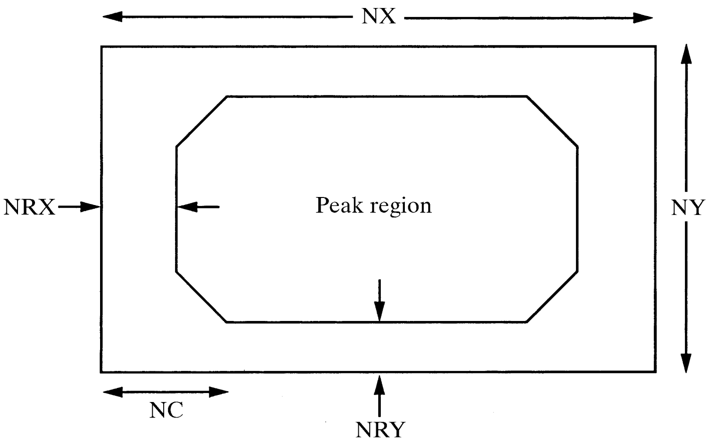

Fig. 11.2.4.1. The measurement-box definition used in *MOSFLM*. The measurement box has overall dimensions of NX by NY pixels (both odd integers). The separation between peak and background pixels is defined by the widths of the background rims (NRX and NRY) and the corner cutoff (NC). The size of the peak region is optimized separately for each of the standard profiles.

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 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. 11.2.4.1. The background parameters NRX, NRY and NC can be optimized automatically by maximizing the ratio of the intensity divided by its standard deviation, 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.

## 11.2.5. Integration by simple summation

### 11.2.5.1. *Determination of the best background plane*

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

$$R_1 = \sum_{i=1}^{n} w_i(\rho_i - ap_i - bq_i - c)^2, \qquad (11.2.5.1)$$

where $\rho_i$ is the total counts at the pixel with coordinates $(p_i q_i)$ 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 $\rho_i$. Assuming that the variance is determined by counting statistics, this gives

$$w_i = 1/GE(\rho_i), \qquad (11.2.5.2)$$

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

$$\begin{pmatrix} \sum p^2 & \sum pq & \sum p \\ \sum pq & \sum q^2 & \sum q \\ \sum p & \sum q & n \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \sum p\rho \\ \sum q\rho \\ \sum \rho \end{pmatrix}, \qquad (11.2.5.3)$$

where all summations are over the $n$ background pixels.

#### 11.2.5.1.1. *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:

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

(2) Evaluate the fit of all background pixels to this plane, rejecting those that deviate by more than three standard deviations.

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

(4) 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 (1) 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 $c$ due to this procedure will be corrected in step (3). Poisson statistics are used to evaluate the standard deviations used in outlier rejection, and the standard deviation used in step (2) is increased to allow for the choice of background pixels in step (1).

### 11.2.5.2. *Evaluating the integrated intensity and standard deviation*

The summation integration intensity $I_s$ is given by

$$I_s = \sum_{i=1}^{m} (\rho_i - ap_i - bq_i - c), \qquad (11.2.5.4)$$

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

$$I_s = \sum_{i=1}^{m} (\rho_i - c). \qquad (11.2.5.5)$$

To evaluate the standard deviation, this can be written as

$$I_s = \sum_{i=1}^{m} \rho_i - (m/n)\sum_{j=1}^{n} \rho_j, \qquad (11.2.5.6)$$

where the second summation is over the $n$ background pixels.

