to further optimize the PSG designs for these effects experimentally. This would then be done by making a test reticle containing PSGs for different combinations of design and processing parameters. Next, the amount of energy in the forbidden orders is assessed by measuring the diffraction pattern, for example using an AIMS microscope. The combination of parameters for which the forbidden order strength is minimal is then used to manufacture the actual measurement reticle.

In reality, the forbidden diffraction orders will never be completely zero. The first order effect of a small to moderate non-zero forbidden order is that it reduces the sensitivity of our proposed method for aberrations having an even azimuthal dependence (α_n^m with m is even), while not influencing the sensitivity of odd aberrations (m is odd). However, since we can, in principle, measure the relative forbidden order strengths of all PSGs with an AIMS microscope, this effect can also be accounted for in the model by adjusting the sensitivity matrix (E in Eq. (21)) accordingly. Nevertheless, it remains best practice to strive for the lowest possible amount of energy in the forbidden orders to avoid higher order effects and to have the highest possible sensitivity for all aberrations.

Even in the case of perfectly forbidden orders, there is one other reticle characteristic that can impact the OVL measured in our method and has not been discussed yet. This is the relative phase of the allowed diffraction orders, φ_1 and φ_2, as occurring in Eq. (1). Looking at Eq. (1) it can be understood that if $\varphi_1 - \varphi_2 \neq 0$ this directly adds up to, and cannot be distinguished from, $\Delta\Phi_{1,2}$, the aberration induced phase contribution and consequently contributes directly to the error of the Zernikes reconstructed by our method. Furthermore, the diffraction order phase is not easily measured using existing semiconductor equipment, which implies that a model based correction of this artifact is not feasible. But again, this reticle effect can also be assumed static assuring that it is also effectively removed by the calibration procedure described in Subsec. 3.B.

3.E. Stochastic overlay variations

The final topic we want to discuss in this section is the uncertainty in the overlay measurement. Two distinct sources generating stochastic variations in the measured overlay signal are recognized. The first, is the uncertainty coming from the DBO measurement itself and is extremely small, in the order of a few tens of an Angstrom. This very high measurement reproducibility was, in fact, the main reason why DBO was selected as the overlay methodology of choice for our application. It basically means that, no significant contribution to the method error is expected from the DBO measurement. The second contribution is caused by small differences in the resist gratings used in the overlay measurement. These differences may, for example, originate from the complicated and random interaction between

exposure radiation and light sensitive molecules in the resist layer or any other stochastic process during exposure or processing of the wafers. The uncertainty introduced by these mechanisms is much larger than that of the DBO measurement and experiments have also shown that their magnitude may vary as a function of the PSG parameters.

Under these conditions, the ordinary least squares minimization problem given in Eq. (21) is no longer optimal and should be replaced by the weighted least squares minimization problem given by

$$\tilde{\alpha}_{\min} = \arg\min_{\alpha} \|Q(E\alpha - (OVL_{\mathrm{meas}} - OVL_{\mathrm{ref}}))\|^2 \;,$$

(24)

where OVL_{ref} denotes the calibration measurement on the reference system, which we assume to be independent of OVL_{meas}, in which case K simply becomes the sum of the experimentally determined covariance matrices

$$K = K_{meas} + K_{ref} \;,$$

(25)

and Q denotes the Cholesky decomposition of K^{-1} according to

$$K^{-1} = QQ^* \;.$$

(26)

The optimal generalized least squares estimator, $\tilde{\alpha}_{\min}$, for the Zernike coefficients describing the aberration state of the lithographic system, can then be explicitly written as

$$\tilde{\alpha}_{min} = \left(E^* K^{-1} E\right)^{-1} E^* K^{-1} \left(OVL_{meas} - OVL_{ref}\right).$$

(27)

where we have now correctly accounted for the PSG target specific overlay measurement uncertainty and the additional uncertainty introduced by the calibration measurement.

An additional uncertainty related problem that we are facing for the current measurement reticles is that the 420 targets included in the measurement modules, do not sample the pupil optimally. As a result, the conditioning of the resulting linear system is not very good and this gives rise to a relatively large variance for specific (spherical) Zernikes. In this case, the least-squares estimation can be further improved by exploiting a Tikhonov-like regularization approach which introduces an additional term in the cost function:

$$\tilde{\alpha}_{min}\left(\zeta\right) = \arg\min_{\alpha} \|Q\left(E\alpha - (OVL_{meas} - OVL_{ref})\right)\|^2$$
$$+\zeta^2 \|D\left(\alpha - \alpha_p\right)\|^2 \;.$$

(28)

Here D is the regularization operator and α_p the prior mean, which are both set based on prior knowledge of typical aberration states of lithographic systems. The parameter $\zeta \geq 0$ determines the importance of the regularization term. Obtaining a good recipe for an optimal choice of ζ is a non-trivial task. For a more detailed description of the applied regularization techniques and the determination of the regularization parameter, the reader is referred to [7].

