Integration of macromolecular diffraction data

A. G. W. Leslie

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

Correspondence e-mail: andrew@mrc-lmb.cam.ac.uk

Diffraction intensities can be evaluated by two distinct procedures: summation integration and profile fitting. Equations are derived for evaluating the intensities and their standard errors for both cases, based on Poisson statistics. These equations highlight the importance of the contribution of the X-ray background to the standard error and give an estimate of the improvement which can be achieved by profile fitting. Profile fitting offers additional advantages in allowing estimation of saturated reflections and in dealing with incompletely resolved diffraction spots.

1. Introduction

Data integration refers to the process of obtaining estimates of diffracted intensities (and their standard errors) from the raw images recorded by an X-ray detector. As two-dimensional area detectors are almost universally used to collect macromolecular diffraction data, only this type of detector will be considered in the following analysis.

When collecting data with a two-dimensional area detector, a decision has to be taken about the magnitude of the angular rotation of the crystal during the recording of each image. Two distinct modes of operation are possible: the rotation per image can be comparable to or greater than the angular reflection range of a typical reflection (coarse φ slicing), or it can be much less than the reflection width (fine φ slicing). The latter approach allows the use of three-dimensional profile fitting and, providing the detector is relatively noise-free, will improve the quality of the resulting data by minimizing the contribution of the X-ray background to the total measured intensity. However, there are significant overheads associated with recording, storing and processing the relatively large number of images that are required. Three-dimensional profile fitting is described in the article by Pflugrath (1999) and will not be discussed here.

2. Prerequisites for accurate integration

2.1. Crystal parameters

Only the integration procedure itself will be described in detail in this article. However, in order to obtain the highest quality data possible from a given set of images, there are a number of parameters which need to be determined in advance of or during the integration. The most important of these are the unit-cell parameters, which should be determined to an accuracy of a few parts in a thousand (or better). Post-refinement procedures (Winkler *et al.*, 1979; Rossmann *et al.*, 1979), which make use of the estimated φ centroids of observed spots rather than their detector coordinates, gener-

ally provide more accurate estimates than methods based on the spot positions. This is because spot positions are affected by residual spatial distortions (after applying appropriate corrections) and, additionally, the unit-cell parameters are correlated with the crystal-to-detector distance, which is not always accurately known. For either method, it is necessary to include data from widely separated regions of reciprocal space (ideally φ values 90° apart) in order to determine all unit-cell parameters accurately. This is particularly important for lower symmetry space groups.

The crystal orientation also needs to be known to an accuracy which corresponds to a few percent of the reflection width. For crystals with low mosaicity (*e.g.* 0.1°), this corresponds to a hundredth of a degree or better. Fortunately, it is a feature of post-refinement that the error in determining the orientation is typically a few percent of the reflection width, and so this condition can generally be met. It is important to allow for movement of the crystal by continuously updating the crystal orientation during integration. This is even true when using cryocooled crystals, as the magnetic couplings which attach the pin (holding the crystal) to the goniometer head are not strong enough to prevent small movements, particularly with the high angular rotation rates employed on intense synchrotron beamlines. Non-orthogonality of the incident X-ray beam and the rotation axis (if not allowed for) or an off-centred crystal will also give rise to apparent changes in crystal orientation with spindle rotation.

The crystal mosaicity can be estimated by visual inspection and refined by post-refinement. Refined values are quite reliable when the mosaic spread is less than about 0.5°, but becomes more dependent on the rocking curve model for the high mosaicities which are often associated with frozen crystals. The presence of diffuse scatter, which appears as haloes around the Bragg diffraction spots, presents further difficulties in determining the correct mosaic spread. When processing coarse-sliced images, it is preferable to slightly overestimate the mosaic spread (rather than underestimate it). This will result in an increase in random errors (by adding in the X-ray background from an image on which the spot is not actually present), whereas using too small a value can give systematic errors (by underestimating the number of images on which the spot lies).

