# Cramér-Rao lower bound and maximum-likelihood estimation in ptychography with Poisson noise

Wei, Xukang; Urbach, H. Paul; Coene, Wim M.J.

**Important note**
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

# Cramér-Rao lower bound and maximum-likelihood estimation in ptychography with Poisson noise

Xukang Wei [1,*] H. Paul Urbach,[1] and Wim M. J. Coene[1,2]

[1]*Optics Research Group, Imaging Physics Department, Delft University of Technology, 2628 CJ, Delft, The Netherlands*
[2]*ASML Netherlands B.V., 5504 DR, Veldhoven, The Netherlands*

We investigate the performance of ptychography with noisy data by analyzing the Cramér-Rao lower bound. The lower bound of ptychography is derived and numerically computed for both top-hat plane wave and structured illumination. The influence of Poisson noise on the ptychography reconstruction is discussed. The computation result shows that, if the estimator is unbiased, the minimum variance for Poisson noise is mostly determined by the illumination power and the transmission function of the object. Monte Carlo analysis is conducted to validate our calculation results for different photon flux numbers. Furthermore, the performance of the maximum-likelihood method and the approach of amplitude-based cost-function minimization is studied in the Monte Carlo analysis.

## I. INTRODUCTION

Ptychography [1–6] is a scanning coherent diffraction imaging method for reconstructing a complex-valued object function from intensity measurements recorded in the Fraunhofer or Fresnel diffraction region. In ptychography the object is partially illuminated multiple times with varying position of the illumination spot, so that the entire object is covered and adjacent illuminations partially overlap [7]. The technique is found very suitable for extreme ultraviolet (EUV) [8,9] and x-ray imaging applications [10–13] due to its high fidelity and its minimum requirement on optical imaging elements. Moreover, abundant studies show that ptychography is able to provide a wide field of view and retrieve the illumination probe also [14,15]. During the last two decades, ptychography has been successfully demonstrated with x-ray radiation sources [11,16,17], electron beams [18], and visible light sources [19]. More recently, many extensions of ptychography have been proposed, including Fourier ptychography [20–22], spatially partial coherent ptychography [23–25], broadband ptychography [26,27], three-dimensional (3D) ptychography [28–30], on-the-fly scanning ptychograhy [31,32], and interference probe ptychography [33].

For retrieving the object from a ptychographic data set, the key is to find a solution which fulfills both the ptychographic illumination condition in real space and the corresponding measured diffraction intensities in reciprocal space. A commonly used approach for solving the problem is the ptychography iterative engine (PIE) [5,15], which can be derived by sequentially minimizing the distance between the estimated amplitude of the diffracted wave field and the measurements [6]. Another popular choice is the difference map algorithm, which can be formulated in terms of finding the intersection of two constraint sets [14,34]. Based on the augmented Lagrangian methods for solving the conventional constrained optimization problems, several interesting ptychographic algorithms have been developed during the past 10 years [35–38].

However, obtaining a unique reconstruction and a reconstruction with minimum error in ptychography is considered difficult and there is still room for improvement. On the one hand, ambiguities due to a constant scaling factor, a global phase shift, and raster grid pathology, occur in particular when the probe is unknown [39]. Although many algorithms have been presented to enhance the robustness of ptychography [19,35,37,38], a good starting point and proper parameter settings (e.g., update step size, regularization factor, etc.) are needed in general. Furthermore, noise in the measurements of the diffracted intensity cause inaccuracies in the reconstructions [40–42]. To prevent the effect caused by the saturation of the detector, dark-field and near-field ptychography have been introduced [43,44]. Moreover, it was shown that adaptive step-size strategies are able to improve the performance of ptychography in the presence of noise [19,45]. In general, the most powerful and robust denoising methods are based on the maximum-likelihood principle [21,40–42,46]. The likelihood function used in the maximum-likelihood method depends on the noise model. Common choices for the noise model in ptychography are Poisson noise, Gaussian noise, and the mixed Poisson-Gaussian model. It has been demonstrated [22,40,41,47] that, by using the variance stabilization transform given by Bartlett [48] and Anscombe [49], one can approximate the maximum-likelihood method of Poisson noise by the amplitude-based cost minimization algorithm. Therefore, both the approach of maximum-likelihood and the amplitude-based cost-minimization algorithms can be used as a refinement method in ptychography with noisy data.

In this paper our work contains two parts. In the first part, we investigate the Cramér-Rao lower bound (CRLB) for the variance of any unbiased estimator in ptychography [50–52]. We study the lower bound for Poisson distributed photon-counting noise, which is the most dominant source of noise which occurs even under the best experimental

---

*x.wei-2@tudelft.nl

conditions [40,41]. In Sec. II, we briefly discuss ptychography, Poisson photon-counting noise, and the maximum-likelihood method. We compute the Fisher information matrix of ptychography with Poisson noise and introduce the CRLB. In Sec. III, the CRLB is numerically computed and the influence of illumination and of the object is discussed in detail. To validate the obtained CRLB, Monte Carlo analysis is implemented in Sec. IV.

For the second part of this paper, the performance of the maximum-likelihood method and the approach of amplitude-based cost-function minimization is also compared using Monte Carlo simulations. Details of the implementation of the algorithms can be found in the Appendix. We investigate the statistical property of the algorithms for various photon counts in Sec. IV. The paper is concluded with a summary and outlook in the last section.

## II. THEORY

### A. Ptychography, Poisson noise, and maximum-likelihood method

The goal of ptychography is to reconstruct a complex-valued object $O$ from a set of diffraction intensity patterns which are recorded in the Fraunhofer or Fresnel region. Let $\mathbf{r}$ and $\mathbf{r}'$ be two-dimensional (2D) coordinates in the object plane and the detector plane, respectively. The exit wave immediately behind the object is denoted by $\Psi(\mathbf{r})$ and the measured diffraction intensity measurement $I(\mathbf{r}')$. According to the thin object model, the exit wave $\Psi(\mathbf{r})$ for an illumination with a probe function $P(\mathbf{r})$ which is centered on position $\mathbf{R}_m$ is given by

$$
\begin{aligned}
\Psi_m(\mathbf{r}) &= P(\mathbf{r} - \mathbf{R}_m) \cdot O(\mathbf{r}) \\
&= P_m(\mathbf{r}) \cdot O(\mathbf{r}),
\end{aligned} \tag{1}
$$

where the object $O(\mathbf{r})$ can be decomposed into two real-valued functions $A(\mathbf{r})$ and $\phi(\mathbf{r})$:

$$
O(\mathbf{r}) = A(\mathbf{r}) \cdot e^{i\phi(\mathbf{r})}, \tag{2}
$$

where $A$ is the object's local transmission function and $\phi$ stands for the phase of the exit wave immediately behind the object. The probe function is assumed to have a finite support with, for instance, circular boundary:

$$
P(\mathbf{r}) = \begin{cases} P(\mathbf{r}), & |\mathbf{r}| \leqslant r_0 \\ 0, & |\mathbf{r}| > r_0. \end{cases} \tag{3}
$$

For a detector located at distance $z$ in the far field, the diffraction intensity pattern $I(\mathbf{r}')$ for the $m$th illumination

is [53]

$$
\begin{aligned}
I_m(\mathbf{r}') &= \left| \mathcal{F}(\Psi_m)\left(\frac{\mathbf{r}'}{\lambda z}\right) \right|^2 \\
&= \left| \sum_{\mathbf{r}} \Psi_m(\mathbf{r}) \exp\left(-i\frac{2\pi}{\lambda z}\mathbf{r} \cdot \mathbf{r}'\right) \right|^2, \tag{4}
\end{aligned}
$$

where $\mathcal{F}$ is the discrete Fourier transform operator.

The task of ptychography is to find an object function which takes account of the *a priori* knowledge, while a cost function $\mathcal{E}$ is minimized. In our case, the *a priori* knowledge is the exact information of the probe function and the set of relative positions $\mathbf{R}_m$. The cost function $\mathcal{E}$ is defined as the $l_2$ distance between the modulus of the far-field diffraction pattern $|\mathcal{F}(\Psi_m)(\boldsymbol{\xi})|$ and the squared root of the measured intensity $I_m^{\text{mea}}(\boldsymbol{\xi})$:

$$
\mathcal{E} = \sum_m \sum_{\boldsymbol{\xi}} \left[ \sqrt{I_m^{\text{mea}}(\boldsymbol{\xi})} - |\mathcal{F}(\Psi_m)(\boldsymbol{\xi})| \right]^2, \tag{5}
$$

where $\boldsymbol{\xi} = \mathbf{r}'(\lambda z)^{-1}$ is the spatial spectrum coordinate.

From $I_m^{\text{mea}}$, one can estimate the number of detected photons:

$$
n_m(\boldsymbol{\xi}) = \frac{I_m^{\text{mea}}(\boldsymbol{\xi})}{\hbar\omega}, \quad \text{where } \omega = \frac{2\pi c}{\lambda}. \tag{6}
$$

Among a variety of noise models, we consider Poisson noise. The probability distribution of detecting $n_m(\boldsymbol{\xi})$ photons by the detector at every $\boldsymbol{\xi}$ for all measurements ($m = 1, 2, \ldots, M$) is given by

$$
\mathcal{P}_P = \prod_m \prod_{\boldsymbol{\xi}} \frac{N_m(\boldsymbol{\xi})^{n_m(\boldsymbol{\xi})}}{n_m(\boldsymbol{\xi})!} e^{-N_m(\boldsymbol{\xi})}, \tag{7}
$$

where the cumulative product is over both the 2D coordinate $\boldsymbol{\xi}$ and the probe position $\mathbf{R}_m$. The negative log-likelihood functional is defined by

$$
\begin{aligned}
\mathcal{L}_P &= -\ln \mathcal{P}_P \\
&= -\sum_m \sum_{\boldsymbol{\xi}} [n_m \ln N_m - N_m - \ln n_m!]. \tag{8}
\end{aligned}
$$

The average number of photons $N_m(\boldsymbol{\xi})$ depends on the object function $O(\mathbf{r})$ through Eqs. (4) and (6). To find the object function for which the negative log-likelihood functional is maximum, the derivative of $\mathcal{L}_P$ with respect to $O$ is set equal to zero. Hence, for any small perturbation $\delta O$ of the object function it should hold

$$
\mathcal{L}_P(\delta O) = 0, \tag{9}
$$

where

$$
\begin{aligned}
\delta\mathcal{L}_P(\delta O) &= -\sum_m \sum_{\boldsymbol{\xi}} \left(\frac{n_m(\boldsymbol{\xi})}{N_m(\boldsymbol{\xi})} - 1\right) \delta N_m(\delta O(\mathbf{r})) = -\frac{1}{\hbar\omega} \sum_m \sum_{\boldsymbol{\xi}} \left(\frac{n_m(\boldsymbol{\xi})}{N_m(\boldsymbol{\xi})} - 1\right) \delta I_m(\delta O(\mathbf{r})) \\
&= -\frac{2}{\hbar\omega} \sum_m \sum_{\boldsymbol{\xi}} \left(\frac{n_m(\boldsymbol{\xi})}{N_m(\boldsymbol{\xi})} - 1\right) \text{Re}[\mathcal{F}(P_m O)(\boldsymbol{\xi})\mathcal{F}(P_m \delta O(\mathbf{r}))^*(\boldsymbol{\xi})] \\
&= -\frac{2}{\hbar\omega} \sum_m \sum_{\mathbf{r}_i} \text{Re}\left\{\mathcal{F}^{-1}\left[\left(\frac{n_m}{N_m} - 1\right)\mathcal{F}(P_m O)\right](\mathbf{r}_i)P_m^*(\mathbf{r}_i)\delta O^*(\mathbf{r}_i)\right\}. \tag{10}
\end{aligned}
$$

