Letter Vol. 48, No. 2 / 15 January 2023 / Optics Letters 485

Remote referencing strategy for high-resolution coded ptychographic imaging

Tianbo Wang,[1,†] Pengming Song,[1,†] Shaowei Jiang,[1,†] Ruihai Wang,[1] Liming Yang,[1] Chengfei Guo,[1,2] Zibang Zhang,[3] and Guoan Zheng[1,*]

[1]Department of Biomedical Engineering, University of Connecticut, Storrs, Connecticut 06269, USA
[2]Hangzhou Institute of Technology, Xidian University, Hangzhou, China
[3]Department of Optoelectronic Engineering, Jinan University, Guangzhou 510632, China
*Corresponding author: guoan.zheng@uconn.edu
†These authors contributed equally to this work

Received 18 November 2022; revised 16 December 2022; accepted 17 December 2022; posted 19 December 2022; published 11 January 2023

The applications of conventional ptychography are limited by its relatively low resolution and throughput in the visible light regime. The new development of coded ptychography (CP) has addressed these issues and achieved the highest numerical aperture for large-area optical imaging in a lensless configuration. A high-quality reconstruction of CP relies on precise tracking of the coded sensor's positional shifts. The coded layer on the sensor, however, prevents the use of cross correlation analysis for motion tracking. Here we derive and analyze the motion tracking model of CP. A novel, to the best of our knowledge, remote referencing scheme and its subsequent refinement pipeline are developed for blind image acquisition. By using this approach, we can suppress the correlation peak caused by the coded surface and recover the positional shifts with deep sub-pixel accuracy. In contrast with common positional refinement methods, the reported approach can be disentangled from the iterative phase retrieval process and is computationally efficient. It allows blind image acquisition without motion feedback from the scanning process. It also provides a robust and reliable solution for implementing ptychography with high imaging throughput. We validate this approach by performing high-resolution whole slide imaging of bio-specimens.

https://doi.org/10.1364/OL.481395

Ptychography was first developed for solving the phase problem in electron crystallography [1]. It has grown rapidly in recent years and attracted attention from different research communities. In a typical implementation, the object is translated through a spatially confined probe beam and the corresponding diffraction patterns are recorded in the far-field [2]. The reconstruction process iteratively imposes two sets of constraints. In the spatial domain, the spatially confined probe beam serves as the support constraint to limit the physical extent of the object for each measurement. In the Fourier domain, the diffraction measurements serve as the Fourier magnitude constraints for the estimated solution. Ptychography does not require a reference beam as in holography. It also lifts the isolated-object requirement of conventional coherent diffraction imaging approaches [3]. In the past few years, it has become an indispensable imaging tool in most synchrotron and national laboratories worldwide [4].

In the visible light regime, the new development of coded ptychography (CP) enables high-resolution, high-throughput optical imaging in a lensless configuration [5,6]. Figure 1(a) shows the schematic of a typical implementation of CP, where the light waves propagate from the object plane to the coded surface plane, and then to the detector plane. As shown in Fig. 1(b), the coded surface, denoted as $cs(x', y')$, can be formed by coating a layer of microbeads [7,8] or by smearing a monolayer of blood cells on the sensor's coverglass [6,9]. Planes (x, y), (x', y'), and (x'', y'') in Fig. 1 denotes different defocus planes. In the following, we use them interchangeably for representing the lateral coordinates. The coded surface serves as a high-resolution scattering lens with a theoretically unlimited field of view. We note that the sensitivity of the pixel array underneath the coded surface falls off quickly for light waves with large incident angles [5]. With CP, the coded surface can redirect the large-angle diffracted waves into smaller angles for detection. As such, the otherwise inaccessible high-resolution details can be acquired using the pixel array underneath. Previous demonstrations of CP rely on a clear region on the sensor surface for positional tracking [5,6,8,9], thereby sacrificing the valuable imaging area and throughput. Furthermore, the tracking process becomes unreliable when imaging a sparse sample or the edge part of a dense sample. In these cases, there will be no object at the clear region for positional tracking. Here we aim to resolve the positional tracking problem caused by the coded surface in CP and provide a reliable solution for implementing CP by researchers in different fields.