4. Experimental verification of the method

In this section we present the experimental validation of our method. We shall start by discussing the hardware used and we will describe the applied experimental procedures in detail, aimed at avoiding the experimental pitfalls discussed in the previous section. Subsequently, aberration retrieval results are presented for the case that a known aberration was intentionally introduced to the projection lens of a lithographic system and, additionally, these wafer based reconstruction results are compared to the aberrations measured by the on-tool aberration sensors.

4.A. Reticle design

The most important components required by the PSG-OVL aberration metrology method are the two matched PSG and REF reticles required to construct the PSG-OVL targets. For the reticle set used in the experiments presented in this paper, the top-level layout is identical. They contain 7×7 measurement modules, where each module consists off 420 PSG-based overlay targets. In turn, each overlay target consists of several grating stacks, that together allow one to determine a single overlay value (see also Fig. 3). The grating stacks themselves are built up from two stacked $40 \times 40 \mu m^2$ gratings, where the top and bottom grating are constructed from the PSG and REF reticle, respectively. Basically, the only difference between both reticle designs is the unit cell used to fill the $40 \times 40 \mu m^2$ grating areas (Note that dimensions are given at wafer level (1X)). For the PSG reticle, the unit cell definitions are given in Subsecs. 2.B and 2.E for the 1D- and 2D-PSGs case, respectively. The gratings used in the REF reticle design are simple line-space gratings, which have the same effective pitch, P, and grating orientation, θ, as the corresponding grating on the PSG reticle.

As mentioned above, the modules on the aberration measurement reticles contain 420 PSG-OVL targets. This set is composed from 28 different PSG-OVL targets (3 1D-PSGs and 25 2D-PSGs) all included at 15 different orientations (rotation angles in the reticle plane). For the current reticles, we have included 3 grating stacks per PSG-OVL target, and this results in a minimum area required for a complete module that is a little bit larger than 2mm^2. Note that on our test reticles the module size is substantially larger due to the inclusion of visually observable labels and additional test and calibration features. Nevertheless, our test reticles already includes 7×7 modules thus allowing one to also measure the field dependence of the aberrations in a single exposure.

4.B. Experiment description

Experimental verification of the method is executed as followes. First, a set of reference wafers is created; using the complimentary binary REF reticle, coated silicon wafers are exposed and subsequently etched, stripped and cleaned. The chosen illumination condition used to print the reference gratings was a wide conventional setting (0.94σ at $1.35NA$), this in order to minimize the impact of possible aberrations on the reference features. The etch depth for the reference gratings is 30nm.

The chosen resist process for all our experiments is 105nm JSR resist AIM5484 on top of 95nm Brewer BARC ARC29SR, this should result in sufficient planarization of the etched reference gratings. All coatings and development are done in-line on a Sokudo Duo track fitted to an ASML NXT:1950Ai scanner. Before starting experiments we determined the smallest safe conventional illumination possible on the scanner (0.122σ at $1.35NA$). We then determined the dose at which all required phase gratings would print simultaneously to an acceptable pattern by qualitative SEM inspection. That exposure dose was then used for all experiments.

Each experiment consists of a set of exposures on the reference wafers using the above mentioned resist process and tools. The use of so-called image-tuner sub-recipes allowed us to dial in specific aberration offsets per wafer. The validity of each sub-recipe was first verified by measuring the wavefront using the scanner's own integrated interferometer when applying the offsets requested in the sub-recipe. For the exposures themselves we used a specific scanner test that would also allow us to measure the wavefront before and after each exposure with the scanner's interferometer while still working in-line with the track to bake and develop each wafer immediately after exposure.

Measurements were then performed on each wafer on a YieldStar S200 and raw pupil data was gathered that was subsequently processed off-line to determine the overlay error for each PSG target. These measured overlay values, together with uncertainty information gathered from multiple measurements, then serve as input to the minimization problems defined in Eq. (24) and Eq. (28) which, upon solving, provide us with estimates of the aberration state of the lithographic scanner at the time of the PSG reticle exposure.

