Title
Quantitative phase retrieval with arbitrary pupil and illumination

Permalink
https://escholarship.org/uc/item/90j7x68r

Journal
Optics Express, 23(20)

Authors
Claus, RA
Naulleau, PP
Neureuther, AR
et al.

Publication Date
2015-10-05

DOI
10.1364/OE.23.026672

Peer reviewed

Quantitative phase retrieval with arbitrary pupil and illumination

Rene A. Claus,† Patrick P. Naulleau,‡ Andrew R. Neureuther,† and Laura Waller†

†Applied Science & Technology, University of California, Berkeley, Berkeley, CA, 94709, USA
‡CXRO, Lawrence Berkeley National Lab, Berkeley, CA, 94709, USA
†reneclaus@gmail.com

Abstract: We present a general algorithm for combining measurements taken under various illumination and imaging conditions to quantitatively extract the amplitude and phase of an object wave. The algorithm uses the weak object transfer function, which incorporates arbitrary pupil functions and partially coherent illumination. The approach is extended beyond the weak object regime using an iterative algorithm. We demonstrate the method on measurements of Extreme Ultraviolet Lithography (EUV) multilayer mask defects taken in an EUV zone plate microscope with both a standard zone plate lens and a zone plate implementing Zernike phase contrast.

© 2015 Optical Society of America

OCIS codes: (110.1758) Computational imaging; (100.5070) Phase retrieval; (110.3010) Image reconstruction techniques.

References and links
1. Introduction

Phase, the cumulative delay of light as it passes through a sample, can provide a wealth of information about a sample. In biology, phase contains information about the thickness and composition of cells. In photolithography, phase delays can negatively affect the accuracy of features being printed [1]. For example, defects in multilayer mask blanks for Extreme Ultraviolet Lithography (EUVL) appear as phase objects. Quantitative phase measurements are useful in modeling and possibly repairing the defects, improving yield, and reducing the cost of the masks [2, 3]. Here, we present a new phase retrieval method that is well suited for such applications and we apply it to measurements of EUVL mask defects.

Phase measurement techniques that work on existing microscopes reduce cost and complexity. In a typical microscope, there are two things that can be changed to enable quantitative phase retrieval. The first is the pupil function, which can be modified by varying focus [4, 5], changing the objective lens [6], or using a Spatial Light Modulator (SLM) [7]. The second is the illumination, which can be changed using a condenser aperture or an SLM [8–11]. Nearly all quantitative phase recovery techniques in conventional microscopes involve multiple measurements taken as either the pupil function, the illumination, or both are changed.

One of the simplest ways to couple phase information into intensity measurements is through defocus, which can be considered a modification to the pupil function. The Transport of Intensity Equation (TIE) linearizes the phase problem by assuming that the defocus is small [4, 12]. Alternatively, iterative algorithms solve the nonlinear problem by propagating the field between the images and imposing constraints based on the measurements [5, 13–17]. These approaches generally assume coherent illumination, and partially coherent extensions [18–20] do not fully consider the interactions between Numerical Aperture (NA) and partial coherence.

Instead of varying the pupil, it is possible to introduce phase contrast by varying the illumination. Differential Phase Contrast (DPC), for example, switches the illumination from one direction to the other, computing phase quantitatively for objects that are weakly scattering [8, 10, 11]. Fourier Ptychography (FP), an iterative technique that uses measurements of the object illuminated from different angles, uses illumination coding [9] with arbitrary pupil functions, but requires coherent illumination and a large number of measurements.

In each of these techniques, the algorithm dictates how the measurements must be taken. If there are aberrations in the system, or if partially coherent illumination is used, there will be errors in the recovered phase. One area where existing algorithms are inadequate is the case of complicated illumination patterns, which are often used in photolithography to print patterns smaller than the conventional resolution limit. As a result, photomask inspection systems are designed to produce these complicated illumination patterns, which interact with the pupil function. To accurately use measurements taken under such conditions requires a new phase retrieval algorithm.

