

pinkIndexer – a universal indexer for pink-beam X-ray and electron diffraction snapshots

Yaroslav Gevorkov,^{a,b*} Anton Barty,^a Wolfgang Brehm,^a Thomas A. White,^a Aleksandra Tolstikova,^{a,c} Max O. Wiedorn,^{a,c,d} Alke Meents,^a Rolf-Rainer Grigat,^b Henry N. Chapman^{a,c,d} and Oleksandr Yefanov^a

Received 2 September 2019

Accepted 18 November 2019

Edited by A. Altomare, Institute of Crystallography - CNR, Bari, Italy

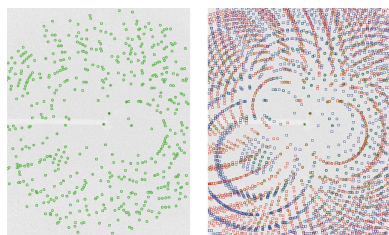
Keywords: indexing; *pinkIndexer*; *CrystFEL*; pink X-ray beam; serial electron diffraction.

^aCenter for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, 22607 Hamburg, Germany, ^bVision Systems, Hamburg University of Technology, 21071 Hamburg, Germany, ^cDepartment of Physics, Universität Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany, and ^dThe Hamburg Center for Ultrafast Imaging, Universität Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. *Correspondence e-mail: yaroslav.gevorkov@desy.de

A crystallographic indexing algorithm, *pinkIndexer*, is presented for the analysis of snapshot diffraction patterns. It can be used in a variety of contexts including measurements made with a monochromatic radiation source, a polychromatic source or with radiation of very short wavelength. As such, the algorithm is particularly suited to automated data processing for two emerging measurement techniques for macromolecular structure determination: serial pink-beam X-ray crystallography and serial electron crystallography, which until now lacked reliable programs for analyzing many individual diffraction patterns from crystals of uncorrelated orientation. The algorithm requires approximate knowledge of the unit-cell parameters of the crystal, but not the wavelengths associated with each Bragg spot. The use of *pinkIndexer* is demonstrated by obtaining 1005 lattices from a published pink-beam serial crystallography data set that had previously yielded 140 indexed lattices. Additionally, in tests on experimental serial crystallography diffraction data recorded with quasi-monochromatic X-rays and with electrons the algorithm indexed more patterns than other programs tested.

1. Introduction

Protein crystallography is a vibrant and continually evolving field spurred by the development of new radiation sources, detectors, measurement techniques and analysis methods. One example is the relatively recent development of serial crystallography using femtosecond-duration X-ray pulses from free-electron lasers (SFX), which is suited to the study of micron-sized and smaller macromolecular crystals (Chapman *et al.*, 2011; Boutet *et al.*, 2012; Schlichting, 2015; Gati *et al.*, 2017). With pulses that out-run atomic motions initiated by photoabsorption, doses may far exceed conventional radiation damage limits to provide structures of radiation-sensitive proteins free of obvious radiation damage, permitting time-resolved studies of biomolecular dynamics at physiologically relevant temperatures (Suga *et al.*, 2014; Tenboer *et al.*, 2014; Kang *et al.*, 2015; Pande *et al.*, 2016; Stagno *et al.*, 2016). The approach of measuring only a single snapshot diffraction pattern from each of many crystals also allows for a lower overall exposure per crystal, and hence lower doses than would be accrued in conventional rotation measurements. This, and the potential for high-throughput measurements, has motivated the development of serial crystallography at



OPEN ACCESS

synchrotron radiation facilities (Stellato *et al.*, 2014; Nogly *et al.*, 2015) and using electron microscopes (Smeets *et al.*, 2018; Bückner *et al.*, 2019).

The speed of serial crystallography measurements, and often the corresponding consumption of sample, is primarily limited by the radiation fluence on the sample and the detector frame rate. At synchrotron radiation sources, higher fluences can be obtained by foregoing the monochromator and using a polychromatic beam. Combined with the enhanced coverage of reciprocal space, a moderate bandwidth of the order of a few per cent (referred to as a 'pink' beam) may offer the additional advantage of fewer necessary diffraction patterns for a serial crystallography measurement (White *et al.*, 2013; Dejoie *et al.*, 2013). For example, Meents *et al.* (2017) demonstrated room-temperature serial crystallography with 100 ps exposure times using the full spectrum of an undulator harmonic (5% relative bandwidth), with the ability to determine structures from as little as 50 indexed diffraction patterns. However, the automated analysis of pink-beam diffraction patterns has been found to be problematic, with only 15% of patterns successfully indexed in the demonstration of Meents *et al.* (2017). We were therefore motivated to create a new robust algorithm to index snapshot diffraction patterns recorded with a quasi-collimated beam of arbitrary bandwidth, with the requirement to index weak or incomplete patterns, using approximately known unit-cell parameters. In meeting this goal, we produced an algorithm '*pinkIndexer*'. We found that *pinkIndexer* can also be applied to several other data collection methods. In addition to superior performance in processing pink-beam diffraction compared with the state-of-the-art algorithms, the algorithm indexes more patterns in monochromatic serial crystallography data sets than all other programs tested, and is successful in indexing snapshot crystal diffraction patterns recorded with electrons.

The determination of the 3D macromolecular crystal structure requires the measurement of diffraction intensities at reciprocal-lattice points throughout a volume of reciprocal space, yet a single snapshot diffraction pattern accesses just a cut through this space. Conservation of photon energy and momentum dictates that for a specific X-ray wavelength this cut is given by a spherical surface – the Ewald sphere – which passes through the origin of reciprocal space. In a serial diffraction experiment the reciprocal-space volume (reduced by the symmetry of the crystal) is sampled by many such patterns recorded from crystals at various random orientations. Usually the orientation of each crystal is initially unknown, and therefore so too is the orientation of its reciprocal lattice. A key analysis step is to identify the crystal orientation, which is equivalent to providing the correct indices to the observed diffraction spots. Furthermore, the distribution of crystal orientations is usually assumed to be random, precluding the use of correlations between successive patterns to deduce the crystal orientations. In the case of a broad-bandwidth X-ray beam, indexing is complicated by the uncertainty of the particular incident wavelength that gave rise to a given Bragg spot, while for electron diffraction the

short wavelength results in an almost flat Ewald sphere for which the determination of unknown 3D lattice parameters is ill conditioned. Indeed, the main bottleneck in pink-beam and electron serial crystallography analysis has been the indexing step.

Automatic indexing algorithms implemented in widely used software including *MOSFLM* (Powell, 1999), *XDS* (Kabsch, 1993, 2010) and *DirAx* (Duisenberg, 1992) were originally devised for data collected in a rotation series with monochromatic radiation. They typically perform poorly when presented with individual pink-beam or electron snapshot diffraction patterns due to their reliance on the particular conditions of monochromatic rotation measurements. Recent algorithms designed for indexing snapshot diffraction patterns encountered in serial crystallography include *TakeTwo* (Ginn *et al.*, 2016), *FELIX* (Beyerlein *et al.*, 2017) and *XGANDALF* (Gevorkov *et al.*, 2019). These all assume monochromatic radiation and do not fare much better than other indexers when processing polychromatic diffraction patterns. Several indexing approaches have been developed for polychromatic crystal diffraction, also referred to as Laue diffraction (Moffat, 1997). These include an approach due to Jacobson (1986) that requires the use of an energy-resolving position-sensitive detector; the Daresbury software suite for indexing Laue patterns (Helliwell *et al.*, 1989; Campbell *et al.*, 1998) and the *Precognition* software (Ren *et al.*, 1999) based on searching arcs of reflections so that prominent zone axes can be identified; geometric approaches of Carr *et al.* (1993) and Wenk *et al.* (1997); and the *LaueUtil* toolkit (Kalinowski *et al.*, 2011) which carries out a clustering analysis of possible orientations that map lattice vectors to observed peaks. The latter algorithm requires measurements of a crystal at several known relative orientations and is therefore not suited to serial crystallography. Of these, the current state-of-the-art software for indexing single wide-bandwidth diffraction patterns of macromolecular crystals is *Precognition*. However, while this works well for patterns recorded with a very wide spectrum (*e.g.* that of a wiggler or bending magnet where the bandwidth is more than 10% of the nominal X-ray energy), it becomes less reliable as the number of Bragg spots decreases as occurs with either reduced spectral width (less than 5% of the nominal X-ray energy) or with small crystals, where only several tens of Bragg reflections are observed.

