International Tables for Crystallography (2006). Vol. F, Section 11.3.2.2, pp. 218–219.

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 $\{\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3\}$. 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 $\varphi = 0°$, 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 $\varphi = 0°$. $\mathbf{p}_0^*$ 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

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

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

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

which leads to a diffraction spot recorded at detector coordinates

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

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 $\mathbf{S}$,

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

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 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 $\mathrm{d}\varepsilon_1\mathrm{d}\varepsilon_2\mathrm{d}\varepsilon_3$ at $\varepsilon_1, \varepsilon_2, \varepsilon_3$ can be approximated by Gaussian functions:

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

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

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

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 $\lim_{\Delta_\varphi\to 0}Z(\varphi) = \varphi$.

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

references