International Tables for Crystallography (2018). Vol. H, Chapter 3.4, pp. 270–281.

3.4. Indexing a powder diffraction pattern

A. Altomare, C. Cuocci, A. Moliterni and R. Rizzi

3.4.1. Introduction

The crystal structure solution process presupposes that the crystal cell and the space group are known. In other words, the first step in the solution pathway is the identification of the unit-cell parameters. Knowledge of the crystal structure strongly depends on the determination of the cell: a cell incorrectly defined does not lead to the solution. The cell-determination process (which operates in a 6-dimensional continuous parameter space) is also called 'indexing' because it consists of assigning the appropriate triple (*hkl*) of Miller indices of the lattice plane to each of the N_l experimental diffraction lines (in a $3N_l$-dimensional integer-valued index space) (Shirley, 2003). (In this chapter, 'line' and 'peak' are used synonymously.) In the case of powder diffraction, the determination of the cell parameters is not a trivial task, and it is much more difficult than in the single-crystal case. This is because the information about the three-dimensional reciprocal space is compressed into the one-dimensional experimental powder pattern. Whatever the method used, working in the parameter or index space, powder pattern indexing aims to recover the three-dimensional information from the positions of the diffraction peaks in the observed profile. In particular, the experimental information used for carrying out the indexing process is the d_{hkl} interplanar spacings, which are related to the diffraction angles by the well known Bragg law:

$$2d_{hkl} \sin \theta_{hkl} = \lambda.$$

In theory, if we had available an experimental pattern at infinite resolution with well resolved peaks with no overlapping, the determination of the six cell parameters corresponding to a problem with six degrees of freedom would be easy (Shirley, 2003). In practice, only the first 20–30 observed lines are useful for two main reasons: (1) they are less sensitive to small changes in the cell parameters than the higher-angle lines; and (2) higher-angle lines (even if they seem to be single peaks) actually often consist of more than one overlapping peak and their positions cannot be accurately evaluated. Using higher-angle lines is therefore unwise. The successful outcome of powder pattern indexing is correlated to which and how many d_{hkl} values derived from the peaks in the diffraction pattern are selected and how reliable they are. Precision and accuracy in detecting peak positions are essential conditions for successful indexing (Altomare *et al.*, 2008). Unfortunately they can be degraded by different sources of errors: peak overlap, poor peak resolution, 2θ zero shift, errors in measurement, and a low peak-to-background ratio. Moreover, impurity lines (*i.e.* peaks from a different chemical phase in the sample to the compound being studied) can hinder the attainment of the correct result. The history of indexing, having its origin in the early 20th century (Runge, 1917), has produced several methods and software packages (Shirley, 2003; Werner, 2002) with surprising progress in strengthening and automating the cell-determination process. Innovative approaches that aim to reduce the dependence on the d_{hkl} values (by avoiding the peak-search step and considering the full information contained in the diffraction pattern) have also been developed. Despite great advances, powder pattern indexing is still a challenge in many cases. Factors that affect the success or failure of the process include: the presence of diffraction peaks from unexpected phases, the precision in the peak-position value, the size of the unit cell to be identified (indexing is easier if the unit cell is not too big) and the symmetry (indexing a pattern from a compound with high symmetry is generally more reliable than for a compound with lower symmetry). Before the zeroth step of the indexing process (the searching for peaks in the experimental pattern) it is always necessary to obtain good-quality diffraction data. Of course, the use of synchrotron radiation is preferable, but conventional laboratory X-ray data are usually suitable. Whether automated or manual, the peak search and each successive step of the indexing process must be carefully checked. For example, in a first attempt the positions that correspond to overlapping peaks could be set aside. If one attempt fails, the most useful tactic is to try another software package, since the programs available at present are based on different approaches.

The aim of this chapter is first to illustrate the background of the topic and the main theoretical approaches used to carry out the powder pattern indexing, and then to give some examples of applications. Section 3.4.2 is mainly devoted to the basic concepts of a crystalline lattice, the main indexing equations and figures of merits; Section 3.4.3 discusses the traditional and non-traditional methods developed for indexing a powder pattern, and Section 3.4.4 discusses some applications, referring to the most widely used indexing programs.

3.4.2. The basic concepts of indexing

We now describe some concepts that are fundamental in crystallography and useful for understanding the indexing process. The measured diffraction intensities correspond to the reciprocal-lattice points

$$\mathbf{r}^*_{hkl} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*.$$

The Miller indices (*hkl*) identify the plane of the direct lattice and $\mathbf{a}^*$, $\mathbf{b}^*$ and $\mathbf{c}^*$ are the three vectors of the reciprocal lattice, which are related to the direct lattice by

$$\mathbf{a} = \frac{\mathbf{b}^* \times \mathbf{c}^*}{V^*}, \quad \mathbf{b} = \frac{\mathbf{c}^* \times \mathbf{a}^*}{V^*}, \quad \mathbf{c} = \frac{\mathbf{a}^* \times \mathbf{b}^*}{V^*},$$

where

$$V^* = \mathbf{a}^* \cdot \mathbf{b}^* \times \mathbf{c}^*$$

is the reciprocal-cell volume (V^* is the inverse of the direct-unit-cell volume V).

In case of single-crystal data, the three-dimensional nature of the experimental diffraction data makes it easy to identify $\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$, from which the direct-space unit-cell vectors are derived (Giacovazzo, 2011).

In case of powder diffraction, the three-dimensional nature of the diffraction data is compressed into one dimension in the experimental pattern, and the accessible experimental information is the d_{hkl} values involved in the Bragg law and related to the

Table 3.4.1
Expressions for $Q(hkl)$ for different types of symmetry

Symmetry	$Q(hkl)$
Cubic	$(h^2 + k^2 + l^2)A_{11}$
Tetragonal	$(h^2 + k^2)A_{11} + l^2A_{33}$
Hexagonal	$(h^2 + hk + k^2)A_{11} + l^2A_{33}$
Orthorhombic	$h^2A_{11} + k^2A_{22} + l^2A_{33}$
Monoclinic	$h^2A_{11} + k^2A_{22} + l^2A_{33} + hlA_{13}$
Triclinic	$h^2A_{11} + k^2A_{22} + l^2A_{33} + hkA_{12} + hlA_{13} + klA_{23}$

diffraction angles by

$$d_{hkl} = \lambda/(2\sin\theta_{hkl}).$$

d_{hkl}, the spacing of the planes (hkl) in the direct lattice, is obtained by the dot products of the reciprocal-lattice vectors with themselves:

$$(\mathbf{r}^*_{hkl})^2 = \frac{1}{d^2_{hkl}} = h^2a^{*2} + k^2b^{*2} + l^2c^{*2} + 2hka^*b^*\cos\gamma^* + 2hla^*c^*\cos\beta^* + 2klb^*c^*\cos\alpha^*, \quad (3.4.1)$$

where α^* is the angle between $\mathbf{b}^*$ and $\mathbf{c}^*$, β^* is the angle between $\mathbf{c}^*$ and $\mathbf{a}^*$, and γ^* is the angle between $\mathbf{a}^*$ and $\mathbf{b}^*$. If we introduce

$$Q(hkl) = \frac{10^4}{d^2_{hkl}}$$

[where $Q(hkl)$ differs from $\sin^2\theta_{hkl}$ by a scale factor of $(200/\lambda)^2$], (3.4.1) becomes

$$Q(hkl) = h^2A_{11} + k^2A_{22} + l^2A_{33} + hkA_{12} + hlA_{13} + klA_{23}, \quad (3.4.2)$$

where

$$A_{11} = 10^4a^{*2},\ A_{22} = 10^4b^{*2},\ A_{33} = 10^4c^{*2},$$
$$A_{12} = 2\times10^4a^*b^*\cos\gamma^*,\ A_{13} = 2\times10^4a^*c^*\cos\beta^*,$$
$$A_{23} = 2\times10^4b^*c^*\cos\alpha^*.$$

The number of parameters A_{ij} in (3.4.2) depends on the type of symmetry: from 1 in the case of cubic symmetry to 6 for triclinic symmetry (see Table 3.4.1).

The quadratic form (3.4.2) relates the observed $Q(hkl)$ values to the reciprocal cell parameters and, consequently, to the direct cell. It is the basic equation used in powder-indexing procedures. Therefore the indexing problem (Werner, 2002) is to find A_{ij} and, for each observed $Q(hkl)$ value, three crystallographic indices (hkl) satisfying (3.4.2) within a suitable tolerance parameter Δ:

$$Q(hkl) - \Delta < h^2A_{11} + k^2A_{22} + l^2A_{33} + hkA_{12} + hlA_{13} + klA_{23} < Q(hkl) + \Delta. \quad (3.4.3)$$

The importance of using accurate $Q(hkl)$ values in (3.4.3) is obvious. Moreover, it is worth noticing that (3.4.3) must lead to physically reasonable indexing – low-angle peaks should correspond to small integer values for h, k and l and the values of the cell parameters and cell volume should be reasonable.

3.4.2.1. *Figures of merit*

An important task is the introduction of a figure of merit (FOM) that is able to (*a*) describe the physical plausibility of a trial cell and its agreement with the observed pattern, and (*b*) select the best cell among different possible ones. de Wolff (1968) made an important contribution in this direction. He developed the M_{20} figure of merit defined by

$$M_{20} = \frac{Q_{20}}{2\langle\varepsilon\rangle N_{20}}, \quad (3.4.4)$$

where Q_{20} is the Q value corresponding to the 20th observed and indexed peak, N_{20} is the number of different calculated Q values up to Q_{20}, and $\langle\varepsilon\rangle$ is the average absolute discrepancy between the observed and the calculated Q values for the 20 indexed peaks; the factor 2 is a result of statistics, explained by the larger chance for an observed line to sit in a large interval as compared with sitting in a small interval. The rationale behind M_{20} is as follows: the better the agreement between the calculated and the observed peak positions (the smaller the $\langle\varepsilon\rangle$ value) and the smaller the volume of the unit cell (the smaller the N_{20} value), the larger the M_{20} value and consequently the confidence in the proposed unit cell. A rule of thumb for M_{20} is that if the number of unindexed peaks whose Q values are less than Q_{20} is not larger than 2 and if $M_{20} > 10$, then the indexing process is physically reasonable (de Wolff, 1968; Werner, 2002). This rule is often valid, but exceptions occur. The use of the first 20 peaks is a compromise (coming from experience) between introducing a quite large number of observed peaks (depending on the number of parameters of the unit cell) and avoiding the use of high-angle peak positions, which are more affected by errors. M_{20} is statistically expected to be 1 in case of completely arbitrary indexing. It has no upper limit (it can be very large when $\langle\varepsilon\rangle$ is very small).

