electronic reprint

JOURNAL OF SYNCHROTRON RADIATION

ISSN: 1600-5775
journals.iucr.org/s

Seed-skewness algorithm for X-ray diffraction signal detection in time-resolved synchrotron Laue photocrystallography

Dariusz Szarejko, Radosław Kamiński, Piotr Łaski and Katarzyna N. Jarzembska

J. Synchrotron Rad. (2020). **27**, 405–413

IUCr Journals
CRYSTALLOGRAPHY JOURNALS ONLINE

Copyright © International Union of Crystallography

Author(s) of this article may load this reprint on their own web site or institutional repository provided that this cover page is retained. Republication of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see https://journals.iucr.org/services/authorrights.html

JOURNAL OF SYNCHROTRON RADIATION

ISSN 1600-5775

Seed-skewness algorithm for X-ray diffraction signal detection in time-resolved synchrotron Laue photocrystallography

Dariusz Szarejko, Radosław Kamiński, Piotr Łaski and Katarzyna N. Jarzembska*

Department of Chemistry, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland.
*Correspondence e-mail: katarzyna.jarzembska@uw.edu.pl

Received 25 September 2019
Accepted 6 January 2020

Edited by V. Favre-Nicolin, CEA and Université Joseph Fourier, France

Keywords: data processing; background estimation; statistical analysis; skewness; X-ray diffraction.

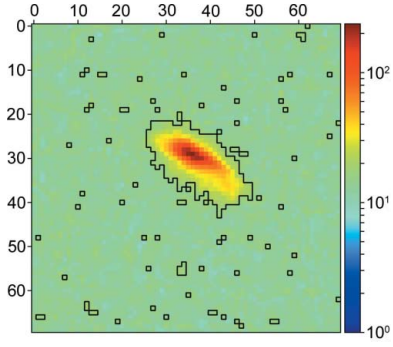
Supporting information: this article has supporting information at journals.iucr.org/s

A one-dimensional seed-skewness algorithm adapted for X-ray diffraction signal detection is presented and discussed. The method, primarily designed for photocrystallographic time-resolved Laue data processing, was shown to work well for the type of data collected at the Advanced Photon Source and European Synchrotron Radiation Facility. Nevertheless, it is also applicable in the case of standard single-crystal X-ray diffraction data. The reported algorithm enables reasonable separation of signal from the background in single one-dimensional data vectors as well as the capability to determine small changes of reflection shapes and intensities resulting from exposure of the sample to laser light. Otherwise, the procedure is objective, and relies only on skewness computation and its subsequent minimization. The new algorithm was proved to yield comparable results to the Kruskal–Wallis test method [Kalinowski, J. A. *et al.* (2012). *J. Synchrotron Rad.* **19**, 637], while the processing takes a similar amount of time. Importantly, in contrast to the Kruskal–Wallis test, the reported seed-skewness approach does not need redundant input data, which allows for faster data collections and wider applications. Furthermore, as far as the structure refinement is concerned, the reported algorithm leads to the excited-state geometry closest to the one modelled using the quantum-mechanics/molecular-mechanics approach reported previously [Jarzembska, K. N. *et al.* (2014). *Inorg. Chem.* **53**, 10594], when the t and s algorithm parameters are set to the recommended values of 0.2 and 3.0, respectively.

1. Introduction

Objective determination of whether the data contains signal, and, if so, its proper detection and separation from noise, constitutes a general problem in experimental science. Among others, it is a very important issue in processing of diffraction data in crystallography. Although a lot has been done with regards to this matter, especially involving standard crystallographic experiment outcomes (Winter *et al.*, 2018), in some more specific cases, new efficient, *i.e.* robust and fast, solutions for the data processing are still welcome. Such tools, for instance, would be desirable for processing small-molecule laser-pump/X-ray-probe time-resolved (TR) diffraction data, in which a large part of important signal is rather weak and datasets are usually large.

The Laue diffraction method (Amorós *et al.*, 1975) has been reconsidered years after the pioneering experiments (Friedrich *et al.*, 1912, 1913), not only as far as neutron diffraction is concerned (Keen *et al.*, 2006; O'Dell *et al.*, 2016; Ouladdiaf *et al.*, 2011) but also due to the development of TR X-ray diffraction techniques, in which it has some particular advantages over monochromatic methods (Coppens, Bene-



© 2020 International Union of Crystallography

electronic reprint

dict, *et al.*, 2010; Ren *et al.*, 1999). The approach has been successfully used to elucidate the dynamics of proteins in crystals (Schmidt, 2017; Šrajer & Schmidt, 2017) and, more recently, to small-molecule systems in order to determine subtle structural changes occurring in crystals upon light-induced electronic excitation (Trzop *et al.*, 2014; Jarzembska *et al.*, 2014, 2019; Makal *et al.*, 2012; Benedict *et al.*, 2011). Owing to the increased interest in the method, a number of computer programs for handling of the X-ray Laue data have been introduced, such as the Daresbury *LAUE* software suite (Helliwell *et al.*, 1989; Campbell, 1995), *LAUEVIEW* and *PRECOGNITION* (Šrajer *et al.*, 2000), *LAUEGUI* (Messerschmidt & Tschentscher, 2008) or *LAUEUTIL* (Kalinowski *et al.*, 2012, 2011). Furthermore, to circumvent application of wavelength-dependent corrections to the TR X-ray Laue data, Coppens *et al.* (2009) developed the so-called RATIO method. In this approach, the intensity ratios of laser-ON and laser-OFF frames ($I_{\rm ON}/I_{\rm OFF}$) are computed for synchrotron-collected data. Such ratios can be used to estimate the laser-induced temperature rise of the sample (Schmøkel *et al.*, 2010; Vorontsov & Coppens, 2005), to compute photodifference maps (Fournier & Coppens, 2014), or to perform structure refinement using the specific software (Vorontsov *et al.*, 2010).

Nevertheless, most of the existing procedures for signal finding and integrating, implemented in the above-mentioned software (with the exception of *LAUEGUI* and *LAUEUTIL*), utilize two-dimensional profile-fitting techniques (Ren & Moffat, 1995; Helliwell *et al.*, 1989; Bourgeois *et al.*, 1998). In practice, this hampers processing of TR data, where spot-shape variation *versus* pump–probe delay time is frequently observed (Collet *et al.*, 2012; Coppens; Benedict, *et al.*, 2010). Furthermore, the majority of approaches are based on the predicted spot position. Thus, the processing sequence is dependent on *a priori* knowledge of crystal orientation, determination of which, in the case of Laue diffraction, may be rather challenging and/or time-consuming. Therefore, the algorithm presented here deals with the general case of Laue data processing, where the integration is performed prior to reflection indexing, and coupled with the statistical approach of signal detection. Our contribution constitutes an extension of algorithms presented by Kalinowski *et al.* (2012) and coded in the *LAUEUTIL* software. It is also a part of a wider project dedicated to facilitating small-molecule laser-pump/X-ray-probe TR Laue diffraction data processing and analysis.

