DTU Library

GrainSpotter
a fast and robust polycrystalline indexing algorithm

Schmidt, Søren

Link back to DTU Orbit

GrainSpotter: a fast and robust polycrystalline indexing algorithm

Søren Schmidt

Department of Physics, Technical University of Denmark, Denmark. Correspondence e-mail: ssch@fysik.dtu.dk

A new approach for indexing multigrain diffraction data is presented. It is based on the use of a monochromatic beam simultaneously illuminating all grains. By operating in sub-volumes of Rodrigues space, a powerful vertex-finding algorithm can be applied, with a running time that is compatible with online analysis. The resulting program, *GrainSpotter*, is sufficiently fast to enable online analysis during synchrotron sessions. The program applies outlier rejection schemes, leading to more robust and accurate data. By simulations it is shown that several thousand grains can be retrieved. A new method to derive partial symmetries, called pseudo-twins, is introduced. Uniquely, *GrainSpotter* includes an analysis of pseudo-twins, which is shown to be critical to avoid erroneous grains resulting from the indexing.

1. Introduction

During the past decade, several X-ray-based methods have materialized for identifying and potentially mapping grains in three dimensions in polycrystals. We mention the micro-beam Laue diffraction scanning technique pioneered by Larsson *et al.* (2002), and the two monochromatic beam based diffraction imaging methods known as three-dimensional X-ray diffraction (3DXRD) microscopy (Poulsen *et al.*, 2001; Juul-Jensen *et al.*, 2006) and diffraction contrast tomography (Ludwig *et al.*, 2008, 2009).

In this article we focus on 3DXRD. Within materials science, this technique has found widespread applications for three-dimensional studies of dynamic phenomena such as nucleation (West *et al.*, 2009), recrystallization (Lauridsen *et al.*, 2000; Schmidt *et al.*, 2004), grain growth (Schmidt *et al.*, 2008), phase transformations (Offerman *et al.*, 2002), dislocation dynamics (Jakobsen *et al.*, 2006) and plastic deformation (Margulies *et al.*, 2001; Winther *et al.*, 2004). Within chemistry, pharmacy and even structural biology, 3DXRD and its extension TotalCryst (Sørensen, Schmidt *et al.*, 2012) have provided a way to achieve structural solution and refinement on polycrystals (Schmidt *et al.*, 2003; Vaughan *et al.*, 2004).

A cornerstone in the analysis of multigrain data is a polycrystalline indexing program. In this article we present a new algorithm for indexing and an associated program, *Grain-Spotter*. It is intended for use with a conventional diffraction setup employing a two-dimensional detector with a pixel size of the same order as, or larger than, the grain size. In this case, the diffraction signal comprises distinct diffraction spots positioned in the vicinity of the Debye–Scherrer rings typical of powder diffraction.

The program is one of several (Wright, 2006; Moscicki *et al.*, 2009; Benier *et al.*, 2011; Sharma *et al.*, 2012b) that have been introduced within the past few years to replace the original multi-indexing program *GRAINDEX* (Lauridsen *et al.*, 2001). The underlying algorithm differs in several aspects, such as the use of back projection *versus* forward projection, the use of Friedel pairs and the reliance on an initial assumption that all grains are positioned at the origin.

GrainSpotter has been conceived with the aim of being able to index typical 3DXRD data with ~1000 grains sufficiently quickly that online analysis during synchrotron experiments becomes possible. Furthermore, the focus is on robustness with respect to incomplete data, for example, owing to small grains, grains rotating in and out of the beam *etc*. The basic approach of *GrainSpotter* is to identify grains by finding vertices in orientation space. As demonstrated below, this can be done using Boolean algebra and with a running time that goes as $O(N)$, where N is the number of grains. Once grains have been identified, in a second step the position and orientation of each grain is optimized, and at the same time outliers in the associated set of reflections are removed. Again a fast algorithm is chosen for the fitting. For ultimate accuracy in terms of determining the position, orientation and in addition the strain tensor of each grain, the proposed procedure is to feed the output of *GrainSpotter* into the (slower) optimization program *FitAllB* (Oddershede *et al.*, 2010).

Below we first introduce a new hybrid representation of orientation space, which is required in order to linearize the space. Following a detailed description of the algorithm, simulations are used to verify the program and to test its limitations, in particular in terms of the number of simultaneously illuminated grains. As examples of real experiments where *GrainSpotter* has already been used, we mention multi-grain studies of metals (West *et al.*, 2009; Oddershede *et al.*, 2010, 2011, 2012; Poulsen *et al.*, 2011; Beaudoin *et al.*, 2012), macromolecules (Paithankar *et al.*, 2011), minerals (Sørensen, Hakim *et al.*, 2012) and high-pressure science (Zhanga *et al.*, 2013).

The article also contains a section on pseudo-twins. Unexpectedly, the multigrain work has revealed that diffraction data are subject to an additional symmetry, which to our knowledge has not been considered in conventional crystallography. In analysing multigrain diffraction data of high-symmetry structures, we find it critical to discriminate against erroneous grains that may emerge because of this symmetry. *GrainSpotter* is unique in including such a discrimination. Below we define the pseudo-twin symmetry, explore its properties for various group symmetries and describe how the pseudo-twin analysis is incorporated into *GrainSpotter*.

The *GrainSpotter* source code is publicly available at http://sourceforge.net/apps/trac/fable/browser/GrainSpotter (Schmidt, 2007).

2. Geometry

2.1. Experimental setup

The setup is the conventional setup for single-crystal diffraction using a monochromatic beam (see Fig. 1). We shall assume that the beam at all times fully illuminates the (polycrystalline) sample and that the diffracted beam is transmitted through the sample and acquired on a two-dimensional far-field detector.

During data acquisition a series of diffraction images covering an ω interval $[\omega_{\min}, \omega_{\max}]$ in steps of $\Delta\omega$ are recorded.

2.2. Orientation space

Central to the algorithm presented here is the mathematical representation of orientation space. Conventionally, the choice is between Euler angles, Rodrigues vectors or quater-

Figure 1
Experimental setup. The laboratory and sample reference systems are denoted by (x_l, y_l, z_l) and (x_s, y_s, z_s), respectively. In the figure, the two coordinate systems are related through a rotation around the z_l axis. However, the direction of the rotation axis can be chosen arbitrarily (see details on the Γ rotation matrix in Appendix A). Furthermore, $\mathbf{x}_s$ denotes the position of a grain in the sample, and $\mathbf{x}_{l,i}$ and $\mathbf{x}_{l,i}\Gamma_i^{-1}$ denote the position of a recorded diffraction spot in the laboratory reference system and sample reference system, respectively. $\mathbf{L}_i$ denotes the direction of the diffracted ray in the sample reference system (see details in Appendix A). In addition, $\mathbf{L}_{0,i}$ denotes the direction of the diffracted ray assuming that the grain is located at the origin of the sample reference system. The direction of the ray can also be parametrized through the angles $(2\theta_i, \eta_i, \omega_i)$ (see Appendix A).

nions (Kocks *et al.*, 2000). We argue that the superior solution is a hybrid of the two latter representations.