Smith & Snyder (1979) proposed the F_N criterion in order to overcome the limits of M_{20} with respect to its dependence on the 20 lines and on crystal class and space group. The F_N figure of merit is given by

$$F_N = \frac{1}{\langle|\Delta 2\theta|\rangle}\frac{N}{N_{\rm poss}},$$

where $\langle|\Delta 2\theta|\rangle$ is the average absolute discrepancy between the observed and calculated 2θ peak position values and $N_{\rm poss}$ is the number of possible diffraction lines up to the Nth observed line. The values of $\langle|\Delta 2\theta|\rangle$ and $N_{\rm poss}$, $(\langle|\Delta 2\theta|\rangle, N_{\rm poss})$, are usually given with F_N. With respect to M_{20}, F_N is more suitable for ranking the trial solutions and less for indicating their physical plausibility (Werner, 2002).

Both M_{20} and F_N, being based on the discrepancies between observed and calculated lines, are less reliable if there are impurity peaks; if the information about the unindexed lines is not taken into account, the risk of obtaining false solutions increases. Alternative FOMs based on joint probability have also been proposed (Ishida & Watanabe, 1967, 1971). Among the recently developed FOMs, we mention:

(1) $Q_{\rm partial}$ (Bergmann, 2007):

$$Q_{\rm partial} = \sum_i \min\left[w_i, \left(\frac{x_i - \hat{x}_i}{\delta_i}\right)^2\right],$$

where the summation is over the number of observed lines, w_i is the observed weight of line i, x and $\hat{x}_i$ are the observed and simulated line positions, respectively, and δ_i is the observed random error of line i. $Q_{\rm partial}$ is multiplied by a factor that depends on the symmetry of the simulated lattice (triclinic, ..., cubic), on the unit cell volume and on the number of ignored peaks.

(2) McM_{20} (Le Bail, 2008):

$$McM_{20} = [100/(R_pN_{20})]B_rS_y,$$

where N_{20} is the number of possible lines that might exist up to the 20th observed line (for a primitive P lattice). R_p is

the profile R factor (Young, 1993). B_r is a factor arbitrarily set to 6 for F and R Bravais lattices, 4 for I, 2 for A, B and C, and 1 for P. S_y is a factor equal to 6 for a cubic or a rhombohedral cell, 4 for a trigonal, hexagonal or tetragonal cell, 2 for an orthorhombic cell, and 1 for a monoclinic or triclinic cell.

(3) WRIP20 (Altomare *et al.*, 2009):

$$\mathrm{WRIP20} = \mathrm{RAT}_{Rp}^{2} \times \mathrm{RAT}_{\mathrm{Ind}} \times \mathrm{RAT}_{\mathrm{Pres}} \times w_u \times \mathrm{RAT}_{M20}^{1/2}. \quad (3.4.5)$$

Based on M_{20} (M_{20} and F_N remain the most widely used FOMs), WRIP20 has been developed for exploiting the full information contained in the diffraction profile. The factors that appear in (3.4.5) are

$$\mathrm{RAT}_{Rp} = \frac{1 - R_p}{1 - (R_p)_{\min}}, \quad \mathrm{RAT}_{\mathrm{Ind}} = \frac{\mathrm{PERC}_{\mathrm{Ind}}}{(\mathrm{PERC}_{\mathrm{Ind}})_{\max}},$$

$$\mathrm{RAT}_{\mathrm{Pres}} = \frac{(\mathrm{PERC}_{\mathrm{Pres}})_{\min}}{\mathrm{PERC}_{\mathrm{Pres}}}, \quad w_u = (N_{\mathrm{obs}} - N_u)/N_{\mathrm{obs}},$$

$$\mathrm{RAT}_{M20} = \frac{M_{20}}{(M_{20})_{\max}}, \quad \mathrm{PERC}_{\mathrm{Pres}} = \sum_{\mathrm{Pres}} \mathrm{mult} / \sum_{\mathrm{all}} \mathrm{mult}.$$

R_p is the profile-fitting agreement calculated after the Le Bail (Chapter 3.5) decomposition of the full pattern using the space group with the highest Laue symmetry compatible with the geometry of the current unit cell and no extinction conditions. $\mathrm{PERC}_{\mathrm{Ind}}$, the percentage of independent observations in the experimental profile, is estimated according to Altomare *et al.* (1995). For each extinction symbol compatible with the lattice geometry of the current unit cell, normalized intensities are calculated and subjected to statistical analysis in order to obtain a probability value associated with each extinction symbol in accordance with Altomare *et al.* (2004, 2005). For the extinction symbol with the highest probability value, the value of $\mathrm{PERC}_{\mathrm{Pres}}$ is calculated: $\sum_{\mathrm{all}} \mathrm{mult}$ is the total number of reflections (symmetry-equivalent included) for the space group having the highest Laue symmetry and no extinction conditions. (It varies with the volume of the unit cell and the data resolution.) $\sum_{\mathrm{Pres}} \mathrm{mult}$, which varies according to the extinction rules of the current extinction symbol, coincides with the number of non-systematically absent reflections (with the symmetry equivalents included). The subscripts min and max mark the minimum and the maximum values of each factor respectively, calculated for the possible unit cells that are to be ranked. N_{obs} and N_u are the number of observed and unindexed lines, respectively. All the terms in (3.4.5) are between 0 and 1, so ensuring that WRIP20 also lies between 0 and 1. In addition, WRIP20 has the following properties: (*a*) it is continuous, that is, definable in any interval of the experimental pattern; (*b*) it takes into account the peak intensities, the number of generated peaks and their overlap, and the systematically absent reflections (through the extinction-symbol test); and (*c*) it is not very sensitive to the presence of impurity lines (these usually have low intensities). WRIP20 is effective in finding the correct cell among a number of possible ones and selecting the corresponding most probable extinction symbol (see Example 3 in Section 3.4.4.6.2).

(4) Two new figures of merit based on de Wolff's method, the reversed figure of merit (M_n^{Rev}) and the symmetric figure of merit (M_n^{Sym}), have recently been proposed (Oishi-Tomiyasu, 2013). As observed by Oishi-Tomiyasu, the de Wolff figure of merit M_n does not use the observed and calculated lines in a symmetrical way, consequently it is (*a*) insensitive to computed but unobserved lines (*i.e.*, extinct peaks) and (*b*) sensitive to unindexed observed lines (*e.g.*, impurity peaks). M_n^{Rev} and M_n^{Sym} aim to compensate for the disadvantages of M_n. In particular, M_n^{Rev} has characteristics opposite to those of M_n with regard to sensitivity to extinct reflections and impurity peaks, and M_n^{Sym} has intermediate properties between M_n and M_n^{Rev}. They prove useful in selecting the correct solution, particularly in case of presence of impurity peaks. (See also Section 3.4.4.3.)

3.4.2.2. *Geometrical ambiguities*

Before discussing the concept of geometrical ambiguity in indexing, it is useful to introduce the definition of a reduced cell. While a unit cell defines the lattice, a lattice can be described by an unlimited number of cells. The Niggli reduced cell (Niggli, 1928) is a special cell able to uniquely define a lattice. Methods and algorithms have been derived for identifying the reduced cell starting from an arbitrary one (Buerger, 1957, 1960; Santoro & Mighell, 1970; Mighell, 1976, 2001). The reduced cell has the advantage of introducing a definitive classification, making a rigorous comparison of two lattices possible in order to establish whether they are identical or related (Santoro *et al.*, 1980). An algorithm based on the converse-transformation theory has been developed and implemented in the Fortran program *NIST*LATTICE* for checking relationships between any two cells (Karen & Mighell, 1991).

It is very important to recognize that two lattices are derivative of each other, because many crystallographic problems (twinning, indexing of powder patterns, single-crystal diffractometry) stem from the derivative properties of the lattices. Derivative lattices are classified as super-, sub- or composite according to the transformation matrices that relate them to the lattice from which they are derived (Santoro & Mighell, 1972).

A further obstacle to the correct indexing of a powder pattern is the problem of geometrical ambiguities. It may occur when '*two or more different lattices, characterized by different reduced forms, may give calculated powder patterns with the identical number of distinct lines in identical 2θ positions*' (Mighell & Santoro, 1975). The number of planes (*hkl*) contributing to each reflection may differ, however. Such ambiguity, due to the fact that the powder diffraction pattern only contains information about the length of the reciprocal-lattice vector and not the three-dimensional vector itself, is geometrical. It mainly occurs for high-symmetry cells (from orthorhombic up). The lattices having this property are related to each other by rotational transformation matrices. In Table 3.4.2 some examples of lattices giving geometrical ambiguities and the corresponding transformation matrices are given (Altomare *et al.*, 2008). Where there are geometrical ambiguities, additional prior information (*e.g.*, a single-crystal study) may be useful in order to choose one of the two possible lattices.

A recent procedure developed by Kroll *et al.* (2011) aims to reveal numerical and geometrical relationships between different reciprocal lattices and unit cells. The procedure is based on the assumption that distinct unit cells with lines in the same 2θ positions are derivatives of each other. However, two non-derivative lattices can have identical peak positions. Very recently, Oishi-Tomiyasu (2014*a*, 2016) has developed a new algorithm able to obtain all lattices with computed lines in the same positions as a given lattice. (See also Section 3.4.4.3.)

Table 3.4.2
Examples of lattices leading to geometrical ambiguities

$\boldsymbol{P} = \{P_{ij}\}$ is the transformation matrix from lattice I to lattice II, described by the vectors $\{\mathbf{a}_i\}$ and $\{\mathbf{b}_i\}$, respectively, with $\mathbf{b}_i = \sum_j P_{ij}\mathbf{a}_j$.

