Robust Ptychographic X-ray Speckle Tracking with Multilayer Laue lenses

Nikolay Ivanov,[1,*] J. Lukas Dresselhaus,[2] Jerome Carnis,[1] Martin Domaracky,[1] Holger Fleckenstein,[1] Chufeng Li,[1] Tang Li,[2] Mauro Prasciolu,[1] Oleksandr Yefanov,[1] Wenhui Zhang,[1] Saša Bajt,[1,2] and Henry N. Chapman[1,2,3,†]

[1] Center for Free-Electron Laser Science CFEL, Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany.

[2] The Hamburg Centre for Ultrafast Imaging, Luruper Chaussee 149, 22761 Hamburg, Germany.

[3] Department of Physics, Universität Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany.

[*] nikolay.ivanov@desy.de

[†] henry.chapman@desy.de

Abstract: In recent years, X-ray speckle tracking techniques have emerged as viable tools for wavefront metrology and sample imaging applications, and have been actively developed for use at synchrotron light sources. Speckle techniques can recover an image free of aberrations and can be used to measure wavefronts with a high angular sensitivity. Since they are compatible with low-coherence sources they can be also used with laboratory X-ray sources. A new implementation of the ptychographic X-ray speckle tracking method, suitable for the metrology of highly divergent wavefields, such as those created by multilayer Laue lenses, is presented here. This new program incorporates machine learning techniques such as Huber and non-parametric regression and enables robust and quick wavefield measurements and data evaluation even for low brilliance X-ray beams, and the imaging of low-contrast samples. To realize this, a software suite was written in Python 3, with a C back-end capable of concurrent calculations for high performance. It is accessible as a Python module and is available as source code under Version 3 or later of the GNU General Public License.

1. Introduction

With the development of X-ray sources of ever increasing brightness and coherence there is an ongoing drive to design and develop optics able to focus X-ray beams down to about 1 nm [1]. Such optics would have a major impact in the field of X-ray microscopy. Also they would add great benefit for various modalities of X-ray imaging that investigate nano-structured materials and biological samples in-situ and in-operando, or which require extremely high X-ray intensities [2–6]. However, diffraction limited X-ray optics are very challenging to make and the highest achievable resolution is ultimately limited by aberrations of the optical elements. Thus, there is an increasing need for at-wavelength and in-situ wavefront metrology techniques of sufficient accuracy. The measurement of the wavefront is critical to understand the source of aberrations and to subsequently improve the fabrication of optical elements.

In this paper we focus on an approach, ptychographic X-ray speckle tracking (PXST) [7–9], that belongs to a family of phase imaging and lens monitoring techniques called X-ray speckle tracking (XST). Despite the name, PXST does not necessarily require speckle patterns, but any pattern with modulations in intensity will do, such as that formed by a beam transmitting through a Siemens star test structure [8]. Here, instead of a speckle pattern, we refer to the measured intensity pattern as an in-line or projection hologram. In any case, the contrast and smallest sizes of the features in the measured patterns are important properties that influence the quality of the reconstructed images and wavefronts obtained via speckle-based techniques [7, Appendix D].
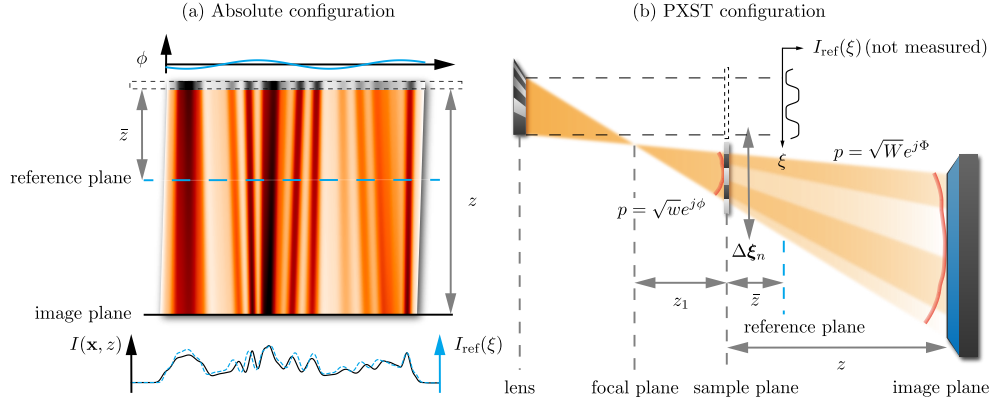
Fig. 1. Illustration of the X-ray speckle tracking principle for the the absolute (a) and ptychographic (b) configurations. (a): the wavefront with the phase profile ϕ produced by the optics is measured at two distances downstream from the mask. Bottom: the reference and the image are compared to determine the displacements. (b): the illuminating beam propagates from left to right. The sample is scanned across the beam. The red lines depict the illumination's wavefront in the sample and image planes. The virtual reference profile generated by the PXST algorithm is depicted at the reference plane.

The XST technique was introduced by Bérujon [10] and Morgan, Paganin, and Siu [11] as a way of obtaining a quantitative measurement of the phase gradient of a wavefront. This technique is based on the principles of geometric optics, where in the eikonal approximation light rays are directed normal to the wavefront's surface (Fig. 1). The principle of XST is simple: a local tilt of the wavefront (that is, a deviation of the direction of a local ray) can be observed as a shift of the "shadow" of a particular feature of an object placed in the beam. By tracking shifts of particular features as seen in different planes downstream or as caused by removing or shifting the refracting object of interest, the ray deflections and thus the phase gradient of the X-ray wavefront can be obtained. Typically, a sheet of sandpaper, consisting of small silicon carbide grains, or a biological filter membrane with micrometer-sized pores, is used as the analyzer to produce the intensity-modulated beam. To retrieve the wavefront created by a lens, the intensity is often measured with and without the lens in the analyzer beam [12]. These two detector images are usually labeled the "image" or "reference image", depending on whether the object (the lens in this case) is inserted in the beam or not. By comparing these, the displacements at positions $\mathbf{x}$ in the beam can be measured in the horizontal and vertical directions separately, and the deflection angle $\alpha(\mathbf{x}) = \{\alpha_x(\mathbf{x}), \alpha_y(\mathbf{x})\}$ in both directions is obtained at each position. The deflection angle is directly related to the phase gradient as

$$\frac{\partial \Phi(\mathbf{x})}{\partial x} = \frac{2\pi}{\lambda}\alpha_x(\mathbf{x}), \quad \frac{\partial \Phi(\mathbf{x})}{\partial y} = \frac{2\pi}{\lambda}\alpha_y(\mathbf{x}), \tag{1}$$

where λ is the X-ray wavelength. The phase $\Phi(\mathbf{x})$ can then be obtained by integration.

Since its introduction, a number of XST-based methods were developed for use with synchrotron and laboratory light sources, which are described and compared in a review paper by Zdora [13]. Some of these methods were devised to characterize focusing X-ray optics [7, 14]. The advantage of the XST method is that it offers a high angular sensitivity and low requirements on transverse and longitudinal coherence of the X-ray beam. However, high angular sensitivity measurements of the wavefront of high numerical aperture (NA) X-ray lenses, such as multilayer Laue lenses

(MLLs) [15], is very challenging. In the absolute configuration of XST, where the reference and the image planes are located a distance $z - \bar{z}$ apart (Fig. 1), the angular sensitivity depends on the ratio of the effective pixel size to the distance between the reference and the image planes: $\Delta\theta(\mathbf{x}) = \Delta x(\mathbf{x})/(z - \bar{z})$. Here, the effective pixel size is defined as the smallest resolvable displacement of a speckle or feature (including effects such as fringe visibility, finite pixel size, beam coherence and noise). Hence, the best accuracy is obtained by maximizing the distance between the reference and image planes. However, for highly divergent wavefields, there arises an inevitable trade-off between the wavefront sampling frequency and the angular sensitivity. Namely, the ideal location of the image plane is as far downstream of the lens focus as required to cover the detector array with the beam since this maximizes the wavefront sampling frequency. In order to minimize $\Delta\theta$ (maximizing the distance z) one should place the reference plane as close as possible to the beam focus. But in this plane the footprint of the beam on the detector may be much smaller than that in the image plane because of the beam divergence. This leads to a poorly sampled reference, as only a few pixels will span the wavefront's footprint.

Our method of PXST [7] overcomes this problem of sampling highly divergent beams. As schematically described in Fig. 1 (b), a sample is stepped across the beam wavefront while in-line projection holograms are collected with the detector at a fixed distance from the sample. The main advantage of this method, as compared to the absolute XST configuration, is that it alleviates the need to measure the reference image. Instead, the virtual reference image can be reconstructed from the collected holograms with nanometer precision, as long as there is sufficient overlap of the illuminated regions of the sample. In the simplest case of one step of the object (that is, two object positions) the displacements $\Delta x(\mathbf{x})$ are obtained from the differences in positions of features from one frame to the other, minus the average displacement of all features. By keeping the measurement and object in fixed planes, the approach is applicable to highly divergent wavefields produced by high NA X-ray optics such as multilayer Laue lenses (MLLs), which we primarily consider in this paper.

MLLs are diffractive, multilayer-based X-ray optics that can be understood as an extension of a conventional Fresnel zone plate in the form of a volume diffracting element [15, 16]. MLLs are cut from a deposited multilayer, which may consist of many thousands of layers of two alternating materials with different refractive properties. A bi-layer of the two materials forms a period. To focus X-rays to a common point the period thickness must decrease with the distance from the optical axis since smaller periods diffract X-rays by larger angles. The numerical aperture of the lens is thus set by the smallest period in the structure. A common way to fabricate the multilayer structure is by magnetron sputtering, in which a substrate is moved sequentially through the plasmas of two different materials. The time spent in each determines the individual layer thickness. Multilayers with periods below 1 nm are feasible using this method. Unlike with lithographically prepared Fresnel zone plates, there is no substantial limit to the thickness of the lens (in the direction of propagation) that can be prepared, allowing structures of high aspect ratio and high diffracting efficiency for hard X-rays. However, the angular acceptance of such structures for efficient diffraction (the rocking curve) is usually narrower than the NA, and so the Bragg condition cannot be simultaneously fulfilled throughout the entire lens aperture unless the layers vary in their tilt to the incident beam. This wedging of the layers is achieved with a straight-edge mask placed between the sputter source and the substrate [17], which creates a shadow of the extended source and thus a thickness gradient in the penumbra of this shadow. Knowing the thickness gradient one can precisely determine where to cut an MLL that focuses X-rays of particular energy at a desired focal length [18]. The deposited multilayer consists of tens of thousands of nanometer thick layers and deposition times can take many days [15]. Long deposition times require a stable and reproducible sputtering process. However, since the sputtering process removes material from the target, it gets thinner and the sputtering rate diminishes over time. An uncorrected linear decrease in rate, for example, manifests in coma, an

aberration that depends on the third power of the pupil coordinate [19].

PXST has been successfully used in our laboratory to characterize the wavefronts of individual MLLs. Nevertheless, as first implemented, it has some limitations. One of them becomes prominent if the sample is too close to the focus. In that case the reference profile is sampled sparsely, and PXST fails to successfully infer the reference image. However, this is also where the angular sensitivity of the PXST method is maximised and therefore one would like to work in this condition. In addition, when the noise of the intensity measurements is too high, PXST is not able to resolve the sample features and therefore fails to measure the feature displacements from one frame to another. In order to curtail these limitations we devised an improved implementation of the PXST analysis method. This new analysis pipeline, robust speckle tracking (R-PXST), is applicable over a wider range of experimental parameters and is highly robust against noise in the intensity measurements. We devised a software suite called `pyrost` for the whole R-PXST data processing pipeline in Python [20]. Also, using this new software, we explored the impact of experimental parameters on the angular sensitivity of R-PXST and compared it with PXST. In order to confirm the parametric analysis, we simulated speckle tracking scans using the Rayleigh-Sommerfeld diffraction theory.

