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)

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

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

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

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 $\Delta V = 400$ 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