Lattice I	Lattice II	$\boldsymbol{P}$
Cubic *P*	Tetragonal *P*	$\begin{pmatrix} 0 & -1/2 & 1/2 \\ 0 & 1/2 & 1/2 \\ -1 & 0 & 0 \end{pmatrix}$
Cubic *I*	Tetragonal *P*	$\begin{pmatrix} 0 & -1/2 & 1/2 \\ 0 & -1/2 & -1/2 \\ 1/2 & 0 & 0 \end{pmatrix}$
	Orthorhombic *F*	$\begin{pmatrix} -1/3 & -1/3 & 0 \\ 0 & 0 & -1 \\ 1 & -1 & 0 \end{pmatrix}$
	Orthorhombic *P*	$\begin{pmatrix} 1/4 & -1/4 & 0 \\ 0 & 0 & 1/2 \\ -1/2 & -1/2 & 0 \end{pmatrix}$
Cubic *F*	Orthorhombic *C*	$\begin{pmatrix} -1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ -1/4 & 0 & -1/4 \end{pmatrix}$
	Orthorhombic *I*	$\begin{pmatrix} -1/6 & 0 & -1/6 \\ 1/2 & 0 & -1/2 \\ 0 & -1 & 0 \end{pmatrix}$
Hexagonal	Orthorhombic *P*	$\begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & -1/2 & 0 \\ 0 & 0 & -1 \end{pmatrix}$
Rhombohedral	Monoclinic *P*	$\begin{pmatrix} -1/2 & 0 & -1/2 \\ 1/2 & 0 & -1/2 \\ 0 & -1 & 0 \end{pmatrix}$

3.4.3. Indexing methods

Indexing methods aim to reconstruct the three-dimensional direct lattice from the one-dimensional distribution of d_{hkl} values. Systematic or accidental peak overlap, inaccuracy of peak positions, zero shift in the $2\theta_{hkl}$ Bragg angles and/or the presence of impurity peaks make the reconstruction difficult. Data accuracy is fundamental for increasing the probability of success; as emphasized by de Wolff: '*The 'indexing problem' is essentially a puzzle: it cannot be stated in rigorous terms (...). It would be quite an easy puzzle if errors of measurements did not exist*' (de Wolff, 1957).

Different approaches have been proposed for solving the indexing puzzle since the pioneering work of Runge (1917). As suggested by Shirley (2003), indexing procedures work in parameter space, or in index space, or in both spaces. As a general consideration, the parameter space allows the inclusion of the cell information and constraints, while the index space is more suitable in cases where there are accidental or systematic absences (Shirley, 1980). In this section an outline of the strategies and search methods adopted by the main traditional and non-traditional indexing approaches is given. For more details see the papers by Shirley (2003) and Bergmann *et al.* (2004).

Among the main indexing procedures, zone indexing (Section 3.4.3.1.1), SIW heuristic (Section 3.4.3.1.2), successive dichotomy (Section 3.4.3.1.5), the topographs method (Section 3.4.3.2.1) and global-optimization methods (Section 3.4.3.2.2) operate in the parameter space; index heuristics (Section 3.4.3.1.3) and index permutation (Section 3.4.3.1.4) work in the index space; and scan/covariance (Bergmann, 2007) operates both in index and parameter space. Each method can be classified as exhaustive or not. An exhaustive method systematically and rigorously searches in the solution space; a non-exhaustive method exploits coincidences and relations between the observed lines with the aim of finding the solution quickly. The classification is not rigorous: approaches that try to combine rigour and speed can be defined as semi-exhaustive (Table 3.4.3).

Indexing procedures can also be classified as traditional and non-traditional. Each indexing method generates a list of possible cells. Their reliability is assessed by FOMs with the aim of selecting the correct one (see Section 3.4.2.1).

3.4.3.1. *Traditional indexing methods*

The traditional indexing approaches adopted over the last century are based on the following strategies and search methods: (1) zone indexing, (2) SIW heuristic, (3) index heuristics, (4) index permutation and (5) successive dichotomy. All of them exploit information about a limited number of observed peak positions.

3.4.3.1.1. *Zone-indexing strategy*

The zone-indexing strategy was originally developed by Runge (1917), successively proposed by Ito (1949, 1950), generalized by de Wolff (1957, 1958) and enhanced by Visser (1969). This approach is based on the search for zones, *i.e.*, crystallographic planes, in the reciprocal lattice, defined by the origin O and two lattice points. If $\mathbf{r}^*_{hkl}$ and $\mathbf{r}^*_{h'k'l'}$ are two vectors in reciprocal space, *i.e.* the positional vectors of the lattice points A and A′, they describe a zone containing any lattice point B whose positional vector is of type $m\mathbf{r}^*_{hkl} \pm n\mathbf{r}^*_{h'k'l'}$, where m and n are positive integers. If ω is the angle between $\mathbf{r}^*_{hkl}$ and $\mathbf{r}^*_{h'k'l'}$, the squared distance of B from O (*i.e.*, $Q_{m,n}$) can be expressed by (de Wolff, 1958; Visser, 1969)

$$Q_{m,n} = m^2 Q_{\rm A} + n^2 Q_{\rm A'} \pm mnR, \qquad (3.4.6)$$

where $Q_{\rm A} = Q(hkl)$ and $Q_{\rm A'} = Q(h'k'l')$ are the squared distances of A and A′ from O, respectively, and $R = 2(Q_{\rm A}Q_{\rm A'})^{1/2}\cos\omega$. R can be derived as

$$R = |Q_{m,n} - m^2 Q_{\rm A} - n^2 Q_{\rm A'}|/mn. \qquad (3.4.7)$$

The method is applied as follows: $Q_{\rm A}$ and $Q_{\rm A'}$ are chosen among the first experimental Q_i values; the $\{Q_i\}$, up to a reasonable resolution, are introduced in (3.4.7) in place of $Q_{m,n}$; and a few positive integer values are assigned to m and n. Equation (3.4.7) provides a large number of R values; equal R values (within error limits) define a zone, for which the ω angle can easily be calculated. The search for zones is performed using different $(Q_{\rm A}, Q_{\rm A'})$ pairs. The R values that are obtained many times identify the most important crystallographic zones. The zones are sorted according to a quality figure, enabling selection of the best ones. In order to find the lattice, all possible combinations of the best zones are tried. For every pair of zones the intersection line is found, then the angle between them is determined and the lattice is obtained.

The method has the advantage of being very efficient for indexing low-symmetry patterns. The main disadvantage is its sensitivity to errors in the peak positions, particularly in the low 2θ region.

3.4.3.1.2. *Shirley–Ishida–Watanabe (SIW) heuristic strategy*

This needs only one single well-established zone, then it arbitrarily chooses the 001 line from the first-level lines. The indexing problem is thus lowered to two dimensions and an exhaustive search is carried out.

3.4.3.1.3. *Index-heuristics strategy*

The index-heuristics strategy searches for the correct cell *via* a trial-and-error approach, assigning tentative Miller indices to a few experimental peak positions (basis lines), usually belonging to the low 2θ region of the experimental pattern. It was first proposed by Werner (1964), then successively refined (Werner *et al.*, 1985) and made more robust and effective (Altomare *et al.*, 2000, 2008, 2009). This approach, which works in the index space, was defined by Shirley as semi-exhaustive (Shirley, 1980). The search starts from the highest-symmetry crystal system (cubic) and, if no plausible solution is found, it is extended to lower symmetry down to triclinic. The number of selected basis lines increases as the crystal symmetry lowers. A dominant zone occurs when one cell axis is significantly shorter than the other two; in this case most of the first observed lines (in terms of increasing $2\theta_{hkl}$ values) can be indexed with a common zero Miller index. Special short-axis tests, aimed at finding two-dimensional lattices, have been proposed for monoclinic symmetry in order to detect the presence of dominant zones (Werner *et al.*, 1985). The index-heuristics method is based on the main indexing equation [see equation (3.4.2)] that can be rewritten (Werner *et al.*, 1985) as

$$Q(hkl) = h^2x_1 + k^2x_2 + l^2x_3 + hkx_4 + hlx_5 + klx_6,$$

where $\{x_i\} = X$ is the vector of unknown parameters, which are derived by solving a system of linear equations

$$\mathbf{MX} = \mathbf{Q}, \tag{3.4.8}$$

where $\mathbf{M}$ is a matrix of Miller indices and $\mathbf{Q}$ is the vector of the selected $Q(hkl)$ values corresponding to the basis lines. The dimensions of $\mathbf{M}$, $\mathbf{X}$ and $\mathbf{Q}$ change according to the assumed symmetry. From the inverse matrix $\mathbf{M}^{-1}$ the corresponding $\mathbf{X}$ is obtained *via* $\mathbf{X} = \mathbf{M}^{-1}\mathbf{Q}$. In the case of monoclinic and higher symmetry, $\{x_i\}$ are calculated by Cramer's rule. Different $\mathbf{X}$ vectors are derived by using a different selection of basis lines. The possible solutions are checked by using the full list of peak positions (up to the first 25 experimental lines). The method is sensitive to errors on peak positions and to the presence of impurities (the presence of only one impurity peak is not critical). The correctness of the $\{x_i\}$ strongly depends on the accuracy of the observed Q values, especially for low-2θ region lines, which are the most dominant ones for this indexing procedure. The possibility of testing different combinations of basis-line sets enables the correct cell to be found by bypassing the cases for which errors in the basis lines occur.

The method has been recently enhanced (Altomare *et al.*, 2000, 2009) by introducing new procedures that are able to increase the probability of successful indexing (see Section 3.4.4.2.1); among them are: (1) a correction for systematic errors in the experimental 2θ values (positive and negative trial 2θ zero shifts are taken into account); this correction should, in principle, describe a real diffractometer error; in practice, it also approximates the specimen displacement error well (perhaps coupled with transparency for organic samples); (2) a more intensive search in solution space for orthorhombic and monoclinic systems; (3) an improvement of the triclinic search; (4) a new figure of merit, WRIP20, which is more powerful than M_{20} in identifying the correct solution among a set of possible ones (see Section 3.4.2.1); (5) a check for geometrical ambiguities; (6) an automatic refinement of the possible cells; and (7) a statistical study of the parity of the Miller indices, performed at the end of the cell refinement, aimed at detecting doubled axes or additional lattice points (for *A*-, *B*-, *C*-, *I*-, *R*- or *F*-centred cells) (such information is used in the successive steps).

