

11.3. Integration, scaling, space-group assignment and post refinement

BY W. KABSCH

11.3.1. Introduction

Key steps in the processing of diffraction data from single crystals involve: (a) accurate modelling of the positions of all the reflections recorded in the images; (b) integration of diffraction intensities; (c) data correction, scaling and post refinement; and (d) space-group assignment. Much of the theory and many of the methods for carrying out these steps were developed about two decades ago for processing rotation data recorded on film and were later extended to exploit fully the capabilities of a variety of electronic area detectors; some CCD (charge-coupled device) and multiwire detectors allow the recording of finely sliced rotation data because of their fast data read-out. In this chapter, the principles of the methods are described as they are employed by the program *XDS* (Section 25.2.9). These apply equally well to rotation images covering small or large oscillation ranges. A large number of other systems have been developed which differ in the details of the implementations. Some of these packages are described in Chapter 25.2. The theory and practice of processing fine-sliced data have recently been discussed by Pflugrath (1997).

11.3.2. Modelling rotation images

The observed diffraction pattern, *i.e.*, the positions of the reflections recorded in the rotation-data images, is controlled by a small set of parameters which must be accurately determined before integration can start. Approximate values for some of these parameters are given by the experimental setup, whereas others may be completely unknown and must be obtained from the rotation images. This is achieved by automatic location of strong diffraction spots, extraction of a primitive lattice basis that yields integer indices for the observed reflections, and subsequent refinement of all parameters to minimize the discrepancies between observed and calculated spot positions in the data images.

11.3.2.1. Coordinate systems and parameters

In the rotation method, the incident beam wave vector \mathbf{S}_0 of length $1/\lambda$ (λ is the wavelength) is fixed while the crystal is rotated around a fixed axis described by a unit vector \mathbf{m}_2 . \mathbf{S}_0 points from the X-ray source towards the crystal. It is assumed that the incident beam and the rotation axis intersect at one point at which the crystal must be located. This point is defined as the origin of a right-handed orthonormal laboratory coordinate system $\{\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3\}$. This fixed but otherwise arbitrary system is used as a reference frame to specify the setup of the diffraction experiment.

Diffraction data are assumed to be recorded on a fixed planar detector. A right-handed orthonormal detector coordinate system $\{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$ is defined such that a point with coordinates X, Y in the detector plane is represented by the vector $(X - X_0)\mathbf{d}_1 + (Y - Y_0)\mathbf{d}_2 + F\mathbf{d}_3$ with respect to the laboratory coordinate system. The origin X_0, Y_0 of the detector plane is found at a distance $|F|$ from the crystal position. It is assumed that the diffraction data are recorded on adjacent non-overlapping rotation images, each covering a constant oscillation range Δ_φ with image No. 1 starting at spindle angle φ_0 .

Diffraction geometry is conveniently expressed with respect to a right-handed orthonormal goniostat system $\{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}$. It is constructed from the rotation axis and the incident beam direction such that $\mathbf{m}_1 = (\mathbf{m}_2 \times \mathbf{S}_0)/|\mathbf{m}_2 \times \mathbf{S}_0|$ and $\mathbf{m}_3 = \mathbf{m}_1 \times \mathbf{m}_2$. The origin of the goniostat system is defined to coincide with the origin of the laboratory system.

Finally, a right-handed crystal coordinate system $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ and its reciprocal basis $\{\mathbf{b}_1^*, \mathbf{b}_2^*, \mathbf{b}_3^*\}$ are defined to represent the unrotated crystal, *i.e.*, at rotation angle $\varphi = 0^\circ$, such that any reciprocal-lattice vector can be expressed as $\mathbf{p}_0^* = h\mathbf{b}_1^* + k\mathbf{b}_2^* + l\mathbf{b}_3^*$ where h, k, l are integers.

Using a Gaussian model, the shape of the diffraction spots is specified by two parameters: the standard deviations of the reflecting range σ_M and the beam divergence σ_D (see Section 11.3.2.3). This leads to an integration region around the spot defined by the parameters δ_M and δ_D , which are typically chosen to be 6–10 times larger than σ_M and σ_D , respectively.

Knowledge of the parameters $\mathbf{S}_0, \mathbf{m}_2, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, X_0, Y_0, F, \mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \varphi_0$ and Δ_φ is sufficient to compute the location of all diffraction peaks recorded in the data images. Determination and refinement of these parameters are described in the following sections.

11.3.2.2. Spot prediction

It is assumed here that accurate values of all parameters describing the diffraction experiment are available, permitting prediction of the positions of all diffraction peaks recorded in the data images. Let \mathbf{p}_0^* denote any arbitrary reciprocal-lattice vector if the crystal has not been rotated, *i.e.*, at rotation angle $\varphi = 0^\circ$. \mathbf{p}_0^* can be expressed by its components with respect to the orthonormal goniostat system as

$$\mathbf{p}_0^* = \mathbf{m}_1(\mathbf{m}_1 \cdot \mathbf{p}_0^*) + \mathbf{m}_2(\mathbf{m}_2 \cdot \mathbf{p}_0^*) + \mathbf{m}_3(\mathbf{m}_3 \cdot \mathbf{p}_0^*).$$

Depending on the diffraction geometry, \mathbf{p}_0^* may be rotated into a position fulfilling the reflecting condition. The required rotation angle φ and the coordinates X, Y of the diffracted beam at its intersection with the detector plane can be found from \mathbf{p}_0^* as follows.

Rotation by φ around axis \mathbf{m}_2 changes \mathbf{p}_0^* into \mathbf{p}^* .

$$\begin{aligned} \mathbf{p}^* &= D(\mathbf{m}_2, \varphi)\mathbf{p}_0^* = \mathbf{m}_2(\mathbf{m}_2 \cdot \mathbf{p}_0^*) + [\mathbf{p}_0^* - \mathbf{m}_2(\mathbf{m}_2 \cdot \mathbf{p}_0^*)] \cos \varphi \\ &\quad + \mathbf{m}_2 \times \mathbf{p}_0^* \sin \varphi \\ &= \mathbf{m}_1(\mathbf{m}_1 \cdot \mathbf{p}_0^* \cos \varphi + \mathbf{m}_3 \cdot \mathbf{p}_0^* \sin \varphi) + \mathbf{m}_2 \mathbf{m}_2 \cdot \mathbf{p}_0^* \\ &\quad + \mathbf{m}_3(\mathbf{m}_3 \cdot \mathbf{p}_0^* \cos \varphi - \mathbf{m}_1 \cdot \mathbf{p}_0^* \sin \varphi) \\ &= \mathbf{m}_1(\mathbf{m}_1 \cdot \mathbf{p}^*) + \mathbf{m}_2(\mathbf{m}_2 \cdot \mathbf{p}^*) + \mathbf{m}_3(\mathbf{m}_3 \cdot \mathbf{p}^*). \end{aligned}$$

The incident and diffracted beam wave vectors, \mathbf{S}_0 and \mathbf{S} , have their termini on the Ewald sphere and satisfy the Laue equations

$$\mathbf{S} = \mathbf{S}_0 + \mathbf{p}^*, \quad \mathbf{S}^2 = \mathbf{S}_0^2 \implies \mathbf{p}^{*2} = -2\mathbf{S}_0 \cdot \mathbf{p}^* = \mathbf{p}_0^{*2}.$$

If $\rho = [\mathbf{p}_0^{*2} - (\mathbf{p}_0^* \cdot \mathbf{m}_2)^2]^{1/2}$ denotes the distance of \mathbf{p}_0^* from the rotation axis, solutions for \mathbf{p}^* and φ can be obtained in terms of \mathbf{p}_0^* as

$$\begin{aligned} \mathbf{p}^* \cdot \mathbf{m}_3 &= [-\mathbf{p}_0^{*2}/2 - (\mathbf{p}_0^* \cdot \mathbf{m}_2)(\mathbf{S}_0 \cdot \mathbf{m}_2)]/\mathbf{S}_0 \cdot \mathbf{m}_3 \\ \mathbf{p}^* \cdot \mathbf{m}_2 &= \mathbf{p}_0^* \cdot \mathbf{m}_2 \\ \mathbf{p}^* \cdot \mathbf{m}_1 &= \pm[\rho^2 - (\mathbf{p}^* \cdot \mathbf{m}_3)^2]^{1/2} \\ \cos \varphi &= [(\mathbf{p}^* \cdot \mathbf{m}_1)(\mathbf{p}_0^* \cdot \mathbf{m}_1) + (\mathbf{p}^* \cdot \mathbf{m}_3)(\mathbf{p}_0^* \cdot \mathbf{m}_3)]/\rho^2 \\ \sin \varphi &= [(\mathbf{p}^* \cdot \mathbf{m}_1)(\mathbf{p}_0^* \cdot \mathbf{m}_3) - (\mathbf{p}^* \cdot \mathbf{m}_3)(\mathbf{p}_0^* \cdot \mathbf{m}_1)]/\rho^2. \end{aligned}$$

In general, there are two solutions according to the sign of $\mathbf{p}^* \cdot \mathbf{m}_1$. If $\rho^2 < (\mathbf{p}^* \cdot \mathbf{m}_3)^2$ or $\mathbf{p}_0^{*2} > 4\mathbf{S}_0^2$, the Laue equations have no solution and the reciprocal-lattice point \mathbf{p}_0^* is in the 'blind' region.