In Eq. (10), the generalized Parseval's theorem was used. Re denotes the real part and $\mathcal{F}^{-1}$ the inverse Fourier transform. The local perturbation of the value of $O$ on a discretized grid $\mathbf{r}_i$ is written as

$$\delta O(\mathbf{r}) = \sum_{\mathbf{r}_i} \delta O(\mathbf{r}_i)\delta(\mathbf{r} - \mathbf{r}_i) = \sum_{\mathbf{r}_i} \begin{bmatrix} \delta A(\mathbf{r}_i) \\ iA(\mathbf{r}_i)\delta\phi(\mathbf{r}_i) \end{bmatrix} e^{i\phi(\mathbf{r}_i)}\delta(\mathbf{r} - \mathbf{r}_i). \tag{11}$$

The solution of Eq. (10) can be found by the method of steepest descent [27,46,54]:

$$\begin{cases} A_{k+1}(\mathbf{r}) = A_k(\mathbf{r}) + \alpha_A \sum_m \text{Re}\{P_m^* e^{-i\phi_k} \mathcal{F}^{-1}[(\frac{n_m}{N_m} - 1)\mathcal{F}(P_m O_k)]\}(\mathbf{r}), \\ \phi_{k+1}(\mathbf{r}) = \phi_k(\mathbf{r}) + \alpha_\phi \sum_m \text{Im}\{P_m^* A_k e^{-i\phi_k} \mathcal{F}^{-1}[(\frac{n_m}{N_m} - 1)\mathcal{F}(P_m O_k)]\}(\mathbf{r}), \end{cases} \tag{12}$$

where $k$ is the iteration number, and $\alpha_A$ and $\alpha_\phi$ are the step sizes, which are normally chosen to be a constant, i.e., they are independent on the iteration number. Im denotes the imaginary part. Alternatively, the projection-based method or conjugate-gradient method can be applied to achieve maximum likelihood [40].

## B. CRLB and the Fisher matrix

In estimation theory, the CRLB gives a lower bound on the variance of any unbiased estimator for the parameter which must be estimated. The estimators that can reach the lower bound are called the minimum variance unbiased estimators. Minimum variance unbiased estimators are often not available [50,55].

We recall the definition of the CRLB, using the notation as in [50]. Suppose we wish to retrieve a real-valued vector parameter $\mathbf{\Theta} = [\theta_1, \theta_2, \dots]^T$ from a set of measurements $\mathbf{X} = [X_1, X_2, \dots]^T$. There are an infinite number of possible outcomes $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_s, \dots$ occurring with probabilities $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_s, \dots$, respectively. To determine the lower bound on the variance of estimator $\hat{\mathbf{\Theta}}$, one computes the Fisher information matrix $I_F$, given by

$$I_F(\mathbf{\Theta}) = -E\left[\frac{\partial^2 \ln \mathcal{P}(\mathbf{X}_s; \mathbf{\Theta})}{\partial \mathbf{\Theta}^2}\right], \tag{13}$$

where $\mathcal{P}(\mathbf{X}_s; \mathbf{\Theta}) = \mathcal{P}_s$ is the conditional probability distribution function and $E$ is the expectation operator. The element $i, j$ of $I_F(\mathbf{\Theta})$ is given by

$$I_F(\mathbf{\Theta})_{ij} = -\sum_s \frac{\partial^2 \ln \mathcal{P}(\mathbf{X}_s; \mathbf{\Theta})}{\partial \theta_i \partial \theta_j} \mathcal{P}(\mathbf{X}_s; \mathbf{\Theta}). \tag{14}$$

The CRLB is then given by the diagonal elements of the inverse of matrix $I_F$, i.e.,

$$\text{Var}(\hat{\theta}_i) \geqslant [I_F^{-1}(\mathbf{\Theta})]_{ii}, \tag{15}$$

where $\text{Var}(\hat{\theta}_i)$ stands for the variance of estimator $\hat{\theta}_i$ for the unknown parameter $\theta_i$.

It is important to note that the estimator based on the maximum-likelihood principle $\hat{\theta}_{\text{ML}}$ asymptotically becomes unbiased and achieves the CRLB for large data sets [50], that is,

$$\hat{\mathbf{\Theta}}_{\text{ML}} \overset{a}{\sim} \mathcal{N}\{\mathbf{\Theta}, \text{diag}[I_F^{-1}(\mathbf{\Theta})]\}, \tag{16}$$

where $\mathcal{N}$ stands for the normal distribution and diag takes the diagonal elements of a matrix.

## C. Fisher matrix with Poisson noise in ptychography

To find the Fisher information matrix, we start by computing the second-order derivative of the likelihood functional $\mathcal{L}_P$ with respect to $O(\mathbf{r})$:

$$\delta^2 \mathcal{L}_P(\delta O, \delta\tilde{O}) = \frac{1}{(\hbar\omega)^2} \sum_m \sum_{\boldsymbol{\xi}} \frac{n_m}{N_m^2} [\delta I_m(\delta O)][\delta I_m(\delta\tilde{O})] - \frac{1}{\hbar\omega} \sum_m \sum_{\boldsymbol{\xi}} \left(\frac{n_m}{N_m} - 1\right)\delta^2 I_m(\delta O, \delta\tilde{O}), \tag{17}$$

where $\delta\tilde{O}$ is the local perturbation of the value of $O$ on a discretized grid as well:

$$\delta\tilde{O} = \sum_{\mathbf{r}_j} \begin{bmatrix} \delta\tilde{A}(\mathbf{r}_j) \\ i\tilde{A}(\mathbf{r}_j)\delta\tilde{\phi}(\mathbf{r}_j) \end{bmatrix} e^{i\tilde{\phi}(\mathbf{r}_j)}\delta(\mathbf{r} - \mathbf{r}_j). \tag{18}$$

By taking the expectation of Eq. (17), we get

$$E(\delta^2 \mathcal{L}_P)(\delta O, \delta\tilde{O}) = \frac{1}{(\hbar\omega)^2} \sum_m \sum_{\boldsymbol{\xi}} E\left\{\frac{n_m}{N_m^2} [\delta I_m(\delta O)][\delta I_m(\delta\tilde{O})]\right\} - \frac{1}{\hbar\omega} \sum_m \sum_{\boldsymbol{\xi}} E\left[\left(\frac{n_m}{N_m} - 1\right)\delta^2 I_m(\delta O, \delta\tilde{O})\right], \tag{19}$$

in which we commute the expectation and summation because the measurements $n_m(\boldsymbol{\xi})$ are independent photon measurements for all pixels $\boldsymbol{\xi}$. Using the properties of the Poisson distribution [52]

$$\sum_{n_m} \frac{N_m^{n_m}}{n_m!} e^{-N_m} = 1, \tag{20a}$$

$$\sum_{n_m} n_m \frac{N_m^{n_m}}{n_m!} e^{-N_m} = N_m, \tag{20b}$$

and using Eq. (10), we find

$$
\begin{aligned}
E(\delta^2 \mathcal{L}_P)(\delta O, \delta \tilde{O}) &= \frac{1}{(\hbar\omega)^2} \sum_m \sum_{\boldsymbol{\xi}} \frac{1}{N_m(\boldsymbol{\xi})} \delta I_m(\delta O) \delta I_m(\delta \tilde{O}) \\
&= \frac{2}{\hbar\omega} \sum_m \sum_{\boldsymbol{\xi}} \mathrm{Re}\left[ \frac{[\mathcal{F}(\Psi_m)(\boldsymbol{\xi})]^2}{I_m(\boldsymbol{\xi})} \mathcal{F}(P_m \delta O)^* \mathcal{F}(P_m \delta \tilde{O})^* \right] + \frac{2}{\hbar\omega} \sum_m \sum_{\boldsymbol{\xi}} \mathrm{Re}[\mathcal{F}(P_m \delta O) \mathcal{F}(P_m \delta \tilde{O})^*] \quad (21)
\end{aligned}
$$

in which we use the following relation:

$$\mathrm{Re}(z_1)\mathrm{Re}(z_2) = \tfrac{1}{2}[\mathrm{Re}(z_1 z_2) + \mathrm{Re}(z_1 z_2^*)], \tag{22}$$

where $z_1$, $z_2$ are arbitrary complex numbers.

From Eqs. (11), (18), and (21) we can derive the discretized Fisher information matrix with respect to the transmission and the thickness function of the object:

$$
\begin{aligned}
I_{F,ij} &= \begin{bmatrix} (I_F)_{AA,ij} & (I_F)_{A\phi,ij} \\ (I_F)_{\phi A,ij} & (I_F)_{\phi\phi,ij} \end{bmatrix} \\
&= \frac{2}{\hbar\omega} \sum_m \begin{bmatrix} \mathrm{Re}[f_m(\mathbf{r}_i, \mathbf{r}_j)] & \mathrm{Im}[A(\mathbf{r}_j)f_m(\mathbf{r}_i, \mathbf{r}_j)] \\ \mathrm{Im}[A(\mathbf{r}_i)f_m(\mathbf{r}_i, \mathbf{r}_j)] & -\mathrm{Re}[A(\mathbf{r}_i)A(\mathbf{r}_j)f_m(\mathbf{r}_i, \mathbf{r}_j)] \end{bmatrix} + \frac{2}{\hbar\omega} \sum_m \begin{bmatrix} |P_m(\mathbf{r}_i)|^2 \delta_{ij} & 0 \\ 0 & A^2(\mathbf{r}_i)|P_m(\mathbf{r}_i)|^2 \delta_{ij} \end{bmatrix}, \quad (23)
\end{aligned}
$$

where the auxiliary function $f$ is given by

$$f_m(\mathbf{r}_i, \mathbf{r}_j) = \mathcal{F}^{-1}\left[ \frac{\mathcal{F}(\Psi_m)}{\mathcal{F}(\Psi_m)^*} \right](\mathbf{r}_i + \mathbf{r}_j) \cdot P_m^*(\mathbf{r}_i) P_m^*(\mathbf{r}_j) e^{-i[\phi(\mathbf{r}_i) + \phi(\mathbf{r}_j)]}, \tag{24}$$

where we used Eq. (4) and the Kronecker's symbol $\delta_{ij}$.

In Eq. (23) we see that the first term is symmetric and the second one is diagonal. The analytical expression for the CRLB, which is obtained from the inverse of the Fisher matrix, cannot be easily derived, but this inverse can be computed numerically. Detailed examples are presented in the next section.