3.4.3.1.4. *Index-permutation strategy*

This strategy was proposed by Taupin (1973), and is based on a systematic permutation of indices associated to observed lines for obtaining candidate cells. Because this trial-and-error strategy is similar to the index-heuristics approach, we do not describe it here.

3.4.3.1.5. *Successive-dichotomy search method*

The successive-dichotomy method, first developed by Louër & Louër (1972), is based on an exhaustive strategy working in direct space (except for triclinic systems, where it operates in reciprocal space) by varying the lengths of the cell axes and the interaxial angles within finite intervals. The search for the correct cell is performed in an n-dimensional domain D (where n is the number of cell parameters to be determined). If no solution belongs to D, the domain is discarded and the ranges for the allowed values of cell parameters are increased; on the contrary, if D contains a possible solution, it is explored further by dividing the domain into 2^n subdomains *via* a successive-dichotomy procedure. Each subdomain is analyzed and discarded if it does not contain a solution. The method was originally applied to orthorhombic and higher-symmetry systems (Louër & Louër, 1972), but it has been successively extended to monoclinic (Louër & Vargas, 1982) and to triclinic systems (Boultif & Louër, 1991). The search can be performed starting from cubic then moving down to lower symmetries (except for triclinic) by partitioning the space into shells of volume $\Delta V = 400$ A^3. For triclinic symmetry ΔV is related to the volume $V_{\rm est}$ suggested by the method proposed by Smith (1977), which is able to estimate the unit-cell volume from only one line in the pattern:

$$V_{\rm est} \simeq \frac{0.60 d_N^3}{\frac{1}{N} - 0.0052},$$

where d_N is the value for the Nth observed line; in the case $N = 20$ the triclinic cell volume is $V_{\rm est} \simeq 13.39 d_{20}^3$.

Let us consider, as an example, the monoclinic case; in terms of direct cell parameters, $Q(hkl)$ is given by (Boultif & Louër, 1991)

$$Q(hkl) = f(A, C, \beta) + g(B),$$

where $f(A, C, \beta) = h^2/A^2 + l^2/C^2 - 2hl\cos\beta/(AC)$, $A = a\sin\beta$, $C = c\sin\beta$, $g(B) = k^2/B^2$ and $B = b$. The search using the successive-dichotomy method is performed in a four-dimensional space that is covered by increasing the integer values i, l, m and n in the intervals $[A_-, A_+] = [A_- = A_0 + ip, A_+ = A_- + p]$, $[B_-, B_+] = [B_- = B_0 + lp, B_+ = B_- + p]$, $[C_-, C_+] = [C_- = C_0 + mp, C_+ = C_- + p]$ and $[\beta_-, \beta_+] = [\beta_- = 90 + n\theta, \beta_+ = \beta_- + \theta]$, where the step values of p and θ are 0.4 Å and 5°, respectively, and A_0, B_0 and C_0 are the lowest values of A, B and C (based on the positions of the lowest-angle peaks), respectively. Each quartet of intervals

defines a domain D and, by taking into account the current limits for the parameters A, B, C and β, a calculated pattern is generated, not in terms of discrete $Q(hkl)$ values but of allowed intervals $\left[Q_-(hkl), Q_+(hkl)\right]$. D is retained only if the observed Q_i values belong to the range $[Q_-(hkl) - \Delta Q_i, Q_+(hkl) + \Delta Q_i]$, where ΔQ_i is the absolute error of the observed lines (*i.e.*, impurity lines are not tolerated). If D has been accepted, it is divided into 2^4 subdomains by halving the original intervals $[A_-, A_+]$, $[B_-, B_+]$, $[C_-, C_+]$ and $[\beta_-, \beta_+]$ and new limits $\left[Q_-(hkl), Q_+(hkl)\right]$ are calculated; if a possible solution is found, the dichotomy method is applied iteratively. In case of triclinic symmetry the expression for $Q(hkl)$ in terms of direct cell parameters is too complicated to be treated *via* the successive-dichotomy method; therefore the basic indexing equation (3.4.2) is used. In this case, the $[Q_-(hkl), Q_+(hkl)]$ intervals are set in reciprocal space according to the A_{ij} parameters of (3.4.2). To reduce computing time the following restrictions are put on the (hkl) Miller indices associated with the observed lines: (1) maximum h, k, l values equal to 2 in case of the first five lines; (2) $h + k + l < 3$ for the first two lines.

The outcome of the successive-dichotomy method is not strongly influenced by the presence of a dominant zone. New approaches have been devoted to overcome the limitations of the method with a strict dependence on data accuracy and on impurities (Boultif & Louër, 2004; Louër & Boultif, 2006, 2007), see Section 3.4.4.2).

3.4.3.2. *Non-traditional indexing methods*

New indexing procedures that provide alternatives to the traditional approaches outlined in Section 3.4.3.1 have recently been proposed.

3.4.3.2.1. *The topographs method*

This method (Oishi *et al.*, 2009) is based on the Ito equation (de Wolff, 1957):

$$Q(\mathbf{h}_1 + \mathbf{h}_2) + Q(\mathbf{h}_1 - \mathbf{h}_2) = 2[Q(\mathbf{h}_1) + Q(\mathbf{h}_2)], \qquad (3.4.9)$$

where $Q(\mathbf{h})$ is the length of the reciprocal vector $\mathbf{r}^*_{hkl}$ corresponding to the Miller index vector $\mathbf{h} = (hkl)$. It uses Conway's topograph (Conway & Fung, 1997), a connected tree obtained by associating a graph to each equation of type (3.4.9) and consisting of infinite directed edges. According to Ito's method, if quadrupoles (Q_1, Q_2, Q_3, Q_4) detected among the observed Q_i values satisfy the condition $2(Q_1 + Q_2) = Q_3 + Q_4$, two Miller-index vectors $\mathbf{h}_1$ and $\mathbf{h}_2$ are expected to exist such that $Q_1 = Q(\mathbf{h}_1)$, $Q_2 = Q(\mathbf{h}_2)$, $Q_3 = Q(\mathbf{h}_1 - \mathbf{h}_2)$ and $Q_4 = Q(\mathbf{h}_1 + \mathbf{h}_2)$. If an additional value Q_5 satisfying the condition $2(Q_1 + Q_4) = Q_2 + Q_5$ is found, the graph of the quadrupole (Q_1, Q_2, Q_3, Q_4) grows *via* the addition of the Q_5 contribution; this procedure is iterated. If topographs share a Q value that corresponds to the same reciprocal-lattice vector, then a three-dimensional lattice is derived containing the two-dimensional lattices associated with the original topographs. Three-dimensional lattices are also obtained by combining topographs. The probability that topographs correspond to the correct cell increases with the number of edges of the graph structure. The method is claimed by the authors to be insensitive to the presence of impurity peaks.

3.4.3.2.2. *Global-optimization methods*

Global-optimization methods, widely adopted for solving crystal structures from powder data, have also been successfully applied to indexing. Among them, we provide brief descriptions of genetic algorithms, and Monte Carlo and grid-search methods.

3.4.3.2.2.1. Genetic-algorithm search method

The use of genetic algorithms (GAs) for solving the indexing problem was proposed by Tam & Compton (1995) and Paszkowicz (1996). Since then, Kariuki and co-workers (Kariuki *et al.*, 1999) have combined GAs with a whole-profile-fitting procedure for indexing powder diffraction patterns. This approach exploits the information of the full powder diffraction pattern. It is inspired by the Darwinian evolutionary principle based on mating, mutation and natural selection of the member of a population that survives and evolves to improve future generations. The initial population consists of a set of trial cell parameters, chosen randomly within a given volume range; a full pattern-decomposition process is performed using the Le Bail algorithm (Chapter 3.5) and the agreement between the calculated and observed profiles is derived and used for assessing the goodness of an individual member (*i.e.*, a set of unit-cell parameters). The most plausible cell is therefore found by exploring a six-dimensional hypersurface $R'_{wp}(a, b, c, \alpha, \beta, \gamma)$ and searching for the global minimum of R'_{wp} (see Section 3.4.4.3.2). In contrast to the main traditional methods, whose outcomes depend on the reliability of a set of peak positions, this procedure has the advantage of being insensitive to the presence of small impurity peaks that have a negligible influence on the agreement factor between the experimental and calculated profiles: the global minimum of R'_{wp} is reached if the majority phase is correctly indexed. The main disadvantage of the method is the computing time required, in particular in the case of low symmetry.

3.4.3.2.2.2. Monte Carlo search method

The Monte Carlo approach has also been applied to indexing powder diffraction patterns (Le Bail, 2004; Bergmann *et al.*, 2004; Le Bail, 2008). It exploits all the information contained in the full pattern, randomly generates and selects trial cell parameters, and calculates peak positions to which it assigns the corresponding Miller indices. An idealized powder pattern consisting of peak positions d and extracted intensities I is considered to test the trial cell. The cell reliability is assessed by suitable figures of merit (*e.g.* R_p and McM_{20}, see Section 3.4.2.1). The main drawback of this approach is the significant computing time required, in particular for triclinic systems.

3.4.3.2.2.3. Grid-search method

This performs an iterated 'step-and-repeat search' in the parameter space. It has the advantage of being flexible, exhaustive and not particularly sensitive to impurities or errors, and the disadvantage of being slow (Shirley, 2003).