In the image acquisition process of CP, the object (or the coded image sensor) is translated to different positions (x_i, y_i) and the corresponding diffraction measurements $I_i(x, y)$ are captured for reconstruction. The imaging model can be expressed as

$$I_i(x, y) = |[w(x - x_i, y - y_i) \cdot cs(x, y)] * psf_{d2}|^2, \quad (1)$$

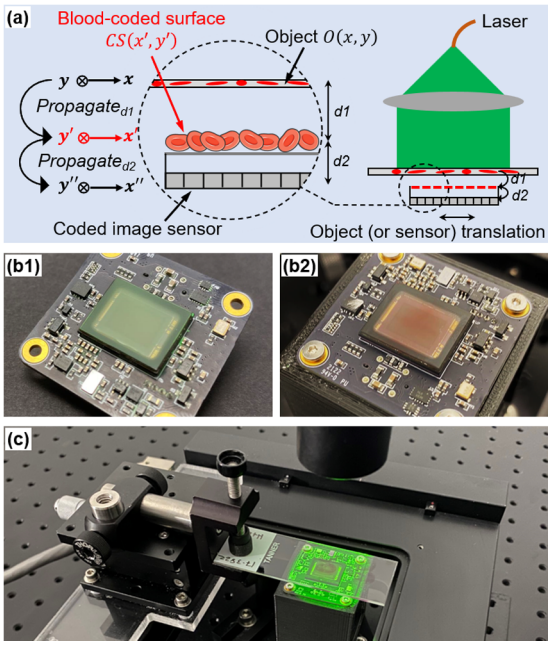

Fig. 1. (a) Schematic of CP. The coded surface can be made (b1) by coating the sensor's coverglass with a thin layer of microbeads, or (b2) by smearing a drop of blood on the sensor's coverglass followed by alcohol fixation. (c) Prototype setup of CP.

where $w(x, y)$ is the object exit wavefront at the coded surface plane, psf_{d2} presents the free-space propagation kernel for a distance of d_2, and "*" represents convolution.

In our experiment, we used the Sony IMX 226 sensor for image acquisition and the ASI MS-2000 stage for sample translation. The scanning step size is 1–3 microns between adjacent acquisitions with $d_2 = 840$ µm. A 10-mW collimated laser beam is used for sample illumination and the exposure time is ~1 ms. By acquiring images at 30 frames per second, the continuous scanning process generates a negligible motion blur of ~40 nm. In contrast, conventional ptychography has a large scanning step size. One needs to fully stop the stage for image acquisition, preventing its operation at the full camera frame rate.

A high-quality reconstruction of CP relies on the precise tracking of the object's positional shifts (x_i, y_i)s. However, the coded surface $cs(x, y)$ modulates the object's light waves and prevents the use of cross correlation analysis for motion tracking. To derive the closed-form correlation map between the first image $I_1(x, y)$ and the ith image $I_i(x, y)$, we make a first-order approximation to Eq. (1):

$$\begin{aligned} &I_i(x, y) \\ &\approx |((1 + \Delta w(x - x_{ref}, y - y_{ref})) \cdot (1 + \Delta cs(x, y))) * psf_{d2}|^2 \\ &\approx 1 + \Delta w'(x - x_i, y - y_i) + \Delta cs'(x, y), \end{aligned} \tag{2}$$

where Δw, Δcs are the first-order expansions of the object exit wavefront and the coded surface profile, and $\Delta w'(x - x_i, y - y_i) = 2Re(\Delta w(x - x_i, y - y_i) * psf_{d2} \cdot e^{-i \cdot d_2 \cdot \frac{2\pi}{\lambda}})$, $\Delta cs'(x, y) = 2Re(\Delta cs(x, y) * psf_{d2} \cdot e^{-i \cdot d_2 \cdot \frac{2\pi}{\lambda}})$. Here "$Re$" represents the real part of the complex expression, and λ is the wavelength. With Eq. (2), we can calculate the correlation map $R_{I_1 I_i}(x, y)$ between $I_1(x, y)$ and $I_i(x, y)$ as follows:

$$\begin{aligned} &R_{I_1 I_i}(x, y) = \mathcal{F}^{-1}\{\mathcal{F}(I_1(x, y)) \cdot conj\{\mathcal{F}(I_i(x, y))\}\} = Constant \\ &+ R_{\Delta w'}(x - x_i, y - y_i) + R_{\Delta cs'}(x, y), \end{aligned} \tag{3}$$