Here we present the principles and performance of our general indexing algorithm, *pinkIndexer*. As described in Section 2, the algorithm maps observed Bragg reflections into trajectories of possible lattice orientations. The most likely orientation is then determined as the orientation in which most trajectories intersect. As such, *pinkIndexer* covers the cases of monochromatic serial X-ray crystallography, X-ray crystallography using the unmodified spectrum of an undulator of 1% to 25% bandwidth and approximately 1 Å wavelength, and serial electron crystallography at approximately 0.01 Å wavelength. These cases are evaluated in Section 3. The algorithm can be employed in automated processing of serial crystallography data sets, for example using the *CrystFEL* software suite (White *et al.*, 2012; White, 2019).

2. The *pinkIndexer* algorithm

2.1. Diffraction geometry

Consider elastic scattering from an object by a plane monochromatic wave characterized by a wavevector $\mathbf{k}_0$ with a wavelength $\lambda = 1/|\mathbf{k}_0|$. In the kinematic approximation, the strength of scattering in a direction $\mathbf{k}$ is given by the magnitude of the Fourier component of the object at a spatial frequency $\mathbf{q}$ equal to the momentum transfer $\mathbf{k} - \mathbf{k}_0$ (James, 1950). Elastic scattering ($|\mathbf{k}| = |\mathbf{k}_0|$) confines the observable spatial frequencies $\mathbf{q}$ to the Ewald sphere, shown as a circle in Fig. 1. Each pixel in a detector placed in the far field measures a particular direction given by the unit vector $\hat{\mathbf{k}}$, which unambiguously maps to a point $\mathbf{q}$ in reciprocal space. The spatial frequency spectrum of a crystal of infinite extent is a lattice of points that are commonly referred to as reciprocal-lattice points (RLPs), shown as black dots in Fig. 1. As can be seen in that figure, a diffraction spot observed in a particular direction $\hat{\mathbf{k}}$ is unambiguously mapped to a particular RLP (green dot in Fig. 1).

Consider now the case where the radiation source emits a finite but continuous distribution of wavelengths within some known range. Instead of a single Ewald sphere as in the case of monochromatic illumination, each incident wavelength produces an Ewald sphere with a radius inversely proportional to that wavelength. Thus a volume of reciprocal space can be excited in a diffraction experiment, contributing to the 2D diffraction pattern. This volume is depicted in Fig. 2, bounded by the red and blue Ewald spheres (longest to shortest wavelengths in the range). There is a significant difference to the monochromatic case: with a polychromatic source, a particular scattering direction $\hat{\mathbf{k}}$ no longer maps to a single point in reciprocal space. There may be many diffracted wavevectors, each with a different wavelength (and hence different wavevector magnitude and different placement in the Ewald-sphere construction), but pointing in the same direction $\hat{\mathbf{k}}$ and thus arriving at the same point on the detector. These wavevectors are depicted by the red, purple and blue

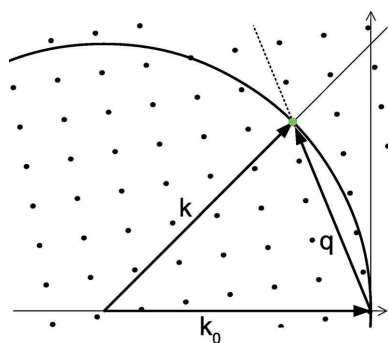


Figure 1

A 2D representation of reciprocal space, showing RLPs of a crystal and the Ewald sphere for monochromatic diffraction. $\mathbf{k}_0$ is the wavevector of the incident beam, $\mathbf{k}$ is the outgoing wavevector elastically scattered from the structure in the object with spatial frequency $\mathbf{q}$. The diffraction intensity measured by a detector in the direction of vector $\mathbf{k}$ corresponds to the intensity at the green RLP.

arrows in Fig. 2. Turning this around, for a given diffraction direction $\hat{\mathbf{k}}$, there are many points $\mathbf{q}$ in reciprocal space that contribute to the diffracted intensity. All these points lie on a straight line segment (green line in Fig. 2), the extension of which passes through the origin of reciprocal space. The line segment can be described by $(1/\lambda)(\hat{\mathbf{k}} - \hat{\mathbf{k}}_0) \mid \lambda \in [\lambda_{\min}, \lambda_{\max}]$.

We therefore see that, in the case of broad bandwidth, a point on the detector integrates signal from a line segment in reciprocal space, in contrast to a single point in the monochromatic case. The RLP which generates a Bragg peak observed at some position on the detector may therefore lie anywhere on the corresponding line segment. We call this line segment the uncertainty line segment (ULS), shown in green in Fig. 2. The main challenge for analyzing broad-bandwidth snapshot crystal diffraction patterns is to determine where along the ULS is the RLP which generated the observed Bragg peak. This is equivalent to identifying the wavelength that excited the measured RLP. Note that if more than one RLP lies on the ULS, they will contribute to the observed intensity, excited by different wavelengths. The bandwidth in that case would be too broad to distinguish those particular reflections in the peak-finding stage without energy-resolving detectors. It is nevertheless possible to separate the summed intensities after indexing and integration (Zurek *et al.*, 1985; Shrive *et al.*, 1990).

Since the crystal orientation is not known, and thus the orientation of the reciprocal lattice is also not known, candidate RLPs may lie anywhere in the volume between $\mathbf{q}$ shells centered at the origin with radii set by the scattering direction and range of wavelengths as depicted by the dashed circles in Fig. 2. We call the RLPs that can match a ULS by rotation of the reciprocal lattice ‘candidate RLPs’ (candidates to predict the particular Bragg spot). The candidate RLPs are plotted in dark green in Fig. 2.

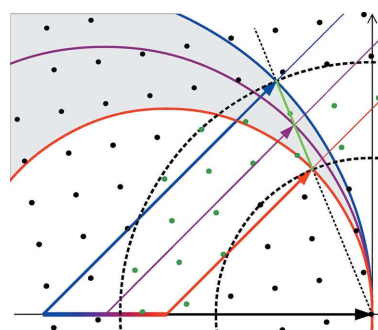


Figure 2

A 2D representation of reciprocal space depicting Ewald spheres for polychromatic diffraction spanning the longest wavelength shown in red to the shortest in blue. The volume between the red and blue spheres is excited and the intensity measured at the detector in the direction of the highlighted vectors corresponds to the integral of diffraction intensities along an uncertainty line segment (ULS) (green line). Candidate RLPs (dark green dots) lie in the volume between shells depicted by the dashed circles. Note that the red, purple and blue vectors are parallel and therefore the corresponding photons arrive at the same pixel of the detector. However, since the photons have different wavelengths, they do not interfere coherently and their intensities are simply added together.

2.2. Determining the crystal orientation

The task of indexing is to find the crystal orientation which gives rise to a particular measured diffraction pattern and then to assign indices to predicted reflection locations. In practice this is achieved by finding the crystal orientation which best predicts the set of Bragg peaks observed on the detector. We assume that the unit-cell parameters of the crystal are known. *pinkIndexer* determines the likely crystal orientation as follows. (i) For each Bragg spot observed on the detector, find all RLPs of a crystal that can be intersected by the Bragg spot's ULS by rotation around the origin (candidate RLPs). (ii) For each observed Bragg spot, find all rotations of the crystal that place at least one candidate RLP onto the corresponding ULS. This is equivalent to finding all orientations of the reciprocal lattice that could predict the measured Bragg spot. (iii) Find the orientation which predicts the most Bragg spots from the list of candidate orientations for all Bragg peaks observed in the pattern. The orientation which correctly predicts the most observed Bragg spots will be the chosen indexing solution. (iv) Refine the lattice parameters and other experimental parameters to further improve the agreement of predicted and observed Bragg peaks [if the original parameters were not accurate, one could repeat steps (i) to (iv) using the refined parameters]. Once the crystal orientation is determined it is of course possible to predict the location and wavelength of all potential reflections including absent or weak reflections not present in the set of observed Bragg peaks. These can then be included in the integration of the observed intensities for structure determination.

