Spatially varying aberration calibration using a pair of matched periodic pinhole array masks

Shao, Yifeng; Loktev, Mikhail; Tang, Ying; Bociort, Florian; Urbach, H. Paul

DOI
10.1364/OE.27.000729

Publication date
2019

Document Version
Final published version

Published in
Optics Express

Citation (APA)
Shao, Y., Loktev, M., Tang, Y., Bociort, F., & Urbach, H. P. (2019). Spatially varying aberration calibration using a pair of matched periodic pinhole array masks. *Optics Express*, *27*(2), 729-742. https://doi.org/10.1364/OE.27.000729

Research Article | Vol. 27, No. 2 | 21 Jan 2019 | OPTICS EXPRESS 729
Optics EXPRESS

Spatially varying aberration calibration using a pair of matched periodic pinhole array masks

YIFENG SHAO,[1,*] MIKHAIL LOKTEV,[2] YING TANG,[1] FLORIAN BOCIORT[1], AND H. PAUL URBACH,[1]

[1]*Optics Research Group, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands*
[2]*Kulicke & Soffa Liteq BV, Esp 314, 5633 AE Eindhoven, The Netherlands*
[*]*y.shao@tudelft.nl*

Abstract: For advanced imaging systems, e.g., projection systems for optical lithography, spatially varying aberration calibration is of utmost importance to achieve uniform imaging performance over the entire field-of-view (FOV). Here we present an efficient, accurate, and robust spatially varying aberration calibration method using a pair of 2-dimensional periodic pinhole array masks: the first mask in the object plane and the second mask in the image plane. Our method divides the entire FOV of the imaging system into partially overlapping subregions by using a measurement system consisting of an additional imaging system and a camera sensor. Each subregion, which covers several mask periods, is imaged onto a distinct camera pixel by the measurement system. Our method measures "Airy disc"-like patterns simultaneously in all subregions by scanning the second mask relative to the first mask over one mask period. The number of subregions is equal to the number of camera pixels, and the sampling number of the measured patterns is equal to the scanning step number. The aberrations can be retrieved from the patterns measured in through-focus planes using an iterative optimization algorithm. In this paper, we performed experimental validation on a realistic lithography machine and demonstrate that our method is capable of retrieving the coefficients of 37 aberration terms, expressed as Zernike polynomials, with a sensitivity at nanometer scale.

1. Introduction

The function of an imaging system is to generate a perfect image (an Airy disc) of every point source in its field-of-view (FOV). However, the wavefront error in the pupil of the imaging system produces the aberrations of various types that blur the images. For large FOV imaging system such as the one we studied in this paper, the wavefront error depends on the location of the point source and hence varies spatially over the entire FOV of the imaging system. Spatially varying aberration calibration is therefore particularly important for essential industrial applications, e.g. projection systems for optical lithography, to achieve uniform imaging performance over the entire FOV.

Traditional methods, such as interferometry [1, 2] and wavefront sensing [3, 4], require a point source in the object plane for providing an ideal reference wavefront (either planar or spherical), and an extra imaging system to image the wavefront in the pupil onto a detector. The extra imaging system has its own aberrations that will be mixed with the aberrations of the imaging system that is to be calibrated. To measure the spatial variation of the aberrations, traditional methods require to position the point source at a large number of locations in the FOV and to perform measurement for one location at a time. Therefore, the whole measurement process is time-consuming. Furthermore, the movements can introduce positioning errors that will influence the calibration of the aberrations.

Alternative methods are based on the blurred images. For imaging systems such as microscopes and telescopes, one can use the blurred images measured by the camera sensor to retrieve the aberrations directly [5–9]. For lithography systems, since the features of the patterns printed

#345500 https://doi.org/10.1364/OE.27.000729
Journal © 2019 Received 12 Sep 2018; revised 7 Nov 2018; accepted 19 Nov 2018; published 9 Jan 2019

Optics EXPRESS

on the wafer are usually very small, one needs to retrieve the blurred features of the printed patterns indirectly using wafer metrology techniques like scatterometry [10] or using aerial image sensors [11, 12]. However, to avoid the overlap between the printed patterns, the characterization of the aberrations can only be performed at very few FOV locations in parallel. It is also reported that scanning electron microscopy can be used [13, 14] to acquire high-resolution images of the blurred printed patterns in a photoresist, but the acquisition of these images in the entire FOV is extremely time-consuming. The influence of the photo-resist on imaging must also be modeled.

In industrial applications, an efficient, accurate and robust spatially varying aberration calibration method is desired because spatially varying aberrations may be caused by e.g. heating, vibration or contamination of the optical elements during operation. Imaging systems need to be calibrated and corrected occasionally to maintain a good imaging performance. In this paper, we therefore propose a spatially varying aberration calibration method, which uses a pair of 2-dimensional periodic pinhole array masks, with the following features:

- The entire FOV is divided into partially overlapping subregions by the measurement system. Each subregion covers several mask periods, and in each subregion, the spatially varying aberrations can be considered to be shift-invariant.
- The intensities (in analogy to the point-spread function (PSF) intensities) of the transmitted images are measured simultaneously in all subregions via a scanning process, and the scanning process is repeated in several through-focus planes. At least two of these planes are needed to be able to retrieve the phase (the aberrations) in the pupil.
- The shift-invariant aberrations in each subregion are retrieved from the measured through-focus intensities, and are combined to determine the spatially varying aberrations in all subregions.

Spatially varying aberrations are represented by a 4-dimensional function of both the FOV coordinates $\mathbf{r}_c$ and the pupil coordinates $\boldsymbol{\rho}$. In order to determine this function, in each subregion corresponding to a FOV location indicated by $\mathbf{r}_c$, we acquire a set of 2-dimensional intensity distributions (each intensity distribution is a function of the scanning positions $\mathbf{r}_s$) in several through-focus image planes at various locations along the z coordinate. We remark that the samplings of $\boldsymbol{\rho}$ and $\mathbf{r}_s$ should obey the Shannon-Nyquist sampling theorem. So, we actually need to deal with a 5-dimensional dataset of $(\mathbf{r}_c, \mathbf{r}_s, z)$.

Our method provides an attractive scheme for parallel data collection and processing. We utilize in particular the fact that the required sampling number of the intensity distributions in each subregion (which equals to the scanning step number) is much less than the number of subregions (which equals to the number of camera pixels). Therefore, it is more efficient to use a scanning process to measure the intensities at one point in every subregions simultaneously, than to measure the entire intensity distribution in every subregions one by one.

Our method has been presented previously in [15] for calibrating only the geometrical aberration terms such as distortion, field curvature and telecentricity (tip-tilt). Here we calibrate the wavefront aberrations in terms of the coefficients of the Zernike polynomials. Namely, we determine a set of coefficients, each of which is a function of FOV location, of the spatially varying aberrations. Our method requires the imaging system to be telecentric (the magnification/demagnification should be constant in through-focus planes). By adapting the design of the masks and the scanning process, our method can be applied to arbitrary wavelengths.

