Supplemental Document

Optics Letters

Optimizing illumination for precise multi-parameter estimations in coherent diffractive imaging: supplement

DORIAN BOUCHET,[1,2,*] JACOB SEIFERT,[1] AND ALLARD P. MOSK[1]

[1]*Nanophotonics, Debye Institute for Nanomaterials Science, Utrecht University, P.O. Box 80000, 3508 TA Utrecht, The Netherlands*
[2]*Present address: Université Grenoble Alpes, CNRS, LIPhy, 38000 Grenoble, France*
[*]*Corresponding author: dorian.bouchet@univ-grenoble-alpes.fr*

Supplement DOI: https://doi.org/10.6084/m9.figshare.13315301

Parent Article DOI: https://doi.org/10.1364/OL.411339

Optimizing illumination for precise multi-parameter estimations in coherent diffractive imaging: supplemental document

This document provides supplementary information to "Optimizing illumination for precise multi-parameter estimations in coherent diffractive imaging".

1. CHOICE OF THE OBJECTIVE FUNCTION

Finding optimal incident fields requires one to clearly define a criterion for optimality, which needs to be attached to a scalar quantity for this criterion to constitute a suitable objective function. In the single-parameter case, the Fisher information is already a scalar quantity and minimizing the CRLB (calculated as the inverse of the Fisher information) constitutes a straightforward objective function [1]. In contrast, for multiple parameter estimations, the Fisher information is a matrix, and the scalar quantity that is to be constructed from this matrix depends on the metrological specifications imposed on the precision that must be achieved in the estimation of each parameter.

Whenever the variance of the estimates averaged over all parameter needs to be minimized, a relevant objective function is constituted by the trace of the inverse of the Fisher information matrix, noted $\mathrm{Tr}(\mathcal{J}^{-1})$. This criterion is invariant under orthonormal transformations, which is a desirable property since it ensures that the optimization procedure yields the same optimal incident field for two equivalent parameterizations. However, this objective function does not guarantee that a controlled threshold value bounds the CRLB for every parameter. Instead, a high CRLB for one parameter can in principle be compensated by a low CRLB for other parameters, since the objective function is constructed as a sum of the $\mathcal{C}_i^2$. This is not a desirable feature for several practical applications, when a tolerance is specified and when all parameters need to be estimated with a precision that is equal or better than the tolerance.

For this reason, in the manuscript, we opted for a different objective function constituted by the spectral radius of the inverse of the Fisher information matrix, noted $\rho(\mathcal{J}^{-1})$. This criterion is also invariant under orthonormal transformations (as opposed for instance to an objective function that would be defined as the largest diagonal coefficient of $\mathcal{J}^{-1}$). Moreover, it directly yields an upper bound that applies to the CRLB of all parameters. Thus, it allows one to easily specify a single tolerance associated with the maximum error that can be made in estimating every parameters.

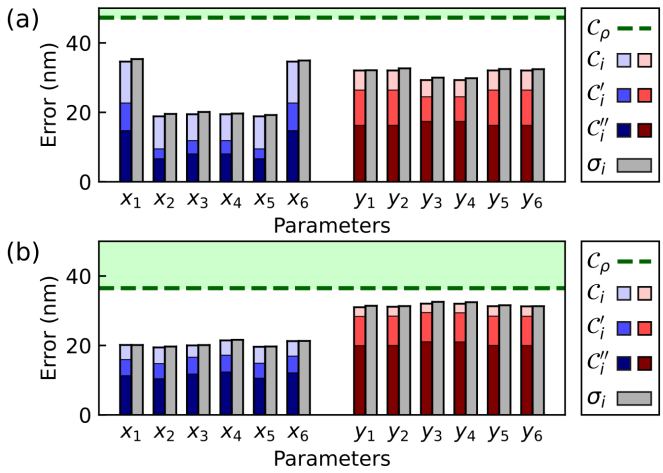

Fig. S1. CRLB for each parameter after the minimization of $\mathrm{Tr}(\mathcal{J}^{-1})$ along with the RMS error obtained by performing ML estimations on 10^4 numerically-generated diffraction patterns, in the case of a) the optimization of the positions of four Gaussian beams and b) the optimization of a zone plate located upstream of the sample. These figures can be compared to Fig. 3i and Fig. 4d of the manuscript, which show the same results after a minimization of $\rho(\mathcal{J}^{-1})$.

