

# Indexing of multi-particle diffraction data in a high-pressure single-crystal diffraction experiment

Hui Li,<sup>a\*</sup> Xiaodong Li,<sup>b</sup> Meng He,<sup>c</sup> Yanchun Li,<sup>b</sup> Jing Liu,<sup>b</sup> Guoyin Shen<sup>d</sup> and Ze Zhang<sup>e</sup>

<sup>a</sup>Beijing University of Technology, Beijing 100124, People's Republic of China, <sup>b</sup>Institute of High Energy Physics, CAS, Beijing 100049, People's Republic of China, <sup>c</sup>National Center for Nanoscience and Technology, Beijing 100190, People's Republic of China, <sup>d</sup>Advanced Photon Source, HP-CAT, Argonne National Laboratory, IL 60439, USA, and <sup>e</sup>Zhejiang University, Hangzhou 310014, People's Republic of China. Correspondence e-mail: huilicn@yahoo.com

High-pressure single-crystal diffraction experiments often suffer from the crushing of single crystals due to the application of high pressure. Consequently, only diffraction data resulting from several particles in random orientations is available, which cannot be routinely indexed by commonly used methods designed for single-crystal data. A protocol is proposed to index such diffraction data. The techniques of powder pattern indexing are first used to propose the possible lattice parameters, and then a genetic algorithm is applied to determine the orientation of the reciprocal lattice for each of the particles. This protocol has been verified experimentally.

© 2013 International Union of Crystallography  
 Printed in Singapore – all rights reserved

## 1. Introduction

Pressure has an important effect on the structure and properties of materials, and high-pressure techniques have been attracting more and more interest from researchers in diverse fields. Up to now, one of the most well developed crystallographic techniques under high-pressure conditions has been powder diffraction with a diamond anvil cell. Nevertheless, this technique has many shortcomings: usually strong preferred orientation, overlapping of diffraction peaks, low resolution, low ratio of signal to noise, a sharp decrease of diffraction intensities with pressure, and so on. More importantly, positional information on reflections is lost to a great extent in powder diffraction, and thus full structural characterization is usually difficult for phases existing only under high-pressure conditions. Single-crystal diffraction data are very desirable, but it is a great challenge to collect single-crystal data under high-pressure conditions. Since the mid-1960s, great effort has been made in the development of high-pressure single-crystal diffraction facilities, and significant achievements have been gained. In the development of such an experimental setup for the Beijing Synchrotron Radiation Facility (BSRF), we suffer from the fact that single crystals are usually crushed into many particles under high pressure, and the diffraction data we have collected are not from a real single crystal but from many particles at random orientations with respect to each other. The well developed methods for the indexing of single-crystal data, such as the Niggli reduction method (Křivý & Gruber, 1976; Andrews & Bernstein, 1988) and the Fourier transform method (Steller *et al.*, 1997; Rossmann, 2006), cannot be applied directly in such cases. Here, we propose a new protocol to handle such multi-particle diffraction data and verify it experimentally.

## 2. Principle and method

### 2.1. Outline of principle

In routine single-crystal diffraction experiments, reflections result from a single crystal and the translational symmetry of the reciprocal lattice is retained. Applying the Niggli reduction method (Křivý & Gruber, 1976; Andrews & Bernstein, 1988) and the Fourier transform method (Steller *et al.*, 1997; Rossmann, 2006) to such a data set will determine the unit cell and its orientation simultaneously. Multi-particle diffraction data consist of several groups of reflections, each of which results from a crystallite. Since the crystallites are at random orientations with respect to each other, reflections of different crystallites are mixed up and cannot be distinguished in a straightforward way. The translational symmetry of the reciprocal lattice is hidden by the random mixing up of reflections of different crystallites. The main idea of our protocol is to propose the unit cell first by applying techniques of powder pattern indexing to multi-particle diffraction data, and then to determine the orientation matrix for each crystallite one by one using nonlinear optimization procedures.

The orientation matrix, represented by UB, relates a reciprocal vector to the instrument coordinate system in the following way:

$$UB \begin{pmatrix} h \\ k \\ l \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad (1)$$

$$\begin{pmatrix} h \\ k \\ l \end{pmatrix} = UB^{-1} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad (2)$$

where  $UB^{-1}$  is the inverse of the orientation matrix. By definition,  $UB$  is a  $3 \times 3$  square matrix:

$$UB = \begin{pmatrix} ub_{11} & ub_{12} & ub_{13} \\ ub_{21} & ub_{22} & ub_{23} \\ ub_{31} & ub_{32} & ub_{33} \end{pmatrix}, \quad (3)$$

where  $ub_{ij}$  represents the  $j$ th component of the  $i$ th reciprocal basis vector in the instrument coordinate system.