2. Description of the algorithm

Two already-known approaches have recently caught our attention, thus constituted inspiration for this contribution: the orientation-matrix-less signal finding procedure (Kalinowski *et al.*, 2012), based on the Kruskal–Wallis (KW) non-parametric test; and the seed-skewness (SS) method (Bolotovsky *et al.*, 1995; Bolotovsky & Coppens, 1997; Darovsky & Kezerashvili, 1997) (for more information see the supporting information). Our algorithm, in general, combines the two ideas. Is it worth noting that the two-dimensional SS algorithm was originally applied to experimental charge-density distribution determination (Iversen *et al.*, 1999, 1998) and to TR Laue X-ray diffraction data (Makal *et al.*, 2011; Kamiński *et al.*, 2010), whereas a three-dimensional case was tested on a monochromatic neutron dataset (Peters, 2003).

The entire procedure reported here consists of two parts. The first part is dedicated to signal detection in a one-dimensional dataset independently for every pixel on the detector. The second part covers construction and filtering of masks for each diffraction frame, and the subsequent intensity integration. The first part, which encompasses realization of our algorithm, is described below in detail; the respective flowcharts are presented in Figs. 1 and 2.

(i) *Selection of pixels containing significant signal.* To determine if any signal is observed in a data vector for a given detector pixel, the initial value of skewness, $\mu_{3,\rm init}$, and its standard deviation, $\sigma(\mu_{3,\rm init})$ (these values are computed for the entire vector), are computed. It is assumed that the signal is present in the vector only if

$$|\mu_{3,\rm init}| > t \cdot \sigma(\mu_{3,\rm init}), \qquad (1)$$

where t (the 'trust level') is the first adjustable parameter (Fig. 1). If this condition is fulfilled, the next steps of the algorithm are invoked. Otherwise, the procedure for this particular pixel is aborted, and the background for this pixel is computed as a simple mean value over all frames.

(ii) *Initial background estimation for a particular pixel.* At this stage our procedure diverges from the two- and three-dimensional cases, where initial background, $B_{\rm init}$, is computed as a mean over the perimeter pixels (assuming initially that a reflection does not intersect the specified integration box). Instead, the median over all values for a given vector is computed. This is because calculating the mean of two perimeter values available for a one-dimensional data vector may introduce bias. In the worst case, the first and last frames could contain relatively strong signal for the same pixel position, and thus the initial background would be highly overestimated.

(iii) *Construction of the seed.* The seed in a vector is constructed by including every value with intensity higher than

$$\langle B_{\rm init}\rangle + s \cdot \sigma(B_{\rm init}), \qquad (2)$$

where s (the 'signal level') constitutes the second adjustable parameter (Fig. 1). The seed does not have to be contiguous (*i.e.* multiple signals can be found across the vector index). If no values higher than the specified threshold are found, the procedure is aborted for this particular pixel, and the background for this pixel constitutes a simple mean of its intensity values from all registered frames.

(iv) *Skewness evaluation and minimization.* Values in a vector, which are not included in the seed, are used to compute skewness (μ_3, Fig. 2); the value of μ_3 is stored in the $\mu_{3,\rm pre}$ variable. Subsequently, the largest intensity value not present in the seed is found and included in it, and again the skewness, μ_3, of the background distribution is computed. If $\mu_3 < \mu_{3,\rm pre}$ (*i.e.* new skewness is not larger than the previous one), the whole step is repeated and new intensity values are

electronic reprint

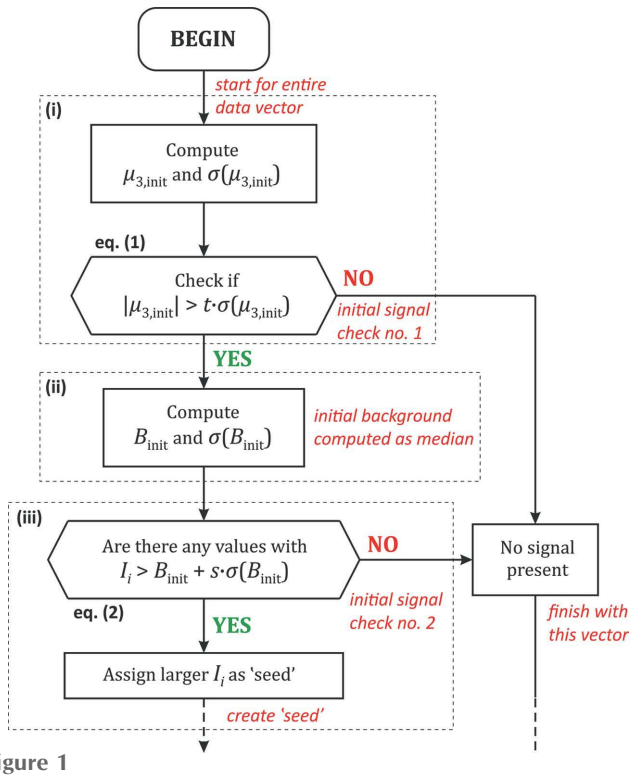


Figure 1
First part of the schematic algorithm flowchart presenting the initial signal detection and seed construction; t ('trust level') and s ('signal level') constitute two adjustable parameters in the algorithm. Symbols (i), (ii) and (iii) or eq. (1) and eq. (2) refer to sections of the main text and equation numbering, respectively.

