## 11.3. INTEGRATION, SCALING, SPACE-GROUP ASSIGNMENT AND POST REFINEMENT

$$FS/S \cdot \mathbf{d}_3 = (FS \cdot \mathbf{d}_1/S \cdot \mathbf{d}_3)\mathbf{d}_1 + (FS \cdot \mathbf{d}_2/S \cdot \mathbf{d}_3)\mathbf{d}_2 + F\mathbf{d}_3$$
$$= (X - X_0)\mathbf{d}_1 + (Y - Y_0)\mathbf{d}_2 + F\mathbf{d}_3,$$

which leads to a diffraction spot recorded at detector coordinates

$$X = X_0 + FS \cdot \mathbf{d}_1/S \cdot \mathbf{d}_3,$$
$$Y = Y_0 + FS \cdot \mathbf{d}_2/S \cdot \mathbf{d}_3.$$

### 11.3.2.3. *Standard spot shape*

A reciprocal-lattice point crosses the Ewald sphere by the shortest route only if the crystal happens to be rotated about an axis perpendicular to both the diffracted and incident beam wave vectors, the '$\beta$-axis' $\mathbf{e}_1 = \mathbf{S} \times \mathbf{S}_0/|\mathbf{S} \times \mathbf{S}_0|$, as introduced by Schutt & Winkler (1977). Rotation around the fixed axis $\mathbf{m}_2$, as enforced by the rotation camera, thus leads to an increase in the length of the shortest path by the factor $1/|\mathbf{e}_1 \cdot \mathbf{m}_2|$. This has motivated the introduction of a coordinate system $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, specific for each reflection, which has its origin on the surface of the Ewald sphere at the terminus of the diffracted beam wave vector $\mathbf{S}$,

$$\mathbf{e}_1 = \mathbf{S} \times \mathbf{S}_0/|\mathbf{S} \times \mathbf{S}_0|, \quad \mathbf{e}_2 = \mathbf{S} \times \mathbf{e}_1/|\mathbf{S} \times \mathbf{e}_1|,$$
$$\mathbf{e}_3 = (\mathbf{S} + \mathbf{S}_0)/|\mathbf{S} + \mathbf{S}_0|.$$

The unit vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ are tangential to the Ewald sphere, while $\mathbf{e}_3$ is perpendicular to $\mathbf{e}_1$ and $\mathbf{p}^* = \mathbf{S} - \mathbf{S}_0$. The shape of a reflection, as represented with respect to $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, then no longer contains geometrical distortions resulting from the fixed rotation axis of the camera and the oblique incidence of the diffracted beam on a flat detector. Instead, all reflections appear as if they had followed the shortest path through the Ewald sphere and had been recorded on the surface of the sphere.

A detector pixel at $X'$, $Y'$ in the neighbourhood of the reflection centre $X$, $Y$, when the crystal is rotated by $\varphi'$ instead of $\varphi$, is mapped to the profile coordinates $\varepsilon_1, \varepsilon_2, \varepsilon_3$ by the following procedure:

$$\mathbf{S}' = [(X' - X_0)\mathbf{d}_1 + (Y' - Y_0)\mathbf{d}_2 + F\mathbf{d}_3]$$
$$\times \{\lambda \cdot [(X' - X_0)^2 + (Y' - Y_0)^2 + F^2]^{1/2}\}^{-1}$$
$$\varepsilon_1 = \mathbf{e}_1 \cdot (\mathbf{S}' - \mathbf{S})180/(|\mathbf{S}|\pi),$$
$$\varepsilon_2 = \mathbf{e}_2 \cdot (\mathbf{S}' - \mathbf{S})180/(|\mathbf{S}|\pi)$$
$$\varepsilon_3 = \mathbf{e}_3 \cdot [D(\mathbf{m}_2, \varphi' - \varphi)\mathbf{p}^* - \mathbf{p}^*]180/(|\mathbf{p}^*|\pi) \simeq \zeta \cdot (\varphi' - \varphi)$$
$$\zeta = \mathbf{m}_2 \cdot \mathbf{e}_1.$$

$\zeta$ corrects for the increased path length of the reflection through the Ewald sphere and is closely related to the reciprocal Lorentz correction factor

$$L^{-1} = |\mathbf{m}_2 \cdot (\mathbf{S} \times \mathbf{S}_0)|/(|\mathbf{S}| \cdot |\mathbf{S}_0|) = |\zeta \cdot \sin \angle(\mathbf{S}, \mathbf{S}_0)|.$$

Because of crystal mosaicity and beam divergence, the intensity of a reflection is smeared around the diffraction maximum. The fraction of total reflection intensity found in the volume element $d\varepsilon_1 d\varepsilon_2 d\varepsilon_3$ at $\varepsilon_1, \varepsilon_2, \varepsilon_3$ can be approximated by Gaussian functions:

$$\omega(\varepsilon_1, \varepsilon_2, \varepsilon_3)d\varepsilon_1 d\varepsilon_2 d\varepsilon_3$$
$$= \frac{\exp(-\varepsilon_1^2/2\sigma_D^2)}{(2\pi)^{1/2}\sigma_D}d\varepsilon_1 \cdot \frac{\exp(-\varepsilon_2^2/2\sigma_D^2)}{(2\pi)^{1/2}\sigma_D}d\varepsilon_2 \cdot \frac{\exp(-\varepsilon_3^2/2\sigma_M^2)}{(2\pi)^{1/2}\sigma_M}d\varepsilon_3.$$