## III. DIRECT CALCULATION OF THE CRLB

As shown in Eq. (15), the CRLB is given by the diagonal elements of the inverse of matrix $I_F$, which can be obtained by numerical computations. In this section, we present the results of some computed CRLB. To investigate how the illumination (i.e., the probe function $P$) and the object $O$ influence the CRLB, we study four cases separately, as described in Table I. Note that only Poisson noise is applied throughout our simulations. Other noise models (e.g., Gaussian noise or Poisson-Gaussian noise [22]) should be included when these are dominant. All of the calculation results given in this section are compared to the Monte Carlo experiment result that are presented in the next section.

For all cases shown in Table I, the probe moves over the object by a $2 \times 2$ regular grid. In line with the conventional ptychography configuration, the overlap ratio between adjacent illuminated areas is 70%. The overlap ratio in each

dimension is defined as follows. Suppose the diameter of the circular support is $L$, and the distance between corresponding points in adjacent illumination positions is $d$, where $0 < d < L$. The overlap ratio is then defined by

$$\text{overlap ratio} = 1 - \frac{d}{L} \tag{25}$$

TABLE I. Four cases that are considered in the computation of the CRLB.

| | |
|---|---|
| Case 1 | Both the transmission and thickness function of the object are uniform. The probe has structured wavefront but uniform illumination power in the circular support. |
| Case 2 | Both the transmission and thickness function of the object are uniform. The probe has structured wavefront and structured illumination power in the circular support. |
| Case 3 | The object has nonuniform transmission but uniform thickness function. The probe is a plane wave with circular support. |
| Case 4 | The object has uniform transmission but nonuniform thickness function. The probe is a plane wave with circular support. |

TABLE II. The characteristic parameters for the simulations.

| Probe | Grid size | Grid spacing | Wavelength | Scanning grid | Overlap ratio | Radius of circular support |
|---|---|---|---|---|---|---|
| | $60 \times 60$ | $1\ \mu$m | 30 nm | $2 \times 2$ | 70% | $30\ \mu$m |
| Object | Grid size | Grid spacing | | Pixel number | Pixel size | Propagation distance |
| | | | Detector | | | |
| | $70 \times 70$ | $1\ \mu$m | | $60 \times 60$ | $50\ \mu$m | 5 cm |

which is usually chosen between 60% and 85% to achieve optimal performance of the reconstruction algorithm [56]. The overlap ratio and the actual probe function are regarded as *a priori* knowledge and employed in the reconstruction algorithm.

The characteristic parameters for the numerical computations are shown in Table II. The object is discretized and zero padded by a $70 \times 70$ square grid with grid spacing 1 $\mu$m. The total illuminated area is roughly $40 \times 40\ \mu$m². The circular probe has radius of 30 $\mu$m and is discretized by a square grid of $60 \times 60$ grid points with grid spacing of 1 $\mu$m. The wavelength is 30 nm. The far-field intensities are measured with a detector at propagation distance of 5 cm behind the object. The detector consists of an array of $60 \times 60$ pixels with a pixel size of 50 $\mu$m. Hence, the maximum spatial frequency (without factor $2\pi$) that is measured is 1 $\mu$m$^{-1}$ and the frequency is sampled with distance $(30)^{-1}\mu$m$^{-1}$.

To compute the CRLB, we first construct the Fisher information matrix $I_F$ using Eq. (23). Although the number of degrees of freedom used to describe the object is small, namely, $70 \times 70 \times 2$ elements, where the factor 2 is due to the fact that the object function is complex, the discretized Fisher matrix already includes $9800 \times 9800$ elements. The CRLB is obtained by numerically computing the inverse of $I_F$. Since $I_{F,ij}$ is a symmetric matrix with real entries, one can apply the eigenvalue decomposition to find the inverse of the Fisher matrix. We select the eigenvalues of $I_F$ that are bigger than a default tolerance, then use these eigenvalues and the corresponding eigenvectors to compute the inverse of $I_F$. This calculation is done by utilizing the "pinv" routine in MATLAB. The diagonal elements of the inverse matrix $I_F^{-1}$ consist of an array of $70 \times 70 \times 2$ elements, of which the first $70 \times 70$ elements correspond to the CRLB of $A(\mathbf{r})$ and the last $70 \times 70$ elements contain the CRLB of $\phi(\mathbf{r})$.

We define the illumination power by means of the total photon-number (PN) counting over the cross section of the probe, given by

$$\mathrm{PN} = \frac{\sum_{\mathbf{r}} |P(\mathbf{r})|^2}{\hbar\omega}. \tag{26}$$

An important property of the CRLB is that it is proportional to the reciprocal of the illumination power. This property follows from the fact that Eqs. (23) and (24) are proportional to the input power or the square of the probe $P_m$. The observation that the CRLB scales with the reciprocal of the illuminating power is confirmed by the computations discussed below.

In the remainder of this section we show the computed CRLB for high illumination power, i.e., PN $= 10^9$, and for low illumination power, i.e., PN $= 10^3$, as examples. The influence of the object and the probe on the CRLB will be discussed separately.

### A. Influence of the illumination on the CRLB

In order to investigate the influence of the illumination on the CRLB, we start by studying cases 1 and 2 described in Table I. For these cases, the actual object, the actual illumination and the computed CRLB are shown in Figs. 1 and 2. We let the object have uniform transmission and thickness function for the time being. For case 1, the probe function $P$ has uniform power throughout its circular support and zero value outside its support, but the phase of the probe has variation in the form of two characters "P" as shown in Fig. 1(a4). On the other hand, the illumination in case 2 has the shape of the character "P" and truncated by the circular support as shown in Fig. 2(a3), and its phase has the same features consisting of two characters P as in case 1 [see Figs. 2(a3) and 2(a4)]. Considering that a perfectly collimated beam is difficult to obtain, we have chosen the wavefront of the illumination to be nonuniform for both cases 1 and 2.

It is seen in Fig. 1 that the CRLB of the object resembles the normalized sum of the intensities of the illuminations shown in Fig. 1(a5). In particular, the part of the object which is illuminated four times reaches a variance approximately four times smaller than the part which is illuminated only once, and this conclusion holds for both the object's
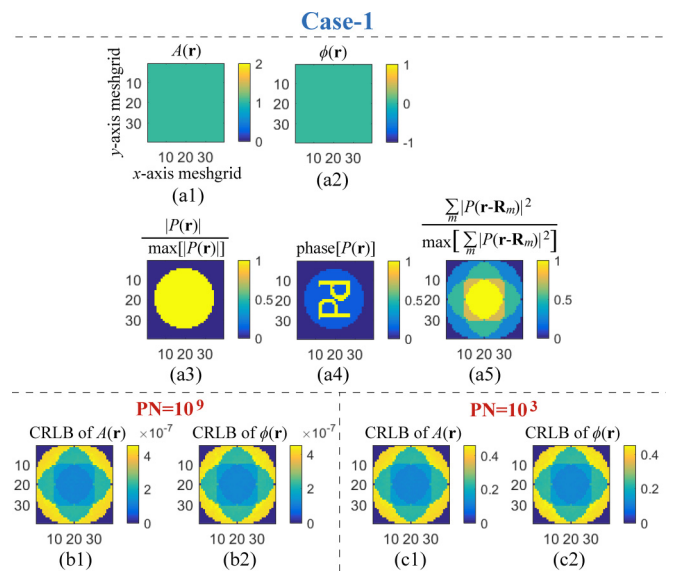


FIG. 1. The CRLB computed from the Fisher matrix for case 1. (a1), (a2) The object's actual transmission $A(\mathbf{r})$ and actual phase function $\phi(\mathbf{r})$, respectively. (a3), (a4) Show the actual amplitude and phase of the probe function, respectively. (a5) Shows the normalized sum of the intensities of the illuminations. (b1), (b2) Show the CRLB of $A(\mathbf{r})$ and $\phi(\mathbf{r})$, respectively, for the case of PN $= 10^9$. (c1), (c2) The CRLB for the case of PN $= 10^3$.
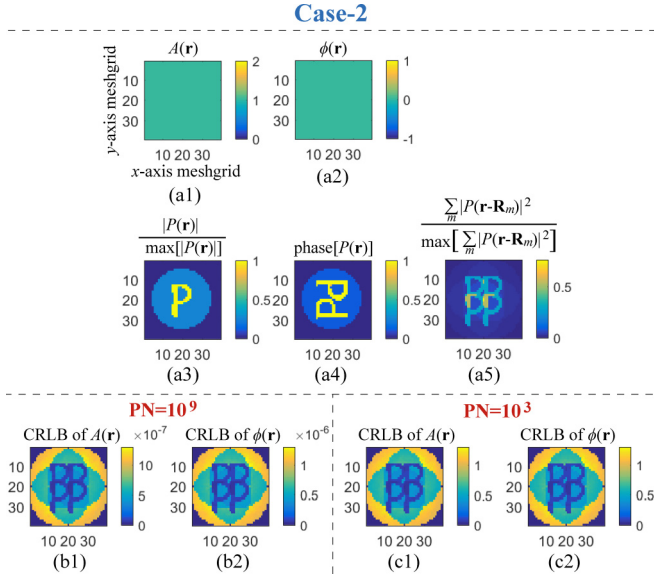
FIG. 2. The calculated CRLB for case 2. (a1)–(a5) The actual object, probe, and the normalized sum of the intensities of the illuminations, respectively. (b1), (b2) The CRLB of $A(\mathbf{r})$ and $\phi(\mathbf{r})$, respectively, for PN $= 10^9$. (c1), (c2) The CRLB for the case of PN $= 10^3$.



FIG. 3. The CRLB for case 3. (a1)–(a5) The actual object, probe, and the normalized sum of the intensities of the illuminations, respectively. (b1), (b2) The CRLB of $A(\mathbf{r})$ and $\phi(\mathbf{r})$, respectively, for PN $= 10^9$. (c1), (c2) The CRLB when PN $= 10^3$.

local transmission $A(\mathbf{r})$ and phase function $\phi(\mathbf{r})$. Interestingly, when the dose distribution of the illumination is more complicated as given in Figs. 2(a3) and 2(a4), the CRLB shown in Figs. 2(b) and 2(c) again resemble the overall illumination pattern shown in Fig. 2(a5). In other words, the more illumination power we apply to the object, the lower the minimum variance of the obtained reconstruction. One can notice that the maximum of the CRLB in Figs. 2(c) and 2(d) is in the yellow corner and is larger than the CRLB in Fig. 1. This is because for case 2 the illuminating power is concentrated in the P character, as shown in Fig. 2(a3). Around the yellow corner there are parts of the object where the computed CRLB is zero. These parts of the object are not illuminated. For the areas where $I_F$ is zero, the computed CRLB is also set equal to zero because we ignore these singular values of $I_F$. In reality, the CRLB there is infinite.

Moreover, we can see in Figs. 1 and 2 that the CRLB is linearly proportional to the inverse of PN (i.e., the illumination power). This calculation result is in agreement with Eq. (23) because the probe function $P(\mathbf{r})$ can be written as the factor $\sqrt{\text{PN}}$ times the normalized $P(\mathbf{r})$. On the other hand, the computed CRLB of both $A(\mathbf{r})$ and $\phi(\mathbf{r})$ do not show any influence due to the spatial variation of the phase of the probe. Therefore, we conclude that it is the illumination intensity pattern, i.e., the dose distribution, which strongly determines the CRLB in ptychography for Poisson noise.