4.C. Experimental results

In this subsection we give an overview of the experimental work done to validate the method. We begin with an experiment in which a single non-zero Zernike has been dialed in as an aberration offset of the lithographic system. This intentionally de-tuned system is subsequently used to print the PSG reticle on top of the reference features already on the wafers. Next, the overlay is measured for all targets in a single module and, based on these measured overlay values, we try to recover the magnitude of the dialed in Zernike term. In Fig. 15 the Zernike estimates obtained with the PSG-OVL method are shown. The bar-plot on the top, pertains to retrieval based on the raw measured overlay values, while the finite source is correctly accounted for according to Subsec. 3.A. It may be observed that, although a Z_7 with a magnitude between 2 and 3nm is detected, the method erroneously reconstructs large values for several other
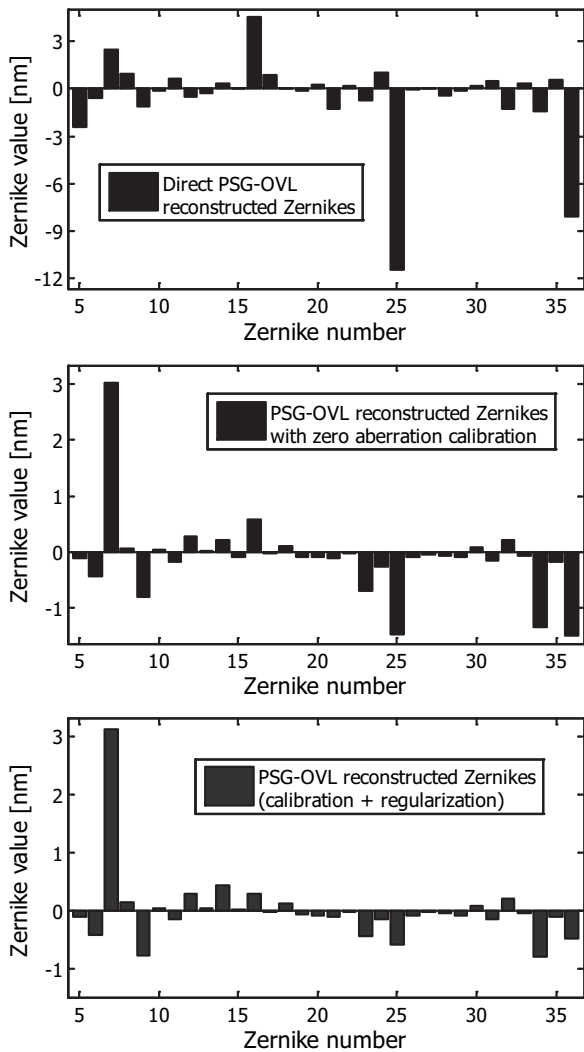
Fig. 16. PSG-OVL reconstructed Zernikes for a requested cocktail of 3 non-zero coefficients ($Z_8 = 2$nm, $Z_{10} = -2$nm and $Z_{27} = 2$nm).

Fig. 15. Zernikes estimated with the PSG-OVL method for the case that a single non-zero coefficient ($Z_7 = 3$nm) was dialed in during exposure of the PSG layer onto the wafer. On the top, estimates for the Zernikes ($Z_5 - Z_{36}$), obtained from the raw overlay measurements, are shown. In the middle, the corresponding retrieval result is shown for the case that the measured overlay values are first calibrated with respect to an aberration-free reference system. The bottom row shows the calibrated result when additionally regularization is applied in the fitting procedure.

ditioning of the linear system being solved is not very good and, in this case, results in relatively large uncertainty, specifically for the spherical Zernikes. A proven concept to improve estimates from poorly conditioned systems is the use of regularization and doing so further improves the Zernikes estimates obtained with the PSG-OVL method as is shown in the bottom row of Fig. 15. In the remainder of this subsection, all presented reconstruction results will be obtained with regularization and are based on calibrated overlay data.

The results of a second experiment, in which a cocktail of three non-zero Zernikes ($Z_8 = 2$nm, $Z_{10} = -2$nm and $Z_{27} = 2$nm) was dialed in, is shown in Fig. 16. Also in this experiment, a very good correlation is observed between the requested and retrieved Zernike terms and all residual Zernikes are smaller than 1nm (1nm ≈ 5mλ for a lithographic system operating at 193nm). This gives a good indication that the PSG-OVL method can determine arbitrary aberration states and does not suffer from a strong cross correlation between different aberration terms.

So far, only reconstruction results are shown for odd Zernike terms. As explained in Subsec. 3.D, the aberration induced overlay sensitivity of even Zernikes is reduced when the intended forbidden orders of a PSGs are non-zero. The impact of this phenomenon may be observed in the retrieval results shown in Fig. 17, where a cocktail of even Zernikes ($Z_5 = 2$nm, $Z_{13} = -2$nm and $Z_{17} = 2$nm) is dialed in during exposure. In the top row, the PSG-OVL based estimates are shown when this effect is neglected, indeed showing Zernike estimates that are too small. A significant improvement is observed when this effect is accounted for in the model using AIMS measured forbidden order information (see bottom row Fig. 17). Note that, AIMS measurements were only available for a subset of all PSGs and that missing data was supplemented by extrapolated values. Therefore, the improvement achieved by this correction is expected to be even more drastic when forbidden order information is available for all PSGs used in the reconstruction.

