Ptychographic X-ray speckle tracking[1]

Andrew J. Morgan,[a,b]* Harry M. Quiney,[a] Saša Bajt[c,d] and Henry N. Chapman[b,d,e]

[a]ARC Centre of Excellence in Advanced Molecular Imaging, School of Physics, University of Melbourne, Parkville, Victoria 3010, Australia, [b]CFEL, Deutsches Elektronen-Synchrotron DESY, Notkestrasse 85, 22607 Hamburg, Germany, [c]DESY, Notkestrasse 85, 22607 Hamburg, Germany, [d]The Hamburg Centre for Ultrafast Imaging, Luruper Chaussee 149, 22761 Hamburg, Germany, and [e]Department of Physics, University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. *Correspondence e-mail: morganaj@unimelb.edu.au

A method is presented for the measurement of the phase gradient of a wavefront by tracking the relative motion of speckles in projection holograms as a sample is scanned across the wavefront. By removing the need to obtain an undistorted reference image of the sample, this method is suitable for the metrology of highly divergent wavefields. Such wavefields allow for large magnification factors that, according to current imaging capabilities, will allow for nanoradian angular sensitivity and nanoscale sample projection imaging. Both the reconstruction algorithm and the imaging geometry are nearly identical to that of ptychography, except that the sample is placed downstream of the beam focus and that no coherent propagation is explicitly accounted for. Like other X-ray speckle tracking methods, it is robust to low-coherence X-ray sources, making it suitable for laboratory-based X-ray sources. Likewise, it is robust to errors in the registered sample positions, making it suitable for X-ray free-electron laser facilities, where beam-pointing fluctuations can be problematic for wavefront metrology. A modified form of the speckle tracking approximation is also presented, based on a second-order local expansion of the Fresnel integral. This result extends the validity of the speckle tracking approximation and may be useful for similar approaches in the field.

1. Introduction

New facilities are providing ever more brilliant X-ray sources. To access the full potential of these sources we need X-ray optics that are capable of focusing light to meet the requirements of various imaging modalities. Thus there is an increasing need for at-wavelength and *in situ* wavefront metrology techniques that are capable of measuring the performance of these optics to the level of their desired performance. This is a challenging task, as current X-ray optics technologies are attaining focal spot sizes below 10 nm (Mimura *et al.*, 2010; Huang *et al.*, 2013; Morgan *et al.*, 2015; Bajt *et al.*, 2018; Murray *et al.*, 2019). Furthermore, adaptive optics are being employed to correct for wavefront aberrations by altering the physical state of a lens system in response to real-time measurements of wavefront errors (Mercère *et al.*, 2006; Zhou *et al.*, 2019). Such systems therefore benefit from fast and accurate wavefront metrology for rapid feedback.

Wavefront metrology techniques generally fall into one of three categories (Wilkins *et al.*, 2014; Wang *et al.*, 2015): (i) direct phase measurements, such as interferometry using crystals (Bonse & Hart, 1965); (ii) phase gradient measurements, such as Hartmann sensors (Lane & Tallon, 1992), coded aperture methods (Olivo & Speller, 2007) and grating-based interferometry (David *et al.*, 2002); and (iii) propagation-based methods sensitive to the second derivative of the
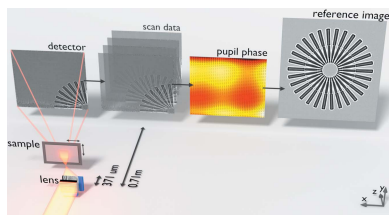
wavefront's phase (Wilkins *et al.*, 1996; Wang *et al.*, 2015; Bérujon *et al.*, 2014).

One such method, falling into the second category above, was introduced by Bérujon, Ziegler *et al.* (2012) and Morgan *et al.* (2012) (no relation to the current author). This method is a wavefront metrology technique based on near-field speckle-based imaging, which was termed the 'X-ray speckle tracking' (XST) technique. In XST, the 2D phase gradient of a wavefield can be recovered by tracking the displacement of localized 'speckles' between an image and a reference image produced in the projection hologram of an object with a random phase/absorption profile. Additionally, XST can be employed to measure the phase profile of an object's transmission function. Thanks to the simple experimental setup, high angular sensitivity and compatibility with low-coherence sources, this method has since been actively developed for use in synchrotron and laboratory light sources [see Zdora (2018) for a recent review].

In ptychography, a sample is scanned across the beam wavefront (typically at or near the focal plane of a lens) while diffraction data are collected in the far field of the sample. An iterative algorithm is usually employed to update initial estimates for the complex wavefront of the illumination and the sample transmission functions. If illuminated regions of the sample overlap sufficiently, then it is possible for a unique solution for both of these functions to be obtained (Hüe *et al.*, 2010). Thus, ptychography is an imaging modality that performs both aberration-free sample imaging and wavefront metrology simultaneously. This is in contrast to XST where these two imaging modalities correspond to separate imaging geometries.

Ptychography can also be performed in the near-field diffraction regime, as reported for example by Nugent *et al.* (1996). Stockmar *et al.* (2013) found that the illumination must be sufficiently inhomogeneous to allow for a successful reconstruction, and those authors suggested the use of an additional diffuser as a means of achieving this. Consequently, this near-field ptychographic imaging approach closely resembles that of many XST-based approaches in its experimental configuration. The key distinction here is that, in ptychography, a fully coherent wave model is employed. This can lead to non-unique solutions in situations where an XST-based approach would yield a well defined solution, although typically this solution will contain less information at lower resolution than a successful ptychographic reconstruction.

We propose a combined approach, which we term ptychographic X-ray speckle tracking (PXST). In this approach, near-field in-line holograms are recorded as an unknown sample is scanned across an unknown wavefield. Estimates for the undistorted sample projection image and the wavefield are then updated on the basis of the observed speckle displacements. There is no reference image and no additional speckle-producing object is required. This imaging geometry allows for XST to be used for highly divergent X-ray beams, thus expanding the applicability of this simple and robust method to include next-generation high-numerical-aperture X-ray lenses.

Table 1
Symbols.

$I_n(\mathbf{x}, z)$	n^{th} recorded image
$I_{\text{ref}}(\mathbf{x}, z)$	Reference projection image of the sample
$\Delta\mathbf{x}_n$	Displacement of sample in transverse plane
$T(\mathbf{x})$	Transmission function of the quasi-2D sample
z_1	Source-to-sample distance
z	Sample-to-detector distance
$\bar{z} = zz_1/(z_1 + z)$	Effective propagation distance
$M = (z_1 + z)/z_1$	Geometric magnification factor
λ	Wavelength of radiation
σ_{eff}	Smallest resolvable speckle displacement in the plane of the detector
i	$(-1)^{1/2}$
$\mathbf{a} \cdot \mathbf{b}$	Dot product between vectors $\mathbf{a}$ and $\mathbf{b}$
$p(\mathbf{x}, 0) = w^{1/2}(\mathbf{x}) \exp[i\phi(\mathbf{x})]$	Illumination wavefront in the sample plane; w and ϕ are the intensity and phase, respectively
$p(\mathbf{x}, z) = W^{1/2}(\mathbf{x}) \exp[i\Phi(\mathbf{x})]$	Illumination wavefront in the detector plane; W and Φ are the intensity and phase, respectively
$\mathbf{x} \equiv (x, y)$	Transverse coordinate
$\nabla \equiv (\partial/\partial x, \partial/\partial y)$	Transverse gradient operator

Bérujon *et al.* (2014) have proposed a similar method, also based on XST and compatible with highly divergent beams. In their approach, the second derivative of the wavefront phase is measured. Additionally, nanoradian angular sensitivity can be achieved with relatively small step sizes in the scan of the sample on a piezo-driven stage (discussed further in the next section). In contrast, PXST more closely aligns with current XST-based techniques, such as the 'unified modulated pattern analysis' method of Zdora *et al.* (2017, 2018), that do not rely on small sample translations.

In Section 2, we briefly review the XST method and its extension to PXST. In Section 3 we present the governing equation, which is based on a second-order expansion of the Fresnel diffraction integral (presented in Appendix *A*). The region of validity for the speckle tracking approximation determines the applicable imaging geometries, which are presented in Section 4 and Appendix *B*. We present the iterative reconstruction algorithm and the target function, which is to be minimized by the algorithm, in Section 5. Conditions for the uniqueness of the solution are discussed in Appendix *C*. Finally, the theoretically achievable angular sensitivity of the wavefront reconstruction and the imaging resolution of the sample projection image are presented in Appendix *D*. For reference, we define the principal mathematical symbols used throughout the paper in Table 1. In Table 2, we summarize the main results of this article and refer the reader to the relevant sections.

2. Background

The problem with wavefront metrology is that it is much more difficult to measure a wavefront's phase than its intensity; the intensity can be measured directly by placing a photon-counting device at any point in the wavefront's path, whereas the phase information is indirectly encoded in the wavefront's intensity profile as it propagates through space. For plane-wave illumination, no measurement of the wavefront's intensity alone will reveal its direction of propagation. One solution

Table 2
The PXST method.

Governing equation	$I_n(\mathbf{x}, z) \simeq W(\mathbf{x})I_{ref}[\mathbf{u}(\mathbf{x}) - \Delta\mathbf{x}_n, \overline{z}]$	See Section 3 and Appendix A
	$w(\mathbf{x})(\overline{z}/z)^2 I_{ref}(\mathbf{x}, \overline{z}) \simeq I_n[\mathbf{u}^{-1}(\mathbf{x} + \Delta\mathbf{x}_n), z]$	Reciprocal form for the above equation; $\mathbf{u}^{-1}$ is the inverse of $\mathbf{u}$
Target function	$\varepsilon = \sum_n \int \int d\mathbf{x} \, [1/\sigma_I^2(\mathbf{x})]\{I_n(\mathbf{x}, z)$	Equation (26) in Section 5; to be minimized with respect to I_{ref}, $\nabla\Phi$ and $\Delta\mathbf{x}_n$
	$\quad - W(\mathbf{x})I_{ref}[\mathbf{u}(\mathbf{x}) - \Delta\mathbf{x}_n, \overline{z}]\}^2$	
Geometric mapping	$\mathbf{u}(\mathbf{x}) = \mathbf{x} - (\lambda z/2\pi)\nabla\Phi(\mathbf{x})$	See equation (54) in Appendix A
	$\mathbf{u}^{-1}(\mathbf{x}) = \mathbf{x} + (\lambda z/2\pi)\nabla\phi(\mathbf{x})$	Reciprocal form for the above equation; see equation (37) in Appendix A
Imaging geometry	See Fig. 3	Described in Section 2
Iterative update algorithm	See Fig. 6	Described in Section 3
Angular sensitivity	$\Delta\Theta_\phi = \sigma_{eff}/z$	In the plane of the sample; see equation (123)
	$\Delta\Theta_\Phi = \sigma_{eff}/zM$	In the plane of the detector; see equation (126)
Phase sensitivity	$\Delta\phi = \Delta\Phi = (2\pi/\lambda)(\sigma_{eff}^2/zM)$	Sample/detector plane; see equation (130)

to this problem is to place an absorbing object at a known point in the path of the light, from which the direction of propagation can then be inferred from the relative displacement between the centre of the object and the shadow cast on a screen some distance away, just as the angle of the sun can be estimated by following the line from a shadow to its object.

This simple idea forms the basis of the Hartmann sensor (Daniel & Ghozeil, 1992), shown schematically in Fig. 1. Originally designed to measure aberrations in telescopes and later for atmospheric distortions, the Hartmann sensor can be used as an X-ray wavefront metrology tool (Mayo & Sexton, 2004)[2] by cutting a regular grid of small holes, spaced at known intervals (say x_i where i is the hole index), in a mask and then recording the shadow image on a detector, which is placed a small distance downstream of the mask.

Provided that each hole can be matched with each shadow image, the angle made between them, $\Theta(x_i) = \arctan[\Delta x(x_i)/z]$ (in one dimension), is equal to the average direction of propagation of that part of the wavefront passing through each hole $\Theta(x_i) = (\lambda/2\pi d) \int_{x_i-d/2}^{x_i+d/2} [\partial\phi(x)/\partial x] \, dx$, where $\Delta x(x_i)$ is the observed displacement along x of the ith shadow, z is the distance between the mask and the detector, and d is the hole width.

With a suitable interpolation routine, $\Theta(x)$ can be estimated from the set of $\Theta(x_i)$ and the phase profile can be obtained up to a constant with

$$\phi(x) \simeq (2\pi/\lambda) \int \Theta(x) \, dx. \tag{1}$$

One limitation of this technique is that the resolution obtained is limited by the spacing between each hole in the mask. For example, Mercère et al. (2006) used a Hartmann sensor in an active optic system with a grid of 75×75 holes over a 10×10 mm area, whereas the CCD detector had a 1024×1024 grid of pixels over a 13×13 mm area. Thus the Hartman sensor had a resolution 10.5 times worse than the CCD detector.

The maximum density of the holes in the grid is limited. This is because the task of uniquely matching each shadow image with each hole becomes more difficult as the hole density is increased – a problem that is easier to appreciate in two

dimensions. In 2012, Bérujon and co-workers realized a simple yet elegant solution to this problem, one that allowed for an arbitrarily fine grid of 'masks' with a resolution and sensitivity limited only by the CCD pixel array and the signal-to-noise ratio (Bérujon, Ziegler et al., 2012). Their solution, XST, is to
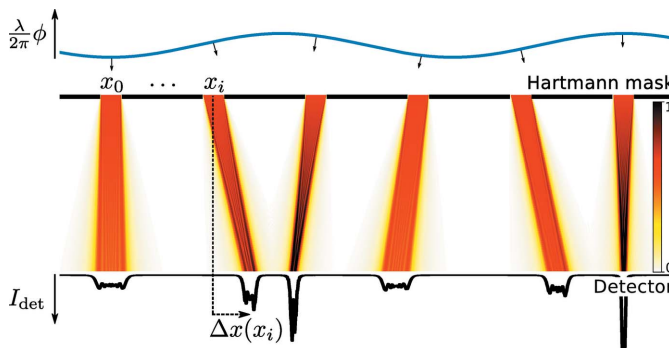

Figure 1
Illustration of the Hartmann sensing principle. Top: phase of the wavefront incident on the entrance surface of the mask. The phase has been scaled by $\lambda/2\pi$ so that the normal to the tangent is parallel to the local direction of propagation of the wavefront. The arrows indicate the direction of propagation at the centre of each mask hole. Middle: intensity of the wavefront as it propagates from the mask (top) to the detector (bottom). The colour scale is shown on the right. Bottom: the one-dimensional intensity profile of the wavefront as measured by the detector.
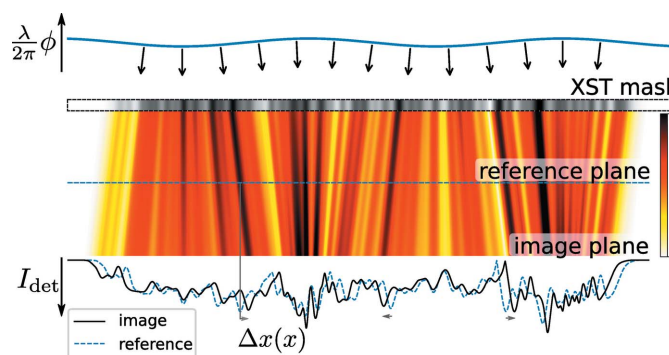

Figure 2
Illustration of the X-ray speckle tracking (XST) principle. Top: as in Fig. 1. Middle: as in Fig. 1, with the binary mask replaced by a random phase/absorption mask (dashed outline). Bottom: sub-regions of the measured shadow image (solid black line) are compared with the reference shadow image (dashed blue line) to determine displacements (black arrows).

[2] In this work, an array of refractive elements was used rather than a grid of mask holes.

replace the binary mask of identical holes with a thin random phase object, such as a diffuser, as shown in Fig. 2. Because the diffuser is random (in the sense that the modulation of the beam by the diffuser is both detailed and non-repeating over the relevant spatial frequencies of the image), the shadow from each sub-region of the diffuser is unique – encoded by the speckle pattern seen on the detector – so that one can therefore consider any point in the diffuser to be the centre of a virtual Hartmann hole. In this sense, the random object serves as a high-density fiducial marker for each of the light rays that pass from the reference or mask plane to the detector. Note that this approach requires a greater degree of beam coherence than the Hartman sensor, to the extent necessary to provide sufficient visibility of the speckles, so that each speckle pattern can be distinguished from its neighbour.

In the Hartmann sensor, it is assumed that the mask is well characterized, so that the shadow positions can be compared with their ideal positions, which are known *a priori*. However, since the mask is no longer a simple geometric object (in the sense that it is difficult to know *a priori* the modulation function of the mask with a sufficient degree of precision), it is now necessary to record a reference image of the mask with which to compare the distorted image.