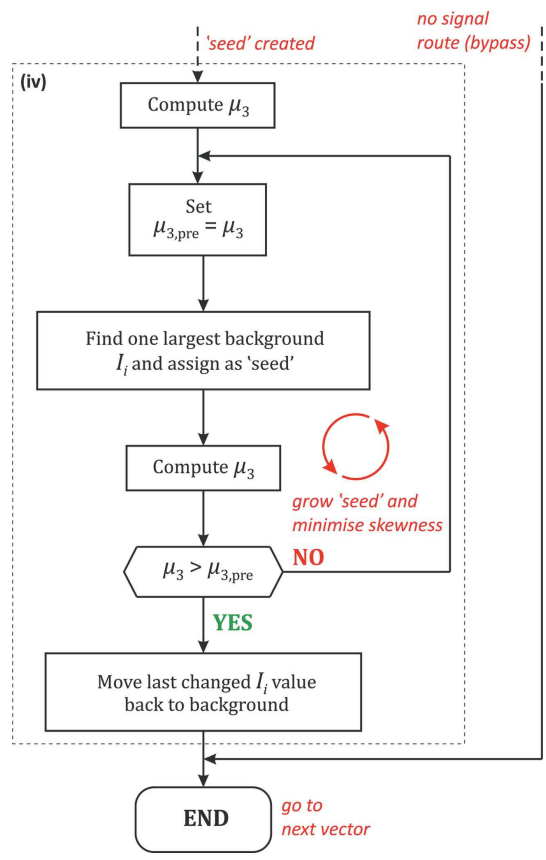


Figure 2
Second part of the schematic algorithm flowchart, illustrating the seed-growing procedure based on the skewness minimization. Symbol (iv) refers to the section of the main text.

moved from the background to the seed (this is the seed-growing part). However, when $\mu_3 > \mu_{3,\mathrm{pre}}$ (*i.e.* the skewness has increased), further actions are aborted and the last-found intensity value is moved back to the background. Furthermore, the mean background for this pixel is computed as a mean value of the intensity values not included in the seed. In the case of our TR datasets the procedure usually reaches the minimum of the background skewness after several to about a dozen cycles (the required number of cycles is dataset-dependent). Finally, it should be noted that essentially only the skewness-minimization part of the original algorithm is used, *i.e.* the grown-seed region does not have to be contiguous. Since in our case vectors span the entire angular scan, implementation of the criterion assuring contiguity of a single seed, which was well grounded in the two-dimensional case where a single diffraction spot was considered, is no longer justified. Such a condition could result in losing relatively weak signals. For a similar reason, the data contained in the one-dimensional vector not smoothed in the algorithm presented.

(v) *Mask preparation and integration.* Further stages of the data processing follow the published procedure (Kalinowski *et al.*, 2012), in which the key steps are (*a*) construction of two-dimensional integration masks for each frame, (*b*) their filtering using morphological operations (Pierre, 2003), (*c*) connected-component labelling to enumerate independent reflection masks and (*d*) reflection integration. In the last step, the integrated intensity, I, of a particular reflection is computed as

$$I = \sum_{(i,j)} (I_{ij} - \langle B_{ij} \rangle), \qquad (3)$$

where summation covers all pixel coordinate pairs belonging to the mask at this particular frame, and B_{ij} is the mean background determined at the stage of signal analysis procedure (note the background may vary over the reflection area).

3. Performance tests on TR Laue data

First, the algorithm performance on single vectors was briefly examined. As mentioned previously, the t parameter controls whether to start analysing the vector, whereas the s parameter modifies the size of the initial seed, which is further expanded via background skewness minimization. Similarly, as described by Straasø *et al.* (2013), a simple model of the diffraction signal consists of some background level and a peak, summed together. The peak is modelled by a simple Gaussian function, $g(k) = A \exp\left[-(k - k_0)^2/2\sigma^2\right]$, where A indicates its height, and which is centred at a frame k_0 (k is the frame index) with standard deviation σ. Both the background and the peak are Poisson-randomized. As a test example, a relatively narrow peak was chosen (A = 60 counts, σ = 2 frames) and summed

electronic reprint

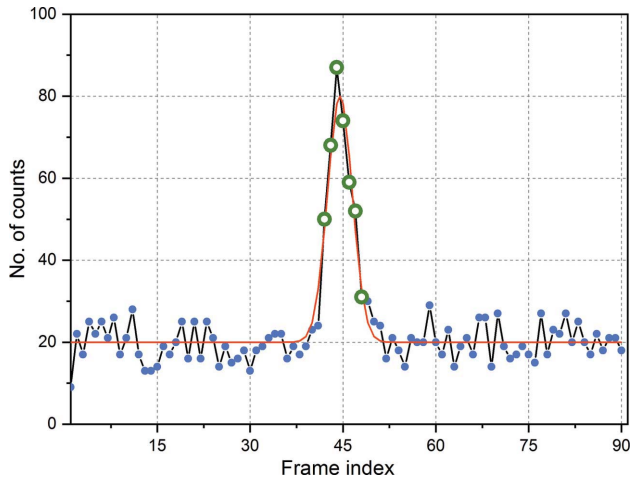


Figure 3
Performance of the algorithm on a single Poisson-randomized simulated data vector (Gaussian peak; A = 60 counts, σ = 2 frames; B = 20 counts). Solid red line – simulated noiseless Gaussian-shape peak; small blue-filled circles – determined background; large green-edge circles – SS-algorithm-determined signal.

with a background of 20 counts on average (Fig. 3). This represents a medium-intensity peak found in the Laue data where reflections occur only on several consecutive frames at best. In Fig. 3 the large green-edge circles represent the signal as recognized by the algorithm. It seems that the algorithm performance is satisfactory, as far as finding and separating the signal from the background are concerned. However, as pointed out in the literature (Kalinowski *et al.*, 2012; Straasø *et al.*, 2013), the derived signal appears to be underestimated. Hence, two dilations at further stages of the integration mask evaluation to account for the reflection rim appear to be justified. Nevertheless, it should be kept in mind that all masks are constructed 'perpendicularly' to the data vector, *i.e.* for a given frame in two dimensions. The dilations used introduce additional signal to a given data vector only if the signal was found earlier for the neighbouring pixels on the same diffraction frame. In this view, careful inspection of two-dimensional masks is of prime importance while processing the data.

In order to check if the method provides reasonable results for real experimental data, several TR Laue X-ray datasets collected for a single-crystal of a model tetranuclear silver(I)–copper(I) complex (Jarzembska *et al.*, 2014) on two synchrotron beamlines [ID09 (Wulff *et al.*, 2002) at the European Synchrotron Radiation Facility (ESRF) and BioCARS 14-ID-B (Graber *et al.*, 2011) at the Advanced Photon Source (APS)] were selected. These sets contain both *dark*-type (only light-OFF frames, single exposure per crystal orientation) and *laser*-type (both light-ON and light-OFF frames, multiple exposures per crystal orientation) data.