Zernikes, with most profound values for the spherical aberration terms (Z_{16}, Z_{25} and Z_{36}). These large spherical terms are mainly due to the PSG asymmetry effect discussed in Subsec. 3.B. Therefore, by applying the calibration proposed in Subsec. 3.B, a strong improvement can be achieved, generating the results shown in the middle row of Fig. 15. Nonetheless, even for the calibrated result, some reconstructed spherical Zernikes remain substantial. A systematic analysis of this phenomenon has shown that it is caused by the sub-optimal sampling of the aberration function achieved by the current set of measurement reticles. As a result, the con-
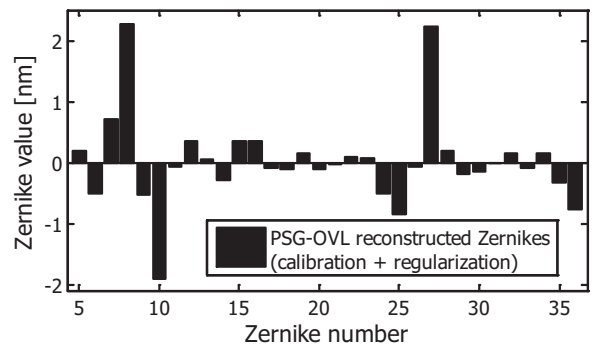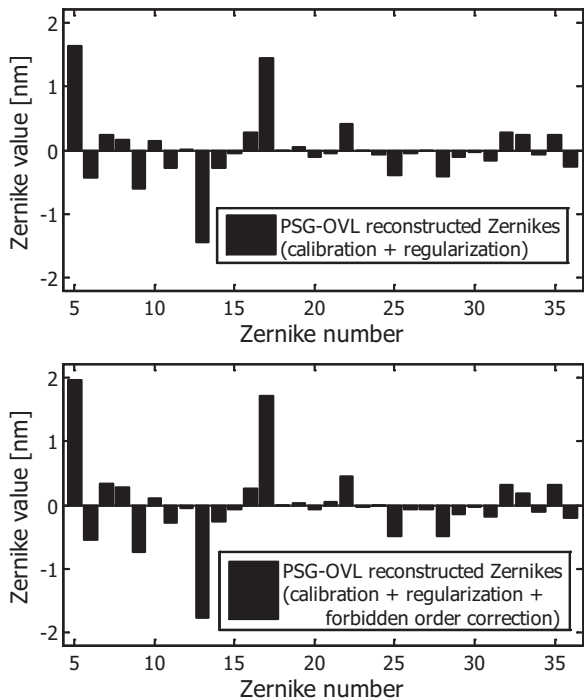
Fig. 17. Zernikes estimated with the PSG-OVL method for the case that a cocktail of 3 even Zernikes ($Z_5 = 2$nm, $Z_{13} = -2$nm and $Z_{17} = 2$nm) is dialed in. The top axis shows the Zernike estimates obtained using both calibration and regularization. The results in the bottom figure are obtained using an additional correction based on the AIMS measured residual forbidden order intensities generated by the PSG reticle.

Finally, we present in Fig. 18 a comparison between the aberration coefficients measured with the on-tool aberration sensors of the exposure system and those estimated by the PSG-OVL method. Also here, a very good agreement between both measurements is observed for the imposed non-zero Zernike coefficients. A comparison of the residuals of both methods shows larger differences between both methods, where it should be pointed out that the residuals of the PSG-OVL method are slightly larger than those of the on-tool sensor. Nevertheless, it is remarkable that such a good agreement, down to the sub-nanometer regime, can be achieved between a wafer based method and a dedicated on-tool sensor that is based on shearing interferometry.

5. Conclusion and discussion

In this paper we have presented a new methodology to derive the aberration state of a lithographic projection system based on wafer metrology data. This method uses the aberration induced image shift of specially designed overlay targets as a measurand based on which the phase aberration function in the exit pupil of the lithographic system can be reconstructed. In Sec. 2 it was explained how PSG targets can be designed such that a very simple relation emerges between their relative image position and the difference of the phase
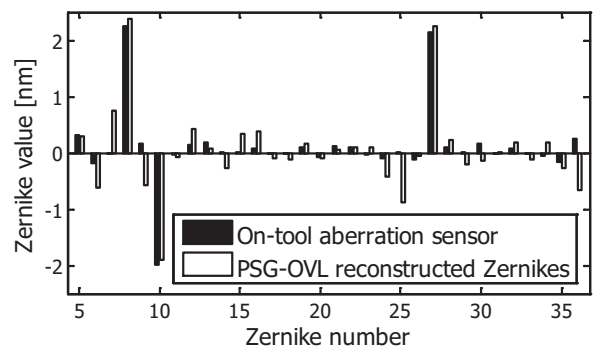

Fig. 18. Comparison between the aberration coefficients measured by the on-tool sensors of the exposure tool and the PSG-OVL method (identical experimental settings as in Fig. 16).

aberration function evaluated at the location of their diffraction orders. Moreover, it was shown that, using a combination of 1D- and 2D-PSGs, the phase difference between different points throughout the pupil can be probed, eventually providing enough information to accurately estimate the Zernikes representing the phase aberration function. At the end of Sec. 2 this methodology was proven to be feasible in simulation via a numerical experiment based on a lithographic simulator and showed an aberration reconstruction accuracy of better than 5mλ.