2.2. Detector parameters

Detector calibration is essential for high data quality. Both the spatial distortion and the non-uniformity of response of the detector must be known accurately, and it is equally important that these corrections are stable over the time scale of the experiment (and preferably for much longer).

Finally, the crystal-to-detector distance, the detector orientation and the direct-beam position must be refined and continuously updated during integration, using observed spot positions. The crystal-to-detector distance can vary during data collection if the crystal is not exactly centred on the rotation axis, and the direct-beam position can move after a beam refill at a synchrotron. For image-plate detectors with

two (or more) plates, the direct-beam position and detector distance often differ slightly for different plates.

With appropriate care, it is normally possible to predict reflection positions on the detector to an accuracy of 20–30 μm, or a fraction of the pixel size, particularly for highly collimated X-ray beams available at synchrotron sources. This level of accuracy is necessary to minimize possible systematic errors, particularly in the case of profile fitting.

3. Methods of integration

There are two quite distinct procedures available for determining the integrated intensities: summation integration and profile fitting. Summation integration involves simply adding the pixel values for all pixels lying within the area of a spot and then subtracting the estimated background contribution to the same pixels. Profile fitting (Diamond, 1969; Ford, 1974; Rossmann, 1979) assumes that the actual spot shape or profile is known (in two or three dimensions), and the intensity is derived by finding the scale factor which, when applied to the known (or standard) profile, gives the best fit to the observed spot profile. In practice, profile fitting requires two separate steps: determination of the standard profiles and evaluation of the profile-fitting intensities. As will be shown later, profile fitting results in a reduction in the random error associated with weak intensities, but offers no improvement for very high intensities.

4. X-ray background

In the complete absence of X-ray background and detector noise, the integration of the diffraction images becomes very straightforward. Using the geometry of the Ewald sphere construction, it is possible to map every pixel on the detector into reciprocal space and assign to each pixel the indices of the nearest reciprocal-lattice point. The diffracted intensity could then be found by summing the pixel values for all pixels which have been assigned to that particular reciprocal-lattice point. In the absence of background and detector noise, these intensities are as accurate an estimate as it is possible to obtain, and methods like profile fitting offer no advantage. The inclusion of pixels which lie outside the physical extent of the spot on the detector does not compromise the signal-to-noise ratio. In practice, it is extremely rare for the background to be negligible, even for the relatively strong low-resolution diffraction spots.

4.1. The measurement box

X-ray scattering from air, the sample holder and the specimen itself give rise to a general background in the images which has to be subtracted in order to obtain the Bragg intensities. Ideally, the background should be measured for the same pixels used to record the Bragg diffraction spot, but this is not usually practical and the background is determined using pixels immediately adjacent to the spot. In practice, the pixels to be used for the determination of the background

(background pixels) and those to be used for evaluating the intensity (peak pixels) are defined using a 'measurement box'. This is a rectangular box of pixels which is centred on the predicted spot position. Each pixel within the box is classified as being a background or a peak pixel (or neither). This mask can either be defined by the user or the classification can be made automatically by the program. An example of a possible measurement-box definition is given in Fig. 1. The background parameters NRX, NRY and NC can be optimized automatically by maximizing the ratio of the intensity divided by its standard error, in a manner analogous to that described by Lehmann & Larsen (1974). It is generally assumed that the background can be adequately modelled as a plane, and the plane constants are determined using the background pixels. This allows the background to be estimated for the peak pixels, so that the background-corrected intensity can be calculated.