At first, the influence of the algorithm adjustable parameters on the final mask was checked. For this purpose, different s and t parameter values were used to process the ESRF *dark*-type dataset. The t-parameter scan with the s value fixed to 3.0 is shown in Fig. 4. The filtering based on morphological operations was the same throughout the whole analysis and followed the sequence of (i) removing of single-pixel masks, (ii) erosion, (iii) two consecutive dilations and, finally, (iv) filling out mask holes (if needed). Since the t parameter controls one of the very first steps of the algorithm, as expected, the lower its value, the more noise-like spots are introduced into the raw mask [see Figs. S2 and S3 of the supporting information]. In such cases, however, the filtering procedures appear to work very well, yielding clean background and reasonable spot shapes. Consequently, the algorithm appears to work sensibly for relatively small t values, such as 0.1–0.3. In contrast, increasing the t parameter over 1.0 results in rather pathological shapes of strong spots (Fig. 4). Furthermore, weak spots are not detected when t exceeds 0.4. The t value of 0.2 seems to be a rational choice regarding both time of processing and final results.

In turn, inspection of the s parameter, keeping t fixed to 0.2, showed it is responsible for introducing noise when its values are too low (s = 1.0; see Fig. 5). Furthermore, in the vicinity of strong reflections this noise cannot be attenuated by standard filtering techniques. Taking $s > 2.5$ makes the 'parasite' spots disappear, whereas s values greater or equal to 4.0 lead to severely underestimated weak spot mask areas. Consequently, setting s between 2.5 and 3.5 seems to be a good compromise for both strong and weak spots' shapes.

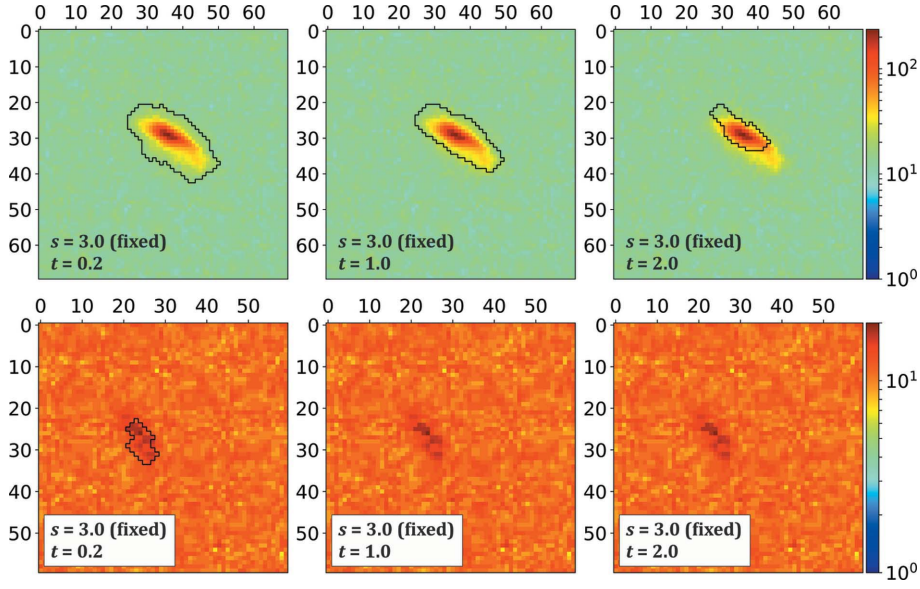


Figure 4
Filtered masks obtained for the *dark*-type dataset with various t-parameter settings (with s = 3.0) for strong (top panels) and weak spots (bottom panels) (filtering sequence: removing of single-pixel masks, erosion, two consecutive dilations, filling mask holes – this sequence is retained for all figures). For more details see Figs. S2 and S3.

electronic reprint

Table 1
Benchmark datasets compared in this contribution.

	Dataset source		
	ID09 at ESRF		14-ID-B at APS
Data type	Dark†	Laser‡	Laser‡
Number of frames	91	530	800
Angular coverage (°)	91	106	160
Angle increment (°)	1	2	2
Detector	Rayonix MX170-HS		Rayonix MX340-HS
Detector shape	Square		Square
Detector dimensions§ (mm)	170		340
Frame dimensions§ (pixels)	3840		3840
Pixel size§ (μm)	44		87
Detector distance (mm)	40		100
2θ (°)	0		0

† Dark – 'normal' dataset with no laser exposure. ‡ Laser – TR dataset collected according to the laser-pump/X-ray-probe strategy with light-ON and light-OFF block data. § Both dimensions (vertical and horizontal) are the same.

In the next step, our algorithm was tested on full laser-type datasets, as shown in Table 1. The first dataset was collected at ESRF and the second one at the APS synchrotron. The performance of the algorithm was examined, with the 'signal level' s parameter set to 3.0, and the 'trust level' t parameter set either to 0.2 or 1.0. The seed-skewness outcomes were compared with the results obtained by applying the KW integration scheme, used by *LAUEUTIL* and also implemented in our code. Numerical parameters of all integrations are presented in Table 2. As expected, the SS method yields more reflections with t set to 0.2 than with t = 1.0, which is about 36% more reflections integrated on average per single frame in the former case. This result is consistent with the previous analyses, which showed that establishing higher t values may result in unsatisfactory evaluation of weaker reflections. In turn, the KW test provides roughly 10–15% more reflections compared with the SS method with t = 1.0.

Table 2
Total number of reflections integrated (after filtering) with three different methods for two laser-type datasets (see Table 1).

Values in brackets indicate the average number of integrated spots per frame.

	Data set source	
Method	ID09 at ESRF	14-ID-B at APS
KW test	132969 (250.9)	147877 (184.8)
SS method (t = 0.2, s = 3.0)	158349 (298.8)	180775 (226.0)
SS method (t = 1.0, s = 3.0)	115295 (217.5)	133664 (167.1)

The algorithm dedicated to time-resolved Laue data should also clearly distinguish between frames collected with and without laser exposure coming from the same dataset. In the case of our approach, the derived masks clearly reflect the differences between the respective light-ON and light-OFF frames (see Figs. S6 and S7). As such, these outcomes are similar to those reported in the literature (Collet *et al.*, 2012; Coppens, Benedict *et al.*, 2010). The differences between light-ON and light-OFF frames are visible regardless of the integration scheme used.