This paper is organized as follow. We describe the concept and the experimental setup of our method in Section 2. We then discuss the computation of the point-spread function (PSF) in Section 3, and the optimization algorithm in Section 4. Finally we validate our method by simulations in Section 5 and by experiments in Section 6.

Optics EXPRESS

2. The concept and the mathematical description

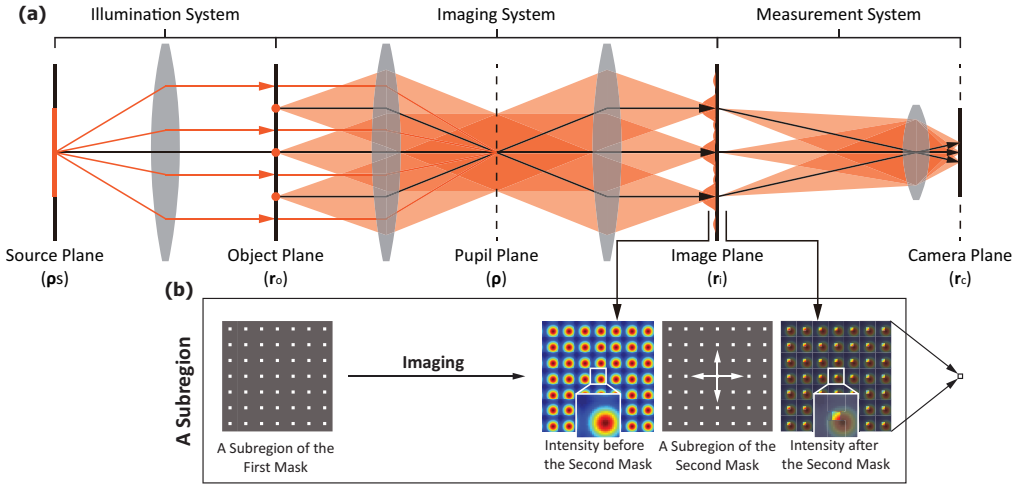

Fig. 1. Experimental setup and concept of the method. **(a)**: Schematic plot of the experimental setup. A lithographic system consists of a illumination system (Kohler illumination) and a telecentric imaging system. The first mask and the second mask are in the object plane and in the image plane, respectively. The measurement system consists of an additional imaging system and a camera sensor, of which the second mask is in the object plane. **(b)**: Demonstration of the imaging and measurement process in a subregion, which is defined as a region the first mask that is imaged by the combined lithographic system and measurement system onto a single pixel of the camera. The second mask is superposed with the image of the first mask. Because their pitches are identical, a point of the "PSF-like" pattern in each period of the transmitted image is measured (see the insertions) by the pinhole in each period of the second mask. To measure the entire "PSF-like" pattern, the second mask needs to be scanned over a one period, along the directions depicted by the arrows, relative to the image of the first mask.

A schematic plot of the experimental setup is shown in Fig.1(a). We consider a lithographic system which consists of a Köhler illumination system and a telecentric imaging system. We place the first mask in the object plane and the second mask in the image plane . Both masks are transmissive 2-dimensional periodic pinhole array masks. The measurement system consists of an additional imaging system and a camera sensor. The object plane of the measurement system coincides with the second mask. In this way, the entire FOV of the imaging system is divided into partially overlapping subregions by the measurement system. A subregion is defined as a region in the first mask that is imaged by the combined lithographic system and the measurement system onto a distinct pixel of the camera sensor. We need to guarantee that on one hand each subregion covers several mask periods, and on the other hand the spatially varying aberrations in each subregion can be considered to be shift-invariant.

The imaging and measurement process of one subregion is illustrated in Fig. 1(b). The second mask is superposed with the image of the first mask. Because their pitches are identical, measuring the total intensity of the transmitted image in one subregion is equivalent to sampling the image of a pinhole in the first mask by another pinhole in the second mask as shown by the insertions in Fig. 1(b). We remark that the sampled intensity in our method is enhanced by several folds compared to the case when one subregion covers only one mask period. The folds of enhancement is equal to the number of mask period covered by one subregion.

In order to measure the entire "PSF-like" pattern, we need to scan the location of the sampling

Optics EXPRESS

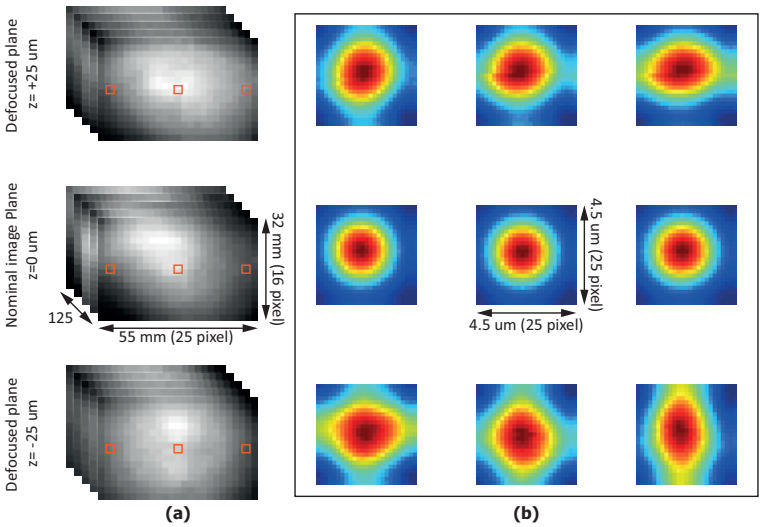

Fig. 2. Demonstration of the scanning measurement process in various through-focal planes using experiment data. **(a)**: The sequence of camera measurements at 125 scanning positions in the focal plane (z = 0 um) and two defocused planes (z = ±25 um). The FOV of the lithographic system with size 32 mm × 55 mm is projected onto the camera sensor. This area is sampled by 16 × 25 pixels as we downsample the original camera measurements by a factor of 20. **(b)**: The PSF-like" patterns measured by the corresponding pixels depicted by the orange squared boxes in **(a)**. Each pattern with size 4.5 μm × 4.5 μm is sampled by 25 × 25 scanning positions. The patterns measured by different pixels show significantly more difference in the defocused planes than in the focal plane.

point over one period of the mask period. This is achieved by scanning the second mask relative to the first mask. Since the image of the first mask stays at the same position with respect to the camera sensor while scanning the second mask, each subregion is always imaged onto the same camera pixel. Measuring "PSF-like" patterns by scanning is advantageous because we avoid mixing of the aberrations of the measurement system with the aberration of the lithographic system. Furthermore, the requirement on the measurement system in our experimental setup is much less critical compared to the imaging system that is used to magnify the "PSF-like" pattern since the "PSF-like" patterns is typically very small.