5. Integration by simple summation

5.1. Determination of the best background plane

The background-plane constants a, b and c are determined by minimizing

$$R_1 = \sum_{i=1}^{n} w_i(\rho_i - ap_i - bq_i - c)^2, \qquad (1)$$

where ρ_i is the total number of counts at the pixel with co-ordinates (p_i, q_i) with respect to the centre of the measurement box and the summation is over the n background pixels. w_i is a weight which should ideally be the inverse of the variance of ρ_i. Assuming that the variance is determined by counting statistics, this gives

$$w_i = 1/GE(\rho_i), \qquad (2)$$

where G is the detector gain, which converts pixel counts to equivalent X-ray photons, and $E(\rho_i)$ is the expectation value of the background counts ρ_i. In practice, the variation in
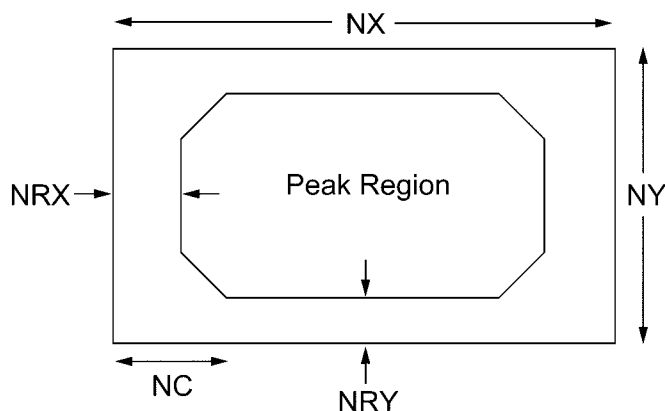

Figure 1
The measurement-box definition used in *MOSFLM*. The measurement box has overall dimensions NXS by NYS pixels (both odd integers). The separation between peak and background pixels is defined by the widths of the background rims (NRX and NRY) and the corner cutoff (NC). The size of the peak region is optimized separately for each of the standard profiles.

background across the measurement box is usually sufficiently small that all weights can be considered to be equal.

This gives the following equations for a, b and c, as given in Rossmann (1979)

$$\begin{pmatrix} \sum p^2 & \sum pq & \sum p \\ \sum pq & \sum q^2 & \sum q \\ \sum p & \sum q & n \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \sum p\rho \\ \sum q\rho \\ \sum \rho \end{pmatrix}, \qquad (3)$$

where all summations are over the n background pixels.

5.1.1. Outlier rejection. It is not unusual for the diffraction pattern to display features other than the Bragg diffraction spots from the crystal of interest. Possible causes are the presence of a satellite crystal or twin component, white-radiation streaks, cosmic rays or zingers. In order to minimize their effect on the determination of the background-plane constants, the following outlier-rejection algorithm is employed.

(i) Determine the background-plane constants using a fraction (say 80%) of the background pixels selecting those with the *lowest* pixel values.

(ii) Evaluate the fit of all background pixels to this plane, rejecting those which deviate by more than three standard errors.

(iii) Re-determine the background plane using all accepted pixels.

(iv) Re-evaluate the fit of all accepted pixels and reject outliers. If any new outliers are found, re-determine the plane constants.

The rationale for using a subset of the pixels with the lowest pixel values in step (i) is that the presence of zingers or cosmic rays or a strongly diffracting satellite crystal can distort the initial calculation of the background plane so much that it becomes difficult to identify the true outliers. Such features will normally only affect a small percentage of the background pixels and will invariably give higher than expected pixel counts. Selecting a subset with the lowest pixel values will facilitate identification of the true outliers. The initial bias in the resulting plane constant c owing to this procedure will be corrected in step (iii). Poisson statistics are used to evaluate the standard errors used in outlier rejection, and the standard error used in step (ii) is increased to allow for the choice of background pixels in step (i).

5.2. Evaluating the integrated intensity and standard error

The summation integration intensity I_s is given by

$$I_s = \sum_{i=1}^{m} (\rho_i - ap_i - bq_i - c), \qquad (4)$$

where the summation is over the m pixels in the peak region of the measurement box. If the peak region has mm symmetry, this simplifies to

$$I_s = \sum_{i=1}^{m} (\rho_i - c). \qquad (5)$$

To evaluate the standard error, this can be written

$$I_s = \sum_{i=1}^{m} \rho_i - (m/n) \sum_{j=1}^{n} \rho_j, \qquad (6)$$

where the second summation is over the n background pixels. The variance in I_s is

$$\sigma_{I_s}^2 = \sum_{i=1}^{m} \sigma_i^2 + (m/n)^2 \sum_{j=1}^{n} \sigma_j^2. \qquad (7)$$