Ideally, a general transfer function could incorporate any pupil function and illumination. This transfer function could then be inverted to recover the field. However, measured intensity is nonlinear, so some form of linearization must be employed to do this. Under the assumption of a weakly scattering object, it is possible to calculate the Weak Object Transfer Function (WOTF) / Contrast Transfer Function (CTF) to relate the intensity and field. Note that this transfer function describes the spatial frequency content of the phase information which can be measured by intensity images. While several techniques already use the WOTF with defocus measurements [21–25], we suggest here that it is possible to use the WOTF with arbitrary measurements where the pupil function or illumination is varied, and we further extend the method to some non-weak objects.

By making an approximation on the object instead of the measurement, we are able to consider arbitrary pupil functions and illumination. To mitigate errors due to non-weak objects,
our algorithm is extended beyond the weak object approximation by iteratively removing the nonlinear term which does not satisfy the WOTF. The result is an iterative algorithm that is applicable to many (though not all) objects and can incorporate nearly arbitrary combinations of measurements. The inversion of the transfer function is performed using least squares, such that combining many measurements improves noise tolerance. A second advantage of combining both the imaging system and illumination conditions into a single transfer function is that it can aid with optical system design by enabling quantitative evaluation of the performance of different experimental schemes in covering all spatial frequencies of phase and amplitude [26, 27].

We have previously applied our algorithm to EUV photomasks to study multilayer defects [28, 29]. In this paper we describe the algorithm in depth, providing a description of convergence and validity. To demonstrate the accuracy of the algorithm experimentally we use EUV photomask results taken on the EUV microscope SHARP, which uses a zone plate lens to examine EUV photomasks for 13.5 nm photolithography. SHARP has fully programmable illumination, achieved by rapidly scanning the EUV source over the desired angles to define an effective 2D K"ohler source [30]. EUV photomasks are multilayer mirrors which contain weak phase roughness and particle defects that appear as phase and amplitude objects. The programmable illumination, ease of changing the pupil function using zone plates, and the generally weak phase nature of features on EUV masks makes this a well suited application for our algorithm. We capture data with two different pupil functions: a standard zone plate and a zone plate designed for Zernike phase contrast imaging. We use partially coherent illumination to improve resolution and flux.

The notation used throughout this paper denotes functions in frequency space with a tilde ($\tilde{X}$). Conjugation is noted as $X^*$. Function arguments ($x, f, x',$ and $f'$) are 2D-vectors and are omitted where it is not ambiguous. All integrals are double integrals from $-\infty$ to $\infty$. Convolution is denoted by $*$ while cross-correlation is denoted by $\star$. The coherence parameter $\sigma$ is the ratio of the illumination NA to the imaging NA. Spatially coherent illumination corresponds to $\sigma = 0$.

2. Algorithm

Our algorithm is based on the paraxial transmission cross coefficient model, which assumes K"ohler Illumination [31] and a thin object. Imaging of a scalar electric field, $E$, under this model is captured by Eq. (1), where $\tilde{J}$ is the 2D source intensity distribution and $P$ is the coherent point spread function. Each point in $\tilde{J}$ corresponds to a plane wave illuminating the object from a different direction. These plane waves are incoherent with respect to each other and contribute to the intensity independently,

$$ I = \iint E(x')e^{i\mathbf{f}' \cdot \mathbf{x}'} P(x-x')d\mathbf{x}' \left| \tilde{J}(f') \right| df'. \quad (1) $$

Since this equation is linear in neither the phase nor amplitude of $E$, we apply the Weak Object Approximation (WOA) to create a linear relationship between the intensity, which can be measured, and the field, which contains the phase and amplitude of the object. The WOA is defined by separating the object into scattered and unscattered light. The unscattered light corresponds to the DC of the field ($0^{th}$ diffracted order) or, alternatively, to the light that passes directly through a transparent object. The scattered light consists of everything else and describes the structure of the object. This can be understood by considering the simple case of the intensity at the image plane under coherent illumination. Considering the electric field at the image plane, the unscattered light is constant, $E_u = 1$, and the scattered light is $E_s$. The intensity can then be expressed as $I = |E_u + E_s|^2 = |E_u|^2 + 2Re\{E_uE_s^*\} + |E_s|^2$. The WOA requires
that $|E_s|^2$ can be ignored, making this a linear equation. Section 2.1 contains the derivation of an expression for the $2Re\{E_sE_{s*}\}$ term that incorporates the effect of the pupil function and illumination. Section 2.2 shows how the equation can be inverted using multiple measurements.