### B. Influence of the object on the CRLB

The Fisher matrix in Eq. (23) is in fact a function of the object, and hence so is the CRLB. To find the influence of $A(\mathbf{r})$ and $\phi(\mathbf{r})$ on the CRLB, we focus on cases 3 and 4 from now on. To reduce the influence of the illumination to a minimum, we let the probe function be a plane wave with
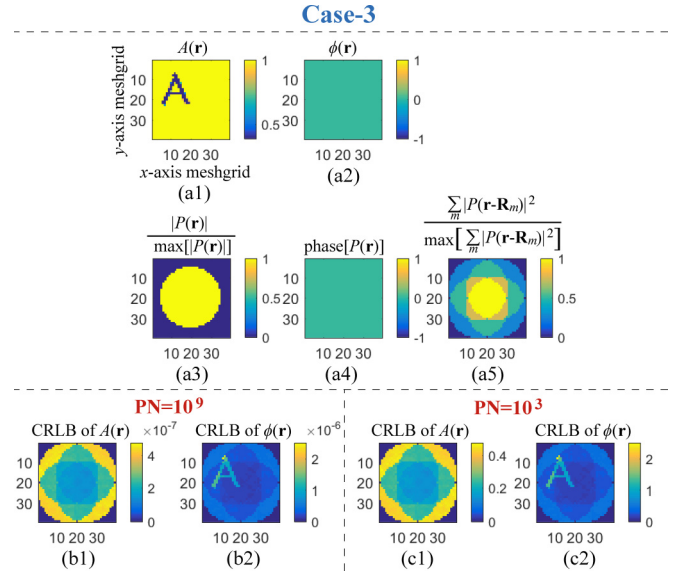
circular support. The influence of the object's transmission and phase function is investigated separately. In case 3 we let the function $A(\mathbf{r})$ have the shape of the character A while $\phi(\mathbf{r})$ is kept uniform, as shown in Fig. 3. The minimum value of $A(\mathbf{r})$ is 0.1. For case 4, the function $A(\mathbf{r})$ is uniform whereas the phase function $\phi(\mathbf{r})$ has the shape of the character T as shown in Fig. 4.

The computed CRLB of the object for cases 3 and 4 is illustrated in Figs. 3(b), 3(c), 4(b), and 4(c), respectively. It is clear that our conclusion in Sec. III A still holds, i.e., the CRLB
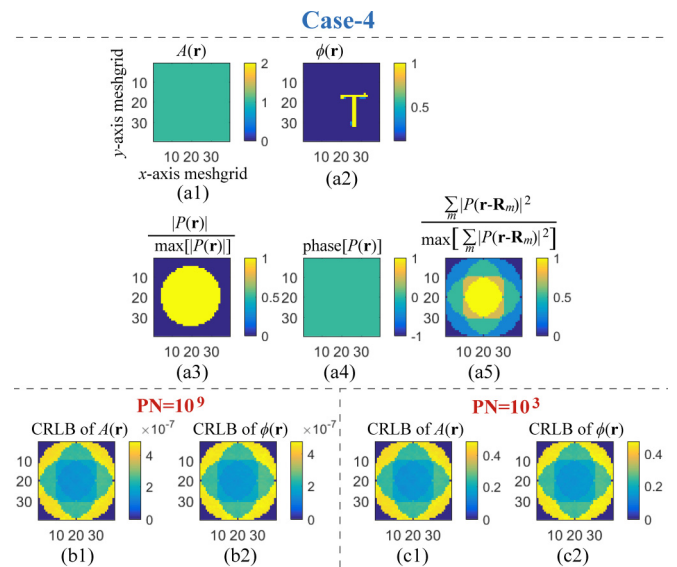


FIG. 4. The CRLB for case 4. (a1)–(a5) The actual object, probe, and the normalized sum of the intensities of the illuminations, respectively. (b1), (b2) The CRLB of $A(\mathbf{r})$ and $\phi(\mathbf{r})$, respectively, when PN $= 10^9$. (c1), (c2) The CRLB for PN $= 10^3$.

is very similar to the pattern of the sum of the intensities of the illuminations. On the other hand, we can see also that the object's local transmission $A$ is predominant in determining the CRLB of $\phi$, as shown in Figs. 3(b2) and 3(c2). This result agrees with Eq. (23) because the function $A$ appears in the terms of $I_F$ which relates to $\phi$. However, the influence of $\phi$ on the CRLB is much less than $A$. Therefore, we conclude that the second term in Eq. (23) is dominant. In other words, when the estimator of ptychography is unbiased, the variance of the object's transmission $A(\mathbf{r})$ is strongly determined by the illumination power and dose distribution, whereas the variance of the object's phase $\phi(\mathbf{r})$ is influenced by both of the transmission $A(\mathbf{r})$, the illumination power, and the dose distribution.

In the next section, the CRLB shown in Figs. 1–4 are used as references for Monte Carlo experiments.

## IV. MONTE CARLO ANALYSIS

To validate our calculation of the CRLB, Monte Carlo computations have been performed. For consistency, we discretize the probe and the object in the same way as described in Table II. The wavelength, object, probe, far-field measurements, and grid sizes are as described in Table II also. The Fresnel number of the system is 0.15. Hence, for this configuration the detector is in the Fraunhofer region.

The ptychographic data with various level of noise is generated as follows. For every ptychography simulation and for every probe position, we first assign the probe function with corresponding photon numbers in accordance with the PN that is chosen. Then, the noise-free diffracted wave field in the far field is calculated, and the Poisson random number generator in MATLAB is applied to generate the noisy data.

To verify the asymptotic property of the maximum-likelihood method of Eq. (16), we developed and implemented Algorithm 1 as described in the Appendix. To mitigate ambiguity problems of ptychography [27], e.g., the global phase shift, the conjugate reconstruction, and the raster grid pathology, it is assumed that the probe used in the Monte Carlo experiment is known. To shorten the computation time and to improve the convergence, the conjugate-gradient method [40,57] is implemented in Algorithm 1.

For comparison, the performance of another popular method, namely, the amplitude-based cost-function minimization approach [6], was investigated in the Monte Carlo experiment also. This is implemented in Algorithm 2. The idea of this algorithm is to retrieve the object by minimizing the cost function defined in Eq. (5). We remark that one can alternatively derive Algorithm 2 from the maximum-likelihood method by using the variance stabilization transform [22,40,41,48,49]. Algorithm 2 is also described in the Appendix.

To investigate the performance of the above-mentioned algorithms, the variance and the squared bias of the estimator are evaluated in our Monte Carlo analysis. Explicitly, the variance of an estimator $\hat{O}(\mathbf{r})$ is defined by [50]

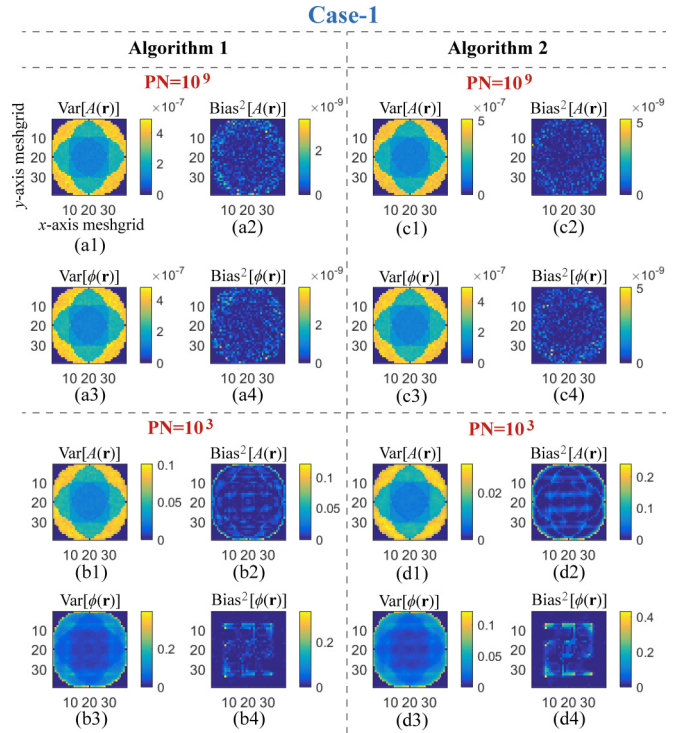$$\text{Var}[\hat{O}(\mathbf{r})] = E\{[\hat{O}(\mathbf{r}) - \langle\hat{O}(\mathbf{r})\rangle]^2\}, \quad (27)$$



FIG. 5. The result of Monte Carlo experiment for Case-1. (a1) and (a2) are the variance and bias squared of the object's transmission $A$ when PN $= 10^9$, respectively, obtained with Algorithm 1. (a3) and (a4) are the variance and bias squared of the object's thickness $\phi$, respectively. (b1)-(b4) show the variance and bias squared when PN $= 10^3$, respectively, obtained with Algorithm 1. (c1)-(c4) and (d1)-(d4) show the results obtained with Algorithm 2 when PN $= 10^9$ and PN $= 10^3$, respectively.

where

$$\langle\hat{O}(\mathbf{r})\rangle = E[\hat{O}(\mathbf{r})], \quad (28)$$

and the squared bias of the estimator is given by

$$\text{Bias}^2[\hat{O}(\mathbf{r})] = |\langle\hat{O}(\mathbf{r})\rangle - O_o(\mathbf{r})|^2, \quad (29)$$

where $O_o$ is the actual object function.

In order to compute the expectation accurately, 2000 individual ptychographic data sets have been generated for all for cases mentioned in Table I and for different value of PN. These data sets have been post-processed by Algorithms 1 and 2, respectively, and the results are discussed next.

### A. Statistic properties of the maximum-likelihood method and the amplitude-based cost-minimization method, and the influence of the illumination

We begin with the case of uniform object function and structured illumination, i.e., cases 1 and 2. For these cases, the actual object and probe function are as in Figs. 1(a) and Fig. 2(a).

When the illumination has a uniform dose distribution but a structured wavefront, the variance and bias of both Algorithms 1 and 2 are shown in Fig. 5. In line with the CRLB given in Fig. 1(b), we see that both algorithms that asymptotically achieve the CRLB when PN $= 10^9$. The squared bias of the

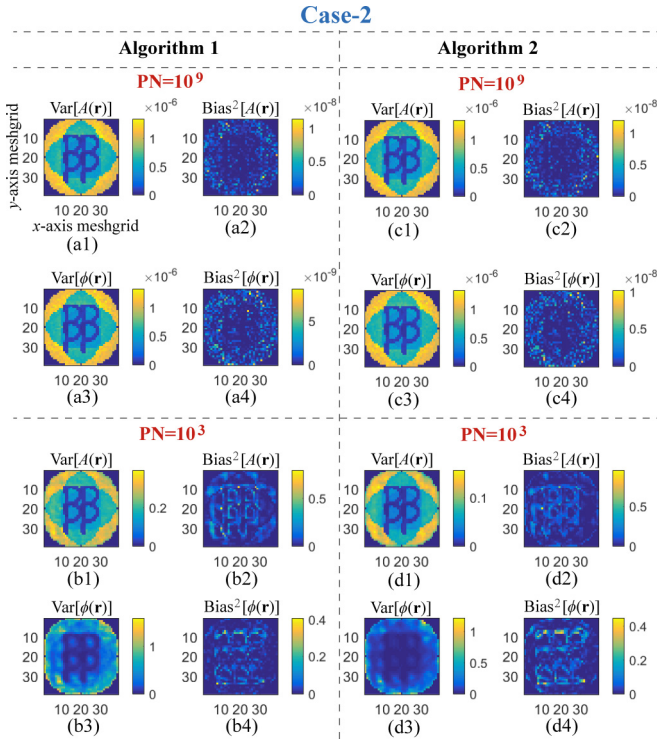FIG. 6. The Monte Carlo experiment result for case 2.