From Poisson statistics, this becomes

$$\sigma_{I_s}^2 = \sum_{i=1}^{m} G\rho_i + (m/n)^2 \sum_{j=1}^{n} G\rho_j \qquad (8)$$

$$= G\left[I_s + I_{bg} + (m/n)(m/n) \sum_{j=1}^{n} \rho_j \right], \qquad (9)$$

where I_{bg} is the background summed over all peak pixels. We can also write

$$I_{bg} \simeq (m/n) \sum_{j=1}^{n} \rho_j \qquad (10)$$

(this is only strictly true if the background region has mm symmetry). Then,

$$\sigma_{I_s}^2 = G[I_s + I_{bg} + (m/n)I_{bg}]. \qquad (11)$$

This expression shows the importance of the background (I_{bg}) in determining the standard error in the intensity. For weak reflections, the Bragg intensity (I_s) is often much smaller than the background (I_{bg}) and the error in the intensity is determined entirely by the background contribution.

5.3. The effect of instrument or detector errors

Standard error estimates calculated using (11) are generally in quite good agreement with observed differences between the intensities of symmetry-related reflections for weak or medium intensities. This is particularly true if other sources of systematic error are minimized by measuring the *same* reflections five or more times by taking multiple exposures of the same small oscillation range and then processing the data in space group $P1$. However, even in this case, the agreement between strong intensities is significantly worse than that predicted using (11). This is consistent with the observation that it is very unusual to obtain merging R factors lower than 1%, even for very strong reflections where Poisson statistics would suggest merging R factors should be in the range 0.2–0.3%.

An experiment in which a diffraction spot recorded on photographic film was scanned many times on an optical microdensitometer showed that the r.m.s. variation in individual pixel values between the scans was greatest for those pixels immediately surrounding the centre of the spot, where the gradient of the optical density was greatest. One explanation for this observation is that these optical densities will be most sensitive to small errors in positioning the reading head, owing to vibration or other mechanical defects. A simple model for the instrumental contribution to the standard error of the spot intensity is obtained by introducing an additional term for each pixel in the spot peak,

$$\sigma_{ins} = K(\delta\rho/\delta x), \qquad (12)$$

where $\delta\rho/\delta x$ is the average gradient and K is a proportionality constant. Taking a triangular-shaped reflection profile, the gradient and integrated intensity are related by the equation

$$I_s = (1/12)(x^3 + 3x^2 + 5x + 3)(\delta\rho/\delta x), \qquad (13)$$

where x is the half-width of the reflection (in pixels). Writing

$$A = (1/12)(x^3 + 3x^2 + 5x + 3), \qquad (14)$$

this gives

$$\sigma_{ins} = (K/A)I_s, \qquad (15)$$

where the factor A allows for differences in spot size and K is ideally a constant for a given instrument.

The total variance in the integrated intensity is then

$$\sigma_{tot}^2 = \sigma_{I_s}^2 + m\sigma_{ins}^2 \qquad (16)$$

$$= G[I_s + I_{bg} + (m/n)I_{bg}] + m(K/A)^2 I_s^2. \qquad (17)$$

A value for K can be determined by comparing the goodness-of-fit of the standard profiles to individual reflection profiles (of fully recorded reflections) with that calculated from combined Poisson statistics and the instrument-error term. Standard errors estimated using (17) give much more realistic estimates than those based on (11), even for data collected with CCD detectors, where the physical model for the source of the error is clearly not appropriate.

6. Integration by profile fitting

Providing the background and peak regions are correctly defined, summation integration provides a method for evaluating integrated intensities which is both robust and free from systematic error. For weak reflections, however, many of the pixels in the peak region will contain very little signal (Bragg intensity), but will contribute significantly to the noise because of the Poissonian variation in the background [as shown by the I_{bg} term in (11)]. Profile fitting provides a means of improving the signal-to-noise ratio for this class of reflection (but will provide no improvement for reflections where the background level is negligible).

