Ronchi shearing interferometry for wavefronts with circular symmetry

Bob Nagler,[a]* Eric C. Galtier,[a] Shaughnessy B. Brown,[a,b] Philip Heimann,[a] Gilliss Dyer[a] and Hae Ja Lee[a]

[a]SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA, and [b]Stanford University, 450 Serra Mall, Stanford, CA 94305, USA. *Correspondence e-mail: bnagler@slac.stanford.edu

Ronchi testing of a focused electromagnetic wave has in the last few years been used extensively at X-ray free-electron laser (FEL) facilities to qualitatively evaluate the wavefront of the beam. It is a quick and straightforward test, is easy to interpret on the fly, and can be used to align phase plates that correct the focus of aberrated beams. In general, a single Ronchigram is not sufficient to gain complete quantitative knowledge of the wavefront. However the compound refractive lenses that are commonly used at X-ray FELs exhibit a strong circular symmetry in their aberration, and this can be exploited. Here, a simple algorithm that uses a single recorded Ronchigram to recover the full wavefront of a nano-focused beam, assuming circular symmetry, is presented, and applied to experimental measurements at the Matter in Extreme Conditions instrument at the Linac Coherent Light Source.

1. Introduction

Shearing interferometry and other forms of wavefront sensing have in the last few years become widely used to determine the focal characteristics of X-ray free-electron lasers (Rutishauser *et al.*, 2012; Nagler *et al.*, 2017; Liu *et al.*, 2018; Seaberg *et al.*, 2019; Makita *et al.*, 2020). Indeed, the unique scientific opportunities at X-ray FELs (Bostedt *et al.*, 2016) have required ever more precise knowledge of the focal properties of the X-ray source, especially in fields where the highest X-ray intensities are needed, such as in anomalous nonlinear X-ray Compton scattering (Fuchs *et al.*, 2015), the formation of hollow atoms (Tamasaku *et al.*, 2014), radiation damage studies on protein micro-crystals (Nass *et al.*, 2015), dense plasma science (Vinko *et al.*, 2012) and phase contrast imaging (Schropp *et al.*, 2015; Nagler *et al.*, 2016).

At the same time, a simple Ronchi test has frequently been used to gather quick, easy interpretable qualitative data about the wavefront, and therefore the focus, of X-ray beams (Nilsson *et al.*, 2012). They have been especially useful to align phase plates that correct the aberrations of compound refractive lenses (CRLs) of beryllium, and achieve near diffraction limited foci of approximately 100 nm (Seiboth *et al.*, 2017).

Measurements of the wavefront of X-ray beams focused with CRLs have revealed that the predominant aberrations have a circular symmetry (Schropp *et al.*, 2013; Seaberg *et al.*, 2019). This is not surprising, considering the processes involved in the manufacturing of a Be lens (Lengeler *et al.*, 2002). This circular symmetry can be exploited when determining the wavefront of the beam. While in general at least four Ronchigrams are needed to measure the wavefront of circular beams (Nagler *et al.*, 2017), we show that, by
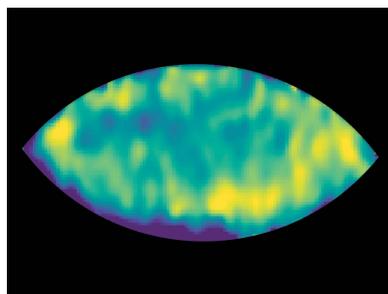
exploiting the circular symmetry, a single Ronchigram is sufficient. We will start by giving an overview of the principle of Ronchi testing, summarizing the method explained in detail by Nagler *et al.* (2017). We then introduce the algorithm to retrieve the phase that exploits the circular symmetry of the wavefront. We conclude with an example based on measurements of the nano-focus at the Matter in Extreme Conditions (MEC) (Nagler *et al.*, 2015) instrument at Linac Coherent Light Source (LCLS).

2. Principle of Ronchi shearing interferometry

The principles of Ronchi shearing interferometry are treated in detail by Nagler *et al.* (2017). A conceptual sketch of the setup can be seen in Figs. 1(*a*) and 1(*b*). The Ronchi target is a one-dimensional (1D) diffraction grating, with a pitch chosen such that the first orders overlap with the zeroth order (*i.e.* the transmitted beam), but do *not* overlap with each other. This is achieved when the pitch of the grating $D = 2f_\# \lambda$, with $f_\#$ the *f*-number of the optic that is used to focus the X-rays (*i.e.* the CRL), and λ the wavelength of the X-ray photons. Where the first orders overlap with the zeroth order (*i.e.* the transmitted beam), interference fringes appear. The properties of the interference fringes can be easily understood, by tracing back the first-order diffractions to their virtual foci. This is shown in Fig. 1(*b*), where only the +1 diffraction order is shown for clarity. The virtual foci can be viewed as secondary sources

of spherical waves. The resulting interference pattern [see Fig. 1(*c*)] is nothing but the interference of two spherical wave sources, separated by a distance *d*. This interference pattern is therefore identical to the pattern of the famous Young's double slit experiment, and the fringes will have the same spatial frequency, with the wavenumber equal to

$$\frac{2\pi z_1}{z_c D}. \tag{1}$$

However, the focus (and virtual focus) is not a perfect source of spherical waves, since the beam has aberrations. Therefore, the interference fringes seen on the detector will be distorted. From the above, it follows that the phase difference between the first order and the zeroth order, also called the differential phase, which uniquely determines the fringe pattern, can be written as

$$\Delta\phi = \text{Cte} + \hat{S}_{X_D}[\phi_{c0}(x, y)] - \frac{2\pi z_1}{z_c D} x \tag{2}$$