The main challenge lies in making the search outlined above tractable and robust. As we will now discuss, for each candidate RLP $\mathbf{h}$ there is an infinite set of reciprocal-lattice rotations which place it onto its particular ULS. We identify all in

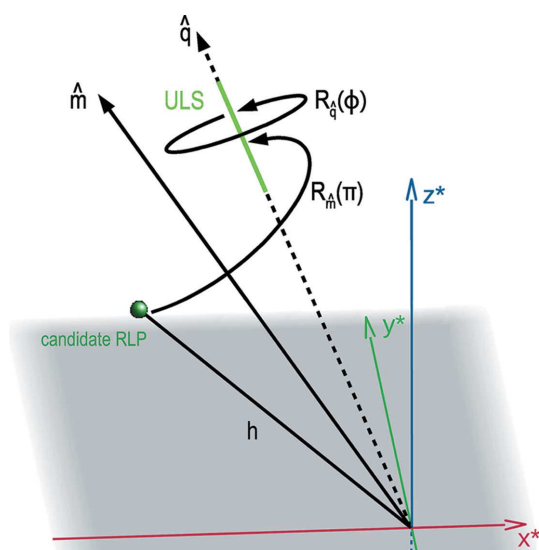


Figure 3

All possible rotations for which a candidate RLP intersects with a given ULS are uniquely described by the rotation $R = R_{\hat{\mathbf{q}}}(\phi)R_{\hat{\mathbf{m}}}(\pi)$. The first rotation $R_{\hat{\mathbf{m}}}(\pi)$ rotates the lattice by the constant angle of π around the bisecting vector $\hat{\mathbf{m}}$ of $\mathbf{h}$ and $\hat{\mathbf{q}}$. The second rotation $R_{\hat{\mathbf{q}}}(\phi)$ rotates the lattice around the axis $\hat{\mathbf{q}}$ of the ULS by an angle ϕ .

this family of rotations by constructing a rotation operation in two steps: first the reciprocal lattice is rotated such that the vector $\mathbf{h}$ of the RLP is rotated by an angle π around the axis $\hat{\mathbf{m}}$ that bisects $\mathbf{h}$ and $\hat{\mathbf{q}}$ as shown in Fig. 3. This rotation brings the candidate RLP onto the ULS. Next, the reciprocal lattice is rotated around $\hat{\mathbf{q}}$ by a rotation of ϕ (see Fig. 3). Since the rotated candidate RLP lies on the ULS it is invariant to the second rotation and thus all rotations ϕ are potential orientations of the lattice. This construction is only valid for one particular candidate RLP and a particular ULS. The particular RLP might not actually give rise to the Bragg spot which generated the ULS. That is, none of the orientations of the lattice generated by the operations $R_{\hat{\mathbf{q}}}(\phi)R_{\hat{\mathbf{m}}}(\pi)$ might be the indexing solution. In a triclinic lattice only one candidate RLP, as well as other RLPs lying on the same straight line through the origin, generate the correct indexing solution. In lattices with higher symmetry, multiple candidate RLPs generate correct indexing solutions.

To determine common orientations that bring a number of RLPs onto ULSs we define a 3D vector space that contains curves parameterized by $\hat{\mathbf{m}}$, $\hat{\mathbf{q}}$ and the rotation angle ϕ , satisfying the reflection conditions stated above that place a particular candidate RLP onto a particular ULS (corresponding to an observed Bragg spot). The vector space consists of 3D since it is spanned by three variables describing rotations, such as the three Euler angles. In this 3D space, all candidate RLPs for a particular Bragg spot will form a set of non-intersecting curves. We call this collection a rotogram. By combining rotograms for all measured Bragg spots, a total rotogram for a diffraction pattern is formed, depicted schematically in Fig. 4. The point in a rotogram with the highest density of overlapping curves provides the lattice orientation that predicts most of the observed Bragg spots. This point represents the rotation of the lattice onto that of the measured crystal, i.e. it is the indexing solution. The task of crystal orientation determination is therefore now reduced to one of finding the point in rotation space with the largest number of intersecting lines.

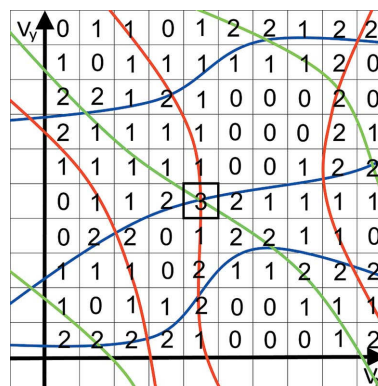


Figure 4

Schematic sketch of a rotogram in a 2D vector space. Each point in the rotogram corresponds to a rotation (defined by the constants v_x and v_y) of a sample reciprocal lattice. Each color marks the positions in the rotogram that predict one particular Bragg spot. The position that is marked by most colors corresponds to the orientation that predicts most Bragg spots correctly – a good indexing solution.

2.3. Algorithm details

In practice, many additional issues arise when dealing with data from a real experiment that complicate the indexing process, such as spurious intensity peaks resulting from experimental noise, or multiple crystals in the beam contributing to the same diffraction pattern. The robustness of the algorithm to these factors becomes a critical issue.

pinkIndexer uses the same basic approach as another indexing method for monochromatic crystal diffraction patterns, *FELIX* (Beyerlein *et al.*, 2017), which similarly parameterizes possible orientations as curves in a 3D rotation space. Both methods are similar to the Hough and Radon transforms that operate on 2D parameter spaces. With such approaches, the choice of the mapping function is crucial for the performance and simplicity of the algorithm. Well-known mappings for 3D rotations to a 3D space are: the Euler-angles representation, the axis-angle representation, the Gibbs representation and the modified Rodrigues parameters (Terzakis *et al.*, 2018). We employ a novel mapping function by which we achieve a drastic reduction of complexity and the number of necessary parameters compared with *FELIX* (which uses Rodrigues parameters), while at the same time increasing the noise tolerance. The following features of the transform are desired for robustness and efficient construction of the rotoqram: (a) adjacent voxels in the rotoqram correspond to similar rotations; (b) rotations are distributed uniformly across a volume of the 3D rotation space; (c) the results of the transformation are efficiently discretized in cuboid samples (*i.e.* on an orthogonal lattice); (d) the transform is calculable in an efficient way.

Since none of the well-known examples sufficiently fulfill these requirements, we propose another transform that better fulfills the requirements and is the major factor in the quality of the *pinkIndexer* algorithm. In our scheme, a single rotation operation $R_{\hat{\mathbf{e}}}(\theta)$ is determined from the composite rotation $R_{\hat{\mathbf{q}}}(\phi)R_{\hat{\mathbf{m}}}(\pi) = R_{\hat{\mathbf{e}}}(\theta)$. This rotation then is mapped to the

point $\mathbf{v}$ in the rotoqram given by $\mathbf{v} = \arctan(\theta/4)\hat{\mathbf{e}}$, where $\hat{\mathbf{e}}$ is the rotation axis, θ is the rotation angle and $\arctan(\theta/4)$ is a nonlinear scaling factor. Compared with the well-known axis-angle representation which maps a rotation to $\theta\hat{\mathbf{e}}$ (*i.e.* the length of the vector encodes the rotation angle), this definition only slightly increases the computational burden and inherits its property of adjacent voxels corresponding to similar rotations. The nonlinear scaling of rotation angle to the length of the vector gives a more uniform distribution of points in the rotoqram than the axis-angle representation. Our transform maps all possible rotations to a finite-size ball of radius $\arctan(\pi/4)$ and is in some sense the opposite of the modified Rodrigues parameters, $\tan(\theta/4)\hat{\mathbf{e}}$.

The construction of $R_{\hat{\mathbf{e}}}(\theta)$ from $R_{\hat{\mathbf{q}}}(\phi)R_{\hat{\mathbf{m}}}(\pi)$ is achieved in a computationally inexpensive way by employing the composition law for finite rotations first derived by Olinde Rodrigues (Altmann, 1989; Pujol, 2013). This describes the consecutive operations of two general rotations $R_{\hat{\mathbf{m}}}(\psi)$ and $R_{\hat{\mathbf{q}}}(\phi)$ to give $R_{\hat{\mathbf{e}}}(\theta) = R_{\hat{\mathbf{q}}}(\phi)R_{\hat{\mathbf{m}}}(\psi)$ by solving

$$\begin{aligned}\cos\frac{\theta}{2} &= \cos\frac{\phi}{2}\cos\frac{\psi}{2} - \sin\frac{\phi}{2}\sin\frac{\psi}{2}\hat{\mathbf{q}} \cdot \hat{\mathbf{m}} \\ \sin\frac{\theta}{2}\hat{\mathbf{e}} &= \sin\frac{\phi}{2}\cos\frac{\psi}{2}\hat{\mathbf{q}} - \cos\frac{\phi}{2}\sin\frac{\psi}{2}\hat{\mathbf{m}} + \sin\frac{\phi}{2}\sin\frac{\psi}{2}\hat{\mathbf{q}} \times \hat{\mathbf{m}}.\end{aligned}\quad (1)$$