In Rodrigues space an orientation is represented as a three-dimensional vector, the Rodrigues vector

$$\mathbf{r} = \tan(\varphi/2)\mathbf{n}. \qquad (1)$$

$\mathbf{n}$ is the rotation axis and φ is the rotation angle around the rotation axis. Here orientation means the rotation that allows the transformation of the orthonormal basis representation (Cartesian coordinate system) of the unit cell that coincides with the sample reference system into the orthonormal basis representation of the unit cell of the grain. Each grain corresponds to a single orientation and therefore a single point in Rodrigues space. Let $\mathbf{u}$ be a vector in the reference reciprocal lattice and $\mathbf{v}$ the corresponding vector in the reciprocal lattice of the grain. The set of points in the Rodrigues space (orientations) that rotates $\mathbf{u}$ onto $\mathbf{v}$ is called the geodesic.

A crucial property of this space is that the geodesics are straight lines (Moraviec & Field, 1996). This line can be parameterized by a line (Moraviec & Field, 1996)

$$\mathbf{r}(t) = \mathbf{r}_o^g + t\,\mathbf{r}_s^g, \quad -\infty < t < \infty, \qquad (2)$$

where the origin, $\mathbf{r}_o^g$, and the slope, $\mathbf{r}_s^g$, are given by

$$\mathbf{r}_o^g = \frac{(\mathbf{u} \times \mathbf{v})}{(1 + \mathbf{u} \cdot \mathbf{v})}, \quad \mathbf{r}_s^g = \frac{(\mathbf{u} + \mathbf{v})}{(1 + \mathbf{u} \cdot \mathbf{v})}. \qquad (3)$$

Neglecting the complication of crystal symmetry for the moment, it appears that the task of indexing corresponds to identifying vertices in Rodrigues space. More specifically, for each grain it is necessary to determine a common intersection point (orientation) of multiple geodesics.

The invariant volume in the Rodrigues space can be written as (Moraviec & Field, 1996)

$$dV = (1/\pi)\cos^2(\varphi)\,dr^1\,dr^2\,dr^3, \qquad (4)$$

i.e. being radially symmetric in φ. Hence, for small values of φ,

$$dV \simeq (1/\pi)\,dr^1\,dr^2\,dr^3, \qquad (5)$$

to first order in φ. Consequently, near the origin the Rodrigues space is approximately Euclidean. It is well known that powerful algorithms exist for finding vertices in three-dimensional Euclidian space. Unfortunately, the Rodrigues space rapidly become non-Euclidian with increasing distance to the origin, and for low-symmetry space groups the space is infinite. This implies that the Rodrigues formulation is not generally useful.

The *GrainSpotter* approach is to use quaternion space (which is finite) for general sampling in orientation space, but restrict the key elements in the algorithm to local searches around a certain fixed-point Rodrigues vector. By a simple rotation, the fixed point can be moved to the origin and the local search is restricted to the Euclidean part of the orientation space, hereafter named the local Rodrigues space. This way the global orientation space, *i.e.* all space groups, can be accommodated by subdividing the global orientation space into a series of local Rodrigues spaces. This is the core principle of *GrainSpotter*. The local Rodrigues space may contain
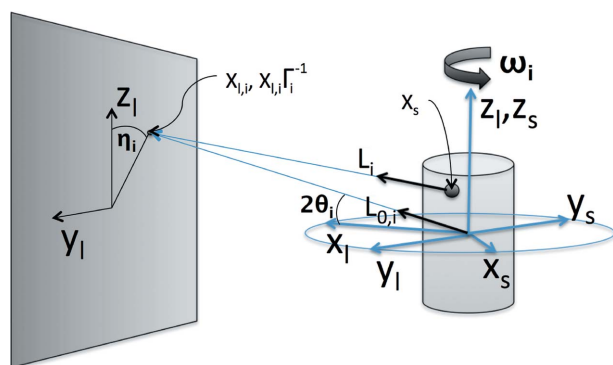
orientations of several grains in the sample. The algorithm for obtaining a fast identification of these vertices is outlined in the following.

3. Algorithm

GrainSpotter operates on a set of diffraction spots, $\{\mathbf{x}_{1,i}\Gamma_i^{-1}\}$ (see Fig. 1). Initially, the set of reciprocal **G** vectors are calculated, assuming that all grains are located at the origin of the sample reference system, *i.e.*

$$\mathbf{G}_{0,i}^{r} = \tfrac{1}{2}\left[\mathbf{L}_{0,i} - \left(\Gamma_i^{-1}\right)|_1\right] \qquad (6)$$

[see equation (32) in Appendix A]. The output is a list of grains and associated properties: the orientations $\{U\}$ (see Appendix A) and (optionally) centre-of-mass positions $\{\mathbf{x}_s\}$. After refinement of the orientation and the position, $\mathbf{L}_{0,i}$ has propagated into $\mathbf{L}_i$ (see Fig. 1). Furthermore, for each grain, descriptors of the quality of indexing are provided, such as the completeness: the ratio between the observed number of reciprocal **G** vectors associated with a grain and the number expected. In the current version *GrainSpotter* assumes all grains to belong to the same phase, and furthermore the space group must be known *a priori*.

As stated previously *GrainSpotter* subdivides the global orientation space into a series of local Rodrigues spaces. Within each subdivision vertices are located by using the initial set of the reciprocal **G** vectors $\{\mathbf{G}_{0,i}^{r}\}$. Only the geodesics from a small fraction of the reciprocal **G** vectors pass through the sub-volume. The subset of **G** vectors can be identified prior to the determination of the geodesics: see the description of fast vertex detection below. Afterwards, one geodesic is calculated for each of the selected **G** vectors. The user-specified size of the local Rodrigues space is given by $\Delta R = \tan(\Delta\varphi/2)$. In order to stay in the Euclidian part of the Rodrigues space $\Delta\varphi$ should be less than $15°$. Consequently, when splitting up the local Rodrigues space into voxels, the density of orientations within each voxel is approximately constant. This property facilitates robust identification of vertices as the size of the voxels must be large enough to capture the deviation from the initial set $\{\mathbf{G}_{0,i}^{r}\}$ to the real set of reciprocal **G** vectors $\{\mathbf{G}_i^{r}\}$. Outside the Euclidian part of the sub-volume the density of orientations in the voxels drops off. Hence, owing to the aforementioned deviation in the reciprocal **G** vectors, the geodesics will no longer cross within a voxel, making identification of vertices less robust. The user-defined parameters including the size of the voxels are discussed in §5.