3.4.4. Software packages for indexing and examples of their use

The different strategies and methods described in Section 3.4.3 have been implemented in a variety of automatic indexing programs (Bergman *et al.*, 2004). Almost all use one of the two different approaches working in parameter space (*i.e.*, unit-cell parameters) or index space (*i.e.*, reflection indices). Only the *EFLECH/INDEX* program (Bergman, 2007), applying the scan/covariance strategy, works in both spaces: in parameter space from cubic down to monoclinic, switching to index space for triclinic. The different indexing methods are classified according to Shirley (2003) in Table 3.4.3. Alternative classifications can be

Table 3.4.3
Classification of indexing methods

Method (strategy/search)	Space	Exhaustive
Zone indexing	Parameter	No
SIW heuristic	Parameter	Semi
Scan/covariance	Parameter (for cubic to monoclinic); index (for triclinic)	To monoclinic
Index heuristics	Index	Semi
Index permutation	Index	Yes
Successive dichotomy	Parameter	Yes
Grid search	Parameter	Yes
Genetic algorithms	Parameter	No
Monte Carlo	Parameter	No

made by considering whether a program works in direct or reciprocal space, or uses Bragg diffraction line positions or the whole experimental diffraction profile. Using the whole experimental diffraction profile requires a lot of computing time, but has become possible as a consequence of recent increases in the speed of computers. In the following, descriptions of the principal (default) steps of the most widely used indexing programs are given. Several non-default options are available for each program. The chances of success of the indexing step increase if more than one program is used.

3.4.4.1. *Traditional indexing programs*

3.4.4.1.1. *ITO (Visser, 1969)*

This program is based on the zone-indexing strategy and uses the Runge–Ito–de Wolff–Visser method of decomposition of the reciprocal space into zones, as described in Section 3.4.3.1.1.

The following steps are executed by the program:

(1) The potential zones are found always using the 20 lowest Bragg angle peaks. The program does not work with fewer peaks.
(2) All possible combinations of the six best zones (including the combination of each zone with itself) are found by searching for trial zones that share a row of common points. For every pair of such zones, the angle between them is found, thus giving a trial reciprocal lattice.
(3) The reduction of the resulting unit cells is carried out using the Delaunay–Ito method (Pecharsky & Zavalij, 2009).
(4) The program tries to index the first 20 lines and repeats this check after least-squares refinement of the unit-cell parameters.
(5) The figures of merit are calculated to assess the quality of each trial unit cell and the four best lattices are provided.

ITO is very efficient at indexing patterns with low symmetry and is only weakly sensitive to impurity peaks, if they occur at high angles. The most frequent causes of failure are inaccuracy or incompleteness of the input data.

3.4.4.1.2. *TREOR90 (Werner et al., 1985)*

Classified by Shirley (1980, 2003) as semi-exhaustive, *TREOR90* is based on the index-heuristics strategy (see Section 3.4.3.1.3) and uses a trial-and-error approach. It performs the following steps:

(1) Some basis lines are selected among the experimental d values, generally from the low-2θ region of the powder diffraction pattern. Five sets of basis lines are generally sufficient for orthorhombic tests, whereas more than seven sets may be necessary for the monoclinic system. At least 20–25 experimental d values are potentially required.
(2) The trial unit cells in the index space are searched by varying the Miller indices that are tentatively assigned to the basis lines.
(3) The analysis starts with cubic symmetry and, in a stepwise manner, tests for lower-symmetry crystal systems are performed. In the case of monoclinic symmetry, a special short-axis test is carried out.
(4) The solution of the linear system in equation (3.4.8) gives the possible cell parameters. Different combinations of the basis lines are tested.
(5) Each possible solution is checked by using the full list of experimental lines.
(6) The quality of the trial cell parameters is mainly assessed by using the M_{20} figure of merit [see equation (3.4.4)]. An effective rule for identifying a reliable solution is $M_{20} > 10$ and no more than one unindexed line.

The success of the program is related to the use of some suitable standard sets of parameter values (maximum unit-cell volume, maximum cell axis, tolerance of values *etc*.) arising from the accumulated experience of the authors; they can be easily changed by the user *via* suitable keywords in the input file.

3.4.4.1.3. *DICVOL91 (Boultif & Louër, 1991)*

This program works in direct space (down to the monoclinic system) by using the successive-dichotomy search method (see Section 3.4.3.1.5), which was introduced for the automatic indexing of powder diffraction patterns by Louër & Louër (1972). *DICVOL91* has been defined as exhaustive by Shirley (1980, 2003). Its main steps are:

(1) The unit-cell volume is partitioned by moving from high to low symmetry. Shells of 400 Å^3 of volume are scanned for all the symmetry systems except for triclinic; for the triclinic system the shells are based on the definition of volume proposed by Smith (1977) (see Section 3.4.3.1.5).
(2) A search using the successive-dichotomy method, based on suitable intervals (see Section 3.4.3.1.5), is carried out. In the case of the triclinic system the general expression for $Q(hkl)$ as function of the direct cell parameters is too complicated, and the reciprocal-space parameters in equation (3.4.2) are used for setting the intervals.
(3) The derived cell parameters are refined using the least-squares method.
(4) The quality of each trial unit cell is evaluated by using the M_N and F_N figures of merit (see Section 3.4.2.1).

The program is fast at performing exhaustive searches in parameter space (except for the triclinic case); on the other hand, its efficiency is strongly related to the quality of the data and to the presence of impurities (in fact, impurities are not permitted).

3.4.4.2. *Evolved indexing programs*

3.4.4.2.1. *N-TREOR09 (Altomare et al., 2009)*

Implemented in the *EXPO* program (Altomare *et al.*, 2013) to perform the powder pattern indexing step, *N-TREOR09* is an update of *N-TREOR* (Altomare *et al.*, 2000), which in turn is an evolution of *TREOR90*, and preserves the main strategies with some changes introduced to make the program more exhaustive and powerful. In particular:

(*a*) If the default indexing process fails, the unit-cell search is automatically repeated by changing some default choices, *e.g.*, increasing the tolerance value on the observed *d* values. If still no solution is obtained, the maximum (*hkl*) Miller indices assigned to the orthorhombic or monoclinic base lines are increased and the tolerance limits of the default values are halved in order to avoid the generation of wrong large unit cells.

(*b*) At the end of the first run, whatever the obtained results, a possible 2θ zero-position shift is taken into account: the indexing process starts again by applying positive and negative 2θ zero-position shifts to the original peak search result.

(*c*) An exhaustive triclinic search is performed. The dominant-zone tests that are usually carried out for the monoclinic system have been extended to include the triclinic case.

(*d*) A new figure of merit, WRIP20, more powerful than the classical M_{20}, is used. It is calculated when more than one possible cell is found and takes into account the M_{20} value, the full experimental pattern, the degree of reflection overlap, the systematically absent reflections and the number of unindexed lines (see Section 3.4.2.1).

This program is also able to index powder patterns from small proteins: see Example 4 in Section 3.4.4.6.2.

3.4.4.2.2. *DICVOL06 (Louër & Boultif, 2006, 2007) and DICVOL14 (Louër & Boultif, 2014)*

The most recent of a series of versions, *DICVOL14* is the successor of *DICVOL04* (Boultif & Louër, 2004) and *DICVOL06*. *DICVOL06* includes *DICVOL04* with its optimized search procedure and an extended search in shells of volumes. *DICVOL04* represented an improvement of *DICVOL91*. Among the features of *DICVOL06* are:

(*a*) A tolerance for unindexed lines that can result from the presence of unwanted additional phases or inaccurately measured peaks. The program can tolerate a user-defined number of unindexed lines. Care must be taken when using this option to avoid the possibility of generating erroneous cells. It is worth noting that the inclusion of the possibility of at least one unindexed peak has markedly increased the success rate of *DICVOL06*.

(*b*) A correction of the zero-point error in the measured data. *Via* an *a priori* zero-origin evaluation, two different approaches can be followed: (i) if there is a non-negligible zero shift (*i.e.*, ~0.1°), the reflection-pair method is adopted (Dong *et al.*, 1999); (ii) if the shift is small (<0.03°), a refinement of the experimental data zero point together with the cell parameters is carried out as soon as a solution is found. In the monoclinic and triclinic systems, a reduced-cell analysis is performed to choose among equivalent solutions.

(*c*) When a solution is found in a 400 Å^3 shell of volume, the exhaustive search is extended to the whole domain.

No formal limits on the number of input Bragg peaks have been established but, for reliable indexing, it is recommended that 20 or more peaks (in the low-2θ region) are used.

Compared to *DICVOL04*/*DICVOL06*, *DICVOL14* includes: an optimization of filters in the final stages of the convergence of the successive dichotomy process; an optimization and extension of scanning limits for the triclinic case; a new approach for zero-point offset evaluation; a detailed review of the input data from the resulting unit cells; and cell centring tests. *DICVOL14* has been improved particularly for triclinic cases, which are generally the most difficult to solve with the dichotomy algorithm.

3.4.4.3. *Non-traditional indexing programs*

The indexing programs described above are based on using, for a limited number of lines, the measured positions of peak maxima as directly obtained from the experimental powder diffraction pattern. *Conograph* (Oishi-Tomiyasu, 2014*b*), which has been more recently proposed, also belongs to that group of programs. A brief description of *Conograph* follows

3.4.4.3.1. *Conograph: indexing via the topographs method*

Conograph is based on the topographs method, and its main functions are the determination of the primitive unit cell and lattice symmetry, and refinement of lattice parameters. Among the main features we note:

(1) A new Bravais-lattice determination algorithm (Oishi-Tomiyasu, 2012), which has been proved to be stable with respect to peak-position errors under very general conditions. The algorithm applies the Minkowski reduction to primitive cells and the Delaunay reduction (Delaunay, 1933) to face-centred, body-centred, rhombohedral and base-centred cells in such a way that the computational efficiency of the process is better than the Andrews & Bernstein (1988) method.

(2) The two figures of merit $M_n^{\rm Rev}$ and $M_n^{\rm Sym}$ proposed by Oishi-Tomiyasu (2013) are used for for selecting the true unit cells. They are also used to estimate the zero-point shift.

(3) The use of many observed peaks in the default setting, which aims to make *Conograph* robust against dominant zones and missing or false peaks (Oishi-Tomiyasu, 2014*b*).

(4) The method for exhaustively searching unit cells that involve geometrical ambiguity (Oishi-Tomiyasu, 2014*a*, 2016). The geometrical ambiguities that are detected also include lattices with very similar calculated lines, because of the error tolerance in the *d* spacings.

Programs that use only the measured positions of peak maxima are particularly vulnerable to experimental errors in the measured peak positions and to the presence of impurity peaks. For these reasons, at the end of the 1990s new indexing strategies were developed that do not require the peak locations in the experimental pattern. These approaches are completely different from the methods described above because they use the whole diffraction profile. They try to explore the parameter space (direct space) exhaustively by applying different optimization techniques in order to find the cells in best agreement with the experimental powder diffraction pattern. Some of the most widely used indexing programs in direct space are described here.