For our problem, the first rotation axis is the bisector $\hat{\mathbf{m}} = (\hat{\mathbf{h}} + \hat{\mathbf{q}})/\|\hat{\mathbf{h}} + \hat{\mathbf{q}}\|_2$, and $\hat{\mathbf{q}}$ is the direction of the ULS. This choice of $\hat{\mathbf{m}}$ allows setting $\psi = \pi$ such that the equations simplify to

$$\begin{aligned}\cos\frac{\theta}{2} &= -\sin\frac{\phi}{2}\hat{\mathbf{q}} \cdot \hat{\mathbf{m}} \\ \sin\frac{\theta}{2}\hat{\mathbf{e}} &= -\cos\frac{\phi}{2}\hat{\mathbf{m}} + \sin\frac{\phi}{2}\hat{\mathbf{q}} \times \hat{\mathbf{m}}.\end{aligned}\quad (2)$$

Setting the parameters $c_1 = \sin(\phi/2)$, $c_2 = -\cos(\phi/2)$, $\mathbf{d}_1 = \hat{\mathbf{q}} \cdot \hat{\mathbf{m}}$ and $\mathbf{d}_2 = \hat{\mathbf{q}} \times \hat{\mathbf{m}}$ we obtain

$$\begin{aligned}\cos\frac{\theta}{2} &= -c_1 d_1 \\ \sin\frac{\theta}{2}\hat{\mathbf{e}} &= c_2 \hat{\mathbf{m}} + c_1 \mathbf{d}_2,\end{aligned}\quad (3)$$

which can be solved as

$$\begin{aligned}\theta &= 2 \arccos(-c_1 d_1) \\ \hat{\mathbf{e}} &= \frac{c_2 \hat{\mathbf{m}} + c_1 \mathbf{d}_2}{\sin(\theta/2)}.\end{aligned}\quad (4)$$

For machines where $1/x^{1/2}$ is implemented in hardware, replacing $1/\sin(\theta/2)$ by $1/[1 - (c_1 d_1)^2]^{1/2}$ can lead to faster execution.

An example of a rotoqram for a particular Bragg peak is shown in Fig. 5, for 56 candidate RLPs on a cubic lattice and a relative bandwidth of 6.5%. Each of the non-intersecting 56 colored curves is a plot of the vector $\mathbf{v} = \arctan(\theta/4)\hat{\mathbf{e}}$ for a full rotation of the lattice. For a given point $\mathbf{v}$ in the plot, the corresponding rotation angle of the lattice to bring the RLP onto the ULS defined by $\hat{\mathbf{q}}$ is $\theta = 4 \tan(\|\mathbf{v}\|_2)$, and the rotation axis is $\hat{\mathbf{v}}$. As seen from equation (2), each trajectory lies in the

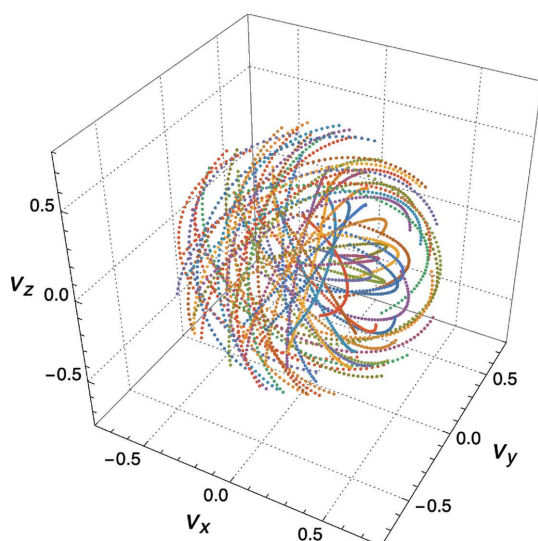


Figure 5
Plot of the curves $\mathbf{v} = \arctan(\theta/4)\hat{\mathbf{e}}$ for a particular Bragg-point direction $\hat{\mathbf{q}}$ and a set of 56 candidate RLPs in a cubic lattice.

plane containing the orthogonal vectors $\hat{\mathbf{m}}$ and $\hat{\mathbf{q}} \times \hat{\mathbf{m}}$, which is to say the plane normal to $\hat{\mathbf{m}} \times \hat{\mathbf{q}} \times \hat{\mathbf{m}} = \hat{\mathbf{q}} - d_1 \hat{\mathbf{m}}$. The trajectories form closed curves in the vector space over the range of ϕ from 0 to 4π , but we only require a single rotation of the lattice. To keep the rotogram volume as small as possible we choose the range $\phi \in [-2\pi, 0)$. In the example of Fig. 5 the curves were sampled over that range at steps of 0.1 rad and while the curves are not necessarily uniformly sampled in the vector space, the choice of $\mathbf{v} = \arctan(\theta/4) \hat{\mathbf{e}}$ generates curves that more uniformly fill the space than the axis-angle representation $\theta \hat{\mathbf{e}}$ or any other construction that we tried.

In the implementation of *pinkIndexer*, rotograms are not calculated continuously as shown in Fig. 5 but computed on discrete sets of $N \times N \times N$ voxels that circumscribe the ball of radius $\arctan(\pi/4)$. For each Bragg spot the voxel array is initialized with zeros and voxels are set to 1 that are intersected by the curve $\mathbf{v} = \arctan(\theta/4) \hat{\mathbf{e}}$ for each of the candidate RLPs, with a uniform sampling of ϕ that is chosen to ensure that the curve is contiguous across the voxels. This is accomplished by computing the parameters $c_1 = \sin(\phi/2)$ and $c_2 = -\cos(\phi/2)$ at those values once for the whole rotogram and using values of the parameters $d_1 = \hat{\mathbf{q}} \cdot \hat{\mathbf{m}}$ and $d_2 = \hat{\mathbf{q}} \times \hat{\mathbf{m}}$ that need to be computed once per curve. To make the discretization of ϕ smoother, the flagged voxels are dilated by setting all of their 26 neighboring voxels to 1. This reduces the effective resolution of the rotogram, but increases the noise tolerance. The rotogram indicates all orientations of the crystal that predict the respective Bragg spot.

By adding each Bragg spot's rotogram, a total rotogram is created where the value of each voxel gives the number of Bragg spots predicted by the corresponding orientation. The voxel with the maximum value thus indicates the most likely lattice orientation that provides the correct indexing solution. The task of indexing is thus reduced to finding the location of this maximum. Since the rotogram is discrete the determined indexing solution is approximate. A subsequent refinement is carried out to increase the precision of the indexing solution, in which the lattice basis is refined to minimize the mean Euclidean distance between the ULS and the respective closest RLP using a gradient descent approach. Only the RLPs close enough to a ULS are used for this refinement to improve noise tolerance.

2.4. Implementation details

The algorithm has been implemented such that parameters have effects that are easy to understand. Besides experimental settings like detector distance and pixel geometry, beam parameters and crystal lattice parameters, *pinkIndexer* requires a relative tolerance, set by the parameter *tolerance*, to decide when a peak is correctly fitted. Additional parameters trade-off fitting performance against execution time. The parameter *consideredPeaksCount* specifies the number of found Bragg spots that are used to compute the initial indexing solution from the maximum of the rotogram. All Bragg spots are considered during refinement. The parameter

angleResolution sets the resolution of the rotogram in terms of number of voxels N spanning $-\arctan(\pi/4)$ to $\arctan(\pi/4)$. Choosing larger voxels (lower resolution) leads to a faster calculation but lower precision in the initial step of determining the orientation from the rotogram. The second step of refining the orientation is controlled by the parameter *refinementType*. Refinement can be performed by a gradient descent method, fitting all parameters of the lattice or keeping the cell parameters constant and just refining the orientation. All parameters take descriptive values wherever possible.

3. Evaluation of algorithm performance

We evaluated the performance of the *pinkIndexer* algorithm on data from macromolecular crystal diffraction experiments utilizing three different types of radiation: monochromatic X-rays, pink X-rays and electrons. For the evaluation we used the *CrystFEL* (White *et al.*, 2012) software suite 0.8.0+50a3cb06 with modifications to include the *pinkIndexer* library and enable prediction for wide-bandwidth and electron beams.