Following the identification of vertices in a local Rodrigues space the orientations and (optionally) grain centre-of-mass positions are fitted using the reciprocal **G** vectors belonging to each vertex (see details in §§3.3 and 3.4).

A central part of *GrainSpotter* is fast vertex identification. For a given local Rodrigues space the following two steps are carried out:

(1) Identification of the subset of reciprocal **G** vectors in $\{\mathbf{G}_{0,i}^{r}\}$ from which the corresponding geodesics pass through the local Rodrigues space. This is done without searching through the full list of reciprocal **G** vectors. The potential reciprocal **G** vectors must be in the neighbourhood of the predicted directions $\{\mathbf{G}_{\text{centre},k}^{r}\}$ derived from the centre point in the local Rodrigues space, U_{centre}. During startup of *Grain-Spotter*, a look-up table is constructed for $\{\mathbf{G}_{0,i}^{r}\}$, ensuring that for a given predicted direction, $\mathbf{G}_{\text{centre},k}^{r}$, all potential reciprocal **G** vectors are identified.

(2) Finding the vertices. These are obtained by mapping all geodesics into the local Rodrigues space, but for each voxel keeping track of the reciprocal **G** vectors that have already passed through this voxel. This way it is possible to group the reciprocal **G** vectors that belong to the same vertex without searching through the local Rodrigues space itself. This is a two-step selection procedure (see details in §§3.1 and 3.2). In the first selection step a grouping of reciprocal **G** vectors potentially belonging to the same vertex is made. The second selection step identifies the vertices and verifies the reciprocal **G** vectors belonging to different vertices. The complexity of this method scales with the number of reciprocal **G** vectors.

3.1. First selection

The reciprocal **G** vectors with geodesics passing through the local Rodrigues space are labelled 1 to N_g. Next a four-dimensional data structure is built, with the first three dimensions being identical to the voxels in the local Rodrigues space. The fourth dimension comprises a list of the geodesics (lines) passing through each voxel. To save memory and increase speed the list is implemented as an array of unsigned integers (16 bits), as illustrated in Fig. 2. For each voxel hit by
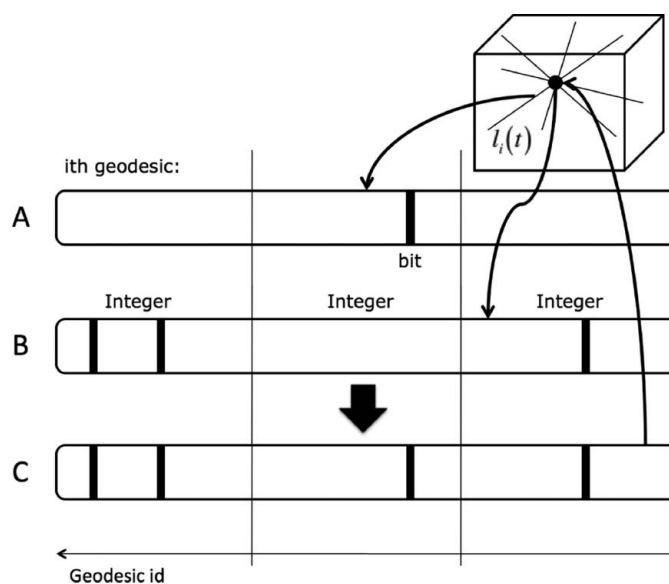
Figure 2
Illustration of the first search. The selection is carried out in a four-dimensional space with the first three dimensions being the local Rodrigues space (voxelized) and the fourth dimension a mask of the geodesic labels passing through the voxels. Shown is the Boolean OR operation between a bit indicating the number of the geodesic (A) and the fourth dimension for a given voxel (B), leading to an update of this fourth dimension, representing the geodesics that have been found to pass the voxel so far (C).

Figure 3
Illustration of the second selection.
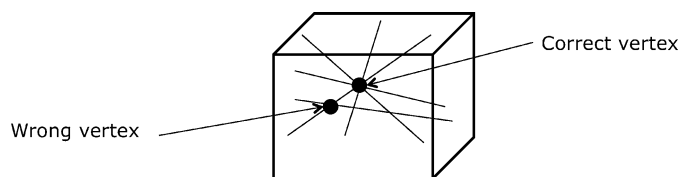
the ith geodesic a logical OR operation with a mask is carried out. This keeps track of the geodesics visiting this particular voxel.

In addition, the set of geodesics that the ith geodesic has crossed are also stored in a separate two-dimensional structure by making a logical OR operation with all the voxels that the geodesic passed through. This way, the ith **G** vector keeps track of the possible intersections of the ith geodesic with the previous $i-1$ geodesics.

3.2. Second selection

The two-dimensional structure lists the possible intersections of geodesics. However, not all crossings of geodesics are bound (see Fig. 3). Consequently, a second selection is done. For each table entry the geodesics are mapped into the modified four-dimensional space, keeping track of the voxel with the maximum hits, giving a candidate orientation for further investigation. Only the last $N_g - N_{min}$ table entries are processed, where N_{min} is a lower limit of measurements for a grain specified by the user. Note, for both the first and second selection, the vertices are obtained without searching through the whole local Rodrigues space, implying that all candidate grains have been identified with an $O(N)$ complexity ($2N_g - N_{min}$ geodesics are mapped in the combined selection).

