Gradient descent algorithm applied to wavefront retrieval from through-focus images by an extreme ultraviolet microscope with partially coherent source

Kenji Yamazoe, Iacopo Mochi, and Kenneth A. Goldberg

Abstract

The wavefront retrieval by gradient descent algorithm that is typically applied to coherent or incoherent imaging is extended to retrieve a wavefront from a series of through-focus images by partially coherent illumination. For accurate retrieval, we modeled partial coherence as well as object transmittance into the gradient descent algorithm. However, this modeling increases the computation time due to the complexity of partially coherent imaging simulation that is repeatedly used in the optimization loop. To accelerate the computation, we incorporate not only the Fourier transform but also an eigenfunction decomposition of the image. As a demonstration, the extended algorithm is applied to retrieve a field-dependent wavefront of a microscope operated at extreme ultraviolet wavelength (13.4 nm). The retrieved wavefront qualitatively matches the expected characteristics of the lens design.

1. INTRODUCTION

The phase retrieval approach reconstructs the wavefront of an imaging system from image intensity measurements alone [1,2]. Usually, the input consists of a series of through-focus images since the evolution of the intensity through focus depends on the wavefront. To perform the phase retrieval, all we need is an image-capturing device. External phase measurement systems, such as Shack–Hartman test systems [3] or interferometry [4,5], are not required. Therefore, phase retrieval is compatible with quick wavefront measurement for instrumentation calibration and alignment [1]. This technique can be especially useful in fields in which independent wavefront measurements are difficult to achieve. Such is the case with optical systems operating in the extreme ultraviolet (EUV) wavelength range, where calibrated null lenses and optical systems of sufficient accuracy are unavailable.

The SEMATECH-Berkeley actinic inspection tool (AIT) is an EUV microscope developed at Lawrence Berkeley National Laboratory (LBNL) [6,7]. Similar to other digital microscopes, the AIT consists of imaging optics (Fresnel zone plate) and a charge coupled device (CCD) camera to capture the image of the object that is illuminated by partially coherent source. Accurate measurement of the wavefront provides reliable feedback to perform the alignment of the AIT, allowing us to achieve diffraction-limited imaging performance. Phase retrieval is perhaps the best method to measure the AIT wavefront since it does not require any customization of the microscope's optical system.

Here we review existing phase retrieval methods to see which one can be applied most effectively to the AIT, or more generally, to any digital microscope. Properly modeling the illumination coherence properties and the object transmittance are important components of the phase retrieval process. However, adding partial coherence is not straightforward and is currently the subject of active research.

For example, researchers have studied how to incorporate such effects into the Gerchberg–Saxton algorithm [8,9], the curvature sensing method [10], the transport of intensity equation [11], and phase retrieval by the extended Nijboer–Zernike theory [12]. Although the extended Nijboer–Zernike theory can incorporate the coherence properties and the object transmittance [12], solving the inverse problem is not simple. Therefore, next, we consider combining the partially coherent imaging calculation and a numerical optimization method. Repeating the partially coherent imaging calculation while changing the Zernike coefficients of the wavefront allows us to estimate the Zernike coefficients from a series of through-focus images [7,13]. Although this optimization is computationally demanding, a simulated annealing algorithm combined with the eigenfunction decomposition of the image significantly reduced the computational burden [14]. This method yields a set of Zernike coefficients, not a two-dimensional map of the wavefront. Furthermore, only a finite subset of the Zernike coefficients is estimated in this method, and higher-order Zernike terms are not recovered.

In this paper, we combine the gradient descent algorithm [2] and partially coherent imaging simulations to retrieve not the Zernike coefficients but the two-dimensional wavefront map of the AIT lens. In most cases the gradient descent method is used with coherent or incoherent illumination [1]; therefore the incorporation of partially coherent imaging is the first step of this paper. After the incorporation, the gradient descent algorithm is accelerated by combining the fast Fourier transform (FFT) and the eigenfunction decomposition of the image [14,15]. The simulation speed of the image calculation by eigenfunction decomposition depends on the object characteristics. The retrieved wavefront accuracy also depends on the object. We show that the smaller the object is, the more the image calculation is accelerated and the retrieval accuracy is improved. In Section 2 we describe the formulation of the gradient descent method for partially coherent imaging. The optimized condition is introduced by simulation in Section 3. The AIT wavefront measurement, the ultimate goal of this paper, is shown in Section 4, where the retrieval result of field-dependent aberrations is shown as well. In Section 5, a qualitative assessment of the retrieval accuracy and future perspective are presented.

2. FORMULATION

A. Optical Model

A simplified optical model of the AIT is illustrated in Fig. 1. The *Z* axis is defined as the optical axis. The object and image planes are conjugate, and points on these planes are identified with the coordinate notation (x,y). The light source and pupil planes are also conjugates, and points on these planes are identified with the coordinate notation (f,g) and are normalized for the pupil radius to be unity. The coordinates (x,y) and (f,g) are linked by the Fourier transform relationship. The source $S(f,g)$ is defined at the aperture stop of a Koehler illumination system, which illuminates the object whose transmittance (reflectance for the AIT) is $a(x,y)$. The spectrum $a\hat{}(f,g)$ is defined in the pupil plane of the imaging optics as the Fourier transform of $a(x,y)$. After the spectrum is truncated by the pupil function $P(f,g)$, it propagates to the image plane to form the image $I(x,y,z)$. The best focus plane is

defined at $z=0$ with the fourth Zernike coefficient (defocus) being zero, using the fringe Zernike polynomial notation [16].

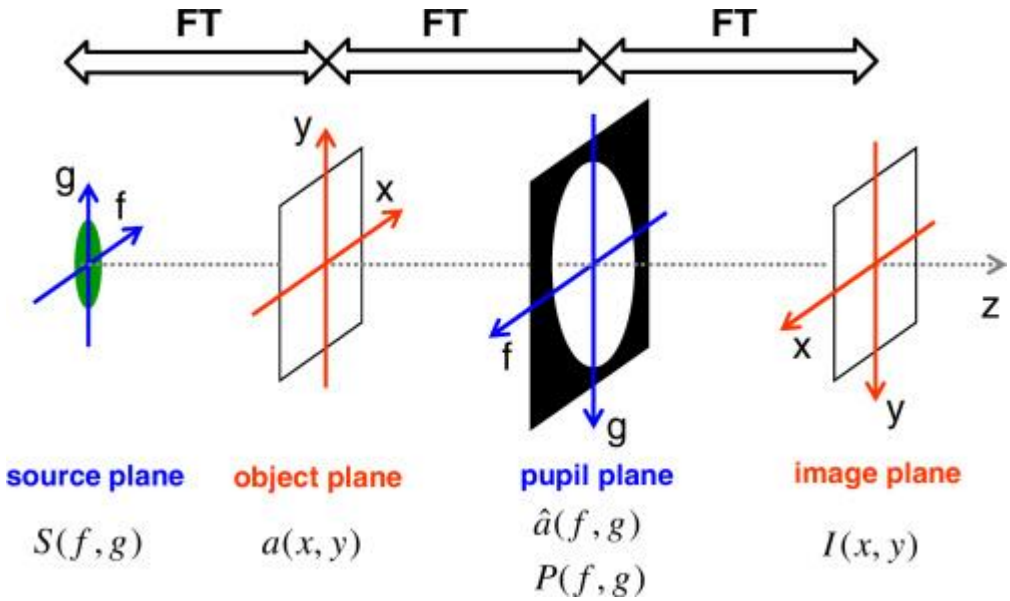

Fig. 1. Imaging system configuration. Due to the Fourier transform relationship in each plane, source and pupil planes as well as object and image planes are under a conjugate relationship.