By repeating the scanning process in several through-focus planes, we thus measure through-focus "PSF-like" patterns in each subregion (see Fig. 2 and Visualization 1). As mentioned in the Introduction, at least two of these planes are needed to be able to retrieve the phase (the aberrations) in the pupil. Retrieving the aberrations in each subregion only requires the "PSF-like" pattern to be sampled by hundreds of points. Therefore, by taking hundreds of camera measurements in one scanning process, we can simultaneously measures "PSF-like" pattern in millions of subregions, which means that we sample the entire FOV of the imaging system by millions of points. Therefore, by combining the shift-invariant aberrations retrieved from the through-focus measurements in all subregions, the spatially varying aberrations can be calibrated with high-resolution in both FOV coordinates and pupil coordinates.

2.1. Mathematical description

In this paper, we model spatially partially coherent imaging and the scanning measurement process using the mutual coherent function (MCF), which represents the correlation between monochromatic fields at a pair of locations. Let the MCF in the object plane and the image plane

Optics EXPRESS

be denoted by $W_o(\mathbf{r}_{o1}, \mathbf{r}_{o2})$ and $W_{i,z}(\mathbf{r}_{i1}, \mathbf{r}_{i2})$ respectively, where $\mathbf{r}_o$ and $\mathbf{r}_i$ denote the coordinates of the relative planes, and z denotes the location of the image plane ($z = 0$ denotes the location of the focal plane). The relation between $W_o(\mathbf{r}_{o1}, \mathbf{r}_{o2})$ and $W_{i,z}(\mathbf{r}_{i1}, \mathbf{r}_{i2})$ is given by the Hopkin's formula for spatially partially coherent imaging [16] as follows:

$$W_{i,z}(\mathbf{r}_{i1}, \mathbf{r}_{i2}) = \iint W_o(\mathbf{r}_{o1}, \mathbf{r}_{o2}) T_o(\mathbf{r}_{o1}) T_o(\mathbf{r}_{o2})^* H_z(\mathbf{r}_{i1}, \mathbf{r}_{o1}) H_z(\mathbf{r}_{i2}, \mathbf{r}_{o2})^* \mathrm{d}^2\mathbf{r}_{o1} \mathrm{d}^2\mathbf{r}_{o2}, \quad (1)$$

where $T_o(\mathbf{r}_o)$ is the complex-valued transmission function of the first mask in the object plane and $H_z(\mathbf{r}_i, \mathbf{r}_o)$ is the PSF of the lithographic system in the plane at location z, which represents the field evaluated at $\mathbf{r}_i$ generated by a point source at $\mathbf{r}_o$.

Let the coordinates of the camera plane be denoted by $\mathbf{r}_c$. The intensity distribution measured by the camera is given by

$$I_z(\mathbf{r}_c) = \iint W_{i,z}(\mathbf{r}_{i1}, \mathbf{r}_{i2}) T_i(\mathbf{r}_{i1}) T_i(\mathbf{r}_{i2})^* H'(\mathbf{r}_c, \mathbf{r}_{i1}) H'(\mathbf{r}_c, \mathbf{r}_{i2})^* \mathrm{d}^2\mathbf{r}_{i1} \mathrm{d}^2\mathbf{r}_{i2}, \quad (2)$$

where $T_i(\mathbf{r}_i)$ is the complex-valued transmission function of the second mask in the image plane, and $H'(\mathbf{r}_c, \mathbf{r}_i)$ is the PSF of the measurement system that represents the contribution of the field at $\mathbf{r}_i$ to a pixel at $\mathbf{r}_c$. By substituting Eq. (1) into Eq. (2), we obtain

$$I_z(\mathbf{r}_c) = \iint W_o(\mathbf{r}_{o1}, \mathbf{r}_{o2}) T_o(\mathbf{r}_{o1}) T_o(\mathbf{r}_{o2})^* G_z(\mathbf{r}_{o1}, \mathbf{r}_c) G_z(\mathbf{r}_{o2}, \mathbf{r}_c)^* \mathrm{d}^2\mathbf{r}_{o1} \mathrm{d}^2\mathbf{r}_{o2}, \quad (3)$$

where

$$G_z(\mathbf{r}_o, \mathbf{r}_c) = \int H_z(\mathbf{r}_i, \mathbf{r}_o) T_i(\mathbf{r}_i) H'(\mathbf{r}_c, \mathbf{r}_i) \mathrm{d}^2\mathbf{r}_i. \quad (4)$$

In our method, we assume the measurement system to be shift-invariant in the entire FOV and the lithographic system to be shift-invariant only in each subregion. We can therefore write the PSF of the measurement system $H'(\mathbf{r}_c, \mathbf{r}_i)$, which determines the size of the image of each subregion in the image plane of the lithographic system, as $H'(\mathbf{r}_i)$, and write the PSF of the lithographic system $H_z(\mathbf{r}_i, \mathbf{r}_o)$ as $H_z(\mathbf{r}_i - \mathbf{r}_o; \mathbf{r}_c)$, which differs in subregions that are imaged onto pixels at different location $\mathbf{r}_c$. Consequently, we can rewrite Eq. (4) as

$$G_z(\mathbf{r}_o; \mathbf{r}_c) = \int H_z(\mathbf{r}_i - \mathbf{r}_o; \mathbf{r}_c) T_i(\mathbf{r}_i) H'(\mathbf{r}_i) \mathrm{d}^2\mathbf{r}_i. \quad (5)$$

We can compute Eq. (5) as a correlation using efficient numerical methods like the fast Fourier transform (FFT) algorithm. $G_z(\mathbf{r}_o; \mathbf{r}_c)$ can be considered as the effective PSF of the combined lithographic system and the measurement system, which is shift-invariant in each subregion, but depends on the location $\mathbf{r}_c$ of the pixel that each subregion is imaged onto.

2.2. Köhler illumination

Now suppose we scan the first mask relative to the second mask. We denote the intensity distribution measured by the pixel at $\mathbf{r}_c$ as function of the scanning positions $\mathbf{r}_s$ by $I_z(\mathbf{r}_s; \mathbf{r}_c)$. According to the Van Cittert-Zernike theorem [17], the MCF in the object plane generated by Kohler illumination is given by $W_o(\mathbf{r}_{o1}, \mathbf{r}_{o2}) = \hat{S}(\mathbf{r}_{o1} - \mathbf{r}_{o2})$, where $\hat{S}(\mathbf{r}_o)$ is the Fourier transform of a uniform planar monochromatic spatially incoherent source with intensity distribution $S(\mathbf{k})$, where $\mathbf{k}$ is the source plane coordinates. It is shown in Fig. 1(a) that $\mathbf{k}$ and $\mathbf{r}_o$, $\mathbf{r}_i$, $\mathbf{r}_s$ are Fourier transform coordinates. The measured intensity distribution $I_z(\mathbf{r}_s; \mathbf{r}_c)$ can be written as

$$I_z(\mathbf{r}_s; \mathbf{r}_c) = \iint \hat{S}(\mathbf{r}_{o1} - \mathbf{r}_{o2}) T_o(\mathbf{r}_{o1} - \mathbf{r}_s) T_o(\mathbf{r}_{o2} - \mathbf{r}_s)^* G_z(\mathbf{r}_{o1}, \mathbf{r}_c) G_z(\mathbf{r}_{o2}, \mathbf{r}_c)^* \mathrm{d}^2\mathbf{r}_{o1} \mathrm{d}^2\mathbf{r}_{o2}. \quad (6)$$