2. Speckle Tracking algorithm

In this paper, we consider the general case of the projection imaging geometry shown in Fig. 1 (b). Here, the detector records a magnified near-field projection hologram $I(\mathbf{x}, z)$ of the sample, at a plane a distance z from the object (referred to as the image plane), as illuminated by an X-ray beam diverging from an effective point source a distance z_1 before the sample. The phase profile of the wavefront illuminating the sample is $\phi(\mathbf{x})$ and the amplitude is given by $w^{1/2}(\mathbf{x})$. Morgan *et al.* [7] devised a speckle tracking approximation which relates the projection hologram $I(\mathbf{x}, z)$ with a virtual reference hologram $I_{\text{ref}}(\mathbf{x}, \bar{z})$ formed under plane-wave illumination. The reference hologram is that which would be recorded on a detector with a demagnified pixel size and placed a distance $\bar{z}$ downstream of an object illuminated by a perfect plane. In particular,

$$I(\mathbf{x}, z) \simeq W(\mathbf{x}) I_{\text{ref}}\left[\mathbf{x} - \frac{\lambda z}{2\pi} \nabla \Phi(\mathbf{x}), \bar{z} \right], \tag{2}$$

$$\left(\frac{\bar{z}}{z} \right) w(\boldsymbol{\xi}) I_{\text{ref}}(\boldsymbol{\xi}, \bar{z}) \simeq I\left[\boldsymbol{\xi} + \frac{\lambda z}{2\pi} \nabla \phi(\boldsymbol{\xi}), z \right], \tag{3}$$

where $\bar{z} = z/M = zz_1/z+z_1$ is the effective propagation distance, $M = z+z_1/z_1$ is the geometric magnification factor, $\nabla \phi(\boldsymbol{\xi})$ and $\nabla \Phi(\mathbf{x})$ are the transverse gradients of the illuminating wavefield phase in the sample and image planes, and $w(\boldsymbol{\xi})$ and $W(\mathbf{x})$ are the amplitude profiles in the sample and image planes. The approximation was derived under the assumption of monochromatic light and in the near-field regime.

From their speckle tracking approximation, Morgan *et al.* [7] developed an iterative algorithm capable of reconstructing the reference image and the phase profile of the illumination from a series of projection holograms of the same sample, each recorded for a different displacement $\Delta \boldsymbol{\xi}_n$ of the sample in the transverse plane. According to Eqs. (2) and (3), the relationship between the measured images I_n and the virtual reference hologram I_{ref} can be expressed as

$$I_n(\mathbf{x}) \approx W(\mathbf{x}) I_{\text{ref}}(\mathbf{x} - \frac{\lambda z}{2\pi} \nabla \Phi(\mathbf{x}) - \Delta \boldsymbol{\xi}_n) = W(\mathbf{x}) I_{\text{ref}}(\mathbf{u}(\mathbf{x}) - \Delta \boldsymbol{\xi}_n), \tag{4}$$

where $\mathbf{u}(\mathbf{x})$ maps the measured intensity profile in the detector (image) plane $(\mathbf{x})$ to the reference (sample) plane $(\boldsymbol{\xi})$. This mapping is dependent upon the magnification of the reference image but it also includes a contribution caused by wavefront errors. The mapping is proportional to the phase gradient of the wavefront. The intensity profile $W(\mathbf{x})$ can be measured without the sample

in place, but in practice this "white-field" can be obtained simply by averaging all $I_n(\mathbf{x})$. For a uniform white-field profile W and within the limits of the approximation in Eq. (4), the reference image is simply a merging of each recorded image in the scan after correcting for the geometric distortions induced by the lens aberrations. The displacements applied to these overlaid images are given by the demagnified sample positions $\Delta\boldsymbol{\xi}_n/M$.
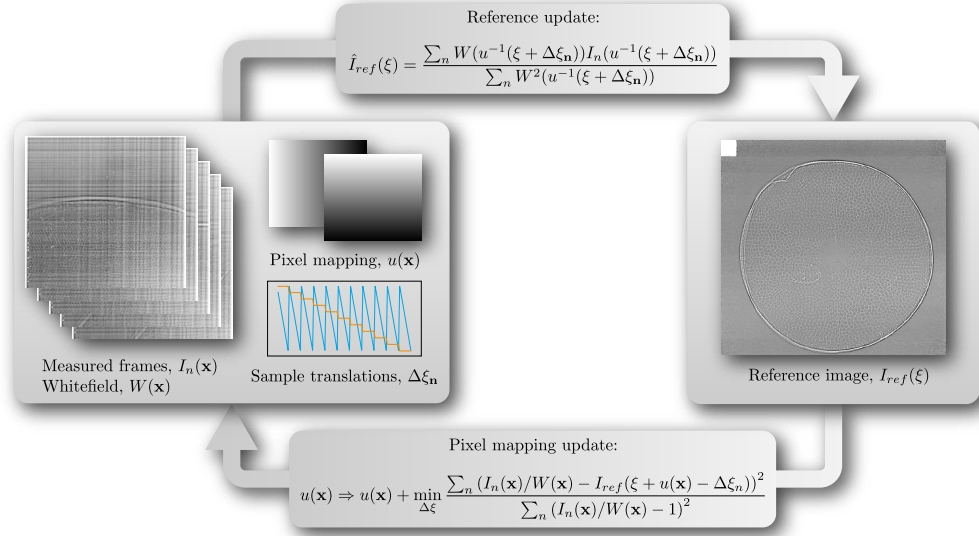

Fig. 2. Flow diagram for the PXST iterative update algorithm [9]. The algorithm starts with the measured data, sample translations, and initial aberration-free pixel mapping. During the update procedure at each iteration the reference image is reconstructed from the current estimation of the pixel mapping, whereupon the pixel mapping is updated so as to minimize the difference between the right-hand and left-hand sides of Eq. (4).

As described in [21, Appendix B], the Fresnel scaling theorem states that the hologram of a thin object formed using a beam diverging from a point source, recorded a distance z from the object, is equal to the hologram of a magnified version of that object formed under plane-wave illumination and recorded at a shorter propagation distance $\bar{z} = z/M$. Therefore, as an initial step, the mapping between detector and reference planes is approximated as linear:

$$I(\mathbf{x}, z) = M^{-2}I_{\text{ref}}(\mathbf{x}/M, z/M) \Rightarrow \mathbf{u}(\mathbf{x}) = \mathbf{x}/M \ . \tag{5}$$

Starting from this preliminary step, the mapping (u) and the reference image (I_{ref}) are iteratively updated following the flowchart illustrated in Fig. 2. The update procedures are accomplished by minimizing the target function given by [7, 9]:

$$\varepsilon(\mathbf{u}, I_{\text{ref}}, \Delta\boldsymbol{\xi}_n) = \int \frac{\sum_n \left(I_n(\mathbf{x}) - W(\mathbf{x})I_{\text{ref}}(\mathbf{u}(\mathbf{x}) - \Delta\boldsymbol{\xi}_n)\right)^2}{\sum_n \left(I_n(\mathbf{x}) - W(\mathbf{x})\right)^2} d\mathbf{x} \ . \tag{6}$$

The procedure performs accurately in the case of low noise and a high density of intensity measurements mapped to the reference plane, but it fails to yield plausible results when the noise is too high (such as for low exposures obtained using a less brilliant source) or when the intensity samples are sparsely spread out in the reference plane. Robust speckle tracking (R-PXST) aims to attain more accurate results from noisy and sparse data by employing robust machine learning and non-parametric techniques such as Nadaraya-Watson kernel regression [22, 23] and

local weighted linear regression [24, 25] for the estimation of the reference image, and Huber regression [26, 27] for the geometrical mapping update. All of these aspects were compiled in the software suite called `pyrost`.

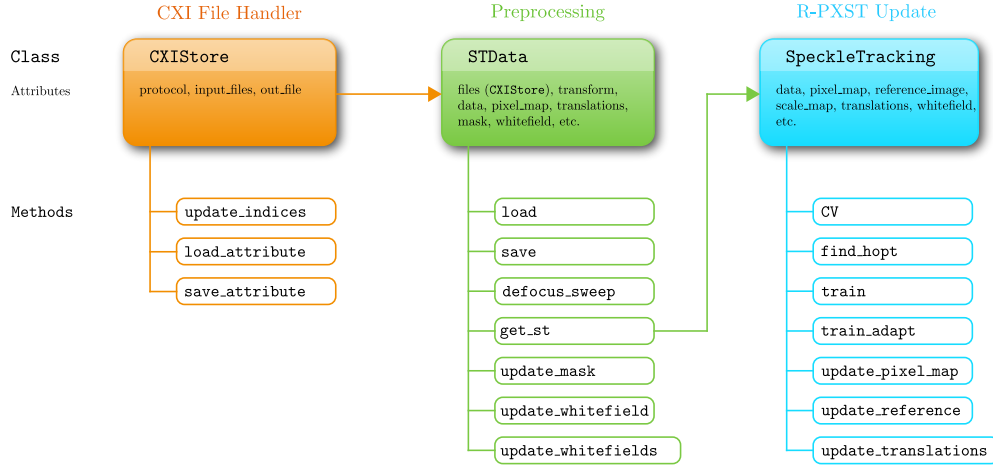## 3. Main functions and overview

Fig. 3. A diagram of the main classes and their attributes and methods in `pyrost` software [20].

The software was devised with an intent to be accessible to imaging and data scientists. Therefore, `pyrost` was designed with a user-friendly interface written in Python 3 and provides a high-level documentation for the project, including tutorials and an installation instruction [28]. The most intensive parts of calculations are executed in C with a support of concurrent calculations to yield the best performance. In order to maximize the utility and accessibility of the software suite, pyrost is available under Version 3 or later of the GNU General Public Licence. This allows other software projects to incorporate and modify all or part of this program into their own processing pipeline. The software is available through the anaconda package distribution system for the ease of installation or can be compiled locally.

The software has an object-oriented architecture and stores all of the necessary data in data containers with an interface that includes various utility functions. The organization of the main containers and their main methods is described in Fig. 3. The following classes are central to the suite:

1. `CXIProtocol` (not shown in Fig. 3) stores a list of paths for each of the data attributes necessary for the data processing pipeline together with their corresponding data types.

2. `CXIStore` takes a `CXIProtocol` protocol, list of input files, and an output file. It offers an interface to extract from the input files all the data attributes defined in the protocol (`read_indices`), load an attribute (`load_attribute`), and save a new attribute's data to the output file (`save_attribute`). The class is designed to work with HDF5 or CXI files.

3. `STData` is the main data container, which provides an interface to edit, load, and save data to and from the files in `CXIStore` file handler. Also it provides a set of tools for the preprocessing of a PXST dataset and can create a `SpeckleTracking` object that performs the main loop of R-PXST update.

4. `SpeckleTracking` performs the reference image and geometrical mapping updates needed for the iterative R-PXST update.

Additionally the suite contains a set of classes to perform common image transformations such as cropping and down-scaling, tools to simulate a PXST scan dataset based on scalar diffraction theory, and a class to integrate phase gradient maps and to perform polynomial fitting (see [28] for more information).