B. Problem Setting

Assume that the image is captured at multiple focal positions (i.e., through focus). Let the number of the captured images be $J(\geq 1)$; then, the captured image is written as $I_c(x,y,z_j)$, where $1\leq j\leq J$. When we calculate the image, we have to consider the wavefront $\alpha(f,g)$, defined in the pupil plane. Therefore, the calculated image is a function of $\alpha(f,g)$, which can be written as $I(x,y,z_j,\alpha)$. Here, as we compare the calculated images to the measured data, we define a cost function based on the point-by-point difference as

$$F(\alpha)=\sum_{j=1}^{J}\iint |I_c(x,y,z_j)-I(x,y,z_j,\alpha)|^2 \mathrm{d}x\mathrm{d}y.$$

(1)

In the absence of noise, if $\alpha(f,g)$ is equal to the wavefront in the imaging optics and the illumination conditions are well known, $F(\alpha)$ is minimized. Therefore, our final goal is to find $\alpha(f,g)$ that achieves $F(\alpha)=0$ or $F(\alpha)\approx 0$. To solve this problem we use the gradient descent method due to its compatibility with the partially coherent image calculation.

C. Partially Coherent Imaging Formulation for the Gradient Descent Method

We will select an efficient calculation method of partially coherent image formation for the gradient descent method. Let the defocus-induced aberration be $\beta(f,g)$. For a quasi-monochromatic source of wavelength λ, we can simplify the Hopkins transmission cross-coefficient equation [17,18] as

$$I(x,y,z,\alpha)=\iint S(f',g')|\mathbf{FT}[P(f+f',g+g')a\hat{}(f,g)]|^2 df'dg',$$

(2)

$$P(f,g)=c(f,g)\exp[-i\alpha(f,g)]\exp[-iz\beta(f,g)],$$

(3)

$$\beta(f,g)=2\pi\lambda[1-\sqrt{1-\mathrm{NA}^2(f^2+g^2)}].$$

(4)

Here $\mathbf{FT}$ is the two-dimensional Fourier transform operator, $c(f,g)$ represents modulation in the pupil plane, NA is the numerical aperture, and we referred to [19] for $\beta(f,g)$. In this paper, we assume no modulation, so that $c(f,g)$ reduces to the circ function $\mathrm{circ}(\sqrt{f^2+g^2})$. Often, the Fourier transform is computed by the FFT. We made a key change to Eq. (2): shifting the object spectrum instead of the pupil function; the reason for this is discussed later in this section. Then, Eq. (2) can be changed to

$$I(x,y,z,\alpha)=\iint S(f',g')|\mathbf{FT}[a\hat{}(f-f',g-g')P(f,g)]|^2 df'dg'=\iint\iint Q(f_1,g_1,f_2,g_2)P(f_1,g_1)P_*(f_2,g_2)e^{-i2\pi[(f_1-f_2)x+(g_1-g_2)y]}df_1dg_1df_2dg_2,$$

(5)

where

$$Q(f_1,g_1,f_2,g_2)=\iint S(f,g)a\hat{}(f-f_1,g-g_1)a\hat{}_*(f-f_2,g-g_2)dfdg.$$

(6)

Note that $Q(f_1,g_1,f_2,g_2)$ represents the correlation between the two points of the pupil function as the transmission cross coefficient represents the correlation between two points of the diffracted light. A set of eigenvalues and eigenfunctions for efficient computation is obtained by decomposing the correlation $Q(f_1,g_1,f_2,g_2)$ [14,15].

The decomposed result is shown assuming computation. For the computation, we may regard the source as a collection of N mutually incoherent point sources. Then, Eq. (5) is

$$I(x,y,z,\alpha)=\sum_{n=1}^{N} S(f_n,g_n)|\mathbf{FT}[a\hat{}(f-f_n,g-g_n)P(f,g)]|^2,$$

(7)

where (f_n,g_n) is the nth point source position. From Eq. (7), we can obtain discretized $Q(f_1,g_1,f_2,g_2)$, which is decomposed into eigenvalues μ'_j and eigenfunctions $\psi'_j(f,g)$ [14,15]; then

$$I(x,y,z,\alpha)=\sum_{n=1}^{N'} |\mathbf{FT}\{\mu'_n\psi'_n(f,g)\exp[-i\alpha(f,g)]\exp[-iz\beta(f,g)]\}|^2.$$

(8)

Here $N'\leq N$ [15]. Equation (8) is the final form of partially coherent imaging for the gradient descent method.

We will use Eq. (8) for the following reasons. First, the image calculation time depends on the repetition of the FFT, and Eq. (8) has a possibility to finish the image calculation with less than N FFTs. Second, although the eigenvalues μ'_j and eigenfunctions $\psi'_j(f,g)$ depend on the object and the source, they do not depend on the wavefront $\alpha(f,g)$. During the gradient descent optimization, the object and source are fixed. Once the eigenvalues and eigenfunctions are derived, they are reusable even as the wavefront $\alpha(f,g)$ changes. Therefore, Eq. (8) will be an efficient, fast calculation method for the gradient descent optimization. The benefit of the eigenvalue methods is well examined in [14], where the wavefront is changed in simulated annealing loops.

Decomposing Eq. (2) into eigenvalues μ_j and eigenfunctions $\psi_j(f,g)$, the image function is written as [20,21]

$$I(x,y,z,\alpha)=\sum_{n=1}^{N} |\mathbf{FT}[\mu_n(z,\alpha)\psi_n(f,g,z,\alpha)a\hat{}(f,g)]|^2.$$

(9)

Computationally, in Eq. (9), we have to repeat N FFTs [21]. In addition, the eigenvalues μ_j and eigenfunctions $\psi_j(f,g)$ depend on the wavefront $\alpha(f,g)$. If we changed the wavefront $\alpha(f,g)$, we would have to recalculate the eigenvalues and eigenfunctions, which would require additional computation time for eigenvalue decomposition in each loop. The difference between Eqs. (8) and (9) tells us that it is more efficient for the gradient descent algorithm to shift the object spectrum rather than the pupil function.