with Cte a constant, X_D the displacement of the first order of the beam with respect to the zeroth order (*i.e.* half the beam diameter) at the camera plane, $\phi_{c0}(x, y)$ the phase of the beam at the detector position after removing its spherical component, z_1 the distance between the Ronchi target and the focus, z_c the distance between the focus and the camera position, and $\hat{S}_{X_D}$ the shearing operator defined as
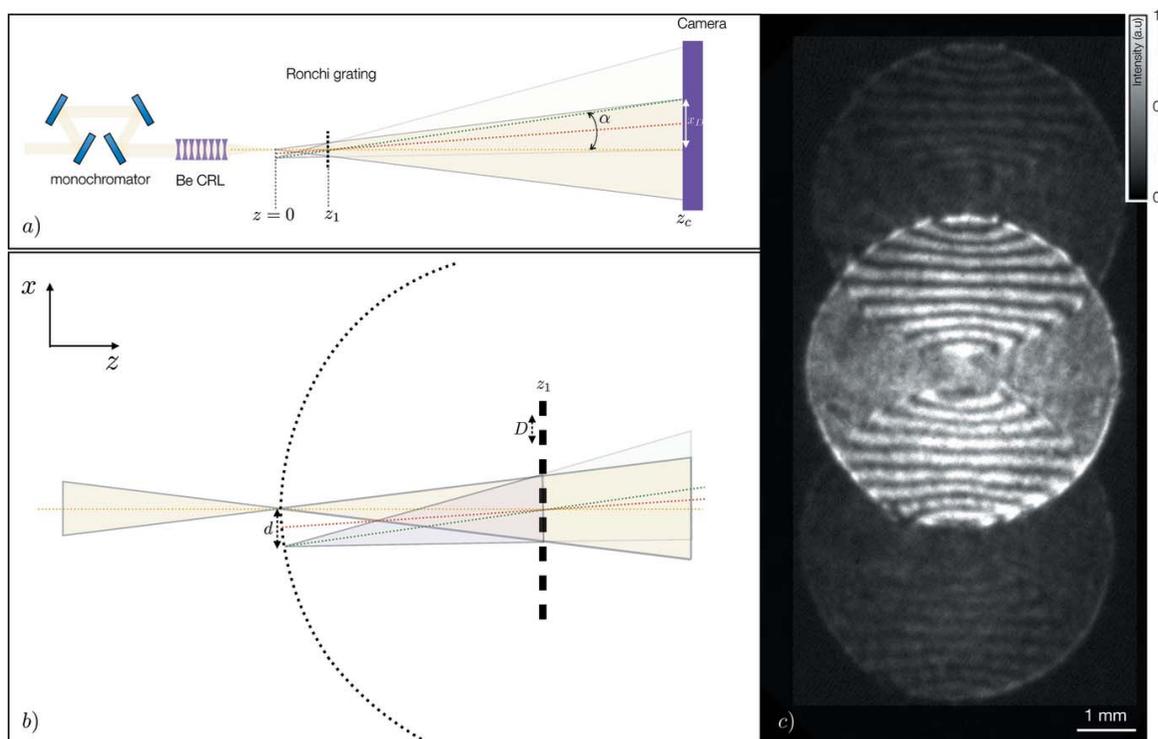

Figure 1
(*a*) Conceptual sketch (not to scale) of the setup. A 1D Ronchi grating is placed somewhat behind the focus of the X-rays [*e.g.* 5 mm in the data presented in (*c*)]. The period of the Ronchi grating is chosen such that orders 1 and −1 do not overlap, while maintaining a large overlap between order 0 and the first orders. Only the +1 order is drawn for clarity. A monochromator [*e.g.* Si(111) for data shown in (*c*)] may be necessary due to the chromaticity of the CRL. (*b*) Closeup of the rays close to the focus and the Ronchi grating, showing the focus of the beam and the virtual focus of the +1 order, that act as secondary sources. (*c*) Image of the interference pattern on the camera, showing the distorted linear fringes.

$$\hat{S}_{X_D}\big[f(x,y)\big] = f(x,y) - f(x - X_D, y). \tag{3}$$

A mathematically rigorous justification of equation (2) is given by Nagler *et al.* (2017), and in Appendix *A* for completeness.

The phase difference in equation (2) can be easily recovered from the experimental data in Fig. 1(*c*) using standard Fourier transform and phase unwrapping methods (Takeda *et al.*, 1982). Inverting the shearing operator is therefore the only remaining problem. As elaborated by Nagler *et al.* (2017), it usually requires multiple Ronchigrams. We will show below that only one Ronchigram is necessary if we assume a circular symmetry in $\phi_{c0}(x,y)$.

3. Inverting the shearing operator exploiting circular symmetry

3.1. The cost function

To invert the shearing operator we use the same optimization method as in Nagler *et al.* (2017) and Makita *et al.* (2020): we minimize a cost function that expresses the difference between our experimental measured differential phase and the differential phase that is calculated from an initial guess of the X-ray wavefront. It is this guess that will be optimized iteratively [*e.g.* by using a conjugate gradient descent method, or a limited memory Broyden–Fletcher–Goldfard–Shanno algorithm (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970; Nocedal, 1980)]. We define the cost function as

$$U^2 = U_{\bar{s}}^2 + \mu U_{\circ}^2 + \lambda\big(R_x^2 + R_y^2\big), \tag{4}$$

where μ and λ are regularization parameters that need to be chosen judiciously. The difference with Makita *et al.* (2020) is that we only use one interferogram and we add a term to the cost function, $U_{\circ}^2$. This new term will be chosen in such a way that it will drive the solution of optimization towards cylindrical symmetric phases.

As in Makita *et al.* (2020), we have

$$U_{\bar{s}}^2 = \sum_{k,l}\Big(\hat{S}_{\bar{s}}[\phi_{k,l}] - \phi_{k,l}^{\bar{s}} + \phi_{c}^{\bar{s}}\Big)^2 P_{k,l}^{\bar{s}} \tag{5}$$

where $\phi_{k,l}$ is the phase of the beam that needs to be optimized at pixel (k,l), $\phi_{k,l}^{\bar{s}}$ is the measured differential phase, $\phi_{c}^{\bar{s}}$ is a constant phase that needs to be optimized and $P_{k,l}^{\bar{s}}$ is the analysis mask.