3.1. Initialization

A file protocol (`CXIProtocol`) must be created before loading the files. `pyrost` offers a default protocol, but it can be customized to accommodate different data files. The protocol, including file paths, is passed to the `CXIStore` object, which searches the input files for data defined by the protocol. The PXST dataset is then preprocessed by generating an `STData` object from `CXIStore` and loading the set of attributes defined by the user using the `load` method.

$\quad$`STData` contains tools to create maps and quantities needed prior to the main speckle tracking update procedure. In particular, `update_mask` generates a pixel mask that excludes bad and hot pixels of the detector from the subsequent analysis. `defocus_sweep` is used to estimate the defocus distance z_1 by generating reference images and computing their mean local variances, for a set of assumed focal distances. The mean local variance serves as a figure of merit of how sharp or blurry the reference image is. At the end of this process the defocus distance with the sharpest reference image is chosen. `update_whitefield` generates a white-field image by taking a pixel-wise median through a stack of measured frames.

$\quad$Measurements at very high magnification, made by positioning the sample close to the focus, often suffer from the background varying in time during the measurement. `update_whitefield` can alleviate this by generating a separate white-field for each frame, based on a principal component analysis (PCA) approach [29] or by applying pixel-wise median filtering over a stack of frames. Figure 4 displays the results of the PCA-based dynamic white-field correction applied to data collected from a biomineralized shell of a marine planktonic diatom measured at P11
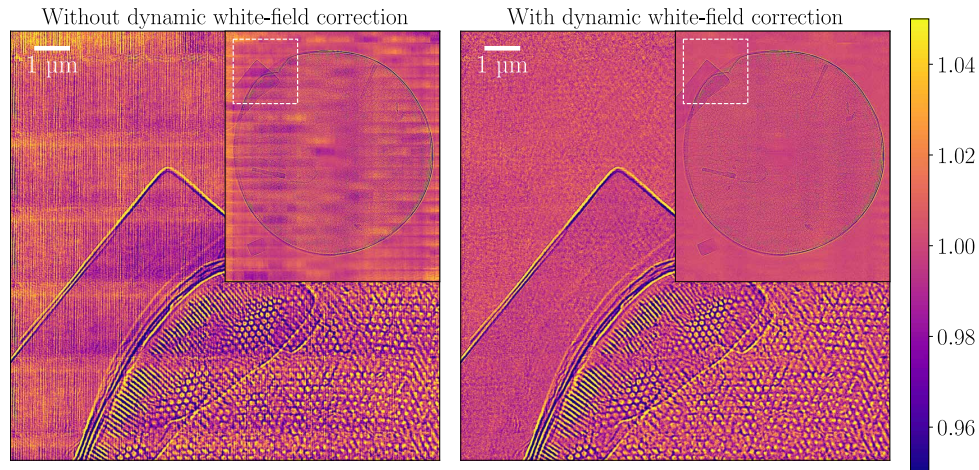

Fig. 4. The reference image obtained by stitching projection holograms obtained from a diatom, processed with (right panel) and without (left) dynamic white-field correction. The insert shows the entire field of view of the reference image and the region displayed at higher magnification is given by the dashed white lines. Data were collected at a photon energy of 17.5 keV and a magnification of 4837.

beamtime of the PETRA III synchrotron radiation facility as previously described [8]. The beam was focused with a pair of MLLs with focal lengths of 1.25 mm and 1.15 mm and a NA's of 0.014 and 0.015, in the vertical and horizontal directions, respectively. The X-ray beam photon energy was 17.5 keV and holograms were recorded at a magnification of 4837.

3.2. Speckle tracking update loop

Having formed an initial estimate for the defocus distance and the white-field (or a set of white-fields, if needed), a `SpeckleTracking` object with all data attributes necessary for the R-PXST update can be generated. `SpeckleTracking` provides an interface to iteratively refine the reference image and lens wavefront. It offers two methods (`train` and `train_adapt`) to perform iterations until the Huber criterion converges to a minimum (see Eq. (16) in Section 3.5). The difference between these two methods is that `train_adapt` performs an update of the optimal kernel bandwidth used in the estimation of the reference profile at each iteration, where `train` keeps the kernel bandwidth constant. The typical reconstruction cycle consists of: (i) the optional optimisation of the kernel bandwidth used for reference image estimation (`find_hopt`, in `train_adapt`); (ii) generation of the reference image (`update_reference`); (iii) an update of the discrete (pixel) mapping between a stack of frames and the generated reference image (`update_pixel_map`); (iv) an update of the sample translation vectors (`update_translations`); and (v) calculation of figures of merit (`update_error`). These five steps are described in the sections below.

In Fig. 5 we display the results after the first and 14$^{\text{th}}$ iteration of the outlined update procedure for the case of a Siemens star test sample. These data were obtained using the same setup as described above, and a magnification of 4837. The evolution of the error metric in this example is described below.

3.3. Reference image update

The reference image is recovered from a set of N measurements $I_n(\mathbf{x})$ with the help of the speckle tracking approximation as follows [7]:

$$\hat{I}_{\text{ref}}(\boldsymbol{\xi}) = \frac{\sum_n W(\mathbf{u}^{-1}(\boldsymbol{\xi} + \Delta\boldsymbol{\xi}_n)) I_n(\mathbf{u}^{-1}(\boldsymbol{\xi} + \Delta\boldsymbol{\xi}_n))}{\sum_n W^2(\mathbf{u}^{-1}(\boldsymbol{\xi} + \Delta\boldsymbol{\xi}_n))} \ . \tag{7}$$

The error in the reference image, caused by noise in the measured intensities, plays a key role in the robustness of the whole speckle tracking algorithm and directly affects the quality of the update of the geometric mapping $\mathbf{u}(\mathbf{x})$ as discussed below in Section 3.5. In Appendix A we derive the variance of the reference image estimator of Eq. (7) to examine how this depends upon and correlates with noise in the measured photon counts, which follows the Poisson distribution. We find that the variance in the reference profile reduces with the inverse of the redundancy factor $N_0(\boldsymbol{\xi})$. This factor is equal to the number of times that the given feature in the reference profile is present in the measured stack of frames. On average it is equal to the ratio of the width of a single image to the size of the step by which the sample is shifted from one frame to another.

In order to reduce the variance in the reference image estimator we employed non-parametric techniques [30, 31] in our implementation of the reference image update. Given random pairs of variables $\{x_i, y_i\}_{i=\{1..N\}}$, the non-parametric regression model can be formulated as [32]:

$$y_i = f(x_i) + \varepsilon_i, \quad f(x_i) = \text{E}[y_i], \quad \text{E}[\varepsilon_i] = 0, \tag{8}$$

where $f(x)$ is the regression function we try to estimate and ε is the error term. Here, we consider the Nadaraya-Watson kernel regression estimator [22, 23]:

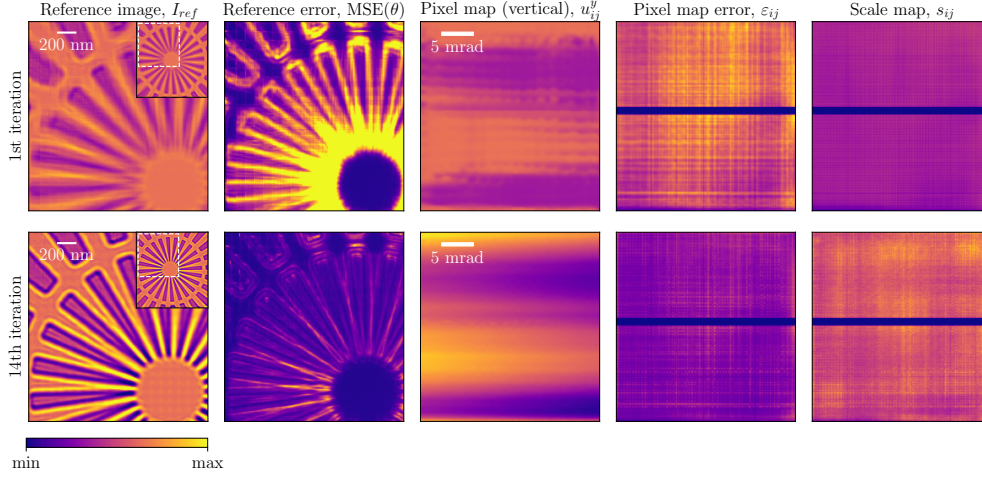$$\hat{f}(x) = \frac{\sum_{i=1}^N K((x - x_i)/h) y_i}{\sum_{i=1}^N K((x - x_i)/h)}, \tag{9}$$

Fig. 5. Maps of the reference image and maps pertaining to the wavefront, obtained after the first iteration (top row) and 22$^{\text{nd}}$ iteration (bottom row) of the R-PXST reconstruction using a Siemens star object. The first two columns show cropped regions of the reference image and reference image MSE. The full area of the scan is shown in the insert of the reference images of the first column. The third and fourth columns show maps of the vertical component of the displacements $\mathbf{u}_{ij}$ and the Huber errors for this displacement component. The final column gives maps of the scaling factor s_{ij}. The colour scale is indicated with (minimum, maximum) values of (0.6, 1.3), (0, 1.2), (-60, 60), (0, 1.2), and (0, 100) for columns 1–5, respectively. Data were collected at a photon energy of 17.5 keV and a magnification of 4837.

where $K(x)$ is the kernel function and h is the kernel bandwidth. The kernel regression estimator minimizes the mean-squared-error $\text{MSE}(\hat{f}(x))$ at the point x, which is given by:

$$\text{MSE}(\hat{f}(x)) = \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right)(y_i - \hat{f}(x))^2 = \sum_{i=1}^{N} K(\frac{x - x_i}{h})(f(x_i) + \varepsilon_i - \hat{f}(x))^2. \quad (10)$$

By applying the kernel regression approach to the speckle tracking approximation in Eq. (7), the reference image estimator can be given by:

$$\hat{I}_{\text{ref}}^{h}(\boldsymbol{\xi}) = \frac{\sum_n \sum_i W(\mathbf{x}_i) I_n(\mathbf{x}_i) K((\boldsymbol{\xi} - \boldsymbol{\xi}_{ni})/h)}{\sum_n \sum_i W^2(\mathbf{x}_i) K((\boldsymbol{\xi} - \boldsymbol{\xi}_{ni})/h)}, \quad (11)$$

where we used the Gaussian kernel $K(\boldsymbol{\xi}) = e^{-\xi^2/2}/\sqrt{2\pi}$ and where $\boldsymbol{\xi}_{ni} = u(\mathbf{x}_i) + \Delta\boldsymbol{\xi}_n$ are the intensity measurement coordinates mapped to the reference plane. The variance of this estimator is inversely proportional to the kernel bandwidth (see Appendix A) which can therefore be reduced by increasing the bandwidth at the cost of band-limiting the reconstructed reference image.