If $F\mathbf{S} \cdot \mathbf{d}_3 > 0$, the diffracted beam intersects the detector plane at the point

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

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

which leads to a diffraction spot recorded at detector coordinates

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

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 ' β -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} ,

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

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 φ' instead of φ , is mapped to the profile coordinates $\varepsilon_1, \varepsilon_2, \varepsilon_3$ by the following procedure:

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

ζ 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:

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

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

$$\begin{aligned} R_j &= \int_{-\infty}^{\infty} d\varepsilon_1 \int_{-\infty}^{\infty} d\varepsilon_2 \int_{\zeta[\varphi_0+(j-1)\Delta_\varphi-\varphi]}^{\zeta(\varphi_0+j\Delta_\varphi-\varphi)} d\varepsilon_3 \omega(\varepsilon_1, \varepsilon_2, \varepsilon_3) \\ &= \{[1/(2\pi)^{1/2}\sigma_M]/|\zeta|\} \\ &\quad \times \int_{\varphi_0+(j-1)\Delta_\varphi}^{\varphi_0+j\Delta_\varphi} \exp[-(\varphi' - \varphi)^2/2(\sigma_M/|\zeta|)^2] d\varphi' \\ &= \left(\operatorname{erf}[|\zeta|(\varphi_0 + j\Delta_\varphi - \varphi)/(2)^{1/2}\sigma_M] \right. \\ &\quad \left. - \operatorname{erf}[|\zeta|[\varphi_0 + (j-1)\Delta_\varphi - \varphi]/(2)^{1/2}\sigma_M] \right) / 2. \end{aligned}$$

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 φ angle of the maximum. Its accuracy depends strongly on the value of φ and the size of the oscillation range Δ_φ relative to the mosaicity σ_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 φ accidentally happens to be close to the centre of the rotation range of the image. In contrast, the φ 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 φ when the Laue equations are satisfied; in fact, in the limiting case of infinitely fine-sliced data, it can be shown that $\lim_{\Delta_\varphi \rightarrow 0} Z(\varphi) = \varphi$.

Most refinement routines minimize the discrepancies between the predicted φ 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

11. DATA PROCESSING

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 ($i = 1, \dots, n$) 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 ($i = 1, \dots, n$). 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

$$\begin{aligned} \mathbf{p}_0^* &= D(\mathbf{m}_2, -Z')(\mathbf{S}' - \mathbf{S}_0) \\ \mathbf{S}' &= [(X' - X_0)\mathbf{d}_1 + (Y' - Y_0)\mathbf{d}_2 + F\mathbf{d}_3] \\ &\times \left\{ \lambda \cdot [(X' - X_0)^2 + (Y' - Y_0)^2 + F^2]^{1/2} \right\}^{-1}. \end{aligned}$$

Unfortunately, the reciprocal-lattice vectors \mathbf{p}_{0i}^* ($i = 1, \dots, n$) 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}^* ($i = 1, \dots, n$) is first reduced to a small number m of low-resolution difference-vector clusters \mathbf{v}_μ^* ($\mu = 1, \dots, m$). f_μ is the population of a difference-vector cluster \mathbf{v}_μ^* , 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}_μ^* . 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 :

$$\begin{aligned} Q(\mathbf{b}_1^*, \mathbf{b}_2^*, \mathbf{b}_3^*) &= \sum_{\mu=1}^m f_\mu q(\xi_1^\mu, \xi_2^\mu, \xi_3^\mu) \\ q(\xi_1^\mu, \xi_2^\mu, \xi_3^\mu) &= \exp\left(-2 \sum_{k=1}^3 \left\{ [\max(|\xi_k^\mu - h_k^\mu| - \varepsilon, 0)/\varepsilon]^2 \right. \right. \\ &\quad \left. \left. + [\max(|h_k^\mu| - \delta, 0)]^2 \right\} \right) \\ \xi_k^\mu &= \mathbf{v}_\mu^* \cdot \mathbf{b}_k, \quad \mathbf{v}_\mu^* = \sum_{k=1}^3 \xi_k^\mu \mathbf{b}_k^*, \quad \mathbf{b}_k \cdot \mathbf{b}_l = \begin{cases} 1 & \text{if } k = l; \\ 0 & \text{otherwise} \end{cases} \\ h_k^\mu &= \text{nearest integer to } \xi_k^\mu. \end{aligned}$$

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}^* ($i = 1, \dots, n$). 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

$$\begin{aligned} \ell_{ij} &= 1 - \exp\left(-2 \sum_{k=1}^3 \left\{ [\max(|\xi_k^{ij} - h_k^{ij}| - \varepsilon, 0)/\varepsilon]^2 \right. \right. \\ &\quad \left. \left. + [\max(|h_k^{ij}| - \delta, 0)]^2 \right\} \right) \\ \xi_k^{ij} &= (\mathbf{p}_{0i}^* - \mathbf{p}_{0j}^*) \cdot \mathbf{b}_k, \quad h_k^{ij} = \text{nearest integer of } \xi_k^{ij}, \quad k = 1, 2, 3. \end{aligned}$$

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 ℓ_{\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 $h_i, k_i, l_i, X'_i, Y'_i, Z'_i$ ($i = 1, \dots, n$). In this method, the parameters are chosen to minimize a weighted sum of squares of the residuals

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

$$E = w_X \sum_{i=1}^n (\Delta_X^i)^2 + w_Y \sum_{i=1}^n (\Delta_Y^i)^2 + w_Z \sum_{i=1}^n (\Delta_Z^i)^2.$$

The residuals between the calculated (X_i, Y_i, Z_i) and observed spot centroids are

$$\begin{aligned} \Delta_X^i &= X_i - X'_i = X_0 + F\mathbf{S}_i \cdot \mathbf{d}_1/\mathbf{S}_i \cdot \mathbf{d}_3 - X'_i \\ \Delta_Y^i &= Y_i - Y'_i = Y_0 + F\mathbf{S}_i \cdot \mathbf{d}_2/\mathbf{S}_i \cdot \mathbf{d}_3 - Y'_i \\ \Delta_Z^i &= Z_i - Z'_i = \varphi_0 + \Delta_\varphi \sum_{j=-\infty}^{\infty} (j - 1/2)R_j^i - Z'_i. \end{aligned}$$

Let s_μ ($\mu = 1, \dots, k$) denote the k independent parameters for which initial estimates are available. Expanding the residuals to first order in the parameter changes δs_μ gives

$$\Delta(s_\mu + \delta s_\mu) \approx \Delta(s_\mu) + \sum_{\mu=1}^k \frac{\partial \Delta}{\partial s_\mu} \delta s_\mu.$$

The parameters should be changed in such a way as to minimize $E(\delta s_\mu)$, which implies $\partial E/\partial \delta s_\mu = 0$ for $\mu = 1, \dots, k$. The δs_μ are found as the solution of the k normal equations

$$\begin{aligned} \sum_{\mu'=1}^k \left(w_X \sum_{i=1}^n \frac{\partial \Delta_X^i}{\partial s_\mu} \frac{\partial \Delta_X^i}{\partial s_{\mu'}} + w_Y \sum_{i=1}^n \frac{\partial \Delta_Y^i}{\partial s_\mu} \frac{\partial \Delta_Y^i}{\partial s_{\mu'}} + w_Z \sum_{i=1}^n \frac{\partial \Delta_Z^i}{\partial s_\mu} \frac{\partial \Delta_Z^i}{\partial s_{\mu'}} \right) \delta s_{\mu'} \\ = - \left(w_X \sum_{i=1}^n \Delta_X^i \frac{\partial \Delta_X^i}{\partial s_\mu} + w_Y \sum_{i=1}^n \Delta_Y^i \frac{\partial \Delta_Y^i}{\partial s_\mu} + w_Z \sum_{i=1}^n \Delta_Z^i \frac{\partial \Delta_Z^i}{\partial s_\mu} \right). \end{aligned}$$

The parameters are corrected by δs_μ and a new cycle of refinement is started until a minimum of E is reached. The weights

$$w_X = 1/\sum_{i=1}^n (\Delta_X^i)^2, \quad w_Y = 1/\sum_{i=1}^n (\Delta_Y^i)^2, \quad w_Z = 1/\sum_{i=1}^n (\Delta_Z^i)^2$$

are calculated with the current guess for s_μ at the beginning of each cycle.