3.4.4.3.2. *GAIN: indexing via a genetic-algorithm search method*

The use of genetic algorithms (GAs) for indexing powder diffraction data by exploiting the diffraction geometry (as in the traditional indexing methods) was firstly proposed by Tam & Compton (1995) and Paszkowicz (1996). Subsequently, Kariuki *et al.* (1999) applied GA techniques by using whole profile fitting with the aim of exploring the parameter space $\{a, b, c, \alpha, \beta, \gamma\}$ and finding the global minimum of the *R*-factor $\{a, b, c, \alpha, \beta, \gamma\}$ hypersurface, yielding the parameter set able to generate the best agreement between the observed and calculated powder diffraction patterns.

This new strategy has been implemented in the program *GAIN* (Harris *et al.*, 2000), whose main features are:

(1) Starting from a population of N_p sets of lattice parameters and using the evolutionary operations of mating, mutation and natural selection, the population is allowed to evolve

through several generations, with the aim of generating sets of possible trial cell parameters.
(2) The search procedure, using a GA, is performed in restricted, sensible cell-volume ranges consistent with the knowledge of the system under study.
(3) For each set of trial parameters a calculated powder diffraction pattern is constructed. The peak positions and parameters describing the shape and width of each peak are used in the Le Bail profile-fitting procedure (Chapter 3.5).
(4) The pattern is split into different regions (defined by the user), and the weighted profile R factor is calculated for each region; all the values are summed to obtain the overall $R'_{\rm wp}$:

$$R'_{\rm wp} = \sum_{\rm regions} \left[\frac{\sum_i w_i (y_i - y_{ci})^2}{\sum_i w_i y_i^2} \right]^{1/2},$$

where the summation is over the regions, i runs over the experimental points belonging to each region and y_i and y_{ci} are the observed and calculated profile at the ith experimental step, respectively. *Via* the $R'_{\rm wp}$ formula the residual for each region is scaled according to the total intensity in the region, so a region with only low-intensity peaks can make an important contribution to $R'_{\rm wp}$.

This approach is robust at handling the problems that may affect the experimental powder pattern: peak overlap, (*hkl*)-dependent effects and zero-point errors. It is time consuming (particularly in the case of low symmetry) but not very sensitive to the presence of minority impurity phases.

3.4.4.3.3. *McMaille: indexing via a Monte Carlo search method*

The information in the whole powder diffraction profile is exploited by the program *McMaille* (Le Bail, 2004), which is based on the random generation of cell parameters and uses the Monte Carlo optimization technique. Once the trial cell parameters have been generated and the Miller indices and the peak positions have been calculated, the quality of the cell is assessed by using, as figure of merit, the conventional Rietveld profile reliability factor R_p (Young, 1993) or McM_{20} (see Section 3.4.2.1). The program uses some tricks that can increase the success of the Monte Carlo algorithm:
(1) Only the trial cells corresponding to a value of R_p that is smaller than a user-defined value (∼50%) are retained for successive refinement.
(2) If all the observed peaks, except for a user-defined number of tolerated impurity peaks, are 'explained' whatever the R_p value, the cell is retained for successive examination.
(3) If either of the conditions (1) or (2) is fulfilled, the cell parameters are randomly changed in 200 to 5000 attempts (for cubic to triclinic cases, respectively) in which small random parameter variations *via* the Monte Carlo algorithm are carried out. The new parameters are preserved if an improvement of R_p is verified in 85% of the attempts.

This procedure is not sensitive to impurity lines, provided that the sum of their intensities is less than 10–15% of the total intensity. A zero-point error up to 0.05° is tolerated. To reduce the long computing time required to successfully complete the procedure, a significant increase in speed has been obtained by using idealized profiles generated by applying simplified line profiles to extracted line positions. A parallelized version of *McMaille* has also been developed. The indexing problem can usually be solved in few minutes if: (*a*) no triclinic symmetry is handled (because this requires more computing time); (*b*) the cell volume is less than 2000 Å^3; (*c*) no cell length is longer than 20 Å.

3.4.4.4. *Crysfire: a suite of indexing programs*

The *Crysfire* suite (Shirley, 2002) is a multi-program indexing facility. It can perform a self-calibration, which is aimed at detecting and correcting 2θ zero errors, and is able to strip out weak lines. Its single unified user interface and data-file format make a wide set of indexing packages accessible with minimal effort (especially to non-specialists). *Crysfire* provides a list of the possible cells suggested by each indexing program, suitably ranked. The *Crysfire* 2003 suite supports a total of 11 programs (Bergmann *et al.*, 2004), among which are *ITO*, *TREOR90*, *DICVOL91* and *McMaille*. The possibility of using different indexing programs, working in parameter space or index space and adopting different indexing approaches increases the probability of finding the correct cell.

3.4.4.5. *Two commercial programs*

3.4.4.5.1. *SVD-Index*

This commercial indexing program (Coelho, 2003*a*), which uses the Monte Carlo method, is part of the *TOPAS* (Coelho, 2003*b*) suite from Bruker AXS. The reciprocal-cell parameters in equation (3.4.2) are found by using, in an iterative way, the singular value decomposition (SVD) approach (Nash, 1990) to solve linear equations relating (*hkl*) values to d spacings. The method is particularly useful in cases for which there are more equations than variables. All the observed lines in the powder pattern are involved in the indexing procedure. It is claimed that the program is relatively insensitive to impurity peaks and missing high d spacings; it performs well on data with large diffractometer zero errors.

More recently, two indexing methods have been introduced in *TOPAS*: LSI (least-squares iteration), an iterative least-squares process which operates on the d-spacing values extracted from reasonable-quality powder diffraction data, and LP-Search (lattice parameter search), a Monte Carlo based whole-powder-pattern decomposition approach independent of the knowledge of the d-spacings (Coelho & Kern, 2005).

3.4.4.5.2. *X-CELL*

This commercial program is part of the *Materials Studio* suite from Accelrys (Neumann, 2003). To perform an exhaustive search, like *DICVOL*, the program uses the successive-dichotomy approach. Its principal features are:
(1) the user can define how many impurity lines can be tolerated;
(2) a search for the zero-point shift of the diffraction pattern; and
(3) systematic absences are taken into account.

The program is described as 'virtually exhaustive'; it is expected to work well when faced with missing lines, impurities and errors.

3.4.4.6. *Examples of applications of indexing programs*

3.4.4.6.1. *Indexing using DICVOL06*

The program *DICVOL06*, as implemented in the *WinPLOTR*/*FULLPROF* suite (Roisnel & Rodríquez-Carvajal, 2001) and recently introduced into *EXPO*, was applied to two experimental diffraction patterns.

Example 1

Norbornene (Brunelli *et al.*, 2001). Published information: C_7H_{10}, monoclinic, a = 7.6063 (9), b = 8.6220 (1), c = 8.749 (1) Å, β = 97.24 (1)°, $P2_1/c$, experimental range 5–60° 2θ, λ = 0.85041 Å, RES = 1.0 Å (where RES is the data resolution), synchrotron data, indexed by *Fzon* (Visser, 1969).

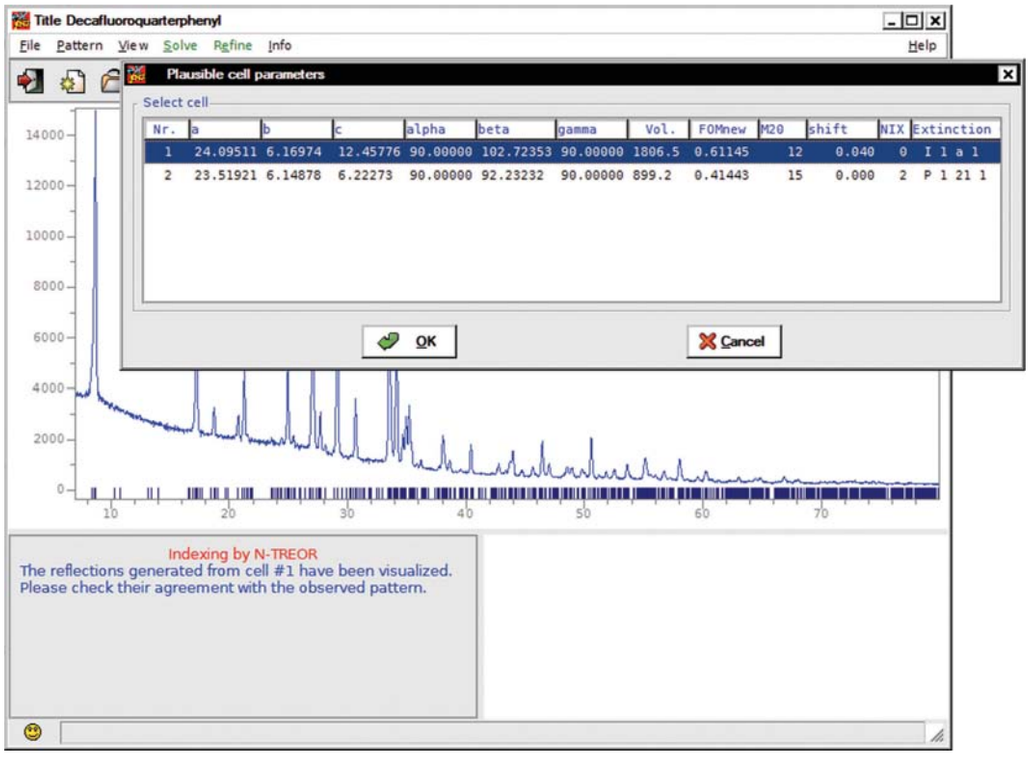

Figure 3.4.1
The list of possible cells for the decafluoroquarterphenyl structure automatically found using *N-TREOR09*.