6.1. Forming the standard profiles

In order to apply profile-fitting methods, the first requirement is to derive a 'standard' profile which accurately represents the true reflection profile. Although analytical functions can be used, it is difficult to define a simple function which will cope adequately with the wide variation in spot shapes which can arise in practice. Most programs therefore rely on an empirical profile derived by summing many different spots. The optimum profile is that which provides the best fit to all the contributing reflections, *i.e.* that which minimizes

$$R_2 = \sum_{h} w_j(h)[K_h P_j - \rho_j(h)_{corr}]^2, \qquad (18)$$

where P_j is the profile value for the jth pixel, $\rho_j(h)_{corr}$ is the observed background-corrected counts at that pixel for

reflection h, K_h is a scale factor and $w_j(h)$ is a weight for the jth pixel of reflection h. The summation extends over all reflections contributing to the profile. The weight $w_j(h)$ is given by

$$w_j(h) = 1/\sigma_{hj}^2 \tag{19}$$

and, from Poisson statistics, the expectation value of the counts at pixel j is given by

$$\sigma_{hj}^2 = K_h P_j + (a_h p_j + b_h q_j + c_h). \tag{20}$$

After Rossmann (1979), the summation-integration intensity $I_s(h)$ can be used to derive a value for K_h,

$$I_s(h) = K_h \sum_{j=1}^m P_j. \tag{21}$$

In (20) and (21), as the profile values P_j are not yet determined, a preliminary profile derived, for example, from simple summation of strong reflections used in the detector-parameter refinement can be used, which will give acceptable weights for use in (18).

This method of deriving the standard profile is only appropriate for fully recorded reflections. However, in many cases there will be very few or no fully recorded reflections on each image. In such cases, the profile is determined by simply adding together the background-corrected pixel counts from all contributing reflections. In the program *MOSFLM* (Leslie, 1992), the profiles are determined using reflections on, typically, ten or more successive images, so that partials will be summed to give the correct fully recorded profile for the majority of the contributing reflections. Tests carried out using standard profiles derived using only fully recorded reflections and (18), or using both fully recorded and partially recorded reflections and simple summation, give data of the same quality as judged by the merging statistics.

The reflection profile changes across the face of the detector, owing to obliquity of incidence, changes in the projected diffracting volume and geometric factors. In the *MOSFLM* program, this variation is accommodated by determining several standard profiles (typically 9 or 25) for different regions of the detector. When evaluating the profile-fitted intensity for a given reflection, a weighted sum of the nearest standard profiles is calculated to provide the best estimate of the true profile at that position on the detector. For the central regions of the detector, there will be four contributing profiles, while at the edges there will be between one and three. The weights assigned to each profile vary linearly with the distance from the reflection to the centres of the regions used in determining the standard profiles. An alternative procedure, used in *DENZO* (Otwinowski & Minor, 1997), is to evaluate a new profile for each reflection based on spots lying within a pre-specified radius.

6.2. Evaluation of the profile-fitted intensity

Given an appropriate standard profile, the reflection intensity for fully recorded reflections is evaluated by determining the scale factor K and background-plane constants a, b and c which minimize

$$R_3 = \sum w_i (KP_i + ap_i + bq_i + c - \rho_i)^2, \tag{22}$$

where the summation is over all valid pixels in the measurement box. As before,

$$w_i = 1/\sigma_i^2 \tag{23}$$

and the expectation value of the counts at pixel i is given by

$$\sigma_i^2 = ap_i + bq_i + c + JP_i. \tag{24}$$

In order to calculate the weights, the background plane constants and summation-integration intensity I_s are evaluated as described in §5, at the same time identifying any outliers in the background. The summation-integration intensity is used to evaluate the scale factor J in (24) using

$$I_s = J \sum_i P_i. \tag{25}$$

In (22), the summation is over all valid pixels within the measurement box. This excludes pixels which are overlapped by neighbouring spots (if any) and any outliers identified in the background region.