D. Gradient Descent Algorithm for Partially Coherent Imaging

The gradient descent algorithm iteratively searches for an extremum of a cost function. If $h(\xi)$ is the function to minimize, the $k+1$ value of ξ in the iterative procedure is given by

$$\xi_{k+1}=\xi_k-wdhd\xi.$$

(10)

Here $w\ (>0)$ is an arbitrarily chosen weight factor. As we have to minimize $F(\alpha)$ in Eq. (1), we have to differentiate $F(\alpha)$ with respect to $\alpha(f,g)$. In this case, the functional derivative should be used to define the derivative at (f',g'). Then, the gradient descent formula for $\alpha(f,g)$ is

$$\alpha_{k+1}(f',g')=\alpha_k(f',g')-w\iint dFd\alpha\delta(f-f',g-g')\mathrm{d}f\mathrm{d}g=\alpha_k(f',g')-wdFd\alpha|||_{f=f',g=g'}.$$

(11)

Here $\delta(f,g)$ is the Dirac delta function. According to Eq. (11), we would have to repeat the iteration at every point inside the pupil, which would be computationally demanding. However, the compatibility between the gradient descent algorithm and the imaging calculation allows us to overcome this obstacle [2]. Let us combine Eqs. (1), (8), (11), and the Fourier transform to obtain

$$\alpha_{k+1}(f,g)=\alpha_k(f,g)-4w\ \mathrm{Im}\sum_{j=1}^{J}\sum_{n=1}^{N}\mu'_n\psi'_n(f,g)\exp[-i\alpha_k(f,g)]\exp[-iz_j\beta(f,g)]\mathbf{FT}\{[I(x,y,z_j,\alpha_k)-I_c(x,y,z_j)]A_{*n}(x,y,z_j,\alpha_k)\},$$

(12)

$$A_n(x,y,z,\alpha)=\mathbf{FT}\{\mu'_n\psi'_n(f,g)\exp[-i\alpha_k(f,g)]\exp[-iz_j\beta(f,g)]\}.$$

(13)

where * represents the complex conjugate, Im takes the imaginary part of the argument, and $\mathbf{FT}^{-1}$ is the inverse Fourier transform. This step replaces the point-by-point iteration, and the whole wavefront gets updated on each iteration. Therefore, the final output is not a limited number of Zernike coefficients, but the complete two-dimensional wavefront map.

3. SIMULATION

A. Parameter Setup

To match the experimental conditions of the AIT microscope, we run a simulation in which the wavelength is 13.4 nm and the NA is 0.0625. The source coherence factor, σ, defined as the ratio of the illumination NA and the object side NA, is 0.2. A random wavefront error is created as follows. First, we randomly generate Zernike coefficients from the fifth up to the 36th term (coefficients 1–4, for piston, x tilt, y tilt, and defocus are set to zero). Then, the nth Zernike coefficient C_n is divided by $\sqrt{n}$ to imitate common behavior of real-world aberrations. Finally, the whole wavefront error was normalized to $150\ \mathrm{m}\lambda$ root-mean-square (RMS), which is estimated from the zone plate design. These parameters are fixed during the following simulation.

To extract the phase information, and reduce the uncertainties introduced by noise in the experimental data, multiple images must be included in the fit. Here, we use four images, which are assumed to be captured at $z=-1.5z'$, $-0.5z'$, $0.5z'$, and $1.5z'$, where z' represents a single focus step.

For the object, we use a $d\times d$ square "contact," a bright feature on a dark background, which is a commonly used pattern in lithography masks for semiconductor manufacturing.

B. Contact Size and Defocus Matrix

The retrieval accuracy depends on optical parameters. The optical parameters for partially coherent imaging include the wavelength, NA, source, object, aberration, and defocus. Although in the AIT it is possible to tune the wavelength and the NA, these are not common features for other microscopes. Thus, here we change the object and defocus step to investigate an experimental framework for better retrieval accuracy.

We vary the side length d of the square contact from $0.25\lambda/\mathrm{NA}$ to $1.00\lambda/\mathrm{NA}$. In each d, the defocus step z' is varied from 0.1 to 3.0 R, where $R=\lambda/[4(1-\sqrt{1-\mathrm{NA}^2})]$ is the Rayleigh unit of defocus. At each combination of d and z', we simulated an image $I_c(x,y,z_j)$ affected by the given wavefront error. Let us define the convergence τ as in Eq. (14):

$$\tau=\sum_{j=1}^{4}\iint|I_c(x,y,z_j)-I(x,y,z_j,\alpha_k)|^2\mathrm{d}x\mathrm{d}y\sqrt{\Big/\sum_{j=1}^{4}\iint|I_c(x,y,z_j)|^2\mathrm{d}x\mathrm{d}y}.$$

(14)

We set $\alpha_1(f,g)=0$ and repeated 100 iterations to obtain the retrieved wavefront $\alpha_{100}(f,g)$. The number of the iteration was determined because several tests showed that the decrease of τ at the 100th iteration was less than 1.0×10^{-4} in most cases. Figure 2 shows the convergence τ after 100 iterations.

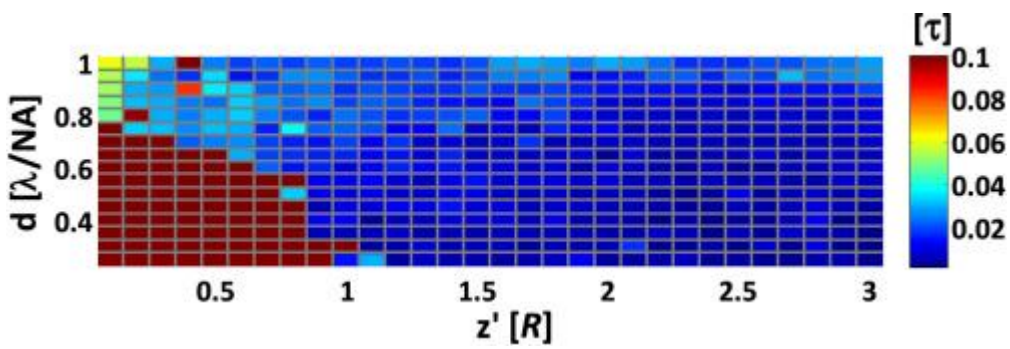

Fig. 2. Contact size and defocus matrix to find an experimental framework. The color scale represents the convergence value τ defined in Eq. (14).

To obtain Fig. 2, the image calculation was done by Eq. (7) with the given wavefront error, and then the wavefront was retrieved by Eq. (12), which originates from Eq. (8). Therefore, different simulation engines were used for imaging simulation and the wavefront retrieval, though Eqs. (7) and (8) are mathematically equal and thus produce an identical image within the range of numerical error.

Note that the best convergence is achieved around $d=0.25\lambda/\mathrm{NA}$ (the smallest size studied) and $z'=1.5\mathrm{R}$. When both d and z' are small, our gradient descent algorithm diverges. The reason is that even if the sign of the 180 deg rotation symmetric wavefront changes, the resultant images at best focus are unchanged. For example, assume that the wavefront is expressed only by the fifth Zernike polynomial (astigmatism). If the fifth Zernike coefficient C_5 is either 1 or -1, the resultant images at best focus are identical. Therefore, we cannot determine the sign of the phase from the best focus image. Furthermore, 180 deg rotation symmetric aberration distorts the image less than 180 deg rotation asymmetric aberration does. Thus, the image near best focus contains less phase information than the large out-of-focus image. Although we can avoid the divergence by carefully adjusting the weight factor w, it is not a trivial task. It is much simpler to avoid unstable contact size and defocus combinations.

If we take a focus step of 0.75 R, a smaller contact shows worse convergence than a larger contact. We do not have a clear reason for this because this is a combined result of illumination and object. There are two possible main reasons. First, the focus step is still too small to distort the image enough. Second, if the illumination is combined appropriately, the image from a bigger contact is more distorted from the image without aberration. The counterintuitive result that occurs at a defocus of 0.75 R is perhaps due to the mixed results of these two reasons.

The important result is that, if combined with a proper defocus step, smaller contacts can achieve better convergence. The reason is that the diffracted light from the smaller contact can fill the pupil more uniformly, improving the algorithm's ability to sample the nature of the wavefront. The trade-off of using a smaller contact is less light flux, which may increase the shot noise.