In the protocol proposed here, the orientation matrix is rewritten in a different form:

$$UB = UR, \quad (4)$$

where  $U$  is the matrix that relates the reciprocal basis vector to a specially orientated Cartesian coordinate system, while  $R$  describes the orientation relationship between this Cartesian coordinate system and the instrument coordinate system. The specially orientated Cartesian coordinate system is set up in the following way: its  $z'$  axis is parallel to the reciprocal basis vector  $\mathbf{c}^*$ , and the  $x'$  axis is perpendicular to both reciprocal basis vectors  $\mathbf{b}^*$  and  $\mathbf{c}^*$ . Then each component of matrix  $U$  consists of only reciprocal lattice parameters. On the other hand, the components of matrix  $R$  are only related to the angles between the axes of the Cartesian coordinate system and those of the instrument coordinate system.

The matrices  $U$  and  $R$  are given, respectively, as follows:

$$U = \begin{pmatrix} a_{x'}^* & a_{y'}^* & a_{z'}^* \\ b_{x'}^* & b_{y'}^* & b_{z'}^* \\ c_{x'}^* & c_{y'}^* & c_{z'}^* \end{pmatrix}, \quad (5)$$

where

$$a_{x'}^* = a^*(\sin^2 \beta^* - \cos^2 \alpha^* - \cos^2 \gamma^* + 2 \cos \alpha^* \cos \beta^* \cos \gamma^*)^{1/2} / \sin \alpha^*, \quad (6)$$

$$a_{y'}^* = a^*(\cos \gamma^* - \cos \alpha^* \cos \beta^*) / \sin \alpha^*, \quad (7)$$

$$a_{z'}^* = a^* \cos \beta^*, \quad (8)$$

$$b_{x'}^* = 0, \quad (9)$$

$$b_{y'}^* = b^* \sin \alpha^*, \quad (10)$$

$$b_{z'}^* = b^* \cos \alpha^*, \quad (11)$$

$$c_{x'}^* = 0, \quad (12)$$

$$c_{y'}^* = 0, \quad (13)$$

$$c_{z'}^* = c^*, \quad (14)$$

and  $a^*$ ,  $b^*$ ,  $c^*$ ,  $\alpha^*$ ,  $\beta^*$  and  $\gamma^*$  are reciprocal lattice parameters.

$$R = \begin{pmatrix} \cos \theta \cos \varphi & \cos \theta \sin \varphi \sin \omega & \sin \theta \sin \omega \\ & + \sin \theta \cos \omega & - \cos \theta \sin \varphi \cos \omega \\ - \sin \theta \cos \varphi & \cos \theta \cos \omega & \cos \theta \sin \omega \\ & - \sin \theta \sin \varphi \sin \omega & + \sin \theta \sin \varphi \cos \omega \\ \sin \varphi & - \cos \varphi \sin \omega & \cos \varphi \cos \omega \end{pmatrix}, \quad (15)$$

where  $\omega$ ,  $\varphi$  and  $\theta$  are the rotation angles of the specially orientated Cartesian coordinate system with respect to the  $x$ ,  $y$  and  $z$  axes, respectively, of the instrument coordinate system with the assumption that the two coordinate systems coincide with each other initially.

The lattice parameters can be deduced by applying techniques of powder pattern indexing to the multi-particle diffraction data, and then the matrix  $U$  is determined since it depends only on the lattice parameters.

The matrix  $R$  can be determined by nonlinear optimization procedures. Parameters  $\omega$ ,  $\varphi$  and  $\theta$  can be varied in a systematic way, and when a group of values of these parameters correspond to the orientation of one of the crystallites, the reflections of the crystallite will be indexed. This method can be applied repeatedly until the orientation of all crystallites is determined, and then all reflections of the multi-particle diffraction data are indexed.

The details of the methods to determine the lattice parameters and matrix  $R$  are presented in the following sections.

## 2.2. Determination of the unit cell

Many algorithms have been developed to deduce the lattice parameters based on powder diffraction data, in which the only positional information on the reflection available is the Bragg angle. All such algorithms can be directly applied to the multi-particle diffraction data because the positional information on the reflections needed for such algorithms, namely the Bragg angle of each reflection, is available in the multi-particle data.