FIG. 7. The Monte Carlo experiment result for case 3.

two algorithms is 100 times smaller than the variance, hence, both Algorithms 1 and 2 are asymptotically unbiased when the photon number is high. Meanwhile, by inspecting Figs. 5(a) and 5(c), one can infer that the variance of both algorithms is related to the local illuminating power as mentioned in Sec. III A, i.e., the parts of the object that are illuminated four times have a variance that is four times smaller than the parts that are illuminated only once. A very similar conclusion can be made for case 2, i.e., when the illumination's local dose distribution is not uniform. As shown in Figs. 6(a) and 6(c), the variance of both algorithms agrees with the CRLB given in Fig. 2(b) and is inversely proportional to the local illumination power given in Fig. 2(a5).

When the photon number is low, i.e., $PN = 10^3$, Algorithms 1 and 2 behave differently with the current data set. In particular, we see in Figs. 5 and 6 that Algorithm 1 in fact reaches smaller bias than Algorithm 2 when the photon number is low. This suggests that the approach based on the maximum-likelihood principle can provide less bias than the amplitude-based cost-function minimization method. Meanwhile, the variance of the estimator Algorithm 2 tends to be smaller than Algorithm 1. This can be explained from the fact that minimizing the amplitude-based cost-function minimization can approximately be regarded as a variance stabilizing denoising algorithm [22,40,41,48]. On the other hand, the two algorithms share certain properties. For low photon count, both Algorithms 1 and 2 have lower variance than the CRLB, which indicates they cannot converge to unbiased estimators and cannot reach the CRLB with the current Monte Carlo data set. More discussion about the case of low photon count is given in Sec. V A.
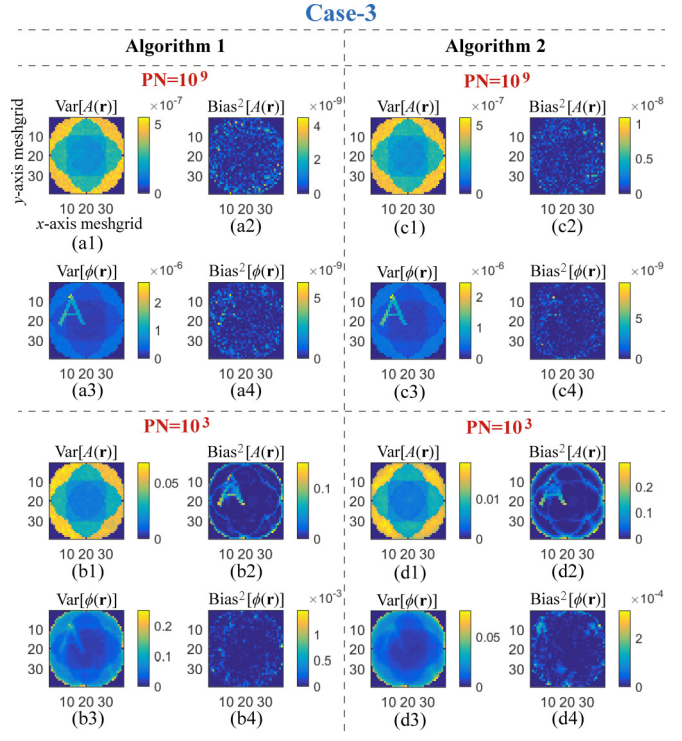
In Figs. 5 and 6 we see that the wavefront profile of the probe only appears in the bias of the reconstruction when the photon count is low. The local illumination power determines the bias for cases 3 and 4 for $PN = 10^3$ as well. For higher photon number, e.g., $PN = 10^9$, there is no trace of the illumination in the bias for case 3 and only negligible trace of illumination's local power for case 4. Therefore, we conclude that the illumination's wavefront profile only influences the statistic property of the algorithms when the photon count is low, whereas the illumination's local power always influences the variance.

### B. Influence of the object on the variance and bias

Next, we consider case 3 where the object has a spatially varying amplitude but the phase is uniform and case 4, where the amplitude is uniform but the phase has variation. In both cases, the probe is a plane wave truncated by a circular aperture. We use the object and probe as in Figs. 3(a) and 4(a). The Monte Carlo results obtained with Algorithms 1 and 2 for case 3 are shown in Fig. 7 and for case 4 in Fig. 8.

When $PN = 10^9$, the variance shown in Figs. 7 and 8 agree with the computed CRLB in Figs. 3 and 4. To be explicit, the variance of the phase of the object $\phi(\mathbf{r})$ is determined by both the object's transmission $A(\mathbf{r})$ and the power of the illumination. The part of the object with lower local transmission will have high variance in reconstruction of the phase. On the other hand, the variance of $A(\mathbf{r})$ is influenced by the sum of the intensities of the illuminations only. These conclusions are true for both algorithms. Meanwhile, we see that the object itself does not influence the bias of the reconstruction when the photon count is high, which means that both algorithms are unbiased for high photon count.
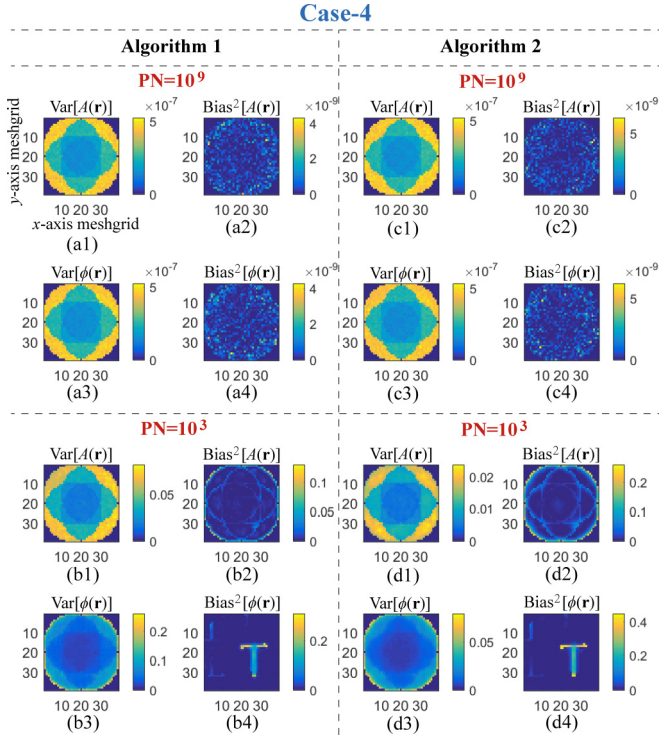
FIG. 8. The Monte Carlo experiment result for case 4.



FIG. 9. The CRLB, variance, and bias-variance ratio of two algorithms for various of values of PN. The value of BVR($\phi$) in (c4) is much smaller than BVR($A$) in (c3), which agrees with the Monte Carlo result shown in Figs. 7(b) and 7(d).

When the photon number is low, i.e., PN $= 10^3$, the profile of the variance deviates from the computed CRLB which is given in Sec. III B. This statement is true for both Algorithms 1 and 2, and is particularly obvious for $\phi(\mathbf{r})$ as shown in Figs. 7 and 8. We can see that there is trace of the actual $A(\mathbf{r})$ in Figs. 7(b2) and in 7(d2), and trace of the actual $\phi(\mathbf{r})$ in Figs. 8(b2) and 8(d2), respectively. This trace indicates that, with the current data set, both algorithms cannot converge to the CRLB for low photon counts.

Interestingly, although the object's transmission $A(\mathbf{r})$ predominantly determines the variance of the object's phase function $\phi(\mathbf{r})$, there is no effect of $A$ on the bias of $\phi$ for any value of PN. In the meantime, we see that $\phi$ do not influence the bias of $A$ for any value of PN, as shown in Figs. 7 and 8. Together with Figs. 5 and 6 in the previous section, we conclude that the profile of the illumination and the object has more influence on the variance of the solutions obtained with Algorithms 1 and 2, more strongly than on the amount of bias.

### C. The CRLB, variance, and bias-variance ratio in ptychography

It is seen in Figs. 5–8 that the ratio of the bias and the variance, as obtained with both algorithms, tend to increase when the photon count is lower. To further investigate this trend and the property of the two algorithms, we define the bias-variance ratio (BVR) of the estimator $\hat{O}$ by

$$\text{BVR}(\hat{O}) = \frac{\sum_{\mathbf{r}} \text{Bias}^2[\hat{O}(\mathbf{r})]}{\sum_{\mathbf{r}} \text{Var}[\hat{O}(\mathbf{r})]}. \quad (30)$$

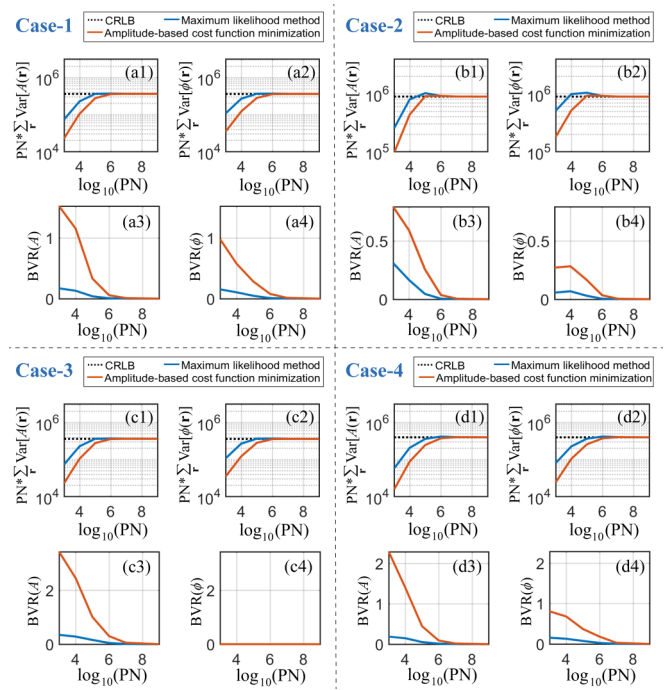In Fig. 9 we show the BVR of Algorithms 1 and 2 for various photon counts and for case 1 to case 4. The overall

CRLB and variance of $A(\mathbf{r})$ and $\phi(\mathbf{r})$ obtained from both algorithms are also shown. We see that the overall variance of both algorithms is the same as the computed CRLB asymptotically when the photon number is high. For lower photon counts, the variance becomes lower than the CRLB, meanwhile, the BVR of both algorithms increases. For our current configuration, this threshold is at PN $= 10^6$. When PN $< 10^6$, the variance of Algorithm 1 is higher than Algorithm 2 for all case 1 to case 4. On the other hand, the BVR of Algorithm 1 is higher than Algorithm 2, which indicates that the Algorithm 1 generally has lower bias than Algorithm 2.