Before continuing with the experimental verification of the method, first a number of practical complications were addressed in Sec. 3. There it was shown that for all complications currently recognized, their impact can either be accounted for in the PSG-OVL model or their negative effect can be strongly reduced through a calibration step. This was illustrated in Sec. 4 through a number of experiments, with and without the mentioned corrections, which showed the most drastic improvement when the reference scanner calibration was included. Altogether, using all model refinement and corrections discussed in Sec. 3, the accuracy of the wafer based aberration estimation method introduced in this paper, was shown to be remarkably good. The retrieval error, with respect to the dialed in aberration and independent measurement with the on-tool aberration sensors, is below 1nm, which corresponds to approximately 5mλ at the exposure wavelength of the lithographic system being evaluated.

Nonetheless, we still observe residuals in the reconstructed Zernikes that are between 0.5 and 1nm. Although extremely small, aberrations of this order of magnitude are nowadays considered significant by the lithographic industry. Therefore, the aim of the current research was to achieve an aberration estimation accuracy better than 0.5nm and in this respect the current work still requires an additional effort to reach the desired accuracy. It should be noted here that metrology solutions operating in the sub-nm regime always suffer

from the lack off a well defined absolute reference. In a sense, one can no longer speak of an error for a given method, but can only compare results between different independent methods and, when differences are observed, it is impossible to say which method is most accurate. Apart from this merely philosophical issue encountered in advanced metrology, we do still see room for further improvement of our method and we shall discuss this here briefly.

For the experiments presented in this paper, we have used a set of reticles that was designed based on our initial understanding and ideas regarding wafer based aberration metrology. Since that time, we have gained a lot more insight on the physical processes and error sources relevant for our method. Basically, we now know that our reticles are, in many ways, not optimal for the aimed aberration metrology application. For example, the current set of 420 PSG targets does not sample the pupil in an optimal way and this leads to poor conditioning and significant cross-correlation in the linear system being solved to estimate Zernikes. In fact, this is the main reason why regularization does improve the aberration reconstruction results obtained with the current reticles. Apart from selecting the optimal PSG targets, there is also still room to optimize the targets themselves. It was discussed in Subsec. 3.D that, although the PSG designs are optimized to compensate mask 3D-effects, there will always remain differences between the design and realized features, leading to non-zero forbidden diffraction orders. An experimental optimization approach, in which PSGs having slightly different design parameters are evaluated in terms of their measured forbidden order strength, could be a way to compensate these manufacturing defects which are hard to parameterize.

Parallel to improving the PSG set and the individual designs, also the analysis and corrections presented in this paper can be further improved. For example, the forbidden order correction mentioned in Subsec. 3.D was based on AIMS measurements on just a subset off all PSGs, with the missing data being filled in by extrapolated values. Also, determining the target specific variance for the overlay via a large number of experiments, which then serves as input to the generalized least-squares problem, will contribute to an improved stability of the PSG-OVL method.

Altogether, we are convinced that, with a set of new optimized measurement reticles, complete reticle metrology data and accurate knowledge of the stochastic overlay behavior of the PSGs, the aimed aberration reconstruction accuracy for the method presented in this paper is well in reach. Consequently, we see this method as an excellent candidate to study dynamic aberration effects occurring during exposure of a reticle and to serve as an independent reference for on-tool aberration sensors.

Appendix A: Conversion between Z_t and Z_n^m

In this appendix we provide a conversion rule to go from a Zernike function in single index notation, Z_t, (FRINGE convention) to a Zernike function in double index notation, Z_n^m, and vice versa.

a. Conversion: $Z_t \rightarrow Z_n^m$

For $t = 1,\ 2,\ \cdots$, we let

$$q = \left\lfloor \sqrt{t-1} \right\rfloor , \quad p = \left\lfloor \frac{t - q^2 - 1}{2} \right\rfloor , \quad \text{(A1)}$$

and

$$n = q + p , \quad m = q - p , \quad \text{(A2)}$$

then a single index Zernike function, $Z_t(\rho, \theta)$ equals

$$
\begin{aligned}
Z_t(\rho, \theta) = R_n^m(\rho) &\left\{ (t - q^2 - 2p) \cos m\theta \right. \\
&\left. + \left[1 - (t - q^2 - 2p) \right] \sin m\theta \right\} \\
= Z_n^m(\rho, \theta) .& \quad \text{(A3)}
\end{aligned}
$$

b. Conversion: $Z_n^m \rightarrow Z_t$

When n, m are positive integers such that $n - m$ is even and non-negative, then

$$R_n^m(\rho) \cos m\theta = Z_{(\frac{n+m}{2})^2 + n - m + 1} , \quad \text{(A4)}$$

$$R_n^m(\rho) \sin m\theta = Z_{(\frac{n+m}{2})^2 + n - m + 2} , \quad \text{(A5)}$$

with Eq.(A5) only valid for $m \neq 0$.

Appendix B: 2D-PSG unit cell definition

In this appendix we discuss the relation between the 2D-PSG unit cell definition, its corresponding pupil positions of its allowed diffraction orders and the parameters we use to define them.