The shear operator is defined as

$$\hat{S}_{\bar{s}}[\phi_{k,l}] = \phi_{k,l} - \hat{d}_{\bar{s}}[\phi_{k,l}], \tag{6}$$

with the shift operator $\hat{d}_{\bar{s}}$ given by

$$\hat{d}_{\bar{s}}[\phi(x,y)] = \phi\big(x - s_x, y - s_y\big). \tag{7}$$

R_x and R_y are added to the cost function to regularize for smoothness and are defined as

$$R_x^2 = \sum_{k,l}\big(\phi_{k-1,l} - 2\phi_{k,l} + \phi_{k+1,l}\big)^2 P_{k-1,l}\, P_{k+1,l} \tag{8}$$

$$R_y^2 = \sum_{k,l}\big(\phi_{k,l-1} - 2\phi_{k,l} + \phi_{k,l+1}\big)^2 P_{k,l-1}\, P_{k,l+1}\ . \tag{9}$$

Since the shift operator $\hat{d}_{\bar{s}}$ is unitary, we can show that the partial derivatives of the cost function are equal to

$$\frac{\partial U^2}{\partial \phi_{k,l}} = \frac{\partial U_{\bar{s}}^2}{\partial \phi_{k,l}} + \lambda\left(\frac{\partial R_x^2}{\partial \phi_{k,l}} + \frac{\partial R_y^2}{\partial \phi_{k,l}}\right) + \mu\frac{\partial U_{\circ}^2}{\partial \phi_{k,l}}, \tag{10}$$

$$\frac{\partial U^2}{\partial \phi_{c}^{\bar{s}}} = 2\sum_{k,l} P_{k,l}^{\bar{s}}\big(\hat{S}_{\bar{s}}[\phi_{k,l}] - \phi_{k,l}^{\bar{s}} + \phi_{c}^{\bar{s}}\big) + \mu\frac{\partial U_{\circ}^2}{\partial \phi_{c}^{\bar{s}}}, \tag{11}$$

with

$$\frac{\partial U_{\bar{s}}^2}{\partial \phi_{k,l}} = 2\hat{S}_{-\bar{s}}\Big[\big(\hat{S}_{\bar{s}}[\phi_{k,l}] - \phi_{k,l}^{\bar{s}} + \phi_{c}^{\bar{s}}\big)P_{k,l}^{\bar{s}}\Big], \tag{12}$$

$$\begin{aligned}
\frac{\partial R_x^2}{\partial \phi_{k,l}} = {} & 2\big(\phi_{k,l} - 2\phi_{k+1,l} + \phi_{k+2,l}\big)P_{k,l}\, P_{k+2,l} \\
& - 4\big(\phi_{k-1,l} - 2\phi_{k,l} + \phi_{k+1,l}\big)P_{k-1,l}\, P_{k+1,l} \\
& + 2\big(\phi_{k-2,l} - 2\phi_{k-1,l} + \phi_{k,l}\big)P_{k-2,l}\, P_{k,l}\ ,
\end{aligned} \tag{13}$$

$$\begin{aligned}
\frac{\partial R_y^2}{\partial \phi_{k,l}} = {} & 2\big(\phi_{k,l} - 2\phi_{k,l+1} + \phi_{k,l+2}\big)P_{k,l}\, P_{k,l+2} \\
& - 4\big(\phi_{k,l-1} - 2\phi_{k,l} + \phi_{k,l+1}\big)P_{k,l-1}\, P_{k,l+1} \\
& + 2\big(\phi_{k,l-2} - 2\phi_{k,l-1} + \phi_{k,l}\big)P_{k,l-2}\, P_{k,l}\ .
\end{aligned} \tag{14}$$

These equation can be used for non-integer shear values, if we use an interpolated form of the shift function (*e.g.* bi-linear, or bi-quadratic interpolation). The introduction of $U_{\circ}^2$ and its partial derivatives are the new elements compared with the previous work of Nagler *et al.* (2017), and are explained in detail below.

3.2. Cost function for circular symmetry

To impose circular symmetry, we can define a cost function that expresses the variance of the wavefront over the angle θ, in the polar coordinates (ρ, θ). This cost function has the form

$$U_{\rho}^2 = \langle \phi^2 \rangle_{\rho} - \langle \phi \rangle_{\rho}^2 \tag{15}$$

with

$$\langle \phi^2 \rangle_{\rho} = \frac{1}{2\pi}\int\limits_{0}^{2\pi} \phi^2(\rho, \theta)\, \mathrm{d}\theta, \tag{16}$$

$$\langle \phi \rangle_{\rho} = \frac{1}{2\pi}\int\limits_{0}^{2\pi} \phi(\rho, \theta)\, \mathrm{d}\theta. \tag{17}$$

The total cost function can be defined as

$$U_{\circ}^2 = \int U_{\rho}^2\, \mathrm{d}\rho, \tag{18}$$

or, alternatively, if the circumference is used to weight the function, as

$$U_\circ^2 = \int U_\rho^2 \rho \, d\rho. \tag{19}$$

In practice, since different radii optimize independently, both cost functions (18) and (19) optimize to the same minimum, although the path of the optimizer would differ due to the different gradients. We choose a numerical implementation of (18).

3.3. Numerical implementation

A numerical implementation will require resampling the wavefront in polar coordinates, with center (x_c, y_c), using the relation

$$x - x_c = \rho \cos(\theta), \tag{20}$$

$$y - y_c = \rho \sin(\theta), \tag{21}$$

where x and y are expressed in pixels. The sample spacing along the radius, $d\rho$, can be constant, but the spacing in angle will depend on the radius, since we want to keep $dl_i = \rho_i \, d\theta_i$ pseudo-constant. It is only pseudo-constant since we need the sampling to be uniform over 2π, which implies dl_i will vary slightly with ρ. We choose the size of dl_i to be close to, but larger than, one pixel. If this value is too large, we risk introducing circular harmonic artifacts, although this is somewhat mitigated by the Tikonov regularization for smoothness. We choose ρ_{max}, the largest radius in the apperture that will be considered, dl, the nominal distance in pixels between sample points in the polar angle and $d\rho$, the radial distance in pixels between sample points. We can then calculate

$$I_{max} = \left\lfloor \frac{\rho_{max}}{d\rho} \right\rfloor - 1, \tag{22}$$

$$\rho_i = (i + 1) \, d\rho, \tag{23}$$

$$J_i = \left\lceil 2\pi \rho_i \, dl^{-1} \right\rceil - 1, \tag{24}$$

$$dl_i = \frac{2\pi \rho_i}{J_i + 1}, \tag{25}$$

$$\theta_{ij} = j \frac{2\pi}{J_i + 1} = j \rho_i^{-1} \, dl_i. \tag{26}$$

We use the floor function (*i.e.* $\lfloor \ldots \rfloor$) to calculate the integer I_{max}, the largest index in the radial sampling. This is to ensure that $\rho_I \leq \rho_{max}$. The integer J_i is the number of sample points in the polar angle for the radius ρ_i and is defined using the ceiling function (*i.e.* $\lceil \ldots \rceil$). This ensures that $J_i \geq 0$. In equations (22) to (26), we have

$$i \in \{0, 1, 2, \ldots, I_{max}\}, \tag{27}$$

$$j \in \{0, 1, 2, \ldots, J_i\}. \tag{28}$$

We will express the sampling of the polar coordinates in a matrix. Since the number of samples in θ is different for every ρ, we will zero pad them to keep the matrix rectangular. We define

$$J_{max} = \max\{J_i\}, \tag{29}$$

and

$$\mathcal{W}_{ij} = \phi(\rho_i, \theta_{ij}), \qquad \text{for} \quad j \in \{0, 1, 2, \ldots, J_i\}, \tag{30}$$

$$\mathcal{W}_{ij} = 0, \qquad \text{for} \quad j \in \{J_i + 1, \ldots, J_{max}\}. \tag{31}$$