3.1. Monochromatic X-ray beam crystallography

The performance of *pinkIndexer* in treating monochromatic serial femtosecond X-ray diffraction data was compared with the indexers *MOSFLM* (Powell, 1999), *XDS* (Kabsch, 1993, 2010), *DirAx* (Duisenberg, 1992), *TakeTwo* (Ginn *et al.*, 2016), *FELIX* (Beyerlein *et al.*, 2017) and *XGANDALF* (Gevorkov *et al.*, 2019) using the *indexamajig* program from the *CrystFEL* (White *et al.*, 2012) software suite. For the test, all *CrystFEL* optimizations were turned off by using the options `--no-retry --no-refine --no-multi --no-check-cell --no-check-peaks`. Only one indexing solution per pattern was accepted. Indexing solutions that differed from the original indexing solution by less than 3° were counted as correct. The diffraction data set was retrieved from the CXIDB (Maia, 2012), entry 21, from SFX measurements of a G-protein-coupled receptor (the serotonin 5-HT_{2B} receptor bound to ergotamine) (Liu *et al.*, 2013).

Comparing algorithms using real data provides results that indicate their performance under real conditions. However, unlike when using simulated data, the true indexing solution is unknown. The indexing solutions can be tested for correctness by comparing the predicted Bragg spots with the found ones. This is a precise method when there are many Bragg spots, but when the number of found spots is small there can be several incorrect orientations of a crystal that predict the found spots well enough to pass the indexing test. Following a practice we introduced earlier to compare indexing algorithms (Gevorkov *et al.*, 2019) we created semi-simulated data sets with different numbers of Bragg spots by removing spots from patterns with large numbers of spots (which have reliable indexing solutions). As previously, we tested the indexers in two modes of Bragg-peak removal. In one mode the sets contained patterns with only five to 50 Bragg spots selected randomly from the patterns. In the other mode the sets of patterns contained five

to 50 Bragg spots only at low resolution. The comparison is given in Fig. 6. All algorithms performed well when there were sufficient measured Bragg spots to determine the crystal lattice. With both cases of randomly distributed Bragg spots and low-resolution Bragg spots, the *pinkIndexer* algorithm outperformed all others over the whole range of Bragg-spot counts. The settings of *pinkIndexer* in these tests were chosen to favor precision over speed (*angleResolution* = 'dense'). In all cases the lattice parameters were specified to the indexing algorithm. No additional tuning of the indexing algorithms was performed apart from an option that allows *FELIX* to index patterns with as few as five peaks.

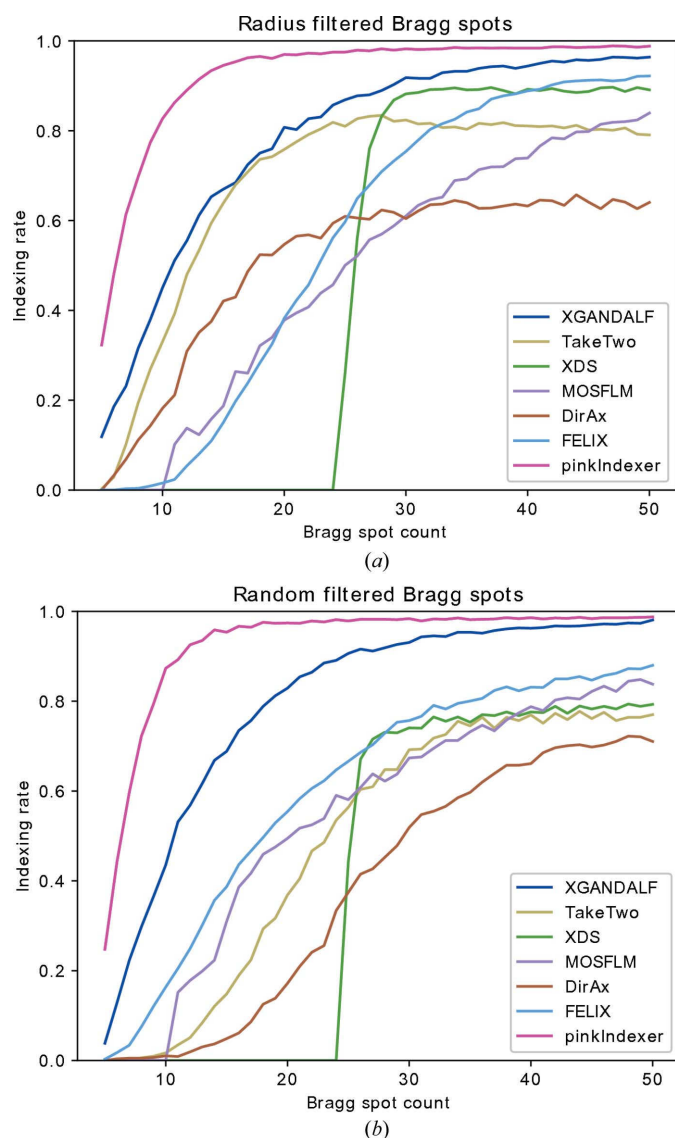


Figure 6
Comparison of *pinkIndexer* with other indexing algorithms, using the *CrystFEL* software suite. Each original experimental diffraction pattern had more than 50 Bragg spots and was indexable with *MOSFLM*. From every pattern, a number of Bragg peaks were selected to be used for indexing (indicated along the abscissa) either (a) by selecting those with the smallest scattering angles or (b) randomly. The indexing results were compared with the original *MOSFLM* indexing solution using all Bragg spots. Indexing solutions that differed from the original indexing solution by less than 3° were counted as correct.

Table 1

Indexing results and execution times for monochromatic indexers and *pinkIndexer*.

Indexer name	Indexed patterns	Total execution time [(mm:ss)/per pattern (ms)]
<i>MOSFLM</i>	452	00:17/17
<i>XDS</i>	400	00:22/22
<i>DirAx</i>	394	00:12/12
<i>TakeTwo</i>	545	11:02/662
<i>FELIX</i>	656	01:05/65
<i>XGANDALF</i>	724	00:19/19
<i>pinkIndexer</i> fast mode	757	04:15/255
<i>pinkIndexer</i> precise mode	816	22:16/1336

The average times for the various algorithms to index monochromatic diffraction patterns are given in Table 1, computed by indexing a set of 1000 diffraction patterns chosen randomly from the same data set as used above. To ensure a fair comparison, all indexers were called from *CrystFEL* with the `--no-retry` flag set. No attempt was made to index multiple crystals per pattern. The program was executed on a dual-socket Intel Xeon E5-2698 v4 CPU (2.20 MHz, 20 cores, 50 MB cache, 512 GB RAM). *pinkIndexer* was tested with settings to maximize the speed ('fast mode' in Table 1, *angleResolution* = 'loose') and with settings to maximize the yield ('precise mode', *angleResolution* = 'dense') which took about five times longer. Settings in-between are also possible. Even in the fast mode the algorithm takes considerably longer than other algorithms except for *TakeTwo*. The slower speed of *pinkIndexer* is because the algorithm is memory intense. Nevertheless, due to its high indexing success rate, *pinkIndexer* can be profitably used as a fallback option for monochromatic diffraction patterns that cannot be indexed by other indexers. This can be implemented in *CrystFEL* by placing *pinkIndexer* last in the list of indexers.

3.2. Pink-beam serial crystallography

Diffraction patterns collected using pink-beam radiation (1% to 25% relative bandwidth) contain many observed peaks, which means that indexing solutions can easily be verified by comparing the predicted with the observed Bragg spots. To evaluate *pinkIndexer* using real pink-beam serial crystallography data we used the data set of proteinase K crystal diffraction from Meents *et al.* (2017) measured at the 14-ID-B (BioCARS) beamline at the Advanced Photon Source (APS), using the full polychromatic spectrum of an undulator harmonic. A representative diffraction pattern is depicted in Fig. 7. Although the FWHM of the incident X-ray beam spectrum was 5% of the mean photon energy, the tails of the spectrum extended up to 25% of the mean photon energy. The data set contained 999 patterns that had been classified as crystal diffraction 'hits' based on the detection of at least 35 Bragg spots in the original work of Meents *et al.* (2017). Of these, 667 patterns were successfully indexed by *pinkIndexer*, with 428 determined to contain a single lattice, 168 with two lattices, and 71 with three or more lattices. This gave a total of 1005 indexed lattices. A vast majority of the 332 patterns that

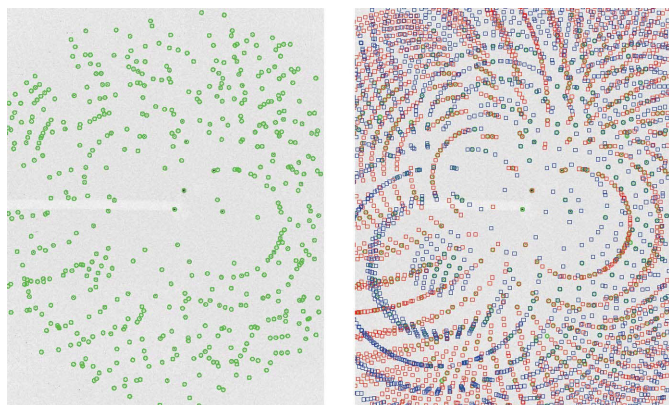


Figure 7

A snapshot pink-beam diffraction pattern of two proteinase K crystals measured coincidentally at the 14-ID-B (BioCARS) beamline at the APS. The pattern is overlaid with the detected peaks (left) and with detected as well as predicted peaks (right). Both crystals were indexed by *pinkIndexer* (red and blue). Note the agreement between prediction and measurement.

could not be indexed appeared to be falsely identified as crystal diffraction patterns due to fitting peaks to noise. The *pinkIndexer* parameters used in this test are given in Appendix A.