In order to compare the results obtained by minimizing $\rho(\mathcal{J}^{-1})$ and $\mathrm{Tr}(\mathcal{J}^{-1})$, we calculated the CRLB for each parameter after the minimization of $\mathrm{Tr}(\mathcal{J}^{-1})$ in the case of an optimization of the positions of four Gaussian beams (Fig. S1a) and in the case of an optimization of the zone plate located upstream of the sample (Fig. S1b). It clearly appears that, in both cases, the resulting CRLB for each parameter is very close to the CRLB obtained by minimizing $\rho(\mathcal{J}^{-1})$ (see Fig. 3i and Fig. 4d of the manuscript). Nevertheless, the CRLB for the first principal component $\mathcal{C}_\rho$ is slightly higher when $\mathrm{Tr}(\mathcal{J}^{-1})$ is minimized. Indeed, the value of $\mathcal{C}_\rho$ is 47 nm (instead of 44 nm) when the four Gaussian beams are optimized, and the value of $\mathcal{C}_\rho$ is 36 nm (instead of 34 nm) when the zone plate is optimized. Thus, assuming that the metrological specifications involve a single tolerance value that applies to all parameters, there is here a slight disadvantage of minimizing $\mathrm{Tr}(\mathcal{J}^{-1})$ instead of $\rho(\mathcal{J}^{-1})$.

2. NUMERICAL METHODS

We implemented a numerical model based on scalar wave propagation at a wavelength of $\lambda = 561\,\mathrm{nm}$. In the detection plane ($z = z_{\mathrm{det}}$), the field $E_l^{\mathrm{det}}(x', y')$ associated with the l-th measured diffraction pattern is represented by a 128×128 complex-valued array, with a pixel size of 6.45 µm. The field in the object plane ($z = 0$) is oversampled by a factor of 8 and thereby represented by a 1024×1024 complex-valued array. The distance between the object and the camera is assumed to be $z_{\mathrm{det}} = 10\,\mathrm{mm}$. The object under consideration $O(x, y)$ is composed of three vertical lines, described with 12 parameters $\boldsymbol{\theta} = (x_1, \ldots, x_6, y_1, \ldots, y_6)$ that correspond to the coordinates of the edges of the lines. The numerical approach that we employ to evaluate the Fisher information matrix requires that the 1024×1024 array representing the object function is differentiable with respect to $\boldsymbol{\theta}$. For this reason we constructed each line by multiplying four sigmoid functions ranging between 0 and 1. With this strategy, the components of $\boldsymbol{\theta}$ are then defined as being the coordinates for which the sigmoid functions take the value $1/2$.

This object function $O(x, y)$ is multiplied by the incident field $E_l^{\mathrm{inc}}(x, y)$ evaluated in the object plane. Within the projection approximation, this procedure yields a correct estimate of the transmitted field $E_l^{\mathrm{obj}}(x, y)$ in the object plane. Introducing the wavenumber $k_0 = 2\pi/\lambda$ and using the angular spectrum representation of plane waves [2], we can calculate the field in the detection plane as follows:

$$E_l^{\mathrm{det}}(x', y') = \frac{1}{4\pi^2} \iint \tilde{E}_l^{\mathrm{obj}}(\alpha, \beta) \exp\left(i\gamma z_{\mathrm{det}}\right) \exp\left[i(\alpha x' + \beta y')\right] \mathrm{d}\alpha \mathrm{d}\beta \,, \tag{S1}$$

where we noted $\gamma = \sqrt{k_0^2 - \alpha^2 - \beta^2}$ and where $\tilde{E}_l^{\mathrm{obj}}(\alpha, \beta)$ is the Fourier transform of $E_l^{\mathrm{obj}}(x, y)$ expressed by

$$\tilde{E}_l^{\mathrm{obj}}(\alpha, \beta) = \iint E_l^{\mathrm{inc}}(x, y) O(x, y) \exp\left[-i(\alpha x + \beta y)\right] \mathrm{d}x \mathrm{d}y \,. \tag{S2}$$