In addition to measurements of a wavefront's phase, the XST principle can be extended to incorporate phase imaging of samples. This can be achieved by recording an image of the wavefront with the diffuser (acting as a mask) in the beam path – this image is called the 'reference' image. Then, another image is recorded with an additional sample (the one to be imaged) placed in the beam path, in addition to the diffuser – this is referred to simply as the 'image'. Here the relative displacements between the 'reference' and the 'image' are due not to the phase profile of the wavefront, which affects both images equally, but to the phase profile of the sample transmission function.

The following two XST imaging configurations were suggested by Bérujon *et al.*, one for imaging samples and the other for wavefront metrology:

(i) in the differential configuration a speckle image is recorded with and without the addition of a sample, and

(ii) in the absolute configuration a speckle image is recorded at two detector distances with respect to the mask.

In (i), the relative motion of speckles reveals the local phase gradient of the sample in the beam, whereas in (ii), the total wavefront phase is recovered and this is therefore useful for characterizing X-ray beamline optics (this is the configuration shown in Fig. 2). Of course, it is still possible to characterize beamline optics in (i) (just not *in situ*) by placing the optical element in the sample position. This approach has been useful, for example, in measuring the phase profile of compound refractive lens systems (Bérujon, Wang & Sawhney, 2012) but is impractical for larger systems such as Kirkpatrick–Baez mirrors.

Since the proposal by Bérujon *et al.* there have been a number of substantial improvements: see for example the extensive review by Zdora (2018). For example, Zanette *et al.* (2014) developed a method where a diffuser is scanned so as to obtain a number of reference/image pairs at different diffuser positions. This step can add a great deal of redundancy, which improves the angular sensitivity of the method and even allows for multi-modal imaging of the sample when employed in the differential configuration. In subsequent publications, this approach has been termed the unified modulated pattern analysis (UMPA) method (Zdora *et al.*, 2017, 2018).

In the absolute configuration, where the reference and image have been recorded at two detector distances, the smallest resolvable angular displacement (the angular sensitivity) is given by the ratio of the effective pixel size, which is the smallest resolvable displacement of a speckle (including effects such as fringe visibility, finite pixel size, beam coherence and noise), to the distance between the reference and image planes: $\Delta\theta = d/\Delta z$. Therefore, the best accuracy is obtained by maximizing the distance between the reference and image planes. However, for highly divergent wavefields, as would be produced (for example) by a high-numerical-aperture lens system, there arises an unavoidable trade-off between the wavefront sampling frequency and the angular sensitivity. In this situation the ideal location of the image plane is as far downstream of the lens focus as is required to fill the detector array with the beam, as this maximizes the wavefront sampling frequency. In order to minimize $\Delta\theta$ (maximize Δz) one should then 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. Therefore, the smallest resolvable speckle shift will be larger than that obtainable by plane-wave illumination, by a factor proportional to the beam divergence.

Realizing this, Bérujon *et al.* (2014) devised an XST technique, X-ray speckle scanning (XSS), that relies on small displacements of the XST mask between acquired images. No reference is required and the diffraction data are recorded in a single plane. This enables the sampling frequency to be maximized by placing the detector such that the divergent beam fills the pixel array. Without a reference, however, the speckle locations in one image are instead compared with the locations observed in neighbouring images. As the speckle displacements in each image are proportional to the phase gradient, the differentials of the speckle locations between images are proportional to the second derivative of the phase; thus this approach can be viewed as a wavefront curvature measurement. The achievable angular sensitivity is now proportional to the step size of the mask, which can be substantially smaller than the effective pixel size. Interestingly, this approach is similar in principle to the Wigner-distribution deconvolution approach described by Chapman (1996).

In the following section, we describe an approach that is similar in principle to the one described above:

(iii) ptychographic XST: shadow images are recorded as the mask/object is translated across the wavefront.

In this method (see Fig. 3) the unknown object acts as both the imaging target and the speckle mask simultaneously. There
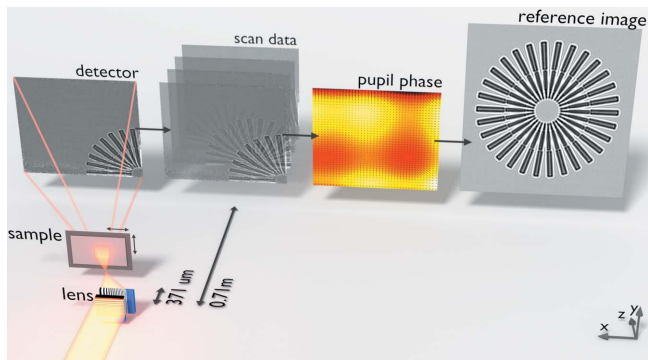
research papers

Figure 3
Illustration of the ptychographic XST method. The beamline illumination was focused (off-axis) in two dimensions by two linear focusing lenses, with numerical apertures of 0.015 (horizontal) and 0.014 (vertical). The Siemens star sample was placed 371 μm downstream of the focal plane. Images were recorded on a CCD pixel array detector 0.71 m downstream of the focus. The scan data consist of 49 shadow images, recorded as the sample was translated across the beam profile. The wavefront phase and reference maps were refined iteratively.

is no special reference image; rather each image serves as a reference for all other images. Both the wavefront phase (without the influence of the object) and the object image (without the influence of wavefront distortions) are determined in an iterative update procedure. At each iteration, speckles[3] in the recorded images are compared with the current estimate of the reference (in contrast to the XSS method). Images are recorded at a fixed detector distance and there is no trade-off between phase sensitivity and the wavefront sampling frequency, making this method suitable for highly divergent beams. Because the speckle displacements are compared between the image and the estimated reference, large angular distortions can be accommodated. This is advantageous because it allows for the sample to be placed very near the beam focus, where the phase gradients across the sample surface are largest and where the magnification factor allows for high imaging resolution and angular sensitivity.

3. The speckle tracking approximation

In this section we describe the governing equation that relates the measured intensities in each image and the reference in terms of the wavefront phase. For monochromatic light, in the Fresnel diffraction regime the image formed on a detector placed a distance z downstream of an object is given by

$$I_{\text{ref}}(\mathbf{x}, z) = \frac{1}{(\lambda z)^2} \left| \int \int T(\mathbf{x}') \exp\left(i\pi \frac{|\mathbf{x} - \mathbf{x}'|^2}{\lambda z}\right) d\mathbf{x}' \right|^2, \quad (2)$$

where $T(\mathbf{x})$ represents the exit-surface wave of the light in the plane $z = 0$. For plane-wave illumination, under the projection approximation [see equation 2.39 of Paganin (2006)], $T(\mathbf{x})$ also represents the transmission function of the object.

[3] In this article we use the word 'speckle' loosely, to mean any localized diffraction feature recorded by the detector. Indeed, our method could just as well have been referred to as 'ptychographic X-ray *feature* tracking'.

Now let us suppose that, rather than plane-wave illumination, the object is illuminated by a wavefront with an arbitrary phase (ϕ) and amplitude ($w^{1/2}$) profile given by $p(\mathbf{x}, 0) = w^{1/2}(\mathbf{x}) \exp[i\phi(\mathbf{x})]$. The observed intensity is now given by

$$I(\mathbf{x}, z) = \frac{1}{(\lambda z)^2} \left| \int T(\mathbf{x}') p(\mathbf{x}', 0) \exp\left(i\pi \frac{|\mathbf{x} - \mathbf{x}'|^2}{\lambda z}\right) d\mathbf{x}' \right|^2. \quad (3)$$

For XST-based techniques, the challenge is to relate the image (I) to the reference (I_{ref}) via a geometric transformation. Here, the reference as defined in equation (3) represents an image of the sample, a distance z downstream of the sample plane, that is neither distorted by wavefront aberrations nor magnified by beam divergence. For the purposes of this section, I_{ref} could be a recorded image, but in subsequent sections we will see that this image can be estimated from a set of distorted images.

Note that at this point the mathematical description is rather general. For example, in the differential configuration of XST, $T(\mathbf{x})$ would represent the wavefront generated by the diffuser in the plane of the object and $p(\mathbf{x}, 0)$ would represent the transmission function of the object. In what follows, however, we will continue to describe $T(\mathbf{x})$ as the object or mask transmission function and $p(\mathbf{x}, z)$ as the X-ray beam profile (unmodulated by the object).

A common approach to this problem is outlined by Zanette *et al.* (2014). There, ϕ is expanded to first order, and $w^{1/2}$ to zeroth order, in a Taylor series about the point $\mathbf{x}$:

$$\phi(\mathbf{x}') = \phi(\mathbf{x}) + (\mathbf{x}' - \mathbf{x}) \cdot \nabla\phi(\mathbf{x}) + \phi_{\text{H}}(\mathbf{x}'), \quad (4)$$

$$w^{1/2}(\mathbf{x}') = w^{1/2}(\mathbf{x}) + w_{\text{H}}^{1/2}(\mathbf{x}'), \quad (5)$$

where $\phi_{\text{H}}(\mathbf{x}')$ and $w_{\text{H}}^{1/2}(\mathbf{x}')$ are the higher-order terms in the expansion. Now we have, for ϕ_{H} and $w_{\text{H}}^{1/2} \simeq 0$,

$$I(\mathbf{x}, z) \simeq \frac{w(\mathbf{x})}{\lambda z} \left| \int \int T(\mathbf{x}') \exp[i(\mathbf{x}' - \mathbf{x}) \cdot \nabla\phi(\mathbf{x})] \right.$$
$$\left. \times \exp\left(i\pi \frac{|\mathbf{x} - \mathbf{x}'|^2}{\lambda z}\right) d\mathbf{x}' \right|^2$$
$$= \frac{w(\mathbf{x})}{\lambda z} \left| \int \int T(\mathbf{x}') \exp\left[\frac{i\pi}{\lambda z} \left|\mathbf{x} - \mathbf{x}' - \frac{\lambda z}{2\pi}\nabla\phi(\mathbf{x})\right|^2\right] d\mathbf{x}' \right|^2$$
$$= w(\mathbf{x}) I_{\text{ref}}\left[\mathbf{x} - \frac{\lambda z}{2\pi}\nabla\phi(\mathbf{x}), z\right]. \quad (6)$$

This confirms the intuitive assumption that the local gradient of ϕ at each position along the sample is converted into a lateral displacement of the speckles observed in the reference. Equation (6) serves well in the limit where ϕ_{H} and $w_{\text{H}}^{1/2}$ approach 0 and is employed in a number of XST-based techniques. For example, in the UMPA approach [see equation (9) of Zdora (2018)] the governing equation is given by

$$I(\mathbf{x}, z) = w(\mathbf{x})\left(\bar{I}_{\text{r}} + D(\mathbf{x})\left\{I_{\text{ref}}\left[\mathbf{x} - \frac{\lambda z}{2\pi}\nabla\phi(\mathbf{x}), z\right] - \bar{I}_{\text{r}}\right\}\right), \quad (7)$$

where $\bar{I}_{\text{r}}$ is the mean intensity of the reference pattern and $D(\mathbf{x})$ is a term the authors refer to as the 'dark-field signal'. This term is related to a reduction in fringe visibility due to

fine features in $w(\mathbf{x})$ and, in fact, serves as an alternative contrast mechanism when solved for in addition to the phase gradients. Putting this term aside by setting $D = 1$, one can see that equation (7) reduces to equation (6).

Given the restrictive nature of the approximations employed, however, it is not surprising that equation (6) quickly fails to serve as a valid approximation for larger phase gradients. To see this, let us consider a well known analytical solution to I in terms of I_{ref} called the 'Fresnel scaling theorem', which is described in, for example, Appendix B of Paganin (2006). Simply put, it states that

The projected image of a thin scattering object from a point source of monochromatic light is equivalent to a magnified defocused image of the object illuminated by a point source of light infinitely far away.

The derivation is rather simple and so we shall present it here using the current notation. Let us say that the image, I, is formed by the point source of illumination a distance z_1 along the optical axis (the z axis) and that this distance is large enough that we can ignore intensity variations of the illumination across the sample surface, so that $w^{1/2}(\mathbf{x}) = 1$. The probing illumination in the plane of the sample is then given by $p(\mathbf{x}, 0) = \exp(i\pi\mathbf{x}^2/\lambda z_1)$. Substituting this into equation (3) and completing the square in the exponent, we have

$$I(\mathbf{x}, z) = \frac{1}{(\lambda z)^2}\left|\int\int T(\mathbf{x}') \exp\left(\frac{i\pi\mathbf{x}'^2}{\lambda z_1}\right) \exp\left(i\pi\frac{|\mathbf{x}-\mathbf{x}'|^2}{\lambda z}\right) d\mathbf{x}'\right|^2$$

$$= \frac{1}{(\lambda z)^2}\left|\int\int T(\mathbf{x}') \exp\left(\frac{i\pi}{\lambda}\frac{z_1+z}{zz_1}\left|\frac{z_1}{z_1+z}\mathbf{x}-\mathbf{x}'\right|^2\right) d\mathbf{x}'\right|^2$$

$$= \left(\frac{z_1}{z_1+z}\right)^2 I_{\mathrm{ref}}\left(\frac{z_1}{z_1+z}\mathbf{x}, \frac{zz_1}{z_1+z}\right)$$

$$= M^{-2} I_{\mathrm{ref}}(\mathbf{x}/M, z/M), \qquad (8)$$

where the geometric magnification factor $M = (z_1 + z)/z_1$ and z/M is the effective propagation distance ($\bar{z}$). But according to equation (6) we would have

$$I(\mathbf{x}, z) = I_{\mathrm{ref}}\left(\mathbf{x} - \frac{\lambda z}{2\pi}\frac{2\pi\mathbf{x}}{\lambda z_1}, z\right) \qquad (9)$$

$$= I_{\mathrm{ref}}\left(\frac{z_1 - z}{z_1}\mathbf{x}, z\right), \qquad (10)$$

with a geometric magnification factor $M' = z_1/(z_1 - z)$, in contradiction to the result from the Fresnel scaling theorem. As expected, the results agree in the limit $z_1 \rightarrow \infty$, i.e. in the limit where the phase gradient approaches 0. Current formulations for XST based on equation (6) (in the absolute configuration) are expected to perform badly when the effective source distance, z_1, approaches the propagation distance, z, or (in the differential configuration) when the sample transmission function departs significantly from the weak phase approximation.

In a notable departure from this approach, Paganin *et al.* (2018) have recently developed an alternative description of the speckle tracking approximation based on a 'geometric flow' equation:

$$I(\mathbf{x}, z) \simeq I_{\mathrm{ref}}(\mathbf{x}, z) - \frac{\lambda z}{2\pi}\nabla \cdot \left[I_{\mathrm{ref}}(\mathbf{x}, z)\nabla\phi(\mathbf{x})\right]. \qquad (11)$$

This approximation, which closely resembles the transport of intensity equation (Teague, 1983), has the remarkable property that ϕ may be determined analytically from a reference–image pair, thus permitting the rapid and simple processing of large tomographic data sets. This approach also assumes small and local distortions of the reference and is, therefore, ill suited as an approximation for larger phase gradients. For example, substituting the quadratic phase for a diverging wavefield, $\phi = \pi\mathbf{x}^2/\lambda z_1$, into equation (11) yields

$$I(\mathbf{x}, z) \simeq \frac{z_1 - z}{z_1}\left[I_{\mathrm{ref}}(\mathbf{x}, z) - \frac{z}{z_1 - z}\mathbf{x} \cdot \nabla I_{\mathrm{ref}}(x, z)\right]. \qquad (12)$$

This corresponds to a geometric magnification factor of $M'' = (z_1 - z)/(z_1 - 2z)$, once again in contradiction to the analytical result $M = (z_1 + z)/z_1$.

To see this more clearly, let us examine the exact result of equation (8) in the limit where $M \rightarrow 1$. First, we set $1/M = 1 + m$, so that $m \rightarrow 0$ as $M \rightarrow 1$. Then we expand $I_{\mathrm{ref}}(\mathbf{x}/M, z/M)$ to first order in a Taylor series about $\mathbf{x}$:

$$I(\mathbf{x}, z) = M^{-2}I_{\mathrm{ref}}(\mathbf{x} + m\mathbf{x}, z/M)$$
$$\simeq M^{-2}\left[I_{\mathrm{ref}}(\mathbf{x}, z/M) + m\mathbf{x} \cdot \nabla I_{\mathrm{ref}}(\mathbf{x}, z/M)\right]. \qquad (8a)$$

Comparing the above equation with equation (12), we have $m = -z/(z_1 - z)$. Solving for the geometric magnification factor yields $M'' = (z_1 - z)/(z_1 - 2z)$ as above.