where "$conj$" represents the complex conjugate of the expression, "$\mathcal{F}$" represents the Fourier transform operation, and $R_{\Delta w'}$ and $R_{\Delta cs'}$ represent the autocorrelation maps of $\Delta w'(x, y)$ and $\Delta cs'(x, y)$, respectively. The goal of the motion tracking process is to recover the positional shift (x_i, y_i) from the correlation map $R_{I_1 I_i}(x, y)$ in Eq. (3). We can see that the map contains two peaks, one at (x_i, y_i) and the other at (0, 0). The strong peak caused by the coded surface prevents us to precisely locate the peak at the position (x_i, y_i).

To address this issue, we developed the following pipeline with two main steps: (1) reduce the impact of the coded surface profile via a remote referencing strategy; and (2) enhance the object profile by smearing out the coded surface profile. The unique combination of these two steps allows us to minimize the impact caused by the coded surface in Eq. (3). Figure 2 shows the overview of the proposed pipeline. In step 1, we capture a reference image $I_{ref}(x, y)$ by moving the object to a remote location (typically >2 mm away from other scanning positions). Similar to Eq. (2), this reference image can be approximated as

$$I_{ref}(x, y) \approx 1 + \Delta w'(x - x_{ref}, y - y_{ref}) + \Delta cs'_{remote}(x, y). \tag{4}$$

We can then calculate the correlation map between $I_{ref}(x, y)$ and $I_i(x, y)$ as follows:

$$\begin{aligned} &R_{I_{ref} I_i}(x, y) = \mathcal{F}^{-1}\left\{\mathcal{F}\left(I_{ref}(x, y)\right) \cdot conj\left\{\mathcal{F}\left(I_i(x, y)\right)\right\}\right\} \approx Constant \\ &+ R_{\Delta w'}\left(x - \left(x_i + x_{ref}\right), y - \left(y_i + y_{ref}\right)\right) + R_{\Delta cs'_{remote} \Delta cs'}(x, y), \end{aligned} \tag{5}$$

where $R_{\Delta cs'_{remote} \Delta cs'}$ represents the cross correlation map between $\Delta cs'(x, y)$ and $\Delta cs'_{remote}(x, y)$. In contrast with Eq. (3), this term is a constant because different regions of the blood-coded surface are not correlated with each other. Therefore, Eq. (5) only contains one peak at the position (x_i, y_i). In Figs. 2(a) and 2(b), we can then recover the initial positional shift by locating the peak of $R_{I_{ref} I_i}(x, y)$ as follows:

$$(x_i + x_{ref}, y_i + y_{ref}) = \underset{(x,y)}{\operatorname{argmax}} R_{I_{ref} I_i}(x, y), \tag{6}$$

where x_{ref} and y_{ref} are constants and can be subtracted to obtain x_i and y_i. The second step in Fig. 2(c) is to enhance the object

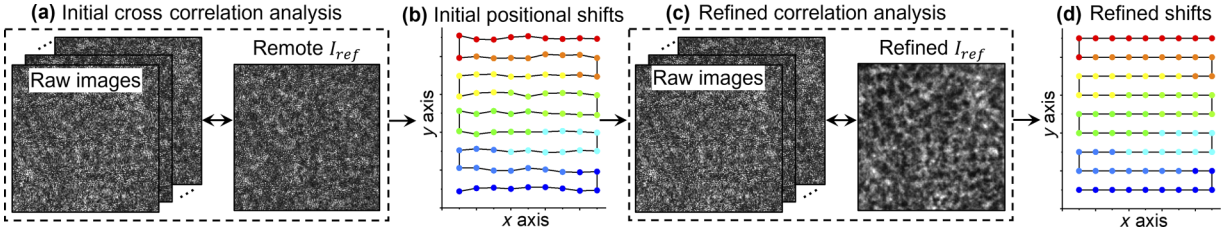

Fig. 2. (a) Cross correlation analysis between the raw measurements and a remote reference image for obtaining the initial positional shifts. (b) Initial positional shifts. (c) Reference is updated with Eq. (7). (d) Updated position shifts based on Eq. (8).