The Fourier transform operations involved in Eqs. (S1) and (S2) can be numerically implemented with a fast Fourier transform (FFT) algorithm. Nevertheless, in order to avoid ringing artifacts, $\tilde{E}_l^{\mathrm{obj}}(\alpha, \beta)$ is first multiplied by a circular aperture function before integration in Eq. (S1). The radius of this aperture function is determined from the effective numerical aperture of the detection apparatus, and we convolve this aperture function with a 2-dimensional Hann function in order to avoid a hard truncation of the field in the frequency domain. Once the field $E_l^{\mathrm{det}}(x', y')$ in the detection plane is calculated according to Eq. (S1), the intensity in this plane is then simply expressed by $I_{k,l} = |E_l^{\mathrm{det}}(x'_k, y'_k)|^2$. The Fisher information matrix is then numerically estimated using $\mathcal{J} \simeq \boldsymbol{H}^{\mathsf{T}} \boldsymbol{H}$, where

$$[\boldsymbol{H}]_{ki} = \sum_l \frac{1}{\sqrt{I_{k,l}(\theta_i) + \epsilon_{\mathrm{r}}}} \left[\frac{I_{k,l}(\theta_i + \Delta\theta) - I_{k,l}(\theta_i - \Delta\theta)}{2\Delta\theta} \right] \,. \tag{S3}$$

All results presented in this work are obtained using a step size $\Delta\theta = 1\,\mathrm{nm}$ and a regularization parameter $\epsilon_{\mathrm{r}} = 0.01$, which has the physical interpretation of being the expected value of an additive noise with Poisson statistics. Results are then found to be insensitive to $\Delta\theta$ and ϵ over several orders of magnitude, attesting that these values yield here an accurate estimate of the Fisher information matrix.

The optimization was performed using a NVIDIA GeForce RTX 2070, which is a commercial graphics processing unit (GPU). The optimization procedure relies on the Adam optimizer [3],

called from TensorFlow libraries. This optimizer takes one main hyperparameter (the learning rate) which must be determined heuristically. We found that learning rates between 0.3 and 0.4 are appropriate to efficiently minimize the objective function, defined as the logarithm of the largest eigenvalue of the inverse of the Fisher information matrix. We used the default values provided by TensorFlow for other the hyperparameters of the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-7}$).

At first, we have found the optimized fields that minimize $\mathcal{C}_\rho$ by tuning the positions of four Gaussian beams and their spatial extent. The full width at half maximum (FWHM) was initialized at 28 µm, and the coordinates of the probes were initialized at ± 0.4 µm (see Fig. 2 of the manuscript). Using the Adam optimizer, we performed 400 iterations with a learning rate of 0.4 to simultaneously optimize the probe positions and the FWHM of the probes. On our GPU, this was achieved in a time of 149 s.

Then, we have found the optimized zone plate that minimize $\mathcal{C}_\rho$ by tuning the amplitudes of the 1024×1024 array representing a zone plate located upstream of the sample. We considered a zone plate with a radius of 180 µm, and we supposed that the distance between the zone plate and the sample is 10 mm. The amplitudes defining the design of the zone plate were initialized with random values taken from a uniform distribution. Using the Adam optimizer, we performed 300 iterations with a learning rate of 0.3 to optimize these amplitudes. On our GPU, this was achieved in a time of 32 s. Note that this time is significantly lower than the time needed to find the optimized probe positions. This difference arises from the fact that four Fisher information matrices need to be evaluated per iteration to find optimal probe positions (one for each probe position), whereas only one Fisher information matrix needs to be evaluated per iteration to identify the optimized zone plate.

3. INFLUENCE OF AN INACCURATE PRIOR KNOWLEDGE

In general, the Fisher information matrix depends on the value taken by all parameters $\boldsymbol{\theta}$ that describe the object. Consequently, optimal illumination schemes depend on the values of $\boldsymbol{\theta}$ that were assumed during the optimization process. These values must be inferred from an *a priori* knowledge of the object, which can be obtained for instance through design considerations or using a low-intensity plane-wave illumination. In order to test what kind of *a priori* knowledge of the object is required for the optimization process to be effective, we study here the precision that can be achieved on the estimation of the parameter y_2 (top edge of the left line) for different illumination schemes. In the following, the origin of the coordinate system in the object plane $(x = 0, y = 0)$ is defined as being located in the center of the middle line.