We use 4 parameters that completely define the 2D-PSG targets used by our method; the pitch, P, of the resulting 1D-pattern generated from the interference between the two orders close the pupil edge, the distance ρ_1 between the pupil center and the allowed diffraction order closest to the pupil edge, the fraction ξ being the ratio between ρ_3 and ρ_0 that are defined in Fig. 19 and finally the grating orientation ϕ. All distances defined
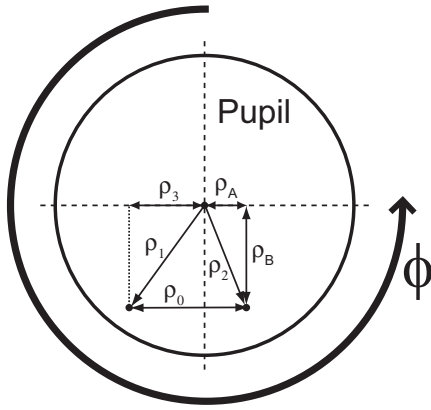
Fig. 19. Definition of the relevant distances in reciprocal (pupil) space.

in Fig. 19 are then given by

$$\rho_0 = \frac{\lambda}{NA\,P}\,, \tag{B1}$$

$$\rho_1 = \rho_1\,, \tag{B2}$$

$$\rho_2 = \sqrt{(1-\xi)^2\rho_0^2 + \rho_1^2 - \xi^2\rho_0^2}\,, \tag{B3}$$

$$\rho_3 = \xi\rho_0\,, \tag{B4}$$

$$\rho_A = (1-\xi)\rho_0 \tag{B5}$$

$$\rho_B = \sqrt{(\rho_1^2 - \xi^2\rho_0^2)}\,, \tag{B6}$$

where λ and NA are again the wavelength and numerical aperture of the lithographic system.
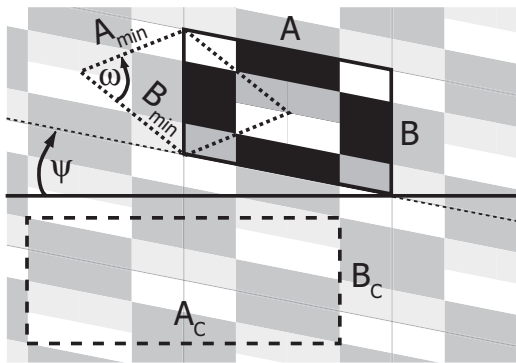

Fig. 20. Definition of the real space variables and relation between the minimum unit cell (dotted line), computational unit cell (dashed line) and the unit cell as defined in Fig. 6.

Next, we can also express the real-space parameters defined in Fig. 20 in terms of the 2D-PSG parameters

and the distances defined in Fig. 19

$$\omega = \arccos\left[\frac{\rho_1^2 + \rho_2^2 - \rho_0^2}{2\rho_1\rho_2}\right]\,, \tag{B7}$$

$$\Psi = \arccos\left[\frac{\rho_0^2 + (2\sin\left(\frac{\omega}{2}\right)\rho_2)^2 - (\rho_1-\rho_2)^2}{4\sin\left(\frac{\omega}{2}\right)\rho_0\rho_2}\right]\,, \tag{B8}$$

$$A_{\min} = \frac{\lambda}{2NA\,\rho_1}\sqrt{\left(\frac{1}{\cos\left(\frac{\omega}{2}\right)}\right)^2 + \left(\frac{1}{\sin\left(\frac{\omega}{2}\right)}\right)^2}\,, \tag{B9}$$

$$B_{\min} = \frac{\rho_1}{\rho_2}A_{\min}\,, \tag{B10}$$

$$A_C = \frac{\Xi\lambda}{NA\,\xi\rho_0}\,, \tag{B11}$$

$$B_C = \frac{\lambda}{NA\,\sqrt{\rho_1^2 - \xi^2\rho_0^2}}\,, \tag{B12}$$

where ξ is a rational number between $\frac{1}{2} \le \xi \le 1$ and Ξ is the numerator of the smallest possible rational representation of $\left(\frac{\xi}{1-\xi}\right)$. Note that the angles ω and Ψ, together with the minimum unit cell rib lengths, are sufficient to define a 2D-PSG in a reticle design. However, the super cell defined by A_C and B_C is also provided because lithographic simulators commonly require such a manhatten like unit-cell to simulate the proposed 2D-PSGs.