A lot of programs are available for the indexing of powder patterns: among the most well known are *Dicvol* (Boultif & Louër, 2004), *TREOR* (Werner *et al.*, 1985), *ITO* (Visser, 1969), *X-Cell* (Neumann, 2003) and *DASH* (David *et al.*, 2006). We deduce the information needed by such programs from the multi-particle diffraction data with the high-pressure single-crystal diffraction package *HPSXD*, developed for the high-pressure single-crystal diffraction facility at BSRF. This information is then input into the powder data indexing programs. Usually, the programs will propose several possible unit cells. In routine powder diffraction work, the correct unit cell might be selected on the basis of figures of merit. For the multi-particle diffraction data, the positional information on the reflections is still available, in contrast to the case for 'real' powder data. This positional information can be utilized in the subsequent procedures to verify the unit cells proposed by the programs for the indexing of powder data.

### 2.3. Determination of matrix $R$

The matrix  $R$  can be determined by any nonlinear optimization procedure. Here we applied the genetic algorithm (GA) technique, which has found extensive applications in fields of crystallography such as structure determination (Chang & Lewis, 1997; Harris *et al.*, 1998; Zhou & Su, 2004; Feng & Dong, 2007), indexing of powder diffraction patterns and whole-profile fitting (Kariuki *et al.*, 1999).

$N$  individual solutions are generated randomly to form the initial population, each of which has a group of values of  $\omega$ ,  $\varphi$  and  $\theta$ . The fitness is defined as the number of reflections that can be indexed with a certain individual solution. The fitness of every solution in the initial population is evaluated, and then solutions are selected based on their fitness to breed the new generation of solutions. A solution is selected to be a 'parent' with the probability

$$p = f_i / \sum_i^N f_i, \quad (16)$$

where  $f_i$  is the fitness of the  $i$ th solution and  $N$  is the size of the population.

New 'child' solutions are created through crossover and/or mutation. A pair of 'child' solutions created through crossover can be given as

$$C_1 = P_1 c_p + P_2 (1 - c_p), \quad (17)$$

$$C_2 = P_1 (1 - c_p) + P_2 c_p, \quad (18)$$

where  $C_1$ ,  $C_2$  and  $P_1$ ,  $P_2$  represent 'child' and 'parent' solutions, respectively, and  $c_p$  is a randomly generated number in the range 0–1.

A random number in the range 0–1 is generated for every solution. The solution is forced to experience mutation if the random number is less than a preset constant which defines the probability of mutation. 'Child' solutions created through mutation will be set to random values.

The top  $n$  solutions with the highest fitness survive into the new generation without having to compete with their offspring and without experiencing mutation. The other  $N - n$  solutions in the new generation of population are offspring created through crossover and/or mutation.

The above-mentioned procedure is repeated until the termination conditions have been reached. The termination conditions are

- (1) a fixed number of generations has been reached and/or
- (2) a plateau has been reached such that successive iterations no longer produce better results.

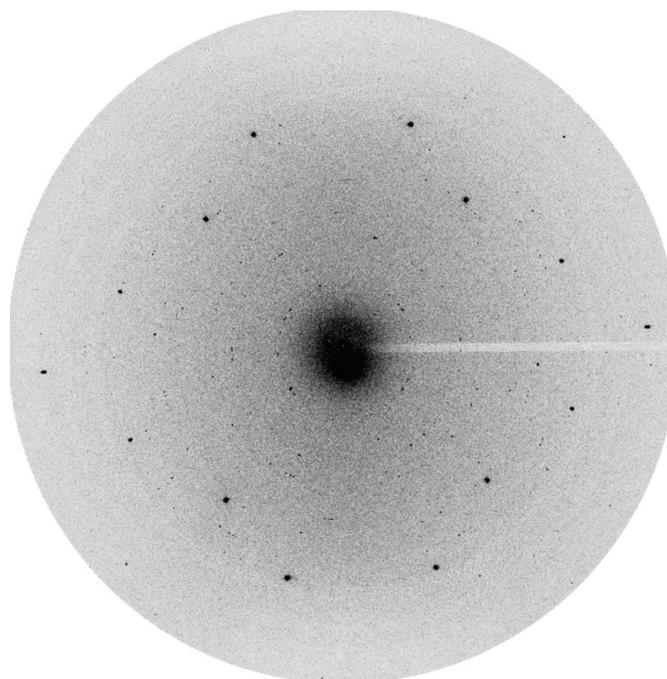
A program (*HP-UB*) has been developed in Fortran90 in order to carry out the GA calculations. It runs under MS-DOS or Windows operating systems. The parameters of the genetic algorithm are set empirically and no attempts have been made to optimize these parameters. The size of the population  $N = 1000$ , the probability of mutation is 0.05, the top ten solutions in each generation survive into the next generation and the maximum number of iterations is 2000. With an Intel Core2 Duo CPU 2.30G  $\times$  2 and 1 GB of memory, the running time of

