evaluated and saved in a file. To make room for new strong pixels as the spot search proceeds, all entries of strong pixels that are no longer needed are removed from the hash table and the remaining ones are rehashed. On termination, a list X_i', Y_i', Z_i' $(i = 1, \ldots, n)$ of the centroids of strong spots is available.

11.3.2.6. *Basis extraction*

Any reciprocal-lattice vector can be written in the form $\mathbf{p}_0^* = h\mathbf{b}_1^* + k\mathbf{b}_2^* + l\mathbf{b}_3^*$ where h, k, l are integer numbers and $\mathbf{b}_1^*, \mathbf{b}_2^*, \mathbf{b}_3^*$ are basis vectors of the lattice. The basis vectors which describe the orientation, metric and symmetry of the crystal, as well as the reflection indices h, k, l, have to be determined from the list of strong diffraction spots X_i', Y_i', Z_i' $(i = 1, \ldots, n)$. Ideally, each spot corresponds to a reciprocal-lattice vector $\mathbf{p}_0^*$ which satisfies the Laue equations after a crystal rotation by φ. Substituting the observed value Z' for the unknown φ angle (see Section 11.3.2.4), $\mathbf{p}_0^*$ is found from the observed spot coordinates as

$$\mathbf{p}_0^* = D(\mathbf{m}_2, -Z')(\mathbf{S}' - \mathbf{S}_0)$$

$$\mathbf{S}' = [(X' - X_0)\mathbf{d}_1 + (Y' - Y_0)\mathbf{d}_2 + F\mathbf{d}_3] \times \left\{\lambda \cdot \left[(X' - X_0)^2 + (Y' - Y_0)^2 + F^2\right]^{1/2}\right\}^{-1}.$$

Unfortunately, the reciprocal-lattice vectors $\mathbf{p}_{0i}^*$ $(i = 1, \ldots, n)$ derived from the above list of strong diffraction spots often contain a number of 'aliens' (spots arising from fluctuations of the background, from ice, or from satellite crystals) and a robust method has to be used which is still capable of recognizing the dominant lattice. One approach, suggested by Bricogne (1986) and implemented in a number of variants (Otwinowski & Minor, 1997; Steller *et al.*, 1997), is to identify a lattice basis as the three shortest linear independent vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$, each at a maximum of the Fourier transform $\sum_{i=1}^{n} \cos(2\pi\mathbf{b} \cdot \mathbf{p}_{0i}^*)$. Alternatively, a reciprocal basis for the dominant lattice can be determined from short differences between the reciprocal-lattice vectors (Howard, 1986; Kabsch, 1988*a*). As implemented in *XDS*, a lattice basis is found by the following procedure.

The list of given reciprocal-lattice points $\mathbf{p}_{0i}^*$ $(i = 1, \ldots, n)$ is first reduced to a small number m of low-resolution difference-vector clusters $\mathbf{v}_\mu^*$ $(\mu = 1, \ldots, m)$. f_μ is the population of a difference-vector cluster $\mathbf{v}_\mu^*$, that is the number of times the difference between any two reciprocal-lattice vectors $\mathbf{p}_{0i}^* - \mathbf{p}_{0j}^*$ is approximately equal to $\mathbf{v}_\mu^*$. In a second step, three linear independent vectors $\mathbf{b}_1^*, \mathbf{b}_2^*, \mathbf{b}_3^*$ are selected among all possible triplets of difference-vector clusters that maximize the function Q:

$$Q(\mathbf{b}_1^*, \mathbf{b}_2^*, \mathbf{b}_3^*) = \sum_{\mu=1}^{m} f_\mu q(\xi_1^\mu, \xi_2^\mu, \xi_3^\mu)$$

$$q(\xi_1^\mu, \xi_2^\mu, \xi_3^\mu) = \exp\left(-2\sum_{k=1}^{3}\left\{[\max(|\xi_k^\mu - h_k^\mu| - \varepsilon, 0)/\varepsilon]^2 + [\max(|h_k^\mu| - \delta, 0)]^2\right\}\right)$$

$$\xi_k^\mu = \mathbf{v}_\mu^* \cdot \mathbf{b}_k, \quad \mathbf{v}_\mu^* = \sum_{k=1}^{3} \xi_k^\mu \mathbf{b}_k^*, \quad \mathbf{b}_k \cdot \mathbf{b}_l^* = \begin{cases} 1 \text{ if } k = l; \\ 0 \text{ otherwise} \end{cases}$$