3.3. Fitting position and orientation

The orientation space (local Rodrigues space) and direct space basically have the same geometrical properties: The fitted grain position is by definition the point in the sample reference system that minimizes the distance to the back-projected rays from the centre-of-mass diffraction spots associated with the grain. Equivalently, the fitted orientation is by definition the point in the local Rodrigues space that minimizes the distances to the geodesics from the scattering vectors.[1]

The position of the grain in the sample reference system is denoted by x_s (see Fig. 1). Each ray is parametrized by the position of the diffraction spot in the sample reference system, $x_{l,i}\Gamma_i^{-1}$, and the direction of the ray in the sample reference system, L_i. More specifically, x_s is obtained by minimizing the following expression:

$$\chi^2(x_s) = \sum_i \left[|\Delta x_i|^2 - (L_i \cdot \Delta x_i)^2 \right], \qquad (7)$$

where $\Delta x_i = x_{l,i}\Gamma_i^{-1} - x_s$. Hence,

[1] Note that there is only one geodesic per reciprocal scattering vector in the local Rodrigues space.

$$x_s = S^{-1}b, \qquad (8)$$

where

$$S_{pq} = \sum_i \left[L_{p,i}L_{q,i} - \delta_{pq} \right], \quad b = \sum_i \left[L_i \left(L_i \cdot x_{l,i}\Gamma_i^{-1} \right) - x_{l,i}\Gamma_i^{-1} \right]. \qquad (9)$$

The case for orientations is completely equivalent. Equation (7) can be reused by replacing $x_{l,i}\Gamma_i^{-1}$ with the origin of the geodesics, r_o^g, and L_i with the direction of the geodesic, r_s^g [see equation (3)].

3.4. Fitting procedure and outlier removal

The set of collected **G** vectors may contain falsely assigned vectors, typically originating from other grains or caused by noisy data. The following outlier rejection procedure aims at removing these wrongly assigned measurements:

Following the minimization of equation (7), the quality or the agreement of the individual measurement i with the overall fit is evaluated by the estimator f_i,

$$f_i = \frac{\chi_i^2}{\psi_{max}^2} \frac{\chi_i^2 N}{\chi^2}, \qquad (10)$$

where χ_i^2 is the contribution to χ^2 from the ith measurement, N is the number of measurements in the fit and ψ_{max} is a predefined maximal angle deviation, which is related to the measurement uncertainty and is specified by the user. Generally, the lower the value of f_i the better the quality or agreement of the overall fit for the ith measurement. Note that f_i consists of two factors. The χ_i^2/ψ_{max}^2 factor punishes measurements with deviations larger than an absolute value specified by ψ_{max}. The second factor, $\chi_i^2 N/\chi^2$, punishes measurements that are larger than the mean χ^2 contribution, *i.e.* a relative limit. Both factors taken together result in a good quality for a given measurement if the deviation is smaller than ψ_{max} or if the deviation is smaller than the mean deviation (and potentially larger than ψ_{max}). Consequently, by keeping measurements with $f_i \leq 1$ the outliers with deviations larger than expected and larger than the mean deviation for the whole fit are removed.

Removing outliers and fitting the orientation is an iterative procedure. For a given set of measurements an orientation estimate is obtained through equations (8) and (9). Afterwards measurements with

$$f_i > 1 \qquad (11)$$

are removed. A new fit with the remaining measurement is done. The procedure is repeated until a stable solution is found or the number of remaining measurements goes below a user-specified minimum.

If the position is estimated as well, the procedure alternates between the orientation fit and the position fit. Following the position fit the directions of the **G** vectors are updated according to the new origin. No selection is made. Following the orientation fit the outlier filtering is done according to the procedure described above.

research papers

3.5. Implementation

The source code is written in C, and apart from standard libraries the software package *SgInfo* (Grosse-Kunstleve, 1994) is used to generate equivalent matrices and theoretical scattering vectors for a given space group.

4. Partial symmetry analysis

It is well known that crystal symmetry implies that orientations can be equivalent: that is they are associated with exactly the same **G** vectors. When analysing multigrain data, partial symmetries do also occur: in this case the orientations only share some **G** vectors [see *e.g.* work by Grimmer *et al.* (1974) on coincidence-site lattices for the cubic system]. As an example, the partial symmetries appearing for the first five *hkl* families for a face-centred cubic material are shown in Fig. 4. For a given orientation, 274 other orientations exist with a partial overlap of reflections – all within the fundamental zone (where there are no symmetry equivalents). We shall call these pseudo-twins. Out of a total of 112 **G** vectors, pseudo-twins with 34, 24, 16, ten, eight, six, four and two reflections occur. The ones with 34 reflections correspond to regular first-order twins.

Experience shows that for high-symmetry structures it is crucial to include an analysis of pseudo-twins in the polycrystalline indexing algorithm. For the above example, one may argue that a completeness threshold of 1/3 or higher would remove these. However, in real data **G** vectors may be wrongly assigned to grains owing to either closely positioned measurements on the same diffraction rings or closely positioned diffraction rings. In both cases, the completeness of the pseudo-twins may end up systematically higher. To overcome this problem, it is relevant for each orientation found to
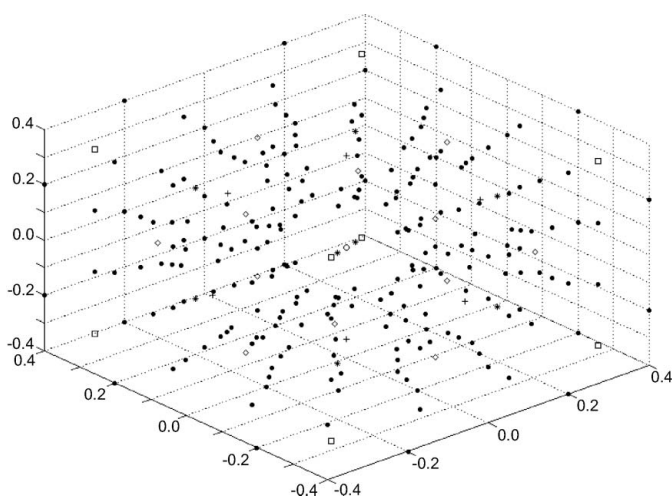

Figure 4
Partial symmetries shown in Rodrigues space for space group $Fm\bar{3}m$, first five *hkl* families (in total 112 theoretical reflections). The true orientation is located at $\mathbf{r} = (0, 0, 0)$. The remaining spots identify the partial symmetries, *i.e.* where only a fraction of the theoretical scattering vectors are mapped upon themselves. Of the 112 reflections, pseudo-twins with 34, 24, 16, ten, eight, six, four and two reflections occur.

investigate the pseudo-twin orientation and see if these are associated with a higher completeness or better fits. Such an analysis is also helpful in cases where real twinning occurs and in general for noisy data. In the following, a method to derive pseudo-twins for all space groups is presented.