## V. DISCUSSION

### A. Discussion about improving the sufficiency of the ptychographic data set

It is seen in the Monte Carlo results that, for low photon counts, the variance with both Algorithms 1 and 2 is lower than the computed CRLB. This observation indicates that, with the current data set, the two estimators are unbiased for high photon count but cannot converge to the CRLB when the photon count is low.

One may argue that the variances shown in Fig. 9 are lower than the CRLB when PN $< 10^6$ because the current data set is insufficient [50]. In particular, if sufficient amount of data are given, the maximum-likelihood estimator should be asymptotically unbiased and achieves the CRLB if sufficient amount of data are given, as shown in Eq. (16). Indeed, we see in the simulation that Eq. (16) holds when PN $> 10^6$, which indicates that the current data set is already sufficient when PN $> 10^6$. However, for low photon counts, the current
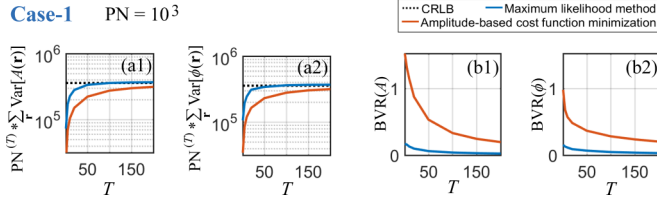
FIG. 10. The CRLB, variance and bias-variance ratio of two algorithms for various number of measurements $T$. This plot is for case 1 and for PN $= 10^3$.

data set is insufficient for the maximum-likelihood estimator to converge to the CRLB.

To explain this fact, we first investigate the signal-to-noise ratio (SNR) of each $m$th ptychographyic measurement with Poisson noise:

$$\text{SNR}_{P,m}(\boldsymbol{\xi}) = \sqrt{n_m(\boldsymbol{\xi})}. \tag{31}$$

For typical far-field diffraction patterns the intensities are not uniform. Hence, the SNR should be a function of $\boldsymbol{\xi}$ and the value of SNR should vary per pixel on the detector. Nevertheless, we can still see that the SNR will in general decrease when the number of photons detected is decreased. Therefore, for Poisson noise, one can extract less and less information about the actual signal when the photon counts are decreasing.

Moreover, we note that the measurement $n_m(\boldsymbol{\xi})$ is discrete and contains natural numbers only, which is associated with the particle nature of light leading to a quantization error at the detector. This discreteness has more disruptive effect on the measurement for the case of low photon counts than the case of high photon counts. Taking an extreme example, suppose only one photon is detected, this photon will most likely appear at $\boldsymbol{\xi} = 0$. Therefore, almost all of the spatial information about the object is lost in the measurement, and hence it is more difficult for estimators to converge to the CRLB.

If we want to increase the size of data set while keeping the current characteristic simulation parameters, one way is to take multiple measurements for each $m$th probe's position. Suppose for each probe's position we take $T$ measurements, denoted by $n_{m,t}(\boldsymbol{\xi})$, where $t = 1, 2, \ldots, T$. A straightforward way to process the data is simply to compute the mean of the measurements:

$$n_m^{(T)}(\boldsymbol{\xi}) = \frac{\sum_t n_{m,t}(\boldsymbol{\xi})}{T}. \tag{32}$$

It has been shown that, when $T$ is large enough, Eq. (32) is a sufficient statistic for Poisson distribution. That is, $n_m^{(T)}$ carries all the information as in the data set: $n_{m,t}$, $t = 1, 2, \ldots, T$. In Fig. 10 the Monte Carlo result with data set $n_{m,t}(\boldsymbol{\xi})$ is shown. To give an example, we study case 1 for low photon counts, i.e., PN $= 10^3$. We note that, by summing over all $T$ measurements, the total photon number PN$^{(T)}$ counting in the probe is now given by

$$\text{PN}^{(T)} = \frac{\sum_{\mathbf{r}} |P(\mathbf{r})|^2}{\hbar\omega} * T = \text{PN} * T, \tag{33}$$

and the CRLB is proportional to the reciprocal of PN$^{(T)}$ according to Eq. (23).

Figure 10(a) shows the computed CRLB and the variance of reconstruction for various of number of measurements $T$. We see that, for both algorithms, the variances approach the CRLB as the number of measurements is increasing. In particular, the variance of Algorithm 1 have reached the CRLB when $T$ is up to 200. Meanwhile, it is seen in Fig. 10(b) that the bias of Algorithm 1 is considerably small compared to the variance when $T > 200$. Therefore, we confirm that, for low photon counts, Algorithm 1 can be asymptotically unbiased and converge to the CRLB by increasing the number of measurements. We see in Fig. 10 that this conclusion is true for Algorithm 2 also. However, the speed of this convergence for Algorithm 2 is slower than for Algorithm 1.

On the other hand, the method proposed in this section can be regarded as a way to increase the total photon flux for each probe's positions. This method can be implemented by taking multiple acquisitions (with number $T$) for each probe's position. In this way, one can illuminate the sample with relatively high photon flux, while avoiding the influence of the limited dynamic range of the detector. Whether there are other approaches to improve the sufficiency of the ptychography data set is above the scope of this paper, but the subject will be studied in the future.

### B. Comparison with the Wigner distribution deconvolution method and discussion on the minimization of the CRLB

Until now we have investigated the statistic property of the iterative ptychographic algorithms. The Wigner distribution deconvolution (WDD) method [2,3,58,59] on the other hand, is a noniterative ptychographic method and provides an approximate closed-form solution to ptychography. The framework of the WDD method is given as follows. We can regard the ptychographic measurement be a four-dimensional (4D) data set with indices $\boldsymbol{\xi}$ and $\mathbf{R}_m$. By taking the Fourier transform of these 4D data set with respect to $\mathbf{R}_m$ and the inverse Fourier transform with respect to $\boldsymbol{\xi}$, we arrive at a 4D array which is denoted by $H$ [2]:

$$H(\mathbf{r}, \mathbf{K}) \approx W_O(\mathbf{r}, \mathbf{K}) \cdot W_P(\mathbf{r}, -\mathbf{K}), \tag{34}$$

where $W_O$ and $W_P$ are the Wigner distribution of the object $O(\mathbf{r})$ and the probe $P(\mathbf{r})$:

$$W_O(\mathbf{r}, \mathbf{K}) = \sum_{\tilde{\mathbf{r}}} O^*(\tilde{\mathbf{r}}) O(\mathbf{r} + \tilde{\mathbf{r}}) \exp(-i\mathbf{K} \cdot \tilde{\mathbf{r}}), \tag{35}$$

where $\tilde{\mathbf{r}}$ is an auxiliary coordinate in object plane. $\mathbf{K}$ is a 2D coordinate in reciprocal space. Suppose $\mathbf{K}$ and $\mathbf{R}_m$ are meshed in rectangular grids and have the spacing

$$\mathbf{K} = [m_x \Delta K_x, m_y \Delta K_y]^T, \tag{36a}$$

$$\mathbf{R}_m = [m_x \Delta R_x, m_y \Delta R_y]^T, \tag{36b}$$

where $m_x = 1, \ldots, M_x$ and $m_y = 1, \ldots, M_y$ are the indices of probe position along the $x$ and $y$ directions. The relation of $\mathbf{K}$ and $\mathbf{R}_m$ is

$$[\Delta K_x, \Delta K_y]^T = 2\pi[(M_x \Delta R_x)^{-1}, (M_y \Delta R_y)^{-1}]^T. \tag{37}$$

Once $H(\mathbf{r}, \mathbf{K})$ has been computed and if we have the knowledge about $P(\mathbf{r})$, we can obtain $W_O(\mathbf{r}, \mathbf{K})$ through an elementwise division. When the measurement is noisy

or the array $W_P$ contains zero value, this division is replaced by applying Wiener filter. To reconstruct the object $O$ from $W_O(\mathbf{r}, \mathbf{K})$, many strategies have been proposed [2,3,58]. Note that the resolution obtained through the WDD method equals to the spacing of the scanning step of the probe. Therefore, in order to achieve the same resolution and the same field of view as obtained from the iterative ptychographic methods, one needs to take finer scanning steps and to record more diffraction measurements in the WDD scheme. However, the line of thinking of the WDD method may provide an alternative insight of the mechanism of ptychography.

For the WDD scheme, it has been proposed [58] that a probe with a strong curved wavefront or a probe created by a diffuser is preferred in practice. This is because such probes are more evenly distributed over the function $W_P(\mathbf{r}, \mathbf{K})$, hence, it is less likely to divide $H(\mathbf{r}, \mathbf{K})$ by zero in Eq. (34). For the iterative ptychographic methods, it has been shown experimentally that using such probes can give a more promising reconstruction than using plane-wave illumination with a finite support [59,60]. Indeed, comparing to plane-wave illumination, such probes ease up the effect of the limited dynamic range of the detector and meanwhile change the distribution of SNR over the detector plane. However, to the best of our knowledge, it is still inconclusive in theory that whether such probes lead to a more noise-robust scheme in ptychography. This is because of the following: (1) For iterative ptychographic methods, one does not compute the division as given in Eq. (34), but instead one updates the reconstruction of the object by applying optimization algorithms, e.g., gradient descent method or alternative projection method. Although regularization techniques are used in iterative ptychographic methods as shown in the Appendix, the regularization does not directly relate to the spatial-frequency spectrum of the probe. (2) For iterative ptychographic methods, the grid size of the probe's position is determined by the overlap ratio as given in Eq. (25). It has been reported that the preferred overlap ratio is 60%–80% [56] so that one can obtain optimal reconstruction in ptychography. Hence, the grid size of the probe's position is smaller than the grid size of the discretized object function. For instance, we have shown in Table II that we can reconstruct a $70 \times 70$ array object with a $2 \times 2$ grid of $\mathbf{R}_m$. Hence, $\mathbf{K}$ is merely meshed on a $2 \times 2$ grid according to Eq. (37). If we wish to apply the WDD analysis to this case, it is reasonable to design a probe which maximizes $W_P(\mathbf{r}, \mathbf{K})$ rather than use a probe which has a evenly distributed spatial-frequency spectrum.

Although the computed CRLB and the Monte Carlo results presented in this paper show that the CRLB is predominantly determined by the local illuminating power and the object's local transmission, we note that the wavefront of the probe may influence the reconstruction. Referring to the first term on the right-hand side of Eq. (23), we see that the Fisher matrix is a function of the actual object and the complex-valued probe. Therefore, for a certain object, it is possible to engineer the probe so that the CRLB is minimized. This minimization process can be done numerically because the CRLB is obtained by computing the inverse of the Fisher matrix. It is likely that for different shape of object we will arrive at different design of the probe. This

optimization of the CRLB is out of the scope of this paper but is certainly an interesting subject which deserves further research.