$$h_k^\mu = \text{ nearest integer to } \xi_k^\mu.$$

The absolute maximum of Q is assumed if all difference vectors can be expressed as small integral multiples of the best triplet. Deviations from this ideal situation are quantified by the quality measure q. The value of q declines sharply if the expansion coefficients ξ_k^μ deviate by more than ε from their nearest integers h_k^μ or if the indices are absolutely larger than δ. The constraint on the allowed range of indices prevents the selection of a spurious triplet of very short difference vector clusters which might be present in the set. Excellent results have been obtained using $\varepsilon = 0.05$ and $\delta = 5$. The best vector triplet thus found is refined against the observed difference-vector clusters. Finally, a reduced cell is derived from the refined reciprocal-base vector triplet as defined in IT *A* (1995), p. 743.

11.3.2.7. *Indexing*

Once a basis $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ of the lattice is available, integral indices h_i, k_i, l_i must be assigned to each reciprocal-lattice vector $\mathbf{p}_{0i}^*$ $(i = 1, \ldots, n)$. Using the integers nearest to $\mathbf{p}_{0i}^* \cdot \mathbf{b}_k$ $(k = 1, 2, 3)$ as indices of the reciprocal-lattice vectors $\mathbf{p}_{0i}^*$ could easily lead to a misindexing of longer vectors because of inaccuracies in the basis vectors $\mathbf{b}_k$ and the initial values of the parameters describing the instrumental setup. A more robust solution of the indexing problem is provided by the *local indexing method* which assigns only small index differences $h_i - h_j, k_i - k_j, l_i - l_j$ between pairs of neighbouring reciprocal-lattice vectors (Kabsch, 1993).

The reciprocal-lattice points can be considered as the nodes of a tree. The tree connects the n points to each other with the connections as its branches. The length ℓ_{ij} of a possible branch between nodes i and j is defined here as

$$\ell_{ij} = 1 - \exp\left(-2\sum_{k=1}^{3}\left\{[\max(|\xi_k^{ij} - h_k^{ij}| - \varepsilon, 0)/\varepsilon]^2 + [\max(|h_k^{ij}| - \delta, 0)]^2\right\}\right)$$

$$\xi_k^{ij} = (\mathbf{p}_{0i}^* - \mathbf{p}_{0j}^*) \cdot \mathbf{b}_k, \quad h_k^{ij} = \text{ nearest integer of } \xi_k^{ij}, \quad k = 1, 2, 3.$$

Reliable index differences are indicated by short branches; in fact, ℓ_{ij} is 0 if none of the indices h_k^{ij} is absolutely larger than δ and the ξ_k^{ij} are integer values to within ε. Typical values of ε and δ are $\varepsilon = 0.05$ and $\delta = 5$. Defining the length of a tree as the sum of the lengths of its branches, a shortest tree among all n^{n-2} possible trees is determined by the elegant algorithm described by Dijkstra (1976). Starting with arbitrary indices 0, 0, 0 for the root node, the local indexing method then consists of traversing the shortest tree and thereby assigning each node the indices of its predecessor plus the small index differences between the two nodes.

During traversal of the tree, each node is also given a subtree number. Starting with subtree number 1 for the root node, each successor node is given the same subtree number as its predecessor if the length of the connecting branch is below a minimal length $\ell_{\min}$. Otherwise its subtree number is incremented by 1. Thus all nodes in the same subtree have internally consistent reflection indices. Defining the size of a subtree by the number of its nodes, aliens are usually found in small subtrees. Finally, a constant index offset is determined such that the centroids of the observed reciprocal-lattice points $\mathbf{p}_{0i}^*$ belonging to the largest subtree and their corresponding grid vectors $\sum_{k=1}^{3} h_k^i \mathbf{b}_k^*$ are as close as possible. This offset is added to the indices of each reciprocal-lattice point.

11.3.2.8. *Refinement*

For a fixed detector, the diffraction pattern depends on the parameters $\mathbf{S}_0, \mathbf{m}_2, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, X_0, Y_0$ and F. Starting values for the parameters can be obtained by the procedures described above that do not rely on prior knowledge of the crystal orientation, space-group symmetry or unit-cell metric. Better estimates of the parameter values, as required for the subsequent integration step, can be obtained by the method of least squares from the list of n observed indexed reflection centroids h_i, k_i, l_i, X_i', Y_i', Z_i' $(i = 1, \ldots, n)$. In this method, the parameters are chosen to minimize a weighted sum of squares of the residuals

references