The derivatives appearing in the normal equations can be worked out from the definitions given in Sections 11.3.2.2 and 11.3.2.4, and only the form of the gradient of the Z residuals is shown. Assuming $\sigma_i = \sigma_M/|\zeta_i|$ ($i = 1, \dots, n$) is constant for each reflection, the gradients of the Z residuals are obtained from the chain rule and the relation $d \operatorname{erf}(z)/dz = [2/(\pi)^{1/2}] \exp(-z^2)$.

$$\begin{aligned} \frac{\partial \Delta_Z^i}{\partial s_\mu} &= \frac{\partial \Delta_Z^i}{\partial \varphi_i} \frac{\partial \varphi_i}{\partial s_\mu} \\ \frac{\partial \Delta_Z^i}{\partial \varphi_i} &= \frac{\Delta_\varphi}{(2\pi)^{1/2} \sigma_i} \sum_{j=-\infty}^{\infty} \exp[-(\varphi_0 + j\Delta_\varphi - \varphi_i)^2/2\sigma_i^2] \\ \frac{\partial \varphi_i}{\partial s_\mu} &= \cos \varphi_i \frac{\partial \sin \varphi_i}{\partial s_\mu} - \sin \varphi_i \frac{\partial \cos \varphi_i}{\partial s_\mu}. \end{aligned}$$

Obviously, $\partial \Delta_Z^i/\partial s_\mu$ is small for a fully recorded reflection because of the small values of all exponentials appearing in $\partial \Delta_Z^i/\partial \varphi_i$. In contrast, the gradient for a partial reflection, equally recorded on two adjacent images, is most sensitive to parameter variations because one of the exponentials assumes its maximum value. In the limiting case of infinitely fine-sliced data, it can be shown that $\lim_{\Delta_\varphi \rightarrow 0} \partial \Delta_Z^i/\partial \varphi_i = 1$. Thus, the refinement scheme based on observed Z centroids, as described here and implemented in *XDS*, is applicable to fine-sliced data – and to data recorded with a large oscillation range as well.

11.3.3. Integration

A fundamental requirement for a general integration method is that it should distinguish carefully between signal and background

points within its integration domain. For weak reflections, this distinction cannot be made reliably because of the errors superimposed on the signal. The problem can be solved, however, provided that both weak and strong reflections share the same profile shape – an assumption that has been adopted by most data-processing packages.

The intensity distribution of a reflection can be modelled analytically or derived from the observed profiles of neighbouring strong spots. For the rotation method, the profile shape depends strongly on the specific path of the reflection through the Ewald sphere and on variations in the angle of incidence of the diffracted beam on a flat detector. These geometrical distortions can be eliminated by mapping the reflections onto the coordinate system defined in Section 11.3.2.3, which simplifies the task of modelling the expected intensity distribution as all reflection profiles become similar.

11.3.3.1. Spot extraction

The region around a spot is defined by the two parameters δ_D and δ_M , which represent spot diameter and reflecting range, respectively. It is assumed that the coordinates of all image pixels contributing to the intensity of a spot satisfy $|\varepsilon_1| \leq \delta_D/2$, $|\varepsilon_2| \leq \delta_D/2$ and $|\varepsilon_3| \leq \delta_M/2$ when mapped to the profile coordinate system $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ defined in Section 11.3.2.3. Regions of neighbouring reflections may overlap. As implemented in *XDS*, potential overlap is dealt with by a simple strategy: pixels within the overlap region are assigned to the nearest spot. This is carried out in two steps. First, reflections predicted to occur on a given rotation image are found by generating and testing all possible indices h, k, l up to the highest resolution recorded by the detector. Reflection indices, coordinates of the diffracted beam wave vector and the expected fraction of spot intensity recorded on the image are saved in a table. In the second step, each reflection boundary is traced in the image and corrected to exclude pixels belonging to overlapping reflections, which are rapidly located in the table by the hash technique. The image scaling factor obtained from the mean image background and the neighbourhood pixel values belonging to the reflections recorded in the image are saved on a scratch file dedicated to the currently processed data image.

At regular intervals, these files are merged such that all pixel values belonging to a spot found in the contributing images follow each other. Reflections for which contributing pixels are expected further ahead in data processing are just copied to a scratch output file. The other reflections are mapped to the Ewald sphere, as described below, and their three-dimensional profiles and accompanying information are routed to the main output file of the spot-extraction step. After the file-merging procedure, spot extraction continues.

11.3.3.2. Background

The region around a spot is assumed to have been chosen to be large enough to include a sufficient number of pixels which can be used for determination of the background. Background determination, as implemented in *XDS*, begins by sorting all pixels belonging to a reflection by increasing intensity. For weak or absent reflections, these values should represent a random sample drawn from a normal distribution. If this is not the case, the pixel with the largest intensity is removed until the sampling distribution of the remaining smaller items satisfies the expected distribution. This method will also exclude pixels with unexpectedly high values, such as ice reflections. The background, determined as the mean value of the accepted pixels, is systematically overestimated for strong spots because of some residual intensity extending into the accepted background pixels. This residual intensity is estimated

11. DATA PROCESSING

from the expected distribution $\omega(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ defined in Section 11.3.2.3 and removed from the final background value.

11.3.3.3. Standard profiles

Reflection profiles are represented on the Ewald sphere within a domain D_0 comprising $2n_1 + 1, 2n_2 + 1, 2n_3 + 1$ equidistant grid-points along $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, respectively. The sampling distances between adjacent grid points are then $\Delta_1 = \delta_D/(2n_1 + 1)$, $\Delta_2 = \delta_D/(2n_2 + 1)$, $\Delta_3 = \delta_M/(2n_3 + 1)$. Thus, grid coordinate ν_3 ($\nu_3 = -n_3, \dots, n_3$) covers the set of rotation angles

$$\Gamma_{\nu_3} = \{\varphi' | (\nu_3 - 1/2)\Delta_3 \leq (\varphi' - \varphi) \cdot \zeta \leq (\nu_3 + 1/2)\Delta_3\}.$$

Contributions to the spot intensity come from one or several adjacent data images ($j = j_1, \dots, j_2$), each covering the set of rotation angles

$$\Gamma_j = \{\varphi' | \varphi_0 + (j - 1)\Delta_\varphi \leq \varphi' \leq \varphi_0 + j\Delta_\varphi\}.$$

Assuming Gaussian profiles along \mathbf{e}_3 for all reflections (see Section 11.3.2.3), the fraction of counts (after subtraction of the background) contributed by data frame j to grid coordinate ν_3 is

$$f_{\nu_3 j} \approx \int_{\Gamma_j \cap \Gamma_{\nu_3}} \exp[-(\varphi' - \varphi)^2/2\sigma^2] d\varphi' \\ \times \left\{ \int_{\Gamma_j} \exp[-(\varphi' - \varphi)^2/2\sigma^2] d\varphi' \right\}^{-1},$$

where $\sigma = \sigma_M/|\zeta|$. The integrals can be expressed in terms of the error function, for which efficient numerical approximations are available (Abramowitz & Stegun, 1972). Finally, each pixel on data image j belonging to the reflection is subdivided into 5×5 areas of equal size, and $f_{\nu_3 j}/25$ of the pixel signal is added to the profile value at grid coordinates ν_1, ν_2, ν_3 corresponding to each subdivision.

This complicated procedure leads to more uniform intensity profiles for all reflections than using their untransformed shape. This simplifies the task of modelling the expected intensity distribution needed for integration by profile fitting. As implemented in *XDS*, reference profiles are learnt every 5° of crystal rotation at nine positions on the detector, each covering an equal area of the detector face. In the learning phase, profile boxes of the strong reflections are normalized and added to their nearest reference profile boxes. The contributions are weighted according to the distance from the location of the reference profile. Each grid point within the average profile boxes is classified as signal if it is above 2% of the peak maximum. Finally, each profile is scaled such that the sum of its signal pixels normalizes to one. The analytic expression $\omega(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ defined in Section 11.3.2.3 for the expected intensity distribution is only a rough initial approximation which is now replaced by the empirical reference profiles.