Although the kernel-based regression delivers a robust and smooth estimate of the reference profile in the regions where the density of measurement coordinates $\boldsymbol{\xi}_{ni}$ is high as compared to the kernel bandwidth, it suffers from higher bias at the boundaries of the domain of the inputs $\boldsymbol{\xi}_{ni}$. Therefore, we also implemented a local linear regression estimator, which provides a better local linear approximation by minimizing the following criterion as a function of the vector

$\{\hat{\alpha}(x), \hat{\beta}(x)\}$ [24, 25]:

$$\text{MSE}(\hat{\alpha}(x), \hat{\beta}(x)) = \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right)(y_i - \hat{\alpha}(x) - (x - x_i)\hat{\beta}(x))^2. \tag{12}$$

Naturally, each image is discretely sampled by the pixel array of the detector. For pixel i, j of image n, the discrete representation of the image is given by $I_{nij} = I_n(i\,\delta x, j\,\delta y)$, where δx and δy are the extents of a pixel along the x and y coordinates, respectively. Since the reference profile is measured virtually, the sampling interval in the reference plane can be of arbitrary value and highly impacts the computational cost of the speckle tracking update procedure. In our software the sampling intervals f_x, f_y are measured in the units of the demagnified pixel sizes δu and δv, so that $I_{\text{ref}}(f_x i, f_y j) = I_{\text{ref}}(f_x i\,\delta u\mathbf{i} + f_y j\,\delta v\mathbf{j})$, where $\delta u = \delta x/M$, $\delta v = \delta y/M$, and $\mathbf{i}, \mathbf{j}$ are the unit vectors along the horizontal and vertical detector axes, respectively. Likewise, we convert the sample translations to pixel units, so that $\Delta i_n = \Delta \xi_n/\delta u$ and $\Delta j_n = \Delta \chi_n/\delta v$, where $\Delta \boldsymbol{\xi}_n = (\Delta \xi_n, \Delta \chi_n)$. The function $\mathbf{u}(\mathbf{x})$ maps intensities from the detector to the reference plane. Since both the detector and reference intensity profiles are discretized, the mapping is also discrete. We call the discrete representation of the mapping function the "pixel mapping". This mapping is scaled by the demagnified pixel size and is given by $u_{ij}^x = 1/\delta u\ u_x(i\,\delta x, j\,\delta y)$, $u_{ij}^y = 1/\delta v\ u_y(i\,\delta x, j\,\delta y)$, where $\mathbf{u}(\mathbf{x}) = (u_x(\mathbf{x}), u_y(\mathbf{x}))$. Finally, the discrete reference image may be reconstructed from the measured data I_{nij} and pixel mapping (u_{ij}^x, u_{ij}^y) as follows:

$$I_{\text{ref}}(f_x i, f_y j) = \frac{\sum_n \sum_{i'} \sum_{j'} K(f_x i - u_{i'j'}^x + \Delta i_n, f_y j - u_{i'j'}^y + \Delta j_n, h)\ W_{i'j'} I_{ni'j'}}{\sum_n \sum_{i'} \sum_{j'} K(f_x i - u_{i'j'}^x + \Delta i_n, f_y j - u_{i'j'}^y + \Delta j_n, h)\ W_{i'j'}^2}, \tag{13}$$

where the kernel is converted to pixel units $K(i, j, h) = K(\frac{i\,\delta u + j\,\delta v}{h})$.

In the top left panel of Fig. 5, we show the reference image corresponding to the initial estimate for the pixel map with $u_{ij}^x = i$ and $u_{ij}^y = j$, following Eq. (5). Looking closely, one can see that distortions of the individual holograms of the Siemens star, caused by lens aberrations, give rise to the placement errors when stitching together the reference image. These errors are reflected in the map of the MSE, in the second column of Fig. 5. The MSE (Eq. (10)) reaches maximum near the inner ends of the spokes of the sample where the placement errors are most prominent. However, after the 14th iteration, the placement errors are largely reduced and the Fresnel fringes near the edges of the Siemens star structure become clearly visible.

3.4. Optimal kernel bandwidth

The use of a small kernel bandwidth in a non-parametric estimator can introduce a small bias to the estimate. At the same time, less smoothing means that each estimate is obtained by averaging over (in effect) just a few observations, making the estimate noisier. So less smoothing increases the variance of the estimate. Therefore, we want to select an optimal bandwidth that minimizes the MSE metric in Eq. (10). By performing the second-order Taylor expansion of the regression function $f(x)$ in Eq. (10), the MSE as the function of the kernel bandwidth h can be found as

$$\text{MSE}(h) = \frac{R(K)f(x)}{Nh} + \frac{1}{4}h^4\mu_2(K)^2 f''(x)^2 + o\{(Nh)^{-1} + h^4\}. \tag{14}$$

where $R(K) = \int K(x)^2 dx$ and $\mu_2(K) = \int x^2 K(x)dx$ [33]. The error profile is concave and has a single optimal bandwidth h^{opt}, that depends on the unknown regression function $f(x)$ and therefore can not be estimated directly. One way to estimate the optimal bandwidth based on the data is to use the cross-validation (CV) metric [33]. The CV is given by:

$$\text{CV}(h) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{f}_{-i}(x_i))^2, \tag{15}$$

where $\hat{f}_{-i}(x)$ is the "leave-one-out" estimator based on the dataset with x_i deleted, similar to the concept of R_{free} in crystallography [34]. `SpeckleTracking` class divides the dataset into two subsets at the initialization stage. The `CV` method calculates the CV by generating a reference image based on the former "training" subset and calculating the error metric introduced in Section 3.5 (Eq. (16)) for the latter "testing" subset. `find_hopt` estimates the optimal bandwidth by finding a minimum of $\text{CV}(h)$ with the quasi-Newton method of Broyden, Fletcher, Goldfarb, and Shanno (BFGS) [35].
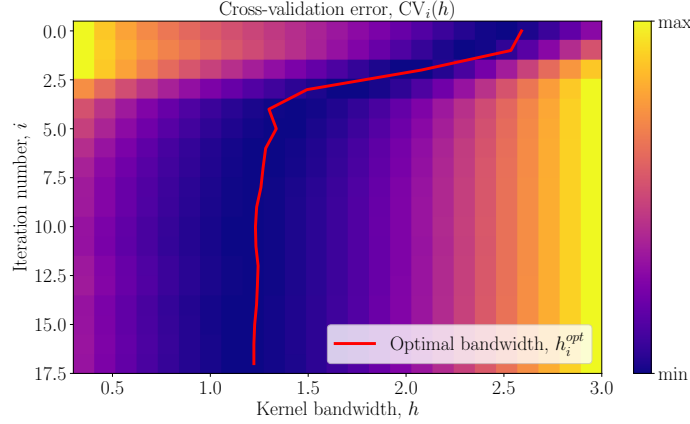

Fig. 6. Evolution of the cross-validation error metric $\text{CV}(h)$ (Eq. (15)) during the R-PXST iterations using the adaptive kernel bandwidth update (`train_adapt`). By updating the optimal kernel bandwidth h_i^{opt} at each iteration, `train_adapt` attains a lower average error at the end. The dataset used here is the same shown in Fig. 5.

It should be noted that the cross validation error CV changes at each iteration of the R-PXST reconstruction, and thus so too does the optimal kernel bandwidth. To account for this we implemented the ability to update the kernel bandwidth at each iteration of R-PXST, using the BFGS method in `train_adapt`. In Fig. 6 we show $\text{CV}_i(h)$ for consecutive iterations of the R-PXST update loop together with the estimate of the optimal kernel bandwidth at each iteration, for the same dataset as in Fig. 5.

3.5. Pixel mapping update

The pixel mapping update is carried out separately at each pixel i, j in the detector grid by minimizing the error metric in Eq. (6) as a function of $\mathbf{u}(\mathbf{x})$. In order to make the prediction robust against the outliers in the intensity measurements we use the Huber regression technique [26, 27]. Following the convention of Eq. (13), the discrete representation of the Huber criterion based on the target function in Eq. (6) is given by:

$$
\begin{aligned}
\varepsilon_{ij}(\delta i, \delta j, s) = \frac{1}{N} \\
\times \sum_{n=1}^{N} \left[s + \mathcal{H}_{1.35} \left(\frac{I_{nij} - W_{ij} I_{\text{ref}}(f_x i - u_{ij}^x - \delta i + \Delta i_n, f_y i - u_{ij}^y - \delta j + \Delta j_n)}{s} \right) s \right],
\end{aligned} \quad (16)
$$

where $\mathcal{H}_M(x)$ is the Huber's function [36]. The parameter M of the Huber function is held fixed and equal to 1.35, following the recommendation of Owen [26], and s is the scaling parameter. The `update_errors` method in the `SpeckleTracking` class calculates ε_{ij} as well as the same error normalised by that obtained for a uniform reference image and the linear pixel

mapping $\mathbf{u}(\mathbf{x}) = \mathbf{x}/M$:

$$\tilde{\varepsilon}_{ij}(\delta i, \delta j, s) = \frac{\varepsilon_{ij}(\delta i, \delta j, s)}{\frac{1}{N} \sum_{n=1}^{N} \left[s + \mathcal{H}_{1.35} \left(\frac{I_{nij} - W_{ij}}{s} \right) s \right]}. \tag{17}$$

The scaling parameter essentially defines a threshold for whether observations are treated as outliers or not. If the scaled argument of $\mathcal{H}_{1.35}$ in Eq. (16) is less than 1.35 then the loss function has a quadratic nature that is similar to the terms in a least-squares minimisation (L2 regression). However, when the scaled argument exceeds 1.35, the loss function becomes linear, which then reduces the contribution of that particular term in the sum, compared with a least-squares approach. That is, when the difference between an observation I_{nij} and its corresponding prediction $W_{ij}I_{\text{ref}}$ is large (on a scale determined by s) the observation is considered an outlier. The scaling itself is a parameter that is optimised in the minimisation, so the threshold between inliers and outliers can evolve through the iterative process. The criterion in Eq. (16) is jointly convex in δi, δj, and s. Thus, starting from the preliminary approximation of the pixel mapping $u_{ij}^x = i$, $u_{ij}^y = j$ and the scale mapping $s_{ij} = \sqrt{W_{ij}}$, we obtain a new estimation $u_{ij}^{x\prime} = u_{ij}^x + \delta\hat{i}$, $u_{ij}^{y\prime} = u_{ij}^y + \delta\hat{j}$, $s_{ij}' = \hat{s}$ by minimizing $\varepsilon_{ij}(\delta i, \delta j, s)$ in Eq. (16). The minimization can be performed by one of three algorithm options implemented in the `update_pixel_map` method: grid search, random search or differential evolution [37–39]. These methods were chosen over the gradient-based minimization algorithms since they minimize the error metric globally instead of descending into an adjacent local minimum in the error space. From our experience and observations the error landscapes of most of the datasets we tested suffer from noise and contain many local minima that surround the global minimum.

The update of the pixel mapping usually requires a further regularisation, since these estimates are affected by the noise of the measured photon counts I_{nij} and are prone to artefacts that may arise from poor initial estimates of the reference image. To address this, we employed a Gaussian-weighted smoothing as follows:

$$\tilde{u}_{ij}^x = \frac{K(i, j, \sigma) * (w_{ij} u_{ij}^{x\prime})}{K(i, j, \sigma) * w_{ij}}, \quad \tilde{u}_{ij}^y = \frac{K(i, j, \sigma) * (w_{ij} u_{ij}^{y\prime})}{K(i, j, \sigma) * w_{ij}}, \tag{18}$$

where σ is the Gaussian kernel bandwidth, $*$ denotes the convolution operation, and $w_{ij} = \max\{\varepsilon_{ij}(0, 0, s_{ij}) - \varepsilon_{ij}(\delta\hat{i}, \delta\hat{j}, \hat{s}), 0\}$ is the error difference before and after the update.