Finally, it is valuable to compare various integration routines through examination of correlation plots of the respective intensities computed for the same diffraction dataset. The reflections obtained using two data processing methods can be matched via simple geometric criterion, *i.e.* nearly identical coordinates on the detector surface. In our case the intensity data derived using the SS method with two different 'trust level' settings, *i.e.* with t = 1.0 and t = 0.2 applied, were confronted with the KW integration outcomes. The results are shown in Fig. 6. The obtained picture confirms that the processed datasets collected at APS and ESRF behave similarly. In all cases the correlation is very high (coefficient of determination R^2 > 99%). Furthermore, the stronger the signal, the better the correlation; the strongest reflections are clearly integrated almost identically in the tested approaches. The most notable differences are observed for the group of weakest reflections, primarily for those with raw intensities not exceeding 1000 ADU (arbitrary detector units). In the case of the first SS integration (t = 1.0), the correlation plots show a rather symmetric spread for low-intensity reflections. However, a closer look at the ratios of SS- *versus* KW-integrated spots' areas indicates that the bottom part of the plot (with stronger KW intensities) exhibits larger spots for the KW method. In contrast, the top part contains stronger SS reflections of significantly larger sizes. Quite interestingly, when the KW-test integration is compared with the SS method with the t parameter set to 0.2 (suggested to be a more rational choice by the above

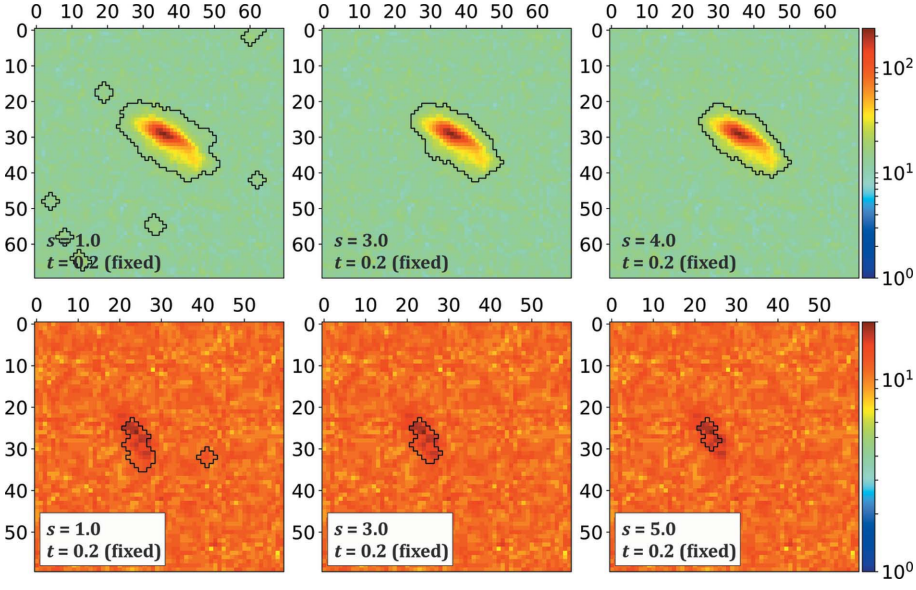


Figure 5
Filtered masks obtained for the *dark*-type dataset with various s -parameter settings (with t = 0.3) for strong (top panels) and weak spots (bottom panels). For more information, see Figs. S3 and S4.

electronic reprint

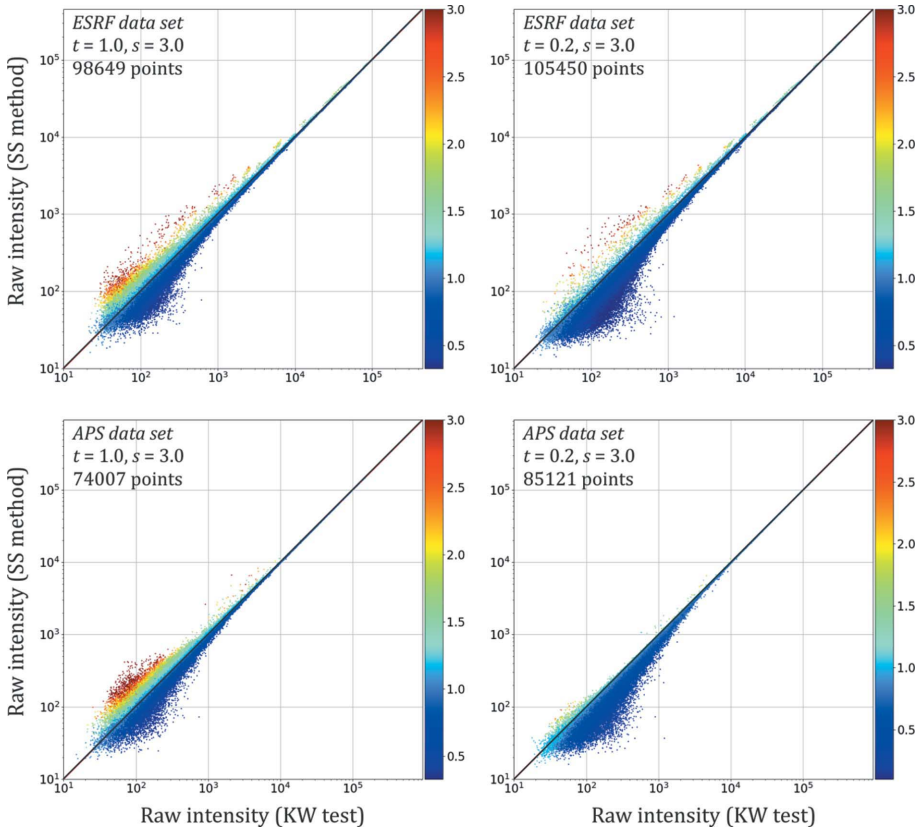


Figure 6
Correlation plots of the SS-integrated (vertical axes) *versus* KW-integrated spot intensities (horizontal axes). Two SS integration schemes were used with t = 1.0 (left panels) and t = 0.2 (right panels) for two laser-type datasets from ESRF (top panels) and APS (bottom panels) are presented. Plots are coloured with reflection area ratios (SS- over KW-reflection area ratios); intensities are given in ADU; a logarithmic scale is used; the black solid straight line is the reference line of $I_{SS} = I_{KW}$. For more details see the supporting information.

analyses), a substantial asymmetry is observed at the low-intensity region. Clearly, the top part of the spread region is missing (except for some isolated spots observed only for the ESRF dataset). At the same time, however, about 7% and 15% more spots match between KW- and SS-integrated sets for SS with t = 0.2 *versus* t = 1.0 for ESRF and APS datasets, respectively. It appears that a great number of weak reflections moved to the bottom of the reference line (Fig. S8), which indicates that they are still paired with the KW-determined spots, but their calculated intensities are different. All this suggests that the reflection area (*i.e.* the size of the integration mask) is a key factor influencing the intensity in this case, leading to the observed differences in signal-finding and integrating methods. Furthermore, the importance of reflection sizes on their intensity is evident when analysing their absolute values (Fig. S10). Out of weaker reflections (raw intensities lower than *ca.* 1000 ADU), the smallest ones (as integrated with the SS method) exhibit lower intensity than the ones evaluated with the KW method (regardless of the t value setting). On the other hand, larger-area low-intensity reflections seem to have lower intensity than the KW-evaluated equivalents when t = 1.0, but not for t = 0.2. However, this effect is more pronounced for the APS data with respect to the ESRF, for which it is very weak, thus may not be general. The other contributing factor, *i.e.* the mean background computed for a given reflection, seems to be very comparable for the examined data-processing methods (Fig. S9). Indeed, the computed mean reflection background ratios are very close to 1.0, which are especially well matched for the APS dataset.