C. Simulation Time

For the image calculation, we require N' FFTs with Eq. (8). The number N' is dominated by the spatial coherence of the light that forms the image. Therefore, N' depends mostly on the object if the illumination is fixed. For example, when the object is an infinitely small contact, we find that $N'=1$ regardless of the source intensity distribution [15]. Thus, we expect that a smaller N' corresponds to a smaller contact. Let us confirm it.

We set a source intensity distribution that contains 749 mutually incoherent point sources, as shown in Fig. 3. According to Eqs. (7) or (9), we have to repeat 749 FFTs to obtain the final image regardless of the object. However, the number of FFTs performed for one image calculation is reduced when we use Eq. (8). Note that no eigenfuction truncation is required. Therefore, image simulation is accelerated without approximation. By setting a $d\times d$ square contact as the object, we simulated the relationship

between the contact size d and N', and the result is shown in Fig. 4. Note that as the contact becomes smaller, N' also becomes smaller. Thus, we can reduce the simulation time with a smaller contact. For example, when d is $0.7\lambda/\mathrm{NA}$, N' is 47 to get a 94% reduction of FFT repetition compared with Eqs. (7) or (9). This result is also preferable because a smaller contact can achieve a higher retrieval accuracy as discussed in Subsection 3.B.

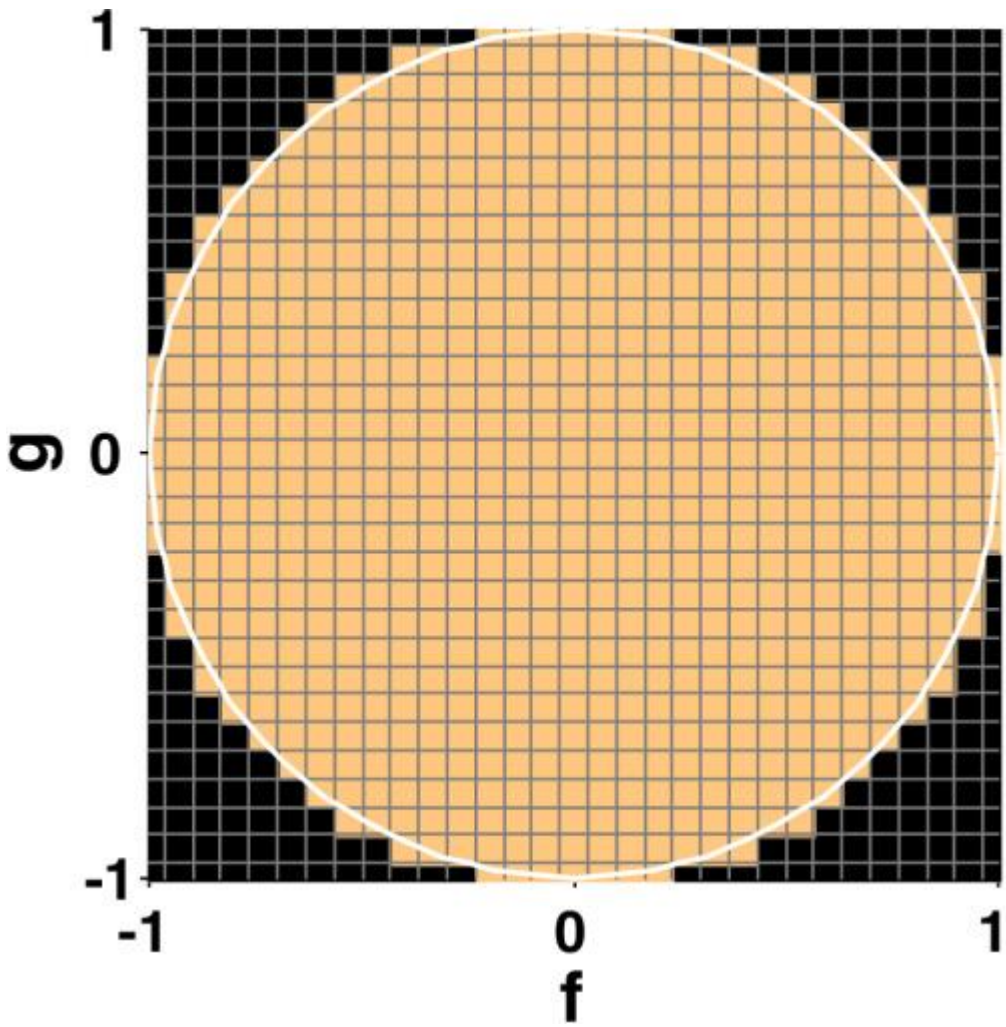

Fig. 3. Partially coherent illumination used for the simulation in Subsection 3.C. The white circle shows $(f^2+g^2)^{1/2}=1$. Each pixel shows a point source with unit intensity that is mutually incoherent with all others.

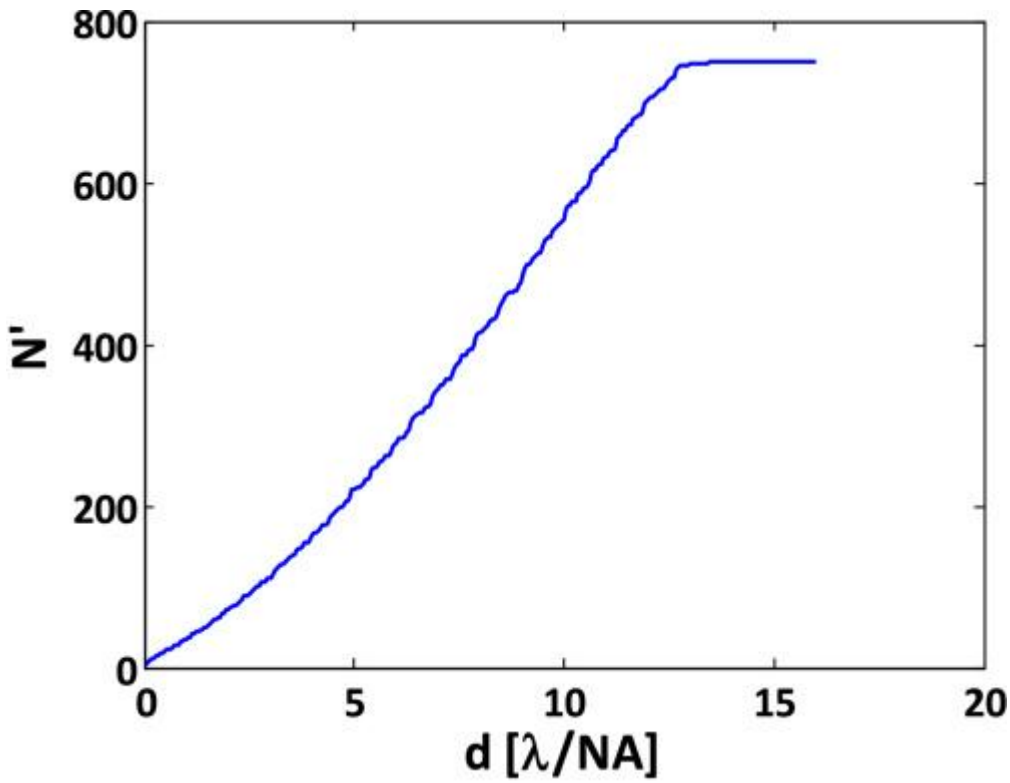

Fig. 4. Relationship between the contact size and the number of eigenfunctions.