Remarkably, with only a minor modification to the speckle tracking formula in equation (6), a second-order expansion of the phase term can be accommodated in the Fresnel integral, leading to the 'speckle tracking approximation':

$$I(\mathbf{x}, z) \simeq W(\mathbf{x})\, I_{\mathrm{ref}}\left[\mathbf{x} - \frac{\lambda z}{2\pi}\nabla\Phi(\mathbf{x}), \bar{z}\right], \qquad (13)$$

$$\left(\frac{\bar{z}}{z}\right)^2 w(\mathbf{x})\, I_{\mathrm{ref}}(\mathbf{x}, \bar{z}) \simeq I\left[\mathbf{x} + \frac{\lambda z}{2\pi}\nabla\phi(\mathbf{x}), z\right], \qquad (14)$$

where $\nabla\phi$ and $\nabla\Phi$ are the transverse gradients of the illuminating wavefield phase in the sample and image planes, respectively (without the influence of the object), and w and W are the intensity profiles of the illuminating wavefield in the reference and image planes, respectively. In Fig. 4 we show a diagram for a hypothetical PXST imaging experiment. This diagram shows the lens, focal, sample, reference and image planes. The reference would have been measured by plane-wave illumination in the plane indicated. A point that is not illustrated in the diagram is that both the image and the reference exhibit propagation effects, such as Fresnel fringes. We note, once again, that the speckle tracking approximation above applies to more imaging geometries/modalities than that displayed in Fig. 4.

Equations (13) and (14) are reciprocal statements of the same approximation and choosing between them is a matter of

convenience depending on the desired application. We note here that this approximation makes a distinction between the phase gradients in the sample and image planes, whereas it is common to assume that they are similar or related by a lateral scaling factor (magnification). This distinction is not important in cases where the separation between these two planes and the beam divergence is small, but becomes critical for highly magnified imaging geometries or long propagation distances. This approximation is not as strong as the 'stationary phase approximation' (Fedoryuk, 1971), which links coherent propagation theory with geometric optics, although the principles used to derive this result are similar. The derivation is straightforward and self-contained but lengthy, and may be found in Appendix A.

Equations (13) and (14) possess two beneficial properties for the current analysis: they relate the image and its reference via a geometric transformation and they are consistent with the Fresnel scaling theorem. In fact, the Fresnel scaling theorem is a special case of the above approximations when $w(\mathbf{x}) = 1$ and $\phi(\mathbf{x}) = \pi\mathbf{x}^2/\lambda z_1$. Evaluating equation (14) for these values of w and ϕ and using $\overline{z} = zz_1/(z + z_1)$ we have

$$I_{\mathrm{ref}}\left(\mathbf{x}, \frac{zz_1}{z + z_1}\right) = \left(\frac{z_1 + z}{z_1}\right)^2 I\left(\mathbf{x} + \frac{\lambda z}{2\pi}\frac{2\pi\mathbf{x}}{\lambda z_1}, z\right)$$
$$= \left(\frac{z_1 + z}{z_1}\right)^2 I\left(\mathbf{x}\frac{z_1 + z}{z_1}, z\right) \quad (15)$$

and so $I(\mathbf{x}, z) = \left(\frac{z_1 + z}{z_1}\right)^{-2} I_{\mathrm{ref}}\left(\mathbf{x}\frac{z_1}{z_1 + z}, \frac{zz_1}{z_1 + z}\right)$,

which yield the correct magnification and scaling factors, in agreement with equation (8). Similarly, we can evaluate equation (13) using

$$W(\mathbf{x}) = \left(\frac{\overline{z}}{z}\right)^2 \quad \text{and} \quad \Phi(\mathbf{x}) = \frac{\pi\mathbf{x}^2}{z_1 + z}, \quad (16)$$

where these values for the illumination's wavefront in the plane of the detector follow from the Fresnel approximation
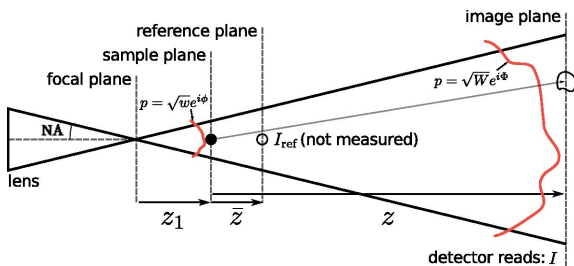
for a point source placed a distance $z + z_1$ upstream and from flux conservation of the beam when $w(\mathbf{x}) = 1$ in the sample plane.

Evaluating equation (13) yields

$$I(\mathbf{x}, z) = \left(\frac{z_1}{z + z_1}\right)^2 I_{\mathrm{ref}}\left(\mathbf{x} - \frac{\lambda z}{2\pi}\frac{2\pi\mathbf{x}}{z_1 + z}, \frac{zz_1}{z + z_1}\right)$$
$$= \left(\frac{z_1}{z + z_1}\right)^2 I_{\mathrm{ref}}\left(\mathbf{x}\frac{z_1}{z_1 + z}, \frac{zz_1}{z + z_1}\right), \quad (17)$$

which is, once again, in agreement with equation (8).

In general, for arbitrary ϕ, the phase curvature of the illumination may vary in direction, as is the case (for example) in an astigmatic lens system, and also with position in the image. Thus, the magnification is also position dependent and directional:

$$M_{\mathbf{v}}(\mathbf{x}) = \left[1 - \frac{\lambda z}{2\pi}\nabla_{\mathbf{v}}^2\Phi(\mathbf{x})\right]^{-1}, \quad (18)$$

where $\nabla_{\mathbf{v}}\Phi(\mathbf{x})$ is the directional derivative of $\Phi(\mathbf{x})$ along the unit normal vector $\mathbf{v}$.

Given the extended validity of equation (13), we suggest that the following modification to the UMPA equation [equation (7)] will achieve better results:

$$I\left[\mathbf{x} + \frac{\lambda z}{2\pi}\nabla\phi(\mathbf{x}), z\right] \simeq w(\mathbf{x})\{\overline{I}_{\mathrm{r}} + D(\mathbf{x})[I_{\mathrm{ref}}(\mathbf{x}, z) - \overline{I}_{\mathrm{r}}]\}, \quad (19)$$

where $\overline{I}_{\mathrm{ref}} = \langle I_{\mathrm{ref}}(\mathbf{x})\rangle_{\mathbf{x}}$, or, using the notation of Zdora (2018),

$$I(x - u_x, y - u_y) \simeq T(x, y)\{\overline{I}_0 + D(x, y)[I_0(x, y) - \overline{I}_0]\}. \quad (20)$$

We also note that, although Paganin et al.'s geometric flow algorithm [equation (11)] is a poor approximation for larger distortion factors (large M), it may be a more general physical description in the limit $M \to 1$. As the authors note, the term $\propto \nabla I_{\mathrm{ref}} \cdot \nabla\phi$ in the expansion of equation (11) accounts for speckle translations that arise from strong intensity gradients of the reference, i.e. that are not generated from $\nabla\phi$ alone.

Figure 4
Schematic diagram for a hypothetical projection imaging experiment. The illuminating beam propagates from left to right and the solid black lines indicate the boundaries of the illumination wavefront. The sample is depicted as a small black filled circle in the sample plane and as a black circle in the reference and image planes. The red lines depict the illumination's wavefront in the sample and image planes, which are not merely related by transverse magnification. The distorted shape of the circle in the image plane represents possible distortions of the speckle produced by the sample and the transverse phase gradients of the illumination.

4. Limits to the approximation

The second-order speckle tracking approximation of equations (13) and (14) is subject to the following approximations:

(1) $g(\mathbf{x}, \mathbf{x}') \simeq g[\mathbf{x}, \mathbf{u}(\mathbf{x})] + \frac{1}{2}\{\nabla^2\phi[\mathbf{u}(\mathbf{x})] + 2\pi/\lambda z\}|\mathbf{x}' - \mathbf{u}(\mathbf{x})|^2$,
(2) $w^{1/2}(\mathbf{x}') \simeq w^{1/2}[\mathbf{u}(\mathbf{x})]$,
(3) $z(\mathbf{x}) \simeq \overline{z} \equiv zz_1/(z + z_1)$,

where these are additional to the approximations necessary for the paraxial approximation to hold, $\mathbf{u}(\mathbf{x}) = \mathbf{x} - (\lambda z/2\pi) \times \nabla\Phi(\mathbf{x})$ and $g(\mathbf{x}, \mathbf{x}') \equiv \pi|\mathbf{x} - \mathbf{x}'|^2/\lambda z + \phi(\mathbf{x}')$. In general, these approximations hold best for smooth wavefront amplitudes $w^{1/2}$, predominantly quadratic phase ϕ and large spatial frequencies of the object.

Here, we examine the speckle tracking approximation, in one dimension, for the imaging geometry depicted in Fig. 4 and with parameters corresponding to a typical experiment

utilizing X-ray multilayer Laue lenses. For this example we choose that the illumination is formed by a lens with a hard-edged aperture and with the sample placed at two possible distances from the focal plane, $z_1 = 500$ and 10 µm. The lens has a numerical aperture of NA = 0.01 and the detector is placed in the far field of the probe and the sample, with $z_1 + z = 1$ m. This imaging geometry leads to an effective propagation for plane-wave illumination that is nearly identical to the distance from the focus to the sample ($\bar{z} \simeq z_1$). The wavelength is 10^{-9} m. The sample has a Gaussian profile so that $T(x) = 1 - n\exp(-x^2/2\sigma^2)$, where $n = 1 - i$ was chosen arbitrarily and would be proportional to the sample thickness and the deviation from unity of the refractive index and σ is the sample width, set to one of 0.15 or 0.01 µm below. The Fresnel number is thus $F \simeq \sigma^2/\lambda\bar{z}$.

The wavefronts in the sample and image planes were simulated using the discrete form of the Fresnel diffraction integral. The illumination's wavefront in the image plane is given by $p(x, z) = cW^{1/2}(x)\exp[i\Phi(x)]$, where c is a complex pre-factor that does not depend on x, $W^{1/2}(x)$ was calculated numerically and Φ is almost quadratic, with $\Phi(x) \simeq \pi x^2/[\lambda(z_1 + z)]$. Note that $\phi(x)$, the phase profile of the illumination in the sample plane, is not given by $\pi x^2/\lambda z_1$ as would be the case for a point source of light (i.e. for NA $\to \infty$). This is because the hard edges of the aperture produce Fresnel fringes that progress from the edge of the wavefront to the focal point at $x = 0$ as one moves from the image to the focal plane.

To test the validity of the speckle tracking approximation, we compare these simulated Fresnel images with those formed by evaluating equation (3). In this case equation (13) can be evaluated analytically with

$$W(x)I_{ref}\left[x - \frac{\lambda z}{2\pi}\nabla\Phi(x), \bar{z}\right] = W(x)\left|1 + n\frac{\sigma}{\sigma'}\exp\left(-\frac{x^2}{2M^2\sigma'^2}\right)\right|^2,$$

(21)

where

$$\sigma'^2 = \sigma^2 + i\frac{\lambda\bar{z}}{2\pi}$$

(22)

and

$$x - \frac{\lambda z}{2\pi}\nabla\Phi(x) = x\frac{z_1}{z_1 + z} = \frac{x}{M}.$$

(23)

In order to arrive at the above result, we have assumed that Φ is purely quadratic across the wavefront, but this approximation has not been used when simulating the image according to Fresnel diffraction theory.

In Appendix B, we suggest a suitable criterion for the speckle tracking approximation to hold for this imaging geometry based on the second criterion above,

$$\frac{(\lambda\bar{z})^{1/2} + \lambda\bar{z}q_T}{z_1\text{NA}} \ll 1,$$

(24)

where $q_T = 1/X$ is the spatial frequency corresponding to full period features of size X. This criterion holds for features within the plateau of the illumination profile.

In the first column of Fig. 5, we have placed the sample in the centre of the illumination profile. Here, the left-hand side of equation (24) evaluates to 0.8 and one can see that the fractional differences between the image and the approximation are small compared with that of the middle column. There, the sample has been shifted to the edge of the illumination profile, where the slope of the illumination amplitude is large. This leads to a breakdown of the second condition $\{w^{1/2}(x') \simeq w^{1/2}[u(x)]\}$, and indeed the discrepancy between the approximation and the image is largest near the edge of the pupil region and slowly reduces for features closer towards the central region.

In the right column of Fig. 5, the sample is smaller, with a σ value of 0.01 µm, and has been moved closer to the focal point. The left-hand side of equation (24) now evaluates to 11.0. As expected, this increase from 0.8 in the first column to 11.0 in the right column corresponds to an increasing discrepancy between the speckle tracking approximation and the image. This image is in the transition region between the near-field and far-field diffraction regimes. Clearly, features in the diffraction outside of the holographic region, where $W^{1/2} \simeq 0$, are not represented at all by the approximation.

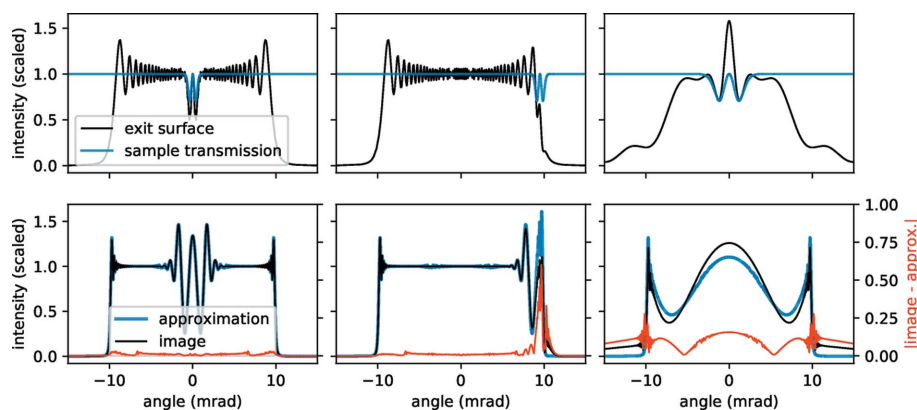In both the second and third examples shown here, the errors in the

Figure 5
Comparison between the images formed according to Fresnel diffraction theory and the speckle tracking approximation. In the left and middle columns, the sample has a σ width of 0.15 µm and is placed 500 µm from the focus. In the left column the sample is centred in the beam profile, whilst in the middle it has been shifted to the edge. In the right column the sample has a σ width of 0.01 µm and is placed 10 µm from the focus. First row: the exit-surface-wave intensities formed by illuminating a small Gaussian object with divergent illumination (black line). The intensities have been scaled by the factor $z/\bar{z}$. The sample transmission amplitudes are shown in blue. The angles along the x axis are given by $\arctan(x/z_1)$ and match those of the second row. Second row: the intensity of the wavefront in the image plane (black line) and the images formed by the speckle tracking approximation (blue line). The fractional differences are shown in red. The angles are given by $\arctan[x/(z_1 + z)]$.

speckle tracking approximation are dominated by the error in the approximation $w^{1/2}(x') \simeq w^{1/2}[u(x)]$. This is not surprising given that the zeroth-order expansion of $w^{1/2}(x')$ about $u(x)$ is a much stronger approximation than the second-order expansion of $\phi(x')$ about $u(x)$ (both approximations are necessary to arrive at the speckle tracking formula).

The increased quality of projection images due to smoother illumination profiles was one of the principle motivations behind Salditt and collaborators' efforts to develop an X-ray single-mode waveguide, in order to improve their tomo-holographic imaging methods [see for example Krenkel et al. (2017)].

5. Reconstruction algorithm

In this section we describe the steps necessary to recover estimates for $\Phi(\mathbf{x})$ and $I_{ref}(\mathbf{x})$ from a series of N measurements of the kind depicted in Fig. 3, where each recorded image on the detector corresponds to a translation of the sample in the transverse plane by $\Delta\mathbf{x}_n$ (here n is the image index). The refinement cycle is illustrated in Fig. 6. According to the speckle tracking approximation of equation (13), the geometric relationship between the recorded images $I_n(\mathbf{x})$ and the unrecorded reference $I_{ref}(\mathbf{x})$ is given by

$$I_n(\mathbf{x}) = W(\mathbf{x}) I_{ref}\left[\mathbf{x} - \frac{\lambda z}{2\pi}\nabla\Phi(\mathbf{x}) - \Delta\mathbf{x}_n, \overline{z}\right]. \quad (25)$$

Translating the sample by $\Delta\mathbf{x}_n$ along the x axis leads to a corresponding translation of the reference, because the convolution integral in equation (2) possesses translational equivariance.