The matrix $\mathcal{W}_{ij}$ has dimensions $(I_{max} + 1) \times (J_{max} + 1)$. With these definitions, the discrete implementation of equations (16) and (17) becomes

$$\langle \mathcal{W}^2 \rangle_{\rho_i} = \frac{1}{J_i + 1} \sum_{j=0}^{J_{max}} \mathcal{W}_{ij}^2, \tag{32}$$

$$\langle \mathcal{W} \rangle_{\rho_i} = \frac{1}{J_i + 1} \sum_{j=0}^{J_{max}} \mathcal{W}_{ij}. \tag{33}$$

Note that the sum can run up to J_{max}, due to our definition of the padding of $\mathcal{W}_{ij}$ in equation (31), but the denominator needs to be $J_i + 1$. With these definitions, the discrete equivalent the total cost function [*i.e.* equation (18)] becomes

$$U_\circ^2 = \frac{1}{I_{max} + 1} \sum_{i=0}^{I_{max}} (\langle \mathcal{W}^2 \rangle_{\rho_i} - \langle \mathcal{W} \rangle_{\rho_i}^2). \tag{34}$$

Since the phase of the beam $\phi(x, y)$ is only expressed at integer pixel values, interpolation will need to be used to obtain the correct values. For example, if we need $\mathcal{W}(\theta = 1.1, \rho = 5)$ with the center of the beam at $(10, 10)$, we need to know $\phi(x = 12.27, y = 14.46)$. This value will need to be calculated by interpolating between the values of the neighboring pixels [for example by using a bilinear interpolation between pixels $(12, 14)$, $(12, 15)$, $(13, 14)$ and $(13, 15)$]. For the moment we will not specify the interpolation method we use, but write in general that

$$\mathcal{W}_{ij} = A_{ijkl} \phi_{kl} \tag{35}$$

and

$$\frac{\partial \mathcal{W}_{ij}}{\partial \phi_{kl}} = A_{ijkl} \tag{36}$$

where we use Einstein's summation convention. Substitution into equation (34) yields

$$U_\circ^2 = \frac{1}{I_{max} + 1} \left[\sum_{ij} \frac{1}{J_i + 1} \left(A_{ijkl} \phi_{kl} \right)^2 - \sum_i \frac{1}{(J_i + 1)^2} \left(\sum_j A_{ijkl} \phi_{kl} \right)^2 \right]. \tag{37}$$

The partial derivatives of $U_\circ^2$ are

$$\frac{\partial U_\circ^2}{\partial \phi_{kl}} = B_{klpq} \phi_{pq} - C_{klpq} \phi_{pq} \tag{38}$$

with

$$B_{pqkl} = \frac{2}{I_{max} + 1} \sum_{ij} (J_i + 1)^{-1} A_{ijpq} A_{ijkl}, \tag{39}$$

$$C_{pqkl} = \frac{2}{I_{max}+1} \sum_{ijm} (J_i+1)^{-2} A_{ijpq} A_{imkl} \ . \qquad (40)$$

The matrices A_{ijkl} and B_{pqkl} may seem prohibitively large: if we sample θ and ρ as densely as x and y, the number of elements in these matrices scales as N^4. However these matrices are in general extremely sparse. Indeed, if we use nearest-neighbor interpolation, we only have one non-zero element for every (i, j), making it no larger in content than W_{kl}.

For bilinear and bicubic interpolation, this would become 4 and 16 non-zero elements per (i, j), respectively. Therefore, sparse matrix routines can be used to implement the calculation of the cost function efficiently and with little memory overhead. Care needs to be taken never to loop over all indices of the matrices, since this will take prohibitively long.

3.4. Flattening and assembling indices: reducing the dimension

To simplify notations, and since sparse matrix routines are generally only defined in 2D matrices in most computational libraries, we will contract two indices to one. In this way, matrices are reshaped by putting all the columns after each other into one large vector with dimension $Y_D = (I_{max}+1) \times (J_{max}+1)$. The mapping of the two-dimensional indices to the 1D one is given by

$$Y = J_D . i + j \qquad (41)$$

and

$$i = Y \bmod J_D, \qquad (42)$$

$$j = Y \operatorname{rem} J_D, \qquad (43)$$

where J_D is the dimension of the index j (e.g. $J_{max} + 1$ in the case of the polar coordinate indices defined above). In the remainder of this paper, we will use capital letters to denote flattened indices. The index Y will generally refer to the flattened indices (ij) of the polar coordinates (ρ, θ). The indices K and P will refer to the flattened Cartesian indices (i.e. kl and pq). K_D is the total number of elements (or the length) of the flattened Cartesian indices (i.e. total number of pixels in the original image). With this convention, we can rewrite the matrices and cost functions in a compact and computationally manageable form.

3.4.1. Defining $\bar{\bar{A}} = A_{YK}$. This matrix has dimension $Y_D \times K_D$. As mentioned above, it interpolates between the original image and the polar sampling of the image, and is sparse. We have

$$\mathcal{W}_Y = A_{YK} \phi_K \qquad (44)$$

or

$$\bar{\mathcal{W}} = \bar{\bar{A}} \bar{\phi}. \qquad (45)$$

3.4.2. Defining $\bar{\bar{B}} = B_{PK}$. This matrix has a dimension $K_D \times K_D$ and is symmetric. To calculate $\bar{\bar{B}}$ we first define the extended matrix J_{ij},

$$J_{ij} = J_i \quad \text{for all} \quad j. \qquad (46)$$

We can then rewrite equation (39) as

$$\bar{\bar{B}} = 2\bar{\bar{A}}^T \bar{\bar{D}} = 2\bar{\bar{D}}^T \bar{\bar{A}} \qquad (47)$$

with

$$D_{YK} = \frac{A_{YK}}{(J_Y+1)(I_{max}+1)}. \qquad (48)$$

3.4.3. Defining $\bar{\bar{C}} = C_{PK}$. This matrix has a dimension $K_D \times K_D$ and is symmetric. We first define

$$G_{iK} = \frac{1}{(I_{max}+1)^{1/2}} \sum_j \frac{A_{ijK}}{J_i+1}. \qquad (49)$$

The matrix G_{iK} is somewhat sparse, and has dimension $(I_{max}+1) \times K_D$.