For a given 2D-PSG design to be applicable, it should satisfy similar requirements as in the 1D-PSG case; non-forbidden higher orders should remain outside the pupil and the effective pitch of the grating image should be observable by the DBO-tool. For a 2D-PSG this translates into the following constraints

$$\frac{2\lambda}{NA\,(1+\sigma)} \le P \le \frac{\lambda_{\text{DBO}}}{CNA_{\text{DBO}}}\,, \tag{B13}$$

$$\sqrt{(1+\sigma)^2 - (4-4\xi)\left(\frac{\lambda}{NA\,P}\right)^2} \le \rho_1 \le 1\,, \tag{B14}$$

$$1 - \frac{\left((1+\sigma)^2 - \rho_1^2\right)(NA\,P)^2}{4\lambda^2} \ge \xi \ge \frac{1}{2}\,. \tag{B15}$$

Appendix C: Semi-analytic computation of phase difference averages

After linearizing the exponential comprising the phase difference, it is required to evaluate the average of the phase difference over a specific subset of the reference pupil. This subset consists of the intersection region of three disks, viz. the pupil disk, a displaced copy of the pupil disk and the supporting disk of the scanner's illumination monopole. For the sake of mathematical convenience we will use in this appendix the double index, complex exponential representation of the Zernike functions and we assume that the disk pertaining to the scanner monopole has radius 1 and center $\underline{o}$ while the two pupil disks have radius r and centers $\underline{\nu}_1$ and $\underline{\nu}_2$, respectively. See Fig. 21 for the configuration in which the hatched region S is the averaging region.
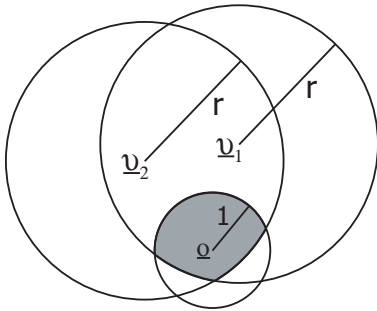
The mathematical problem at hand thus consists of

Fig. 21. Integration range S in Eq. (C1) consisting of the intersection of two pupil disks centered at $\underline{\nu}_1$ and $\underline{\nu}_2$ and a third disk centered at $\underline{o}$ defined by the scanner's illumination monopole, with choice of origin and radii facilitating solving the mathematical problem.

solving the integrals

$$\iint_S Z_{n'}^{m'}\left(\frac{\nu-\nu_{1,2}}{r}\right)\,\mathrm{d}\underline{\nu}\,, \qquad (C1)$$

where $Z_{n'}^{m'}$ denotes the complex Zernike circle polynomial vanishing outside the unit disk, given as

$$Z_{n'}^{m'}(\underline{\nu}) \equiv Z_{n'}^{m'}(\rho e^{i\theta}) = R_{n'}^{|m'|}(\rho)\,e^{im'\theta}\,,$$
$$0 \le \rho \le 1\,,\ 0 \le \theta \le 2\pi\,, \qquad (C2)$$

for integer n', m' such that $n'-|m'|$ is even and non-negative. The evaluation of (C1) can be done using the approach in [8]. Accordingly, focusing on the case $\underline{\nu}_1$, we have

$$\iint_S \quad Z_{n'}^{m'}\left(\frac{\nu-\nu_1}{r}\right)\,\mathrm{d}\underline{\nu}$$
$$= \iint_{\underline{\nu}\le 1} Z_{n'}^{m'}\left(\frac{\nu-\nu_1}{r}\right) Z_0^0(\underline{\nu})\left(Z_0^0\left(\frac{\nu-\nu_2}{r}\right) Z_0^0(\underline{\nu})\right)^*\,\mathrm{d}\underline{\nu}$$
$$= \sum_{n,m}\frac{\pi}{n+1}\,\beta_{n,1}^m\,(\beta_{n,2}^m)^*\,, \qquad (C3)$$

with $\beta_{n,1}^m$ and $\beta_{n,2}^m$ the Zernike coefficients with respect to the unit disk of $Z_{n'}^{m'}\left(\frac{\nu-\nu_1}{r}\right) Z_0^0(\underline{\nu})$ and $Z_0^0\left(\frac{\nu-\nu_2}{r}\right) Z_0^0(\underline{\nu})$, respectively. These β's can be cast into the form of a correlation integral as in [8], Eq. (8),

$$\beta_{n,1}^m = \frac{n+1}{\pi}\iint Z_{n'}^{m'}\left(\frac{\nu-\nu_1}{r}\right)\left(Z_n^m(-\underline{\nu})\right)^*\,\mathrm{d}\underline{\nu}\,, \qquad (C4)$$

$$\beta_{n,2}^m = \frac{n+1}{\pi}\iint Z_0^0\left(\frac{\nu-\nu_2}{r}\right)\left(Z_n^m(-\underline{\nu})\right)^*\,\mathrm{d}\underline{\nu}\,, \qquad (C5)$$

and these correlation integrals have been computed in [8], Eq. (9) and Thm. 2.1. Thus (choosing $c' = r$, $c = 1$, $(\tau',\eta') = -\underline{\nu}_1$, $(\tau,\eta) = \underline{(o)}$)

$$\beta_{n,1}^m = \frac{n+1}{\pi}\sum_{n''} C_{n'n,n''}^{m'm}\,Z_{n''}^{m'-m}\left(\frac{-\nu_1}{r+1}\right)\,, \quad |\underline{\nu}_1| \le r+1\,, \qquad (C6)$$

and a similar formula for $\beta_{n,2}^m$. The summation in Eq. (C6) is over all integers n'' such that $n'' - (n'+n)$ is even and non-negative. The C's are given as