The comparison of polychromatic indexing programs in an automated way for serial crystallography data sets is compli-

cated by these programs not being able to be called from within the *CrystFEL* software suite. Indeed, the analysis of the pink-beam serial crystallography data carried out by Meents *et al.* (2017) could only be achieved in a semi-manual way using the software *Precognition*. This resulted in the indexing of 140 patterns of the 999 (Meents *et al.*, 2017), and only single-crystal diffraction patterns could be indexed. This comparison shows that, for serial crystallography, *pinkIndexer* provides an order of magnitude more indexable patterns than the current state-of-the-art software. Moreover, *pinkIndexer* can deal with multiple crystals per pattern and is fully automatic, thus making serial crystallography with a pink beam much easier. Fig. 7 shows two crystals contributing to the pattern which are both indexed correctly. We have also successfully used *pinkIndexer* for serial crystallography data from Tolstikova *et al.* (2019) measured with X-rays with 2.5% relative bandwidth produced using a multilayer monochromator.

3.3. Serial electron crystallography

Electron crystallography poses a challenge for conventional indexing algorithms due to flatness of the Ewald sphere caused by the short de Broglie wavelength of the electrons. To demonstrate the applicability of *pinkIndexer* to serial electron diffraction data, a rotation series data set from Cruz *et al.* (2017) was treated like serial data by indexing each pattern

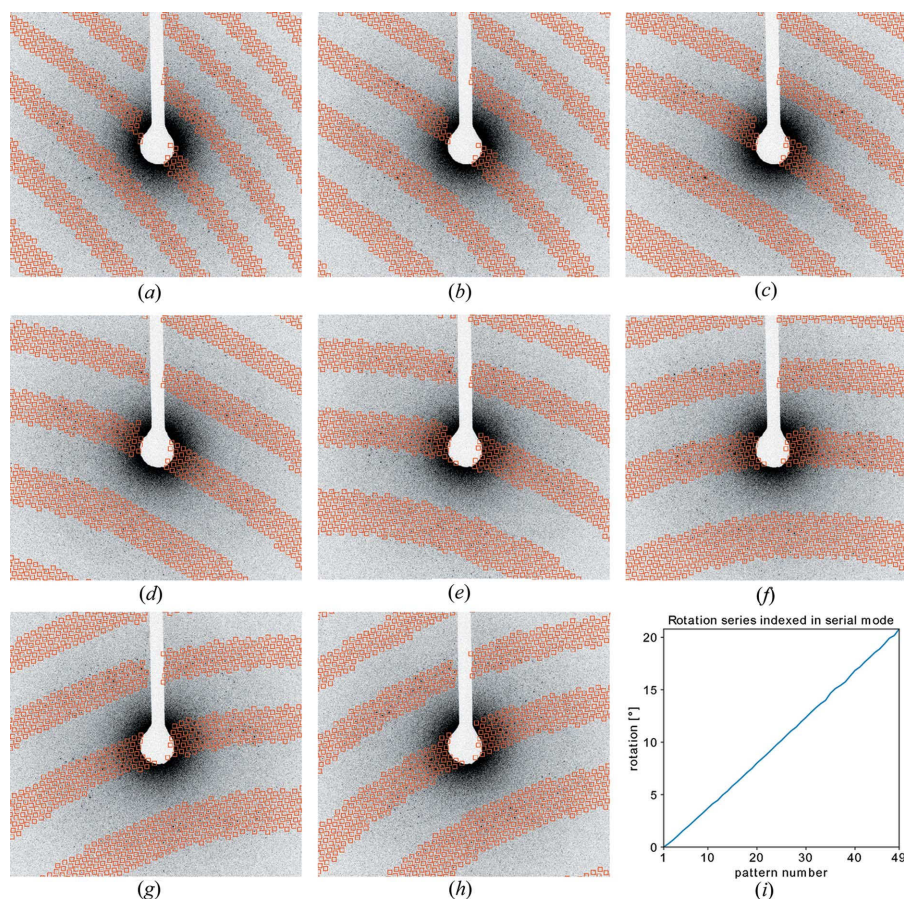


Figure 8

(a)–(h) Patterns 30–37 from the electron diffraction rotation series data of proteinase K in Cruz *et al.* (2017). The red squares mark the locations of the predicted Bragg spots after indexing. (i) Plot of the crystal rotation angle as derived from the indexing result of *pinkIndexer*.

individually. The known rotation increment available for this data set was used to check the correctness of the indexing solutions. The results are displayed in Fig. 8. All patterns from the data set were indexed correctly, as can be seen from the linear increment of the determined rotation angle. The maximum deviation of the angle determined by indexing from a linear fit was 0.14° , which represents an upper bound to the indexing precision since goniometer errors may also contribute. This result opens up the possibility of serial electron crystallography using randomly oriented crystals exposed in individual data frames as performed in SFX measurements.

4. Conclusion

The indexer presented in this paper has been developed for pink-beam serial crystallography using the full polychromatic spectrum of an undulator harmonic at a synchrotron radiation facility. Starting from known unit-cell parameters, the *pinkIndexer* algorithm works by mapping all possible rotations of candidate reciprocal-lattice points onto line segments in reciprocal space that correspond to Bragg peaks of unknown wavelength. By examining these mappings in a novel rotation space the most likely lattice orientation can be found. The main limitation of *pinkIndexer* is its slower speed compared with many other existing algorithms for monochromatic diffraction analysis and the requirement of knowing the cell parameters of the studied crystals. The benefit, however, is its higher success rate in indexing snapshot diffraction patterns than all other algorithms we tested.

Due to the generality of the approach to different wavelengths and spectral characteristics, the algorithm presented here has the ability to open up emerging and as-yet-unexplored avenues of serial crystallography. The higher X-ray flux of the polychromatic beam enables exposure times as short as those emitted from a single electron bunch in the storage ring, while the broad bandwidth leads to a high fraction of fully integrated Bragg peaks recorded in a snapshot pattern and a greater coverage of reciprocal space. These advantages have long been appreciated for time-resolved Laue diffraction experiments at synchrotron facilities (Moffat, 1997), macromolecular crystallography at neutron facilities (Blakeley *et al.*, 2008), and have motivated the generation of pulses with broader bandwidth at free-electron laser facilities (White *et al.*, 2013; Dejoie *et al.*, 2013). As demonstrated here, the *pinkIndexer* program overcomes difficulties previously encountered in automatically analyzing thousands of polychromatic diffraction patterns. Additionally, the generality of the algorithm makes it useful for indexing monochromatic serial crystallography. In this case we found that *pinkIndexer* demonstrates a superior success rate in indexing diffraction patterns, especially for the tricky case of a small number of detected Bragg spots. We also showed that the approach works well for indexing snapshot diffraction patterns recorded with very short wavelengths, which is usually the situation in electron diffraction. The method might additionally find application in neutron diffraction and could be slightly modified to treat the case of convergent-beam diffraction.

5. Code availability

pinkIndexer is implemented in C++ and released as an open-source library under the LGPLv3 licence. This library can be compiled independently or together with the program suite *CrystFEL* (White *et al.*, 2012). The full processing pipeline, including indexing, high-precision prediction and integration, will be realized soon as a part of *CrystFEL*. The source code can be downloaded at <https://stash.desy.de/users/gevorkov/repos/pinkindexer/browse>.