*HP-UB* is not more than 180 s. The *HP-UB* program can run independently or be called by *HPSXD*. In the *HPSXD* package, an icon linking to external programs for powder indexing is embedded in the command bar. Information needed by these powder indexing programs, such as  $d$  values or Bragg angles, may be deduced conveniently from the multi-particle data using the tool supplied by *HPSXD*. The unit cells proposed by the powder indexing programs may be fed back into the *HPSXD* package and the GA module will then calculate the orientation matrix and index the multi-particle diffraction data.

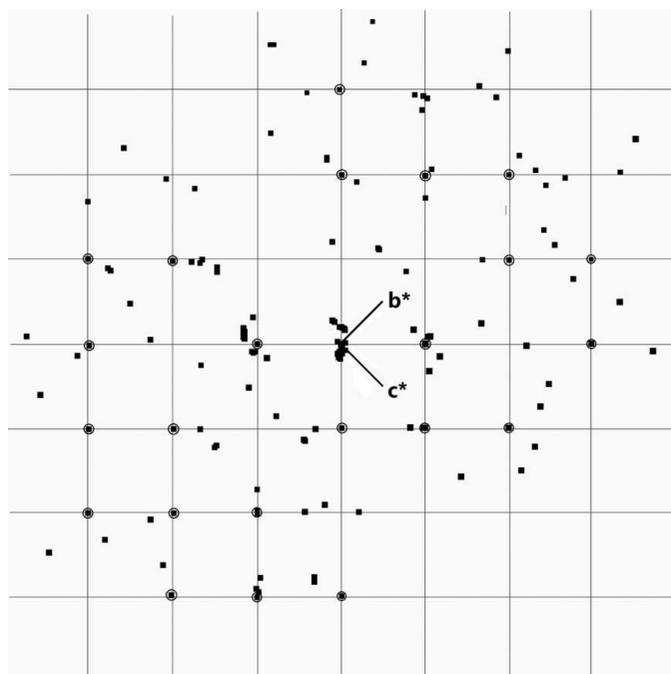
### 3. Application of the method

This method has been applied to several diffraction data sets collected at beamline 4W2 of BSRF. The wavelength of the beam is 0.06199 nm and the cross section is  $40 \times 40 \mu\text{m}$ . A mar345 image plate is used as the detector and the distance to the sample is 182.5 mm.

Two examples of the application of this method are given here. The first data set was collected on an  $\text{LaB}_6$  crystallite of about  $30 \times 40 \times 40 \mu\text{m}$  in size. No pressure was applied in this case and the data set was collected without the diamond anvil cell. The  $\omega$  scan range was from  $-30$  to  $30^\circ$ , which was limited by the apparatus. The scan step was  $2^\circ$  per frame and the exposure time was 10 s for each frame. In total 77 reflections were collected with the highest counts exceeding 1000. Eleven  $d$  values were observed and input into the powder indexing program *Dicvol'04*. Many possible unit cells were proposed, with two of them having much higher figures of merit (de Wolff, 1968) than the others. One is cubic with  $a = 4.1693 \text{ \AA}$ ,



**Figure 1**  
Diffraction data of Ge in a diamond anvil cell collected under a pressure of 1.5 GPa. The data were collected in a continuous  $\omega$  scan in the range from  $-30$  to  $30^\circ$  with the exposure time being 300 s.