The 2θ values of the first 25 peaks, in the range 5–25°, were determined by *WinPLOTR* and supplied to *DICVOL06*. The first 20 peaks were used for searching for the solution. No plausible cell was found when assuming that no impurity was present and exploring all the systems (from cubic to triclinic). *DICVOL06* was also unsuccessful when the non-default strategies of extended search and data correction for zero-point error were considered (by setting some flags to 1 in the input file). If it was supposed that two impurity lines might be present among the peaks (by setting the flag corresponding to the maximum number of accepted impurity/spurious lines to 2), *DICVOL06* was able to find the following monoclinic cell: a = 8.7480 (36), b = 8.6313 (32), c = 7.6077 (26) Å, β = 97.201 (33)°, with two unindexed lines, M_{18} = 41.5, F_{18} = 125(0.0041, 35). The refinement of the cell by considering all the 25 lines gave a = 7.6087 (26), b = 8.6295 (30), c = 8.7459 (34) Å, β = 97.201 (34)°, which is very similar to the published one; 23 indexed lines, M_{20} = 30.1, F_{20} = 102.6(0.0048, 41). The presence of the two impurity lines has been ascribed by the authors to a small amount of hexagonal plastic phase.

Example 2

Cu(II)–Schiff base complex (Banerjee *et al.*, 2002). Published information: $Cu(C_{15}H_{12}NO_2)_2$, triclinic, a = 11.928 (4), b = 12.210 (5), c = 9.330 (5) Å, α = 102.54 (4), β = 111.16 (5), γ = 86.16 (4)°, $P\bar{1}$, experimental range 6–100° 2θ, λ = 1.54056 Å, RES = 1.22 Å, high-quality X-ray laboratory data, indexed by *DICVOL91*. The 2θ values of the first 30 peaks, in the range 6–25°, were determined by *WinPLOTR* and supplied to *DICVOL06*. The first 20 peaks were used for searching for the solution. If it was assumed that no impurity was present, no plausible cell was found down to the monoclinic system. When the triclinic system was explored, *DICVOL06* suggested only one plausible solution: a = 12.2157 (73), b = 12.2031 (77), c = 9.3071 (41) Å, α = 65.798 (46), β = 102.572 (59), γ = 95.711 (61)°, with no unindexed lines, M_{20} = 27.0, F_{20} = 77.0(0.010, 26). The refinement of the cell considering all the 30 lines gave a = 12.2125 (65), b = 12.1989 (61), c = 9.3016 (32) Å, α = 65.826 (33), β = 102.569 (40), γ = 97.755 (44)°, no unindexed lines, M_{20} = 27.9, F_{20} = 72.8(0.0106, 26). For this, the corresponding conventional cell is a = 11.93313 (61), b = 12.2125 (65), c = 9.3016 (32) Å, α = 102.569 (40), β = 111.152 (33), γ = 86.151 (44)°, similar to the published one.

3.4.4.6.2. *Indexing using N-TREOR09*

Two examples of powder diffraction pattern indexing by using *N-TREOR09,* as implemented in the *EXPO* program, will be described. To activate the procedure some specific instructions must be given to *EXPO via* the input file or the graphical interface. As a first step, the peak-search procedure is automatically performed on the experimental powder pattern and the list of corresponding d values are supplied to *N-TREOR09*. During the indexing process a correction for zero-point error is automatically carried out (positive and negative shifts are taken into account). Both the examples below were successfully indexed by a default run of *EXPO*.

Example 3

Decafluoroquarterphenyl (Smrčok *et al.*, 2001). Published information: $C_{24}H_8F_{10}$, monoclinic, a = 24.0519 (9), b = 6.1529 (3), c = 12.4207 (5) Å, β = 102.755 (2)°, $I2/a$, experimental range 7–80° 2θ, λ = 1.79 Å, RES = 1.39 Å, medium-quality X-ray laboratory data. The first 43 peaks (in the range 7–67°) with intensities greater than a default threshold were selected (an intensity-based criterion is automatically adopted). The first 25 lines were used to find a possible cell that was then refined by considering all the 43 peaks. At the end of the automatic indexing procedure, *N-TREOR09* suggested two possible cells ranked according to WRIP20 [equation (3.4.5)], as shown in Fig. 3.4.1 (WRIP20 is denoted as

FOMnew in *N-TREOR09*). The first one in the list is the correct cell. It is worth mentioning that the classical M_{20} figure of merit was not able to pick up the solution. The best cell parameters, found according to FOMnew, were a = 24.0951 (50), b = 6.1697 (21), c = 12.4578 (37) Å, β = 102.724 (18)°, similar to those reported in the literature, with FOMnew = 0.61, M_{20} = 12; all the lines in the pattern were indexed. The program provided the solution thanks to its automatic check for a zero-point correction (2θ zero shift = 0.04°) and was able to correctly identify the extinction group (*I_a_*). For the second suggested cell (the wrong solution) FOMnew = 0.41, M_{20} = 15, and two lines were unindexed.

Example 4

Hexagonal turkey egg-white lysozyme (Margiolaki *et al.*, 2005). Published information: hexagonal, a = 71.0862 (3), c = 85.0276 (5) Å, $P6_122$, experimental range 0.4–12° 2θ, RES = 3.35 Å, synchrotron data. The first 94 peaks (in the range 0.4–6°, λ = 0.700667 Å) with intensities greater than a default threshold were selected. An intensity-based criterion was automatically adopted. The first 25 lines were used to find possible cells that were then refined by considering all 94 peaks. Five possible unit cells were automatically suggested by the program in the following systems: hexagonal (1), orthorhombic (1) and monoclinic (3). The highest value for WRIP20 was 0.99, and was for the correct hexagonal cell parameters: a = 71.0922 (4), c = 85.0269 (7) Å, which are similar to those reported in the literature; all the 94 selected lines in the pattern were indexed. For this cell, the program detected a geometrical ambiguity (see Section 3.4.2.2) between hexagonal and orthorhombic lattices and automatically selected the higher-symmetry one.

3.4.5. Conclusion

Indexing a powder diffraction pattern is sometimes described as a 'gateway technology', because the determination of the cell parameters is so fundamental: if no cell has been identified the execution of the subsequent steps of the structure solution process is impossible, and if a wrong cell has been used the correct solution is unreachable. Therefore extremely close attention must be paid to the indexing step of the process. From the early 1970s, the increasing interest in powder pattern indexing and the progress seen, in terms of both methods and algorithms, have strongly contributed to opening the door to modern applications of powder diffraction techniques. The availability of a quite large number of software packages, based on different indexing strategies, enables the scientist interested in solving crystal structures to switch from one program to another when the first fails, so increasing the possibility of success. In some cases indexing is still a challenging process. Good-quality data are necessary and careful inspection of each indexing step, in particular in the selection of the experimental peak positions to be used, is advisable.

References

Altomare, A., Caliandro, R., Camalli, M., Cuocci, C., da Silva, I., Giacovazzo, C., Moliterni, A. G. G. & Spagna, R. (2004). *Space-group determination from powder diffraction data: a probabilistic approach*. *J. Appl. Cryst.* **37**, 957–966.

Altomare, A., Camalli, M., Cuocci, C., da Silva, I., Giacovazzo, C., Moliterni, A. G. G. & Rizzi, R. (2005). *Space group determination: improvements in EXPO2004*. *J. Appl. Cryst.* **38**, 760–767.

Altomare, A., Campi, G., Cuocci, C., Eriksson, L., Giacovazzo, C., Moliterni, A., Rizzi, R. & Werner, P.-E. (2009). *Advances in powder diffraction pattern indexing: N-TREOR09*. *J. Appl. Cryst.* **42**, 768–775.

Altomare, A., Cascarano, G., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Burla, M. C. & Polidori, G. (1995). *On the number of statistically independent observations in a powder diffraction pattern*. *J. Appl. Cryst.* **28**, 738–744.

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R., Corriero, N. & Falcicchio, A. (2013). *EXPO2013: a kit of tools for phasing crystal structures from powder data*. *J. Appl. Cryst.* **46**, 1231–1235.

Altomare, A., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Rizzi, R. & Werner, P.-E. (2000). *New techniques for indexing: N-TREOR in EXPO*. *J. Appl. Cryst.* **33**, 1180–1186.

Altomare, A., Giacovazzo, C. & Moliterni, A. (2008). *Indexing and space group determination*. In *Powder Diffraction Theory and Practice*, edited by R. E. Dinnebier & S. J. L. Billinge, pp. 206–226. Cambridge: RSC Publishing.

Andrews, L. C. & Bernstein, H. J. (1988). *Lattices and reduced cells as points in 6-space and selection of Bravais lattice type by projections*. *Acta Cryst.* A**44**, 1009–1018.

Banerjee, S., Mukherjee, A., Neumann, M. A. & Louër, D. (2002). *Ab-initio structure determination of a Cu(II)-Schiff base complex from X-ray powder diffraction data*. *Acta Cryst.* A**58**, c264.

Bergmann, J. (2007). *EFLECH/INDEX – another try of whole pattern indexing*. *Z. Kristallogr. Suppl.* **26**, 197–202.

Bergmann, J., Le Bail, A., Shirley, R. & Zlokazov, V. (2004). *Renewed interest in powder diffraction data indexing*. *Z. Kristallogr.* **219**, 783–790.

Boultif, A. & Louër, D. (1991). *Indexing of powder diffraction patterns for low-symmetry lattices by the successive dichotomy method*. *J. Appl. Cryst.* **24**, 987–993.

Boultif, A. & Louër, D. (2004). *Powder pattern indexing with the dichotomy method*. *J. Appl. Cryst.* **37**, 724–731.

Brunelli, M., Fitch, A. N., Jouanneaux, A. & Mora, A. J. (2001). *Crystal and molecular structures of norbornene*. *Z. Kristallogr.* **216**, 51–55.

Buerger, M. J. (1957). *Reduced cells*. *Z. Kristallogr.* **109**, 42–60.

Buerger, M. J. (1960). *Note on reduced cells*. *Z. Kristallogr.* **113**, 52–56.

Coelho, A. A. (2003*a*). *Indexing of powder diffraction patterns by iterative use of singular value decomposition*. *J. Appl. Cryst.* **36**, 86–95.

Coelho, A. A. (2003*b*). *TOPAS*. *Version 3.1 User's Manual*. Bruker AXS GmbH, Karlsruhe, Germany.