11.3.3.4. Intensity estimation

If an expected intensity distribution $\{p_i | i \in D_0\}$ of the observed profile is given in a domain D_0 , the reflection intensity I can be estimated as

$$I = \sum_{i \in D} (c_i - b_i) p_i / v_i \Big/ \sum_{i \in D} p_i^2 / v_i,$$

which minimizes the function

$$\psi(I) = \sum_{i \in D} (c_i - I \cdot p_i - b_i)^2 / v_i, \quad \sum_{i \in D_0} p_i = 1.$$

b_i, c_i, v_i ($i \in D$) are background, contents and variance of pixels observed in a subdomain $D \subseteq D_0$ of the expected distribution. The

background b_i underneath a diffraction spot is often assumed to be a constant which is estimated from the neighbourhood around the reflection. Determination of reflection intensities by profile fitting has a long tradition (Diamond, 1969; Ford, 1974; Kabsch, 1988b; Otwinowski, 1993). Implementations of the method differ mainly in their assumptions about the variances v_i . Ford uses constant variances, which works well for films, which have a high intrinsic background. In *XDS*, which was originally designed for a multiwire detector, $v_i \propto p_i$ was assumed, which results in a straight summation of background-subtracted counts within the expected profile region, $I = \sum_{i \in D} (c_i - b_i) / \sum_{i \in D} p_i$. This particular simple formula is very satisfactory for the low background typical of these detectors. For the general case, however, better results can be obtained by using $v_i = b_i + I p_i$ for the pixel variances as shown by Otwinowski and implemented in *DENZO* and in the later version of *XDS*. Starting with $v_i = b_i$, the intensity is now found by an iterative process which is terminated if the new intensity estimate becomes negative or does not change within a small tolerance, which is usually reached after three cycles. It can be shown that the solution thus obtained is unique.

11.3.4. Scaling

Usually, many statistically independent observations of symmetry-related reflections are recorded in the rotation images taken from one or several similar crystals of the same compound. The squared structure-factor amplitudes of equivalent reflections should be equal and the idea of scaling is to exploit this *a priori* knowledge to determine a correction factor for each observed intensity. These correction factors compensate to some extent for effects such as radiation damage, absorption, and variations in detector sensitivity and exposure times, as well as variations in size and disorder between different crystals.

The usual methods of scaling split the data into batches of roughly the same size, each covering one or more adjacent rotation images, and then determine a single scaling factor for all reflections in each batch. Neighbouring reflections may then receive quite different corrections if they are assigned to different batches. Since the selection of batch boundaries is to some extent arbitrary, a more continuous correction function would be preferable. This function could be modelled analytically (for example by using spherical harmonics) or empirically, as implemented in *XSCALE* and described below.

For each reflection, observational equations are defined as

$$\psi_{hl\alpha} = (I_{hl} - g_\alpha I_h) / \sigma_{hl}.$$

The subscript h represents the unique reflection indices and l enumerates all symmetry-related reflections to h . By definition, the unique reflection indices have the largest h , then k , then l value occurring in the set of all indices related by symmetry to the original indices, including Friedel mates. Thus, two reflections are symmetry-related if and only if their unique indices are identical. I_h is the unknown 'true' intensity and I_{hl}, σ_{hl} are symmetry-related observed intensities and their standard deviations, respectively. The subscript α denotes the coordinates at which the scaling function g_α should be evaluated. As implemented in *XDS* and *XSCALE*, $\alpha = 1, \dots, 9$ denotes nine positions uniformly distributed in the detector plane at the beginning of data collection, $\alpha = 10, \dots, 18$ the same positions on the detector but after the crystal has been rotated by, say, 5° , and so on. The scaling factors g_α and the estimated intensities I_h are found at the minimum of the function

$$\Psi = \sum_{hl\alpha} w_{hl\alpha} \psi_{hl\alpha}^2.$$

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

The main difference from the method of Fox & Holmes (1966) is the introduction of the weights $w_{hl\alpha}$. These weights depend upon the distance between each reflection hl and the positions α . They are monotonically decreasing functions of this distance, implemented as Gaussians in *XDS* and *XSCALE*. This results in a smoothing of the scaling factors since each reflection contributes to the observational equations in proportion to the weights $w_{hl\alpha}$.

Minimization of Ψ is done iteratively. After each step, the g_α are replaced by $\Delta g_\alpha + g_\alpha$ and rescaled to a mean value of 1. The corrections Δg_α are determined from the normal equations

$$\sum_{\beta} A_{\alpha\beta} \Delta g_{\beta} = b_{\alpha},$$

where

$$A_{\alpha\beta} = \sum_h [\delta_{\alpha\beta} I_h^2 u_{h\alpha} + (r_{h\alpha} v_{h\beta} + v_{h\alpha} r_{h\beta} - v_{h\alpha} v_{h\beta}) / u_h]$$

$$b_{\alpha} = \sum_h I_h r_{h\alpha}$$

$$I_h = v_h / u_h$$

$$r_{h\alpha} = v_{h\alpha} - g_{\alpha} u_{h\alpha} I_h$$

$$u_{h\alpha} = \sum_l w_{hl\alpha} / \sigma_{hl}^2$$

$$v_{h\alpha} = \sum_l w_{hl\alpha} I_{hl} / \sigma_{hl}^2$$

$$u_h = \sum_{\alpha} g_{\alpha}^2 u_{h\alpha}$$

$$v_h = \sum_{\alpha} g_{\alpha} v_{h\alpha}.$$

In case a 'true' intensity I_h is available from a reference data set, the non-diagonal elements are omitted from the sum over h in the normal matrix $A_{\alpha\beta}$. The corrections Δg_{α} are expanded in terms of the eigenvectors of the normal matrix, thereby avoiding shifts along eigenvectors with very small eigenvalues (Diamond, 1966). This filtering method is essential since the normal matrix has zero determinant if no reference data set is available.

11.3.5. Post refinement

The number of fully recorded reflections on each single image rapidly declines for small oscillation ranges and the complete intensities of the partially recorded reflections have to be estimated. This presented a serious obstacle in early structural work on virus crystals, as the crystal had to be replaced after each exposure on account of radiation damage. A solution of this problem, the 'post refinement' technique, was found by Schutt, Winkler and Harrison, and variants of this powerful method have been incorporated into most data-reduction programs [for a detailed discussion, see Harrison *et al.* (1985); Rossmann (1985)]. The method derives complete intensities of reflections only partially recorded on an image from accurate estimates for the fractions of observed intensity, the 'partiality'. The partiality of each reflection can always be calculated as a function of orientation, unit-cell metric, mosaic spread of the crystal and model intensity distributions. Obviously, the accuracy of the estimated full reflection intensity then strongly depends on a precise knowledge of the parameters describing the diffraction experiment. Usually, for many of the partial reflections, symmetry-related fully recorded ones can be found, and the list of such pairs of intensity observations can be used to refine the required parameters by a least-squares procedure. Clearly, this refinement is carried out after all images have been processed, which explains why the procedure is called 'post refinement'.

Adjustments of the diffraction parameters s_{μ} ($\mu = 1, \dots, k$) are determined by minimization of the function E , which is defined as the weighted sum of squared residuals between calculated and observed partial intensities.

$$E = \sum_{hj} w_{hj} (\Delta_{hj})^2$$

$$\Delta_{hj} = R_j(\varphi_{hj}) g_j I_h - I_{hj}$$

$$w_{hj} = 1 / \{ \sigma^2(I_{hj}) + [R_j(\varphi_{hj}) g_j]^2 \sigma^2(I_h) \}.$$

Here, I_{hj} is the intensity recorded on image j of a partial reflection with indices summarized as hj , I_h is the mean of the observed intensities of all fully recorded reflections symmetry-equivalent to hj , g_j is the inverse scaling factor of image j , φ_{hj} is the calculated spindle angle of reflection hj at diffraction and R_j is the computed fraction of total intensity recorded on image j .

Expansion of the residuals Δ_{hj} to first order in the parameter changes δs_{μ} and minimization of $E(\delta s_{\mu})$ leads to the k normal equations

$$\sum_{\mu'=1}^k \left(\sum_{hj} w_{hj} \frac{\partial \Delta_{hj}}{\partial s_{\mu}} \frac{\partial \Delta_{hj}}{\partial s_{\mu'}} \right) \delta s_{\mu'} = - \sum_{hj} w_{hj} \Delta_{hj} \frac{\partial \Delta_{hj}}{\partial s_{\mu}}.$$

Often, the normal matrix is ill-conditioned, since changes in some unit-cell parameters or small rotations of the crystal about the incident X-ray beam do not significantly affect the calculated partiality R_j . To take care of these difficulties, the system of equations is rescaled to yield unit diagonal elements for the normal matrix and the correction vector δs_{μ} is filtered by projection into a subspace defined by the eigenvectors of the normal matrix with sufficiently large eigenvalues (Diamond, 1966).