### 11.3.2.4. *Spot centroids and partiality*

The intensity of a reflection can be completely recorded on one image, or distributed among several adjacent images. The fraction $R_j$ of total intensity recorded on image $j$, the 'partiality' of the reflection, can be derived from the distribution function $\omega(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ as

$$R_j = \int_{-\infty}^{\infty} d\varepsilon_1 \int_{-\infty}^{\infty} d\varepsilon_2 \int_{\zeta[\varphi_0 + (j-1)\Delta_\varphi - \varphi]}^{\zeta(\varphi_0 + j\Delta_\varphi - \varphi)} d\varepsilon_3 \, \omega(\varepsilon_1, \varepsilon_2, \varepsilon_3)$$
$$= \{[1/(2\pi)^{1/2}\sigma_M]/|\zeta|\}$$
$$\times \int_{\varphi_0 + (j-1)\Delta_\varphi}^{\varphi_0 + j\Delta_\varphi} \exp[-(\varphi' - \varphi)^2/2(\sigma_M/|\zeta|)^2] \, d\varphi'$$
$$= \Big(\text{erf}[|\zeta|(\varphi_0 + j\Delta_\varphi - \varphi)/(2)^{1/2}\sigma_M]$$
$$- \text{erf}\{|\zeta|[\varphi_0 + (j-1)\Delta_\varphi - \varphi]/(2)^{1/2}\sigma_M\}\Big)/2.$$

The integral is evaluated by using a numerical approximation of the error function, erf (Abramowitz & Stegun, 1972).

While the spot centroids in the detector plane are usually good estimates for the detector position of the diffraction maximum, the angular centroid about the rotation axis,

$$Z = \varphi_0 + \Delta_\varphi \cdot \sum_{j=-\infty}^{\infty} (j - 1/2)R_j \approx \varphi,$$

can be a rather poor guess for the true $\varphi$ angle of the maximum. Its accuracy depends strongly on the value of $\varphi$ and the size of the oscillation range $\Delta_\varphi$ relative to the mosaicity $\sigma_M$ of the crystal. For a reflection fully recorded on image $j$, the value $Z = \varphi_0 + (j - 1/2) \cdot \Delta_\varphi$ will always be obtained, which is correct only if $\varphi$ accidentally happens to be close to the centre of the rotation range of the image. In contrast, the $\varphi$ angle of a partial reflection recorded on images $j$ and $j + 1$ is closely approximated by $Z = \varphi_0 + [j + (R_{j+1} - R_j)/2] \cdot \Delta_\varphi$. If many images contribute to the spot intensity, $Z(\varphi)$ is always an excellent approximation to the ideal angular position $\varphi$ when the Laue equations are satisfied; in fact, in the limiting case of infinitely fine-sliced data, it can be shown that $\lim_{\Delta_\varphi \to 0} Z(\varphi) = \varphi$.

Most refinement routines minimize the discrepancies between the predicted $\varphi$ angles and their approximations obtained from the observed $Z$ centroids, and must therefore carefully distinguish between fully and partially recorded reflections. This distinction is unnecessary, however, if observed $Z$ centroids are compared with their analytic forms instead, because the sensitivity of the centroid positions to the diffraction parameters is correctly weighted in either case (see Section 11.3.2.8).

### 11.3.2.5. *Localizing diffraction spots*

Recognition and refinement of the parameter values controlling the observed diffraction pattern begins with the extraction of a list of coordinates of strong spots occurring in the images. As implemented in *XDS*, this list is obtained by the following procedure. First, each pixel value is compared with the mean value and standard deviation of surrounding pixels in the same image and classified as a strong pixel if its value exceeds the mean by a given multiple (typically 3 to 5) of the standard deviation. Values of the strong pixels and their location addresses and image running numbers are stored in a hash table during spot search [for a discussion of the hash technique, see Wirth (1976)]. After processing a fixed number of images, or when the table is full, all strong pixels are labelled by a unique number identifying the spot to which they belong. By definition, any two such pixels which can be connected by direct strong neighbours in two or three dimensions (if there are adjacent images) belong to the same spot (equivalence class). The labelling is achieved by the highly efficient algorithm for the recording of equivalence classes developed by Rem (see Dijkstra, 1976). At the end of this procedure, the table is searched for spots that have no contributing strong pixel on the current or the previous image. These spots are complete and their centroids are

219

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 $\varphi$. Substituting the observed value $Z'$ for the unknown $\varphi$ 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, 1988a). 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 $\mathbf{v}^*_\mu$ $(\mu = 1, \ldots, m)$. $f_\mu$ 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 $Q$:

$$Q(\mathbf{b}^*_1, \mathbf{b}^*_2, \mathbf{b}^*_3) = \sum_{\mu=1}^{m} f_\mu q(\xi^\mu_1, \xi^\mu_2, \xi^\mu_3)$$

$$q(\xi^\mu_1, \xi^\mu_2, \xi^\mu_3) = \exp\left( -2\sum_{k=1}^{3}\left\{ [\max(|\xi^\mu_k - h^\mu_k| - \varepsilon, 0)/\varepsilon]^2 \right.\right.$$
$$\left.\left. + [\max(|h^\mu_k| - \delta, 0)]^2 \right\} \right)$$

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

$$h^\mu_k = \text{ nearest integer to } \xi^\mu_k.$$

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 $\xi^\mu_k$ deviate by more than $\varepsilon$ from their nearest integers $h^\mu_k$ or if the indices are absolutely larger than $\delta$. 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 $\ell_{ij}$ of a possible branch between nodes $i$ and $j$ is defined here as

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

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

Reliable index differences are indicated by short branches; in fact, $\ell_{ij}$ is 0 if none of the indices $h^{ij}_k$ is absolutely larger than $\delta$ and the $\xi^{ij}_k$ are integer values to within $\varepsilon$. Typical values of $\varepsilon$ and $\delta$ 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 $\sum_{k=1}^{3} h^i_k\mathbf{b}^*_k$ 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