The WOA approximation breaks down when the amplitude or phase of the object varies significantly. Using the iterative extension presented in Section 2.3, the algorithm can be extended to work with up to 100% amplitude modulation and phase variations as high as ±90°.

2.1. Deriving the weak object transfer function

The first step in using the WOA to derive the transfer function is to separate scattered light, $E_s$, from unscattered light. We define the normalized object field to be $E = 1 + E_s + E_{re} + iE_{im}$ where $E_s$, $E_{re}$, and $E_{im}$ have no DC component. This field is normalized so that the constant term is 1. If the object is weakly scattering, the real part, $1 + E_{re}$, approximately corresponds to the amplitude and the imaginary part, $E_{im}$, corresponds to the phase of the object. Substituting $E$ into Eq. (1) we get

$$I = \int \left| \left( (1 + E_{re} + iE_{im}) \cdot e^{iPh} \right) \cdot \tilde{P}(f') \right|^2 d f'$$

The Fourier transform of the intensity is thus:

$$\tilde{I} = \int \left[ \left( \delta(f - f') \cdot \tilde{P}(f) + \tilde{E}_{re}(f - f') \cdot \tilde{P}(f) + i\tilde{E}_{im}(f - f') \cdot \tilde{P}(f) \right) \cdot \tilde{J}(f') df' \right]$$

$$= \int \left[ \left( \delta(f + f') \cdot \tilde{P}^*(f) + \tilde{E}_{re}(f + f') \cdot \tilde{P}^*(f) - i\tilde{E}_{im}(f + f') \cdot \tilde{P}^*(f) \right) \cdot \tilde{J}(f') df' \right]$$

where $\tilde{P}^*$ is the complex conjugate of the coherent point spread function. Expanding Eq. (3) and applying a change of variables results in

$$\tilde{I} = \tilde{I}_0 + \tilde{I}_s + \int \left( \tilde{J}(f') \cdot \tilde{P}(f') \right) \cdot \left( \tilde{E}_{re}(f') \cdot \tilde{P}^*(f - f') - i\tilde{E}_{im}(f') \cdot \tilde{P}^*(f - f') \right) df' + \int \left( \tilde{E}_{re}(f') \cdot \tilde{P}(f + f') + i\tilde{E}_{im}(f') \cdot \tilde{P}(f + f') \right) \cdot \left( \tilde{J}(f') \cdot \tilde{P}^*(f') \right) df'$$

where $\tilde{I}_0$ is the intensity if there were no scattered light and $\tilde{I}_s$ is the intensity if there were only scattered light. $\tilde{I}_0$ is 0th order in $E_s$ and therefore a constant intensity, $\tilde{I}_s$ is 2nd order in $E_s$ and corresponds to the error term for the WOA, which we will drop. This equation can be written more simply using transfer functions $K_{re}$ and $K_{im}$ as

$$\tilde{I} = \tilde{I}_0 + \tilde{E}_{re} \cdot \tilde{K}_{re} + \tilde{E}_{im} \cdot \tilde{K}_{im} + \tilde{I}_s$$

where the transfer functions that relate the real and imaginary parts of the object field to the measured intensity are:

$$\tilde{K}_{re} = \left( \tilde{P} \cdot \tilde{J} \right) \ast \tilde{P} + \tilde{P} \ast \left( \tilde{P} \cdot \tilde{J} \right)$$

$$\tilde{K}_{im} = \left( \tilde{P} \cdot \tilde{J} \right) \ast \tilde{P} - \tilde{P} \ast \left( \tilde{P} \cdot \tilde{J} \right)$$

$K_{re}$ is the transfer function that maps the real part of the object field to intensity, while $K_{im}$ maps the imaginary part of the object field. For weak objects, this corresponds to amplitude and phase, respectively. This set of equations is still nonlinear because of the $I_s$ term.