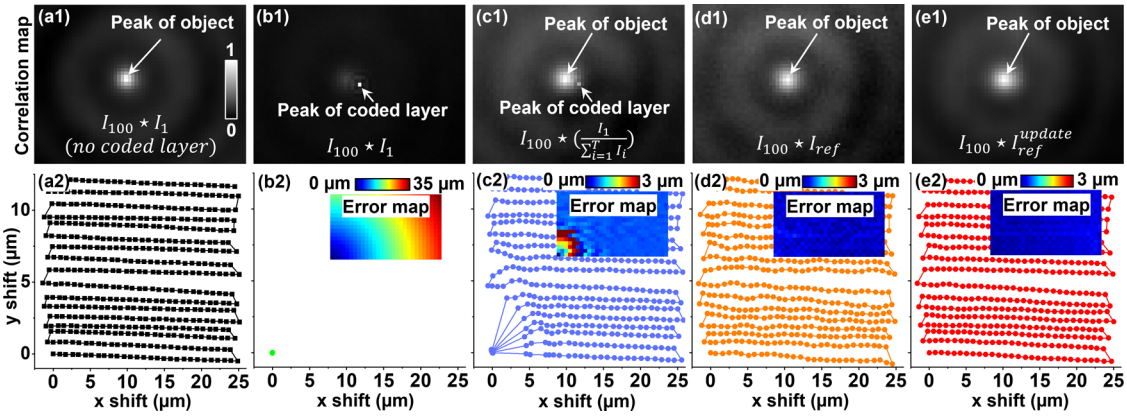

Fig. 3. (a) Ground-truth cross correlation analysis between two raw images captured without the coded layer. (b) Same analysis for two raw measurements captured with the coded layer. (c) Minimizing the impact of the coded layer by dividing the sum of the measurement. (d) Proposed remote referencing strategy by moving the object to a distanced position. (e) With the refinement step of Eqs. (7) and (8).

profile by smearing out the coded surface profile. We use the following equation to generate the updated reference image:

$$I_{ref}^{update}(x, y) = \sum_{i=1}^{T} I_i(x + x_i - x_1, y + y_i - y_1), \tag{7}$$

where measurements are shifted back based on the initial positional shift obtained from Eq. (6). The positional shift can be updated accordingly:

$$(x_i^{update}, y_i^{update}) = \underset{(x,y)}{\operatorname{argmax}} R_{I_{ref}^{update} I_i}(x, y)\ . \tag{8}$$

We typically repeat Eqs. (7) and (8) three times to obtain the final positional shift in Fig. 2(d).

Figures 3(a1)–3(e1) show the correlation maps of different cases and Fig. 3(a2)–3(e2) show the corresponding positional shifts of the object. The sign " $\star$" in Figs. 3(a1)–3(e1) represents the cross correlation operation. In this experiment, we clear part of the coded surface and use it as the ground truth for positional tracking in Fig. 3(a). An efficient sub-pixel registration approach with deep sub-pixel accuracy is adopted in our analysis [10]. Figure 3(b1) shows the correlation map of two captured raw images with the coded layer presented. From this map, we can see one peak at (x_i, y_i) and the other at (0, 0). The strong peak caused by the coded surface prevents us to locate the peak at the position (x_i, y_i), especially when the shifts are small. Figure 3(b2) shows the recovered positional shifts and the error map calculated based on the ground-truth positions in Fig. 3(a2).

In Fig. 3(c1), we use the first raw image as the reference and minimize the impact of the coded surface by dividing it by the sum of all images [11]. This strategy is similar to dividing an image captured in the absence of the object. A better tracking performance can be achieved but the coded surface peak still causes problems for small positional shifts. Figure 3(d) shows the performance of the proposed remote referencing strategy, where the peak of the coded surface has been removed. With the refinement step in Eqs. (7) and (8), we can obtain deep sub-pixel accuracy in Fig. 3(e). Table 1 summarizes the performances of different cases.