4. EXPERIMENT

A. Experimental Setup

The AIT was operated with a wavelength of 13.4 nm and an NA of 0.0625 (the same setup as in Section 3). For the object, we placed approximately 1000 square contacts with 150 nm ($0.70\lambda/\mathrm{NA}$) side length. These contacts are placed across the AIT field on a 900-nm-pitch grid. The contact array is illuminated by partially coherent light with a coherence factor of 0.2. The AIT has a magnification of approximately 900, and the images are captured by a CCD with 13.5 µm pixel size. We captured seven through-focus images with a focus step of 1 µm. One of the images is shown in Fig. 5. In spite of the small contact size, images with enough contrast were obtained.

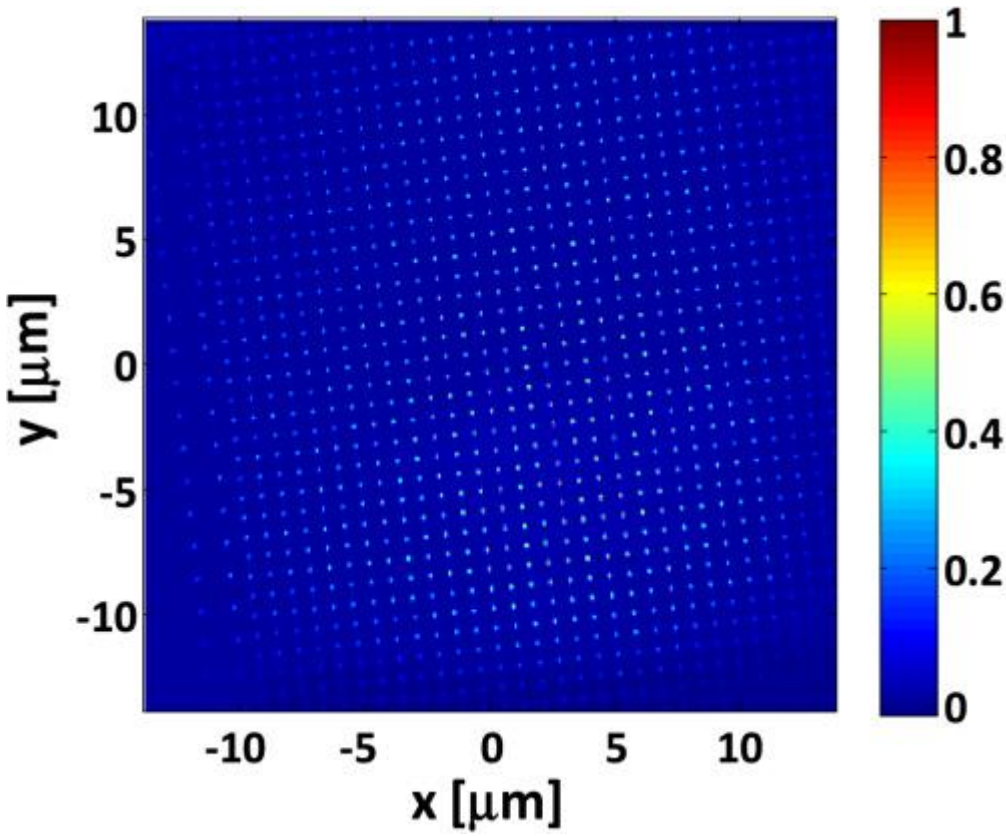

Fig. 5. AIT image of contact array. The maximum intensity is normalized to unity.

B. Data Processing

The strength of coherent interaction between two points in partially coherent imaging can be approximated by $\mathbf{FT}[\mathrm{S}(f,g)]\mathbf{FT}[\mathrm{P}(f,g)]$ normalized by its maximum value [22], which we refer to as $\Omega(x,y)$ in this paper. Here, $\Omega(x,y)$ can be used to determine whether the optical proximity effect from an adjacent contact is significant. As $\Omega(900\ \mathrm{nm},0)\approx\Omega(0,900\ \mathrm{nm})\approx 0$ assuming small aberration, each contact can be regarded as an isolated object in terms of imaging characteristics. Therefore, we can crop a portion of the image that corresponds to one of the contacts, considering it as the image of an isolated contact.

For accurate retrieval, we have to reduce the noise. The maximum spatial frequency of the image corresponds to twice the pupil radius. Thus, we can reduce photon shot noise by applying a low-pass filter that cuts all special frequencies above that threshold.

As defocus adds only phase difference but no absorption as in Eq. (3), every through-focus image should have the same integrated intensity. We normalized every cropped through-focus image to the same value before applying the gradient descent method.

For wavefront retrieval, we have to know the value of the defocus for each image. Although the focus step of 1 μm is known, we do not know *a priori* where the best focus occurs. It is sufficient to roughly

estimate the best focus location because small amounts of focus offset will be calculated by the wavefront retrieval algorithm. Although there are various methods for best focus estimation [23], we used the normalized standard deviation defined as

$$H(z)=\iint_D [I_c(x,y,z)-\bar{I}_c(z)]^2 \mathrm{d}x\mathrm{d}y \bar{I}_c(z) \sqrt{},$$

(15)

$$\bar{I}_c(z)=\iint_D I_c(x,y,z)\mathrm{d}x\mathrm{d}y \iint_D \mathrm{d}x\mathrm{d}y,$$

(16)

where D is the cropped region. According to [23], $H(z)$ has its maximum value at the best focus position. We fit $H(z)$ to a quadratic form in terms of z by least square fitting and assume the best focus position, where $H(z)$ is maximized.

From Fig. 2, images captured with less than 0.50 R defocus sometimes make the optimization diverge in our algorithm. Therefore, it is preferable to use images with more than 0.50 R defocus. The focus step of 1 μm corresponds to 0.58 R. Thus, if we eliminate three images that are captured closest to the best focus, the remaining four images have defocus more than 0.50 R. For this reason, we decided to use four images for the wavefront retrieval. An example series of four images is shown in Fig. 6. The images are cropped around $(x,y)=(3.1\ \mu\mathrm{m},-5.5\ \mu\mathrm{m})$, where the intensity of the images is relatively strong, leading high-contrast images to maximize the visualization.

