9. MONOCHROMATIC DATA COLLECTION

images (*a–d*), hence at high resolution, close to the edge of the detector window, the lunes overlap on images (*e–f*). The reflections, however, do not overlap, as the crystal orientation is diagonal; the lunes are sparsely populated, with large separation between adjacent spots, so the reflections on successive lunes fit between one another. It should be noted that the density of reflections in different regions of the reciprocal lattice is constant, and that the total number of reflections recorded on an image depends only on the rotation range, not on the crystal orientation.

The zero-layer lune containing reflections with indices *hk*0 is especially evident on exposures (*c–d*) directly above the centre of the image. With such a lune close to the centre, the reciprocal lattice shows minimal distortion owing to its projection onto the detector plane, and the lune appears as a 'pseudoprecession' pattern. The systematic absence of every second reflection, with odd index, along the *h*00 and 0*k*0 lines indicates the presence of twofold screw axes of symmetry along the crystal axes *a* and *b*. Images (*e–f*), 90° away, have the *hhl* lune at the centre and, although it is less well separated from higher lunes, the presence of a fourfold screw axis along **c** is confirmed by the presence of only every fourth reflection on the 00*l* line. This allows the identification of the space group as $P4_12_12$ or its enantiomorph, $P4_32_12$. In general, the positions of the reflections define only the Bravais lattice, and it is symmetry of the intensity pattern which reflects the point group. Thus, further confirmation that the symmetry belongs to point group *P*422 rather than *P*4 comes from the symmetric relation of the intensity distribution on either side of each lune in images (*a–d*). This is equivalent to the earlier use of precession photography for space-group elucidation.

Close inspection shows that the reflections at the edges of the lune are also present on the adjacent image. The rotation range was 1.5°, and the mosaicity was estimated at 0.5°, and thus about one-sixth of the reflections are partially recorded at each edge of the lune, giving one-third partially recorded terms in total. The lack of sharpness at the edge of the lunes confirms a substantial level of mosaicity.

9.1.11. Rotation method: qualitative factors

9.1.11.1. *Inspection of reflection profiles*

Reflection profiles should be checked on the first recorded images. Very often a quick inspection of the profiles can disqualify a bad crystal without further loss of time. The profiles should have a single maximum and smooth shoulders. If the crystal shape is irregular, it may be reflected in the spot profile. Profiles should not have double maxima or be substantially elongated or smeared out, which usually arises from crystal splitting. The profiles should certainly be inspected if initial autoindexing of the diffraction pattern is unsuccessful.

Even if the spot profiles appear to be regular on the first image, it is good practice to inspect a second image at a substantially different *φ* rotation angle, preferably 90° away, since crystal splitting may have a similar effect on the appearance of the lunes and profiles as does high mosaicity on a single image (Section 9.1.6.3). High mosaicity and splitting (often incorrectly referred to as twinning) must not be confused. If two parts of a split crystal are slightly rotated with respect to one another around a certain axis, the diffraction patterns will look different depending on the orientation. When such an axis is perpendicular to the detector plane, the spots will be doubled or smeared out. When the axis is parallel to the detector plane, the profiles resulting from the two parts of the crystal will overlap almost perfectly, but the lunes will be broadened, similar to the effect of high mosaicity.

After indexing the diffraction pattern, the integration profiles should be matched with the size and shape of the diffraction spots.

The spots should not extend into the area defined as background. Selection of integration profiles that are too small will lead to incorrect integration of intensities. In contrast, if the profile areas are too large then the standard uncertainties will be wrongly estimated.

9.1.11.2. *Exposure time*

According to the principles of counting statistics, the longer the exposure, the better the signal in the data. The standard uncertainty of the measurement is equal to the square root of the number of counts, and the signal-to-noise ratio increases with the accumulated counts. In practice there are limitations to this rule.

The dynamic range and saturation limit of the detector is one limiting factor. It may be impossible to measure adequately the strongest as well as the weakest reflection simultaneously, since their intensities differ by several orders of magnitude. If the exposure time is long enough to record the weakest intensities, then in general at low resolution the most intense reflections may saturate some pixels within their profile on the detector. Such reflections are termed 'overloads' and this problem will be addressed in Section 9.1.11.3.

Exposure time can be limited by the total time available for the experiment. This is often a particularly acute problem for synchrotron-data collection, with high oversubscription of beam-lines. The decisions concerning exposure time depend on the expected application of the data, since different applications have different requirements, as addressed in Section 9.1.13. Within the given time constraints, the first priority should be data complete-ness, even at the expense of underexposure. In this context it is useful to recall that to increase the statistical signal-to-noise ratio by a factor of two, it is necessary to prolong the exposure time by at least a factor of four.

9.1.11.3. *Overloads*

Some detectors, or their associated read-out systems, are limited in the number of counts they can accumulate in one pixel. The number recorded reaches a maximum number which cannot be further increased, *i.e.* the pixels can become saturated. This means that these pixels retain the same maximum value on longer exposure whilst other, non-saturated, pixels continue to accumulate counts. The intensity in saturated pixels will hence be underestimated compared to the others and any intensities estimated from profiles including such pixels will be biased towards low values. It is essential that pixels that are saturated are flagged and recognized by the processing software. There are several ways to deal with the problem of saturation.