Optics EXPRESS

Computing the intensity distribution $I_z(\mathbf{r}_s, \mathbf{r}_c)$ using Eq. (6) is usually extremely time-consuming (a brief introduction of the computation of Eq. (6) can be found in [18, 19]). However, the computation can be significantly simplified in two special cases:

- Completely spatially coherent illumination when the source is so small that $\hat{S}(\mathbf{r}_{o1} - \mathbf{r}_{o2}) = 1$:
$$I_z(\mathbf{r}_s, \mathbf{r}_c) = \left| \int T_o(\mathbf{r}_o - \mathbf{r}_s) G_z(\mathbf{r}_o, \mathbf{r}_c) \mathrm{d}^2\mathbf{r}_o \right|^2, \tag{7}$$
- Completely spatially incoherent illumination when the source is so large that $\hat{S}(\mathbf{r}_{o1} - \mathbf{r}_{o2}) = \delta(\mathbf{r}_{o1} - \mathbf{r}_{o2})$:
$$I_z(\mathbf{r}_s, \mathbf{r}_c) = \int |T_o(\mathbf{r}_o - \mathbf{r}_s)|^2 |G_z(\mathbf{r}_o, \mathbf{r}_c)|^2 \mathrm{d}^2\mathbf{r}_o. \tag{8}$$

3. Computation of the point-spread function in the focal region

We the PSF of the lithographic system, which is associated with the intensity distribution measured by each camera pixel via the scanning process in the image plane with axial coordinate z, to be a complex-valued function of both the scanning positions $\mathbf{r}_s$ and FOV coordinates $\mathbf{r}_c$, which can be computed using the Debye diffraction integral [17] as follows:

$$H_z(\mathbf{r}_s, \mathbf{r}_c) = \mathcal{F}\left\{\exp[i2\pi\Phi(\boldsymbol{\rho}, \mathbf{r}_c)] \exp(-iz|\boldsymbol{\rho}|^2)\right\}(\mathbf{r}_s), \tag{9}$$

where $\mathcal{F}$ denotes the Fourier transform, $\boldsymbol{\rho}$ is the pupil coordinates normalized by pupil radius, and $\Phi(\boldsymbol{\rho}, \mathbf{r}_c)$ is the wavefront error that produces the shift-invariant aberrations in each subregion.

We can decompose the 4-dimensional wavefront error by the Zernike polynomials as follows:

$$\Phi(\boldsymbol{\rho}, \mathbf{r}_c) = \sum_{m,n} \alpha_n^m(\mathbf{r}_c) Z_n^m(\boldsymbol{\rho}), \tag{10}$$

where $\alpha_n^m(\mathbf{r}_c)$ denotes the aberration coefficient as function of the FOV coordinates $\mathbf{r}_c$, and $Z_n^m(\boldsymbol{\rho})$ denotes the Zernike polynomial [17, 20] as function of the normalized pupil coordinates $\boldsymbol{\rho}$. We aim to combine the aberration coefficients retrieved in each subregion to determine the spatial variation of each aberration term expressed as the Zernike polynomial. The relation between the Zernike polynomials and the primary Seidel aberrations is beneficial for diagnosing and correcting the lithographic system.

In Eq. (9), $\boldsymbol{\rho}$ and $\mathbf{r}_s$ are a pair Fourier transform conjugated coordinates. Their samplings must satisfy the Shannon-Nyquist sampling theorem. Note that Eq. (9) is valid only for imaging system with sufficiently low NA, i.e. NA ≤ 0.6. For imaging system with high NA, i.e. NA > 0.6, polarization effect needs to be taken into account.

4. The optimization algorithm for retrieving the aberration coefficients

In each subregion of the FOV, we retrieve the shift-invariant aberration from the through-focus intensity measurements by solving an non-linear optimization problem. We formulate this non-linear optimization problem by defining a cost function, which is the sum of the squared difference between the measurements $I_z(\mathbf{r}_s)$ and the predictions $I'_z(\mathbf{r}_s; \alpha_n^m)$ modeled using the guessed aberration coefficients α_n^m. We find the solution to this non-linear optimization by updating the guessed aberration coefficients α_n^m iteratively until the cost function reaches a minimum. The cost function is expressed by

$$\mathcal{L}(\alpha_n^m) = \sum_z \int \left[I_z(\mathbf{r}_s) - I'_z(\mathbf{r}_s; \alpha_n^m)\right]^2 \mathrm{d}^2\mathbf{r}_s. \tag{11}$$

Optics EXPRESS

Note that it is not sufficient to retrieve the aberration coefficients from only one measurement in the focal image plane ($z = 0$). Because we measure only the intensity of the Fourier transform of the pupil function, whose phase represents the wavefront error, measurements in the defocused planes ($z \neq 0$) are also required for uniquely determine the aberrations coefficients. Fig. 2 shows that the measurements at different FOV locations have significantly more difference in the defocused planes than in the focal plane. According to studies in [21, 22], the defocused planes are preferred to be located symmetrically on both sides of the focal plane. The z coordinate of the defocused planes should be sufficiently large that the correlation between the measurements are sufficient small as mentioned in [23, 24]. However, the z coordinate cannot be larger than $5\pi\lambda/(\pi \mathrm{NA}^2)$ since otherwise the Debye diffraction Eq. (9) will be invalid.

To minimize the cost function, we need to derive an analytical expression for the gradient of the cost function with respect to the aberration coefficients. By applying the chain rule to the derivation, we obtain:

$$\frac{\partial \mathcal{L}(\alpha_n^m)}{\partial \alpha_n^m} = -\sum_z \int 2\left[I_z(\mathbf{r}_s) - I_z'\left(\mathbf{r}_s; \alpha_n^m\right)\right] \frac{\partial I_z'(\mathbf{r}_s; \alpha_n^m)}{\partial \alpha_n^m} \mathrm{d}^2\mathbf{r}_s. \tag{12}$$

In this formula, since we illuminate the lithographic system by spatial incoherent light, the prediction of the measurement can be modeled by using Eq. (8), which is expressed as

$$I_z'(\mathbf{r}_s; \alpha_n^m) = |T_o(\mathbf{r}_s)|^2 * |G_z(\mathbf{r}_s)|^2, \tag{13}$$

where $*$ denotes the convolution and

$$G_z(\mathbf{r}_s) = H_z(\mathbf{r}_s, \alpha_n^m) \star [T_i(\mathbf{r}_s) H'(\mathbf{r}_s)], \tag{14}$$

where $\star$ denotes the correlation.