APPENDIX A

pinkIndexer settings

The *pinkIndexer* settings for Section 3.2 were chosen as follows (format as used by *CrystFEL*):

```
photon_energy = 11000,
photon_energy_bandwidth = 0.25,
--pinkIndexer-considered-peaks-count=4,
--pinkIndexer-angle-resolution=3,
--pinkIndexer-refinement-type=4,
--pinkIndexer-tolerance=0.03.
```

Unit cell:

```
lattice_type = tetragonal,
unique_axis = c,
centering = P,
a = 69.10 Å,
b = 69.10 Å,
c = 106.60 Å,
a1 = 90.00 deg,
be = 90.00 deg,
ga = 90.00 deg.
```

pinkIndexer settings for Section 3.1 were chosen as follows (format as used by *CrystFEL*):

```
--pinkIndexer-tolerance=0.13,
--pinkIndexer-angle-resolution=3.
```

The *XGANDALF* settings for Section 3 were chosen as follows (format as used by *CrystFEL*):

```
--xgandalf-sampling-pitch=2,
--xgandalf-grad-desc-iterations=2,
--xgandalf-tolerance=0.02.
```

The *FELIX* settings for Section 3 were chosen as follows (format as used by *CrystFEL*):

```
--felix-max-visits=5.
```

Acknowledgements

We thank Vukica Šrajer for discussions about pink-beam indexing and Robert Bückner, Pascal Hogan-Lamarre and Pedram Mehrabi for discussions and collaboration on indexing serial electron diffraction.

Funding information

We acknowledge support from the Cluster of Excellence 'Advanced Imaging of Matter' of the Deutsche Forschungsgemeinschaft (DFG) – EXC 2056 – project ID 390715994; the 'X-probe' project funded by the European Union's 2020

Research and Innovation Program under the Marie Skłodowska-Curie Grant Agreement 637295; and the European Research Council, 'Frontiers in Attosecond X-ray Science: Imaging and Spectroscopy (AXSIS)', ERC-2013-SyG 609920 (2014–2018).

References

- Altmann, S. L. (1989). *Math. Mag.* **62**, 291–308.
- Beyerlein, K. R., White, T. A., Yefanov, O., Gati, C., Kazantsev, I. G., Nielsen, N. F.-G., Larsen, P. M., Chapman, H. N. & Schmidt, S. (2017). *J. Appl. Cryst.* **50**, 1075–1083.
- Blakeley, M. P., Langan, P., Niimura, N. & Podjarny, A. (2008). *Curr. Opin. Struct. Biol.* **18**, 593–600.
- Boutet, S., Lomb, L., Williams, G. J., Barends, T. R. M., Aquila, A., Doak, R. B., Weierstall, U., DePonte, D. P., Steinbrener, J., Shoeman, R. L., Messerschmidt, M., Barty, A., White, T. A., Kassemeyer, S., Kirian, R. A., Seibert, M. M., Montanez, P. A., Kenney, C., Herbst, R., Hart, P., Pines, J., Haller, G., Gruner, S. M., Philipp, H. T., Tate, M. W., Hromalik, M., Koerner, L. J., van Bakel, N., Morse, J., Ghonsalves, W., Arnlund, D., Bogan, M. J., Coleman, C., Fromme, R., Hampton, C. Y., Hunter, M. S., Johansson, L. C., Katona, G., Kupitz, C., Liang, M., Martin, A. V., Nass, K., Redecke, L., Stellato, F., Timneanu, N., Wang, D., Zatsepin, N. A., Schafer, D., Defever, J., Neutze, R., Fromme, P., Spence, J. C. H., Chapman, H. N. & Schlichting, I. (2012). *Science*, **337**, 362–364.
- Bücker, R., Hogan-Lamarre, P., Mehrabi, P., Schulz, E. C., Bultema, L. A., Gevorkov, Y., Brehm, W., Yefanov, O., Oberthür, D., Kassier, G. H. & Miller, R. J. D. (2019). <https://doi.org/10.1101/682575>.
- Campbell, J. W., Hao, Q., Harding, M. M., Nguti, N. D. & Wilkinson, C. (1998). *J. Appl. Cryst.* **31**, 496–502.
- Carr, P. D., Dodd, I. M. & Harding, M. M. (1993). *J. Appl. Cryst.* **26**, 384–387.
- Chapman, H. N., Fromme, P., Barty, A., White, T. A., Kirian, R. A., Aquila, A., Hunter, M. S., Schulz, J., DePonte, D. P., Weierstall, U., Doak, R. B., Maia, F. R. N. C., Martin, A. V., Schlichting, I., Lomb, L., Coppola, N., Shoeman, R. L., Epp, S. W., Hartmann, R., Rolles, D., Rudenko, A., Foucar, L., Kimmel, N., Weidenspointner, G., Holl, P., Liang, M., Barthelmess, M., Coleman, C., Boutet, S., Bogan, M. J., Krzywinski, J., Bostedt, C., Bajt, S., Gumprecht, L., Rudek, B., Erk, B., Schmidt, C., Hömke, A., Reich, C., Pietschner, D., Strüder, L., Hauser, G., Gork, H., Ullrich, J., Herrmann, S., Schaller, G., Schopper, F., Soltan, H., Kühnel, K.-U., Messerschmidt, M., Bozek, J. D., Hau-Riege, S. P., Frank, M., Hampton, C. Y., Sierra, R. G., Starodub, D., Williams, G. J., Hajdu, J., Timneanu, N., Seibert, M. M., Andreasson, J., Rocker, A., Jönsson, O., Svenda, M., Stern, S., Nass, K., Andritschke, R., Schröter, C.-D., Krasniqi, F., Bott, M., Schmidt, K. E., Wang, X., Grotjohann, I., Holton, J. M., Barends, T. R. M., Neutze, R., Marchesini, S., Fromme, R., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Andersson, I., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B. & Spence, J. C. H. (2011). *Nature*, **470**, 73–77.
- Cruz, M. J. de la, Hattne, J., Shi, D., Seidler, P., Rodriguez, J., Reyes, F. E., Sawaya, M. R., Cascio, D., Weiss, S. C., Kim, S. K., Hinck, C. S., Hinck, A. P., Calero, G., Eisenberg, D. & Gonen, T. (2017). *Nat. Methods*, **14**, 399–402.
- Dejoie, C., McCusker, L. B., Baerlocher, C., Abela, R., Patterson, B. D., Kunz, M. & Tamura, N. (2013). *J. Appl. Cryst.* **46**, 791–794.
- Duisenberg, A. J. M. (1992). *J. Appl. Cryst.* **25**, 92–96.
- Gati, C., Oberthuer, D., Yefanov, O., Bunker, R. D., Stellato, F., Chiu, E., Yeh, S.-M., Aquila, A., Basu, S., Bean, R., Beyerlein, K. R., Botha, S., Boutet, S., DePonte, D. P., Doak, R. B., Fromme, R., Galli, L., Grotjohann, I., James, D. R., Kupitz, C., Lomb, L., Messerschmidt, M., Nass, K., Rendek, K., Shoeman, R. L., Wang, D., Weierstall, U., White, T. A., Williams, G. J., Zatsepin, N. A., Fromme, P., Spence, J. C. H., Goldie, K. N., Jehle, J. A., Metcalf, P., Barty, A. & Chapman, H. N. (2017). *Proc. Natl Acad. Sci. USA*, **114**, 2247–2252.
- Gevorkov, Y., Yefanov, O., Barty, A., White, T. A., Mariani, V., Brehm, W., Tolstikova, A., Grigat, R.-R. & Chapman, H. N. (2019). *Acta Cryst. A* **75**, 694–704.
- Ginn, H. M., Roedig, P., Kuo, A., Evans, G., Sauter, N. K., Ernst, O. P., Meents, A., Mueller-Werkmeister, H., Miller, R. J. D. & Stuart, D. I. (2016). *Acta Cryst. D* **72**, 956–965.
- Helliwell, J. R., Habash, J., Cruickshank, D. W. J., Harding, M. M., Greenhough, T. J., Campbell, J. W., Clifton, I. J., Elder, M., Machin, P. A., Papiz, M. Z. & Zurek, S. (1989). *J. Appl. Cryst.* **22**, 483–497.
- Jacobson, R. A. (1986). *J. Appl. Cryst.* **19**, 283–286.
- James, R. (1950). *The Optical Principles of the Diffraction of X-rays*. London: Bell.
- Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
- Kabsch, W. (2010). *Acta Cryst. D* **66**, 125–132.
- Kalinowski, J. A., Makal, A. & Coppens, P. (2011). *J. Appl. Cryst.* **44**, 1182–1189.
- Kang, Y., Zhou, X. E., Gao, X., He, Y., Liu, W., Ishchenko, A., Barty, A., White, T. A., Yefanov, O., Han, G. W., Xu, Q., de Waal, P. W., Ke, J., Tan, M. H. E., Zhang, C., Moeller, A., West, G. M., Pascal, B. D., Van Eps, N., Caro, L. N., Vishnivetskii, S. A., Lee, R. J., Suino-Powell, K. M., Gu, X., Pal, K., Ma, J., Zhi, X., Boutet, S., Williams, G. J., Messerschmidt, M., Gati, C., Zatsepin, N. A., Wang, D., James, D., Basu, S., Roy-Chowdhury, S., Conrad, C. E., Coe, J., Liu, H., Lisova, S., Kupitz, C., Grotjohann, I., Fromme, R., Jiang, Y., Tan, M., Yang, H., Li, J., Wang, M., Zheng, Z., Li, D., Howe, N., Zhao, Y., Standfuss, J., Diederichs, K., Dong, Y., Potter, C. S., Carragher, B., Caffrey, M., Jiang, H., Chapman, H. N., Spence, J. C. H., Fromme, P., Weierstall, U., Ernst, O. P., Katritch, V., Gurevich, V. V., Griffin, P. R., Hubbell, W. L., Stevens, R. C., Cherezov, V., Melcher, K. & Xu, H. E. (2015). *Nature*, **523**, 561–567.
- Liu, W., Wacker, D., Gati, C., Han, G. W., James, D., Wang, D., Nelson, G., Weierstall, U., Katritch, V., Barty, A., Zatsepin, N. A., Li, D., Messerschmidt, M., Boutet, S., Williams, G. J., Koglin, J. E., Seibert, M. M., Wang, C., Shah, S. T. A., Basu, S., Fromme, R., Kupitz, C., Rendek, K. N., Grotjohann, I., Fromme, P., Kirian, R. A., Beyerlein, K. R., White, T. A., Chapman, H. N., Caffrey, M., Spence, J. C. H., Stevens, R. C. & Cherezov, V. (2013). *Science*, **342**, 1521–1524.
- Maia, F. R. N. C. (2012). *Nat. Methods*, **9**, 854–855.
- Meents, A., Wiedorn, M. O., Srajer, V., Henning, R., Sarrou, I., Bergtholdt, J., Barthelmess, M., Reinke, P. Y. A., Dierksmeyer, D., Tolstikova, A., Schaible, S., Messerschmidt, M., Ogata, C. M., Kissick, D. J., Taft, M. H., Manstein, D. J., Lieske, J., Oberthuer, D., Fischetti, R. F. & Chapman, H. N. (2017). *Nat. Commun.* **8**, 1281.
- Moffat, K. (1997). *Methods Enzymol.* **277**, 433–447.
- Nogly, P., James, D., Wang, D., White, T. A., Zatsepin, N., Shilova, A., Nelson, G., Liu, H., Johansson, L., Heymann, M., Jaeger, K., Metz, M., Wickstrand, C., Wu, W., Bähr, P., Berntsen, P., Oberthuer, D., Panneels, V., Cherezov, V., Chapman, H., Schertler, G., Neutze, R., Spence, J., Moraes, I., Burghammer, M., Standfuss, J. & Weierstall, U. (2015). *IUCrJ*, **2**, 168–176.
- Pande, K., Hutchison, C. D. M., Groenhof, G., Aquila, A., Robinson, J. S., Tenboer, J., Basu, S., Boutet, S., DePonte, D. P., Liang, M., White, T. A., Zatsepin, N. A., Yefanov, O., Morozov, D., Oberthuer, D., Gati, C., Subramanian, G., James, D., Zhao, Y., Koralek, J., Brayshaw, J., Kupitz, C., Conrad, C., Roy-Chowdhury, S., Coe, J. D., Metz, M., Xavier, P. L., Grant, T. D., Koglin, J. E., Ketawala, G., Fromme, R., Rajer, V., Henning, R., Spence, J. C. H., Ourmazd, A., Schwander, P., Weierstall, U., Frank, M., Fromme, P., Barty, A., Chapman, H. N., Moffat, K., van Thor, J. J. & Schmidt, M. (2016). *Science*, **352**, 725–729.
- Powell, H. R. (1999). *Acta Cryst. D* **55**, 1690–1695.
- Pujol, J. (2013). *Appl. Mech. Rev.* **65**, 054501.
- Ren, Z., Bourgeois, D., Helliwell, J. R., Moffat, K., Šrajer, V. & Stoddard, B. L. (1999). *J. Synchrotron Rad.* **6**, 891–917.
- Schlichting, I. (2015). *IUCrJ*, **2**, 246–255.