$$C_{n'n,n''}^{m'm} = \left(\frac{r}{r+1}\right)^2 \frac{(-1)^n (n''+1)\,\pi}{(n'+1)(n+1)} \times \qquad (C7)$$
$$\left[S_{n'n}^{n''+1} - S_{n'+2,n}^{n''+1} - S_{n',n+2}^{n''+1} + S_{n'+2,n+2}^{n''+1}\right]\,,$$

with

$$S_{ij}^{k+1} = \frac{(\frac{1}{2}(k+i+j))!\,(\frac{1}{2}(k-i-j))!}{(\frac{1}{2}(k+i-j))!\,(\frac{1}{2}(k-i+j))!} \times$$
$$\frac{r^i}{(r+1)^{i+j}}\left(P_{\frac{1}{2}(k-i-j)}^{(i,j)}\left(\frac{1-r}{1+r}\right)\right)^2\,, \quad (C8)$$

for integers i, j, $k \ge 0$ such that $k-i-j$ is non-negative and even and $S_{i,j}^{k+1} = 0$ otherwise. The $P_l^{(\alpha,\beta)}(x)$ are Jacobi polynomials, see [9], Ch. 22.

The series in Eqs.(C3) and (C6) are rather slowly convergent, and so the computations must be done efficiently. The S_{ij}^{k+1} can be expressed in terms of the generalized Zernike functions of [10] according to

$$\frac{r^i}{(r+1)^{i+j}}\left(P_{\frac{1}{2}(k-i-j)}^{(i,j)}\left(\frac{1-r}{1+r}\right)\right)^2 = (1-\rho^2)^{-\frac{i}{2}}\,R_{k-i}^{j,i}(\rho)\,, \qquad (C9)$$

with $\rho = (1+r)^{-1/2} \in (0,1)$. Finally, according to [10], Thm. 5.2 (correcting two minor typos) the generalized Zernike functions can be computed in a DFT-format as follows. Let n and m be integers such that $n-|m|$ is even and non-negative, and let $q = \frac{1}{2}(n+|m|)$. Furthermore, let $\alpha > -1$ and denote by $C_n^{\alpha+1}$ the Gegenbauer polynomial of [9] Ch. 22. Then for any integer N such that $N > n + |m|$, we have

$$\binom{q+\alpha}{q}(1-\rho^2)^{-\alpha}\,R_n^{|m|,\alpha}(\rho) = \qquad (C10)$$
$$\frac{1}{N}\sum_{k=0}^{N-1} C_n^{\alpha+1}\left(\rho\cos\frac{2\pi k}{N}\right) e^{-2\pi i\frac{km}{N}}\,.$$

The prefactors

$$a_k = \frac{(\frac{1}{2}(k+i+j))!\,(\frac{1}{2}(k-i-j))!}{(\frac{1}{2}(k+i-j))!\,(\frac{1}{2}(k-i+j))!} \qquad (C11)$$

in Eq. (C8) should be computed for all k such that $k-i-j$ is even and non-negative. We can do that recursively according to

$$a_{k=i+j} = \binom{i+j}{i}\,, \qquad (C12)$$

and for $k = i+j,\ i+j+2,\ \cdots$

$$a_{k+2} = \frac{(\frac{1}{2}(k+i+j)+1)(\frac{1}{2}(k-i-j)+1)}{(\frac{1}{2}(k+i-j)+1)(\frac{1}{2}(k-i+j)+1)}\,a_k\,. \qquad (C13)$$

References

[1] M. van de Kerkhof, W. de Boeij, et al., "Full optical column characterization of DUV lithographic projection tools," Proc. SPIE **5377**, p. 1960-1970 (2004).

[2] H. Nomura, "New phase-shift gratings for measuring aberrations," Proc. SPIE **4346**, 25-35 (2001).

[3] C. Kittel, *Introduction to Solid State Physics, 5th edition* (John Wiley & Sons, New York, 1976).

[4] P. Vanoppen, T. Theeuwes, "Lithographic scanner stability improvements through advanced metrology and control," Proc. SPIE **7640**, 764010 (2010).

[5] M. Ebert, H. Cramer, et al., "Combined Overlay, Focus and CD Metrology for leading Edge Lithography," Proc. SPIE **7973**, 797311 (2011).

[6] T. Hastie, R. Tibshirani and J. J. H. Friedman, *The elements of statistical learning. Vol. 1.* (Springer, New York, 2001).

[7] M. Pisarenco, I. Setija, "Compact discrepancy and chi-squared principles for over-determined inverse problems", submitted to Inverse Problems (Dec 2013).

[8] A. J. E. M. Janssen, "Computation of Hopkins' 3-circle integrals using Zernike expansions," J. Europ. Opt. Soc. Rap. Public. **6**, 11059 (2011).

[9] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover Publications, New York, 1972).

[10] A. J. E. M. Janssen, "A generalization of the Zernike circle polynomials for forward and inverse problems in diffraction theory," arXiv: **1110.2369v1** (2011).