## VI. CONCLUSION

In the first part of this paper we have studied the influence of Poisson noise on ptychography by analyzing the CRLB. The CRLB was theoretically derived and numerically computed from the Fisher matrix for four different cases. It was found that if the estimator is unbiased, the minimum variance in the presence of Poisson noise is mostly determined by both the illumination's local dose distribution and the object's local transmission. The calculations of the CRLB indicate that the minimum variance is inversely proportional to the number of photons in the illumination beam. The computations of the CRLB using the Fisher matrix were validated with Monte Carlo analysis. It was confirmed that the local illumination power has a strong effect on the variance of the reconstruction of both object's transmission and phase function. Meanwhile, the object's actual local transmission strongly influences the reconstruction of the object's phase.

In the second part of this work, the statistical properties of the maximum-likelihood method and the amplitude-based cost-function minimization algorithm are studied. Both algorithms were applied in the Monte Carlo simulations, using a conjugate-gradient-based implementation. It was shown that both approaches are asymptotically unbiased with variances that are slightly larger than the CRLB when the photon counts are high. For the case of lower photon number, the Monte Carlo analysis showed that both methods require more measurement to converge to the CRLB. While increasing the number of data, it was shown that the maximum-likelihood method converges to the CRLB faster than the amplitude-based cost-function minimization algorithm.

Our result can help to understand the defects that occur in the ptychograghy reconstruction from noisy data. Our conclusions suggest that more illumination power should be given to the part of object which is of most interest, although this may be difficult to realize in a practical ptychography experiment. As next steps of research, the performance of other ptychographic denoising algorithm [19,35,37,38,47] deserve further investigation. Investigating the CRLB and the statistic properties of the two algorithms for Gaussian noise and the mixed Poisson-Gaussian noise is also an interesting topic for further research.

## APPENDIX

The detail of Algorithm 1 is described in the pseudocode. Unlike Eq. (12), the update step size $\alpha$ is not a constant anymore in Algorithm 1. Instead, an optimal $\alpha$ for every iteration $k$ is obtained in the manner described in [61]: (1) Based on the computed $k$th local gradient, calculate the value

**Algorithm 1** Maximum-likelihood method with Poisson noise.

---

1: $k_{\max} = 10^3$, $\delta_{\mathcal{L}} = 10^{-20}$, $\gamma = 10^{-5}$, $A_1 = A_{\mathrm{o}}$,
    $\phi_1 = \phi_{\mathrm{o}}$, $k = 1$.
2: **repeat**
3: compute the steepest descent gradient of $A$ and $\phi$ using Eq. (12):
    $g_{A,k}$
    $= \sum_m -\mathrm{Re}\{P_m^* e^{-i\phi_k} \mathcal{F}^{-1}[(\frac{n_m}{N_m+\gamma} - 1)\mathcal{F}(P_m O_k)]\}$,
    $g_{\phi,k}$
    $= \sum_m -\mathrm{Im}\{P_m^* A_k e^{-i\phi_k} \mathcal{F}^{-1}[(\frac{n_m}{N_m+\gamma} - 1)\mathcal{F}(P_m O_k)]\}$.
4: **if** $k = 1$ **then**
5: $\Delta_{A,k} = g_{A,k}$, $\quad \Delta_{\phi,k} = g_{\phi,k}$.
6: **else**
7: use the formula of Polak-Ribière:
    $\beta_{A,k}^{\mathrm{PR}} = \frac{\langle (g_{A,k}-g_{A,k-1})|g_{A,k}\rangle}{\|g_{A,k-1}\|_2^2}$,
    $\beta_{\phi,k}^{\mathrm{PR}} = \frac{\langle (g_{\phi,k}-g_{\phi,k-1})|g_{\phi,k}\rangle}{\|g_{\phi,k-1}\|_2^2}$,
8: $\beta_{A,k} = \max(\beta_{A,k}^{\mathrm{PR}}, 0)$, $\quad \beta_{\phi,k} = \max(\beta_{\phi,k}^{\mathrm{PR}}, 0)$,
9: compute the conjugate direction:
    $\Delta_{A,k} = g_{A,k} + \beta_{A,k}\Delta_{A,k-1}$,
    $\Delta_{\phi,k} = g_{\phi,k} + \beta_{\phi,k}\Delta_{\phi,k-1}$.
10: **end if**.
11: optimize the update step size:
    $\alpha_{A,k} = \arg\min_{\alpha_A} \mathcal{L}_P(A_k + \alpha_A\Delta_{A,k})$,
    $\alpha_{\phi,k} = \arg\min_{\alpha_\phi} \mathcal{L}_P(\phi_k + \alpha_\phi\Delta_{\phi,k})$.
12: update the object function:
    $A_{k+1} = A_k + \alpha_{A,k}\Delta_{A,k}$, $\quad \phi_{k+1} = \phi_k + \alpha_{\phi,k}\Delta_{\phi,k}$.
13: **if** $k = 11$ **then**
14: $\gamma = 10^{-20}$,
15: **end if**.
16: $k = k + 1$.
17: **until** $k = k_{\max}$ or $|\mathcal{L}_{P,k} - \mathcal{L}_{P,k-1}| \leqslant \delta_{\mathcal{L}}$.

---

**Algorithm 2** Amplitude-based cost-function minimization approach.

---

1: $k_{\max} = 10^3$, $\delta_{\mathcal{E}} = 10^{-20}$, $\gamma = 10^{-3}$, $A_1 = A_{\mathrm{o}}$, $\phi_1 = \phi_{\mathrm{o}}$,
    $k = 1$,
2: **repeat**
3: compute the steepest descent gradient of $A$ and $\phi$:
    $g_{A,k}$
    $= \sum_m -\mathrm{Re}\{P_m^* e^{-i\phi_k} \mathcal{F}^{-1}[(\frac{\sqrt{n_m}}{\sqrt{N_m}+\gamma} - 1)\mathcal{F}(P_m O_k)]\}$,
    $g_{\phi,k}$
    $= \sum_m -\mathrm{Im}\{P_m^* A_k e^{-i\phi_k} \mathcal{F}^{-1}[(\frac{\sqrt{n_m}}{\sqrt{N_m}+\gamma} - 1)\mathcal{F}(P_m O_k)]\}$.
4: follow 4th–10th steps of Algorithm 1.
5: optimize the update step size:
    $\alpha_{A,k} = \arg\min_{\alpha_A} \mathcal{E}(A_k + \alpha_A\Delta_{A,k})$,
    $\alpha_{\phi,k} = \arg\min_{\alpha_\phi} \mathcal{E}(\phi_k + \alpha_\phi\Delta_{\phi,k})$.
6: follow 12th–16th steps of Algorithm 1.
7: **until** $k = k_{\max}$ or $|\mathcal{E}_k - \mathcal{E}_{k-1}| \leqslant \delta_{\mathcal{E}}$.

---

of the likelihood function $\mathcal{L}_P$ for at least three different values of $\alpha$, e.g., [0.01, 0.5, 1]. (2) Approximate $\mathcal{L}_P$ by a quadratic function of $\alpha$. To do this we apply the "polyfit" routine in MATLAB. (3) Choose the value for $\alpha$ for which the quadratic function is minimum. The parameter $\beta_k$ is chosen such that the update direction of the object function is conjugate between two subsequent iterations, for which many proposals exist [62]. Based on the formula of Polak-Ribière [63], we choose $\beta_k = \max(\beta_k^{\mathrm{PR}}, 0)$, where $\beta_k^{\mathrm{PR}}$ is given by

$$\beta_k^{\mathrm{PR}} = \frac{\langle (g_k - g_{k-1})|g_k\rangle}{\|g_{k-1}\|_2^2}, \tag{A1}$$

where $g_k$ is the gradient of $\mathcal{L}_P$ with respect to $O(\mathbf{r})$ in the $k$th iteration. When the calculated $\beta_k^{\mathrm{PR}}$ has negative value, $\beta_k$ resets the search direction from the conjugate gradient back to the local decent gradient direction, i.e., $\Delta_k = g_k$.

In order to prevent that the algorithm terminates in a local minimum, the initial guess of the object is selected to be the

actual object $A_{\mathrm{o}}(\mathbf{r})$ and $\phi_{\mathrm{o}}(\mathbf{r})$. The denominator $N_m$ in Eq. (12) is a function of $\boldsymbol{\xi}$, and may be close to zero for some $\boldsymbol{\xi}$. Hence, the maximum-likelihood method can be unstable. To avoid the instability, a regularization parameter $\gamma$ is introduced in Algorithm 1, of which the value can be determined in practice depending on the noise level. Throughout this paper, we let $\gamma$ be $10^{-5}$ (note that $N_m$ is non-negative integer) for the first 10 iterations, then reset $\gamma$ to $10^{-20}$ after the 10th iteration. Algorithm 1 terminates when the change of the likelihood function between two subsequent iterations is smaller than a threshold $\delta_{\mathcal{L}}$, or when the number of iteration reaches a maximum $k_{\max}$.

For comparison, the performance of another popular method, namely, the amplitude-based cost-function minimization approach [6], is investigated in the Monte Carlo experiment. The approach is described in Algorithm 2, in which the search of the optimal step size $\alpha_k$ and the method of conjugate gradient is added too. Similar to Algorithm 1, Algorithm 2 stops when the change of the cost function between two subsequent iterations is smaller than a threshold $\delta_{\mathcal{E}}$, or when the number of iteration reaches a maximum $k_{\max}$.

Finally, we note that the characteristic parameters shown in the first step of Algorithms 1 and 2, i.e., $k_{max}$, $\gamma$, $\delta_{\mathcal{L}}$, and $\delta_{\mathcal{E}}$ are chosen through a "trial-and-error" process. For each Monte Carlo simulation, we start with a relative large regularization factor, i.e., $\gamma = 10^{-5}$. After the algorithms have converged, we decrease the value of $\gamma$ for final refinement. The error tolerance $\delta_{\mathcal{L}}$ and $\delta_{\mathcal{E}}$ is visually determined such that the algorithms are indeed converged. Throughout the simulation, we observed that the most influential factor which may devastate the simulation result is the choice of the initial guess of the object. How to obtain a initial guess which is closed to the actual object is an important issue in practice but is out of the scope of this paper.

---

[1] W. Hoppe, Beugung im inhomogenen primärstrahlwellenfeld. I. Prinzip einer phasenmessung von elektronenbeugungsinterferenzen, Acta Crystallogr. Sect. A **25**, 495 (1969).

[2] J. M. Rodenburg and R. H. T. Bates, The theory of super-resolution electron microscopy via Wigner-distribution deconvolution, Philos. Trans. R. Soc. A **339**, 521 (1992).

[3] H. N. Chapman, Phase-retrieval x-ray microscopy by Wigner-distribution deconvolution, Ultramicroscopy **66**, 153 (1996).

[4] H. M. L. Faulkner and J. M. Rodenburg, Movable Aperture Lensless Transmission Microscopy: A Novel Phase Retrieval Algorithm, Phys. Rev. Lett. **93**, 023903 (2004).

[5] J. M. Rodenburg and H. M. L. Faulkner, A phase retrieval algorithm for shifting illumination, Appl. Phys. Lett. **85**, 4795 (2004).

[6] M. Guizar-Sicairos and J. R. Fienup, Phase retrieval with transverse translation diversity: A nonlinear optimization approach, Opt. Express **16**, 7264 (2008).

[7] J. C. da Silva and A. Menzel, Elementary signals in ptychography, Opt. Express **23**, 33812 (2015).