The gradient of the prediction with respect to the aberration coefficients is given by

$$\frac{\partial I_z'(\mathbf{r}_s; \alpha_n^m)}{\partial \alpha_n^m} = |T_o(\mathbf{r}_s)|^2 * 2\Re\left\{G_z(\mathbf{r}_s)^* \left\{\frac{\partial H_z(\mathbf{r}_s; \alpha_n^m)}{\partial \alpha_n^m} \star [T_i(\mathbf{r}_s) H'(\mathbf{r}_s)]\right\}\right\}, \tag{15}$$

where $\Re$ denotes the real part, * denotes the complex-conjugate, and the gradient of the PSF with respect to the aberration coefficients is given by

$$\frac{\partial H_z(\mathbf{r}_s; \alpha_n^m)}{\partial \alpha_n^m} = \mathcal{F}\left\{i2\pi Z_n^m(\boldsymbol{\rho}) \exp[i2\pi\Phi(\boldsymbol{\rho})] \exp(-iz|\boldsymbol{\rho}|^2)\right\}(\mathbf{r}_s). \tag{16}$$

The computation of the gradient of the cost function is the most time-consuming step of the optimization. The computation load is proportional to the number of aberration coefficients. We use an array of zeros as the initial guess of the aberration coefficients, and we find the aberration coefficients that minimize the cost function using the gradient-based optimization routine "fminunc" implemented in Matlab.

5. Simulation results

In this section, we perform a qualitative study on the determination of the number and the locations of the image planes required for retrieving aberration coefficients from through-focus measurements and on the optimization of the design of the mask for optimal sensitivity. Our study is based on simulations, in which we assume the first and the second masks consist of identical periodic squared pinhole array with width w and pitch p. In the simulation, we use 37 aberration coefficients pre-calibrated at seven FOV locations on a realistic lithography machine to simulate the measured intensity distributions. Each simulated measurements is normalized to unity and then convert to 16-bit precision data type to mimic the use of a CCD/CMOS camera sensor in the measurement process.

Optics EXPRESS

5.1. Determining the number and the locations of the measurement planes

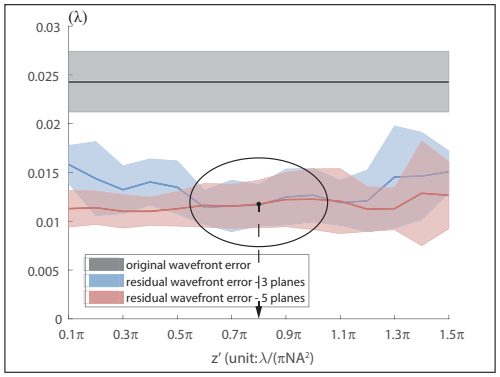

Fig. 3. Variation of the residual wavefront error versus the locations of the measurement planes. The gray graph is the mean and the standard deviation of the original wavefront error used for simulating the measurements. The blue and the red graphs are the mean and the standard deviation of the residual wavefront errors corresponding to the aberration coefficients retrieved from 7 sets of simulated measurements in three planes ($z = 0, \pm z'$) and in five planes ($z = 0, \pm 0.8\pi, \pm z'$) respectively. The vertical axis is in the unit of wavelength λ and the horizontal axis is the normalized z coordinate in the unit of $\lambda/(\pi \mathrm{NA}^2)$.

In the simulation, we use periodic squared pinhole array with pinhole size $w = 2.5$ μm and $p = 7.5$ μm. We scan the second mask relative to the image of the first mask by 20 steps with step size 375 nm in two orthogonal directions perpendicular to the optical axis with the second mask being placed in several through-focus planes with various z coordinates along the optical axis. We use the root-mean-square (RMS) of the residual wavefront error $\Delta\Phi(\boldsymbol{\rho})$ to evaluate the performance of the algorithm. $\Delta\Phi(\boldsymbol{\rho})$ is defined as the squared difference between the original wavefront error, which is used for simulating the measurements, and the wavefront error retrieved from the simulated measurements.

Suppose the measurements are taken in only three planes in the focal region. We set one plane to be the focal plane at $z = 0$ and the other two defocused planes to be located symmetrically on both sides of the focal plane at $z = \pm z'$, where z' is the z coordinate, normalized by $\lambda/(\pi \mathrm{NA}^2)$, that varies in the simulation. We plot the mean and the standard deviation of the RMS of $\Delta\Phi(\boldsymbol{\rho})$ for 7 sets of aberration coefficients as function of z' ranging from 0 to 1.5π in Fig. 3 (blue curve). We can observe that performance of the optimization algorithm is the optimal when we set the two defocused planes to be located in the vicinity of $z' = \pm 0.8\pi$ (see the circle in Fig. 3), where the RMS residual wavefront error is minimal.

We keep the three measurements taken in planes at $z = 0$ and $z = \pm 0.8\pi$. We simulate two additional measurements in two defocused planes located still symmetrically on both sides of the focal plane at $z = \pm z'$. So, the optimization algorithm now uses 5 measurements. The mean and the standard deviation of the RMS of $\Delta\Phi(\boldsymbol{\rho})$ for 7 sets of aberration coefficients as function of z' in Fig. 3 (red curve) shows that the RMS of $\Delta\Phi(\boldsymbol{\rho})$ is independent of locations of the two defocused planes. This indicates that using 5 measurements does not further minimize the RMS residual wavefront error compared to using 3 measurements. Therefore, it is sufficient for us to retrieve the aberration coefficients from measurements taken in only three optimally chosen through-focal planes: the focal plane $z = 0$, and two defocused planes at $z = \pm 0.8\pi$.

Optics EXPRESS

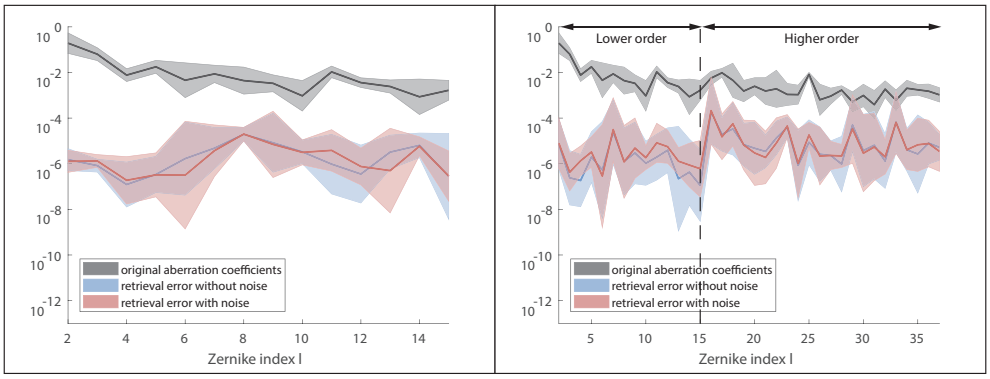

Fig. 4. Comparison of the errors of the retrieved aberration coefficients. The measurements are simulated using 37 aberration coefficients. The gray curve is the mean and the standard deviation of the original aberration coefficients. The red and the blue curves are the errors of the aberration coefficients retrieved from the noisy measurements and from the noise free measurements respectively. In this plot we use 7 sets of aberration coefficients. The vertical axis is in base-10 logarithmic scale and the horizontal axis is the Noll's Zernike index. **Left**: Results using the first 15 coefficients. **Right**:Results using the total 37 coefficients.