4. Structural refinement result analysis

Finally, the mutual performance of the two algorithms was checked via comparison of the resulting excited-state structures. Here, the model copper(I)–silver(I) complex, for which we determined the full three-dimensional structure of the excited state in previous work (Jarzembska *et al.*, 2014), again constitutes the subject of the analyses. We used the already published data (now available in the Repository for Open Data, RepOD; Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Warsaw, Poland; doi: 10.18150/repod.6395772) measured with the Mar165 CCD-type round-shape detector (diameter: 165 cm; frame size: 2048 pixels in the vertical and horizontal directions; pixel size: 80 µm in both directions; Table 3). The four originally examined datasets were integrated applying both KW and SS methods implemented in our software and further processed using the overall procedure of the Laue data treatment (orientation matrix determination, frame-to-frame orientation refinement, indexing, ratio computation) (Coppens & Fournier, 2015). It is worth mentioning that our implementation of the KW-test method is not entirely identical to that in the *LAUEUTIL* software used previously. Thus, for the purpose of the current study the data were processed even more carefully than before (*e.g.* rejected detector areas, such as the beamstop, are now more appropriately defined; spots touching the rejected areas are not taken for further computations, *etc.*). The excited-state structural models were derived with the aid of the program *LASER* (Vorontsov *et al.*, 2010). In all cases the same refinement strategy was applied. The measurement geometry and data processing results are collected in Table 3, the selected refinement and geometrical parameters of the refined crystal structure are gathered in Table 4, while the overlay of the derived metal-core geometries is shown in Fig. 7. It should be noted that all four datasets shared one crystal-structure model during the refinement in *LASER*, whereas only population parameters and temperature-scale factors were refined separately for each case.

electronic reprint

Table 3
Numerical data showing the geometry and processing results of four TR datasets collected for the silver(I)–copper(I) complex.

Original datasets retrieved from the RepOD repository (doi: 10.18150/repod.6395772). Values in parentheses indicate the average number of integrated spots per frame.

Data set	No. 1	No. 2	No. 3	No. 4
Number of frames	670	230	455	230
Angular coverage (°)	66	90	180	90
Angle increment (°)	1	2	2	2
Detector distance (mm)	65	50	50	50
Number of raw spots				
KW test	325130 (485.3)	89725 (390.1)	199212 (437.8)	106028 (461.0)
SS method (t = 1.0, s = 3.0)	318857 (475.9)	95763 (416.4)	214249 (470.9)	103743 (451.1)
SS method (t = 0.2, s = 3.0)	337216 (503.3)	108858 (473.3)	237717 (522.5)	116493 (506.5)
Number of unique reflections				
KW test	3153	3202	5202	3802
SS method (t = 1.0, s = 3.0)	3403	3744	6139	4228
SS method (t = 0.2, s = 3.0)	3237	3533	5539	4014

Table 4
Selected structural and refinement parameters computed for data integrated with three different methods (see Table 3).

For more numerical parameters see Table S1 of the supporting information. R-factor calculated on intensity ratios (Coppens, Kamiński & Schmøkel, 2010): $R[R] = \sum_i |R_{o,i} - R_{c,i}| / \sum_i R_{o,i}$. QM/MM results are taken from the original reference of Jarzembska *et al.* (2014).

Parameter	KW test	SS method (t = 1.0, s = 3.0)	SS method (t = 0.2, s = 3.0)	QM/MM results
$d_{Ag1\cdots Ag2}$ (Å)	2.65 (2)	2.59 (2)	2.70 (2)	2.749
$d_{Ag1\cdots Cu2}$ (Å)	2.85 (2)	2.80 (3)	2.95 (2)	3.114
$d_{Ag2\cdots Cu2}$ (Å)	2.69 (2)	2.66 (3)	2.71 (2)	2.822
Population % (set No. 1)	0.47 (4)	0.37 (3)	0.54 (4)	–
Population % (set No. 2)	1.11 (6)	0.87 (4)	1.21 (7)	–
Population % (set No. 3)	0.77 (4)	0.61 (3)	0.89 (5)	–
Population % (set No. 4)	0.95 (5)	0.75 (4)	1.04 (6)	–
$R[R]$ (all data) (%)	2.34	2.46	2.35	–
No. of reflections (all data)	11797	13196	12508	–

Generally, the number of raw spots is greater for the SS method when compared with the KW approach. This effect is emphasized more when t is set to 0.2 than for t = 1.0. As shown, the SS method with both examined settings of the t parameter yields also a larger number of unique reflections than KW. Quite interestingly, the SS method with t = 1.0 yields roughly 14% more reflections than KW, whereas with t = 0.2 this excess amounts to about 6%, despite the noted larger number of raw spots. The processing times are correlated with the number of detected spots and are comparable for the analysed methods. Regarding the derived excited states' populations and geometrical features of the excited-state species, in all cases the trends are well preserved (Table 4). The most significant bond-shortening was obtained for the SS approach with t set to 1.0, while the most moderate one was again for SS but with t = 0.2. Even though the differences are not very pronounced, at places they exceed the least-squares-derived standard deviation error bars, which certainly reflect better the real accuracy of the refined excited-state geometry from the TR Laue experiment. Importantly, the SS method (t = 0.2) yields results closest to the previously reported theoretically calculated values within the QM/MM approach (Table 4).

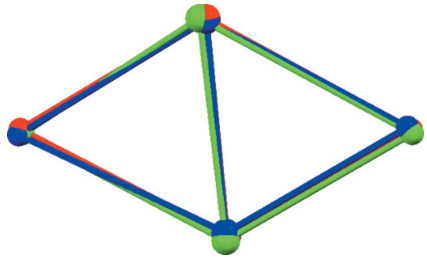


Figure 7
Overlay of the tetranuclear copper(I)–silver(I) metal-core geometry obtained from the tested approaches (red – KW test, blue – SS with t = 1.0, green – SS with t = 0.2).