Minimizing R_3 with respect to K, a, b and c leads to four linear equations from which K, a, b and c can be determined:

$$\begin{pmatrix} \sum wP^2 & \sum wpP & \sum wqP & \sum wP \\ \sum wpP & \sum wp^2 & \sum wpq & \sum wp \\ \sum wqP & \sum wpq & \sum wq^2 & \sum wq \\ \sum wP & \sum wp & \sum wq & \sum w \end{pmatrix} \begin{pmatrix} K \\ a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \sum wP\rho \\ \sum wp\rho \\ \sum wq\rho \\ \sum wP \end{pmatrix} \tag{26}$$

and the profile-fitted intensity I_p is then given by

$$I_p = K \sum_i P_i. \tag{27}$$

The standard error in the profile-fitted intensity is given by

$$\sigma_{I_p}^2 = \sigma_K^2 \left(\sum_i P_i \right)^2 \tag{28}$$

$$= \frac{\sum_i^N w_i \Delta_i^2}{N - 4} A_{KK}^{-1} \left(\sum_i P_i \right)^2, \tag{29}$$

where

$$\Delta_i = KP_i + ap_i + bq_i + c - \rho_i, \tag{30}$$

N is the number of pixels in the summation and A_{KK}^{-1} is the diagonal element for the scale factor K of the inverse normal matrix (used to minimize R_3).

In the case of partially recorded reflections, it is no longer valid to fit the sum of the scaled standard profile and a background plane to all pixels in the measurement box. Partially recorded reflections can have a profile which differs significantly from the standard profile, with the result that the background plane constants take on physically unreasonable values in an attempt to compensate for this difference. Therefore, for partially recorded reflections, the summation in (22) is restricted to pixels in the peak region of the measurement box. Minimizing R_3 with respect to the scale factor K then gives

$$I_p = K \sum P_i \qquad (31)$$

$$= \left(\sum w_i P_i \rho_i - a \sum w_i P_i p_i - b \sum w_i P_i q_i - c \sum w_i P_i\right)\left(\sum P_i / \sum w_i P_i^2\right), \qquad (32)$$

where all summations are over the peak region only.

It is not possible to derive a standard error for partially recorded reflections based on the fit of the scaled standard profile (because partially recorded reflections have a different spot profile). For these reflections, the standard error can be calculated using (17).

6.3. Modifications for very close spots

In order to apply (22), it is necessary to exclude all pixels in the measurement box which are overlapped by a neighbouring spot. This applies not only to the pixels of the reflection being integrated, but also to the pixels of all the reflections used to form the standard profile. Consequently, a pixel should be excluded even if it is only overlapped by a neighbouring spot for one of the reflections used in forming the standard profile. When processing data from large unit cells, this can lead to a very high percentage of the background pixels being rejected and, therefore, a poor determination of the background plane parameters. In these circumstances, the background plane is determined using only background pixels and excluding only those pixels which are overlapped by neighbours for the reflection actually being integrated. The profile-fitted intensity for both fully recorded and partially recorded reflections is then evaluated in the way described for partially recorded reflections in the previous section, with the summation in (32) extending only over peak pixels. The standard error in the intensity for partially recorded reflections is derived from (17) as before. For fully recorded reflections, the standard error has two components; the first is based on the fit of the scaled standard profile to the reflection profile and the second on the contribution from the background:

$$\sigma_I^2 = \sigma_{\text{prof}}^2 + \sigma_{\text{bg}}^2 \qquad (33)$$

$$= \frac{\sum_{i=1}^{m} w_i \Delta_i^2 \left(\sum_{i=1}^{m} P_i\right)^2}{(m-1) \sum_{i=1}^{m} w_i P_i^2} + (m/n) \sum_{i=1}^{n} (\rho_i - ap_i - bq_i - c)^2, \qquad (34)$$

where m and n are the number of pixels in the peak and background, respectively.