Under the WOA, we assume that we can ignore the 2nd order term, $I_s$. This leaves only the DC and the interference between the DC and the scattered light. Thus, Eq. (5) becomes:

$$\tilde{I} \approx \tilde{I}_0 + \tilde{E}_{re} \cdot \tilde{K}_{re} + \tilde{E}_{im} \cdot \tilde{K}_{im}$$

$$I \approx I_0 + E_{re} \ast K_{re} + E_{im} \ast K_{im}$$
From Eq. (9) it is clear that the intensity under the WOA is the sum of two convolutions—one with the real part of the object and one with the imaginary part. The kernels of the convolution can be calculated from the pupil function and the illumination using Eq. (6) and Eq. (7). Since this model is linear, it is possible to solve for the amplitude and phase using a series of intensity measurements where the pupil function and/or illumination source shapes are varied. Collectively we refer to $K_{re}$ and $K_{im}$ as the WOTF.

### 2.2. Inverting the WOTF

To invert the sum of two convolutions, we represent Eq. (9) as a matrix multiplication in Fourier space. If $n$ images, each with different transfer functions (i.e. a different pupil or illumination), are measured, a linear matrix equation can be written for each frequency:

$$
\begin{bmatrix}
\tilde{I}_n(f_i) \\
\vdots \\
\tilde{I}_n(f_i)
\end{bmatrix} =
\begin{bmatrix}
\tilde{K}_{re}^n(f_i) & \tilde{K}_{im}^n(f_i) \\
\vdots & \vdots \\
\tilde{K}_{re}^n(f_i) & \tilde{K}_{im}^n(f_i)
\end{bmatrix}
\begin{bmatrix}
E_{re}(f_i) \\
E_{im}(f_i)
\end{bmatrix}
,$$

(10)

where $\tilde{I}_n(f_i)$ is the Fourier coefficient of measurement $n$ for frequency $f_i$, and $\tilde{K}_{re}^n(f_i)$ and $\tilde{K}_{im}^n(f_i)$ are the corresponding transfer functions. This means that by taking the Fourier transform of the intensity images, this equation can be solved independently for each pixel in frequency space to recover the Fourier transform of the object field.

Under our proposed algorithm, an experiment must capture two or more images where either the pupil or the source are varied, or both. More images result in more rows in the matrix. If the images provide sufficient diversity in the values of $K_{re}$ and $K_{im}$ such that the matrix can be inverted, then that frequency of the electric field can be recovered. In the case of 3 or more images, the overdetermined system is solved using least squares, giving increased noise tolerance. Additionally, by examining the invertibility of the matrix across different frequencies, it is possible to identify optimal measurements to combine.

For a given set of measurements, the matrix may be invertible for some frequencies but not for others. Frequencies well outside the NA of the system, for example, are not generally measured. There will also be frequencies for which either $E_{re}$ or $E_{im}$, but not both, are recoverable. This occurs most commonly for low frequencies of $E_{im}$, because low frequency phase information is often poorly measured and $E_{im}$ is closely related to phase. In this case, Eq. (10) represents an underdetermined system and cannot be inverted directly. Using Truncated Singular Value Decomposition for regularization, we recover only the measured component of the field ($E_{re}$ or $E_{im}$) by suppressing the component that is inadequately measured.