To recover estimates for $\Phi(\mathbf{x})$ and $I_{ref}(\mathbf{x})$, we choose to minimize the target function

$$\varepsilon = \sum_n \int \varepsilon(n, \mathbf{x})\, dx$$
$$= \sum_n \int\int d\mathbf{x} \frac{1}{\sigma_I^2(\mathbf{x})}$$
$$\times \left\{ I_n(\mathbf{x}) - W(\mathbf{x}) I_{ref}\left[\mathbf{x} - \frac{\lambda z}{2\pi}\nabla\Phi(\mathbf{x}) - \Delta\mathbf{x}_n, \overline{z}\right] \right\}^2 \quad (26)$$

in an iterative update procedure with respect to $\nabla\Phi(\mathbf{x})$ and (as needed) $\Delta\mathbf{x}_n$, subject to

$$I_{ref}(\mathbf{x}, \overline{z})$$
$$= \frac{\sum_n w(\mathbf{x} + \Delta\mathbf{x}_n) I_n[\mathbf{x} + (\lambda z/2\pi)\nabla\phi(\mathbf{x} + \Delta\mathbf{x}_n) + \Delta\mathbf{x}_n, z]}{\sum_n w^2(\mathbf{x} + \Delta\mathbf{x}_n)}, \quad (27)$$

where $\sigma_I^2(\mathbf{x})$ is the variance of the recorded intensities at each detector pixel, such that $\sigma_I^2(\mathbf{x}) = \langle I_n^2(\mathbf{x}) - \langle I_n(\mathbf{x})\rangle_n^2 \rangle_n$. In fact equation (27) is the analytical solution for the minimum of $\sum_n \varepsilon(n, \mathbf{x})$ with respect to $I_{ref}(\mathbf{x})$ but for $\sigma_I^2(\mathbf{x}) = 1$. The reason we have set $\sigma_I^2(\mathbf{x}) = 1$ for the reference update is that, in this way, the reference is formed preferentially from parts of the image with larger intensities and thus will not be unduly affected by detector noise. This is also the update procedure that is often employed in single-mode ptychographic recon-

structions [see for example equation (7) of Thibault et al. (2009)].

The update for $\nabla\Phi(\mathbf{x})$ is given by

$$\nabla\Phi(\mathbf{x}) = \text{argmin}_{\nabla\Phi}\left[\sum_n \varepsilon(n, x)\right], \quad (28)$$

while holding $I_{ref}(\mathbf{x})$ and $\Delta\mathbf{x}_n$ constant. Here $\text{argmin}_{\nabla\Phi}$ means 'the argument of the minimum' with respect to $\nabla\Phi$, which is to say, the $\nabla\Phi$ that gives rise to the minimum of $\sum_n \varepsilon(n, x)$. The minimization is performed by evaluating $\sum_n \varepsilon(n, \mathbf{x})$ for possible value of $\nabla\Phi(\mathbf{x})$ within a pre-defined search window.

The update for $\Delta\mathbf{x}_n$ is given by

$$\Delta\mathbf{x}_n = \text{argmin}_{\Delta\mathbf{x}}\left[\int \varepsilon(n, \mathbf{x})\, dx\right], \quad (29)$$

while holding $I_{ref}(\mathbf{x})$ and $\nabla\Phi(\mathbf{x})$ constant. Once again, the minimization is performed by evaluating the possible value of $\Delta\mathbf{x}_n$ within a pre-defined search window.

Additionally, it is often desirable to regularize $\nabla\Phi(\mathbf{x})$ during the update procedure (especially for the first few iterations), according to
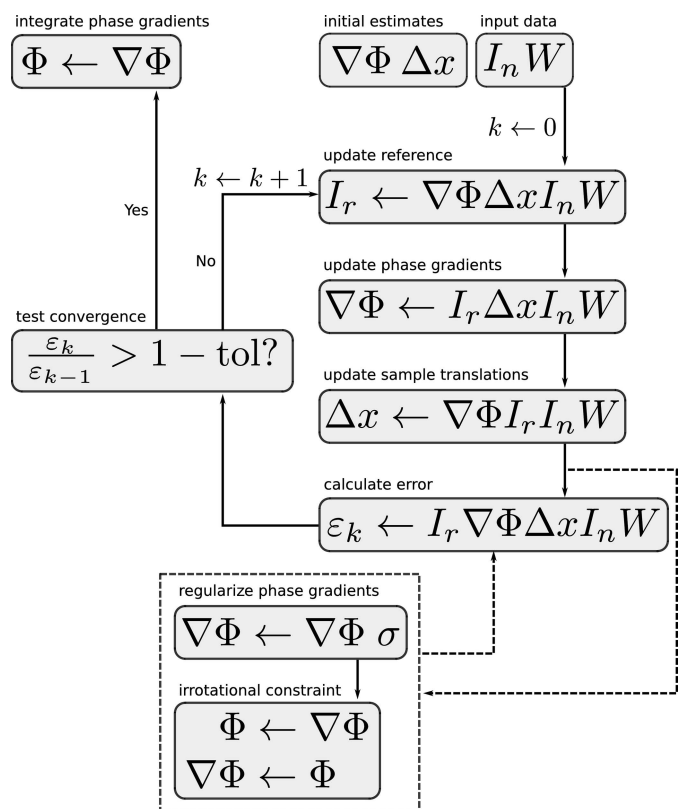

Figure 6
Flow diagram for the PXST iterative refinement cycle. A left arrow ($\leftarrow$) represents an update of the item on the left given the items to the right of the arrow. The dashed line arrows represent optional paths in the algorithm. Each step in the diagram corresponds to an equation in the main text: 'update reference' to equation (27), 'update phase gradients' to equation (28), 'update sample translation' to equation (29), 'calculate error' to equation (26), 'regularize phase gradients' to equation (30), and finally the 'irrotational constraint' and 'integrate phase gradients' steps to equation (31).

$$\nabla\Phi(\mathbf{x}) = \left[\frac{1}{2\pi\sigma^2}\exp\left(-\frac{\mathbf{x}^2}{2\sigma^2}\right)\right] \otimes \nabla\Phi(\mathbf{x}), \qquad (30)$$

where $\otimes$ is the convolution operator and σ is a regularization parameter that can be reduced as the iterations proceed.

Once the iterative procedure has converged, the phase profile of the illumination [$\Phi(\mathbf{x})$] can be recovered from the gradients [$\nabla\Phi(\mathbf{x})$] by numerical integration. For this we follow the method outlined in the supplementary section of Zanette *et al.* (2014). Let us label the final value of the phase gradients by $\delta(\mathbf{x}) \equiv \nabla\Phi(\mathbf{x})$. The procedure is then given by

$$\Phi(\mathbf{x}) = \text{argmin}_\Phi\left[\left|\delta(\mathbf{x}) - \nabla\Phi(\mathbf{x})\right|^2\right], \qquad (31)$$

where $\nabla\Phi(\mathbf{x})$ is evaluated numerically and the minimization is performed via the least-squares conjugate gradient method.

The fact that Φ is given by the numerical integration of $\nabla\Phi$ suggests a further constraint that could be employed in the update procedure. As noted by Paganin *et al.* (2018), $\nabla\Phi$ will be irrotational if Φ is continuous and single valued because $\nabla\Phi$ is given by the gradient of a scalar field. This follows from the Helmholz theorem, which states that any field can be written as the sum of a gradient and a curl. Since we know that $\nabla\Phi$ is, by definition, the gradient of Φ, then the curl must be zero: $\nabla \times \nabla\Phi = 0$. In the work of Paganin *et al.*, this condition is automatically satisfied by the solution. Here, however, we must incorporate this as a separate constraint. An irrotational field $\mathbf{f}$ is one that satisfies

$$\frac{\partial\mathbf{f}_y(\mathbf{x})}{\partial x} = \frac{\partial\mathbf{f}_x(\mathbf{x})}{\partial y}, \qquad (32)$$

where $\mathbf{f}_x(\mathbf{x})$ and $\mathbf{f}_y(\mathbf{x})$ are the x and y components of the vector field, respectively. To ensure that $\nabla\Phi$ is irrotational, one need only apply the numerical integration in equation (32) followed by numerical differentiation as needed during the update procedure. If this condition is not enforced, then the degree to which the recovered $\nabla\Phi$ is irrotational can be used as a measure of the fidelity of the result.

Numerical considerations for the implementation of this iterative update procedure, in addition to the source code developed to implement the PXST algorithm, have been published online (see https://github.com/andyofmelbourne/speckle-tracking).

The algorithm presented here is by no means the only approach to solve for the phase gradients and the reference. Indeed, similar problems emerge in many areas of imaging such as computer vision (Demirci *et al.*, 2006), medical imaging (Thirion, 1998) and military targeting applications (Kechagias-Stamatis *et al.*, 2018). In magnetic resonance imaging, the process of identifying the distortions that relate an image to its reference is often termed the 'image registration' problem and generating the reference from a set of distorted views is termed 'atlas construction'. 'Diffeomorphic image registration' algorithms are popular in that field, many of which are based on Thirion's demons and log-demons algorithm (Thirion, 1998; Lombaert *et al.*, 2014). This approach has been employed in the context of XST by Berto *et al.* (2017) to recover the phase gradients from an image/reference pair.

Others in the XST field use correlation-based approaches, where the geometric mapping between a small region of the distorted image and the reference is determined by the point which provides the greatest correlation coefficient (Zdora, 2018). The approach outlined in this work was employed because of its simplicity and ease of implementation. However, it seems likely (in the authors' view) that one or more of the other approaches mentioned above could be adapted to the current problem in order to produce superior results.

5.1. Example reconstruction

Here we provide a brief example of a PXST reconstruction from a simulated 1D data set. This example is not intended as a realistic simulation of an actual experiment: see Morgan *et al.* (2020) for experimental results in two dimensions. Rather, it serves as a simple illustrative check on the basic principles of PXST.

The simulated sample is similar to that shown in Fig. 2. It was constructed in Fourier space with a Gaussian intensity profile and random phases at each pixel. The real-space object is thus complex valued, so that rays passing through the sample will be both absorbed and deflected in angle. The intensity of the illumination profile, in the plane of the detector, was formed by setting W equal to a top hat function filtered with a Gaussian kernel. This filter produces a smooth tapered fall-off in the intensity near the edges of the beam that helps to avoid aliasing artefacts during numerical propagation of the wavefront. The phase profile, Φ, was constructed with the quadratic function $\pi x^2/[\lambda(z_1 + z)]$, where $\lambda = 1.2$ nm (1 keV), $z = 20$ mm and $z_1 = 40$ mm, so that the focal plane of
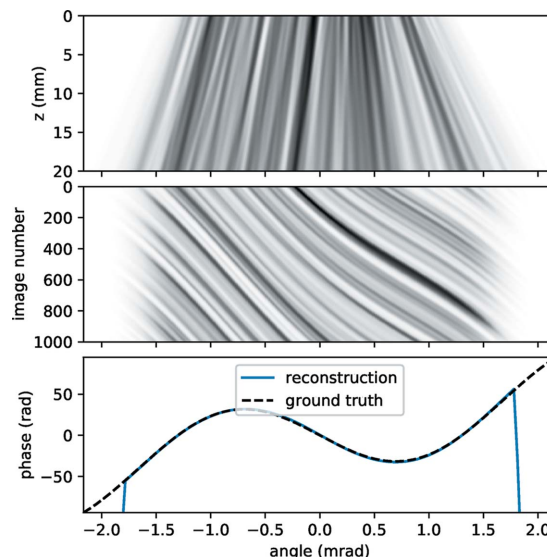
Figure 7
Top: intensity of the wavefront propagating from the sample ($z = 0$) to the detector plane ($z = 20$ mm). The linear colour scale ranges from 0 (white) to the maximum value (black). Middle: stack of the 1D images recorded as the sample is scanned across the wavefield (to the right). The colour scale is the same as in the top panel. Bottom: reconstructed and input phase aberrations in the detector plane. See text for further details.

the illumination is upstream of the sample in the top panel of Fig. 7 by a distance that is twice the sample-to-detector distance. This leads to an average magnification factor of 1.5. In addition to this, a sinusoidal phase profile was added to the phase in order to simulate the result of aberrations in the lens system; this can be seen as the dashed black line in the bottom panel of Fig. 7.

The intensity of the wavefront, $I(x, z)$, propagating from the exit surface of the sample to the detector plane is shown in the top panel. Upon close inspection, one can see that the intensities in the plane of the detector are non-trivially related to those in the exit surface of the sample. As such $I(x, 0)$ cannot be constructed from $I(x, z)$ by a scaling in x (magnification) or indeed by any geometric mapping. We make the point again that in PXST the 'reference' is not the sample transmission profile; rather, it is the intensity profile one would have observed on a detector placed a distance $\bar{z} = zz_1/(z_1 + z) = 13.3$ mm downstream of the sample illuminated by a plane wave. It is the geometric mapping between the reference (not the sample transmission) and the recorded images that is used to reconstruct the phase profile of the illumination.

The advantage of this 1D example is that one can visualize the entire data set in a single 2D image. In the middle panel of Fig. 7 the 1D images formed on the detector, as the sample is scanned across the wavefield, are displayed as an image stack. Along the vertical axis is the image number and the horizontal axis is in angle units, which are the angles made from the point source to each pixel in the image. It is seen that this image stack consists of a series of lines that appear to flow towards positive angles as the image number increases. These are the features in the image that can be obviously tracked through the stack. In this representation, the gradients of the lines at each diffraction angle are proportional the local wavefront curvature. For example, at a diffraction angle of ~ -0.75 mrad, the wavefront aberrations have a negative curvature and so features at this point in the wavefield are demagnified with respect to the mean. At a diffraction angle of ~ 0.75 mrad, the opposite is true (with a greater magnification) and the line gradients are shallow with respect to those at ~ -0.75 mrad. In addition to variations in the geometric magnification, the wavefront aberrations also locally adjust the effective propagation distance of the speckles. This is a non-geometric effect and (unlike the local variations in the magnification) is not accounted for by the PXST reconstruction algorithm. For the current example, we have deliberately set the aberrations such that the local magnification and effective propagation distance vary by a significant fraction across the wavefield. This allows for their effect to be clearly observed in the simulated data, but also leads to some errors in the phases.

The reconstructed phase profile, after 30 iterations of the PXST algorithm, is shown as the blue line in the bottom panel of Fig. 7. The constant, linear and quadratic components of the phase (or pedestal, tilt and defocus terms, respectively) have been removed prior to display, to allow the sinusoidal aberration profile to be clearly visualized. Near the edges of the illumination, $W \simeq 0$ and the phases could not be determined (as expected). Apart from this, the differences between the ground truth and reconstructed phase profile (0.1 rad r.m.s. error) are too small to see in this plot but are still much greater than the theoretical lower limit of ~ 0.0001 (this limit is defined in Appendix D) – owing to the strength of the aberrations (as described in the previous paragraph).

6. Discussion and conclusion

We have presented a modified form of the speckle tracking approximation, valid to second order in a local expansion of the phase term in the Fresnel integral. This result extends the validity of the speckle tracking approximation, thus allowing for greater variation of the unknown phase profile and for greater magnification factors when the wavefield has a high degree of divergence (such as that produced by a high-numerical-aperture lens system) or, when imaging a sample in the differential configuration of XST, allowing for greater phase variation across the transmission function of the sample (such as that produced by a thick specimen). We suggest that this approximation can be used, with little modification, in many of the existing XST applications and suggest such a modification for the UMPA approach.

We have also presented the PXST method, a wavefront metrology tool capable of dealing with highly divergent wavefields (like XSS), but unlike XSS, the resolution does not depend on the step size of the sample translations transverse to the beam. Coupled with a high-numerical-aperture lens, PXST provides access to nanoradian angular sensitivities as well as highly magnified views of the sample projection image. With a suitable scattering object, which in this case is the sample itself, a minimum of two images are required, although more images will improve robustness and resolution.

We must emphasize that it is only the projection image of the sample that is recovered. The phase and transmission profile of the sample must be inferred from the projection image via standard techniques (Wilkins et al., 2014). This is in contrast to other methods that provide multiple modes of imaging of the sample, such as the transmission, phase and 'dark-field' profiles. What distinguishes PXST from these methods is that the sample image is obtained in addition to the wavefield phase in the absolute configuration of XST: that is, both can be obtained from a single scan series of the sample.

A further application of this method is to use it as an efficient prior step to Fourier ptychography, by recording images out of focus. The recovered illumination and sample profiles can be used as initial estimates for a Fourier ptychographic reconstruction. Experimentally, this additional step can be achieved simply by moving the sample towards the focal plane of the lens. In some cases, this additional step would not even be required, so that speckle tracking followed by ptychography could be performed on the same data set.

For experimental results utilizing the PXST method, see Morgan et al. (2020). These results are based on a campaign of measurements for the development of a high-numerical-aperture wedged multi-layer Laue lens systems.

APPENDIX A
Derivation of the speckle tracking approximation

The Fresnel integral of equation (2) is often referred to as a point projection mapping. This is because, when the Fresnel number is $\gg 1$, the dominant contributions to the integral typically arise from values of the sample transmission, $T(x')$, around the point x (*i.e.* for $x' \simeq x$). At this point, the phase term $\pi(x - x')^2/\lambda z$ has a spatial frequency $q_F = (x - x')/\lambda z \simeq 0$. For x' far from x, the phase term causes the integrand to oscillate rapidly between $\pm T(x')$. If $T(x')$ is bandwidth limited, with a maximum spatial frequency of (say) q_{max}, then, for a sufficiently large $|x - x'|$, $q_F \geq q_{max}$ and successive oscillations of the integrand, caused by the phase term, will occur at roughly the same values of $T(x')$ and will thus cancel each other in the integration.