With this definition, we can easily calculate C_{PK},

$$\bar{\bar{C}} = 2\bar{\bar{G}}^T \bar{\bar{G}}. \qquad (50)$$

3.5. Expressions for the cost function and its derivative

With the matrices defined above, we can recast the cost function and its partial derivatives [i.e. equations (37) and (38)] as

$$U_\circ^2 = \frac{1}{2}\bar{\phi}\bar{\bar{B}}\bar{\phi} - \|\bar{\bar{G}}\bar{\phi}\|^2, \qquad (51)$$

$$\frac{\partial U_\circ^2}{\partial \phi_K} = \bar{\bar{F}}\bar{\phi}, \qquad (52)$$

with

$$\bar{\bar{B}} = 2\bar{\bar{A}}^T \bar{\bar{D}}, \qquad (53)$$

$$\bar{\bar{F}} = \bar{\bar{B}} - 2\bar{\bar{G}}^T \bar{\bar{G}}, \qquad (54)$$

$$D_{YK} = \frac{A_{YK}}{(J_Y+1)(I_{max}+1)}, \qquad (55)$$

$$G_{iK} = \frac{1}{(I_{max}+1)^{1/2}} \sum_j \frac{A_{ijK}}{J_i+1}. \qquad (56)$$

3.6. Calculating the matrices $\bar{\bar{A}}$, $\bar{\bar{D}}$ and $\bar{\bar{G}}$

While the matrices $\bar{\bar{B}}$, $\bar{\bar{C}}$, $\bar{\bar{F}}$, the cost function, and its derivatives can be calculated using sparse matrix multiplication, this is not the case for $\bar{\bar{A}}$, $\bar{\bar{D}}$ and $\bar{\bar{G}}$. They need to be constructed, and looping over their indices takes prohibitively long. A procedure to calculate the non-zero elements of these matrices, and their corresponding indices, is outlined in Fig. 2.

loop over (i, j):

for each $(i, j) \xrightarrow{flatten} Y$

$(\rho_i, \theta_{i,j}) \rightarrow (x, y)$
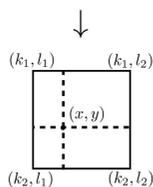
$(k_1, l_1) \xrightarrow{flatten} K_1 \rightarrow \begin{cases} A_{Y,K_1} = (k_2 - x)(l_2 - y) \\ D_{Y,K_1} = A_{Y,K_1}(J_i + 1)^{-1}(I_{\max} + 1)^{-1} \\ G_{i,K_1} + = D_{Y,K_1}\sqrt{I_{\max} + 1} \end{cases}$

$(k_2, l_1) \xrightarrow{flatten} K_2 \rightarrow \begin{cases} A_{Y,K_2} = (x - k_1)(l_2 - y) \\ D_{Y,K_2} = A_{Y,K_2}(J_i + 1)^{-1}(I_{\max} + 1)^{-1} \\ G_{i,K_2} + = D_{Y,K_2}\sqrt{I_{\max} + 1} \end{cases}$

$(k_1, l_2) \xrightarrow{flatten} K_3 \rightarrow \begin{cases} A_{Y,K_3} = (k_2 - x)(y - l_1) \\ D_{Y,K_3} = A_{Y,K_3}(J_i + 1)^{-1}(I_{\max} + 1)^{-1} \\ G_{i,K_3} + = D_{Y,K_3}\sqrt{I_{\max} + 1} \end{cases}$

$(k_2, l_2) \xrightarrow{flatten} K_4 \rightarrow \begin{cases} A_{Y,K_4} = (x - k_1)(y - l_1) \\ D_{Y,K_4} = A_{Y,K_4}(J_i + 1)^{-1}(I_{\max} + 1)^{-1} \\ G_{i,K_4} + = D_{Y,K_4}\sqrt{I_{\max} + 1} \end{cases}$

Figure 2
Procedure to calculate the non-zero elements of $\bar{\bar{A}}$, $\bar{\bar{D}}$ and $\bar{\bar{G}}$, and their indices.

4. Experimental results and conclusions

The phase retrieval and optimization method described above was applied to data obtained during an experiment at the MEC endstation (Nagler *et al.*, 2015) at LCLS. A set of 25 beryllium CRLs with a radius of curvature of 50 µm at their apex was used to focus the X-ray beam. A photon energy of 8.2 keV was used, leading to a focal length for the CRL of 205.9 mm. A four-bounce Si(111) monochromator was used to reduce the bandwidth of the beam to approximately 1 eV. This is necessary, since the chromaticity of the CLR together with the native bandwidth of LCLS (*i.e.* approximately 30 eV at 8.2 keV photon energy) would wash out the fringes of the Ronchigram. A Ronchi grating with a pitch of 250 nm was placed approximately 5 mm behind the focus of the CRL. The Ronchi grating consisted of 1.5 µm-thick diamond with an additional layer of 250 nm tungsten (Uhlén *et al.*, 2011), giving a diffraction efficiency of approximately 13% (Nilsson *et al.*, 2012). A Princeton Instruments PIXIS X-ray camera with 13.5 µm pixels was placed 3.85 m behind the focus. High-quality Ronchigrams were recorded in a single shot. In general, integrating multiple shots is not possible with this method, unless the spatial jitter of the beam is extremely small. Indeed, if the beam jitter on the grating is half of the pitch of the grating (*i.e.* 125 nm in our case), the position of the maxima and minima in intensity is reversed, and the fringes are completely washed out in the integration. A sketch of the setup is shown in Figs. 1(*a*) and 1(*b*). A raw camera image is shown in Fig. 1(*c*).