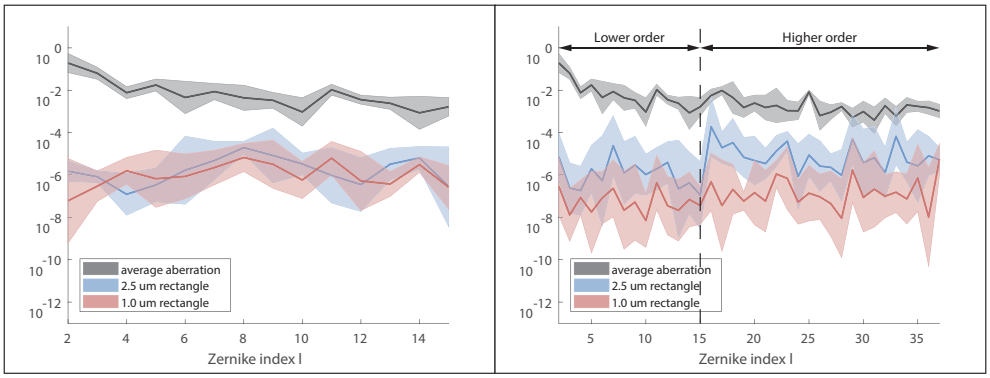

Fig. 5. Comparison of the errors of the retrieved aberration coefficients. The measurements are simulated using 37 aberration coefficients. The gray curve is the mean and the standard deviation of the original aberration coefficients. The red and the blue curves are the errors of the aberration coefficients retrieved from the measurements simulated using periodic squared pinhole array with pinhole size 1.0 μm and 2.5 μm respectively. In this plot we use 7 sets of aberration coefficients. The vertical axis is in base-10 logarithmic scale and the horizontal axis is the Noll's Zernike index. **Left**: Results using the first 15 coefficients. **Right**: Results using the total 37 coefficients.

5.2. Sensitivity analysis and the optimization of the mask design

In Fig. 4 and 5, we plot the retrieval errors of the aberration coefficients for each order of the Zernike polynomials. The retrieval error is defined as the squared difference between the original and the retrieved aberration coefficients. We adopt the Noll's index to enumerate the aberration coefficients. The mapping from Z_n^m, the Zernike polynomials with radial order n and azimuthal order m, to Z_ℓ, the Zernike polynomials with Noll's index ℓ, can be found in [25]. In both figures, we use 37 aberration coefficients to simulated the measurements, and we fit the simulated measurements by the first 15 aberration coefficients (left) and by the total 37 aberration coefficients (right) respectively.

Optics EXPRESS

In practical situations, the complete description of the wavefront error requires infinite number of aberration coefficients, but the optimization algorithm can only use finite number of aberration coefficients to fit the measurements. So the influence of truncating the aberration coefficients on the retrieval errors needs to be analyzed. In Fig. 4 and 5, the left panel shows that we can accurately retrieve only the first 15 aberration coefficients from the measurements simulated using total 37 aberration coefficients, and the right panel shows that although using the total 37 aberration coefficients to fit the measurements, the retrieval of the low order (first 15) aberration coefficients is still more than high order (the rest) aberration coefficients

To investigate the robustness of our method against the noise, we introduce Gaussian distributed random noise with zero mean and $10^{-}3$ variance to the simulated measurements. In Fig. 4, by comparing the aberration coefficients retrieved from the noisy measurements (red) and the noise free measurements (blue) in the left panel and in the right panel respectively, we find that the influence of the Gaussian noise on our algorithm is negligible, which may be because the Gaussian noise does not bias the measurements.

The design parameter of the mask are the pinhole size w and the pitch p. In Fig. 5, we plot the errors of the aberration coefficients retrieved from the measurements simulated using periodic squared pinhole array with pinhole size $w = 1.0$ μm (red) and $w = 2.5$ μm (blue) . The left plot shows that using a smaller pinhole cannot reduce the retrieval error when the optimization algorithm uses only the first 15 aberration coefficients, while the right plot shows that the retrieval error can be significantly reduced by using a smaller pinhole when the optimization algorithm uses the total 37 aberration coefficients.

6. Experimental results

In this section, we validate our method by performing an experiment on a realistic lithography machine. It is a projection lens with 1× magnification, NA 0.128 on both object side and image side, and operating wavelength 355 nm, which allows imaging with <2 μm critical dimension in a rectangle FOV with size 52 mm × 33 mm. The first mask and second mask consist of identical periodic arrays of squared pinholes with pinholes size $w = 2.5$ μm and pitch $p = 4.5$ μm. Both masks are fabricated in chrome on fused silica substrate using e-beam lithography with 25 nm position accuracy in a rectangle area with size 55 mm × 35mm.

The first mask, which is placed in the object plane of the projection lens, is illuminated with a Köhler illumination system. As a source, we use a Q-switched laser operating at 355 nm wavelength, 25 ns pulses and 200 kHz pulse repetition frequency. The beam is expanded to 2 inch diameter and scattered by a 20° engineered diffuser. A condenser lens with 200 mm focal length performs optical Fourier transform between the diffuser and the first mask, which are located in back and front focal planes of the condenser lens, respectively. The scattering angle is sufficient to fully illuminate the entire first mask, and the beam diameter is sufficient to overfill the pupil of the projection lens and make the illumination spatially incoherent.

During the scanning measurement process, the position of the first mask is fixed; the second mask, which is placed in the image plane of the projection lens, is movable, which is mounted on a piezo stage with 3 degrees of freedom: two orthogonal directions perpendicular to the optical axis and one direction along the optical axis. The piezo stage's closed-loop control provides absolute accuracy of 25 nm.

We scan the second mask relative to the first mask by 25 steps with step size 180 nm over one pitch along the two orthogonal directions perpendicular to the optical axis. The second mask is placed in 10 planes with with step size 5.56 μm, or 0.26π in normalized z coordinate, in the direction along the optical axis for taking the through-focus measurements.

In the experiment, the resulting MoirÃľ pattern is captured with a half inch size CCD camera equipped with a camera lens with 25 mm focal length. The FOV is divided into a large number of subregions. The number of subregions equals to the number of pixels in the area where the

Optics EXPRESS

camera is illuminated. Nevertheless, the optimization algorithm is time-consuming, which takes about few minutes on a normal desktop computer, so running optimization algorithm at every FOV locations is not feasible. We downsample camera measurements by a factor of 20 and the FOV is sampled by 16×25 pixels. Therefore, the retrieved aberration coefficients are 15 matrices (15 aberration terms) with size 16×25.