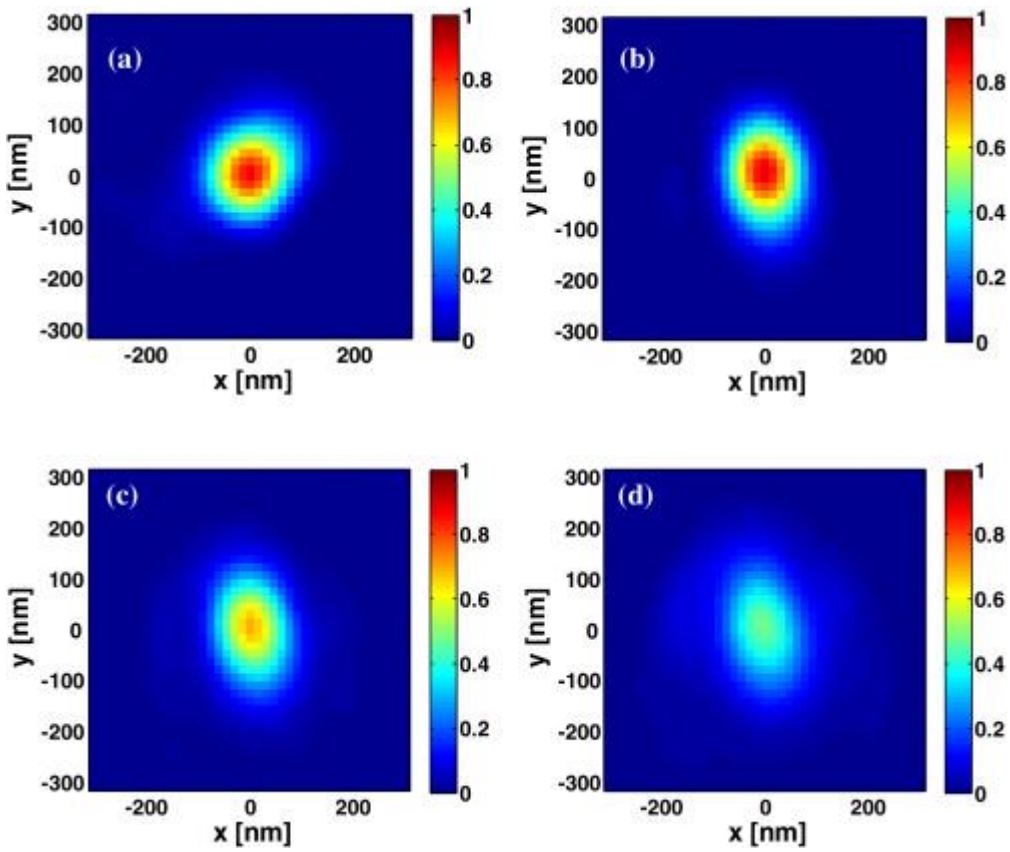

Fig. 6. Four selected through-focus image details, captured at defocus values of (a) -0.7685 R, (b) 0.9806 R, (c) 1.5636 R, and (d) 2.1467 R.

As part of the processing, the CCD camera's background offset is subtracted, and a small, focus-dependent lateral image shift that arises from stage motion and the 6 deg angle of incidence is removed from each frame, based on the intensity centroid. As the cropped image size was 59×59 pixels, zero padding is applied to adjust the image size to 64×64 pixels for FFT compatibility.

C. Wavefront Retrieval Result

For testing, we applied the gradient descent algorithm to the images shown in Fig. 6. The convergence τ defined in Eq. (14) was almost stable at the 250th iteration, where τ was 0.11 and the decrease of τ was less than 1.0×10^{-3}. The images are simulated with the retrieved wavefront, which is shown in Fig. 7. We see from Figs. 6 and 7 that image distortions are reproduced. Therefore, we may find that the principal aberration is retrieved at $\tau=0.11$, which gave us a decision to iterate 250 times for wavefront retrieval.

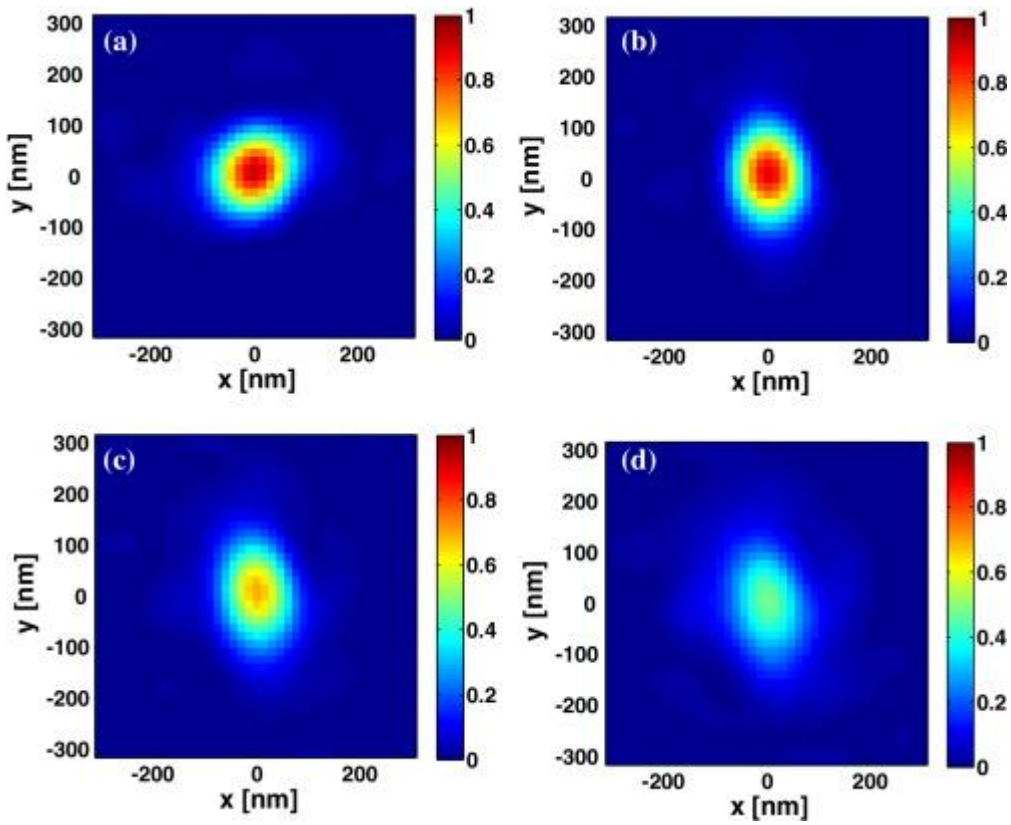

Fig. 7. Simulated images with the retrieved wavefront at the same defocus positions as in Figs. 6(a)–6(d).

The weight factor w is changed in each iteration. At the kth iteration, we set three weight factors, $w_1<w_2<w_3$, to obtain three different $\alpha_{k+1}(f,g)$. We simulated three sets of through-focus aerial images from these three different $\alpha_{k+1}(f,g)$ followed by the calculation of three different τ_1, τ_2, and τ_3. Quadratic fitting of τ_1, τ_2, and τ_3 with respect to w_1, w_2, and w_3 gives a parabolic curve whose vertex position indicates the weight factor at the kth iteration. Including this weight factor determination step, the simulation time for 250 iterations was approximately 2.8 s, in which the eigenfunction generation occupied 0.1 s. Here, we used MATLAB R2013b in Windows 7 with a 2.80 GHz Xeon processer. This weight factor determination step was also used in Subsection 3.B.

After this test, we extracted 926 separate areas in Fig. 5, where the maximum intensity in each area exceeds 15% of the peak intensity. We repeated the gradient descent algorithm independently in each region to retrieve the field-dependent wavefront variation. The wavefront retrieval converged at 905 out of 926 areas. For practical assessment of the wavefront, the wavefront was fit to Zernike polynomials up to the 16th term. The number of the fitting terms is determined by the cropped image size, which corresponds to the sampling period in the spatial frequency domain. The RMS distribution obtained from the fifth to 16th Zernike coefficients of the recovered field-dependent wavefront is shown in Fig. 8(a), which is our final goal. Assuming the field-dependent aberrations vary smoothly, we apply a polynomial fit to the available locations, using polynomials defined as

$$p(x,y)=\sum_{l=0}^{12}\sum_{m=0}^{l} C_{l,m} x^l y^m,$$

(17)

and the smoothed result is shown in Fig. 8(b).