### 2.3. Extending the WOTF beyond weak objects

The limitation of the WOTF model is that it restricts the types of objects that can be imaged. To reduce this limitation, the approximation error term can be estimated and subtracted from the measurements iteratively. If the object is somewhat weak, the model will produce a reasonable estimate of the electric field, $E_s$. This estimate can be used to calculate an estimate of the approximation error term, $I_s$, using Eq. (1). Subtracting the estimate of $I_s$ from the measured intensity, $I$, produces an adjusted intensity, $I_a$, that satisfies the WOA even though the electric field it corresponds to is not necessarily weak. This modified intensity will produce a more accurate estimate of the electric field on the next iteration. The steps of this algorithm are:

1. **Step 1.** Initialize $E_s = 0$; normalize the measured images, $I$
2. **Step 2.** Calculate an estimate of the WOA error term, $I_s$, using the current estimate, $E_s$
3. **Step 3.** Remove the approximation error term to get $I_a = I - I_s$; normalize $I_a$
Step 4. Use the transfer functions to calculate a new $E_s$ from $I_a$
Step 5. Go back to step 2 and use the updated $E_s$

The algorithm assumes that the DC term of the electric field (and of the intensity) is 1. However, the DC of the intensity will in general not be 1 since $I_s$ will contribute to the mean intensity as well. This problem is solved in the iterative algorithm by repeatedly normalizing the adjusted intensity images for each iteration.

An example of the iterative extension is illustrated in Fig. 1 for one image from a simulated focus stack. While in the first iteration there are some strong-phase artifacts, the errors have essentially disappeared by the second iteration. A convergence plot is shown where the error metric is the L2 norm of the difference between the measured image stack and the simulated image stack. The simulated image stack is what the measurements are expected to be, given the estimated complex field. When this error is low, it indicates that the recovered field is consistent with the measurements and implies that the solution is valid.

Fig. 1. Iterative algorithm simulation for an object that does not obey the weak object approximation. A focus stack of 7 images, evenly spaced in the interval $\pm 1.7\lambda/NA^2$, were generated with partial coherence $\sigma = 0.25$. A defocused image ($-1.7\lambda/NA^2$) is shown over a few iterations to illustrate the iterative correction. After one iteration, some error in the recovered phase and amplitude remains, but is mostly removed by the second iteration. The color scale of $I_s$ is offset by 1 to match that of $I$ and $I_a$. 
2.4. Iterative algorithm convergence

Typically, the algorithm converges until noise dominates the remaining error and then it diverges rapidly. Instead of a fixed number of iterations, a non-decreasing error metric over several consecutive iterations indicates the algorithm has started diverging. However, as shown in Fig. 1, the error metric is not always strictly decreasing, but can increase slightly over a few iterations before reducing.

Whether the algorithm converges depends on whether the measured intensity yields a sufficiently close estimate of \( E_s \) such that the approximation is improved during the next iteration. Since the WOA is applied in real space, errors in the solution occur where the field strongly breaks the approximation. As a result, convergence can typically be predicted from the normalized intensity images themselves. Pixels with values close to \((1 - 1)^2 = 0\) and \((1 + 1)^2 = 4\) correspond to locations where \(|E_s|\) has a value close to the DC value. If these pixel values appear in the measurements, the algorithm is likely to produce artifacts near those pixels. It is generally possible to tolerate some of these pixels, if they only occur in a subset of the images. For example, in a focus series, extreme pixels tend to appear mostly near focus for strong amplitude objects and away from focus for strong phase objects.

3. Experiments

To test the algorithm experimentally, we use the SHARP EUV microscope at Lawrence Berkeley National Laboratory. We measured a clear area on an EUV multilayer mask containing a native defect. The defect is most likely a particle embedded in the multilayer and is expected to have both amplitude and phase effects, with the surrounding area having low levels of phase variations from substrate roughness or deposition variations replicated in the multilayer. Since the phase of the roughness and defect are relatively weak, this object satisfies the WOA.

