3.4. INDEXING

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	$(h^2 + k^2)A_{11} + l^2 A_{33}$
Hexagonal	$(h^2 + hk + k^2)A_{11} + l^2 A_{33}$
Orthorhombic	$h^2 A_{11} + k^2 A_{22} + l^2 A_{33}$
Monoclinic	$h^2 A_{11} + k^2 A_{22} + l^2 A_{33} + hl A_{13}$
Triclinic	$h^2 A_{11} + k^2 A_{22} + l^2 A_{33} + hk A_{12} + hl A_{13} + kl A_{23}$

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:

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

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

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

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

$$Q(hkl) = h^2 A_{11} + k^2 A_{22} + l^2 A_{33} + hk A_{12} + hl A_{13} + kl A_{23}, \qquad (3.4.2)$$

where

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

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:

$$Q(hkl) - \Delta < h^2 A_{11} + k^2 A_{22} + l^2 A_{33} + hk A_{12} + hl A_{13} + kl A_{23}$$
$$< Q(hkl) + \Delta. \qquad (3.4.3)$$

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

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

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

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

where $\langle |\Delta 2\theta| \rangle$ is the average absolute discrepancy between the observed and calculated 2θ peak position values and N_{poss} is the number of possible diffraction lines up to the Nth observed line. The values of $\langle |\Delta 2\theta| \rangle$ and N_{poss}, $(\langle |\Delta 2\theta| \rangle, N_{poss})$, 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) $Q_{partial}$ (Bergmann, 2007):

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

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. $Q_{partial}$ 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):

$$McM_{20} = [100/(R_p N_{20})] B_r S_y,$$

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):

$$\text{WRIP20} = \text{RAT}_{Rp}^2 \times \text{RAT}_{\text{Ind}} \times \text{RAT}_{\text{Pres}} \times w_u \times \text{RAT}_{M20}^{1/2}.$$

$$(3.4.5)$$

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

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

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

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

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. PERC_{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 $\text{PERC}_{\text{Pres}}$ is calculated: $\sum_{\text{all}}$ 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.) $\sum_{\text{Pres}}$ 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 (2014a, 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.)