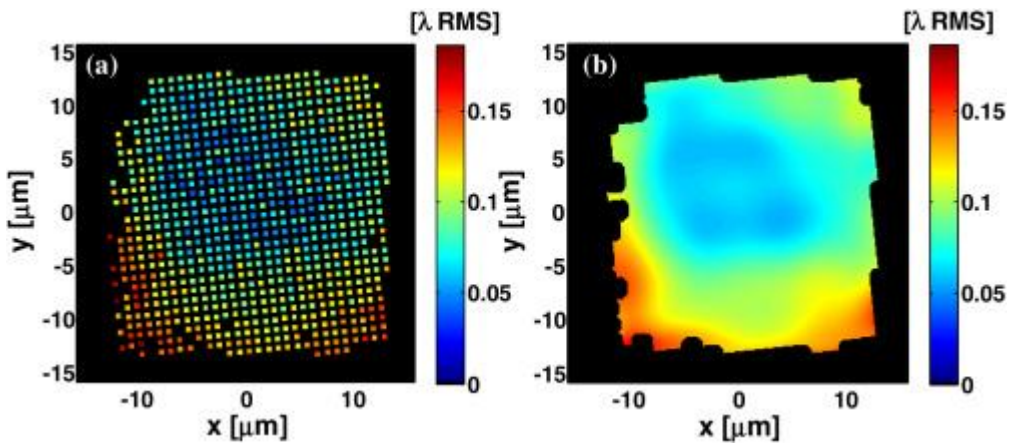

Fig. 8. RMS distribution of the recovered field-dependent wavefront. (a) RMS of wavefront at each discrete contact hole location. (b) Polynomial fit of (a).

5. DISCUSSION

A. Assessment of Wavefront Retrieval

In this subsection, we compare the recovered wavefront with the expected wavefront from the lens (Fresnel zone plate) design [24]. As we have not established a method to quantitatively assess the retrieval accuracy, qualitative assessment will be performed.

First, the best focus distribution was examined. Temporarily, we assumed that the fourth focal plane was the best focus plane. Rough focus offset was estimated as the position at which the normalized standard deviation $H(z)$ took its maximum value. Then, additional defocus offset was estimated by fitting $\beta(f,g)$ in Eq. (4) from the retrieved wavefront. Summing up the rough and additional focus offset, the best focus distribution is obtained as shown in Fig. 9(a). As the lens design predicts [24], the best focus position varies along the y axis, whereas the best focus position is stable along the x axis. Figure 9(a) shows this characteristic.

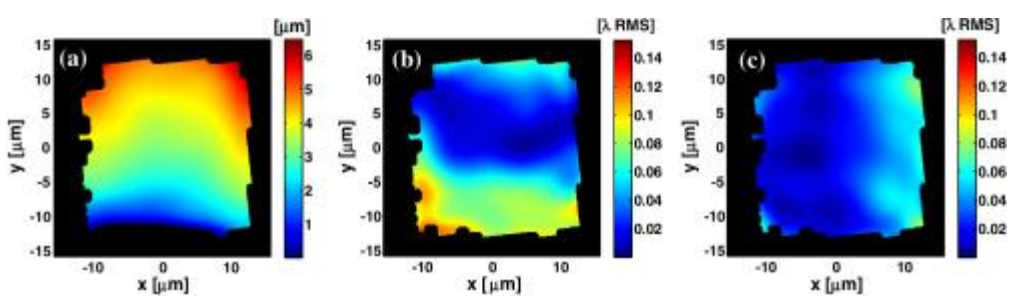

Fig. 9. Field-dependent wavefront fit by Eq. (17) for visualization. (a) Best focus distribution. (b) RMS distribution of 0–90 deg astigmatism. (c) RMS distribution of $\pm 45\ \mathrm{deg}$ astigmatism.

Next, the distribution of the astigmatism is examined. Let C_5 and C_6 be the fifth and sixth Zernike coefficients, respectively. The RMS of 0–90 deg astigmatism is determined by $|C_5|/\sqrt{6}$, and $\pm 45\ \mathrm{deg}$ astigmatism is determined by $|C_6|/\sqrt{6}$. The RMS distribution of astigmatism is shown in Figs. 9(b) and 9(c). According to [24], the fifth Zernike coefficient C_5 varies along the y direction, and the sixth Zernike coefficient C_6 varies along the x direction. Figures 9(b) and 9(c) clearly show this tendency. Furthermore, total astigmatism $\overline{C_{25}+C_{26}}/\sqrt{6}$ around $x=-3\ \mu\mathrm{m}$ and $y=4\ \mu\mathrm{m}$ is almost zero. The astigmatism in the AIT images can be reduced by alignment [7]. As the data used in this paper were taken after the alignment was done, the retrieved wavefront should have a nearly zero astigmatism region. Figures 9(b) and 9(c) show this feature as well.

Clearly, the best performance of the AIT is achieved where astigmatism is minimized. To explore the potential performance of the AIT, two points are selected in a region in which the astigmatism is minimized. These two points are ($-2.5\ \mu\mathrm{m}$, 4.8 μm) and ($-0.2\ \mu\mathrm{m}$, 5.1 μm). The Zernike coefficients up to the 16th term at these two points are evaluated as in Fig. 10. The RMS values up to the 16th term are $34\mathrm{m}\lambda$ RMS at ($-2.5\ \mu\mathrm{m}$, 4.8 μm) and $38\mathrm{m}\lambda$ RMS at ($-0.2\ \mu\mathrm{m}$, 5.1 μm), and it is assumed that the AIT can approximately realize the diffraction-limited imaging system.

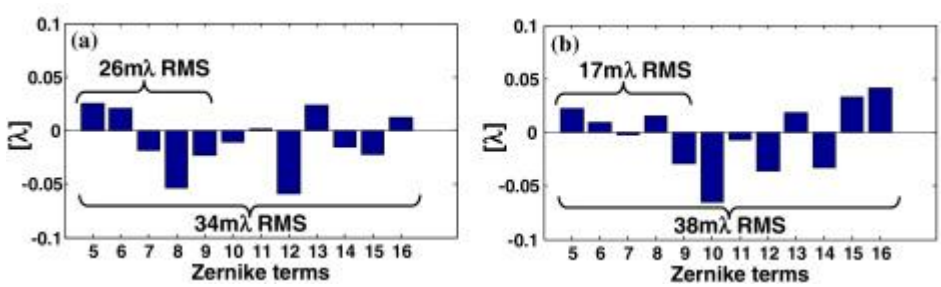

Fig. 10. (a) Zernike coefficients at ($-2.5\ \mu\mathrm{m}$, 4.8 μm). The RMS value up to the ninth term is $26\mathrm{m}\lambda$ RMS, and up to the 16th term is $34\mathrm{m}\lambda$ RMS. (b) Zernike coefficients at ($-0.2\ \mu\mathrm{m}$, 5.1 μm). The RMS value up to the ninth term is $17\mathrm{m}\lambda$ RMS, and up to the 16th term is $38\mathrm{m}\lambda$ RMS.

The wavefront in a small enough area should have some similarities because the wavefront slowly varies across the field. Therefore, to confirm the similarities of the wavefront in a small area can be an additional qualitative assessment of this wavefront retrieval method. Two-dimensional wavefront maps at adjacent 3×3 points (total of 18 adjacent points) are shown in Fig. 11. The two points selected for Fig. 10, highlighted by read squares, are the centers of the 3×3 points. The 18 wavefront maps possess similar profiles, such as negative dot-like peaks. This result gives us an additional qualitative verification of the wavefront retrieval method.