Since the geometric mapping between the reference plane and the detector plane is defined in terms of the gradient of the scalar function Φ, as $\mathbf{u}(\mathbf{x}) = \mathbf{x} - \frac{\lambda z}{2\pi} \nabla \Phi(\mathbf{x})$, the curl of the mapping must be zero: $\nabla \times \mathbf{u}(\mathbf{x}) = 0$ [9,40]. Such a vector field is called irrotational. However, the pixel mapping is updated at each point separately without any examination of the irrotationality. To constrain the mapping field to be irrotational one can first integrate $\mathbf{u}(\mathbf{x})$ and then numerically calculate the gradient to obtain a curl-free version. We implemented this procedure into the `update_pixel_map` method with the cosine transform integration [41] chosen for integrating the pixel mapping.

In the third and the fifth columns of Fig. 5 we show the vertical component of the geometric mapping (u_{ij}^y) and scale mapping (s_{ij}) respectively, after the first and 14th iteration. The pixel map was updated by the random search algorithm with regularization ($\sigma = 8.0$) and the irrotationality constraint. In our example of Fig. 5 the X-ray beam was focused by a pair of MLLs oriented orthogonally to each other and along the x and y axes of the detector. As discussed in the introduction, the aberrations in the wavefront of the lens are primarily due to systematic layer thickness error, and therefore are expected to vary along the lens orientation. Hence, u_{ij}^y is expected to vary along the y axis of the detector and u_{ij}^x along the x axis. This matches what we see in the pixel map shown in Fig. 5.

The scaling parameter estimator based on finding a minimum of Eq. (16) gives an estimate of the square root of MSE [27]:

$$\hat{s}^2 = \frac{1}{N} \sum_{n=1}^{N} \left(I_{nij} - W_{ij} I_{\text{ref}}(f_x i - u_{ij}^x - \delta i + \Delta i_n, f_y i - u_{ij}^y - \delta j + \Delta j_n) \right)^2 . \tag{19}$$

Therefore, if the only source of variance in measured intensities comes from the Poisson noise, we expect the scaling parameter after the reconstruction to be on average smaller than before. However, as shown in fifth column in Fig. 5 the scaling map after the update is larger. This is caused by the intensity variations within in the white field of the lens pupil (as can be seen in the measured frames of Fig. 2, for example) that modulate over the course of the scan as noted in Sec. 3.1.

3.6. Updates of translation vectors

The procedure for updating the sample translation vectors follows the same logic as that for the pixel mapping update. The error function of Eq. (16) is taken as a function of $\Delta i_n, \Delta j_n$ and summed over the detector grid, not over the stack of frames. The sample translation is then updated for each frame separately using the grid search algorithm.

4. Results

4.1. Dependence on defocus distance and number of frames

As discussed in Appendix D of reference [7], the sample position that maximizes the angular sensitivity $\Delta\phi$ of PXST is at the focal plane, where the magnification is the greatest:

$$\Delta\phi = \frac{2\pi}{\lambda} \Delta\theta \, \delta x = \frac{2\pi}{\lambda} \frac{\delta x^2}{M}, \tag{20}$$

for a detector pixel size δx and magnification $M \approx z_1/z$. Therefore, for wavefront measurements it is beneficial to place the test object very close to the focus of the lens. On the other hand, a
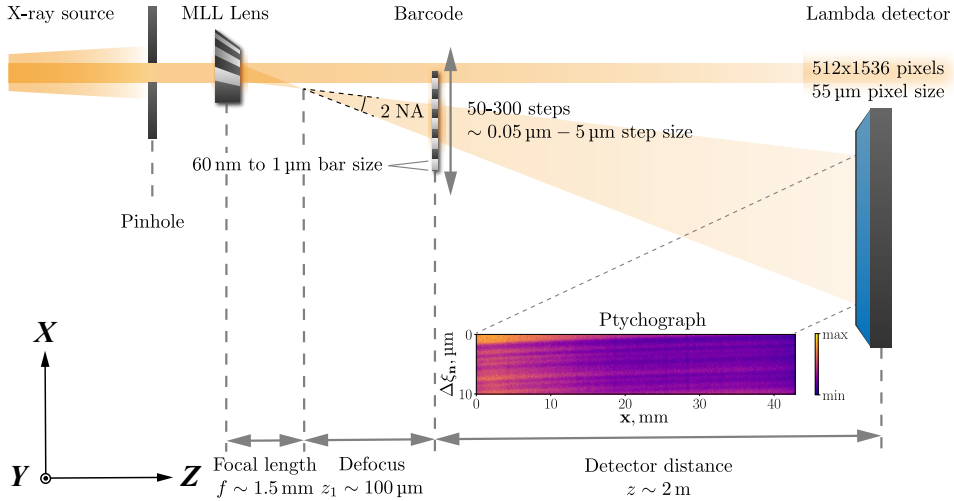

Fig. 7. Experimental geometry of a PXST wavefront metrology experiment used in our laboratory setup and in our simulations. The scheme is provided with one example of the defocus distance, focal length, and sample-to-detector distance.

Source, MLL and propagation distance					
Focal length, f	NA	Total number of photons per PXST scan	Photon energy	Angular width of source	Detector distance, z
1.5 mm	0.013	5×10^6	17 keV	2.5 mrad	2 m
Barcode sample and detector parameters					
Bar size	Transmittance of the first material		Transmittance of the second material	Detector grid size	Pixel size
0.5 µm to 1.5 µm	0.85		0.65	2000×1000	55 µm

Table 1. Parameters used in simulations shown in Figs. 8 and 9.

very high magnification raises the problem of under-sampling the reference profile, since the sampling interval of the intensity measurements mapped to the reference plane is proportional to the magnification. Furthermore, for measurements using laboratory X-ray sources, we wish to understand how the PXST analysis depends on the number of frames collected when the total measurement time (and hence total signal) is limited. To test these dependencies, and to determine the robustness of the R-PXST and PXST algorithms, analyses were carried out on simulated data using the `pyrost` and `speckle_tracking` software.

For the simulations we considered a geometry that is typical for one-dimensional PXST wavefront metrology measurements routinely carried out in our laboratory (Fig. 7), which is the measurement of a cylindrical wavefront formed by focusing a collimated beam to a line focus by a single MLL [42]. Some of the parameters of the geometry are listed in Table 1. To match the line focus, we assumed a test object with a structure that varies only in one direction, which we call a barcode object, consisting of layers of two materials with randomly-varying thicknesses ranging between 0.5 µm and 1.5 µm as detailed in Table 1. The wavefront of the focused beam was generated using a complex-valued lens pupil function $\Pi(x)$ (for a lens coordinate x) similar to that of an actual MLL. The amplitude of the pupil transmission was taken from a measured lens transmission which decreased slightly with increasing diffraction angle and includes random modulations across the pupil of about ±35 %. The phase of the pupil transmission was assumed to vary with x^3, representing a coma aberration often found in high-NA MLLs [8]. The object was taken to be in the beam behind the focus and scanned in the transverse direction. Sample translations $\Delta\xi_n$ were simulated as described in Section 6.2. These holograms are magnified by the diverging wavefield only in one direction. From each scan of simulated one-dimensional holograms, we generated a two-dimensional "ptychograph" in which each row represents the one-dimensional divergent beam at the detector for each of the positions $\Delta\xi_n$ of the object. One such ptychograph is shown in Fig. 7, simulated for $z_1 = 100$ µm giving a magnification $z_1/z = 20\,000$. At this defocus the beam size is 2.6 µm and so only one or two bars of the test object are illuminated. Each row of the ptychograph therefore only covers a few bars of the object. However the simulated object was scanned in 1 µm steps, so the shadow of the object moves by a large fraction of the illuminated region on each step, bringing new bars rapidly into view with increasing $\Delta\xi_n$. Features of the object thereby trace out lines in the ptychograph with a gradient dependent on the defocus. If the lens was free of wavefront aberration, the lines would

be straight. In this case, the lines are slightly curved, indicating aberrations (other than defocus). Distinguishing features and tracking them through the ptychogram is the aim of PXST.

To explore the domain of applicability of the algorithms with regard to the defocus distance, ptychograms were simulated for different defocus distances ranging from 20 μm to 5 mm (resulting in magnifications of 100 000 to 400). To examine whether the performance of the algorithms for different defoci depends on the magnitude of the wavefront aberration, the defocus series of ptychograms was simulated for different magnitudes of the coma wavefront aberration, given by $\phi(x) = \alpha_0(x/f)^3$ where f is the lens focal length. Three different magnitudes were chosen: $\alpha_0 = 0.01\,\mathrm{rad/mrad}^3$, $0.03\,\mathrm{rad/mrad}^3$, and $0.05\,\mathrm{rad/mrad}^3$. For NA = 0.013, a magnitude of $\alpha_0 = 0.01\,\mathrm{rad/mrad}^3$ gives a peak-to-valley wavefront error of 16.9 rad or 2.7 waves. Each ptychogram was simulated at 200 sample positions, and the step size was random, chosen from a range spanning 0.05 μm to 0.45 μm. The total number of detected photons over each entire PXST scan was set to 5×10^6. Poisson noise was added to the simulated holograms.

One simulated ptychograph is shown in Fig. 8 (a) for a very short defocus of $z_1 = 25\,\mu\mathrm{m}$ and $\alpha_0 = 0.03\,\mathrm{rad\,mrad}^{-3}$. This ptychogram was then processed by by the `pyrost` (R-PXST) and `speckle_tracking` (PXST) packages to reconstruct the one-dimensional image of the test object in Fig. 8 (b) and the pixel displacements of features, proportional to the phase gradient of the lens wavefront, in Fig. 8 (c). R-PXST analysis was performed with the optimal kernel bandwidth estimated by finding a minimum of the cross-validation error ($h^{\mathrm{opt}} = 35.89$ in Fig. 8 (b)). The pixel map was updated using the random search algorithm with regularization ($\sigma = 15$). The reconstruction results can be compared with the ground truth shown as red dashed lines in Fig. 8 (b) and (c). It is seen that R-PXST recovers the image and phase gradient with good accuracy in this case, whereas the original PXST routine produces a noisy image and fails to recover the true phase gradient. This is due to the extreme sparsity of intensity measurements mapped to the plane of the reference image.

The results of the PXST and R-PXST reconstructions for the three defocus series for the different aberration magnitudes are summarised in Fig. 8 (d), where the estimated third-order aberration coefficient α is plotted as a function of the defocus z_1. The third-order coefficient was obtained by fitting a parabola to the recovered phase gradient. These are compared with the ground truth (α_0) as horizontal dashed red lines. The example of Fig. 8 (a)–(c) for $z_1 = 25\,\mu\mathrm{m}$ and $\alpha_0 = 0.03\,\mathrm{rad\,mrad}^{-3}$ is shown as the large blue and gold circles in Fig. 8 (d) for R-PXST and PXST, respectively. It is seen that PXST underestimates the wavefront magnitude for defocus distances less than about 200 μm. At these magnifications the pixel locations of the intensity measurements mapped to the reference plane become so sparse that the bilinear subpixel interpolation employed in `speckle_tracking` [9, Appendix A] does not yield an accurate reference image (cf. Fig. 8 (b)). R-PXST achieves good results even at a defocus of 20 μm because it employs the kernel regression to reconstruct the reference profile and adaptively chooses the optimal kernel bandwidth that minimizes the MSE introduced by the estimator. Both algorithms exhibit a degradation at high defocus. A closer look at the displacement profiles show that at higher defoci Poisson noise becomes more prominent since the RMS value of the displacement profile becomes comparable with the smallest resolvable angular displacement.