[8] M. D. Seaberg, B. Zhang, D. F. Gardner, E. R. Shanblatt, M. M. Murnane, H. C. Kapteyn, and D. E. Adams, Tabletop nanometer extreme ultraviolet imaging in an extended reflection mode using coherent fresnel ptychography, Optica **1**, 39 (2014).

[9] M. Odstrcil, J. Bussmann, D. Rudolf, R. Bresenitz, J. Miao, W. S. Brocklesby, and L. Juschkin, Ptychographic imaging with a compact gas–discharge plasma extreme ultraviolet light source, Opt. Lett. **40**, 5574 (2015).

[10] J. M. Rodenburg, A. C. Hurst, A. G. Cullis, B. R. Dobson, F. Pfeiffer, O. Bunk, C. David, K. Jefimovs, and I. Johnson, Hard-X-Ray Lensless Imaging of Extended Objects, Phys. Rev. Lett. **98**, 034801 (2007).

[11] P. Thibault, M. Dierolf, A. Menzel, O. Bunk, C. David, and F. Pfeiffer, High-resolution scanning x-ray diffraction microscopy, Science **321**, 379 (2008).

[12] H. N. Chapman and K. A. Nugent, Coherent lensless x-ray imaging, Nat. Photonics **4**, 833 (2010).

[13] F. Pfeiffer, X-ray ptychography, Nat. Photonics **12**, 9 (2017).

[14] P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer, Probe retrieval in ptychographic coherent diffractive imaging, Ultramicroscopy **109**, 338 (2009).

[15] A. M. Maiden and J. M. Rodenburg, An improved ptychographical phase retrieval algorithm for diffractive imaging, Ultramicroscopy **109**, 1256 (2009).

[16] M. Holler, M. Guizar-Sicairos, E. H. R. Tsai, R. Dinapoli, E. Müller, O. Bunk, J. Raabe, and G. Aeppli, High-resolution non-destructive three-dimensional imaging of integrated circuits, Nature (London) **543**, 402 (2017).

[17] D. F. Gardner, M. Tanksalvala, E. R. Shanblatt, X. Zhang, B. R. Galloway, C. L. Porter, R. K. Jr, C. Bevis, D. E. Adams, H. C. Kapteyn, M. M. Murnane, and G. F. Mancini, Subwavelength coherent imaging of periodic samples using a 13.5 nm tabletop high-harmonic light source, Nat. Photonics **11**, 259 (2017).

[18] Y. Jiang, Z. Chen, Y. Han, P. Deb, H. Gao, S. Xie, P. Purohit, M. W. Tate, J. Park, S. M. Gruner, V. Elser, and D. A. Muller, Electron ptychography of 2d materials to deep sub-ångström resolution, Nature (London) **559**, 343 (2018).

[19] A. Maiden, D. Johnson, and P. Li, Further improvements to the ptychographical iterative engine, Optica **4**, 736 (2017).

[20] G. Zheng, R. Horstmeyer, and C. Yang, Wide-field, high-resolution Fourier ptychographic microscopy, Nat. Photonics **7**, 739 (2013).

[21] L.-H. Yeh, J. Dong, J. Zhong, L. Tian, M. Chen, G. Tang, M. Soltanolkotabi, and L. Waller, Experimental robustness of Fourier ptychography phase retrieval algorithms, Opt. Express **23**, 33214 (2015).

[22] Y. Zhang, P. Song, and Q. Dai, Fourier ptychographic microscopy using a generalized anscombe transform approximation of the mixed Poisson-Gaussian likelihood, Opt. Express **25**, 168 (2017).

[23] P. Thibault and A. Menzel, Reconstructing state mixtures from diffraction measurements, Nature (London) **494**, 68 (2013).

[24] N. Burdet, X. Shi, D. Parks, J. N. Clark, X. Huang, S. D. Kevan, and I. K. Robinson, Evaluation of partial coherence correction in x-ray ptychography, Opt. Express **23**, 5452 (2015).

[25] J. Zhong, L. Tian, P. Varma, and L. Waller, Nonlinear optimization algorithm for partially coherent phase retrieval and source recovery, IEEE Trans. Comput. Imag. **2**, 310 (2016).

[26] D. J. Batey, D. Claus, and J. M. Rodenburg, Information multiplexing in ptychography, Ultramicroscopy **138**, 13 (2014).

[27] X. Wei and P. Urbach, Ptychography with multiple wavelength illumination, Opt. Express **27**, 36767 (2019).

[28] A. M. Maiden, M. J. Humphry, and J. M. Rodenburg, Ptychographic transmission microscopy in three dimensions using a multi-slice approach, J. Opt. Soc. Am. A **29**, 1606 (2012).

[29] M. A. Gilles, Y. S. G. Nashed, M. Du, C. Jacobsen, and S. M. Wild, 3d x-ray imaging of continuous objects beyond the depth of focus limit, Optica **5**, 1078 (2018).

[30] M. Kahnt, J. Becher, D. Brückner, Y. Fam, T. Sheppard, T. Weissenberger, F. Wittwer, J.-D. Grunwaldt, W. Schwieger, and C. G. Schroer, Coupled ptychography and tomography algorithm improves reconstruction of experimental data, Optica **6**, 1282 (2019).

[31] P. M. Pelz, M. Guizar-Sicairos, P. Thibault, I. Johnson, M. Holler, and A. Menzel, On-the-fly scans for x-ray ptychography, Appl. Phys. Lett. **105**, 251101 (2014).

[32] J. Deng, Y. S. G. Nashed, S. Chen, N. W. Phillips, T. Peterka, R. Ross, S. Vogt, C. Jacobsen, and D. J. Vine, Continuous motion scan ptychography: characterization for increased speed in coherent x-ray imaging, Opt. Express **23**, 5438 (2015).

[33] D. E. B. Flaes and S. Witte, Interference probe ptychography for computational amplitude and phase microscopy, Opt. Express **26**, 31372 (2018).

[34] V. Elser, Phase retrieval by iterated projections, J. Opt. Soc. Am. A **20**, 40 (2003).

[35] Z. Wen, C. Yang, X. Liu, and S. Marchesini, Alternating direction methods for classical and ptychographic phase retrieval, Inverse Probl. **28**, 115010 (2012).

[36] S. Marchesini, A. Schirotzek, C. Yang, H. tieng Wu, and F. Maia, Augmented projections for ptychographic imaging, Inverse Probl. **29**, 115009 (2013).

[37] R. Horstmeyer, R. Y. Chen, X. Ou, B. Ames, J. A. Tropp, and C. Yang, Solving ptychography with a convex relaxation, New J. Phys. **17**, 053044 (2015).

[38] M. Pham, A. Rana, J. Miao, and S. Osher, Semi-implicit relaxed Douglas-Rachford algorithm (sDR) for ptychography, Opt. Express **27**, 31246 (2019).

[39] A. Fannjiang and P. Chen, Blind ptychography: Uniqueness and ambiguities, arXiv:1806.02674.

[40] P. Thibault and M. Guizar-Sicairos, Maximum-likelihood refinement for coherent diffractive imaging, New J. Phys. **14**, 063004 (2012).

[41] P. Godard, M. Allain, V. Chamard, and J. Rodenburg, Noise models for low counting rate coherent diffraction imaging, Opt. Express **20**, 25914 (2012).

[42] H. Chang, P. Enfedaque, J. Zhang, J. Reinhardt, B. Enders, Y.-S. Yu, D. Shapiro, C. G. Schroer, T. Zeng, and S. Marchesini, Advanced denoising for x-ray ptychography, Opt. Express **27**, 10395 (2019).

[43] A. Suzuki and Y. Takahashi, Dark-field x-ray ptychography, Opt. Express **23**, 16429 (2015).

[44] M. Stockmar, P. Cloetens, I. Zanette, B. Enders, M. Dierolf, F. Pfeiffer, and P. Thibault, Near-field ptychography: phase retrieval for inline holography using a structured illumination, Sci. Rep. **3**, 1927 (2013).

[45] C. Zuo, J. Sun, and Q. Chen, Adaptive step-size strategy for noise-robust Fourier ptychographic microscopy, Opt. Express **24**, 20724 (2016).

[46] M. Odstrčil, A. Menzel, and M. Guizar-Sicairos, Iterative least-squares solver for generalized maximum-likelihood ptychography, Opt. Express **26**, 3108 (2018).

[47] A. P. Konijnenberg, W. M. J. Coene, and H. P. Urbach, Model-independent noise-robust extension of ptychography, Opt. Express **26**, 5857 (2018).

[48] M. S. Bartlett, The square root transformation in analysis of variance, Suppl. J. R. Stat. Soc. **3**, 68 (1936).

[49] F. J. Anscombe, The transformation of Poisson, binomial and negative-binomial data, Biometrika **35**, 246 (1948).

[50] S. M. Kay, *Fundamentals Of Statistical Signal Processing, Volume 1: Estimation Theory* (Pearson, London, 2009).

[51] J. N. Cederquist and C. C. Wackerman, Phase-retrieval error: a lower bound, J. Opt. Soc. Am. A **4**, 1788 (1987).

[52] 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 (1993).

[53] J. Goodman, *Introduction to Fourier Optics*, McGraw-Hill Physical and Quantum Electronics Series (McGraw-Hill, New York, 2005).

[54] W. Murray, M. H. Wright, and P. E. Gill, *Practical Optimization* (Emerald Publishing, Bingley, UK, 1982).

[55] D. Bouchet, R. Carminati, and A. P. Mosk, Influence of the Local Scattering Environment on the Localization Precision of Single Particles, Phys. Rev. Lett. **124**, 133903 (2020).

[56] O. Bunk, M. Dierolf, S. Kynde, I. Johnson, O. Marti, and F. Pfeiffer, Influence of the overlap parameter on the convergence of the ptychographical iterative engine, Ultramicroscopy **108**, 481 (2008).

[57] A. Tripathi, I. McNulty, and O. G. Shpyrko, Ptychographic overlap constraint errors and the limits of their numerical recovery using conjugate gradient descent methods, Opt. Express **22**, 1452 (2014).

[58] P. Li, T. B. Edo, and J. M. Rodenburg, Ptychographic inversion via Wigner distribution deconvolution: Noise suppression and probe design, Ultramicroscopy **147**, 106 (2014).

[59] M. Guizar-Sicairos, M. Holler, A. Diaz, J. Vila-Comamala, O. Bunk, and A. Menzel, Role of the illumination spatial-frequency spectrum for ptychography, Phys. Rev. B **86**, 100103(R) (2012).

[60] M. Odstrčil, M. Lebugle, M. Guizar-Sicairos, C. David, and M. Holler, Towards optimized illumination for high-resolution ptychography, Opt. Express **27**, 14981 (2019).

[61] W. Coene, A. Thust, M. O. de Beeck, and D. V. Dyck, Maximum-likelihood method for focus-variation image reconstruction in high resolution transmission electron microscopy, Ultramicroscopy **64**, 109 (1996).

[62] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. (Wiley, New York, 1988).

[63] J. R. Shewchuk, An Introduction to the Conjugate Gradient Method Without the Agonizing Pain, Technical Report, Carnegie Mellon University, 1994 (unpublished).