We first consider the illumination scheme involving four Gaussian probes that minimizes $\mathcal{C}_\rho$ under the hypothesis that $y_2 = 50$ µm, which is equivalent to a total line length of 100 µm and corresponds to the situation considered in the manuscript (Fig. S2a, left). For this illumination scheme, we vary the true value taken by y_2 and we calculate the CRLB associated with the estimation of this parameter (Fig. S2b, dark blue curve). We observe that the CRLB is minimized when the true value of y_2 matches the value assumed during the optimization process ($y_2 = 50$ µm), and that the CRLB remains close to this minimum value within a range of the order of the FWHM of the probe field (15 µm). For larger variations of y_2, the critical area of the sample that depends on this parameter is not properly illuminated by the incident field, resulting in a higher CRLB. For comparison purposes, we also identify the illumination scheme involving four Gaussian probes that minimizes $\mathcal{C}_\rho$ under the hypothesis that $y_2 = 60$ µm (Fig. S2a, right). For this new illumination scheme, the CRLB is minimized for this value of y_2 and remains close to this minimum value within a relatively large range (Fig. S2b, light blue curve), which confirms that the optimization procedure is here robust with respect to an imperfect *a priori* knowledge of the sample.

For completeness, we perform the same analysis for the illumination scheme that involves a zone plate minimizing $\mathcal{C}_\rho$, as described in the manuscript (see Fig. 4 of the manuscript). Similarly, we observe that the CRLB is minimized when the true value of y_2 matches the value assumed during the optimization process, which is either $y_2 = 50$ µm (Fig. S2c, dark red curve) or $y_2 = 60$ µm (Fig. S2c, light red curve). Note that all optimized fields that we identified here are optimally shaped for the simultaneous estimation of all 12 parameters describing the object. The complex shape of the resulting illumination patterns in the object plane (see for instance the intensity distribution shown in Fig. 4b of the manuscript) can then give rise to a non-convex dependence of the CRLB upon the value taken by the parameters, as observed in Fig. S2c (light red curve).

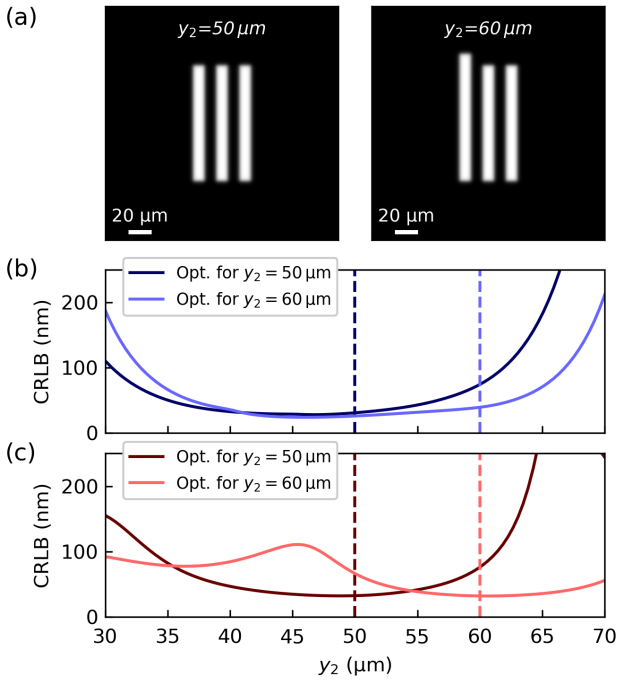

Fig. S2. a) Object function under the hypothesis that $y_2 = 50\,\mu\text{m}$ (left) and under the hypothesis that $y_2 = 60\,\mu\text{m}$ (right). b) CRLB for the parameter y_2 as a function of the true value taken by this parameter in the case of the optimization of the positions of four Gaussian beams. c) CRLB for the parameter y_2 in the case of the optimization of a zone plane located upstream the sample. Opt. stands for Optimized (dark and light curves represent the CRLB obtained after an optimization procedure performed under the hypothesis $y_2 = 50\,\mu\text{m}$ and $y_2 = 60\,\mu\text{m}$, respectively).

REFERENCES

1. D. Bouchet, S. Rotter, and A. P. Mosk, "Maximum information states for coherent scattering measurements," arXiv:2002.10388 (2020).
2. J. W. Goodman, *Introduction to Fourier Optics* (W. H. Freeman, 2017).
3. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *3rd International Conference on Learning Representations, San Diego*, (2015).