Table 1. Summary of the Motion Tracking Performance

Reference Image	Mean Error	Standard Deviation
First raw image	14.11 μm	6.57 μm
Divide the sum	0.75 μm	0.55 μm
Remote referencing	0.22 μm	0.1 μm
With refinement	0.2 μm	0.08 μm

We validate the imaging performance using a mouse kidney slide Fig. 4. The raw image is shown in Fig. 4(a) and CP reconstructions using different motion tracking strategies are shown in Figs. 4(b)–4(e). As expected, the remote referencing strategy with refinement gives us the best overall image quality in Fig. 4(e). With the proposed remote referencing strategy and its

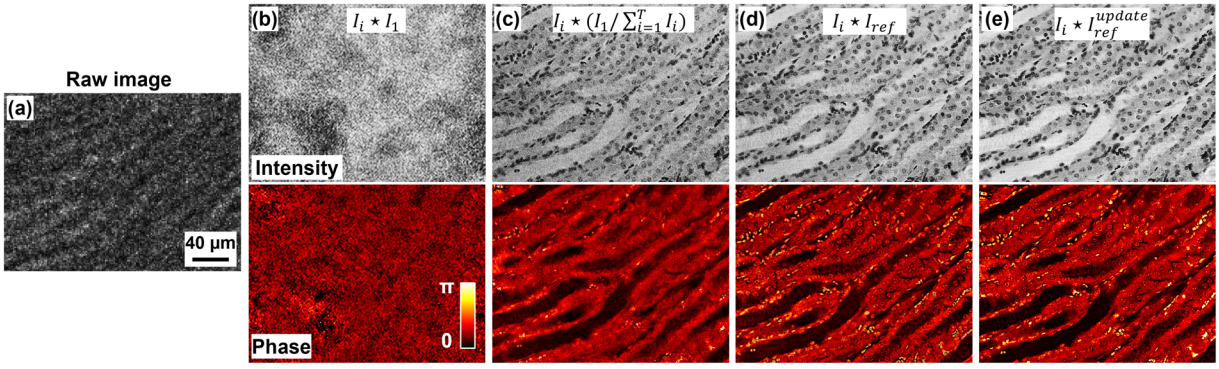

Fig. 4. (a) Raw image of a mouse kidney slide. (b)–(e) CP reconstructions corresponding to Figs. 3(b)–3(e). The first row shows the recovered intensity images and the second row shows the phase images. The remote referencing strategy with the refinement step give us the best reconstruction quality.

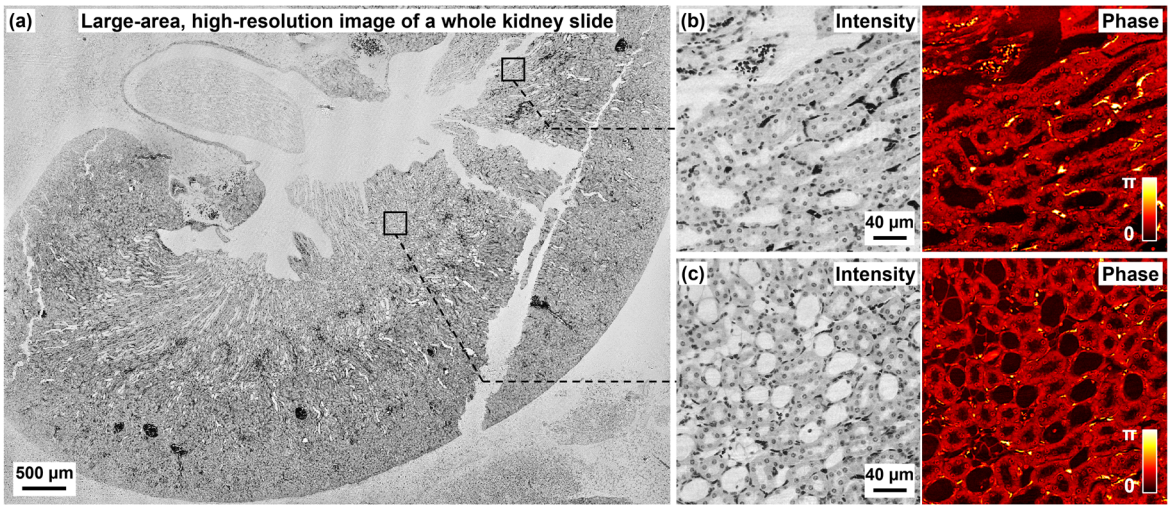

Fig. 5. (a) Large-area, high-resolution lensless CP imaging using the proposed remote referencing strategy with refinement. (b) and (c) Zoomed-in views of two regions. The acquisition time of the entire dataset is ~12 s, with an image acquisition speed comparable to that of a high-end whole slide scanner.