However, in the Fresnel integral of equation (3), the modulation of $T(x')$ by $p(x', 0) = w^{1/2}(x') \exp[i\phi(x')]$ has generated an additional phase term, and we would now expect the dominant contribution to $I(x, z)$ from $T(x')$ to arise at values of x' for which the integrand is smooth. To simplify the analysis, let us gather the phase terms of the Fresnel exponent and the incident illumination into a global phase factor

$$g(x, x') = \phi(x') + \frac{\pi}{\lambda z}(x - x')^2, \qquad (33)$$

so that the complex amplitude of the Fresnel integral becomes[4]

$$\psi(x, z) = \frac{\exp(2\pi i z/\lambda)}{(-i\lambda z)^{1/2}} \int T(x') \, p(x, x') \exp\left[\frac{i\pi}{\lambda z}(x - x')^2\right] dx' \qquad (34)$$

$$= \frac{\exp(2\pi i z/\lambda)}{(-i\lambda z)^{1/2}} \int T(x') \, w^{1/2}(x') \exp[ig(x, x')] \, dx'. \qquad (35)$$

Note that the global phase term $g(x, x')$, does not contain any contribution from the phase of the transmission function $T(x')$. Without any prior knowledge of this phase term, our smoothness condition becomes

$$\frac{\partial g(x, x')}{\partial x'} = \nabla\phi(x') - \frac{2\pi}{\lambda z}(x - x') = 0. \qquad (36)$$

For now, we will define u^{-1} as the solution to this equation for x given x', so that

$$x = u^{-1}(x') = x' + \frac{\lambda z}{2\pi}\nabla\phi(x'). \qquad (37)$$

u^{-1} is the functional inverse of u, which is yet to be defined. So, equation (3) now represents the point projection mapping $x' \to u^{-1}(x')$, rather than $x' \to x$.

This point is illustrated in Fig. 8 with $w^{1/2}(x') = 1$ for three different values of $\phi(x')$. On the vertical axis of each panel, we plot two real top-hat functions representing possible values for $T(x')$. In the two-dimensional domain of the integrand, $T(x')$ is

[4] For simplicity, the following analysis will be presented in one dimension. The difference between one and two dimensions is mostly in the normalization constants. At the end of this section we will generalize the result to two dimensions.
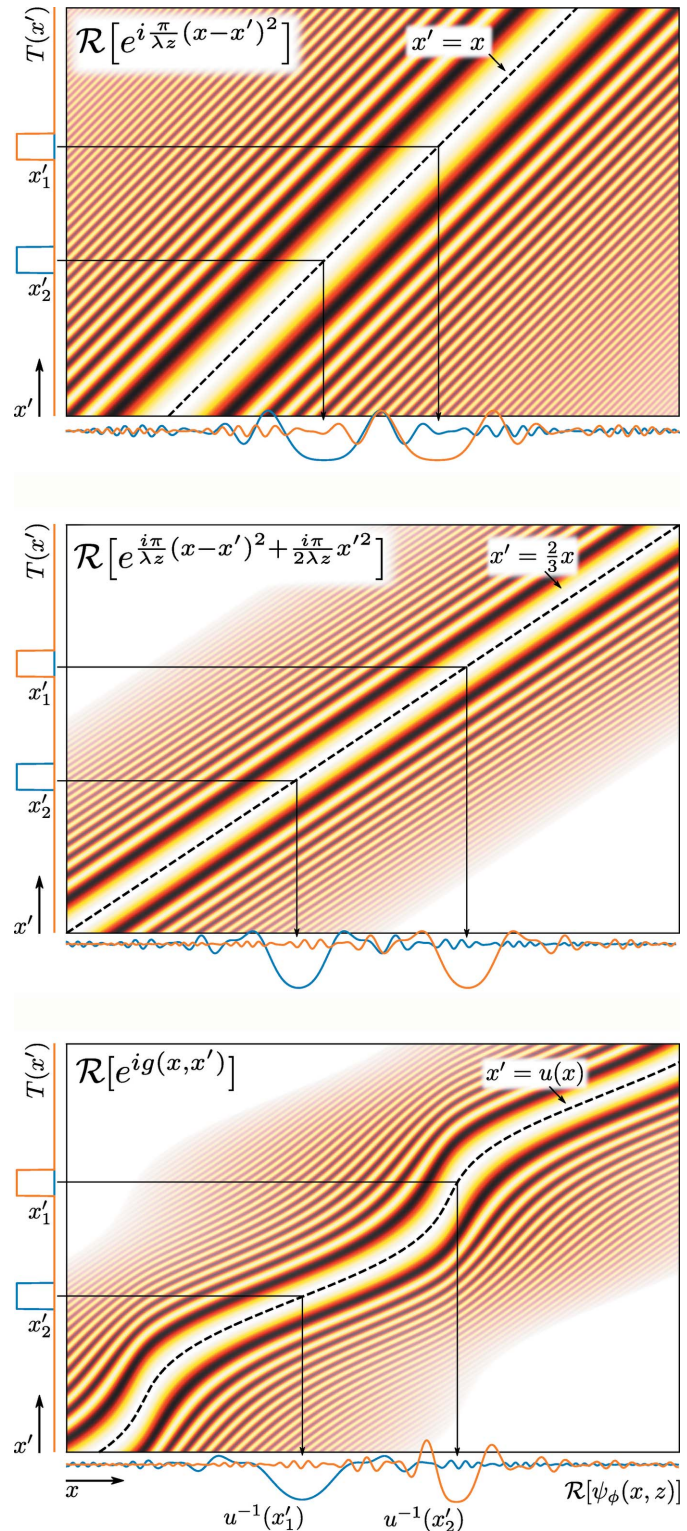

Figure 8
Illustration of the Fresnel integral, with a modulating phase term, for two top-hat functions. The colour maps display the real part of the exponential term, $\mathcal{R} \exp[g(x, x')]$, with the same colour scale as in Fig. 1. The alpha channel of the colour scale has been increased in regions where the spatial frequency of the exponent approaches the pixel size, so as to make transparent pixels that would otherwise be aliased. Phase terms that are constant with respect to x have been subtracted before display. These terms would not affect the final intensity of the image and removing them more clearly shows the line $x' = u(x)$ where $\partial g/\partial x' = 0$. See text for further details.

constant along the x axis. The Fresnel integral is performed by extruding $T(x')$ along the horizontal axis, multiplying by $\exp[ig(x, x')]$ and then integrating along the vertical axis. The real part of this integral is illustrated along the horizontal axis of each panel. We can see that the centroids of the features, before and after the Fresnel integral, follow the point projection mapping delineated by the dashed line $x' = u(x)$, which is defined by the condition $\partial g / \partial x' = 0$. In the top panel $\phi(x') = 0$, leading to $x' = u(x) = x$. This corresponds to Fresnel propagation with plane-wave illumination. Therefore, the separation between the top-hat functions in the sample plane is equal to the separation between the 'speckles' produced by each top-hat function. In the second panel $\phi(x') = \pi x'^2/(2\lambda z)$, corresponding to divergent illumination that would arise from a point source of illumination, or an ideal lens system with an infinite numerical aperture. Here $x' = u(x) = 2x/3$ and, consistent with the Fresnel scaling theorem, this leads to both a geometric magnification (the speckles are separated by a greater distance than the top hats) and a change in the effective propagation distance (which can be observed in the more rapid oscillation of the exponential term). In the final panel, a sinusoidal phase term has been added to the phase from the middle panel. Here $u(x)$ does not have a simple form and both the effective propagation distance and the magnification vary with position along the x axis.

This suggests the following modification to the approach outlined by Zanette and co-workers: instead of expanding $\phi(x')$ and $w^{1/2}(x')$ about x in equation (3), we should shift this expansion about the point $x' = u(x)$. In this way, the Taylor series expansion will be most accurate over the domain of the integrand that contributes most to the integral. The Nth-order Taylor series expansions of $g(x, x')$ and $w^{1/2}(x')$ about $x' = u(x)$ are given by

$$g_N(x, x') = \sum_{n=0}^{n=N} \frac{[x' - u(x)]^n}{n!} \left. \frac{\partial^n g}{\partial x'^n} \right|_{x, x'=u(x)}, \quad (38)$$

$$w_N^{1/2}(x') = \sum_{n=0}^{n=N} \frac{[x' - u(x)]^n}{n!} \left. \frac{\partial^n w}{\partial x'^n} \right|_{x'=u(x)}. \quad (39)$$

Evaluating equation (38) for $N = 1$ and equation (39) for $N = 0$, we have

$$g_1(x, x') = g[x, u(x)] + \left. \frac{\partial g}{\partial x'} \right|_{x, x'=u(x)} [x - u(x)] = g[x, u(x)], \quad (38a)$$

$$w_0^{1/2}(x') = w^{1/2}[u(x)], \quad (39a)$$

where the $n = 1$ term in the expansion of g is zero by construction. With $g \simeq g_1$ and $w^{1/2} \simeq w_0^{1/2}$, equation (34) becomes

$$\psi(x, z) \simeq \frac{\exp(2\pi i z/\lambda)}{(-i\lambda z)^{1/2}} w^{1/2}[u(x)] \exp\{ig[x, u(x)]\} \int T(x') \, \mathrm{d}x'. \quad (34a)$$

Unfortunately, the above expression completely fails to capture the physics upon which XST methods are based, *i.e.* the geometric mapping between I and I_{ref} defined by ϕ. In our attempt to improve the accuracy of the speckle tracking approximation, the first-order expansion of g about $x' = u(x)$ no longer depends on x'. Thus the integral over x' in equation (34) has reduced to the term $\int T(x') \, \mathrm{d}x'$.

With the above result in mind, let us try the following approach:

Instead of expanding $\phi(x')$ to first order and $w^{1/2}(x')$ to zeroth order about the point $x' = x$, expand $g(x, x')$ to second order and $w^{1/2}(x')$ to zeroth order about the point $x' = u(x)$.

This approach leads to

$$g_2(x, x') = g[x, u(x)] + \frac{1}{2} \left\{ \nabla^2 \phi[u(x)] + \frac{2\pi}{\lambda z} \right\} [x' - u(x)]^2, \quad (40)$$

where, once again, the $n = 1$ term for $g(x, x')$ is zero by construction in equation (37). Substituting equations (40) and $w^{1/2}(x') \simeq w^{1/2}[u(x)]$ into (35) then completing the square in the exponent we can recast the Fresnel integral in the following form:

$$\psi(x, z) \simeq \frac{\exp(2\pi i z/\lambda)}{(-i\lambda z)^{1/2}} w[u(x)] \exp\{ig[x, u(x)]\}$$
$$\times \int T(x') \exp\left\{ \frac{i\pi}{\lambda z[u(x)]} [x' - u(x)]^2 \right\} \mathrm{d}x', \quad (41)$$

where we have defined $z(x)$ as

$$z(x) \equiv \left[\frac{1}{z} + \frac{\lambda}{2\pi} \nabla^2 \phi(x) \right]^{-1}. \quad (42)$$

One can interpret $z(x)$ as the propagation distance required to locally reproduce the diffraction features in $\psi(x, z)$ had the illumination been plane wave [*i.e.* $\phi(x') = 1$].

We remind the reader that the 'ref' subscript refers to the wavefront that would have been formed with plane-wave illumination, with $p(x, 0) = 1$. Here we define ψ_{ref} as the complex amplitudes corresponding to the Fresnel integral in equation (2):

$$\psi_{\mathrm{ref}}(x, z) = \frac{\exp(2\pi i z/\lambda)}{(-i\lambda z)^{1/2}} \int T(x') \exp\left[\frac{i\pi}{\lambda z} (x - x')^2 \right] \mathrm{d}x'. \quad (43)$$

Now $\psi(x, z)$ can be related to $\psi_{\mathrm{ref}}(x, z)$ (where $I_{\mathrm{ref}} = |\psi_{\mathrm{ref}}|^2$) by the substitutions $x \to u(x)$ and $z \to z[u(x)]$, yielding

$$\psi(x, z) \simeq \exp(2\pi i\{z - z[u(x)]\}/\lambda)\{z[u(x)]/z\}^{1/2}$$
$$\times w[u(x)] \exp\{ig[x, u(x)]\}\psi_{\mathrm{ref}}\{u(x), z[u(x)]\}, \quad (44)$$

$$I(x, z) \simeq \{z[u(x)]/z\}w[u(x)] I_{\mathrm{ref}}\{u(x), z[u(x)]\}. \quad (45)$$

So far we have avoided a more explicit definition of the geometric mapping factor $u(x)$; it is currently defined by its inverse in equation (37). A more meaningful definition can be obtained by the following consideration. Setting $T(x') = 1$, equation (41) represents the propagation of the incident beam through free space in the absence of the sample, so that

$$p(x, z) \simeq \frac{\exp(2\pi i z/\lambda)}{(-i\lambda z)^{1/2}} w[u(x)] \exp\{ig[x, u(x)]\}$$

$$\times \int \exp\left\{\frac{i\pi}{\lambda z[u(x)]}[x' - u(x)]^2\right\} \mathrm{d}x'$$

$$= \exp(2\pi i z/\lambda)\{z[u(x)]/z\}^{1/2} w^{1/2}[u(x)] \exp\{ig[x, u(x)]\}$$

$$= \exp(2\pi i z/\lambda) W^{1/2}(x) \exp[i\Phi(x)]. \tag{46}$$

Here we have defined

$$W(x) \equiv \{z[u(x)]/z\}w[u(x)] \quad \text{and} \quad \Phi(x) \equiv g[x, u(x))] \tag{47}$$

and $W(x)$ and $\Phi(x)$ are, respectively, the intensity and phase profiles of the undisturbed beam in the plane of the detector.

The benefit of this calculation is that it provides an interpretation of the mapping function $u(x)$ in terms of the phase gradient of the illumination in the z plane. To see this, we first explicitly evaluate $\Phi(x)$ in terms of the incident phase profile $\phi(x)$. Substituting $x' = u(x)$ into equation (33), we have

$$\Phi(x) = \phi[u(x)] + \frac{\pi}{\lambda z}[x - u(x)]^2. \tag{48}$$

Using the definition for $u^{-1}(x)$ in equation (37), we can then evaluate

$$\Phi[u^{-1}(x)] = \phi(x) + \frac{\pi}{\lambda z}\left[x - x - \frac{\lambda z}{2\pi}\nabla\phi(x)\right]^2$$

$$= \phi(x) + \frac{\lambda z}{4\pi}[\nabla\phi(x)]^2. \tag{49}$$

Taking the derivative of both sides of equation (49) with respect to x yields

$$\frac{\partial u^{-1}(x)}{\partial x}\nabla\Phi[u^{-1}(x)] = \frac{\partial}{\partial x}\left\{\phi(x) + \frac{\lambda z}{4\pi}[\nabla\phi(x)]^2\right\}, \tag{50}$$

$$\left[1 + \frac{\lambda z}{2\pi}\nabla^2\phi(x)\right]\nabla\Phi[u^{-1}(x)] = \left[1 + \frac{\lambda z}{2\pi}\nabla^2\phi(x)\right]\nabla\phi(x), \tag{51}$$

$$\nabla\Phi[u^{-1}(x)] = \nabla\phi(x). \tag{52}$$

With this equality and the definition for $u^{-1}(x)$ in equation (37), one can now verify that the following equality holds:

$$x = u^{-1}(x) - \frac{\lambda z}{2\pi}\nabla\Phi[u^{-1}(x)]. \tag{53}$$

But since $u[u^{-1}(x)] = x$ we can identify $u(x)$ with

$$u(x) = x - \frac{\lambda z}{2\pi}\nabla\Phi(x). \tag{54}$$

Furthermore, we can evaluate $z[u(x)]$ in terms of $\Phi(x)$ by making use of the following relation:

$$\frac{\partial}{\partial x}u^{-1}(x) = \left[1 + \frac{\lambda z}{2\pi}\nabla^2\phi(x)\right] = \frac{z}{z(x)}, \tag{55}$$

where we have used the expression for $z(x)$ in equation (42). Then, taking the derivative with respect to x of both sides of equation (53),

$$\frac{\partial}{\partial x}x = 1 = \frac{\partial}{\partial x}\left\{u^{-1}(x) - \frac{\lambda z}{2\pi}\nabla\Phi[u^{-1}(x)]\right\}$$

$$= \frac{z}{z(x)} - \frac{z}{z(x)}\frac{\lambda z}{2\pi}\nabla^2\Phi[u^{-1}(x)], \tag{56}$$

and rearranging, we obtain

$$z_\Phi(x) \equiv z[u(x)] = z\left[1 - \frac{\lambda z}{2\pi}\nabla^2\Phi(x)\right]. \tag{57}$$

Inserting equations (47), (54) and (57) into equation (44) we can then write

$$\psi(x, z) \simeq \exp[-2\pi i z_\Phi(x)/\lambda]$$

$$\times p(x, z)\psi_{\mathrm{ref}}\left[x - \frac{\lambda z}{2\pi}\nabla\Phi(x), z_\Phi(x)\right] \tag{58}$$

or, the inverse relationship,

$$p(x, 0)\psi_{\mathrm{ref}}[x, z(x)] \simeq \exp\{-2\pi i[z - z(x)]/\lambda\}$$

$$\times \left[\frac{z}{z(x)}\right]^{1/2}\psi\left[x + \frac{\lambda z}{2\pi}\nabla\phi(x), z\right]. \tag{59}$$

This formulation of the projected image separates the effects of the geometric and propagation-based distortions induced in the detected image by phase variations in the incident illumination. The geometric distortions are captured by the term $(\lambda z/2\pi)\nabla\Phi(x)$ and the change in the fringe structure of a feature by the term $z_\Phi(x)$.