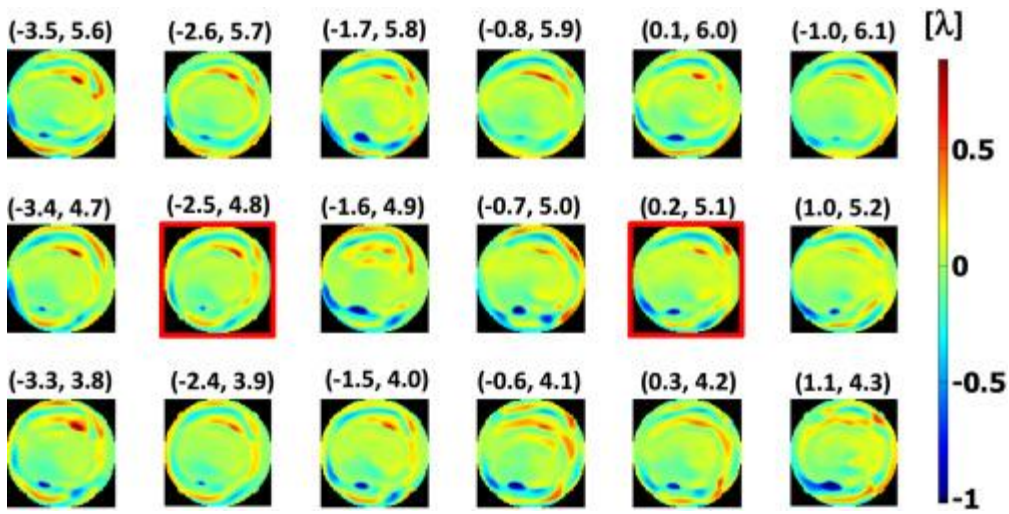

Fig. 11. Two-dimensional wavefront maps retrieved at 18 adjacent points. The coordinates above each wavefront are in units of micrometers. The wavefront highlighted by a red square is obtained at one of the two selected points of Fig. 10.

Figure 11 has another aspect. If we assume in this region that the contact is perfectly fabricated, the illumination is uniform, the best focus position is unchanged, and the wavefront is the same, Fig. 11 gives a repeatability estimation of wavefront retrieval, although a limited one because of the required assumptions.

To calculate the results of Figs. 8–10, an image with size of 64×64 pixels is used by zero padding the original 59×59 pixel image. However, for Fig. 11, we used 512×512 pixel images obtained by extra zero padding to the 59×59 pixel image to increase the number of pixels in the wavefront, which improves the visual evaluation of the wavefront.

B. Further Modeling Ideas

For the sake of modeling simplicity, we neglected several factors in imaging calculation. In this subsection, we discuss possible factors that can be modeled into the gradient descent algorithm.

The first factor is the object transmission function $a(x,y)$. In this paper, we used the Kirchhoff boundary condition [25], in which the contact transmittance was represented by a two-dimensional rect function. If we know the object transmission function $a(x,y)$ more accurately, the wavefront retrieval accuracy will be improved. One possible way is the use of simulation that can predict the electromagnetic field from the contact consisting of an absorber and multilayer film stack to reflect EUV light [26]. The simulation methods are reviewed and studied in [26]. Since the image can be predicted with the simulated object transmittance, we may safely assume that the object transmittance $a(x,y)$ can be obtained by simulation. Once the object transmittance $a(x,y)$ is simulated, we reuse it on every contact assuming that every contact has the same shape.

The EUV light source of AIT has an energy bandwidth $\Delta E/E$ of 1/1450 [7]. Finite bandwidth blurs the image, which can be modeled as an incoherent summation of images from different wavelengths [27]. This process does not change the fundamental of the gradient descent algorithm formulation, although simulation time is increased. As we concentrated on the acceleration of wavefront retrieval, we assumed a quasi-monochromatic source.

We neglected the magnification effect as well. If the magnification is high, the object side NA and image side NA differ, which should be radiometrically corrected. The radiometric correction can be modeled by adding an obliquity factor to $c(f,g)$ in Eq. (3) [28]. Although the magnification of AIT is not small (approximately 900), the NA of AIT is low (0.0625). The obliquity factor with this NA did not affect the imaging result significantly. To simplify the imaging model in this paper, we omitted the obliquity factor.

Another factor that can be modeled but is neglected in this paper is the high-NA effect. We used the scalar approximation that is applicable to low-NA imaging as in this paper. The scalar formulation is applicable even to the newly developed higher-NA EUV microscope with NA of 0.15625 [29]. Nevertheless, for higher simulation accuracy, we may model the high-NA related polarization effect by changing $c(f,g)$ in Eq. (3) [28]. There is an EUV microscope specific issue related to high-NA imaging. As the indent angle to the object is 6 deg, the diffracted light is distorted. We need to obtain the object transmittance more accurately considering the oblique incidence, which can be realized by current simulation tools. The last concern is the contact manufacturability. In this paper, a contact with side length of $0.7\lambda/\mathrm{NA}$ is used, with which acceptable retrieval speed and accuracy are realized. If we scale down the contact size with $\mathrm{NA}=0.15625$, $0.7\lambda/0.15625=60\ \mathrm{nm}$, which approaches the current fabrication limit. We may use a larger contact, though it increases the retrieval time associated with an increased number of eigenfunctions for image calculation. Modeling shot noise would be one way to set specifications on the required number of photons to make the accuracy reach a given level.

6. SUMMARY

The wavefront of the AIT was retrieved by the gradient descent algorithm from a series of through-focus images illuminated by a partially coherent source. To retrieve the field-dependent wavefront, contacts should be placed such that the interaction between the contacts becomes negligible; then we can apply the gradient descent algorithm to each contact image. The retrieved wavefront distribution qualitatively agrees with the expected characteristic from its lens design. We showed that smaller contacts achieved greater accuracy. We attribute this to the larger coverage of diffracted light in the pupil, and the benefits to the partially coherent imaging calculation within this method. We also showed that the smaller contact accelerated image calculation if the image was decomposed into eigenfunctions and eigenvalues determined from the source and object. Technically, the gradient descent algorithm can be applied to retrieve both phase and amplitude. Formulation of the object's phase and amplitude retrieval is a challenging application for extending the capabilities of short-wavelength imaging microscopes where interferometric lens characterization is unavailable. In addition, we incorporated the object transmittance and illumination source intensity. Initial testing shows that our algorithm can retrieve the

wavefront with the object of double contacts that are sensitive to coma aberration (result not shown here). Finding better combinations of objects and sources for phase retrieval will be another way to improve the accuracy of wavefront recovery.

ACKNOWLEDGMENTS

This work started while Yamazoe was a visiting industrial fellow at the University of California, Berkeley, and we thank Professor Andrew R. Neureuther for organizing this collaboration. Yamazoe continued this work at Canon Inc., Japan, and Canon U.S.A. Inc. We thank Yoshinari Higaki, Yoshiyuki Sekine of Canon Inc., and Tokuyuki Honda of Canon U.S.A. Inc. for their technical discussion. Measurements performed on the AIT microscope were conducted by the Center for X-Ray Optics. Portions of this work were funded by SEMATECH and performed by the University of California Lawrence Berkeley National Laboratory under the auspices of the U.S. Department of Energy. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The EUV photomask used in these experiments was created by Intel for the Microfield Exposure Tool (MET) project, also at LBNL.

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.