The parameters are corrected by the filtered δs_{μ} and a new cycle of refinement is started until a minimum of E is reached. The weights, residuals and their gradients are calculated using the current values for s_{μ} and g_j at the beginning of each cycle. The derivatives

$$\frac{\partial \Delta_{hj}}{\partial s_{\mu}} = g_j I_h \left(\frac{\partial R_j}{\partial \varphi_{hj}} \frac{\partial \varphi_{hj}}{\partial s_{\mu}} + \frac{\partial R_j}{\partial \sigma_M} \frac{\partial \sigma_M}{\partial s_{\mu}} + \frac{\partial R_j}{\partial |\zeta_{hj}|} \frac{\partial |\zeta_{hj}|}{\partial s_{\mu}} \right)$$

appearing in the normal equations can be worked out from the definitions given in Sections 11.3.2.2 and 11.3.2.4 (to simplify the following equations, the subscript hj is omitted). The fraction R_j of the total intensity can be expressed in terms of the error function (see Section 11.3.2.4) as

$$R_j = [\operatorname{erf}(z_1) - \operatorname{erf}(z_2)] / 2$$

$$z_1 = |\zeta|(\varphi_0 + j\Delta\varphi - \varphi) / (2)^{1/2} \sigma_M$$

$$z_2 = |\zeta|[\varphi_0 + (j-1)\Delta\varphi - \varphi] / (2)^{1/2} \sigma_M.$$

Using the relation $d \operatorname{erf}(z) / dz = [2/(\pi)^{1/2}] \exp(-z^2)$, the derivatives of R_j are

$$\partial R_j / \partial \varphi = [\exp(-z_2^2) - \exp(-z_1^2)] |\zeta| / [\sigma_M (2\pi)^{1/2}]$$

$$\partial R_j / \partial \sigma_M = [z_2 \exp(-z_2^2) - z_1 \exp(-z_1^2)] / [\sigma_M (\pi)^{1/2}]$$

$$\partial R_j / \partial |\zeta| = [z_1 \exp(-z_1^2) - z_2 \exp(-z_2^2)] / [|\zeta| (\pi)^{1/2}].$$

It remains to work out the derivatives $\partial \varphi / \partial s_{\mu}$, $\partial \sigma_M / \partial s_{\mu}$ and $\partial |\zeta| / \partial s_{\mu}$ (not shown here). As discussed in detail by Greenhough & Helliwell (1982), spectral dispersion and asymmetric beam cross fire lead to some variation of σ_M , which makes it necessary to include additional parameters in the list s_{μ} . The effect of these parameters on the partiality is dealt with easily by the derivatives $\partial \sigma_M / \partial s_{\mu}$.

The refinement scheme described above requires initial scaling factors g_j . With the now improved estimates for the partialities R_j , a new set of scaling factors can be obtained by the method outlined in Section 11.3.4. This alternating procedure of scaling and post-refinement usually converges within three cycles.

The use of error functions for modelling partiality, as implicated by a Gaussian model for describing spot shape, was chosen here for reasons of conceptual simplicity and coherence. This choice is unlikely to alter significantly the results of post-refinement that are based on other functions of similar form [see the discussion by Rossmann (1985)].

11.3.6. Space-group assignment

Identification of the correct space group is not always an easy task and should be postponed for as long as possible. Fortunately, all data processing as implemented in the program *XDS* can be carried out even in the absence of any knowledge of crystal symmetry and cell constants. In this case, a reduced cell is extracted from the observed diffraction pattern and processing of the data images continues to completion as if the crystal were triclinic. Clearly, the reflection indices then refer to the reduced cell and must be reindexed once the space group is known. For all space groups, the required reindexing transformation is linear and involves only whole numbers as shown in Part 9 of IT A. The following description and example are taken from Kabsch (1993).

Space-group assignment is carried out in two steps under control of the crystallographer once integrated intensities of all reflections are available. First, the Bravais lattices that are compatible with the observed reduced cell are identified. In the second step, any of the plausible space groups may be tested and rated according to symmetry R factors and systematic absences of integrated reflection intensities after reindexing. Additional acceptance criteria are obtained from refinement, now using a reduced set of independent parameters describing the conventional unit cell which should not lead to a significant increase of r.m.s. deviations between observed and calculated reflection positions and angles.

11.3.6.1. Determination of the Bravais lattice

The determination of possible Bravais lattices is based upon the concept of the reduced cell whose metric parameters characterize 44 lattice types as described in Part 9 of IT A. A primitive basis $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ of a given lattice is defined there as a reduced cell if it is right-handed and if the components of its metric tensor

$$\begin{array}{lll} A = \mathbf{b}_1 \cdot \mathbf{b}_1, & B = \mathbf{b}_2 \cdot \mathbf{b}_2, & C = \mathbf{b}_3 \cdot \mathbf{b}_3, \\ D = \mathbf{b}_2 \cdot \mathbf{b}_3, & E = \mathbf{b}_1 \cdot \mathbf{b}_3, & F = \mathbf{b}_1 \cdot \mathbf{b}_2 \end{array}$$

satisfy a number of conditions (inequalities). The main conditions state that the basis vectors are the shortest three linear independent lattice vectors with either all acute or all non-acute angles between them. As specified in IT A, each of the 44 lattice types is characterized by additional equality relations among the six components of the reduced-cell metric tensor. As an example, for lattice character 13 (Bravais type *oC*) the components of the metric tensor of the reduced cell must satisfy

$$A = B, \quad B \leq C, \quad D = 0, \quad E = 0, \quad 0 \leq -F \leq A/2.$$

Any primitive triclinic cell describing a given lattice can be converted into a reduced cell. It is well known, however, that the reduced cell thus derived is sensitive to experimental error. Hence, the direct approach of first deriving the correct reduced cell and then finding the lattice type is unstable and may in certain cases even prevent the identification of the correct Bravais lattice.

A suitable solution of the problem has been found that avoids any decision about what the 'true' reduced cell is. The essential

requirements of this procedure are: (a) a database of possible reduced cells and (b) a backward search strategy that finds the best-fitting cell in the database for each lattice type.

The database is derived from a seed cell which strictly satisfies the definitions for a reduced cell. All cells of the same volume as the seed cell whose basis vectors can be linearly expressed in terms of the seed vectors by indices $-1, 0,$ or $+1$ are included in the database. Each unit cell in the database is considered as a potential reduced cell even though some of the defining conditions as given in Part 9 of IT A may be violated. These violations are treated as being due to experimental error.

The backward search strategy starts with the hypothesis that the lattice type is already known and identifies the best-fitting cell in the database of possible reduced cells. Contrary to a forward directed search, it is now always possible to decide which conditions have to be satisfied by the components of the metric tensor of the reduced cell. The total amount by which all these equality and inequality conditions are violated is used as a quality index. This measure is defined below for lattice type 13 *oC* testing a potential reduced cell $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ from the database for agreement. Positive values of the quality index p_{13} indicate that some conditions are not satisfied.

$$p_{13}(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) = |A - B| + \max(0, B - C) + |D| + |E| \\ + \max(0, F) + \max(0, -F - A/2).$$

All potential reduced cells in the database are tested and the smallest value for p_{13} is assigned to lattice type 13. This test is carried out for all 44 possible lattice types using quality indices derived in a similar way from the defining conditions as listed in Part 9 of IT A. For each of the 44 lattice types thus tested, the procedure described here returns the quality index, the conventional cell parameters and a transformation matrix relating original indices with respect to the seed cell to the new indices with respect to the conventional cell. These index-transformation matrices are derived from those given in Table 9.3.1 in IT A.

The results obtained by this method are shown in Table 11.3.6.1 for the example of a 1.5° oscillation data film containing 1313 strong diffraction spots which were located automatically. The space group of the crystal is $C222_1$ and the cell constants are $a = 72.9, b = 100.1, c = 92.6$ Å. The entry for the correct Bravais lattice *oC* with derived cell constants close to the true ones has a low value for its quality index and thus appears as a possible explanation of the observed diffraction pattern.