A common reason for failure of reconstruction of a ptychographic speckle tracking scan is due to measurement noise as a result of low photon counts. This is particularly critical for measurements made with laboratory X-ray sources, where a PXST dataset with a high number of frames (to give high redundancy) together with long exposure times (to reduce noise) may be impractical. We performed a series of simulations of ptychographs made with different numbers of frames and a constant total detected flux to determine how best to fractionate an allotted exposure to achieve best results. The parameters used in the simulations were the same as for the defocus study, as listed in Table 1. Here, the defocus was chosen as 100 μm and a randomized step size varying in a range from 0.05 μm to 0.45 μm was assumed.
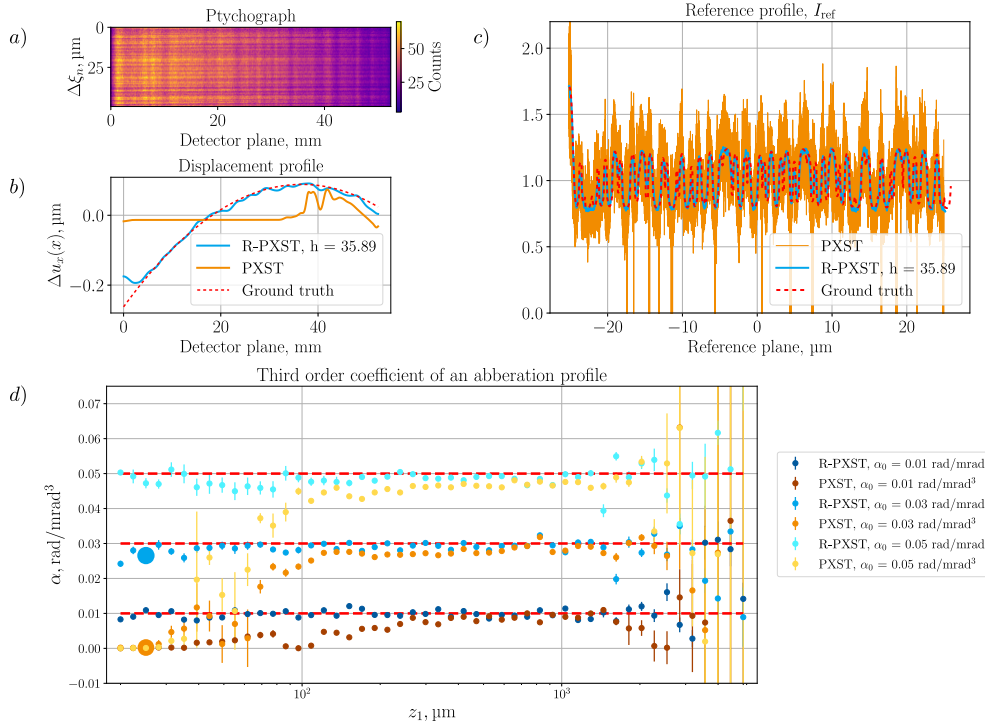
Fig. 8. (a) Simulated ptychograph of the bar object imaged with a one-dimensional MLL for a defocus $z_1 = 25\,\mu m$ and aberration magnitude $\alpha_0 = 0.03\,rad/mrad^3$. Recovered image of the bar object (b) and displacement profile (proportional to the wavefront phase gradient) (c), obtained by R-PXST and PXST. (d) Plots of the estimated aberration magnitude α as a function of defocus, obtained by R-PXST and PXST from simulations with $\alpha_0 = 0.01\,rad/mrad^3$, $0.03\,rad/mrad^3$, and $0.05\,rad/mrad^3$. The ground truth in (b), (c), and (d) is shown with red dashed lines.

A simulated ptychograph is shown in Fig. 9 (a) for a scan that distributes the signal into 500 frames of equal exposure (of 10^4 detected photons). A third-order aberration magnitude of $\alpha_0 = 0.05\,rad\,mrad^{-3}$ was used. The mean number of counts per pixel in the ptychograph is about 10. This appears to be too low for the `speckle_tracking` program since it produces a noisy reference image, shown in Fig. 9 (b), and fails to recover the wavefront phase gradient as shown in Fig. 9 (c). The `pyrost` program obtains reasonable reconstructions, close to the ground truth, shown as red dashed lines. Plots of the recovered wavefront third-order magnitude α are given in Fig. 9 (d) as a function of the number of frames from 2 to 500, and for three different series with $\alpha_0 = 0.01\,rad/mrad^3$, $0.03\,rad/mrad^3$ and $0.05\,rad/mrad^3$. The reconstruction of the simulated data fails for scans with less than 20 frames. This failure is due to a lack of redundancy to reconstruct the undistorted reference image. In contrary, for a large number of frames, the low counts per pixel results in a low signal to noise ratio. It becomes crucial to suppress the noise in the regression of the reference image in order to resolve the fine structure of the sample. As seen in Fig. 9 (b), simulated at the extreme of 500 frames, and the plots of Fig. 9 (d), the PXST algorithm fails to yield accurate results for more than about 100 frames (signals lower than 7 photons per pixel), even when the R-PXST algorithm accurately recovers the image.
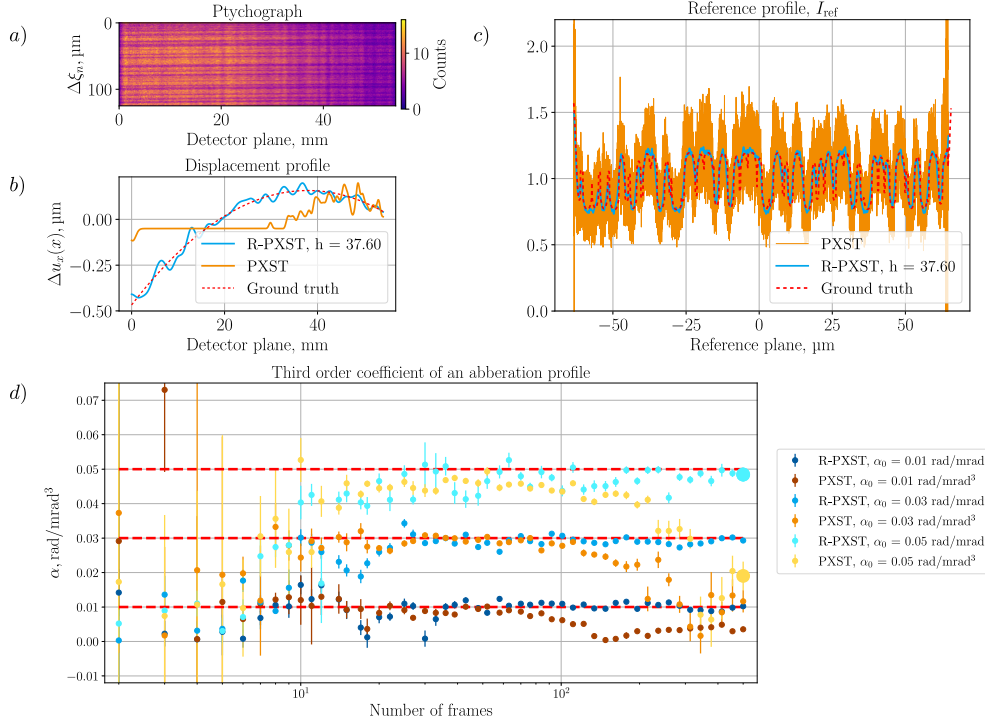
Fig. 9. (a) Simulated ptychograph of a bar object imaged with a one-dimensional MLL with 500 frames, a defocus of $z_1 = 100\,\mu m$, 5×10^6 total photons detected, and an aberration magnitude $\alpha_0 = 0.05\,rad/mrad^3$. Recovered image of the bar object (b) and displacement profile (c), obtained by R-PXST and PXST. (d) Plots of the estimated aberration magnitude α as a function of the number of frames in the scan, keeping the total counts constant. The ground truth in (b), (c), and (d) is shown with red dashed lines.

4.2. Robustness of the the R-PXST algorithm to noise

Here we explore the robustness of R-PXST update procedure in the presence of noise in measured holograms. For these tests we used a dataset measured at the P11 beamline at the PETRA III synchrotron radiation facility with a pair of MLL lenses of NA equal to 0.015, focusing in two dimensions at a photon energy of 17.4 keV, giving a similar geometry to the simulations carried out above. The object was a 2.5 μm-thick piece of nanoporous gold [43]. This object consists of a sponge-like reticulate structure of gold with branches that are of the order of 100 nm in diameter infused with pores that are tens of nanometers in diameter. The focus-to-sample distance was 65 μm where the defocused beam width was 1.9 μm. The detector, with 75 μm-wide pixels was located 2.37 m downstream of the sample, giving a magnification of 36 500. The scan was performed with a step size of 50 nm in a 21×21 grid of sample positions. The mean photon counts per pixel per step was 160. To examine the case of low signal to noise ratio, where the relative noise is comparable to the contrast of the holograms, we also created an artificially "noisy" version of the dataset by dividing the photon counts by 10 and applying random noise generated from the appropriate Poisson distribution.

The reconstructed reference images (top row) and the retrieved phase profile (bottom row) of the measured dataset are shown in Fig. 10. The data without added noise was analysed by the R-PXST procedure as shown in the first column of Fig. 10. The contrast of the recovered reference
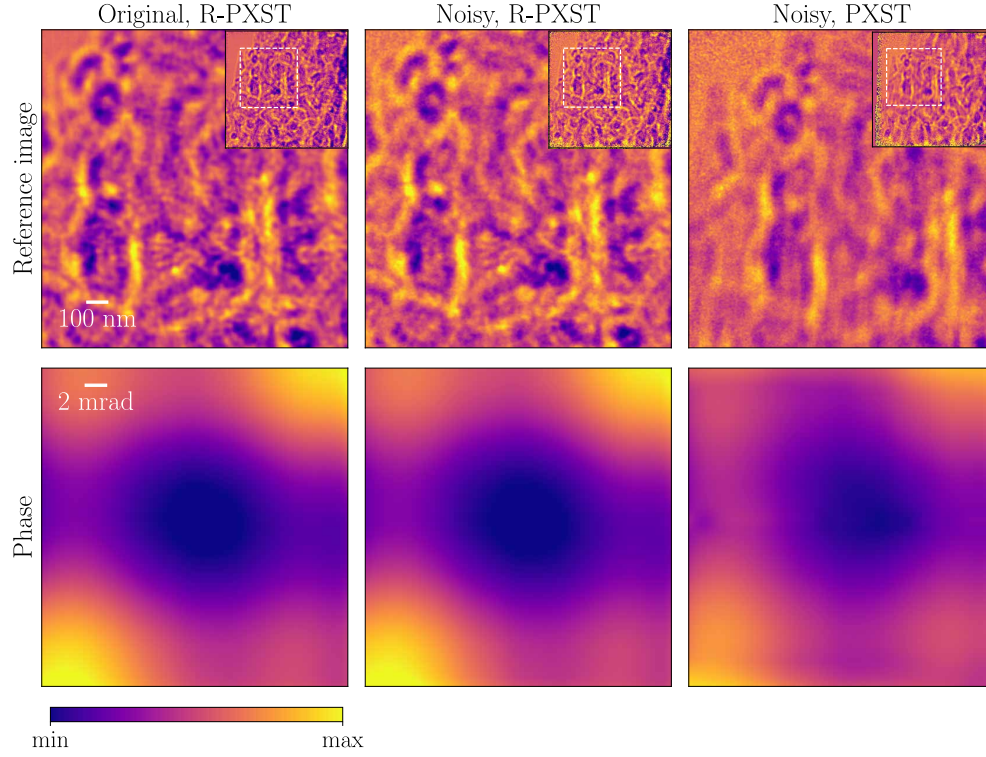
Fig. 10. The reference image of the nanostructured Au sample (top left) and the reference image of the modified dataset with lower average number of counts higher noise ('Noisy') obtained with the R-PXST (top center) and PXST (top right) update algorithms. The lower row shows the retrieved phase profile for the same set of datasets and methods, respectively. Each image is displayed on a colour map, with (min, max) values of (0.8, 1.2), (0.9, 1.1), (0.9, 1.1) for the upper row from left to right and with (min, max) values of (−50 rad, 80 rad) for the lower row. Also this dataset was measured with our MLLs and experimental setup at P11 beamline (PETRA III).