The retrieved phase of the beam is shown in Fig. 3, where we removed the wavefront curvature to make the aberrations
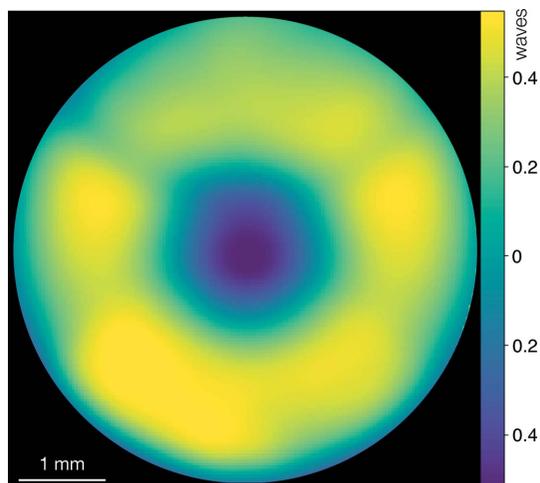
Figure 3
Retrieved wavefront of the beam, after removal of the wavefront curvature. The total RMS of the aberrations over the measurement aperture is approximately 1/4 wave.

more visible. The total RMS of the remaining aberrations is approximately 1/4 wave, with a peak-to-valley of 1 wave.

We then proceeded to test this result in the following manner. The phase difference (or differential phase) of the reconstructed wavefront of the beam was calculated in the same vertical shearing direction as in the experimental data shown in Fig. 1(*c*). This calculated differential phase was compared with the actual measured differential phase, and we call the difference the residual differential phase error. This residual differential phase error is shown in Fig. 4. The residual differential phase error is a good measure of how well the reconstruction algorithm can find a wavefront that matches the measured data, and is therefore a good indication of the overall (minimal) error of the whole phase retrieval procedure. The RMS value of the residual differential phase error over the whole measurement aperture is $\lambda/40$.

Armed with both the phase and the intensity of the beam, we can now calculate the focal properties of the beam by back-propagating the electric field to the focus. This result can be seen in Fig. 5, where we show the beam profile a the best focus
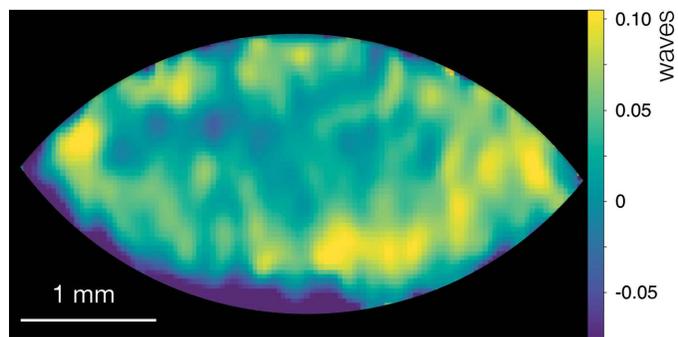
Figure 4
The residual differential phase error of the reconstruction. The RMS of the residual differential phase error over the whole measurement aperture is $\lambda/40$, giving an indication of the reliability of the algorithm (and the symmetry of the beam).
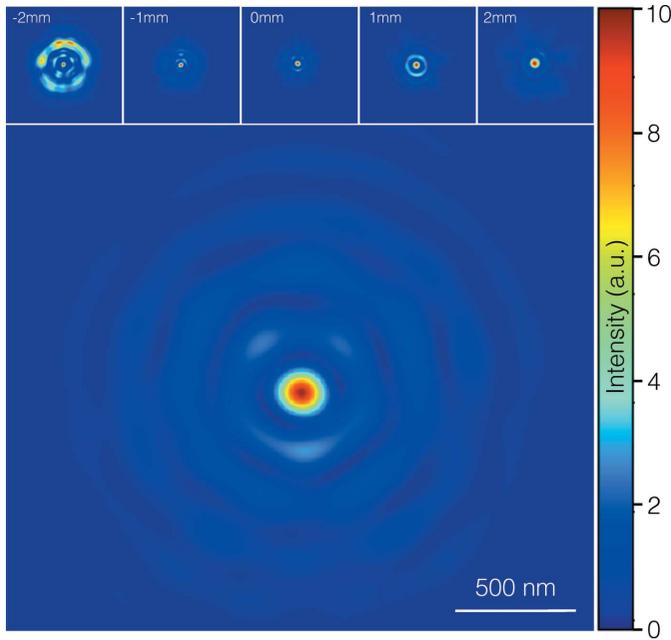
Figure 5
Calculated spatial profile of the X-ray beam at its focal position (bottom) and a scan of the profile around the focal position (top row) in 1 mm steps. Negative values indicate a position before the focal plane.

location, as well as around the focus in 1 mm steps. In Fig. 6 we show lineouts in both the horizontal and vertical direction of the best focus.

Wavefront sensing of X-ray FEL beams focused by CRLs to 100 nm focal spots have been performed in the past with similar lens sets, using a variety of techniques (Schropp *et al.*, 2013; Seaberg *et al.*, 2019). They reported wavefronts with aberration dominated by spherical aberration, of the same order as measured in this paper.

In conclusion, we have presented a method and algorithm that can retrieve the phase of a focused X-ray beam using a single Ronchigram, with a prior of circular symmetry. The algorithm was designed for use on an X-ray free-electron laser, but can be used in general for any wavefront with circular symmetry. We applied it to X-ray beams focused with CRLs to 100 nm size at the MEC endstation at LCLS, and
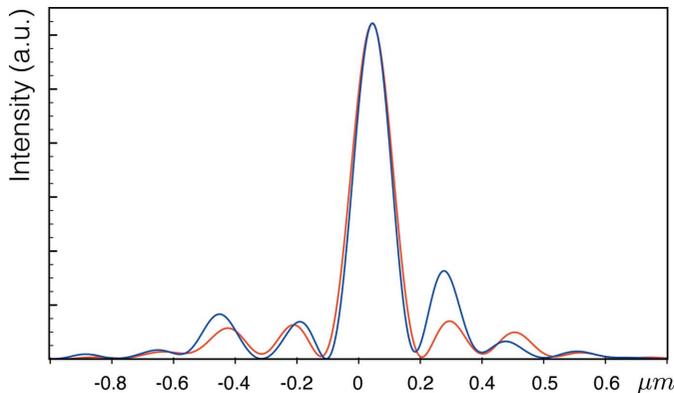

Figure 6
Line-out of the spatial profile of the X-ray beam at its focal position, in both the horizontal (red) and vertical (blue) direction.

obtained results that compare well with previous measurements.