6.4. Profile fitting very strong reflections

For very strong reflections, the background level is very small, and (32) reduces to

$$I_p \simeq \sum w_i P_i \rho_i \left(\sum P_i / \sum w_i P_i^2\right); \qquad (35)$$

the weights are given by

$$w_i \simeq 1/JP_i. \qquad (36)$$

Substituting for w_i in (35) gives

$$I_p \simeq \sum \rho_i. \qquad (37)$$

As pointed out by Otwinowski (personal communication), this shows that for correctly weighted profile fitting, the profile-fitted intensity reduces to the summation-integration intensity for very strong intensities.

6.5. Profile fitting very weak reflections

For very weak reflections, all pixels will have very similar counts and, therefore, all the weights will be the same. For simplicity, consider the case where the profile fit is evaluated only for the peak pixels; (32) then reduces to

$$I_p = \sum p_i (\rho_i - ap_i - bq_i - c)\left(\sum P_i / \sum P_i^2\right). \qquad (38)$$

The last term in this equation depends only on the shape of the standard profile. This shows that the intensity is a weighted sum of the individual background-corrected pixel counts (rather than a simple unweighted sum, as is the case for summation integration). Because the values of P_i are a maximum in the centre of the spot, this will place a higher weight on those pixels where the contribution of the Bragg diffraction is greatest and a very low weight on the peripheral pixels where the Bragg diffraction is weakest. In this way, profile fitting improves the signal-to-noise ratio without the risk of introducing any systematic error which may result from simply reducing the size of the peak region for weak spots.

6.6. Improvement provided by profile fitting weak reflections

For very weak reflections, where all the weights w_i are approximately the same, the variance in I_p using (38) is given by

$$\sigma_{I_p}^2 = \sum \text{Var}(\rho_i - ap_i - bq_i - c)P_i^2 \left(\sum P_i / \sum P_i^2\right)^2. \qquad (39)$$

Assuring a flat background and very weak intensity, from Poisson statistics

$$\text{Var}(\rho_i - ap_i - bq_i - c) \simeq G\rho_i \qquad (40)$$

and, as ρ_i has approximately the same value (ρ) for all pixels,

$$\sigma_{I_p}^2 = G\rho \sum P_i^2 \left(\sum P_i / \sum P_i^2\right)^2 \qquad (41)$$

$$= G\rho[(\sum P_i)^2] / \sum P_i^2. \qquad (42)$$

The variance in the summation-integration intensity is simply

$$\sigma_{I_s}^2 = Gm\rho. \qquad (43)$$

The ratio of the variances is thus

$$\sigma_{I_s}^2 / \sigma_{I_p}^2 = m \sum P_i^2 / (\sum P_i)^2. \qquad (44)$$

For a typical spot profile, the right-hand side (which depends only on the shape of the standard profile) has a value of 2, showing that profile fitting can reduce the standard error in the integrated intensity by a factor of $2^{1/2}$.

6.7. Other benefits of profile fitting

6.7.1. Incompletely resolved spots. If adjacent spots are not fully resolved, there will be a systematic error in the integrated intensity which will be largest for weak spots which are adja-

cent to very strong spots. However, the profile-fitted intensity will be affected less than the summation-integration intensity because the peripheral pixels (where the influence of neighbouring spots is greatest) are down-weighted relative to the central pixels (where the neighbours will have least influence).

Further steps can be taken to minimize the errors caused by overlapping spots. Firstly, when forming the standard profiles, reflections are only included if they are significantly stronger than their nearest neighbours. This will minimize the errors in the standard profiles. Secondly, when evaluating the profile-fitted intensity of a particular reflection, pixels can be omitted if they are adjacent to a pixel which is part of a neighbouring spot (rather than having to be part of that spot).

A more satisfactory approach is to deconvolute spatially overlapping spots as described, for example, by Bourgeois *et al.* (1998).

