3. METHODOLOGY

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 $\Delta V = 400$ Å^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