APPENDIX A
Mathematical description of Ronchigrams

A mathematical rigorous derivation of equation (2) can be derived as follows. Consider an X-ray beam with a focus located at $z = 0$, with electric field $E_0(x, y, 0)$. Using the paraxial approximation, we can propagate the electric field to the camera position, z_c, using the Fresnel intergral,

$$E_{c0}(x, y) = P_{z_c}(x, y) \, E_{c0}^F(x, y), \tag{57}$$

with

$$E_{c0}^F(x, y) = -\frac{ik}{2\pi z_c} \exp(ikz_c) \, \mathcal{F}\big[E_0(x, y) \, P_{z_c}(x, y)\big]_{k_x = \frac{k \cdot x}{z_c}, k_y = \frac{k \cdot y}{z_c}} \tag{58}$$

with k the wavenumber of the electromagnetic field, and the spherical phase factor P_{z_c} defined as

$$P_{z_c}(x, y) = \exp\left[\frac{ik}{2z_c}(x^2 + y^2)\right]. \tag{59}$$

E_{c0}^F represents the electric field at the camera position, after removal of the wavefront curvature due to the propagation distance z_c. The Ronchi grating is placed at position $z = z_1$. We propagate the beam from focus to the grating using the Fresnel integral. We multiply the field with the transmission function of the Ronchi grating. Since the Ronchi grating is periodic along the x-axis, and has no dependence on y, its Fourier transform can be written as

$$\mathcal{R}(k_x, k_y) = 2\pi \sum_{l=-\infty}^{+\infty} R_l \, \delta(k_x - k_l) \, \delta(k_y), \tag{60}$$

with

$$k_l = \frac{2\pi}{D} \, l \qquad \text{with} \quad l \in \mathbb{Z}, \tag{61}$$

with D the pitch of the grating, and R_l a sequence of complex numbers that depend on the grating characteristics. When we propagate the field back to the (now virtual) focus of the beam, we obtain the (virtual) field E_{R0},

$$E_{R0} = \frac{1}{2\pi} \sum_{l=-\infty}^{+\infty} R_l \exp\left(i \, \frac{k_l X_l}{2}\right)$$
$$\times E_0(x + X_l, y, 0) \exp(ik_l x), \tag{62}$$

with

$$X_l = \frac{2\pi z_1}{kD} \, l \qquad \text{with} \quad l \in \mathbb{Z}. \tag{63}$$

We now use the Fresnel integral to propagate this field to $z = z_c$. Substituting equation (58) we obtain

$$E_{Rc} = \frac{1}{2\pi} P_{z_c}(x, y) \sum_{l=-\infty}^{+\infty} R_l \exp(i\phi_l)$$
$$\times E_{c0}^F\left(x - X_D^l, y, z_c\right) \exp\left(ik_l \frac{z_1}{z_c} x\right), \quad (64)$$

with

$$\phi_l = \frac{k_l X_l}{2}\left(\frac{z_1}{z_c} - 1\right), \quad (65)$$

$$X_D^l = \frac{k_l}{k}(z_c - z_1). \quad (66)$$

When different orders (*i.e.* different values of l) overlap, they will interfere, and form predominantly linear fringes due to the linear phase in x. From equation (64) we can calculate the phase difference between the zeroth and first order,

$$\Delta\phi = \phi_{R_0} - \phi_{R_1} - \frac{2\pi^2 z_1}{kD^2}\left(\frac{z_1}{z_c} - 1\right)$$
$$+ \hat{S}_{X_D}[\phi_{c0}(x, y)] - \frac{2\pi z_1}{z_c D} x, \quad (67)$$

with ϕ_{R_0} the phase of R_0, ϕ_{R_1} the phase of R_1, $\phi_{c0}(x, y)$ the phase of $E_{c0}^F(x, y, z_c)$, and $\hat{S}_{X_D}$ the shearing operator defined as

$$\hat{S}_{X_D}[f(x, y)] = f(x, y) - f(x - X_D, y), \quad (68)$$

and $X_D \equiv X_D^1$. The first three terms of equation (67) give the constant phase difference between the orders [corresponding to the undetermined Cte in equation (2)]. The last two terms show how the phase varies in x and y. The linear phase in x of the last term will result in linear fringes in the intensity of E_{RC}, provided that the spatial frequency $2\pi z_1/z_c D$ is large enough. Using standard Fourier methods (Takeda *et al.*, 1982) and phase unwrapping algorithms, we can retrieve $\hat{S}_{X_D}[\phi_{c0}(x, y)]$ as long as the spatial frequency $2\pi z_1/z_c D$ is at least twice as big as the highest spatial frequency that is present in the intensity of E_{RC}; otherwise some aliasing will occur.