Coelho, A. A. & Kern, A. (2005). *Discussion of the indexing algorithms within TOPAS*. *IUCr Commission on Powder Diffraction Newsletter*, **32**, 43–45.

Conway, J. H. & Fung, F. Y. C. (1997). *The Sensual (Quadratic) Form*. Washington, DC: The Mathematical Association of America.

Delaunay, B. (1933). *Neue Darstellung der geometrischen Kristallographie*. *Z. Kristallogr.* **84**, 109–149.

Dong, C., Wu, F. & Chen, H. (1999). *Correction of zero shift in powder diffraction patterns using the reflection-pair method*. *J. Appl. Cryst.* **32**, 850–853.

Giacovazzo, C. (2011). *Crystallographic computing.* In *Fundamentals of Crystallography*, 3rd ed., edited by C. Giacovazzo, pp. 66–156. Oxford: IUCr/Oxford University Press.

Harris, K. D. M., Johnston, R. L., Chao, M. H., Kariuki, B. M., Tedesco, E. & Turner, G. W. (2000). *Genetic algorithm for indexing powder diffraction data*. University of Birmingham, UK.

Ishida, T. & Watanabe, Y. (1967). *Probability computer method of determining the lattice parameters from powder diffraction data*. *J. Phys. Soc. Jpn*, **23**, 556–565.

Ishida, T. & Watanabe, Y. (1971). *Analysis of powder diffraction patterns of monoclinic and triclinic crystals*. *J. Appl. Cryst.* **4**, 311–316.

Ito, T. (1949). *A general powder X-ray photography*. *Nature*, **164**, 755–756.

Ito, T. (1950). *X-ray Studies on Polymorphism*. Tokyo: Maruzen Company.

Karen, V. L. & Mighell, A. D. (1991). *Converse-transformation analysis*. *J. Appl. Cryst.* **24**, 1076–1078.

Kariuki, B. M., Belmonte, S. A., McMahon, M. I., Johnston, R. L., Harris, K. D. M. & Nelmes, R. J. (1999). *A new approach for indexing powder*

diffraction data based on whole-profile fitting and global optimization using a genetic algorithm. *J. Synchrotron Rad.* **6**, 87–92.
Kroll, H., Stöckelmann, D. & Heinemann, R. (2011). *Analysis of multiple solutions in powder pattern indexing: the common reciprocal metric tensor approach*. *J. Appl. Cryst.* **44**, 812–819.
Le Bail, A. (2004). *Monte Carlo indexing with McMaille*. *Powder Diffr.* **19**, 249–254.
Le Bail, A. (2008). *Structure solution*. In *Principles and Applications of Powder Diffraction*, edited by A. Clearfield, J. H. Reibenspies & N. Bhuvanesh, pp. 261–309. Oxford: Wiley-Blackwell.
Louër, D. & Boultif, A. (2006). *Indexing with the successive dichotomy method, DICVOL04*. *Z. Kristallogr. Suppl.* **23**, 225–230.
Louër, D. & Boultif, A. (2007). *Powder pattern indexing and the dichotomy algorithm*. *Z. Kristallogr. Suppl.* **26**, 191–196.
Louër, D. & Boultif, A. (2014). *Some further considerations in powder diffraction pattern indexing with the dichotomy method. Powder Diffr.* **29**, S2, S7–S12.
Louër, D. & Louër, M. (1972). *Méthode d'essais et erreurs pour l'indexation automatique des diagrammes de poudre*. *J. Appl. Cryst.* **5**, 271–275.
Louër, D. & Vargas, R. (1982). *Indexation automatique des diagrammes de poudre par dichotomies successives*. *J. Appl. Cryst.* **15**, 542–545.
Margiolaki, I., Wright, J. P., Fitch, A. N., Fox, G. C. & Von Dreele, R. B. (2005). *Synchrotron X-ray powder diffraction study of hexagonal turkey egg-white lysozyme*. *Acta Cryst.* D**61**, 423–432.
Mighell, A. D. (1976). *The reduced cell: its use in the identification of crystalline materials*. *J. Appl. Cryst.* **9**, 491–498.
Mighell, A. D. (2001). *Lattice symmetry and identification – the fundamental role of reduced cells in materials characterization*. *J. Res. Natl Inst. Stand. Technol.* **106**, 983–995.
Mighell, A. D. & Santoro, A. (1975). *Geometrical ambiguities in the indexing of powder patterns*. *J. Appl. Cryst.* **8**, 372–374.
Nash, J. C. (1990). *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation*, 2nd ed. Bristol: Adam Hilger.
Neumann, M. A. (2003). *X-cell: a novel indexing algorithm for routine tasks and difficult cases*. *J. Appl. Cryst.* **36**, 356–365.
Niggli, P. (1928). *Handbuch der Experimentalphysik*, Vol. 7, Part 1. Leipzig: Akademische Verlagsgesellschaft.
Oishi, R., Yonemura, M., Hoshikawa, A., Ishigaki, T., Mori, K., Torii, S., Morishima, T. & Kamiyama, T. (2009). *New approach to the indexing of powder diffraction patterns using topographs*. *Z. Kristallogr. Suppl.* **30**, 15–20.
Oishi-Tomiyasu, R. (2012). *Rapid Bravais-lattice determination algorithm for lattice parameters containing large observation errors*. *Acta Cryst.* A**68**, 525–535.
Oishi-Tomiyasu, R. (2013). *Reversed de Wolff figure of merit and its application to powder indexing solutions*. *J. Appl. Cryst.* **46**, 1277–1282.
Oishi-Tomiyasu, R. (2014*a*). *Method to generate all the geometrical ambiguities of powder indexing solutions*. *J. Appl. Cryst.* **47**, 2055–2059.
Oishi-Tomiyasu, R. (2014*b*). *Robust powder auto-indexing using many peaks*. *J. Appl. Cryst.* **47**, 593–598.
Oishi-Tomiyasu, R. (2016). *A table of geometrical ambiguities in powder indexing obtained by exhaustive search*. *Acta Cryst.* A**72**, 73–80.
Paszkowicz, W. (1996). *Application of the smooth genetic algorithm for indexing powder patterns – tests for the orthorhombic system*. *Mater. Sci. Forum*, **228–231**, 19–24.
Pecharsky, V. K. & Zavalij, P. Y. (2009). *Determination and refinement of the unit cell*. In *Fundamentals of Powder Diffraction and Structural Characterization of Materials*, 2nd ed., pp. 407–495. New York: Springer.
Roisnel, T. & Rodríquez-Carvajal, J. (2001). *WinPLOTR: a windows tool for powder diffraction pattern analysis*. *Mater. Sci. Forum*, **378–381**, 118–123.
Runge, C. (1917). *Die Bestimmung eines Kristallsystems durch Rontgenstrahlen*. *Phys. Z.* **18**, 509–515.
Santoro, A. & Mighell, A. D. (1970). *Determination of reduced cells*. *Acta Cryst.* A**26**, 124–127.
Santoro, A. & Mighell, A. D. (1972). *Properties of crystal lattices: the derivative lattices and their determination*. *Acta Cryst.* A**28**, 284–287.
Santoro, A., Mighell, A. D. & Rodgers, J. R. (1980). *The determination of the relationship between derivative lattices*. *Acta Cryst.* A**36**, 796–800.
Shirley, R. (1980). *Data accuracy for powder indexing*. In *Accuracy in Powder Diffraction*, edited by S. Block & C. R. Hubbard, *NBS Spec. Publ.* **567**, 361–382.
Shirley, R. (2002). *User Manual, The Crysfire 2002 system for Automatic Powder Indexing*. Guildford: Lattice Press.
Shirley, R. (2003). *Overview of powder-indexing program algorithms (history and strengths and weaknesses)*. *IUCr Comput. Comm. Newsl.* **2**, 48–54. http://www.iucr.org/resources/commissions/crystallographic-computing/newsletters/2.
Smith, G. S. (1977). *Estimating the unit-cell volume from one line in a powder diffraction pattern: the triclinic case*. *J. Appl. Cryst.* **10**, 252–255.
Smith, G. S. & Snyder, R. L. (1979). *F_N: a criterion for rating powder diffraction patterns and evaluating the reliability of powder-pattern indexing*. *J. Appl. Cryst.* **12**, 60–65.
Smrčok, L., Koppelhuber-Bitschau, B., Shankland, K., David, W. I. F., Tunega, D. & Resel, R. (2001). *Decafluoroquarterphenyl – crystal and molecular structure solved from X-ray powder data*. *Z. Kristallogr.* **216**, 63–66.
Tam, K. Y. & Compton, R. G. (1995). *GAMATCH – a genetic algorithm-based program for indexing crystal faces*. *J. Appl. Cryst.* **28**, 640–645.
Taupin, D. (1973). *A powder-diagram automatic-indexing routine*. *J. Appl. Cryst.* **6**, 380–385.
Visser, J. W. (1969). *A fully automatic program for finding the unit cell from powder data*. *J. Appl. Cryst.* **2**, 89–95.
Werner, P.-E. (1964). *Trial-and-error computer methods for the indexing of unknown powder patterns*. *Z. Kristallogr.* **120**, 375–387.
Werner, P.-E. (2002). *Autoindexing*. In *Structure Determination from Powder Diffraction Data*, edited by W. I. F. David, K. Shankland, L. B. McCusker & Ch. Baerlocher, pp. 118–135. Oxford University Press.
Werner, P.-E., Eriksson, L. & Westdahl, M. (1985). *TREOR, a semi-exhaustive trial-and-error powder indexing program for all symmetries*. *J. Appl. Cryst.* **18**, 367–370.
Wolff, P. M. de (1957). *On the determination of unit-cell dimensions from powder diffraction patterns*. *Acta Cryst.* **10**, 590–595.
Wolff, P. M. de (1958). *Detection of simultaneous zone relations among powder diffraction lines*. *Acta Cryst.* **11**, 664–665.
Wolff, P. M. de (1968). *A simplified criterion for the reliability of a powder pattern indexing*. *J. Appl. Cryst.* **1**, 108–113.
Young, R. A. (1993). *Introduction to the Rietveld method*. In *The Rietveld Method*, edited by R. A. Young, pp. 1–38. Oxford: IUCr/Oxford University Press.