The use of zone plates allows easy manipulation of both the illumination and the pupil function, by patterning the appropriate hologram [32]. For example, a zone plate can be used to create a phase contrast objective lens that, unlike a conventional lens, causes phase objects to produce strong contrast near focus [33–35]. To demonstrate the algorithm’s ability to consider an arbitrary pupil function and partial coherence, measurements were taken using two lenses: a standard zone plate that results in conventional imaging and an unapodized phase contrast zone plate with a circular, 90° phase shifting region of radius \(0.3 \times\) the imaging NA. Partially coherent illumination with \(\sigma = 0.25\) was used with both and the NA of both lenses was 0.0825. More details on the zone plates used can be found in [36] and [37].

Twenty one through-focus images of the defect in the range \(\pm 5 \mu m\) were captured. A subset of the images is shown in Fig. 2 along with the corresponding transfer functions. The most noticeable difference between the two measurements is that the defect is bright at focus for the standard zone plate, but dark for the phase contrast zone plate. Also, the weak phase roughness from the substrate causes strong speckle at focus for the phase contrast zone plate, but not for the standard zone plate. This suggests that the roughness is mainly phase variation and that the defect has both a phase and an amplitude component.

Since the WOA provides a linear mapping of \(E_{re}\) and \(E_{im}\) to intensity, we can plot the transfer function for each in order to explain the contrast variations. Examining the transfer functions, this contrast behavior can be explained by \(K_{re}\) and \(K_{im}\) having their values switched relative to the standard zone plate. The low frequencies are not switched due to the size of the phase shifting region. A strong positive or negative value of the transfer function means that the corresponding frequency of the real or imaginary part of the object is strongly measured in the image and will be recovered. As focus is varied, different frequencies produce contrast in the images in a method similar to the Talbot Effect. This produces the rings visible in the transfer functions. High values of the transfer function outside the NA (dashed circle) indicates that the
Fig. 2. Sample measurements taken using the standard zone plate (top) and phase contrast zone plate (bottom). The defocus distance is indicated at the top, along with the real and imaginary transfer functions in frequency space for each image, with the NA marked by a dashed circle. The illumination is a disk source with $\sigma = 0.25$. 
Fig. 3. The recovered phase in degrees (top) and amplitude (bottom) are shown for the standard zone plate and phase contrast zone plate. The differences between the two estimates of the field are shown on the right, with an arbitrary offset.

partial coherence is increasing the resolution of the measurements, giving phase recovery of frequencies as high as \((1 + \sigma)\text{NA}/\lambda\).

Figure 3 shows the recovered amplitude and phase for each zone plate. Despite very different measurements, the algorithm was able to produce consistent results for the same object using both lenses. The maximum difference between the two fields is 8° for phase and 0.09 for amplitude. This is evidence that the algorithm reproduced the actual field of the object, since it would not be expected to coincidentally produce the same wrong result from two measurements that are qualitatively so different.

4. Conclusion

We have presented an algorithm to extract the phase and amplitude of an object using various modifications to an EUV microscope. The algorithm, based on the Weak Object Transfer Function (WOTF), is extended beyond the approximation by iteratively estimating and correcting the error in the approximation. The algorithm is able to consider arbitrary illumination and an arbitrary pupil functions giving high flexibility in what measurements can be taken. This allows phase imaging to benefit from the resolution enhancement of partial coherence as well as the increased light flux from a larger source. For some objects—particularly objects with large phase variations (> 90°)—the algorithm may not produce a reasonable result, even after extending it iteratively, because the object breaks the WOA too strongly during the first iteration.

We demonstrated the algorithm on measurements of an EUV photomask defect taken on the SHARP zone plate microscope. By imaging an EUV multilayer mask defect using two different lenses—a phase contrast lens and a conventional lens—under partially coherent illumination, we have shown that the algorithm can produce consistent results and is able to consider complicated pupil functions and partial coherence.
Acknowledgments

The authors would like to thank Yow-Gwo (Henry) Wang, Kenneth A. Goldberg, Markus P. Benk, and Antoine Wojdyla for support with experimental data. This research is sponsored by IMPACT+ (Integrated Modeling Process and Computation for Technology). This work was performed in part at Lawrence Berkeley National Laboratory which is operated under the auspices of the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.