11.3.6.2. Finding possible space groups

Inspection of the table rating the likelihood of each of the 44 lattice types usually reveals a rather limited set of possible space groups. Furthermore, the absence of parity-changing symmetry operators required for protein crystals restricts the number of possible space groups to 65 instead of 230. Any space group can be tested by repeating only the final steps of data processing. These steps include a comparison of symmetry-related reflection intensities, as well as a refinement of the parameters controlling the diffraction pattern after reindexing the reflections by the appropriate transformation. Low r.m.s. deviations between the observed and refined spot positions, as well as small R factors for symmetry-related reflection intensities, indicate that the constraints imposed by the tentatively chosen space group are satisfied. The space group with highest symmetry compatible with the data is almost certainly correct if the data set is sufficiently complete and redundant, which requires that each symmetry element relates a sufficient number of reflections to one another.

For the example of a 1.5° oscillation data film given above, space-group determination consists of the following steps. Inspection of Table 11.3.6.1 indicates that lattice characters 10, 13, 14 and 34, besides the triclinic characters 31 and 44, are approximately

Table 11.3.6.1. Rating of lattice types implied by a given reduced cell

Lattice type	Quality index	Conventional cell constants (Å, °)						Reindexing transformation
		<i>a</i>	<i>b</i>	<i>c</i>	α	β	γ	
1 <i>cF</i>	999.0	119.3	137.3	119.1	121.1	77.5	122.6	11 $\bar{1}$ 0/1 $\bar{1}$ 10/ $\bar{1}\bar{1}\bar{1}$ 0
2 <i>hR</i>	770.1	74.6	111.7	137.4	103.6	89.1	108.8	1100/ $\bar{1}$ 0 $\bar{1}$ 0/ $\bar{1}\bar{1}$ 10
3 <i>cP</i>	769.7	62.1	63.5	92.9	90.0	90.1	107.2	1000/0100/0010
5 <i>cI</i>	936.0	111.7	74.6	112.6	70.1	53.6	71.2	1010/1100/0110
4 <i>hR</i>	769.5	101.1	111.9	119.1	116.5	89.1	116.4	$\bar{1}\bar{1}$ 00/ $\bar{1}$ 010/ $\bar{1}\bar{1}\bar{1}$ 0
6 <i>tI</i>	999.0	112.6	111.7	74.6	71.2	70.1	53.6	0110/1010/1100
7 <i>tI</i>	999.0	111.7	74.6	112.6	70.1	53.6	71.2	1010/1100/0110
8 <i>oI</i>	999.0	74.6	111.7	112.6	53.6	70.1	71.2	$\bar{1}\bar{1}$ 00/ $\bar{1}$ 0 $\bar{1}$ 0/ $\bar{0}\bar{1}\bar{1}$ 0
9 <i>hR</i>	772.7	62.1	74.6	296.6	90.5	105.8	125.5	1000/ $\bar{1}\bar{1}$ 00/ $\bar{1}\bar{1}$ 30
10 <i>mC</i>	24.0	101.1	74.6	92.9	90.1	90.0	91.3	$\bar{1}\bar{1}$ 00/1100/0010
11 <i>tP</i>	174.8	62.1	63.5	92.9	90.0	90.1	107.2	1000/0100/0010
12 <i>hP</i>	122.8	62.1	63.5	92.9	90.0	90.1	107.2	1000/0100/0010
13 <i>oC</i>	23.8	74.6	101.1	92.9	90.0	90.1	88.7	1100/ $\bar{1}\bar{1}$ 00/0010
15 <i>tI</i>	672.7	62.1	63.5	200.2	77.0	77.6	107.2	1000/0100/1120
16 <i>oF</i>	999.0	74.6	101.1	200.2	90.5	111.8	88.7	$\bar{1}\bar{1}$ 00/1 $\bar{1}$ 00/1120
14 <i>mC</i>	23.4	74.6	101.1	92.9	90.0	90.1	88.7	1100/ $\bar{1}\bar{1}$ 00/0010
17 <i>mC</i>	999.0	101.1	74.6	111.7	71.2	116.4	88.7	$\bar{1}\bar{1}$ 00/ $\bar{1}\bar{1}$ 00/ $\bar{1}$ 0 $\bar{1}$ 0
18 <i>tI</i>	999.0	112.6	119.1	62.1	68.7	99.5	115.4	01 $\bar{1}$ 0/1110/1000
19 <i>oI</i>	999.0	62.1	112.6	119.1	64.6	68.7	80.5	$\bar{1}$ 000/01 $\bar{1}$ 0/ $\bar{1}\bar{1}\bar{1}$ 0
20 <i>mC</i>	746.3	112.6	112.6	62.1	99.5	99.6	111.3	0 $\bar{1}\bar{1}$ 0/0 $\bar{1}$ 10/ $\bar{1}$ 000
21 <i>tP</i>	748.0	63.5	92.9	62.1	90.1	107.2	90.0	0100/0010/1000
22 <i>hP</i>	999.0	63.5	92.9	62.1	90.1	107.2	90.0	0100/0010/1000
23 <i>oC</i>	747.8	112.6	112.6	62.1	80.5	99.6	68.7	0110/0 $\bar{1}$ 10/1000
24 <i>hR</i>	999.0	154.8	112.6	62.1	80.5	80.9	84.3	1210/0 $\bar{1}$ 10/1000
25 <i>mC</i>	746.1	112.6	112.6	62.1	80.5	99.6	68.7	0110/0 $\bar{1}$ 10/1000
26 <i>oF</i>	624.9	62.1	123.9	195.9	86.4	108.4	101.5	1000/ $\bar{1}\bar{2}$ 00/ $\bar{1}$ 0 $\bar{2}$ 0
27 <i>mC</i>	499.7	123.9	62.1	112.6	80.5	119.7	78.5	$\bar{1}\bar{2}$ 00/ $\bar{1}$ 000/01 $\bar{1}$ 0
28 <i>mC</i>	325.0	62.1	195.9	63.5	95.4	107.2	71.6	$\bar{1}$ 000/ $\bar{1}$ 0 $\bar{2}$ 0/0 $\bar{1}$ 00
29 <i>mC</i>	99.8	62.1	123.9	92.9	90.0	90.1	78.5	1000/1200/0010
30 <i>mC</i>	336.4	63.5	196.5	62.1	95.4	107.2	71.1	0 $\bar{1}$ 00/0 $\bar{1}$ 20/ $\bar{1}$ 000
31 <i>aP</i>	0.2	62.1	63.5	92.9	90.0	89.9	72.8	1000/0 $\bar{1}$ 00/00 $\bar{1}$ 0
32 <i>oP</i>	152.0	62.1	63.5	92.9	90.0	90.1	107.2	1000/0100/0010
40 <i>oC</i>	413.0	63.5	196.4	62.1	84.5	107.2	108.9	0 $\bar{1}$ 00/0120/ $\bar{1}$ 000
35 <i>mP</i>	151.8	63.5	62.1	92.9	90.1	90.0	107.2	0 $\bar{1}$ 00/ $\bar{1}$ 000/00 $\bar{1}$ 0
36 <i>oC</i>	400.3	62.1	195.9	63.5	84.6	107.2	108.4	1000/ $\bar{1}$ 0 $\bar{2}$ 0/0100
33 <i>mP</i>	151.2	62.1	63.5	92.9	90.0	90.1	107.2	1000/0100/0010
38 <i>oC</i>	100.1	62.1	123.9	92.9	90.0	90.1	101.5	$\bar{1}$ 000/1200/00 $\bar{1}$ 0
34 <i>mP</i>	1.0	62.1	92.9	63.5	90.0	107.2	90.1	$\bar{1}$ 000/00 $\bar{1}$ 0/0 $\bar{1}$ 00
42 <i>oI</i>	661.3	62.1	63.5	200.2	103.0	102.4	107.2	$\bar{1}$ 000/0 $\bar{1}$ 00/1120
41 <i>mC</i>	412.2	196.4	63.5	62.1	107.2	95.5	71.1	0 $\bar{1}\bar{2}$ 0/0 $\bar{1}$ 00/ $\bar{1}$ 000
37 <i>mC</i>	400.1	195.9	62.1	63.5	107.2	95.4	71.6	1020/1000/0100
39 <i>mC</i>	99.9	123.9	62.1	92.9	90.1	90.0	78.5	$\bar{1}\bar{2}$ 00/ $\bar{1}$ 000/00 $\bar{1}$ 0
43 <i>mI</i>	999.0	74.6	200.2	63.5	103.0	127.3	68.2	1100/1120/0 $\bar{1}$ 00
44 <i>aP</i>	0.0	62.1	63.5	92.9	90.0	90.1	107.2	1000/0100/0010

compatible with the observed diffraction pattern. The highest lattice symmetry is orthorhombic (character 13, Bravais type *oC*), which limits the possible space groups for protein crystals to either $C222_1$ or $C222$. Processing of all films in the data set was completed in space group $P1$ using the cell constants shown for lattice character 44. To test whether the crystal has space-group symmetry $C222$ and conventional cell constants $a = 74.6, b = 101.1, c = 92.9$ Å, the final steps of data processing were repeated after reindexing the

reflections by the transformation $h' = 1 \cdot h + 1 \cdot k + 0 \cdot l + 0$, $k' = -1 \cdot h + 1 \cdot k + 0 \cdot l + 0$, $l' = 0 \cdot h + 0 \cdot k + 1 \cdot l + 0$ as specified for lattice character 13. Note that the transformation also provides a simple tool for correcting the indices if all reflections are misindexed by a constant. The results clearly show that the crystal has space-group symmetry $C222_1$. The presence of the 2_1 axis was deduced from the rather weak intensities observed for reflections of type $00l' = \text{odd}$.