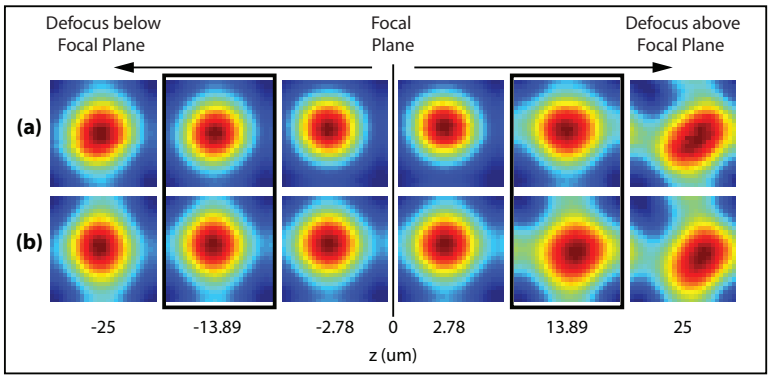

Fig. 6. Comparison between the experimental measurements **(a)** and the predictions calculated using the retrieved aberration coefficients **(b)**. We retrieve the first 15 aberration coefficients from 4 experimental measurements taken in 4 image planes at $z = \pm 2.78$ μm and $z = \pm 25$ μm respectively, and we then model the predictions in these 4 image planes and another 2 image planes at $z = \pm 13.89$ μm (in the black boxes).

In Fig. 6, we compare the experimental measurements and the predictions calculated using the retrieved aberration coefficients in one subregion (at one FOV location). The optimization algorithm takes 4 intensity distributions measured in 4 image planes located at $z = \pm 2.78$ μm ($\pm 0.13\pi$) and $z = \pm 25$ μm (1.15π) to retrieve the coefficients of the shift-invariant aberrations. We then use the retrieved aberration coefficients to model the predictions of the intensity distributions in these 4 image planes and another two image planes at $z = \pm 13.89$ μm (the intensity distributions in the black boxes). The comparison in Fig. 6 shows that the predictions are in good agreement with all 6 experimental measurements. In the optimization algorithm, we use a square as the PSF of the measurement system $H(\mathbf{r}_i)$ in Eq. (5). However, the measurement system is assumed to be shift-invariant and aberration-free, it might be more suitable to use a Bessel function (the Fourier transform of a circular pupil).

In Fig. 7 we made a comparison between the defocus aberration calibrated using our method and reference data (published in [15], which is in agreement with interferometry data). Defocus aberration is relative to the location the focal plane of the imaging system. In the experiment, the peizo translation stage has an absolute accuracy of 25 nm, so the error of the location of each image plane can be neglected. However, there might be a systematic error of their locations, which means that the focal plane may not be exactly at $z = 0$. This systematic error leads to an offset of the defocus aberration retrieved from the experimental through-focus measurements (blue curve). In order to perform a fair comparison, we assume that the defocus aberration calibrated using our method and the reference data have zero mean. In this way, we can correct for the offset of the retrieved defocus aberration (red curve), which is in good agreement with the reference defocus aberration (black curve). Fig. 7 also demonstrates that in order to determine the best focal plane precisely for lithography system, it is necessary to the defocus aberration at a large number of FOV locations since its spatial variation may be fast and dramatic.

Optics EXPRESS

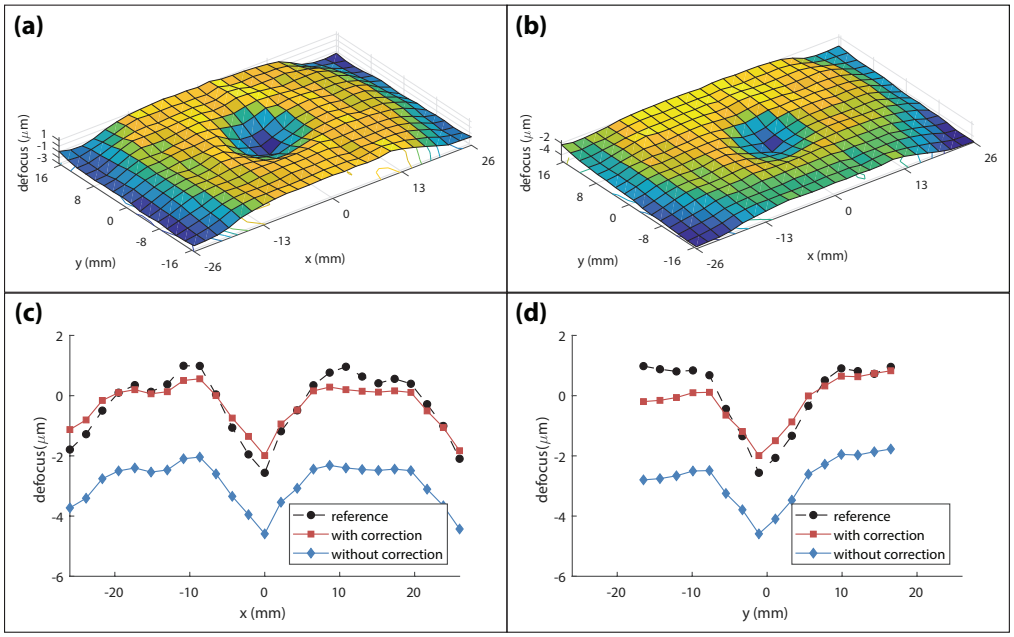

Fig. 7. Comparison between the reference defocus aberration data **(a)** and retrieved from scanning measurements using our method **(b)** in the full FOV. **(c)** and **(d)**: the x and y cross-section of the calibrated result (black curve) in **(a)** and the retrieved results (red/blue curve) in **(b)**. The defocus aberration is relative to the location of the focal plane. The focal plane's location error leads to an offset of the defocus aberration, which can be corrected manually.

7. Conclusion

In this paper, we developed and validated an efficient, accurate and robust method for calibrating the spatially varying aberrations of an imaging system. Our method does not require complex equipments, but only a pair of periodic pinhole array masks, a 3-dimensional precision translation stage, an additional imaging system and a camera sensor. By taking few hundreds of intensity measurements via a scanning process, we can simultaneously measure patterns that are in analogy to the PSF patterns at millions of FOV locations. We demonstrated that we can retrieve the coefficients of 37 aberration terms with a nanometer accuracy.

Because we use only binary masks with transmission/reflection being either 1 or 0, our method can be applied to almost any wavelength. At the current stage, we consider only a low NA imaging system. However, our method also has a potential to be extended to calibrating the spatially varying aberration of a high NA large FOV imaging system.