image is about 20 % and shows fringes and interference due to both the branch network and pore network structure of the nanoporous gold. The wavefront has a peak-to-valley error of 76 rad or 12 waves and is separable into vertical and horizontal components which can be attributed to the two MLLs. Each of those separate components is dominated by a third-order polynomial with a magnitude of $\alpha = 0.127 \, \text{rad/mrad}^3$. The PXST update procedure of `speckle_tracking` performed similarly and gave an equivalent result (not shown). This is to be expected since the signal to noise ratio, redundancy, and defocus were all adequate to achieve good results without having to resort to robust machine learning. However, with artificially reduced signal, we do see a difference between PSXT and R-PXST, as apparent from the middle and right columns of Fig. 10. Both reconstructed images have a lower contrast of only 10 %, but many of the high-frequency features in the reference image are washed out in the PXST reconstruction. The form of the recovered wavefront error is similar but the PXST reconstruction gives a lower magnitude of the aberration. The results in the left and middle columns were obtained by the R-PXST iterative reconstruction with the adaptive kernel bandwidth update. The reference profile was evaluated by the kernel regression estimator (Eq. (13)), and the pixel map was updated by the random search algorithm with regularization ($\sigma = 8.0$) and the irrotationality constraint.

After the update, the phase profiles were retrieved from the phase gradient maps with the cosine transform integration [41].

5. Conclusions

We have presented R-PXST, a modified and more robust analysis procedure of ptychographic X-ray speckle tracking. The PXST method is a wavefront metrology tool capable of dealing with highly divergent wavefields. Coupled with a high-numerical-aperture lens, PXST provides access to nanoradian angular sensitivities as well as highly magnified images of the sample. In this new version of PXST we addressed the problem of the outliers and noise present in the intensity measurements and the problem of sparse sampling of the reference profile. These two problems previously limited the domain of applicability of the PXST method.

In order to remedy the above limitations, we employed machine learning techniques in our implementation of the speckle tracking reconstruction. For the estimation of the reference image we chose kernel-based nonparametric regression algorithms. Nonparametric regression methods, in contrast to the parametric estimation, require fewer assumptions about the estimated function and thus reduce the risk of model misspecification. The bandwidth selection based on a cross-validation approach was implemented. Using this approach, the kernel bandwidth in the reference image update can be automatically tailored to the density of the intensity measurements mapped to the reference plane.

The pixel map update yields the main result of the R-PXST reconstruction, the wavefront of the lens. Computational complexity played a big role in the selection of the update algorithm. In particular, the typical sizes of PXST datasets render an approach of updating the whole pixel map simultaneously computationally implausible. Therefore the pixel map is updated at each pixel in the detector grid separately. The computational burden of this approach is far less but it is more susceptible to the noise in intensity measurements. To increase robustness of the pixel map update procedure we substituted the least-squares criterion in Eq. (6) with the Huber error metric, which is a hybrid of L2 norm penalties (for non-outliers) and L1 norm penalties (for outliers) with a scale parameter that divides the PXST dataset into outliers and non-outliers. The scale map is also updated during the reconstruction, so the Huber regression gives a trade-off between ensuring the necessary level of robustness and maintaining a high level of accuracy.

In order to realize these capabilities and to make them available to other scientists, a software suite called `pyrost` was written in Python with a C backend and with a support of concurrent calculations. The software was designed in object-oriented manner and provides a flexible and customizable interface for the whole R-PXST update and data processing pipelines. It is accessible as a Python module and provided with documentation and tutorials. `pyrost` is available under Version 3 or later of the GNU General Public Licence.

We tested the performance of R-PXST and compared it to PXST by recovering known wavefronts in simulated datasets. R-PXST obtains accurate estimates over a wider range of defocus values, especially at the low defocus values where the very high magnification leads to a very sparse sampling of intensity measurements as mapped to the reference plane. R-PXST also performs better with low photon counts and hence is well suited to wavefront sensing using laboratory X-ray sources. We also compared the methods using experimental data collected from a nanoporous gold sample at a high magnification of 36 500 and with artificially high noise. Due to better error handling `pyrost` yields an accurate wavefront for the noisy dataset (as compared with the reconstruction without additional noise) whereas PXST underestimated the wavefront aberration and failed to recover fine features in the reference image.

There is still a need for an automatic means of choosing a kernel bandwidth for the pixel map regularization. Moreover, feature extraction machine learning techniques such as deep autoencoders may be employed to extract features from the PXST dataset or suppress the noise in measured holograms. This approach would decrease the computational complexity of the

pixel map reconstruction. Also, the implementation of CUDA framework may help to boost the performance by executing more intensive calculation on discrete GPUs.

Funding

Acknowledgements

Disclosures

The authors declare no conflicts of interest.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

6. Appendix

6.1. Variance of the reference image estimator

In this section we present the analysis of the reference image estimators employed in PXST and R-PXST, and derive their variances. We assume that the only source of variance in intensity measures comes from the Poisson noise. Therefore, the measured intensity I_{nij} is considered as a random variable following the Poisson distribution with the expected value and variance equal to:

$$\mathrm{E}[I_{nij}] = W_{ij} I_{\mathrm{ref}}(u_{ij}^x - \Delta i_n, u_{ij}^y - \Delta j_n), \quad \mathrm{Var}\left[I_{nij}\right] = E\left[I_{nij}\right], \tag{21}$$

where the estimated value follows the speckle tracking approximation in Eq. (4). Furthermore for simplicity the white-field is assumed to be uniform, $W_{ij} \equiv W_0$.

The reference image estimator for PXST, $\hat{I}_{\mathrm{ref}}$, is given by Eq. (7). Since the reference image is discretely sampled, we are interested in the discrete representation of $\hat{I}_{\mathrm{ref}}$. As explained in [9, Appendix A], PXST uses bilinear interpolation to estimate the discrete reference profile as follows:

$$\hat{I}_{\mathrm{ref}}(i, j) = \frac{\sum_n \sum_{i'} \sum_{j'} \Lambda(i - u_{i'j'}^x + \Delta i_n, j - u_{i'j'}^y + \Delta j_n) \, W_{i'j'} I_{ni'j'}}{\sum_n \sum_{i'} \sum_{j'} \Lambda(i - u_{i'j'}^x + \Delta i_n, j - u_{i'j'}^y + \Delta j_n, h) \, W_{i'j'}^2}, \tag{22}$$

where $\Lambda(i, j)$ is the two-dimensional triangular kernel given by

$$\Lambda(i, j) = \begin{cases} (1 - |i|) * (1 - |j|), & |i| \leq 1 \wedge |j| \leq 1, \\ 0, & |i| > 1 \cap |j| > 1 \end{cases}, \tag{23}$$

and we sum over the subset Ω given by

$$\Omega = \{n, i, j : |i - u_{ij}^x + \Delta i_n| < 1, j - |u_{ij}^y + \Delta j_n| < 1\}. \tag{24}$$

That is, we sum over intensity measurements that are directly adjacent to the pixel at the point i, j (as mapped to the reference plane). For the case of the uniform white-field the bilinear reference image estimator is given by

$$\hat{I}_{\text{ref}}(i, j) = \frac{1}{N(\Omega)} \sum_{n,i',j' \in \Omega} \Lambda(i - u_{i'j'}^x + \Delta i_n, j - u_{i'j'}^y + \Delta j_n) \frac{I_{ni'j'}}{W_0}. \tag{25}$$

By performing a second-order Taylor expansion of the regression function I_{ref} in Eq. (21) around point i, j, the variance of the estimator can be derived as (see Eq. (14)) [33]

$$\text{Var}\left[\hat{I}_{\text{ref}}(i, j)\right] = \frac{R(\Lambda)I_{\text{ref}}(i, j)}{N(\Omega)} + o\{N(\Omega)^{-1}\}, \tag{26}$$

where $R(\Lambda) = \iint \Lambda(x, y)^2 dx dy = 4/9$. We see that the variance of the bilinear reference image estimator, in the case of Poisson noise, is inversely proportional to the size of the Ω subset, which as a function of the reference coordinate i, j represents the number of times, that a given part of the sample is measured during the scan. We call this quantity $N(\Omega)$ the "redundancy" of the scan.

We now consider how the variance of the kernel regression estimator $\hat{I}_{\text{ref}}^h$ of Eq. (13) compares to the bilinear estimator. For the case of uniform white-field the estimator reduces to

$$\hat{I}_{\text{ref}}^h(i, j) = \frac{1}{N(\Omega_h)} \sum_{n,i',j' \in \Omega_h} K(i - u_{i'j'}^x + \Delta i_n, j - u_{i'j'}^y + \Delta j_n, h) \frac{I_{ni'j'}}{W_0}, \tag{27}$$

where we sum over the subset Ω_h given by

$$\Omega_h = \{n, i, j : \sqrt{(i - u_{ij}^x + \Delta i_n)^2 + (j - u_{ij}^y + \Delta j_n)^2} < 3h\}. \tag{28}$$

The sum in Eq. (27) includes intensity measurements that are no more than $3h$ away from the point i, j (implying a $3h$ cut-off of the Gaussian kernel). Again, following the same procedure as in Eq. (26), the variance of $\hat{I}_{\text{ref}}^h$ can be estimated as

$$\text{Var}\left[\hat{I}_{\text{ref}}^h(i, j)\right] = \frac{R(K)I_{\text{ref}}(i, j)}{N(\Omega_h)h} + o\{(N(\Omega)h)^{-1}\}, \tag{29}$$

where $R(K)$ is evaluated to $1/(2\sqrt{\pi})$. Compared to the bilinear estimator, the variance of the kernel regression estimator is reduced by the factor $1.576\, N(\Omega_h)h/N(\Omega)$. As mentioned above, reducing the kernel size h reduces the variance, at the expense of band-limiting the reconstructed reference image.