For the purposes of the current work, we will ignore variations in the fringing terms $z(x)$ and $z_\Phi(x)$ across the wavefield and use instead the constants

$$z_\Phi(x) \simeq \overline{z}_\Phi \equiv z\left[1 - \frac{\lambda z}{2\pi}\langle\nabla^2\Phi(x)\rangle_x\right], \tag{60}$$

$$z(x) \simeq \overline{z} \equiv \left[\frac{1}{z} - \frac{\lambda}{2\pi}\langle\nabla^2\phi(x)\rangle_x\right]^{-1}. \tag{61}$$

$\langle\nabla^2\phi(x)\rangle_x$ and $\langle\nabla^2\Phi(x)\rangle_x$ are the mean phase curvatures of the illumination in the plane of the sample and the detector, respectively. Under the Fresnel approximation, they can be defined in terms of the effective source distance from the sample or detector planes. If z_1 is the distance between the entrance surface of the sample and the effective source point, then

$$\langle\nabla^2\phi(x)\rangle_x = \frac{2\pi}{\lambda z_1} \quad \text{and} \quad \langle\nabla^2\Phi(x)\rangle_x = \frac{2\pi}{\lambda(z + z_1)}. \tag{62}$$

Using the above expressions, one can now verify that

$$\overline{z} = \overline{z}_\Phi = \frac{zz_1}{z + z_1}. \tag{63}$$

With the above approximations for $z(x)$ and $z_\Phi(x)$ and taking the mod square of equations (58) and (59), we arrive at the 1D speckle tracking approximation:

$$I(x, z) \simeq W(x) I_{\mathrm{ref}}\left[x - \frac{\lambda z}{2\pi}\nabla\Phi(x), \overline{z}\right], \tag{64}$$

$$w(x) I_{\text{ref}}(x, \overline{z}) \simeq \frac{z}{\overline{z}} I\left[x + \frac{\lambda z}{2\pi} \nabla\phi(x), z\right]. \quad (65)$$

In two dimensions, the constant prefactor to the Fresnel integral in equation (34) is $1/(-i\lambda z)$ [rather than $1/(-i\lambda z)^{1/2}$ in one dimension]. This leads to an altered expression for W:

$$W(\mathbf{x}) \equiv (\overline{z}/z)^2 w(\mathbf{x}). \quad (66)$$

Additionally, $\overline{z}$ is now defined by the average wavefront curvature over the 2D transverse plane:

$$\overline{z} \equiv z\left[1 - \frac{\lambda z}{2\pi}\frac{1}{2}\langle\nabla^2\Phi(\mathbf{x})\rangle_{\mathbf{x}}\right], \quad (67)$$

where the Laplacian operator is now also over the 2D plane, $\nabla^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2$. If one accepts the approximation $z(\mathbf{x}) \simeq \overline{z}$ from the beginning, then the analysis presented here in one dimension can be repeated in two by following the above steps, first along the x axis and then along the y axis. The result of this procedure is

$$I(\mathbf{x}, z) \simeq W(\mathbf{x}) I_{\text{ref}}\left[\mathbf{x} - \frac{\lambda z}{2\pi}\nabla\Phi(\mathbf{x}), \overline{z}\right], \quad (68)$$

$$w(\mathbf{x}) I_{\text{ref}}(\mathbf{x}, \overline{z}) \simeq \left(\frac{z}{\overline{z}}\right)^2 I\left[\mathbf{x} + \frac{\lambda z}{2\pi}\nabla\phi(\mathbf{x}), z\right]. \quad (69)$$

APPENDIX B
Derivation of the region of validity of the speckle tracking approximation

In Appendix A, it was shown that the Fresnel integral of $T(x)$ modulated by $p(x, 0)$,[5]

$$\psi(x, z) = \frac{\exp(2\pi i z/\lambda)}{(-i\lambda z)^{1/2}} \int p(x', 0) T(x') \exp\left[\frac{i\pi}{\lambda z}(x - x')^2\right] dx', \quad (70)$$

can be approximated by

$$\psi(x, z) \simeq \frac{p(x, z)}{(-i\lambda\overline{z})^{1/2}} \int T(x') \exp\left\{\frac{i\pi}{\lambda\overline{z}}[x' - u(x)]^2\right\} dx', \quad (71)$$

subject to the approximations

(1) $g(x, x') \simeq g[x, u(x)] + \frac{1}{2}\{\nabla^2\phi[u(x)] + 2\pi/\lambda z\}[x' - u(x)]^2$,
(2) $w^{1/2}(x') \simeq w^{1/2}[u(x)]$,
(3) $z(x) \simeq \overline{z} \equiv zz_1/(z + z_1)$,

where these are additional to the approximations necessary for the paraxial approximation to hold, $p(x, 0) = w^{1/2}(x) \times \exp[i\phi(x)]$, $p(x, z) = \exp(2\pi i z/\lambda)W^{1/2}(x)\exp[i\Phi(x)]$, $u(x) = x - (\lambda z/2\pi)\nabla\Phi(x)$, z_1 is the effective distance between the light source and the sample plane, and $g(x, x') \equiv \pi(x - x')^2/\lambda z + \phi(x')$.

---

[5] For simplicity, the following analysis will be carried out in one dimension. Generalization to two dimensions is not required to support the results of this section.

B1. Requirements

In this section we shall determine the requirements for each of these three approximations to hold.

B1.1. The first approximation. Let us assume for the moment that approximations 2 and 3 are valid. In this case, it is sufficient to require that the Taylor series expansion of $g(x, x')$ is valid within the interval of convergence of the Fresnel integral in equation (71).

In Appendix A, we have intuited that the expansion of $g(x, x')$ need only be valid for values of x' satisfying $q_g(x, x') < q_T$, where $q_g(x, x')$ are the spatial frequencies of g and q_T is the maximum spatial frequency of T. For values of x' outside of this region, successive oscillations of the integrand will occur at roughly the same magnitude $[\pm T(x')]$ with a net zero contribution to the integral.

So, let us first examine the domain of x' over which the expansion is valid. The Lagrange error bound for the Taylor series expansion of $g(x, x')$, about $x' = u(x)$, sets a limit on the magnitude of the residual:

$$|R_N(x'; g)| = |g(x, x') - g_N(x, x')|$$
$$\leq \left|\frac{M}{(N+1)!}[x' - u(x)]^{N+1}\right|, \quad (72)$$

where

$$\left|\frac{\partial^{N+1} g}{\partial x'^{N+1}}(x, x'')\right| \leq M \quad \text{for} \quad |x'' - u(x)| < |x' - u(x)|. \quad (73)$$

With the above expression, we can identify the interval $2d_\phi$ of x' about $u(x)$, such that $x': u(x) - d_\phi \rightarrow u(x) + d_\phi$, for which the magnitude of the residual $|R_N|$ is below some threshold level of tolerance. So, for $N = 2$,

$$\frac{\partial^3 g}{\partial x'^3}(x, x') = \phi^{(3)}(x') \quad (74)$$

and

$$\phi_{\max}^{(3)} \equiv M = \max\left[|\phi^{(3)}(x')| : |x' - u(x)| \leq d_\phi\right], \quad (75)$$

we have

$$E_2 = \frac{1}{6}\phi_{\max}^{(3)}d_\phi^3 < \frac{1}{6}E_{\text{tol}} \quad \text{or} \quad d_\phi < (E_{\text{tol}}/\phi_{\max}^{(3)})^{1/3}, \quad (76)$$

where E_2 is the Lagrange error bound for the second-order expansion of g and E_{tol} is the maximum tolerable error (which has been implicitly defined so as to absorb the factor of 6).

Now we determine the minimum interval $2d_{q_T}$ required for the Fresnel integral in equation (71) to converge near to its true value.

Let us expand $T(x')$ as a Fourier series, such that $T(x') = \int \hat{T}(q)\exp(2\pi i x'q) dq$, where $\hat{T}(q)$ is the complex amplitude of the qth full-period spatial frequency of $T(x')$ corresponding to features of extent $X = 1/q$.

By the superposition principle, we can select the highest spatial frequency of T such that $\hat{T}(q) = 0$ for $|q| > q_T$ and examine the radius of convergence of the integral in equation (71) for $T(x') \rightarrow \exp(2\pi i x'q_T)$:

$$\psi(x, z; q_T) \equiv p(x, z)$$

$$\times \int_{u(x)-d}^{u(x)+d} \exp(2\pi i x' q_T) \exp\left\{i\pi \frac{[x' - u(x)]^2}{\lambda \overline{z}}\right\} dx', \quad (77)$$

where d_{q_T} is given by the value of d for which the above integral has converged within an acceptable error margin. This integral can reduced to the following form:

$$\psi(x, z; q_T) = \cdots \int_{(2/\lambda\overline{z})^{1/2}(\lambda\overline{z}q_T - d)}^{(2/\lambda\overline{z})^{1/2}(\lambda\overline{z}q_T + d)} \exp\left[i(\pi/2)v^2\right] dv$$

$$= \cdots \left\{ E\left[(2/\lambda\overline{z})^{1/2}(\lambda\overline{z}q_T + d)\right] - E\left[(2/\lambda\overline{z})^{1/2}(\lambda\overline{z}q_T - d)\right]\right\}, \quad (78)$$

where the function E is known as the Euler or Cornu spiral and the '$\cdots$' represent terms that are constant with respect to the integration variable v. The Euler spiral can be constructed in the complex plane in terms of the Fresnel integrals C and S:

$$E(x) = C(x) + iS(x) \quad (79)$$

$$= \int_0^x \cos[(\pi/2)x^2] dx + i \int_0^x \sin[(\pi/2)x^2] dx. \quad (80)$$

As both C and S approach their limit of $\frac{1}{2}$ as $x \to \infty$ with equal rapidity, we shall examine only the imaginary part of E for convenience.

In Fig. 9 we plot the imaginary part of the indefinite integral of equation (78) (shown in green in the second row) as a function of d. We can see here that for $d \gtrsim \lambda\overline{z}q_T$ the integral oscillates about 1 and that the amplitude of these oscillations is dominated by the $S[(2/\lambda\overline{z})^{1/2}(\lambda\overline{z}q_T - d)]$ term (shown in orange).
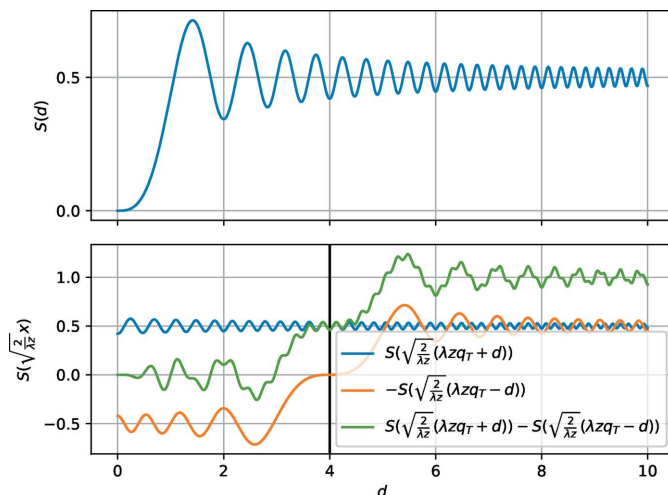
The extrema of $S(x)$ are given by

Figure 9
First row: the Fresnel integral S. Second row: the terms $\propto \psi(x, z; q_T)$ in equation (78). For these plots, $\lambda\overline{z} = 2$ and the vertical black line is at $d = \lambda\overline{z}q_T$.

$$\frac{\partial S(x)}{\partial x} = \sin[(\pi/2)x^2] = 0, \quad (81)$$

and therefore

$$x_m = \pm(2m)^{1/2}. \quad (82)$$

The extrema of $S[(2/\lambda\overline{z})^{1/2}(\lambda\overline{z}q_T - d)]$ are therefore located at

$$d_m = (m\lambda\overline{z})^{1/2} + \lambda\overline{z}q_T \quad \text{for} \quad m = 0, 1, 2, \ldots. \quad (83)$$

So, the residual for the integral in equation (78) is proportional to $1/2 - S[(2m)^{1/2}]$. Therefore, the integer m will serve as a measure of convergence for the integral and we can require that $d_{q_T} > d_m$. Finally, combining equations (76) and (83) we have, for $d_{q_T} < d_\phi$,

$$(m\lambda\overline{z})^{1/2} + \lambda\overline{z}q_T < \left(E_{\text{tol}}/\phi_{\max}^{(3)}\right)^{1/3}. \quad (84)$$

The above inequality can be seen as a limit for the phase variation $\phi_{\max}^{(3)}$ or, for a given $\phi_{\max}^{(3)}$, as a limit on the smallest features that will be resolved ($q_T = 1/X$) according to the approximation in equation (71). Setting, $m^{1/2} = 1$, multiplying both sides of equation (84) by $\phi_{\max}^{(3)}$ and raising both sides by the third power, we have the condition

$$\left[(\lambda\overline{z})^{1/2} + \lambda\overline{z}q_T\right]^3 / \phi_{\max}^{(3)} \ll 1. \quad (85)$$

B1.2. Approximation 2. Assuming for now that approximations 1 and 3 are valid, we require that $w^{1/2}(x') \simeq w^{1/2}[u(x)]$ is valid within the interval $|x' - u(x)| < d_w$, and that this interval is larger that d_{q_T}, which is the interval about $x' = u(x)$ over which the integral in equation (71) will converge. Making use, once again, of the Lagrange error bound we have

$$d_w < \frac{E_{\text{tol}}}{(w^{1/2})_{\max}^{(1)}}. \quad (86)$$

The requirement $d_w > d_{q_T}$ leads to

$$\frac{(\lambda\overline{z})^{1/2} + \lambda\overline{z}q_T}{(w^{1/2})_{\max}^{(1)}} \ll 1. \quad (87)$$

B1.3. Approximation 3. Assuming that approximations 1 and 2 are valid, the governing equation becomes

$$\psi[u^{-1}(x), z] \simeq \frac{p[u^{-1}(x), z]}{(-i\lambda\overline{z})^{1/2}} \int T(x') \exp\left[\frac{i\pi}{\lambda z_\Phi(x)}(x' - x)^2\right] dx', \quad (88)$$

where we have used $u^{-1}(x) = x + (\lambda z/2\pi)\nabla\phi(x)$ and restored $z_\Phi(x)$ in place of $\overline{z}$ in equation (71). This approximation, that $z(x) \simeq z_\Phi(x) \simeq \overline{z}$, is thus valid in the limit where the residual term

$$R = \frac{p[u^{-1}(x), z]}{(-i\lambda\overline{z})^{1/2}} \int T(x')\left\{\exp\left[\frac{i\pi}{\lambda\overline{z}}(x' - x)^2\right]\right.$$

$$\left. - \exp\left[\frac{i\pi}{\lambda z_\Phi(x)}(x' - x)^2\right]\right\} dx' \quad (89)$$

$$= \psi[u^{-1}(x), \overline{z}] - \psi[u^{-1}(x), z_\Phi(x)] \quad (90)$$

approaches zero. The conditions under which $\psi(x, z_1) \simeq \psi(x, z_2)$, in the Fresnel diffraction regime, are well known: the requirement is that the Fresnel number $F = X^2/\lambda|z_2 - z_1| \gg 1$. In this case, $|z_1 - z_2| = |z_\Phi(x) - \overline{z}|$, which depends on x. To generalize this requirement across the entire wavefront, we thus require that

$$X^2/\lambda\sigma(z) \gg 1 \qquad (91)$$

or

$$F \gg \sigma(z)/\overline{z}, \qquad (92)$$

where $\sigma(z)$ is the standard deviation of the effective propagation distance given by $\sigma(z) \equiv \{\langle[z_\Phi(x) - \overline{z}]^2\rangle_x\}^{1/2}$ and $F = X^2/\lambda\overline{z}$ is the Fresnel number for features of size X propagating a distance $\overline{z}$. Under this condition, features of size X (after correcting for the geometric distortions) will produce the same image on the detector regardless of their transverse position along the wavefront.