We recommend to use either complete spatially coherent or complete spatially incoherent illumination for the calibration. This is mainly due to the consideration of computation efficiency. In practice, spatially coherent illumination can always be generated by using a sufficiently small source. However, spatially incoherent illumination cannot be generated when the NA in the object side of the imaging system is larger than 0.5. In this situation, the condition of spatially incoherent imaging cannot be fulfilled even using an infinitely large source.

It is in principle to retrieve the information of source as well. However, this requires more through-focus measurements and consumes significantly more time. From our perspective, this can be surpassed by implementing parallel computation. In our method, it is possible to parallelize the computation of the gradient of the cost function with respect to the each aberration

Optics EXPRESS

coefficients and/or to parallelize the optimization at each FOV locations.

Regarding the shape of the pinhole in the periodic masks, a circular pinhole is easier to fabricate, however, since we use Cartesian coordinates in the computation, a circular pinhole will cause aliasing issue if the sampling of the measurements is not sufficiently high. In our method, the sampling number is equal to the scanning position number, and we want to scan as few positions as possible. So, a squared pinhole array mask is used in our method.

The design parameters of the pinhole array masks are the pinhole size and the pitch. The patterns measured by the scanning process can be computed by convolving the PSF of the lithographic system in each subregion with the correlation of two pinholes. Furthermore, because our method uses a pair of periodic pinhole array mask, the measured pattern suffers from the cross-talk effect. Therefore, we prefer a smaller pinhole size and a larger pitch.

The optimal design parameters and scanning settings, which depend on the imaging system that is to be calibrated and the wavelength, can be determined by simulations. We remark that the determination should base on a sufficiently large data base of all possible combinations of the values of the aberration coefficients to avoid bias.

Funding

Stichting voor de Technische Wetenschappen (12797).

Acknowledgments

This work is part of the research programme "Novel design shapes for complex optical systems" with project number 12797, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO).

References

1. W. J. Bates, "A wavefront shearing interferometer," Proc. Phys. Soc. **59**, 940 (1947).
2. R. Kingslake, "The interferometer patterns due to the primary aberrations," Transactions Opt. Soc. **27**, 94 (1925).
3. D. R. Neal, J. Copland, and D. A. Neal, "Shack-hartmann wavefront sensor precision and accuracy," in *Advanced Characterization Techniques for Optical, Semiconductor, and Data Storage Components,* vol. 4779 (International Society for Optics and Photonics, 2002), pp. 148–161.
4. B. C. Platt and R. Shack, "History and principles of shack-hartmann wavefront sensing," J. Refract. Surg. **17**, S573–S577 (2001).
5. R. A. Gonsalves, "Phase retrieval and diversity in adaptive optics," Opt. Eng. **21**, 215829 (1982).
6. J. R. Fienup, J. C. Marron, T. J. Schulz, and J. H. Seldin, "Hubble space telescope characterized by using phase-retrieval algorithms," Appl. Opt. **32**, 1747–1767 (1993).
7. J. E. Krist and C. J. Burrows, "Phase-retrieval analysis of pre- and post-repair hubble space telescope images," Appl. Opt. **34**, 4951–4964 (1995).
8. R. G. Paxman, T. J. Schulz, and J. R. Fienup, "Joint estimation of object and aberrations by using phase diversity," J. Opt. Soc. Am. A **9**, 1072–1085 (1992).
9. G. Zheng, X. Ou, R. Horstmeyer, and C. Yang, "Characterization of spatially varying aberrations for wide field-of-view microscopy," Opt. Express **21**, 15131–15143 (2013).
10. S. van Haver, W. M. J. Coene, K. D'havé, N. Geypen, P. van Adrichem, L. de Winter, A. J. E. M. Janssen, and S. Cheng, "Wafer-based aberration metrology for lithographic systems using overlay measurements on targets imaged from phase-shift gratings," Appl. Opt. **53**, 2562–2582 (2014).
11. T. Hagiwara, N. Kondo, I. Hiroshi, K. Suzuki, and N. Magome, "Development of aerial image based aberration measurement technique," in *Optical Microlithography XVIII,* vol. 5754 (International Society for Optics and Photonics, 2004), pp. 1659–1670.
12. T. Hagiwara, N. Kondo, I. Hiroshi, K. Suzuki, and N. Magome, "Development of aerial image based aberration measurement technique," in *Optical Microlithography XVIII,* vol. 5754 (International Society for Optics and Photonics, 2004), pp. 1659–1670.
13. C. Van der Avoort, J. J. M. Braat, P. Dirksen, and A. J. E. M. Janssen, "Aberration retrieval from the intensity point-spread function in the focal region using the extended nijboer–zernike approach," J. Mod. Opt. **52**, 1695–1728 (2005).
14. S. van Haver, J. J. Braat, P. Dirksen, and A. J. Janssen, "High-na aberration retrieval with the extended nijboer-zernike vector diffraction theory," J. Eur. Opt. Soc. publications **1** (2006).

Optics EXPRESS

15. M. Loktev and Y. Shao, "Projection lens testing with moiré effect," in *Metrology, Inspection, and Process Control for Microlithography XXXI,* vol. 10145 (International Society for Optics and Photonics, 2017), p. 101452S.
16. H. H. Hopkins, "On the diffraction theory of optical images," Proc. R. Soc. Lond. A **217**, 408–432 (1953).
17. M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light* (Elsevier, 2013).
18. Y. Lian and X. Zhou, "Fast and accurate computation of partially coherent imaging by stacked pupil shift operator," in *Photomask Technology 2009,* vol. 7488 (International Society for Optics and Photonics, 2009), p. 74883G.
19. X. Wu, S. Liu, W. Liu, T. Zhou, and L. Wang, "Comparison of three tcc calculation algorithms for partially coherent imaging simulation," in *Sixth International Symposium on Precision Engineering Measurements and Instrumentation,* vol. 7544 (International Society for Optics and Photonics, 2010), p. 75440Z.
20. R. K. Tyson, "Conversion of zernike aberration coefficients to seidel and higher-order power-series aberration coefficients," Opt. Lett. **7**, 262–264 (1982).
21. D. J. Lee, M. C. Roggemann, and B. M. Welsh, "Cramér–Rao analysis of phase-diverse wave-front sensing," J. Opt. Soc. Am. A **16**, 1005–1015 (1999).
22. B. H. Dean and C. W. Bowers, "Diversity selection for phase-diverse phase retrieval," J. Opt. Soc. Am. A **20**, 1490–1504 (2003).
23. O. El Gawhary, A. Wiegmann, N. Kumar, S. F. Pereira, and H. P. Urbach, "Through-focus phase retrieval and its connection to the spatial correlation for propagating fields," Opt. Express **21**, 5550–5560 (2013).
24. A. Polo, S. F. Pereira, and P. H. Urbach, "Theoretical analysis for best defocus measurement plane for robust phase retrieval," Opt. Lett. **38**, 812–814 (2013).
25. R. J. Noll, "Zernike polynomials and atmospheric turbulence," J. Opt. Soc. Am. **66**, 207–211 (1976).