To derive an algorithm for finding the pseudo-twin orientation for a given trial orientation, we start by considering a pair of theoretical scattering vectors in reciprocal space mapped into a pair of **G** vectors, $\mathbf{h}_1 \to \mathbf{g}_1$ and $\mathbf{h}_2 \to \mathbf{g}_2$, *i.e.* $\mathbf{h}_1 \cdot \mathbf{h}_2 = \mathbf{g}_1 \cdot \mathbf{g}_2$. Following the convention of Busing & Levy (1967), the crystallographic orientation U, fulfilling $\mathbf{g}_1 = U\mathbf{h}_1$ and $\mathbf{g}_2 = U\mathbf{h}_2$, can be expressed as

$$U = T(\mathbf{g}_1, \mathbf{g}_2)T(\mathbf{h}_1, \mathbf{h}_2)^{-1}, \tag{12}$$

given by a product of two orthogonal matrices. The first column vector in $T(\mathbf{g}_1, \mathbf{g}_2)$ is a unit vector along $\mathbf{g}_1$. The third column is a unit vector perpendicular to the plane spanned by $\mathbf{g}_1$ and $\mathbf{g}_2$. The second column is a unit vector perpendicular to the first and third columns, *i.e.* in the plane spanned by $\mathbf{g}_1$ and $\mathbf{g}_2$. Likewise with respect to $T(\mathbf{h}_1, \mathbf{h}_2)$.

Now consider a different pair of theoretical scattering vectors, $(\mathbf{t}_1, \mathbf{t}_2)$, where $|\mathbf{t}_1| = |\mathbf{h}_1|$, $|\mathbf{t}_2| = |\mathbf{h}_2|$ and $\mathbf{t}_1 \cdot \mathbf{t}_2 = \mathbf{h}_1 \cdot \mathbf{h}_2$. The orientation matrix that brings $\mathbf{t}_1 \to \mathbf{g}_1$ and $\mathbf{t}_2 \to \mathbf{g}_2$ is then given by

$$U_2 = T(\mathbf{g}_1, \mathbf{g}_2)T(\mathbf{t}_1, \mathbf{t}_2)^{-1}. \tag{13}$$

If U and U_2 are equivalent orientations, then

$$U_2\mathcal{B} = U\mathcal{B}E_i \tag{14}$$

for some generator E_i of the space group in question, where $\mathcal{B}$ is the correspondence between the Cartesian *hkl* lattice and reciprocal space (see Appendix A). Rearranging equation (14),

$$T(\mathbf{t}_1, \mathbf{t}_2)T(\mathbf{h}_1, \mathbf{h}_2)^{-1} = \mathcal{B}E_i\mathcal{B}^{-1}, \tag{15}$$

we find that the measured **G** vectors, $T(\mathbf{g}_1, \mathbf{g}_2)$, have cancelled out. Consequently, equation (15) provides a general criterion for recovering the same orientation when switching pairs of theoretical reflections. In other words, the orientation that brings one pair of theoretical reflections onto another pair $(\mathbf{h}_1 \to \mathbf{t}_1$ and $\mathbf{h}_2 \to \mathbf{t}_2)$ must be a symmetry-equivalent operation. More specifically,

$$W = T(\mathbf{t}_1, \mathbf{t}_2)T(\mathbf{h}_1, \mathbf{h}_2)^{-1} \begin{cases} = \mathcal{B}E_i\mathcal{B}^{-1}: \text{ full symmetry,} \\ \neq \forall \mathcal{B}E_i\mathcal{B}^{-1}: \text{ partial symmetry.} \end{cases} \tag{16}$$

For a given space group and unit cell the partial symmetries or pseudo-twins, W_j, can be calculated using equation (16). Pseudo-twin orientations are characterized by matching only a fraction of the true measurements. For each candidate orientation, U_c, found in the indexing procedure the true orientation is given by the orientation in the list

$$U_c, U_cW_1, \ldots, U_cW_n \tag{17}$$

that maximizes the completeness estimator, *i.e.* the fraction of measurements found with respect to the number of measurements expected.

5. Simulations

To verify the indexing procedure, Monte Carlo simulations were performed. The sample was assumed to be a $500 \times 500 \times 500\ \mu\text{m}$ Al polycrystal comprising N grains of random orientation. At 50 keV X-ray energy, exposures were made on a perfectly aligned distortion-free two-dimensional detector with $50\ \mu\text{m}$ pixels, placed at a distance of 20 cm from the sample. The simulated data covered a rotation range of $180°$ in steps of $0.25°$ and included only the five hkl families with the highest d spacings.

The associated diffraction peaks are assumed to be ideal intensity spikes in the detector images; in other words the issue of spot broadening is neglected. On the other hand, experimental errors in the centre-of-mass (CMS) positions of the spots were included by adding Gaussian noise to the three angles defining the **G** vector: 2θ, η and ω (see Fig. 1). The

Figure 5
Performance test. The execution time shown as a function of number of grains: dotted line with boxes for $\Delta R = \tan(4°/2)$, and dotted line with diamonds for $\Delta R = \tan(6°/2)$. The purity, *i.e.* the fraction of correctly assigned **G** vectors (solid line), shows a linear and decreasing dependency on the number of grains, due to an increasing probability that **G** vectors from other grains will appear close to the expected location.

Figure 6
Performance test with 3000 grains. Misorientations (top left) with an average error of $0.025°$. Deviations in position for x (top right), y (bottom left) and z (bottom right).

spreads were $\sigma_{2\theta}^{s} = 0.025°$, $\sigma_{\eta}^{s} = 0.05°$ and $\sigma_{\omega}^{s} = 0.125°$. The resulting distribution of the internal angular deviation between the **G** vector and the predicted direction resembles the observed distribution when indexing real data.[2]

Simulations with up to 3000 grains were conducted, where in all cases all grains were retrieved and no erroneous grains were found. *GrainSpotter* was run with the following user-defined parameters: $\sigma_{2\theta} = 0.05°$, $\sigma_{\eta} = 0.1°$, $\sigma_{\omega} = 0.2°$ and $N_{\sigma} = 3$. These quantities have multiple functionalities. At startup each **G** vector will be associated with the set of (hkl) families for which $|2\theta_{\mathbf{G}^{r}} - 2\theta_{(hkl)}| \leq N_{\sigma}\sigma_{2\theta}$. As stated previously, the outlier removal procedure is also based on the uncertainties with $\psi_{\max} = N_{\sigma}(\sigma_{2\theta} + \sigma_{\eta} + \sigma_{\omega})$ in equation (10). The size of the voxels in the local Rodrigues space is given by a conservative estimate: $\delta R = \psi_{\max}/12^{1/2}$. The user-specified size of the local Rodrigues space was $\Delta R = \tan(\Delta\varphi/2)$, yielding $\Delta R/\delta R = 10$ voxels along each dimension for $\Delta\varphi = 4°$. Equivalently, $\Delta\varphi = 6°$ corresponds to $\Delta R/\delta R = 15$. A total of 100 000 random trials were selected for the search. Shown in Fig. 5 is the computing time for $N = 100, 200, \ldots, 1000$ grains using one thread (*i.e.* a single core) on a 2.7 GHz Intel i7 processor and the corresponding purity, defined as the grain average of the ratio between the number of correct reflections associated by *GrainSpotter* and the number of simulated spots for the grain. It is evident that the running time is approximately linear in the number of grains ($\sim 10^{-4}N^2 + 10^{-1}N$) for $\Delta\varphi = 4°$, and the result of less than 4 min for extracting 1000 grains complies well with the aspiration to perform indexing online during synchrotron experiments. The resulting purity of above 99% even at $N = 1000$ is also encouraging. When the size of the local Rodrigues space is enlarged, an enhancement in performance speed is seen for the case $\Delta\varphi = 6°$ in Fig. 5. This is due to the higher rate of grains indexed per trial. However, in terms of complexity the N^2 term has become more pronounced ($\sim 10^{-4}N^2 + 10^{-2}N$).