Funding information

References

Bostedt, C., Boutet, S., Fritz, D. M., Huang, Z., Lee, H. J., Lemke, H. T., Robert, A., Schlotter, W. F., Turner, J. J. & Williams, G. J. (2016). *Rev. Mod. Phys.* **88**, 015007.
Broyden, C. G. (1970). *IMA J. Appl. Math.* **6**, 76–90.
Fletcher, R. (1970). *Comput. J.* **13**, 317–322.
Fuchs, M., Trigo, M., Chen, J., Ghimire, S., Shwartz, S., Kozina, M., Jiang, M., Henighan, T., Bray, C., Ndabashimiye, G., Bucksbaum, P. H., Feng, Y., Herrmann, S., Carini, G. A., Pines, J., Hart, P., Kenney, C., Guillet, S., Boutet, S., Williams, G. J., Messerschmidt, M., Seibert, M. M., Moeller, S., Hastings, J. B. & Reis, D. A. (2015). *Nat. Phys.* **11**, 964–970.
Goldfarb, D. (1970). *Math. C.* **24**, 23–26.
Lengeler, B., Schroer, C. G., Benner, B., Gerhardus, A., Günzler, T. F., Kuhlmann, M., Meyer, J. & Zimprich, C. (2002). *J. Synchrotron Rad.* **9**, 119–124.
Liu, Y., Seaberg, M., Zhu, D., Krzywinski, J., Seiboth, F., Hardin, C., Cocco, D., Aquila, A., Nagler, B., Lee, H. J., Boutet, S., Feng, Y., Ding, Y., Marcus, G. & Sakdinawat, A. (2018). *Optica*, **5**, 967–975.
Makita, M., Seniutinas, G., Seaberg, M. H., Lee, H. J., Galtier, E. C., Liang, M., Aquila, A., Boutet, S., Hashim, A., Hunter, M. S., van Driel, T., Zastrau, U., David, C. & Nagler, B. (2020). *Optica*, **7**, 404–409.
Nagler, B., Aquila, A., Boutet, S., Galtier, E. C., Hashim, A., Hunter, M. S., Liang, M., Sakdinawat, A. E., Schroer, C. G., Schropp, A., Seaberg, M. H., Seiboth, F., van Driel, T., Xing, Z., Liu, Y. & Lee, H. J. (2017). *Sci. Rep.* **7**, 13689.
Nagler, B., Arnold, B., Bouchard, G., Boyce, R. F., Boyce, R. M., Callen, A., Campell, M., Curiel, R., Galtier, E., Garofoli, J., Granados, E., Hastings, J., Hays, G., Heimann, P., Lee, R. W., Milathianaki, D., Plummer, L., Schropp, A., Wallace, A., Welch, M., White, W., Xing, Z., Yin, J., Young, J., Zastrau, U. & Lee, H. J. (2015). *J. Synchrotron Rad.* **22**, 520–525.
Nagler, B., Schropp, A., Galtier, E. C., Arnold, B., Brown, S. B., Fry, A., Gleason, A., Granados, E., Hashim, A., Hastings, J. B., Samberg, D., Seiboth, F., Tavella, F., Xing, Z., Lee, H. J. & Schroer, C. G. (2016). *Rev. Sci. Instrum.* **87**, 103701.
Nass, K., Foucar, L., Barends, T. R. M., Hartmann, E., Botha, S., Shoeman, R. L., Doak, R. B., Alonso-Mori, R., Aquila, A., Bajt, S., Barty, A., Bean, R., Beyerlein, K. R., Bublitz, M., Drachmann, N., Gregersen, J., Jönsson, H. O., Kabsch, W., Kassemeyer, S., Koglin, J. E., Krumrey, M., Mattle, D., Messerschmidt, M., Nissen, P., Reinhard, L., Sitsel, O., Sokaras, D., Williams, G. J., Hau-Riege, S., Timneanu, N., Caleman, C., Chapman, H. N., Boutet, S. & Schlichting, I. (2015). *J. Synchrotron Rad.* **22**, 225–238.
Nilsson, D., Uhlén, F., Holmberg, A., Hertz, H. M., Schropp, A., Patommel, J., Hoppe, R., Seiboth, F., Meier, V., Schroer, C. G., Galtier, E., Nagler, B., Lee, H. J. & Vogt, U. (2012). *Opt. Lett.* **37**, 5046–5048.
Nocedal, J. (1980). *Math. Comput.* **35**, 773–782.
Rutishauser, S., Samoylova, L., Krzywinski, J., Bunk, O., Grünert, J., Sinn, H., Cammarata, M., Fritz, D. M. & David, C. (2012). *Nat. Commun.* **3**, 947.
Schropp, A., Hoppe, R., Meier, V., Patommel, J., Seiboth, F., Lee, H. J., Nagler, B., Galtier, E. C., Arnold, B., Zastrau, U., Hastings, J. B., Nilsson, D., Uhlén, F., Vogt, U., Hertz, H. M. & Schroer, C. G. (2013). *Sci. Rep.* **3**, 1633.
Schropp, A., Hoppe, R., Meier, V., Patommel, J., Seiboth, F., Ping, Y., Hicks, D. G., Beckwith, M. A., Collins, G. W., Higginbotham, A., Wark, J. S., Lee, H. J., Nagler, B., Galtier, E. C., Arnold, B., Zastrau, U., Hastings, J. B. & Schroer, C. G. (2015). *Sci. Rep.* **5**, 11089.
Seaberg, M., Cojocaru, R., Berujon, S., Ziegler, E., Jaggi, A., Krempasky, J., Seiboth, F., Aquila, A., Liu, Y., Sakdinawat, A., Lee, H. J., Flechsig, U., Patthey, L., Koch, F., Seniutinas, G., David, C., Zhu, D., Mikeš, L., Makita, M., Koyama, T., Mancuso, A. P., Chapman, H. N. & Vagovič, P. (2019). *J. Synchrotron Rad.* **26**, 1115–1126.
Seiboth, F., Schropp, A., Scholz, M., Wittwer, F., Rödel, C., Wünsche, M., Ullsperger, T., Nolte, S., Rahomäki, J., Parfeniukas, K., Giakoumidis, S., Vogt, U., Wagner, U., Rau, C., Boesenberg, U., Garrevoet, J., Falkenberg, G., Galtier, E. C., Ja Lee, H., Nagler, B. & Schroer, C. G. (2017). *Nat. Commun.* **8**, 14623.
Shanno, D. F. (1970). *Math. C.* **24**, 647–656.
Takeda, M., Ina, H. & Kobayashi, S. (1982). *J. Opt. Soc. Am.* **72**, 156–160.
Tamasaku, K., Shigemasa, E., Inubushi, Y., Katayama, T., Sawada, K., Yumoto, H., Ohashi, H., Mimura, H., Yabashi, M., Yamauchi, K. & Ishikawa, T. (2014). *Nat. Photon.* **8**, 313–316.

Uhlén, F., Lindqvist, S., Nilsson, D., Reinspach, J., Vogt, U., Hertz, H. M., Holmberg, A. & Barrett, R. (2011). *J. Vac. Sci. Technol. B*, **29**, 06FG03.

Vinko, S. M., Ciricosta, O., Cho, B. I., Engelhorn, K., Chung, H.-K., Brown, C. R. D., Burian, T., Chalupský, J., Falcone, R. W., Graves, C., Hájková, V., Higginbotham, A., Juha, L., Krzywinski, J., Lee, H. J., Messerschmidt, M., Murphy, C. D., Ping, Y., Scherz, A., Schlotter, W., Toleikis, S., Turner, J. J., Vysin, L., Wang, T., Wu, B., Zastrau, U., Zhu, D., Lee, R. W., Heimann, P. A., Nagler, B. & Wark, J. S. (2012). *Nature*, **482**, 59–62.