6.2. Computer simulations

Simulations of the holograms were carried out using the Rayleigh-Sommerfeld convolution (RSC) algorithm [44], which evaluates the general formula of the Rayleigh-Sommerfeld diffraction using an approach based on the fast Fourier transform. In one dimension, the RSC wavefield propagation of monochromatic light over a distance d, is given by:

$$u(x') = \frac{d}{j\sqrt{\lambda}} \int_{-\infty}^{\infty} u(x) \frac{\exp\left[-jkr(x, x')\right]}{r(x, x')^{3/2}} dx, \tag{30}$$

where $r(x, x') \equiv |\Delta \mathbf{r}| = \left[(x' - x)^2 + d^2\right]^{1/2}$, and now j is the imaginary number. The X-ray beam wavefield at the lens exit-surface was assumed to exhibit a second order focusing phase term with a third order coma profile,

$$\Pi(x_0) = \Pi_A(a_x x_0) \exp\left[-\frac{j\pi x_0^2}{\lambda f} + j\alpha\left(\frac{x_0}{f}\right)^3\right], \tag{31}$$

where α is the third order aberration coefficient, x_0 is the coordinate in the lens and a_x is the size of the lens aperture. $\Pi_A(x)$ is the square root of the transmission of the lens. After propagating the wavefield to the sample plane, it was modulated by the transmission profile $t_b(x)$ of the barcode, given by

$$t_b(x) = t_{\text{air}} + \frac{\sqrt{T_1} - t_{\text{air}}}{2} \left(\tanh\left(\frac{x - x_1^b}{\sigma_b}\right) + \tanh\left(\frac{x_N^b - x}{\sigma_b}\right) \right)$$
$$+ \frac{\sqrt{T_2} - \sqrt{T_1}}{2} \sum_{n=1}^{(N-1)/2} \left(\tanh\left(\frac{x - x_{2n}^b}{\sigma_b}\right) + \tanh\left(\frac{x_{2n+1}^b - x}{\sigma_b}\right) \right), \quad (32)$$

where $t_{\text{air}} = 1.0$, T_1, T_2 is a pair of power transmission coefficients of the bilayers, x_n^b is the coordinate of the nth bar, and σ_b is the inter-diffusion length. The wavefield at the sample exit surface is then propagated to the detector plane to generate the intensity data. The simulation also takes into account that the X-ray source is incoherent and has a Gaussian angular distribution given by $\exp(-\theta^2/2\sigma_s^2)$, where θ is the angle made by a ray pointing from the incoherent source point to the lens aperture. This is accomplished by computing the image at the detector plane as

$$I(\mathbf{x}, z, \sigma_s) = \left[\frac{1}{\sqrt{2\pi}\sigma_s\tilde{z}} \exp\left(\frac{\mathbf{x}^2}{2\sigma_s^2\tilde{z}^2}\right) \right] * I_c(\mathbf{x}, z), \quad (33)$$

where $\tilde{z} = f + z + z_1$ is the distance between the lens and the detector, $*$ denotes the convolution operation, and $I_c(\mathbf{x}, z)$ is the intensity of the wavefield in the plane of the detector for an on-axis point source of light. Lastly, we employed a pseudo-random number generator to simulate Poisson noise which was added to the images.

References

1. L. Mino, E. Borfecchia, J. Segura-Ruiz, C. Giannini, G. Martinez-Criado, and C. Lamberti, "Materials characterization by synchrotron x-ray microprobes and nanoprobes," Rev. Mod. Phys. **90**, 025007 (2018).
2. K. T. Murray, A. F. Pedersen, I. Mohacsi, C. Detlefs, A. J. Morgan, M. Prasciolu, C. Yildirim, H. Simons, A. C. Jakobsen, H. N. Chapman, H. F. Poulsen, and S. Bajt, "Multilayer laue lenses at high x-ray energies: performance and applications," Opt. Express **27**, 7120–7138 (2019).
3. P. Villanueva-Perez, H. Fleckenstein, M. Prasciolu, K. T. Murray, M. Domaracký, K. Gregorič, V. Mariani, L. Gelisio, M. Kuhn, J. Hannappel, O. Yefanov, N. Ivanov, I. Sarrou, D. Pennicard, J. Becker, M. von Zimmermann, O. Gutowski, A.-C. Dippel, H. N. Chapman, and S. Bajt, "Scanning compton x-ray microscopy," Opt. Lett. **46**, 1920–1923 (2021).
4. C. Chang, A. Sakdinawat, P. Fischer, E. Anderson, and D. Attwood, "Single-element objective lens for soft x-ray differential interference contrast microscopy," Opt. Lett. **31**, 1564–1566 (2006).
5. A. Sakdinawat and Y. Liu, "Phase contrast soft x-ray microscopy using zernike zone plates," Opt. Express **16**, 1559–1564 (2008).
6. W. Chao, B. D. Harteneck, J. A. Liddle, E. H. Anderson, and D. T. Attwood, "Soft x-ray microscopy at a spatial resolution better than 15 nm," Nature **435**, 1210–1213 (2005).
7. A. J. Morgan, H. M. Quiney, S. Bajt, and H. N. Chapman, "Ptychographic x-ray speckle tracking," J. Appl. Crystallogr. **53**, 760–780 (2020).
8. A. J. Morgan, K. T. Murray, M. Prasciolu, H. Fleckenstein, O. Yefanov, P. Villanueva-Perez, V. Mariani, M. Domaracky, M. Kuhn, S. Aplin, I. Mohacsi, M. Messerschmidt, K. Stachnik, Y. Du, A. Burkhart, A. Meents, E. Nazaretski, H. Yan, X. Huang, Y. S. Chu, H. N. Chapman, and S. Bajt, "Ptychographic x-ray speckle tracking with multi-layer laue lens systems," J. Appl. Crystallogr. **53**, 927–936 (2020).
9. A. J. Morgan, K. T. Murray, H. M. Quiney, S. Bajt, and H. N. Chapman, "*speckle-tracking*: a software suite for ptychographic x-ray speckle tracking," J. Appl. Crystallogr. **53**, 1603–1612 (2020).
10. S. Bérujon, E. Ziegler, R. Cerbino, and L. Peverini, "Two-dimensional x-ray beam phase sensing," Phys. Rev. Lett. **108**, 158102 (2012).
11. K. S. Morgan, D. M. Paganin, and K. K. W. Siu, "X-ray phase imaging with a paper analyzer," Appl. Phys. Lett. **100**, 124102 (2012).
12. S. Berujon, H. Wang, and K. Sawhney, "X-ray multimodal imaging using a random-phase object," Phys. Rev. A **86**, 063813 (2012).
13. M.-C. Zdora, "State of the art of x-ray speckle-based phase-contrast and dark-field imaging," J. Imaging **4** (2018).

14. S. Bérujon, E. Ziegler, and P. Cloetens, "X-ray pulse wavefront metrology using speckle tracking," J. Synchrotron Radiat. **22**, 886–894 (2015).

15. S. Bajt, M. Prasciolu, H. Fleckenstein, M. Domaracký, H. N. Chapman, A. J. Morgan, O. Yefanov, M. Messerschmidt, Y. Du, K. T. Murray, V. Mariani, M. Kuhn, S. Aplin, K. Pande, P. Villanueva-Perez, K. Stachnik, J. P. Chen, A. Andrejczuk, A. Meents, A. Burkhardt, D. Pennicard, X. Huang, H. Yan, E. Nazaretski, Y. S. Chu, and C. E. Hamm, "X-ray focusing with efficient high-NA multilayer laue lenses," Light. Sci. & Appl. **7**, 17162–17162 (2017).

16. H. Yan, H. C. Kang, R. Conley, C. Liu, A. T. Macrander, G. B. Stephenson, and J. Maser, "Multilayer laue lens: A path toward one nanometer x-ray focusing," X-Ray Opt. Instrumentation **2010**, 1–10 (2010).

17. M. Prasciolu, A. F. G. Leontowich, J. Krzywinski, A. Andrejczuk, H. N. Chapman, and S. Bajt, "Fabrication of wedged multilayer laue lenses," Opt. Mat. Express **5**, 748–755 (2015).

18. H. N. Chapman, M. Prasciolu, K. T. Murray, J. L. Dresselhaus, and S. Bajt, "Analysis of x-ray multilayer laue lenses made by masked deposition," Opt. Express **29**, 3097–3113 (2021).

19. H. N. Chapman and S. Bajt, "A ray-trace analysis of x-ray multilayer laue lenses for nanometer focusing," J. Opt. **22**, 115610 (2020).

20. N. Ivanov, "Python robust speckle tracking library," https://github.com/simply-nicky/pyrost (2020). [retrieved 8 April 2021].

21. D. Paganin, *Coherent X-Ray Optics* (Oxford University Press, 2006).

22. E. A. Nadaraya, "On estimating regression," Theory Probab. & Its Appl. **9**, 141–142 (1964).

23. G. S. Watson, "Smooth regression analysis," Sankhyā: The Indian J. Stat. Ser. A pp. 359–372 (1964).

24. C. J. Stone, "Consistent nonparametric regression," The annals statistics pp. 595–620 (1977).

25. H.-G. Müller, "Weighted local regression and kernel methods for nonparametric curve fitting," J. Am. Stat. Assoc. **82**, 231–238 (1987).

26. A. B. Owen, "A robust hybrid of lasso and ridge regression," Tech. rep., Stanford University (2006).

27. P. J. Huber, *Robust statistics*, vol. 523 (John Wiley & Sons, 2004).

28. N. Ivanov, "pyrost documentation," https://robust-speckle-tracking.readthedocs.io/en/latest/index.html (2021). [retrieved 8 April 2021].

29. V. V. Nieuwenhove, J. D. Beenhouwer, F. D. Carlo, L. Mancini, F. Marone, and J. Sijbers, "Dynamic intensity normalization using eigen flat fields in x-ray imaging," Opt. Express **23**, 27975–27989 (2015).

30. A. B. Tsybakov, "Introduction to nonparametric estimation." (2009).

31. K. Takezawa, *Introduction to nonparametric regression*, vol. 606 (John Wiley & Sons, 2005).

32. P. Čížek and S. Sadıkoğlu, "Robust nonparametric regression: A review," Wiley Interdiscip. Rev. Comput. Stat. **12**, e1492 (2020).

33. M. P. Wand and M. C. Jones, *Kernel Smoothing*, Monographs on statistics and applied probability (CRC Press, 1994).

34. A. T. Brünger, "Free r value: a novel statistical quantity for assessing the accuracy of crystal structures," Nature **355**, 472–475 (1992).

35. S. Wright, J. Nocedal *et al.*, "Numerical optimization," Springer Sci. **35**, 7 (1999).

36. P. J. Huber, "Robust Estimation of a Location Parameter," The Annals Math. Stat. **35**, 73 – 101 (1964).

37. R. Storn, "On the usage of differential evolution for function optimization," Biennial Conf. North Am. Fuzzy Inf. Process. Soc. - NAFIPS pp. 519–523 (1996).

38. R. Storn and P. Kenneth, "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," J. Glob. Optim. **11**, 341–359 (1997).

39. S. Das and P. N. Suganthan, "Differential evolution: A survey of the state-of-the-art," IEEE Transactions on Evol. Comput. **15**, 4–31 (2011).

40. D. M. Paganin, H. Labriet, E. Brun, and S. Berujon, "Single-image geometric-flow x-ray speckle tracking," Phys. Rev. A **98**, 053813 (2018).

41. C. Kottler, C. David, F. Pfeiffer, and O. Bunk, "A two-directional approach for grating based differential phase contrast imaging using hard x-rays," Opt. Express **15**, 1175–1181 (2007).

42. J. L. Dresselhaus, H. Fleckenstein, M. Domaracky, M. Prasciolu, N. Ivanov, J. Carnis, K. T. Murray, A. Morgan, H. N. Chapman, and S. Bajt, "Precise wavefront characterization of x-ray optical elements using a laboratory source," (2022). Submitted.

43. S. Shi, Y. Li, B.-N. Ngo-Dinh, J. Markmann, and J. Weissmüller, "Scaling behavior of stiffness and strength of hierarchical network nanomaterials," Science **371**, 1026–1033 (2021).

44. V. Nascov and P. C. Logofătu, "Fast computation algorithm for the rayleigh-sommerfeld diffraction formula using a type of scaled convolution," Appl. Opt. **48**, 4310–4319 (2009).