- Shrive, A. K., Clifton, I. J., Hajdu, J. & Greenhough, T. J. (1990). *J. Appl. Cryst.* **23**, 169–174.
- Smeets, S., Zou, X. & Wan, W. (2018). *J. Appl. Cryst.* **51**, 1262–1273.
- Stagno, J. R., Liu, Y., Bhandari, Y. R., Conrad, C. E., Panja, S., Swain, M., Fan, L., Nelson, G., Li, C., Wendel, D. R., White, T. A., Coe, J. D., Wiedorn, M. O., Knoska, J., Oberthuer, D., Tuckey, R. A., Yu, P., Dyba, M., Tarasov, S. G., Weierstall, U., Grant, T. D., Schwieters, C. D., Zhang, J., Ferré-D'Amaré, A. R., Fromme, P., Draper, D. E., Liang, M., Hunter, M. S., Boutet, S., Tan, K., Zuo, X., Ji, X., Barty, A., Zatsepin, N. A., Chapman, H. N., Spence, J. C. H., Woodson, S. A. & Wang, Y.-X. (2016). *Nature*, **541**, 242–246.
- Stellato, F., Oberthür, D., Liang, M., Bean, R., Gati, C., Yefanov, O., Barty, A., Burkhardt, A., Fischer, P., Galli, L., Kirian, R. A., Meyer, J., Panneerselvam, S., Yoon, C. H., Chervinskii, F., Speller, E., White, T. A., Betzel, C., Meents, A. & Chapman, H. N. (2014). *IUCrJ*, **1**, 204–212.
- Suga, M., Akita, F., Hirata, K., Ueno, G., Murakami, H., Nakajima, Y., Shimizu, T., Yamashita, K., Yamamoto, M., Ago, H. & Shen, J.-R. (2014). *Nature*, **517**, 99–103.
- Tenboer, J., Basu, S., Zatsepin, N., Pande, K., Milathianaki, D., Frank, M., Hunter, M., Boutet, S., Williams, G. J., Koglin, J. E., Oberthuer, D., Heymann, M., Kupitz, C., Conrad, C., Coe, J., Roy-Chowdhury, S., Weierstall, U., James, D., Wang, D., Grant, T., Barty, A., Yefanov, O., Scales, J., Gati, C., Seuring, C., Srajer, V., Henning, R., Schwander, P., Fromme, R., Ourmazd, A., Moffat, K., Van Thor, J. J., Spence, J. C. H., Fromme, P., Chapman, H. N. & Schmidt, M. (2014). *Science*, **346**, 1242–1246.
- Terzakis, G., Lourakis, M. & Ait-Boudaoud, D. (2018). *J. Math. Imaging Vis.* **60**, 422–442.
- Tolstikova, A., Levantino, M., Yefanov, O., Hennicke, V., Fischer, P., Meyer, J., Mozzanica, A., Redford, S., Crosas, E., Opara, N. L., Barthelmess, M., Lieske, J., Oberthuer, D., Wator, E., Mohacsi, I., Wulff, M., Schmitt, B., Chapman, H. N. & Meents, A. (2019). *IUCrJ*, **6**, 927–937.
- Wenk, H. R., Heidelberg, F., Chateigner, D. & Zontone, F. (1997). *J. Synchrotron Rad.* **4**, 95–101.
- White, T. A. (2019). *Acta Cryst. D* **75**, 219–233.
- White, T. A., Barty, A., Stellato, F., Holton, J. M., Kirian, R. A., Zatsepin, N. A. & Chapman, H. N. (2013). *Acta Cryst. D* **69**, 1231–1240.
- White, T. A., Kirian, R. A., Martin, A. V., Aquila, A., Nass, K., Barty, A. & Chapman, H. N. (2012). *J. Appl. Cryst.* **45**, 335–341.
- Zurek, S., Papiz, M., Machin, P. & Helliwell, J. (1985). *Info. Q. Protein Crystallogr.* **16**, 37–40.