6.7.2. Elimination of peak pixel outliers. In the same way that outliers in the background region can be identified and rejected (see §5.1.1), it is possible, in principle, to identify outliers in the peak region of fully recorded reflections as those pixels whose deviation from the scaled standard profile is significantly greater than that expected from counting statistics. This approach works well if the feature which gives rise to the outliers affects only a small fraction of the peak pixels and gives rise to large deviations; this is the case for some zingers, dead pixels and for diffraction from small ice crystals when collecting data from cryo-cooled samples.

Another source of outliers is the encroachment of a strong neighbouring spot into the peak region, as discussed in the previous section. When dealing with peripheral pixels, the outlier test can be applied to both fully recorded and partially recorded reflections, but a high σ cutoff (*e.g.* 10–20) must be used to avoid rejecting pixels which do not fit the profile simply because it is a partially recorded spot.

6.7.3. Estimation of overloaded reflections. Because of the limited dynamic range of current detectors, it is common for many low-resolution spots to contain saturated pixels. Providing the saturation level of the detector is known, such pixels can simply be excluded from the profile fitting, allowing a reasonable estimate of the true intensity (except when the majority of the pixels are saturated). A knowledge of the strong intensities is essential for structure solution based on molecular-replacement techniques, and so this is a very useful additional feature of profile fitting.

6.8. Profile fitting partially recorded reflections

Greenhough & Suddath (1986) have shown that when profile fitting is applied to partially recorded reflections this leads to a systematic error in the individual intensities, but there is no systematic error in the total summed intensity. Although their analysis is strictly only applicable to the case of unweighted profile fitting, experience has shown that even when using weighted profile fitting there is no evidence of systematic errors in the summed profile-fitted intensities of partially recorded reflections. This is particularly important, as many data sets collected from frozen crystals have few, if any, fully recorded reflections.

6.9. Systematic errors in profile-fitted intensities

The fundamental assumption in profile fitting is that the standard profiles accurately reflect the true profile of the reflection being integrated. Errors in the standard profile will result in systematic errors in the profile-fitted intensities. While these errors will often be small compared with the random (Poissonian) error for weak reflections, this is not necessarily the case for strong reflections, as the systematic error is typically a small percentage of the total intensity. Because the standard profiles are derived from the summation of many contributing reflections, small positional errors in spot prediction will lead to a broadening of the standard profile relative to the profile of an individual spot. The same broadening can occur because of the finite sampling interval in the image, which means that a predicted spot position can lie up to half a pixel away from the centre of the measurement box. This error can be minimized by interpolating the pixel values in the image onto a grid which is centred exactly on the predicted position, but the interpolation step itself will inevitably distort the reflection profile. In spite of these difficulties, providing adequate care is taken to determine the crystal and detector parameters accurately (as mentioned in §2) so that the spot positions are predicted to within a small fraction of the overall spot width, then there is no suggestion (from merging statistics at least) for significant systematic error even in the stronger intensities.

References

Bourgeois, D., Nurizzo, D., Kahn, R. & Cambillau, C. (1998). *J. Appl. Cryst.* **31**, 22–35.
Diamond, R. (1969). *Acta Cryst.* A**25**, 43–55.
Ford, G. C. (1974). *J. Appl. Cryst.* **7**, 555–564.
Greenhough, T. J. & Suddath, F. L. (1986). *J. Appl. Cryst.* **19**, 400–409.
Lehmann, M. S. & Larsen, F. K. (1974). *Acta Cryst.* A**30**, 580–584.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACMB. Newslett. Protein Crystallogr.* **26**.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pflugrath, J. (1999). *Acta Cryst.* D**55**, 1718–1725.
Rossmann, M. G. (1979). *J. Appl. Cryst.* **12**, 225–238.
Rossmann, M. G., Leslie, A. G. W., Abdel-Meguid, S. S. & Tsukihara, T. (1979). *J. Appl. Cryst.* **12**, 570–581.
Winkler, F. K., Schutt, C. E. & Harrison, S. C. (1979). *Acta Cryst.* A**35**, 901–911.