5. Test on monochromatic data

In addition, our processing method was briefly tested on the dataset collected using monochromatic X-ray radiation (Mo $K\alpha$) at the laboratory source. The dataset was retrieved from the RepOD data repository (doi: 10.18150/repod.7426818). It was collected using a well known Bruker AXS APEX II square-array CCD detector and contains raw diffraction frames from our previous study on the potassium uridine-5′-monophosphate salt, K(UMP) (Jarzembska *et al.*, 2017). Example filtered masks are presented in Fig. 8. Clearly, the obtained spot shapes look very reasonable when t = 0.2 and s = 3.0 were used, suggesting the algorithm works robustly for this kind of data. The integration of the diffraction spots was not carried out, since our current code is capable of only two-dimensional frame integration, whereas in the case of monochromatic data the three-dimensional integration is indispensable. None-

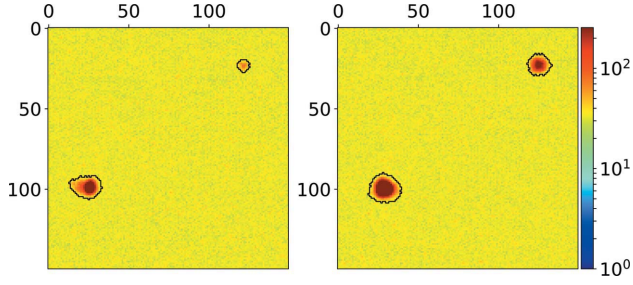


Figure 8
Filtered two-dimensional masks obtained for selected fragments of two consecutive frames (left and right) for monochromatic in-house data for the K(UMP) single crystal (SS algorithm with t = 0.2 and s = 3.0). Mask outlines are marked in black; all panels are drawn with the same intensity scale. For more details see Fig. S11.

electronic reprint

theless, this result provides a future prospect for implementing this method also for accurate charge-density or time-resolved studies.

6. Summary and conclusions

A new one-dimensional seed-skewness algorithm for signal detection and evaluation in X-ray diffraction data, based on the skewness-minimization concept, was described and examined. It was shown that the method enables reasonable separation of signal from the background in single one-dimensional data vectors. As the signal appears underestimated, which is rather common in many approaches of this kind, multiple dilations are used to account for this effect. Importantly, the method is capable of extracting relatively weak reflections from the background. This is possible through adjusting of the 'trust level' (t) and 'signal level' (s) parameters in the algorithm. It was shown that the parameter values of $t \simeq 0.2$ and $s \simeq 3.0$ appear to be a reasonable choice. Otherwise, the procedure is objective and relies only on skewness computation and minimization, which enable rational background estimation.

The method is applicable for full TR X-ray Laue diffraction data, as it is capable of determining small changes of reflection shapes and intensities resulting from exposure of the sample to laser light. In addition, the algorithm performance was compared with the already-published Kruskal–Wallis method, also implemented in our code. It turns out that the proposed method generally yields more reflections than the KW test for the advisable t and s parameters in a comparable time. The intensities of strong reflections evaluated using the examined approaches are very alike, whereas weak reflections are more sensitive to the algorithm setting parameters. Both methods estimate the background level equally well. It is also worth noting that the method should work well in various other cases considered as dynamic structure crystallography, including studies under external electric field (Fertey *et al.*, 2013; Hekstra *et al.*, 2016), in which data with and without external perturbation are collected at various goniometer settings using an area detector.

Regarding the structure refinement, both methods yielded comparable excited-state geometries. However, in places there are statistically significant differences between the evaluated parameters (taking into account the least-squares-derived errors), which indicate the level of real accuracy of the TR Laue method. Nevertheless, the trends in all cases are well preserved. It should also be noted here that the SS-derived molecular geometry is in closest agreement with the theoretical results obtained using the advanced QM/MM approach, when the t and s parameters are set to the recommended values.

The program code used is available from the authors under request. In principle, it allows future implementation of other methods of signal finding, such as the variance-mean-equality method published by Straasø *et al.* (2013). Currently, parallelization of the entire program is in progress, taking advantage of CUDA programming features in order to speed up the data processing. This work shall be presented shortly.

Acknowledgements

RK and KNJ would like to thank Lauren Hatcher (Bath, UK), Martin Pedersen and Michael Wulff (Grenoble, France), Sławomir Domagała and Sylwia Kutniewska (Warsaw, Poland), and Anthony DiChiara (Chicago, USA) for their assistance and/or discussions during initial synchrotron experiments.

Funding information

DS and KNJ, and RK and PŁ would like to thank the SONATA grants (2014/15/D/ST4/02856 and 2016/21/D/ST4/03753, respectively) of the National Science Centre in Poland for financial support. The time-resolved X-ray diffraction experiments were performed at the ID09 beamline of the European Synchrotron Radiation Facility (ESRF), Grenoble, France. The research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory (contract No. DE-AC02-06CH11357). Use of BioCARS was also supported by the National Institute of General Medical Sciences of the National Institutes of Health (NIH) (grant No. R24GM111072) (note: the content is solely the responsibility of the authors and does not necessarily represent the official views of NIH). Time-resolved setup at Sector 14 was funded in part through collaboration with Philip Anfinrud (NIH/NIDDK).

References

Amorós, J. L., Buerger, M. J. & Amorós, M. C. (1975). *The Laue Method*. New York, San Francisco, London: Academic Press.
Benedict, J. B., Makal, A., Sokolow, J. D., Trzop, E., Scheins, S., Henning, R., Graber, T. & Coppens, P. (2011). *Chem. Commun.* **47**, 1704–1706.
Bolotovsky, R. & Coppens, P. (1997). *J. Appl. Cryst.* **30**, 244–253.
Bolotovsky, R., White, M. A., Darovsky, A. & Coppens, P. (1995). *J. Appl. Cryst.* **28**, 86–95.
Bourgeois, D., Nurizzo, D., Kahn, R. & Cambillau, C. (1998). *J. Appl. Cryst.* **31**, 22–35.
Campbell, J. W. (1995). *J. Appl. Cryst.* **28**, 228–236.
Collet, E., Moisan, N., Baldé, C., Bertoni, R., Trzop, E., Laulhé, C., Lorenc, M., Servol, M., Cailleau, H., Tissot, A., Boillot, M.-L., Graber, T., Henning, R., Coppens, P. & Buron-Le Cointe, M. (2012). *Phys. Chem. Chem. Phys.* **14**, 6192–6199.
Coppens, P., Benedict, J., Messerschmidt, M., Novozhilova, I., Graber, T., Chen, Y.-S., Vorontsov, I., Scheins, S. & Zheng, S.-L. (2010). *Acta Cryst.* A**66**, 179–188.
Coppens, P. & Fournier, B. (2015). *J. Synchrotron Rad.* **22**, 280–287.
Coppens, P., Kamiński, R. & Schmøkel, M. S. (2010). *Acta Cryst.* A**66**, 626–628.
Coppens, P., Pitak, M., Gembicky, M., Messerschmidt, M., Scheins, S., Benedict, J., Adachi, S., Sato, T., Nozawa, S., Ichiyanagi, K., Chollet, M. & Koshihara, S. (2009). *J. Synchrotron Rad.* **16**, 226–230.
Darovsky, A. & Kezerashvili, V. (1997). *J. Appl. Cryst.* **30**, 128–132.
Fertey, P., Alle, P., Wenger, E., Dinkespiler, B., Cambon, O., Haines, J., Hustache, S., Medjoubi, K., Picca, F., Dawiec, A., Breugnon, P.,