In Fig. 6 the corresponding histograms for the error in CMS position and orientation components for the case of $N = 3000$ are shown. The error on the position in x and y is $15\ \mu\text{m}$ and that for z is $9\ \mu\text{m}$. Whereas errors on x and y are similar, the error on z is smaller because the z axis is parallel to the rotation axis ω. To index all grains *GrainSpotter* was run several times, saving the unassigned **G** vectors for the following indexing step. Starting with $\Delta R = \tan(2°/2)$ at each step, looser cuts were applied compared to the previous step. The whole process took 50 min on one core. The purity was 97.4%.

6. Discussion

Evidently the simulations represent an idealized situation. In actual experiments phenomena such as grain mosaicity, stress and sample texture will affect performance. A particular concern is the grain size distribution, which in many cases implies that the number of grains visible decreases strongly
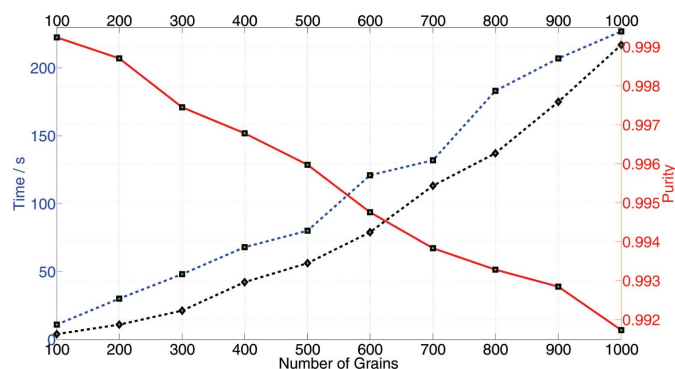
---

[2] The distribution itself is approximately a generalized extreme value probability distribution function of the second type (Coles, 2001).

with decreasing d spacing. Discussions of these effects and suggestions for how to handle them can be found elsewhere (*e.g.* Schmidt *et al.*, 2003; Poulsen, 2004). For a systematic discussion on error propagation in relation to multigrain indexing the reader is referred to the work by Benier *et al.* (2011) and Sharma *et al.* (2012*b*).

Recently several other approaches have been proposed for polycrystalline indexing. We mention the following:

(1) The ImageD11 indexing algorithm is based on sorting of reflections in a combinatorial approach (Wright, 2006): All pairs of **G** vectors are created and the internal angle is compared with a set of theoretical angles. If a mutually consistent subset is found containing a minimum number of pairs it is stored as a grain. This is a simple and effective algorithm and is believed to go as $O(N^2)$, with N being the number of reflections.

(2) Ludwig *et al.* (2009) suggested an approach based on Friedel pairs, through the use of a near-field detector and rotation of the sample by $360°$. This is an effective solution for extended specimens and has been used for a number of grain mapping experiments. As an example, 1008 grains were indexed simultaneously in an experimental data set on β-Ti. In terms of speed, the algorithm again is believed to go as $O(N^2)$.

(3) Moscicki *et al.* (2009) developed an algorithm based on Friedel pairs and far-field data, which was verified on nine grains.

(4) Benier *et al.* (2011) operated on far-field data and proposed a back-projecting algorithm with some similarity to *GrainSpotter*. The focus is here in particular on strain analysis. In short, for selected diffraction spots, a search for candidate grains is performed along the associated geodesic (also known as a fibre). The software was successfully demonstrated on an aggregate of 819 grains.

(5) Sharma *et al.* (2012*b*), operating also on far-field data, put the emphasis on large data sets and introduced an indexing formalism that simultaneously exploits the constraints in both direct and orientation space. They demonstrated the software on a simulated data set with 3000 grains.

Without round-robin tests it is not possible to conclude on relative performance, but it appears that none of the algorithms can be labelled as universally ideal, and they also tend to be motivated, at least in part, by different performance criteria. Nevertheless, *GrainSpotter* is to my knowledge the first indexing routine to be used for online data analysis.

The *GrainSpotter* algorithm can be generalized in several ways:

(i) *Grainspotter* works on segmented data, *i.e.* data where the diffraction spots have been harvested in the diffraction images. To ensure that the total throughput is high, the focus so far has been based on segmentation *via* thresholding and the use of connected components (pixel connectivity search, where detector pixels are grouped into spots) type algorithms. However, *Grainspotter* is directly compatible with any segmentation scheme, such as watersheding, relevant in cases where spot overlap is a concern (Kenesei, 2010; Sharma *et al.*, 2012*a*).

(ii) It is possible to offset the centre of origin in the code. By repeating *Grainspotter* with several origins and picking the best grains from the set of resulting outputs, one can extend the use of *GrainSpotter* in the direction of large N and extended samples.

(iii) *GrainSpotter* is well adapted to perform indexing also in cases of unknown space groups, of interest in particular in geoscience and chemistry. Work on this topic will be reported elsewhere.

(iv) *GrainSpotter* is easy to parallelize, as the computation of candidate grains for local Rodrigues spaces can be performed independently.

A partial symmetry analysis is also relevant in situations where only a fraction of the expected reciprocal space is covered, *e.g.* when the sample is partly illuminated by the beam (grains moving in and out of the beam), or if the active area of the detector is not taken fully into account (partly illuminated diffraction rings, dead areas *etc.*). In these cases the completeness estimator may be lower than expected for the best partial symmetry. However, by carefully examining the set of (hkl)s matched against data one can determine if the orientation is ambiguous or true, as certain combinations of (hkl)s uniquely determine the orientation. For complete data sets the classical way of requiring a high value of the completeness estimator, as an alternative to the pseudo-twin analysis, is also available as a user-defined option in *Grain-Spotter*.