This condition can be expressed as a constraint on the phase profile of the beam. Using the definitions for $z_\Phi(x)$ [equation (57)] and $\overline{z}$ [equation (61)] we have

$$z_\Phi(x) - \overline{z} = \frac{\lambda z^2}{2\pi}\left[\nabla^2\Phi(x) - \langle\nabla^2\Phi(x)\rangle_x\right], \qquad (93)$$

$$\sigma(z) = \frac{\lambda z^2}{2\pi}\left\{\left[\nabla^2\Phi(x) - \langle\nabla^2\Phi(x)\rangle_x\right]^2\right\}^{1/2} \qquad (94)$$

$$= \frac{\lambda z^2}{2\pi}\sigma(\Phi^{(2)}). \qquad (95)$$

Using the above equation we have that approximation 3 is valid in the limit

$$F \gg \frac{\lambda z^2}{2\pi\overline{z}}\sigma(\Phi^{(2)}). \qquad (96)$$

B2. Limit on the defocus for an ideal lens

The illumination formed by an ideal lens, with a hard-edged aperture, has a distinct form (see for example Fig. 4). Within the plateau of the wavefront the intensity oscillates about a mean value, while the phase profile is approximately quadratic: $\phi \simeq \pi x^2/\lambda z_1$. In this case, approximation 2, that $w^{1/2}(x') \simeq w^{1/2}[u(x)]$, is the most onerous of the three. In the above analysis, we had used the Lagrange error bound to estimate the maximum distance along the wavefront ($|x' - x|$) for which this approximation will hold. But for the present case, this estimate is over-bounded given its general nature. Instead, we propose that an acceptable condition for this approximation is that $w^{1/2}(x') \simeq w^{1/2}[u(x)]$ will remain valid for $|x' - x| < z_1\text{NA}$, where NA is the numerical aperture of the lens and $z_1\text{NA}$ is approximately equal to the half-width of the plateau. Thus we replace $d_w = E_{\text{tol}}/(w^{1/2})^{(1)}_{\text{max}}$ with $d_w = z_1\text{NA}$ and equation (87) becomes

$$\frac{(m\lambda\overline{z})^{1/2} + \lambda\overline{z}q_T}{z_1\,\text{NA}} \ll 1, \qquad (97)$$

where this condition applies for features within the plateau of the illumination.

APPENDIX C
Uniqueness

C1. Pedestal and tilt terms are unconstrained in the illumination phase profile

In general, the solution to the target function in equation (26) does not constrain terms proportional to 1, x and y in the recovered phase profile. These terms are sometimes referred to as the 'pedestal' and 'tilt' components of the pupil function. To see this, consider a phase profile $\Phi' = c + \mathbf{d}\cdot\mathbf{x} + \Phi$, where Φ corresponds to the true phase profile in the plane of the detector and where c and $\mathbf{d}$ are constants. This leads to the phase gradients $\nabla\Phi' = \mathbf{d} + \nabla\Phi$. Substitution into equation (25) yields

$$I'_n(\mathbf{x}) = W(\mathbf{x})\,I_{\text{ref}}\left[\mathbf{x} - \frac{\lambda z}{2\pi}\nabla\Phi'(\mathbf{x}) - \Delta\mathbf{x}_n, \overline{z}\right] \qquad (98)$$

$$= W(\mathbf{x})\,I'_{\text{r}}\left[\mathbf{x} - \frac{\lambda z}{2\pi}\nabla\Phi(\mathbf{x}) - \Delta\mathbf{x}_n, \overline{z}\right] \qquad (99)$$

$$= I_n(\mathbf{x}) \quad \text{for} \quad I'_{\text{r}}(\mathbf{x}, \overline{z}) = I_{\text{ref}}\left(\mathbf{x} - \frac{\lambda z}{2\pi}\mathbf{d}, \overline{z}\right), \qquad (100)$$

independent of the pedestal term. Also, the tilt in the phase profile has produced a shift in the reference, which is generally not detectable unless the position of a feature in the object is known *a priori* with respect to the detector.

Alternatively, we could have absorbed the term $(\lambda z/2\pi)\mathbf{d}$ as a constant offset in the sample translation vectors $\Delta\mathbf{x}'_n = \Delta\mathbf{x}_n + (\lambda z/2\pi)\mathbf{d}$. We note that the tilt terms are typically unconstrained in speckle tracking techniques, since it is common to allow for an overall offset in the sample or detector positions.

C2. Speckle patterns of sufficient density are necessary, but not sufficient, for a unique solution to exist

Clearly, in the extreme case where no speckles are recorded the phase is completely unconstrained, so that

$$I'_n(\mathbf{x}) = W(\mathbf{x})\,I_{\text{ref}}\left[\mathbf{x} - \frac{\lambda z}{2\pi}\nabla\Phi'(\mathbf{x}) - \Delta\mathbf{x}_n, \overline{z}\right] \qquad (101)$$

$$= I_n(\mathbf{x}) \quad \text{for} \quad I_{\text{ref}}(\mathbf{x}, \overline{z}) = 1 \qquad (102)$$

and for any Φ'. This condition could also be reached in the limit where the fringe visibility of the speckle pattern approaches zero. The requirement for adequate fringe visibility is a point which is emphasized by Zanette *et al.* (2014) as well as many others in the field (Zdora, 2018). Of course, the above condition could also be reached for any sub-domain of $\mathbf{x}$. If, for example, $I_n(\mathbf{x}') = W(\mathbf{x}')$ for all n, then the phase terms are unconstrained at the points $\mathbf{x}'$. This suggests a more general (necessary) condition for a unique solution to exist: that a speckle of sufficient contrast must be observed at least once at each position in the image. As an example, the Hartmann sensor discussed in Section 2 does not satisfy this

condition and, as such, it is necessary to interpolate values of Φ between the mask holes, rendering the method insensitive to high-order aberrations that lead to rapid variations in $\Delta\Phi$. Note that this is not always an issue, especially in cases where the low-order aberrations of the wavefront are of primary concern, such as when we wish to correct them by some means, or when the low-order aberrations are dominant and dominate the imaging performance of the optic.

In another extreme, the phase is also unconstrained for $N = 1$, when only a single image has been recorded. This is because the unknown I_{ref} can be adjusted to accommodate any Φ':

$$I_0'(\mathbf{x}) = W(\mathbf{x})\, I_{\text{ref}}\left[\mathbf{x} - \frac{\lambda z}{2\pi}\nabla\Phi'(\mathbf{x}) - \Delta\mathbf{x}_0, \overline{z}\right] \quad (103)$$

$$= I_0(\mathbf{x}) \text{ for } I_{\text{ref}}(\mathbf{x}, \overline{z}) = \left(\frac{I_0'}{W}\right)\left[\mathbf{x} + \frac{\lambda z}{2\pi}\nabla\phi'(\mathbf{x}) + \Delta\mathbf{x}_0\right] \quad (104)$$

and for any Φ'. This situation has arisen because the observed speckles are modelled as a function of $\nabla\Phi$ and I_{ref}, both of which are refined in the PXST method. So if a given speckle is observed only once, at a location $\mathbf{x}'$, and the phase gradient at $\mathbf{x}'$ is not otherwise constrained, then multiple solutions for $\nabla\Phi$ and I_{ref} exist. The above two considerations suggest the more general (necessary) conditions for a unique solution to exist:

A speckle of sufficient contrast must be observed at least once at each position in the image, and this speckle must be observed at least twice and at different positions in the image.

Note that the above constraint does not require that every speckle must be observed more than once.

This is the 'speckle density condition' alluded to in the title of this section. The easiest way to satisfy this condition is to use a sample that produces a dense high-contrast array of speckles on the detector, such as a diffuser. With such a sample, the above constraint may be satisfied with just two images (*i.e.* for $N \geq 2$), provided that the sample step size is not greater than half the illuminated region of the sample along the direction of the step.

However, it is possible for a unique solution to exist even when the sample produces only a single observable speckle in each image. In this case, the above condition can be satisfied by scanning the sample such that this speckle is observed at each point in the image. This is a far less efficient means for wavefront sensing than using a diffuser. But this generality allows for nearly any object, such as the Siemens star in Fig. 3 or a Hartmann mask, to be used as a wavefront sensing device.

C3. Ambiguities can arise from unknown sample positions

If the sample translation vectors ($\Delta\mathbf{x}_n$) are unknown, then there exists a family of solutions to equation (25), with each solution corresponding to a set of translation vectors related by an affine transformation. Consider a set of translation vectors $\Delta\mathbf{x}_n' = \mathbf{x}_0 + \mathbf{A} \cdot \Delta\mathbf{x}_n$, where $\mathbf{x}_0$ is an overall offset and the dot product is between the 2×2 linear transformation matrix $\mathbf{A}$ and the true sample translation vectors. As described previously, any overall offset in the translation vectors

generates a corresponding offset in the reference and a tilt term in the recovered phases. So, neglecting the offset term, we generate this family of solutions by the substitution $\Delta\mathbf{x}_n = \mathbf{A}^{-1} \cdot \Delta\mathbf{x}_n'$ into equation (25):

$$I_n(\mathbf{x}) = W(\mathbf{x})\, I_{\text{ref}}\left[\mathbf{x} - \frac{\lambda z}{2\pi}\nabla\Phi(\mathbf{x}) - \mathbf{A}^{-1} \cdot \Delta\mathbf{x}_n', \overline{z}\right] \quad (105)$$

$$= W(\mathbf{x})\, I_{\text{ref}}\left(\mathbf{A}^{-1} \cdot \left\{\mathbf{A} \cdot \left[\mathbf{x} - \frac{\lambda z}{2\pi}\nabla\Phi(\mathbf{x})\right] - \Delta\mathbf{x}_n'\right\}, \overline{z}\right) \quad (106)$$

$$= W(\mathbf{x})\, I_{\text{ref}}'\left[\mathbf{x} - \frac{\lambda z}{2\pi}\nabla\Phi'(\mathbf{x}) - \Delta\mathbf{x}_n', \overline{z}\right], \quad (107)$$

where

$$\Delta\mathbf{x}_n' = \mathbf{A} \cdot \Delta\mathbf{x}_n, \quad (108)$$

$$I_{\text{ref}}'(\mathbf{x}, \overline{z}) = I_{\text{ref}}(\mathbf{A}^{-1} \cdot \mathbf{x}, \overline{z}) \quad (109)$$

and

$$\nabla\Phi'(\mathbf{x}) = \mathbf{A} \cdot \nabla\Phi(\mathbf{x}) + \frac{2\pi}{\lambda z}(\mathbf{x} - \mathbf{A} \cdot \mathbf{x}). \quad (110)$$

If, on the other hand, the true sample translation vectors are given by the input values of $\Delta\mathbf{x}_n$, but with small random offsets of mean 0, then the true solution for the phase and reference can be recovered from the retrieved values by removing the effect of any affine transformation that may have arisen during the reconstruction. This can be accomplished by minimizing $\sum_n |\Delta\mathbf{x}_n^{\text{out}} - \mathbf{A} \cdot \Delta\mathbf{x}_n^{\text{in}}|^2$ with respect to $\mathbf{A}$, where $\Delta\mathbf{x}_n^{\text{in}}$ and $\Delta\mathbf{x}_n^{\text{out}}$ are the input and output values of $\Delta\mathbf{x}_n$, respectively, then generating the corresponding solutions for $\nabla\Phi'$ and I_{r}' from the above equations. This situation can arise, for example, from small relative errors in the translation of a stepper motor or from the pointing jitter of an X-ray free-electron laser pulse.

C4. An unknown rotation of the sample stage axes with respect to the detector axes can be corrected

A common systematic error for the input sample positions is an overall rotation of the axes of the sample translation stages with respect to the pixel axes of the detector. In this case the linear transformation matrix reduces to the rotation matrix:

$$\mathbf{A} \rightarrow \mathbf{R}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \quad (111)$$

Here, we can make use of the fact that in general $\nabla\Phi'$ of equation (110) is not irrotational for $\theta \neq 0$ and $\mathbf{A} = \mathbf{R}(\theta)$. If $(u, v) \equiv \nabla\Phi^{\text{out}}$ then we can require that the vector field $\mathbf{R}^{-1} \cdot (u, v)$ be irrotational. With $\mathbf{R}^{-1} = \mathbf{R}(-\theta)$ and equation (32), we have

$$\theta = \text{argmin}_\theta\left\{\left[\frac{\partial}{\partial y}(u\cos\theta + v\sin\theta) - \frac{\partial}{\partial x}(-u\sin\theta + v\cos\theta)\right]^2\right\}, \quad (112)$$

where the derivatives with respect to x and y are evaluated numerically. However, if the irrotational constraint was enforced during the reconstruction, then the recovered $\Delta\mathbf{x}_n^{\text{out}}$

is free of the erroneous rotation and no further analysis is required.

C5. The raster grid pathology produces artefacts for lattice-like sample translations

A common problem encountered in ptychography is the 'raster grid pathology' (Thibault *et al.*, 2009). The raster grid pathology arises when reconstructing both the illumination and sample profiles from diffraction data acquired while the sample is scanned along a regular grid. In that case, the recovered illumination and sample transmission functions may be modulated by any function, so long as it is periodic on a lattice of points upon which all of the sample positions lie.

In many cases, the governing equation for a ptychographic reconstruction is given by

$$I_n(\mathbf{q}) = \left| \mathcal{F}\left[T(\mathbf{x} - \Delta\mathbf{x}_n)p(\mathbf{x}, 0) \right] \right|^2, \tag{113}$$

where $\mathcal{F}[\cdot]$ is the Fourier transformation operator over the transverse plane and represents the propagation of the exit-surface wavefront $\psi_n(\mathbf{x}) \equiv T(\mathbf{x} - \Delta\mathbf{x}_n)p(\mathbf{x}, 0)$ to the detector (in the far field of the sample). If the sample is translated along a regular grid, for example, with step size $\mathbf{d}$, then $\Delta\mathbf{x}_n = \mathbf{n} \cdot \mathbf{d}$, where the vector $\mathbf{n} = (i_n, j_n)$ is the 2D lattice index corresponding to the nth image (for integer i_n and j_n). If we make the substitution $p'(\mathbf{x}, 0) = f(\mathbf{x})p(\mathbf{x}, 0)$ into the above equation, then we have

$$\psi_n(\mathbf{x}) = T'(\mathbf{x} - \mathbf{n} \cdot \mathbf{d})f(\mathbf{x})p(\mathbf{x}, 0), \tag{114}$$

and hence

$$T'(\mathbf{x}) = \frac{\psi'_n(\mathbf{x} + \mathbf{n} \cdot \mathbf{d})}{p(\mathbf{x} + \mathbf{n} \cdot \mathbf{d}, 0)f(\mathbf{x} + \mathbf{n} \cdot \mathbf{d})} = \frac{T(\mathbf{x})}{f(\mathbf{x})} \tag{115}$$

if $f(\mathbf{x} \pm \mathbf{n} \cdot \mathbf{d}) = f(\mathbf{x})$ for all $\mathbf{n}$. The raster grid pathology can be avoided by ensuring that the sample scan positions lack any translational symmetry, *i.e.* by scanning the sample in non-regular patterns, for example in a spiral grid, or by adding a random offset to every grid position (Fannjiang, 2019). Given that equation (113) is just a special case of the Fresnel integral in equation (2) (from which the speckle tracking approximation is derived) it is natural to consider whether or not the same pathology applies here.

One can show that a similar pathology does indeed arise in the present case. Here, the illumination's intensity is constrained during the reconstruction, so instead we make the substitution $p'(\mathbf{x}, z) = p(\mathbf{x}, z)\exp[ig(\mathbf{x})]$, which is equivalent to $\Phi'(\mathbf{x}) = \Phi(\mathbf{x}) + g(\mathbf{x})$, into equation (25):

$$I_n(\mathbf{x}) = W(\mathbf{x})I'_r\left[\mathbf{x} - \frac{\lambda z}{2\pi}\nabla\Phi'(\mathbf{x}) - \mathbf{n} \cdot \mathbf{d}, \overline{z} \right] \tag{116}$$

$$= W(\mathbf{x})I'_r\left\{ \mathbf{x} - \frac{\lambda z}{2\pi}[\nabla\Phi(\mathbf{x}) + \nabla g(\mathbf{x})] - \mathbf{n} \cdot \mathbf{d}, \overline{z} \right\} \tag{117}$$

$$= W(\mathbf{x})I'_r\left\{ [\mathbf{x} - \nabla g(\mathbf{x} - \mathbf{n}\cdot\mathbf{d})] - \frac{\lambda z}{2\pi}\nabla\Phi(\mathbf{x}) - \mathbf{n}\cdot\mathbf{d}, \overline{z} \right\} \tag{118}$$

and hence

$$I'_r(\mathbf{x}) = I_{\text{ref}}\left[\mathbf{x} + \frac{\lambda z}{2\pi}\nabla g(\mathbf{x}) \right] \tag{119}$$

if $\nabla g(\mathbf{x}) = \nabla g(\mathbf{x} \pm \mathbf{n} \cdot \mathbf{d})$ for all $\mathbf{n}$. So, rather than modulating the reference with a periodic function, the pathology here creates a periodic geometric distortion of the reference.