subsequent refinement pipeline, we do not need any positional feedback from the motion stage. The empty space on the coded surface is also no longer needed for positional tracking, enabling the use of all sensor pixels for diffraction data acquisition. The common iterative motion refinement strategy in ptychography is also not needed in the reported platform [12]. To acquire the large field-of-view images of bio-specimens, we can blindly translate the object to different positions and continuously acquire the corresponding diffraction data for CP reconstruction. Figure 5(a) shows the recovered whole slide image of the kidney slide acquired in 12 seconds, with an image acquisition throughput comparable to that of a high-end whole slide scanner. The zoomed-in views are shown in Figs. 5(b) and 5(c), where we can resolve the cellular structures of the specimen.

In summary, we have analyzed the motion tracking model of CP and reported a novel positional recovery pipeline for blind image acquisition. By using the reported approach, we can suppress the correlation peak caused by the coded surface and recover the positional shifts with deep sub-pixel accuracy. In contrast with common positional refinement methods, the reported approach can be disentangled from the iterative phase retrieval process and is computationally efficient. It allows blind image acquisition without motion feedback from the scanning process. It can also be applied in diffuser-based electron and light microscopy [13,14].

We note that for weak-phase samples, the raw image can be divided by the sum of all images before performing cross-relation analysis, i.e., replacing $I_i(x,y)$ with $I_i(x,y)/\sum_{i=1}^{T} I_i(x,y)$ in Figs. 4(d) and 4(e). Doing so effectively suppresses the impact of the coded surface. If the obtained cross-relation map is not uniform, we can also subtract the ith correlation map with the average of the adjacent maps to better remove the background. Lastly, there are many ways to implement the reported approach. For example, one may not need to move the object to a distant location to capture a reference image. Instead, we can divide the captured images into two sets. The first image can be used as the reference image for the second set. The last image can be used as the reference image for the first set.

Funding. National Science Foundation (2012140).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

REFERENCES

1. W. Hoppe, Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. **25**, 495 (1969).
2. H. M. L. Faulkner and J. Rodenburg, Phys. Rev. Lett. **93**, 023903 (2004).
3. J. R. Fienup, Appl. Opt. **21**, 2758 (1982).
4. F. Pfeiffer, Nat. Photonics **12**, 9 (2018).
5. S. Jiang, C. Guo, P. Song, N. Zhou, Z. Bian, J. Zhu, R. Wang, P. Dong, Z. Zhang, J. Liao, J. Yao, B. Feng, M. Murphy, and G. Zheng, ACS Photonics **8**, 3261 (2021).
6. S. Jiang, C. Guo, T. Wang, J. Liu, P. Song, T. Zhang, R. Wang, B. Feng, and G. Zheng, ACS Sens. **7**, 1058 (2022).
7. S. Jiang, C. Guo, Z. Bian, R. Wang, J. Zhu, P. Song, P. Hu, D. Hu, Z. Zhang, K. Hoshino, B. Feng, and G. Zheng, Biosens. Bioelectron. **196**, 113699 (2022).
8. P. Song, S. Jiang, T. Wang, C. Guo, R. Wang, T. Zhang, and G. Zheng, Photonics Res. **10**, 1624 (2022).
9. S. Jiang, C. Guo, P. Song, T. Wang, R. Wang, T. Zhang, Q. Wu, R. Pandey, and G. Zheng, Lab Chip **22**, 2657 (2022).
10. M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, Opt. Lett. **33**, 156 (2008).
11. P. Song, C. Guo, S. Jiang, T. Wang, P. Hu, D. Hu, Z. Zhang, B. Feng, and G. Zheng, Lab Chip **21**, 4549 (2021).
12. L. Bian, G. Zheng, K. Guo, J. Suo, C. Yang, F. Chen, and Q. Dai, Biomed. Opt. Express **7**, 4543 (2016).
13. S. McDermott and A. Maiden, Opt. Express **26**, 25471 (2018).
14. F. Allars, P.-H. Lu, M. Kruth, R. E. Dunin-Borkowski, J. M. Rodenburg, and A. M. Maiden, Ultramicroscopy **231**, 113257 (2021).