References

11.1

- Bricogne, G. (1986). *Indexing and the Fourier transform*. In *Proceedings of the EEC cooperative workshop on position-sensitive detector software (phase III)*, p. 28. LURE, Paris, 12–19 November.
- Burzlauff, H., Zimmermann, H. & de Wolff, P. M. (1992). *Crystal lattices*. In *International Tables for Crystallography*, Vol. A. *Space-group symmetry*, edited by T. Hahn. Dordrecht: Kluwer Academic Publishers.
- Campbell, J. W. (1997). In *CCP4 Newsletter*, No. 33, edited by M. Winn. Warrington: Daresbury Laboratory.
- Duisenberg, A. J. M. (1992). *Indexing in single-crystal diffractometry with an obstinate list of reflections*. *J. Appl. Cryst.* **25**, 92–96.
- Higashi, T. (1990). *Auto-indexing of oscillation images*. *J. Appl. Cryst.* **23**, 253–257.
- Kabsch, W. (1988). *Automatic indexing of rotation diffraction patterns*. *J. Appl. Cryst.* **21**, 67–71.
- Kabsch, W. (1993). *Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants*. *J. Appl. Cryst.* **26**, 795–800.
- Kim, S. (1989). *Auto-indexing oscillation photographs*. *J. Appl. Cryst.* **22**, 53–60.
- Leslie, A. G. W. (1992). *Recent changes to the MOSFLM package for processing film and image plate data*. In *CCP4 and ESF-EACMB Newsletter on Protein Crystallography*, No. 26. Warrington: Daresbury Laboratory.
- Otwinowski, Z. & Minor, W. (1997). *Processing of X-ray diffraction data collected in oscillation mode*. *Methods Enzymol.* **276**, 307–326.
- Rossmann, M. G. (1979). *Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting*. *J. Appl. Cryst.* **12**, 225–238.
- Rossmann, M. G. & Erickson, J. W. (1983). *Oscillation photography of radiation-sensitive crystals using a synchrotron source*. *J. Appl. Cryst.* **16**, 629–636.
- Sparks, R. A. (1976). *Trends in minicomputer hardware and software. Part I*. In *Crystallographic computing techniques*, edited by F. R. Ahmed, K. Huml & B. Sedláček, pp. 452–467. Copenhagen: Munksgaard.
- Sparks, R. A. (1982). *Data collection with diffractometers*. In *Computational crystallography*, edited by D. Sayre, pp. 1–18. New York: Oxford University Press.
- Steller, I., Bolotovskiy, R. & Rossmann, M. G. (1997). *An algorithm for automatic indexing of oscillation images using Fourier analysis*. *J. Appl. Cryst.* **30**, 1036–1040.
- Strouse, C. E. (1996). Personal communication.
- Tao, Y., Strelkov, S. V., Mesyanzhinov, V. V. & Rossmann, M. G. (1997). *Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain*. *Structure*, **5**, 789–798.
- Vriend, G. & Rossmann, M. G. (1987). *Determination of the orientation of a randomly placed crystal from a single oscillation photograph*. *J. Appl. Cryst.* **20**, 338–343.
- Wonacott, A. J. (1977). *Geometry of the rotation method*. In *The rotation method in crystallography*, edited by U. W. Arndt & A. J. Wonacott, pp. 75–103. Amsterdam: North-Holland Publishing Co.

11.2

- Diamond, R. (1969). *Profile analysis in single crystal diffractometry*. *Acta Cryst.* **A25**, 43–55.
- Ford, G. C. (1974). *Intensity determination by profile fitting applied to precession photographs*. *J. Appl. Cryst.* **7**, 555–564.
- Greenhough, T. J. & Suddath, F. L. (1986). *Oscillation camera data processing. 4. Results and recommendations for the processing of synchrotron radiation data in macromolecular crystallography*. *J. Appl. Cryst.* **19**, 400–409.

- Lehmann, M. S. & Larsen, F. K. (1974). *A method for location of the peaks in step-scan-measured Bragg reflexions*. *Acta Cryst.* **A30**, 580–584.
- Leslie, A. G. W. (1992). *Recent changes to the MOSFLM package for processing film and image plate data*. *CCP4 and ESF-EACMB Newsletter on Protein Crystallography*. Warrington: Daresbury Laboratory.
- Otwinowski, Z. & Minor, W. (1997). *Processing of X-ray diffraction data collected in oscillation mode*. *Methods Enzymol.* **276**, 307–326.
- Rossmann, M. G. (1979). *Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting*. *J. Appl. Cryst.* **12**, 225–238.
- Rossmann, M. G., Leslie, A. G. W., Abdel-Meguid, S. S. & Tsukihara, T. (1979). *Processing and post-refinement of oscillation camera data*. *J. Appl. Cryst.* **12**, 570–581.
- Winkler, F. K., Schutt, C. E. & Harrison, S. C. (1979). *The oscillation method for crystals with very large unit cells*. *Acta Cryst.* **A35**, 901–911.

11.3

- Abramowitz, M. & Stegun, I. A. (1972). *Handbook of mathematical functions*. New York: Dover Publications.
- Bricogne, G. (1986). *Indexing and the Fourier transform*. In *Proceedings of the EEC cooperative workshop on position-sensitive detector software (phase III)*, p. 28. LURE, 12–19 November.
- Diamond, R. (1966). *A mathematical model-building procedure for proteins*. *Acta Cryst.* **21**, 253–266.
- Diamond, R. (1969). *Profile analysis in single crystal diffractometry*. *Acta Cryst.* **A25**, 43–55.
- Dijkstra, E. W. (1976). *A discipline of programming*, pp. 154–167. New Jersey: Prentice-Hall.
- Ford, G. C. (1974). *Intensity determination by profile fitting applied to precession photographs*. *J. Appl. Cryst.* **7**, 555–564.
- Fox, G. C. & Holmes, K. C. (1966). *An alternative method of solving the layer scaling equations of Hamilton, Rollett and Sparks*. *Acta Cryst.* **20**, 886–891.
- Greenhough, T. J. & Helliwell, J. R. (1982). *Oscillation camera data processing: reflecting range and prediction of partiality. I. Conventional X-ray sources*. *J. Appl. Cryst.* **15**, 338–351.
- Harrison, S. C., Winkler, F. K., Schutt, C. E. & Durbin, R. M. (1985). *Oscillation method with large unit cells*. *Methods Enzymol.* **114A**, 211–237.
- Howard, A. (1986). *Autoindexing*. In *Proceedings of the EEC cooperative workshop on position-sensitive detector software (phases I & II)*, pp. 89–94. LURE, 26 May–7 June.
- International Tables for Crystallography* (1995). Vol. A. *Space-group symmetry*, edited by Th. Hahn, pp. 738–749. Dordrecht: Kluwer Academic Publishers.
- Kabsch, W. (1988a). *Automatic indexing of rotation diffraction patterns*. *J. Appl. Cryst.* **21**, 67–71.
- Kabsch, W. (1988b). *Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector*. *J. Appl. Cryst.* **21**, 916–924.
- Kabsch, W. (1993). *Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants*. *J. Appl. Cryst.* **26**, 795–800.
- Otwinowski, Z. (1993). *Oscillation data reduction program*. In *Proceedings of the CCP4 study weekend. Data collection and processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
- Otwinowski, Z. & Minor, W. (1997). *Processing of X-ray diffraction data collected in oscillation mode*. *Methods Enzymol.* **276**, 307–326.
- Pflugrath, J. W. (1997). *Diffraction-data processing for electronic detectors: theory and practice*. *Methods Enzymol.* **276A**, 286–306.
- Rossmann, M. G. (1985). *Determining the intensity of Bragg reflections from oscillation photographs*. *Methods Enzymol.* **114A**, 237–280.

11. DATA PROCESSING

11.3 (cont.)