The variance in $I_s$ is

$$\sigma_{I_s}^2 = \sum_{i=1}^{m} \sigma_i^2 + (m/n)^2 \sum_{j=1}^{n} \sigma_j^2. \qquad (11.2.5.7)$$

From Poisson statistics this becomes

$$\sigma^2_{I_s} = \sum_{i=1}^{m} G\rho_i + (m/n)^2 \sum_{j=1}^{n} G\rho_j \qquad (11.2.5.8)$$

$$= G\left[I_s + I_{\rm bg} + (m/n)(m/n)\sum_{j=1}^{n}\rho_j\right], \qquad (11.2.5.9)$$

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

$$I_{\rm bg} \simeq (m/n)\sum_{j=1}^{n}\rho_j \qquad (11.2.5.10)$$

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

$$\sigma^2_{I_s} = G\big[I_s + I_{\rm bg} + (m/n)I_{\rm bg}\big]. \qquad (11.2.5.11)$$

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

### 11.2.5.3. *The effect of instrument or detector errors*

Standard-deviation estimates calculated using (11.2.5.11) 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 doing multiple exposures of the same small oscillation range and then processing the data in space group $P1$. However, even in this latter case, the agreement between strong intensities is significantly worse than that predicted using equation (11.2.5.11). This is consistent with the observation that it is very unusual to obtain merging $R$ factors lower than 0.01, even for very strong reflections where Poisson statistics would suggest merging $R$ factors should be in the range 0.002–0.003.

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, due to vibration or mechanical defects. A simple model for the instrumental contribution to the standard deviation of the spot intensity is obtained by introducing an additional term for each pixel in the spot peak:

$$\sigma_{\rm ins} = K\frac{\delta\rho}{\delta x}, \qquad (11.2.5.12)$$

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

$$I_s = \frac{1}{12}\left(x^3 + 3x^2 + 5x + 3\right)\frac{\delta\rho}{\delta x}, \qquad (11.2.5.13)$$

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

Writing

$$A = \frac{1}{12}\left(x^3 + 3x^2 + 5x + 3\right) \qquad (11.2.5.14)$$

gives

$$\sigma_{\rm ins} = (K/A)I_s, \qquad (11.2.5.15)$$

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

$$\sigma^2_{\rm tot} = \sigma^2_{I_s} + m\sigma^2_{\rm ins} \qquad (11.2.5.16)$$

$$= G\big[I_s + I_{\rm bg} + (m/n)I_{\rm bg}\big] + m(K/A)^2 I_s^2. \qquad (11.2.5.17)$$

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 deviations estimated using (11.2.5.17) give much more realistic estimates than those based on (11.2.5.11), even for data collected with charge-coupled-device (CCD) detectors where the physical model for the source of the error is clearly not appropriate.

## 11.2.6. Integration by profile fitting

Providing the background and peak regions are correctly defined, summation integration provides a method for evaluating integrated intensities that 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 $I_{\rm bg}$ term in equation (11.2.5.11)]. 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).

### 11.2.6.1. *Forming the standard profiles*

In order to apply profile-fitting methods, the first requirement is to derive a 'standard' profile that accurately represents the true reflection profile. Although analytical functions can be used, it is difficult to define a simple function that will cope adequately with the wide variation in spot shapes that 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

$$R_2 = \sum_h w_j(h)\big[K_h P_j - \rho_j(h)_{\rm corr}\big]^2, \qquad (11.2.6.1)$$

where $P_j$ is the profile value for the $j$th pixel, $\rho_j(h)_{\rm corr}$ is the observed background-corrected count at that pixel for reflection $h$, $K_h$ is a scale factor and $w_j(h)$ is a weight for the $j$th pixel of reflection $h$. The summation extends over all reflections contributing to the profile. The weight is given by

$$w_j(h) = 1/\sigma^2_{hj}, \qquad (11.2.6.2)$$

and from Poisson statistics $\sigma^2_{hj}$ is the expectation value of the counts at pixel $j$, and is given by

$$\sigma^2_{hj} = K_h P_j + \left(a_h p_j + b_h q_j + c_h\right). \qquad (11.2.6.3)$$

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

$$I_s(h) = K_h \sum_{j=1}^{m} P_j. \qquad (11.2.6.4)$$

In equations (11.2.6.3) and (11.2.6.4), 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 equation (11.2.6.1).