APPENDIX *A*
Forward projection

In the following we derive equations for determining the directions of the diffraction vector in the sample reference system given a grain with orientation U and a lattice plane (h, k, l). Generally, each **G** vector gives rise to two directions in the sample reference system, which will be shown in the following. The emphasis is on a formulation that leads to efficient programming. In addition, previous work on establishing the geometry of the forward projection (Poulsen, 2004) is generalized.

Poulsen (2004) gives the diffraction equation as

$$\mathbf{G}_i = \frac{d_i}{2\pi} \Omega_i U \mathcal{B} \mathbf{G}_{hkl,i}. \tag{18}$$

Here $\mathbf{G}_i$ is the scattering vector ($|\mathbf{G}_i| = 1$) in the laboratory system of reflection i, d_i is the spacing between the planes in the atomic lattice for the ith reflection, Ω_i is a right-hand rotation, ω_i, round the z axis for the ith reflection,

$$\Omega_i = \begin{pmatrix} \cos \omega_i & -\sin \omega_i & 0 \\ \sin \omega_i & \cos \omega_i & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{19}$$

and $\mathcal{B}$ is the correspondence between the Cartesian hkl lattice,

$$\mathbf{G}_{hkl} = \begin{pmatrix} h \\ k \\ l \end{pmatrix}, \tag{20}$$

and reciprocal space,

$$\mathcal{B} = \begin{pmatrix} a^* & b^* \cos \gamma^* & c^* \cos \beta^* \\ 0 & b^* \sin \gamma^* & -c^* \sin \beta^* \cos \alpha \\ 0 & 0 & c^* \sin \beta^* \sin \alpha \end{pmatrix}, \tag{21}$$

$$\cos \alpha = \frac{\cos \beta^* \cos \gamma^* - \cos \alpha^*}{\sin \beta^* \sin \gamma^*}, \tag{22}$$

where $(a, b, c, \alpha, \beta, \gamma)$ and $(a^*, b^*, c^*, \alpha^*, \beta^*, \gamma^*)$ are the lattice parameters in direct and reciprocal space.

In the following we generalize equation (18) by introducing the generalized rotation matrix Γ_i,

$$\Gamma_i = \Phi_x \Phi_y \Omega_i = R \Omega_i, \tag{23}$$

where Φ_x and Φ_x are known right-handed rotations around the x and y axis, thus accommodating all possible directions of the rotation axis. Furthermore, the length of the scattering vector is changed to $|\mathbf{G}_i| = \lambda/(2d_i)$, where λ is the wavelength of the monochromatic beam. The scattering vector in the sample system, the reciprocal vector $\mathbf{G}_i^r$, is defined by

$$\mathbf{G}_i^r = \Gamma_i^{-1} \mathbf{G}_i = \frac{\lambda}{4\pi} U \mathcal{B} \mathbf{G}_{hkl,i}. \tag{24}$$

Expressing $\mathbf{G}_i$ as a function of 2θ and η we have

$$\mathbf{G}_i = \frac{1}{2} \begin{pmatrix} \cos 2\theta - 1 \\ -\sin 2\theta \sin \eta \\ \sin 2\theta \cos \eta \end{pmatrix}_i. \tag{25}$$

Since $(\mathbf{G}_i)_1$ is independent of η_i,

$$C_i = A_i \cos \omega_i + B_i \sin \omega_i, \tag{26}$$

where

$$A_i = R_{11}(\mathbf{G}_i^r)_1 + R_{12}(\mathbf{G}_i^r)_2, \tag{27}$$

$$B_i = R_{21}(\mathbf{G}_i^r)_1 - R_{11}(\mathbf{G}_i^r)_2, \tag{28}$$

$$C_i = -\left|\mathbf{G}_i^r\right|^2 - R_{13}(\mathbf{G}_i^r)_3. \tag{29}$$

The solutions are given by

$$\cos \omega_{i,j} = \frac{A_i C_i \pm B_i D_i^{1/2}}{A_i^2 + B_i^2}, \quad \sin \omega_{i,j} = \frac{B_i C_i \mp A_i D_i^{1/2}}{A_i^2 + B_i^2}, \tag{30}$$

$$D_i = A_i^2 + B_i^2 - C_i^2. \tag{31}$$

The solution can easily be verified by substituting $\sin \omega = [\exp(i\omega) - \exp(-i\omega)]/2i$ and $\cos \omega = [\exp(i\omega) + \exp(-i\omega)]/2$ and solving for $\exp(i\omega)$. Consequently, we can determine the direction of the diffraction vector in real space, $\mathbf{L}_{i,j}$,

$$\begin{aligned} \mathbf{L}_{i,j} &= \Gamma_{i,j}^{-1} \begin{pmatrix} \cos 2\theta \\ -\sin 2\theta \sin \eta \\ \sin 2\theta \cos \eta \end{pmatrix}_{i,j} \\ &= 2\Gamma_{i,j}^{-1} \mathbf{G}_{i,j} + \Gamma_{i,j}^{-1} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\ &= 2\mathbf{G}_i^r + \left(\Gamma_{i,j}^{-1}\right)|_1, \end{aligned} \tag{32}$$

where $(\Gamma_{i,j}^{-1})|_1$ is the first column vector of $\Gamma_{i,j}^{-1}$. Note that $|\mathbf{L}_{i,j}| = 1$. For each reciprocal vector $\mathbf{G}_i^r$ there exist (up to) two direct space vectors, $\mathbf{L}_{i,j}$. Also note that $\mathbf{L}_{i,j}$ can be expressed in terms of the reciprocal vector, $\mathbf{G}^r$, without reference to ω, η and 2θ.

Note that the formalism presented here also facilitates an indexing procedure when recording data using multiple rotation axes.

References

Beaudoin, A. J., Obstalecki, M., Storer, R., Tayon, W., Mach, J., Kenesei, P. & Lienert, U. (2012). *Model. Simul. Mater. Sci. Eng.* **20**, 024006.

Benier, J. V., Barton, N. R., Lienert, U. & Miller, M. P. (2011). *J. Strain Anal. Eng. Des.* **46**, 527–547.

Busing, W. R. & Levy, H. A. (1967). *Acta Cryst.* **22**, 457–464.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* London: Springer-Verlag.

Grimmer, H., Bollmann, W. & Warrington, D. H. (1974). *Acta Cryst.* A**30**, 197–207.