electronic reprint

Delpierre, P., Mazzoli, C. & Lecomte, C. (2013). *J. Appl. Cryst.* **46**, 1151–1161.
Fournier, B. & Coppens, P. (2014). *Acta Cryst.* A**70**, 291–299.
Friedrich, W., Knipping, P. & Laue, M. (1912). *Sitzungsber. Kgl. Bayer. Akad. Wiss.* pp. 303–322.
Friedrich, W., Knipping, P. & Laue, M. (1913). *Ann. Phys.* **346**, 971–988.
Graber, T., Anderson, S., Brewer, H., Chen, Y.-S., Cho, H. S., Dashdorj, N., Henning, R. W., Kosheleva, I., Macha, G., Meron, M., Pahl, R., Ren, Z., Ruan, S., Schotte, F., Šrajer, V., Viccaro, P. J., Westferro, F., Anfinrud, P. & Moffat, K. (2011). *J. Synchrotron Rad.* **18**, 658–670.
Hekstra, D. R., White, K. I., Socolich, M. A., Henning, R. W., Šrajer, V. & Ranganathan, R. (2016). *Nature*, **540**, 400–405.
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.
Iversen, B. B., Darovsky, A., Bolotovsky, R. & Coppens, P. (1998). *Acta Cryst.* B**54**, 174–179.
Iversen, B. B., Larsen, F. K., Pinkerton, A. A., Martin, A., Darovsky, A. & Reynolds, P. A. (1999). *Acta Cryst.* B**55**, 363–374.
Jarzembska, K. N., Hapka, M., Kamiński, R., Bury, W., Kutniewska, S. E., Szarejko, D. & Szczęśniak, M. M. (2019). *Crystals*, **9**, 36.
Jarzembska, K. N., Kamiński, R., Fournier, B., Trzop, E., Sokolow, J. D., Henning, R., Chen, Y. & Coppens, P. (2014). *Inorg. Chem.* **53**, 10594–10601.
Jarzembska, K. N., Ślepokura, K., Kamiński, R., Gutmann, M. J., Dominiak, P. M. & Woźniak, K. (2017). *Acta Cryst.* B**73**, 550–564.
Kalinowski, J. A., Fournier, B., Makal, A. & Coppens, P. (2012). *J. Synchrotron Rad.* **19**, 637–646.
Kalinowski, J. A., Makal, A. & Coppens, P. (2011). *J. Appl. Cryst.* **44**, 1182–1189.
Kamiński, R., Graber, T., Benedict, J. B., Henning, R., Chen, Y.-S., Scheins, S., Messerschmidt, M. & Coppens, P. (2010). *J. Synchrotron Rad.* **17**, 479–485.
Keen, D. A., Gutmann, M. J. & Wilson, C. C. (2006). *J. Appl. Cryst.* **39**, 714–722.
Makal, A., Benedict, J., Trzop, E., Sokolow, J., Fournier, B., Chen, Y., Kalinowski, J. A., Graber, T., Henning, R. & Coppens, P. (2012). *J. Phys. Chem. A*, **116**, 3359–3365.
Makal, A., Trzop, E., Sokolow, J., Kalinowski, J., Benedict, J. & Coppens, P. (2011). *Acta Cryst.* A**67**, 319–326.
Messerschmidt, M. & Tschentscher, T. (2008). *Acta Cryst.* A**64**, C611.
O'Dell, W. B., Bodenheimer, A. M. & Meilleur, F. (2016). *Arch. Biochem. Biophys.* **602**, 48–60.
Ouladdiaf, B., Archer, J., Allibon, J. R., Decarpentrie, P., Lemée-Cailleau, M.-H., Rodríguez-Carvajal, J., Hewat, A. W., York, S., Brau, D. & McIntyre, G. J. (2011). *J. Appl. Cryst.* **44**, 392–397.
Peters, J. (2003). *J. Appl. Cryst.* **36**, 1475–1479.
Pierre, S. (2003). *Morphological Image Anaysis.* Berlin, Heidelberg: Springer.
Ren, Z., Bourgeois, D., Helliwell, J. R., Moffat, K., Šrajer, V. & Stoddard, B. L. (1999). *J. Synchrotron Rad.* **6**, 891–917.
Ren, Z. & Moffat, K. (1995). *J. Appl. Cryst.* **28**, 461–481.
Schmidt, M. (2017). *Struct. Dyn.* **4**, 032201.
Schmøkel, M. S., Kamiński, R., Benedict, J. B. & Coppens, P. (2010). *Acta Cryst.* A**66**, 632–636.
Šrajer, V., Crosson, S., Schmidt, M., Key, J., Schotte, F., Anderson, S., Perman, B., Ren, Z., Teng, T., Bourgeois, D., Wulff, M. & Moffat, K. (2000). *J. Synchrotron Rad.* **7**, 236–244.
Šrajer, V. & Schmidt, M. (2017). *J. Phys. D Appl. Phys.* **50**, 373001.
Straasø, T., Müter, D., Sørensen, H. O. & Als-Nielsen, J. (2013). *J. Appl. Cryst.* **46**, 663–671.
Trzop, E., Fournier, B., Jarzembska, K., Sokolow, J., Kaminski, R., Benedict, J., Chen, Y., Henning, R. & Coppens, P. (2014). *Acta Cryst.* A**70**, C776.
Vorontsov, I., Pillet, S., Kamiński, R., Schmøkel, M. S. & Coppens, P. (2010). *J. Appl. Cryst.* **43**, 1129–1130.
Vorontsov, I. I. & Coppens, P. (2005). *J. Synchrotron Rad.* **12**, 488–493.
Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K. & Evans, G. (2018). *Acta Cryst.* D**74**, 85–97.
Wulff, M., Plech, A., Eybert, L., Randler, R., Schotte, F. & Anfinrud, P. (2002). *Faraday Disc.* **122**, 13–26.

electronic reprint