- Schutt, C. & Winkler, F. K. (1977). *The oscillation method for very large unit cells*. In *The rotation method in crystallography*, edited by U. W. Arndt & A. J. Wonacott, pp. 173–186. Amsterdam, New York, Oxford: North-Holland.
- Steller, I., Bolotovskiy, R. & Rossmann, M. G. (1997). *An algorithm for automatic indexing of oscillation images using Fourier analysis*. *J. Appl. Cryst.* **30**, 1036–1040.
- Wirth, N. (1976). *Algorithms + data structures = programs*, pp. 264–274. New York: Prentice-Hall.

11.4

- Arndt, U. W. & Wonacott, A. J. (1977). *The rotation method in crystallography*. Amsterdam: North Holland.
- Blessing, R. H. (1995). *An empirical correction for absorption anisotropy*. *Acta Cryst.* **A51**, 33–38.
- Blum, M., Metcalf, P., Harrison, S. C. & Wiley, D. C. (1987). *A system for collection and on-line integration of X-ray diffraction data from a multiwire area detector*. *J. Appl. Cryst.* **20**, 235–242.
- Bricogne, G. (1987). *The EEC cooperative programming workshop on position-sensitive detector software*. In *Proceedings of the Daresbury study weekend at Daresbury Laboratory*, 23–24 January, edited by J. R. Helliwell, P. A. Machin and M. Z. Papiz, pp. 120–146. Warrington: Daresbury Laboratory.
- EEC Cooperative Workshop on Position-Sensitive Detector Software (1986). Phase I and II, LURE, Paris, 16 May–7 June; Phase III, LURE, Paris, 12–19 November.
- Evans, P. (1993). *Data reduction: data collection and processing*. In *Proceedings of the CCP4 study weekend. Data collection and processing*, 29–30 January, edited by L. Sawyer, N. Isaac & S. Bailey, pp. 114–123. Warrington: Daresbury Laboratory.
- Evans, P. R. (1987). *Postrefinement of oscillation camera data*. In *Proceedings of the Daresbury study weekend at Daresbury Laboratory*, 23–24 January, edited by J. R. Helliwell, P. A. Machin and M. Z. Papiz, pp. 58–66. Warrington: Daresbury Laboratory.
- Fox, G. C. & Holmes, K. C. (1966). *An alternative method of solving the layer scaling equations of Hamilton, Rollett and Sparks*. *Acta Cryst.* **20**, 886–891.
- Gewirth, D. (1996). *HKL manual*. 5th ed. Yale University, New Haven, USA.
- Greenhough, A. G. W. (1987). *Partials and partiality*. In *Proceedings of the Daresbury study weekend at Daresbury Laboratory*, 23–24 January, edited by J. R. Helliwell, P. A. Machin and M. Z. Papiz, pp. 51–57. Warrington: Daresbury Laboratory.
- Hammersley, A. P. (1998). *The FIT2D home page*. <http://www.ccp14.ac.uk/ccp/web-mirrors/fit2d/computing/expg/subgroups/dataanalysis/FIT2D/index.html>
- Higashi, T. (1990). *Auto-indexing of oscillation images*. *J. Appl. Cryst.* **23**, 253–257.
- Howard, A., Nielsen, C. & Xuong, Ng. H. (1985). *Software for a diffractometer with multiwire area detector*. *Methods Enzymol.* **114**, 452–472.
- Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H. & Salemme, F. R. (1987). *The use of an imaging proportional counter in macromolecular crystallography*. *J. Appl. Cryst.* **20**, 383–387.
- International Tables for Crystallography* (1995). Vol. A. *Space-group symmetry*, edited by Th. Hahn. Dordrecht: Kluwer Academic Publishers.
- Kabsch, W. (1988). *Evaluation of single-crystal X-ray diffraction data from a position sensitive detector*. *J. Appl. Cryst.* **21**, 916–924.
- Kabsch, W. (1993). *Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants*. *J. Appl. Cryst.* **26**, 795–800.
- Katayama, C. (1986). *An analytical function for absorption correction*. *Acta Cryst.* **A42**, 19–23.
- Kim, S. (1989). *Auto-indexing oscillation photographs*. *J. Appl. Cryst.* **22**, 53–60.
- Leslie, A. (1993). *Autoindexing of rotation diffraction images and parameter refinement*. In *Proceedings of the CCP4 study weekend. Data collection and processing*, 29–30 January, edited by L. Sawyer, N. Isaac & S. Bailey, pp. 44–51. Warrington: Daresbury Laboratory.
- Leslie, A. G. W. (1987). *Profile fitting*. In *Proceedings of the Daresbury study weekend at Daresbury Laboratory*, 23–24 January, edited by J. R. Helliwell, P. A. Machin and M. Z. Papiz, pp. 39–50. Warrington: Daresbury Laboratory.
- Leslie, A. G. W. & Tsukihara, T. (1980). *A strategy for collecting isomorphous derivative data with the oscillation method*. *J. Appl. Cryst.* **13**, 304–305.
- Messerschmidt, A. & Pflugrath, J. W. (1987). *Crystal orientation and X-ray pattern prediction routines for area-detector diffraction systems in macromolecular crystallography*. *J. Appl. Cryst.* **20**, 306–315.
- Minor, W. & Otwinowski, Z. (1997). *Advances in accuracy and automation of data collection and processing*. In *Proceedings of the IUCr computing school*, Bellingham, 1996.
- Naday, I., Ross, S., Westbrook, E. M. & Zentai, G. (1998). *Charge-coupled device/fiber optic taper array X-ray detector for protein crystallography*. *Opt. Eng.* **37**, 1235–1244.
- Otwinowski, Z. (1993). *Oscillation data reduction program*. In *Proceedings of the Daresbury CCP4 study weekend. Data reduction and processing*, edited by L. Sawyer, N. Isaacs and S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
- Otwinowski, Z. & Minor, W. (1997). *Processing of X-ray diffraction data collected in oscillation mode*. *Methods Enzymol.* **276**, 307–326.
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1989). *Numerical recipes – the art of scientific computing*. Cambridge University Press.
- Rossmann, M. G. (1979). *Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting*. *J. Appl. Cryst.* **12**, 225–238.
- Rossmann, M. G. (1984). *Synchrotron radiation studies of large proteins and supramolecular structures*. In *Proceedings of the study weekend at Daresbury Laboratory. Biological systems: structure and analysis*, 24–25 March, edited by G. P. Diakun & C. D. Garner, pp. 28–40. Daresbury: Science and Engineering Research Council.
- Rossmann, M. G. & Erickson, J. W. (1983). *Oscillation photography of radiation-sensitive crystals using a synchrotron source*. *J. Appl. Cryst.* **16**, 629–636.
- Rossmann, M. G., Leslie, A. G. W., Abdel-Meguid, S. S. & Tsukihara, T. (1979). *Processing and post-refinement of oscillation camera data*. *J. Appl. Cryst.* **12**, 570–581.
- Sakabe, N. (1991). *X-ray diffraction data collection system for modern protein crystallography with a Weissenberg camera and an imaging plate using synchrotron radiation*. *Nucl. Instrum. Methods A*, **303**, 448–463.
- Schutt, C. E. & Evans, P. R. (1985). *Relative absorption correction for rotation film data*. *Acta Cryst.* **A41**, 568–570.
- Schwarzenbach, D., Abrahams, S. C., Flack, H. D., Gonschorek, W., Hahn, Th., Huml, K., Marsh, R. E., Prince, E., Robertson, B. E., Rollett, J. S. & Wilson, A. J. C. (1989). *Statistical descriptors in crystallography*. *Acta Cryst.* **A45**, 63–75.
- Steller, I., Bolotovskiy, R. & Rossmann, M. G. (1997). *An algorithm for automatic indexing of oscillation images using Fourier analysis*. *J. Appl. Cryst.* **30**, 1036–1040.
- Stuart, D. (1987). *Absorption correction*. In *Proceedings of the Daresbury study weekend at Daresbury Laboratory*, 23–24 January, edited by J. R. Helliwell, P. A. Machin & M. Z. Papiz, pp. 25–38. Warrington: Daresbury Laboratory.
- Stuart, D. & Walker, N. (1979). *An empirical method for correcting rotation-camera data for absorption and decay effects*. *Acta Cryst.* **A35**, 925–933.
- Takusagawa, F. (1987). *A simple method of absorption and decay correction in intensities measured by area-detector X-ray diffractometer*. *J. Appl. Cryst.* **20**, 243–245.
- Tanaka, I., Yao, M., Suzuki, M., Hikichi, K., Matsumoto, T., Kozasa, M. & Katayama, C. (1990). *An automatic diffraction data*