APPENDIX D
Angular sensitivity and imaging resolution

D1. The resolution of the reference is given by the demagnified effective pixel size

As discussed in Section 3, the Fresnel scaling theorem states that the projection image of a thin sample formed by a point-like source of coherent light produces a magnified and defocused image of the sample. Similarly, in PXST, the 'reference' is an idealized image that would have been formed if the illumination were plane wave (*i.e.* with a flat phase profile), the detector were placed a distance $\overline{z}$ from the plane of the sample, the detector extended over the entire illuminated region of the sample and the physical pixel size (σ_{det}) were reduced by the magnification factor M. Assuming that the speckle tracking approximation holds, and that the aggregate signal-to-noise level is high, then the resolution of such an image is given by the demagnified effective pixel size of the detector (σ_{ref}).

The effective pixel size (σ_{eff}) can be much smaller than σ_{det} owing to sub-pixel interpolation, which is employed when registering a speckle across many images. We have found, as others have noted (Bérujon, Ziegler *et al.*, 2012), that sub-pixel interpolation can lead to a reduction in the effective pixel size by a factor of 10 or more depending on the point-spread function of the detector, the contrast of the speckles, the signal-to-noise ratio per image and the total number of images. On the other hand, effects such as the finite source size of the X-rays will tend to blur-out speckles and increase the effective pixel size.

Consider the imaging geometry depicted in Fig. 4, for an incoherent source of X-rays with 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. Then the image recorded in the detector plane will be given by

$$I(\mathbf{x}, z; \sigma_s) \simeq \left\{ \frac{1}{2\pi}\left(\frac{z_1}{zf\sigma_s} \right)^2 \exp\left[-\frac{\mathbf{x}^2}{2}\left(\frac{z_1}{zf\sigma_s} \right)^2 \right] \right\} \otimes I(\mathbf{x}, z), \tag{120}$$

where $\otimes$ is the convolution operator, f is the focal length of the lens and $I(\mathbf{x}, z)$ is the intensity of the wavefront in the plane of the detector for an on-axis point source of light. Therefore, the observed speckles will be broadened by a factor $\sigma_s zf/z_1$. In this regime, the effective pixel size is given by

$$\sigma_{\text{eff}} = \delta_{\text{pix}}(\sigma_{\text{det}} + \sigma_s zf/z_1), \tag{121}$$

$$\sigma_{\text{ref}} = \sigma_{\text{eff}}/M, \tag{122}$$

where we have used the symbol δ_{pix} to represent the fractional reduction in the effective pixel size due to numerical interpolation. For example, with a physical pixel size of $\sigma_{\mathrm{eff}} = 50\ \mu\mathrm{m}$, a fully coherent wavefield and $\delta_{\mathrm{pix}} = 1$, the demagnified pixel size for the example shown in Fig. 5 (left and middle columns) would be $\sigma_{\mathrm{ref}} = 25$ nm and for the right column (with the sample 10 μm from the focus) $\sigma_{\mathrm{ref}} = 5$ Å.

D2. The sample position that maximizes the angular sensitivity of the wavefront, in the plane of the sample, depends on the source coherence width

The smallest resolvable angular deviation of the wavefront, in the plane of the sample, is given by the arctangent of the smallest resolvable displacement of a speckle over the distance between the sample and the detector pixel array:

$$\Delta\Theta_\phi = \arctan(\sigma_{\mathrm{eff}}/z) \simeq \sigma_{\mathrm{eff}}/z, \qquad (123)$$

where we have employed the small-angle approximation in the last equality. The smallest resolvable increment in the phase gradient is thus $\Delta(\nabla\phi) = (2\pi/\lambda)\Delta\Theta_\phi$. We note that $\Delta\Theta_\phi$ is a lower bound on the angular sensitivity, since the achieved value will generally be larger than σ_{eff}/z because of detector noise and/or photon counting statistics.

For the imaging geometry depicted in Fig. 4, the optimal position for the detector will be given by the furthest distance from the focus such that the footprint of the illumination is contained within the pixel array, as this maximizes the sampling frequency of the wavefield. This raises the question, how far should one place the sample from the focus in order to maximize the angular resolution (minimize $\Delta\Theta$)? To answer this, let us keep the focus-to-detector distance (z_{t}) fixed, so that $z_{\mathrm{t}} = z_1 + z$, and minimize $\Delta\Theta_\phi$ with respect to z_1. Inserting $M = z_{\mathrm{t}}/z_1$ and equation (122) into equation (123) and minimizing yields

$$\Delta\Theta_\phi = \frac{\sigma_{\mathrm{det}}}{z_{\mathrm{t}}}\left[1 + \left(\frac{f\sigma_{\mathrm{s}}}{\sigma_{\mathrm{det}}}\right)^{1/2}\right]^2 \qquad (124)$$

for

$$z_1 = \frac{z_{\mathrm{t}}}{1 + (\sigma_{\mathrm{det}}/f\sigma_{\mathrm{s}})^{1/2}}. \qquad (125)$$

As the focus-to-sample distance is reduced, the magnification factor increases (improving the angular sensitivity), while at the same time the deleterious effects of the finite source size increase (deteriorating the angular sensitivity). The above value for z_1 represents the optimal compromise between these two effects.

D3. The sample position that maximizes the angular sensitivity of the wavefront, in the plane of the detector, is in the focal plane where the magnification is greatest

The angular resolution for the wavefield in the plane of the detector ($\Delta\Theta_\Phi$) is not, in general, the same as the angular resolution in the plane of the sample. This is because of the difference in extent between the effective pixel size and the

demagnified effective pixel size due to divergent illuminating wavefields. $\Delta\Theta_\Phi$ is given by

$$\Delta\Theta_\Phi = \sigma_{\mathrm{ref}}/z. \qquad (126)$$

For a fixed focus-to-detector distance, once again $z = z_{\mathrm{t}} - z_1$, and we have

$$\Delta\Theta_\Phi = \delta_{\mathrm{pix}}\left[(z_1/z_{\mathrm{t}}^2)\sigma_{\mathrm{det}} + (f/z_{\mathrm{t}})\sigma_{\mathrm{s}}\right]. \qquad (127)$$

An interesting feature of the above equation is that the effect of the source incoherence $\delta_{\mathrm{pix}}f\sigma_{\mathrm{s}}/z_{\mathrm{t}}$ does not vary with z_1; as z_1 is decreased the increase in the magnification factor leads to a corresponding decrease in the effective pixel size due to the finite source size, but this is exactly balanced by the reduction in the angle subtended by the demagnified effective pixel size. Therefore, the optimal position for the sample, in order to minimize $\Delta\Theta_\Phi$, is as close to the focal plane as possible. However, this can be a dangerous limit to approach, as the speckle tracking approximation will begin to break down as a result of the rapidly varying illuminating wavefield – at this point, it would be necessary to employ a fully coherent model for the wavefront propagation, for example, by switching to far-field ptychography. Additionally, it can be beneficial to increase the focus-to-sample distance for practical reasons; for example, increasing z_1 provides a larger field of view of the sample in each image which aids in positioning the region of interest of the sample with respect to the illuminating beam and, typically, increases the speckle visibility. For these reasons, it can be beneficial to approach the limit where

$$\sigma_{\mathrm{det}}/M \simeq \sigma_{\mathrm{s}} \quad \text{or} \quad z_1 \simeq z_{\mathrm{t}}\sigma_{\mathrm{s}}/\sigma_{\mathrm{det}}. \qquad (128)$$

Here z_1 has been increased until the demagnified pixel size is approximately equal to the demagnified feature size one would observe from a point-like object owing to incoherence alone. This represents the transition between modes where $\Delta\Theta_\Phi$ is dominated by the detector pixel size (larger z_1) and by the finite source size (smaller z_1).

D4. Although the angular sensitivities of the wavefields in the sample and detector planes differ, the phase sensitivities are equal and are both minimized by maximizing the magnification

The difference between $\Delta\Theta_\phi$ and $\Delta\Theta_\Phi$ may at first appear to be a curious asymmetry. However, the phase sensitivities in the planes of the sample and the detector are in fact equal for a given sample position. For divergent illumination, the angular distribution of the wavefield in the sample plane is larger than that of the wavefield in the detector plane by a factor of M. On the other hand, the sampling frequency of the wavefield in the sample plane is also larger by the factor M. So, when propagating uncertainties in the angular distributions to the integrated phase profiles, these two effects cancel.

Recall the relationship between the integrated phase profile and the angular distribution of the wavefield in equation (1). Uncertainties in the angular distribution are thus scaled by the step size in x after integration, so that

$$\Delta\phi \simeq \sigma_{\text{ref}} \frac{2\pi}{\lambda} \Delta\Theta_\phi = \sigma_{\text{ref}} \frac{2\pi}{\lambda} \frac{\sigma_{\text{eff}}}{z}, \qquad (129)$$

$$\Delta\Phi \simeq \sigma_{\text{eff}} \frac{2\pi}{\lambda} \Delta\Theta_\Phi = \sigma_{\text{eff}} \frac{2\pi}{\lambda} \frac{\sigma_{\text{ref}}}{z}. \qquad (130)$$

Therefore, with $\sigma_{\text{ref}} = \sigma_{\text{eff}}/M$, we have that $\Delta\phi = \Delta\Phi \simeq (2\pi/\lambda)\sigma_{\text{eff}}^2/M$ and both are minimized by placing the sample as close as possible to the focal plane, as described in the previous subsection.

References

Bajt, S., Prasciolu, M., Fleckenstein, H., Domaracký, M., Chapman, H. N., Morgan, A. J., Yefanov, O., Messerschmidt, M., Du, Y., Murray, K. T., Mariani, V., Kuhn, M., Aplin, S., Pande, K., Villanueva-Perez, P., Stachnik, K., Chen, J. P. J., Andrejczuk, A., Meents, A., Burkhardt, A., Pennicard, D., Huang, X., Yan, H., Nazaretski, E., Chu, Y. S. & Hamm, C. E. (2018). *Light Sci. Appl.* **7**, 17162.

Berto, P., Rigneault, H. & Guillon, M. (2017). *Opt. Lett.* **42**, 5117–5120.

Bérujon, S., Wang, H., Alcock, S. & Sawhney, K. (2014). *Opt. Express*, **22**, 6438–6446.

Bérujon, S., Wang, H. & Sawhney, K. (2012). *Phys. Rev. A*, **86**, 063813.

Bérujon, S., Ziegler, E., Cerbino, R. & Peverini, L. (2012). *Phys. Rev. Lett.* **108**, 158102.

Bonse, U. & Hart, M. (1965). *Appl. Phys. Lett.* **6**, 155–156.

Chapman, H. N. (1996). *Ultramicroscopy*, **66**, 153–172.

Daniel, M.-D. & Ghozeil, I. (1992). *Optical Shop Testing*, 2nd ed., ch. 10. New York: John Wiley & Sons.

David, C., Nöhammer, B., Solak, H. H. & Ziegler, E. (2002). *Appl. Phys. Lett.* **81**, 3287–3289.

Demirci, M. F., Shokoufandeh, A., Keselman, Y., Bretzner, L. & Dickinson, S. (2006). *Int. J. Comput. Vis.* **69**, 203–222.

Fannjiang, A. (2019). *Multiscale Model. Simul.* **17**, 973–995.

Fedoryuk, M. V. (1971). *Russ. Math. Surv.* **26**, 65–115.

Huang, X., Yan, H., Nazaretski, E., Conley, R., Bouet, N., Zhou, J., Lauer, K., Li, L., Eom, D., Legnini, D., Harder, R., Robinson, I. K. & Chu, Y. S. (2013). *Sci. Rep.* **3**, 3562.

Hüe, F., Rodenburg, J. M., Maiden, A. M., Sweeney, F. & Midgley, P. A. (2010). *Phys. Rev. B*, **82**, 121415.

Kechagias-Stamatis, O., Aouf, N., Gray, G., Chermak, L., Richardson, M. & Oudyi, F. (2018). *Aerosp. Sci. Technol.* **73**, 309–317.

Krenkel, M., Toepperwien, M., Alves, F. & Salditt, T. (2017). *Acta Cryst.* A**73**, 282–292.

Lane, R. G. & Tallon, M. (1992). *Appl. Opt.* **31**, 6902–6908.

Lombaert, H., Grady, L., Pennec, X., Ayache, N. & Cheriet, F. (2014). *Int. J. Comput. Vis.* **107**, 254–271.

Mayo, S. C. & Sexton, B. (2004). *Opt. Lett.* **29**, 866–868.

Mercère, P., Idir, M., Moreno, T., Cauchon, G., Dovillaire, G., Levecq, X., Couvet, L., Bucourt, S. & Zeitoun, P. (2006). *Opt. Lett.* **31**, 199–201.

Mimura, H., Handa, S., Kimura, T., Yumoto, H., Yamakawa, D., Yokoyama, H., Matsuyama, S., Inagaki, K., Yamamura, K., Sano, Y., Tamasaku, K., Nishino, Y., Yabashi, M., Ishikawa, T. & Yamauchi, K. (2010). *Nat. Phys.* **6**, 122–125.

Morgan, A. J., Murray, K. T., Prasciolu, M., Fleckenstein, H., Yefanov, O., Villanueva-Perez, P., Mariani, V., Domaracky, M., Kuhn, M., Aplin, S., Mohacsi, I., Messerschmidt, M., Stachnik, K., Du, Y., Burkhart, A., Meents, A., Nazaretski, E., Yan, H., Huang, X., Chu, Y., Andrejczuk, A., Chapman, H. N. & Bajt, S. (2020). *J. Appl. Cryst.* **53**, http://doi.org/10.1107/S1600576720006925.

Morgan, A. J., Prasciolu, M., Andrejczuk, A., Krzywinski, J., Meents, A., Pennicard, D., Graafsma, H., Barty, A., Bean, R. J., Barthelmess, M., Oberthuer, D., Yefanov, O. M., Aquila, A., Chapman, H. N. & Bajt, S. (2015). *Sci. Rep.* **5**, 9892.

Morgan, K. S., Paganin, D. M. & Siu, K. K. W. (2012). *Appl. Phys. Lett.* **100**, 124102.

Murray, K. T., Pedersen, A. F., Mohacsi, I., Detlefs, C., Morgan, A. J., Prasciolu, M., Yildirim, C., Simons, H., Jakobsen, A. C., Chapman, H. N., Poulsen, H. F. & Bajt, S. (2019). *Opt. Express*, **27**, 7120–7138.

Nugent, K. A., Gureyev, T. E., Cookson, D. F., Paganin, D. M. & Barnea, Z. (1996). *Phys. Rev. Lett.* **77**, 2961–2964.

Olivo, A. & Speller, R. (2007). *Appl. Phys. Lett.* **91**, 074106.

Paganin, D. M. (2006). *Coherent X-ray Optics.* Oxford University Press.

Paganin, D. M., Labriet, H., Brun, E. & Bérujon, S. (2018). *Phys. Rev. A*, **98**, 053813.

Stockmar, M., Cloetens, P., Zanette, I., Enders, B., Dierolf, M., Pfeiffer, F. & Thibault, P. (2013). *Sci. Rep.* **3**, 1927.

Teague, M. R. (1983). *J. Opt. Soc. Am.* **73**, 1434.

Thibault, P., Dierolf, M., Bunk, O., Menzel, A. & Pfeiffer, F. (2009). *Ultramicroscopy*, **109**, 338–343.

Thirion, J. P. (1998). *Med. Image Anal.* **2**, 243–260.

Wang, H., Sutter, J. & Sawhney, K. (2015). *Opt. Express*, **23**, 1605–1614.

Wilkins, S. W., Gureyev, T. E., Gao, D., Pogany, A. & Stevenson, A. W. (1996). *Nature*, **384**, 335–338.

Wilkins, S. W., Nesterets, Y. I., Gureyev, T. E., Mayo, S. C., Pogany, A. & Stevenson, A. W. (2014). *Proc. R. Soc. A*, **372**, 20130021.

Zanette, I., Zhou, T., Burvall, A., Lundström, U., Larsson, D. H., Zdora, M., Thibault, P., Pfeiffer, F. & Hertz, H. M. (2014). *Phys. Rev. Lett.* **112**, 253903.

Zdora, M.-C. (2018). *J. Imaging*, **4**, 60.

Zdora, M. C., Thibault, P., Zhou, T., Koch, F. J., Romell, J., Sala, S., Last, A., Rau, C. & Zanette, I. (2017). *Phys. Rev. Lett.* **118**, 203903.

Zdora, M.-C., Zanette, I., Zhou, T., Koch, F. J., Romell, J., Sala, S., Last, A., Ohishi, Y., Hirao, N., Rau, C. & Thibault, P. (2018). *Opt. Express*, **26**, 4989–5004.

Zhou, T., Wang, H., Fox, O. J. & Sawhney, K. J. (2019). *Rev. Sci. Instrum.* **90**, 021706.