**Figure 2**  
Projection of the multi-particle diffraction data along the  $[100]^*$  direction. The indexed reflections are marked with circles. Lines connecting the indexed reflections are used to highlight their periodicity.

while the other is tetragonal with  $a = 2.9509$ ,  $c = 4.1618$  Å. The ratio of  $a/c$  for the tetragonal unit cell is very close to  $\sin 45^\circ$ , and the  $c$  parameter of the tetragonal unit cell is nearly equivalent to the lattice parameter of the cubic one. This means that these two unit cells are closely related to each other. As suggested in §2.2, the positional information on the reflections, which is not available in real powder data, can be utilized to select the correct solution among many possible unit cells proposed by the powder indexing programs. The above-mentioned two unit cells were input into the program *HP-UB*. Of the 77 reflections, 72 were indexed with the cubic unit cell, while only 36 reflections were indexed with the tetragonal one. Clearly, the cubic unit cell is correct and the tetragonal one is a mistake caused by the incompleteness of the positional information on the reflections in the case of powder indexing.

The other data set was collected on a Ge crystallite of size  $10 \times 20 \times 20$  µm. A pressure of 1.5 GPa was applied to the crystallite with a diamond anvil cell. Other experimental parameters were set the same as those in the  $\text{LaB}_6$  case. The result of the continuous  $\omega$  scan is shown in Fig. 1. The strong reflections result from the diamond anvil cell and the weak ones from the Ge sample. The intermittent ring-like nature of the diffraction data suggests that the single crystal of Ge must have been crushed into many randomly orientated crystallites. In total, 180 reflections were observed for Ge crystallites and 12  $d$  values were input into the powder indexing program

*Dicvol'04*. A cubic unit cell with  $a = 5.6135$  Å was suggested as the most probable one on the basis of the figure of merit. With such a unit cell, the program *HP-UB* calculated the orientation matrix and indexed 29 reflections. Visual inspection of the reflections in the reciprocal space revealed that the unit cell and orientation matrix have been determined correctly, as illustrated in Fig. 2. The indexed reflections could be 'removed' from the multi-particle diffraction data set using the *HPSXD* package. Successive application of this method to the remaining reflections would determine the orientation matrix of each crystallite and index all the reflections.

#### 4. Conclusions

A strategy for indexing multi-particle diffraction data has been proposed and the program *HP-UB* based on this strategy has been developed. Possible unit cells are first deduced by applying powder indexing techniques to the multi-particle data, and then the orientation of the unit cell with respect the instrument coordinate system is determined with the genetic algorithm technique. The unit cells proposed by powder indexing procedures may be checked in the process of orientation determination because the positional information on reflections is still available in multi-particle diffraction data, which is contrary to the case for real powder data. The method proposed here has been verified experimentally and will be very helpful for high-pressure single-crystal diffraction, which suffers greatly from the crushing of crystals.

This work was financially supported by the Natural Science Foundation of China under grant No. 11075175.

#### References

- Andrews, L. C. & Bernstein, H. J. (1988). *Acta Cryst.* **A44**, 1009–1018.  
 Boulif, A. & Louër, D. (2004). *J. Appl. Cryst.* **37**, 724–731.  
 Chang, G. & Lewis, M. (1997). *Acta Cryst.* **D53**, 279–289.  
 David, W. I. F., Shankland, K., van de Streek, J., Pidcock, E., Motherwell, W. D. S. & Cole, J. C. (2006). *J. Appl. Cryst.* **39**, 910–915.  
 Feng, Z. J. & Dong, C. (2007). *J. Appl. Cryst.* **40**, 583–588.  
 Harris, K. D. M., Johnston, R. L. & Kariuki, B. M. (1998). *Acta Cryst.* **A54**, 632–645.  
 Kariuki, B. M., Belmonte, S. A., McMahon, M. I., Johnston, R. L., Harris, K. D. M. & Nelmes, R. J. (1999). *J. Synchrotron Rad.* **6**, 87–92.  
 Křivý, I. & Gruber, B. (1976). *Acta Cryst.* **A32**, 297–298.  
 Neumann, M. A. (2003). *J. Appl. Cryst.* **36**, 356–365.  
 Rossmann, M. G. (2006). *International Tables for Crystallography*, Vol. F, 1st online ed., ch. 11.1, pp. 209–211. Chester: International Union of Crystallography.  
 Steller, I., Bolotovskiy, R. & Rossmann, M. G. (1997). *J. Appl. Cryst.* **30**, 1036–1040.  
 Visser, J. W. (1969). *J. Appl. Cryst.* **2**, 89–95.  
 Werner, P.-E., Eriksson, L. & Westdahl, M. (1985). *J. Appl. Cryst.* **18**, 367–370.  
 Wolff, P. M. de (1968). *J. Appl. Cryst.* **1**, 108–113.  
 Zhou, Y. & Su, W.-P. (2004). *Acta Cryst.* **A60**, 306–310.