Grosse-Kunstleve, R. W. (1994). *SgInfo – Space Group Info*, http://cci.lbl.gov/sginfo/.

Jakobsen, B., Poulsen, H. F., Lienert, U., Almer, J., Shastri, S. D., Sorensen, H. O., Gundlach, C. & Pantleon, W. (2006). *Science*, **312**, 889–892.

Juul Jensen, D., Lauridsen, E. M., Margulies, L., Poulsen, H. F., Schmidt, S., Sorensen, H. O. & Vaughen, G. B. M. (2006). *Mater. Today*, **9**, 18–25.

Kenesei, P. (2010). *DIGIgrain*, http://sourceforge.net/apps/trac/digigrain/.

Kocks, U. F., Tome, C. N., Wenk, H.-R. & Mecking, H. (2000). *Texture and Anisotropy: Preferred Orientations in Polycrystals and Their Effect on Materials Properties.* Cambridge University Press.

Larson, B. C., Yang, W., Ice, G. E., Budai, J. D. & Tischler, J. Z. (2002). *Nature*, **415**, 887–890.

Lauridsen, E. M., Jensen, D. J., Poulsen, H. F. & Lienert, U. (2000). *Scr. Mater.* **43**, 561–566.

Lauridsen, E. M., Schmidt, S., Suter, R. M. & Poulsen, H. F. (2001). *J. Appl. Cryst.* **34**, 744–750.

Ludwig, W., Herbig, M., Buffire, J. Y., Ludwig, W., Reischig, P., King, A., Johnson, G., King, A., Johnson, G., Marrow, T. J. & Lauridsen, E. M. (2009). *Rev. Sci. Instrum.* **80**, 033905.

Ludwig, W., Schmidt, S., Lauridsen, E. M. & Poulsen, H. F. (2008). *J. Appl. Cryst.* **41**, 302–309.

Margulies, L., Winther, G. & Poulsen, H. F. (2001). *Science*, **291**, 2392–2394.

Moraviec, A. & Field, D. P. (1996). *Philos. Mag. A*, **73**, 1113–1130.

Moscicki, M., Kenesei, P., Wright, J., Pinto, H., Lippmann, T., Borbely, A. & Pyzalla, A. R. (2009). *Mater. Sci. Eng. A*, **524**, 64–68.

Oddershede, J., Camin, S., Schmidt, S., Mikkelsen, L. P., Sørensen, H. O., Lienert, U., Poulsen, H. F. & Reimers, W. (2012). *Acta Mater.* **60**, 3570–3580.

Oddershede, J., Schmidt, S., Poulsen, H. F., Margulies, M., Wight, J., Moscickic, M., Reimers, W. & Winther, G. (2011). *Mater. Charact.* **62**, 651–660.

Oddershede, J., Schmidt, S., Poulsen, H. F., Sørensen, H. O., Wright, J. & Reimers, W. (2010). *J. Appl. Cryst.* **43**, 539–549.

Offerman, S. E., van Dijk, N. H., Sietsma, J., Grigull, S., Lauridsen, E. M., Margulies, L., Poulsen, H. F., Rekveldt, M. T. & van der Zwaag, S. (2002). *Science*, **298**, 1003–1005.

Paithankar, K. S., Sørensen, H. O., Wright, J. P., Schmidt, S., Poulsen, H. F. & Garman, E. F. (2011). *Acta Cryst.* D**67**, 608–618.

Poulsen, H. F. (2004). *Three-Dimensional X-ray Diffraction Microscopy. Mapping Polycrystals and Their Dynamics*, Tracts in Modern Physics, Vol. 205. Berlin: Springer-Verlag.

Poulsen, H. F., Nielsen, S. F., Lauridsen, E. M., Schmidt, S., Suter, R. M., Lienert, U., Margulies, L., Lorentzen, T. & Juul Jensen, D. (2001). *J. Appl. Cryst.* **34**, 751–756.

Poulsen, S. O., Lauridsen, E. M., Lyckegaard, A., Oddershede, J., Gundlach, C., Curfs, C. & Juul Jensen, D. (2011). *Scr. Mater.* **64**, 1003–1006.

Schmidt, S. (2007). *GrainSpotter*, http://sourceforge.net/apps/trac/fable/browser/GrainSpotter.

Schmidt, S., Nielsen, S. F., Gundlach, C., Margulies, L., Huang, X. & Jensen, D. J. (2004). *Science*, **305**, 229–232.

Schmidt, S., Olsen, U. L., Poulsen, H. F., Sørensen, H. O., Lauridsen, E. M., Margulies, L., Maurice, C. & Juul Jensen, D. (2008). *Scr. Mater.* **59**, 491–494.

Schmidt, S., Poulsen, H. F. & Vaughan, G. B. M. (2003). *J. Appl. Cryst.* **36**, 326–332.

Sharma, H., Huizenga, R. M. & Offerman, S. E. (2012a). *J. Appl. Cryst.* **45**, 693–704.

Sharma, H., Huizenga, R. M. & Offerman, S. E. (2012b). *J. Appl. Cryst.* **45**, 705–718.

Sørensen, H. O., Hakim, S. S., Pedersen, S., Christiansen, B. C., Balogh, Z., Hem, C. P., Pasarin, S., Schmidt, S., Olsen, U. L., Oddershede, J. C. F., Feidenhansĺ, R. & Stipp, S. L. S. (2012). *Can. Mineral.* **50**, 501–509.

Sørensen, H. O., Schmidt, S., Wright, J. P., Vaughan, G. B. M., Techert, S., Garman, E. F., Oddershede, J., Davaasambuu, J., Paithankar, K. S., Gundlach, C. & Poulsen, H. F. (2012). *Z. Kristallogr.* **227**, 63–78.

Vaughan, G. B. M., Schmidt, S. & Poulsen, H. F. (2004). *Z. Kristallogr.* **219**, 813–825.

West, S. S., Schmidt, S., Sørensen, H. O., Winther, G., Poulsen, H. F., Margulies, L., Gundlach, C. & Juul Jensen, D. (2009). *Scr. Mater.* **61**, 875–878.

Winther, G., Margulies, L., Schmidt, S. & Poulsen, H. F. (2004). *Acta Mater.* **52**, 2863–2872.

Wright, J. P. (2006). *ImageD11*, http://sourceforge.net/apps/trac/fable/wiki/imaged11.

Zhanga, L., Meng, Y., Dera, P., Wenge, Y., Wendy, L. M. & Mao, H.-K. (2013). *Proc. Natl Acad. Sci. USA*, **110**, 6292–6295.