(1) Reject all reflections that contain saturated pixels. These will tend to be at low resolution. If more than a very few are rejected, this can be a truly disastrous choice, especially if the data are to be used for molecular replacement. In addition, missing the largest terms degrades the continuity and information content of all electron-density maps derived therefrom. This point is relevant to several applications (Section 9.1.13).

(2) Reject only those pixels that are saturated, and fit average standard profiles estimated from the non-saturated spots. This gives a poorer estimate than if the pixels were not saturated, but for applications such as molecular replacement or direct methods where the high-intensity data are essential, it is certainly better than option (1).

(3) Reduce the exposure time to ensure that there are no overloaded pixels. This is a trade-off, because if there is a large contrast between the intensity of the weakest and the strongest terms in the pattern, then the weaker terms will have a low and possibly unacceptable signal-to-noise ratio under this regime.

190

(4) Use more than one pass through the rotation range, with different exposure times. The longest exposures should be sufficient to ensure that the intensities of the data at the high-resolution limit of the pattern are statistically significant. The shortest should ensure that the number of saturated pixels in the 'low-resolution' pass is minimized. If the contrast between the low- and high-resolution passes is too great, differing by a factor of much more than about ten, then additional passes with intermediate exposure times should be used to allow satisfactory scaling of the data from these images. The CTDD for each pass with shorter exposure should be increased only so as to cover the resolution to which reflections were saturated on the previous pass. The rotation range on individual images can then be increased accordingly, in the wide φ-slicing option. On bright synchrotron beamlines, if the second pass requires exceedingly fast rotation of the spindle-axis motor and rapid opening and closure of the beam shutter beyond the limit of reliability, it may be better to attenuate the beam, for example with a series of aluminium foils. As discussed in Section 9.1.7.1, if high-resolution data are collected in several passes with different exposures and resolution limits, it may not be necessary to cover all of the theoretically required rotation range in the highest-resolution pass. The curvature of the Ewald sphere results in the high-resolution data being completed with a smaller total rotation range than the low. It is vital that the lowest-resolution pass covers the total rotation range required for complete data.

Clearly the optimum solution is to have a detector with a sufficient dynamic range to cover pixels of both weak and strong reflections. The dynamic range has already been increased with recent imaging plates and CCDs. Enhanced dynamic range may prove to be the most important advance of solid-state pixel detectors.

An additional advantage of the fine-slicing approach is that it leads to fewer overloads. Each reflection profile is divided between several separate images and as a result the effective dynamic range of the detector is increased.

9.1.11.4. *R factor, $I/\sigma(I)$ ratio and estimated uncertainties*

It is customary to judge data quality by the overall R_{merge}, calculated using the squares of the structure-factor amplitudes (intensities):

$$R_{\mathrm{merge}} = \sum_{hkl}\sum_i |I_{hkl,\,i} - \langle I_{hkl}\rangle| / \sum_{hkl}\langle I_{hkl}\rangle.$$

R_{merge} provides a measure of the distribution of symmetry-equivalent observed intensities. However, the most popular form of R_{merge} given above is not a proper, statistically valid quantifier. It does not take into account the multiplicity of the measurements and, as a consequence, it actually rises with increased multiplicity, falsely indicating degradation of the data quality when in reality they have a higher accuracy. Modifications of R_{merge} have been proposed to include the effect of multiple measurements properly (Diederichs & Karplus, 1997; Weiss & Hilgenfeld, 1997).

A better quantity for assessing the quality of the X-ray data is the $\sum_{hkl} I_{hkl} / \sum_{hkl}\sigma(I_{hkl})$ ratio, provided the standard uncertainties, $\sigma(I)$, are correctly estimated. Detectors such as imaging plates or CCDs do not measure individual X-ray quanta directly, having a gain factor dependent on the response of the individual detector pixel to a single X-ray photon. If the gain factor is not known accurately for a particular detector, the resulting standard uncertainties of the measured intensities will be estimated at an incorrect level. If the multiplicity of the reflections is higher than unity, it is possible to correct the uncertainties *a posteriori*. This can be done either from a comparison with the expected values using the χ^2 test, or by using the *t*-plot. The latter requires that the ratio of the differences between equivalent intensity measurements to their standard uncertainties, $t = (I_i - \langle I\rangle)/\sigma(I_i)$, follows a normal distribution with a mean of 0.0 and standard uncertainty of 1.0. Both of these methods assume the errors have a normal distribution, and that only the mean and width have been incorrectly estimated and should be appropriately adjusted. They cannot take into account systematic errors of measurement.

The data-merging procedure in addition allows the identification of statistical 'outliers' and their exclusion from the data (Read, 1999). Outliers are defined as those observations that lie sufficiently far from the mean of a set, and assumption of a normal distribution suggests they suffer from substantial systematic errors of measurement. In a crystallographic experiment, outliers are those intensity measurements that deviate unexpectedly from the mean intensity of a set of symmetry-equivalent reflections. In the recording of rotation data, one typical source of such systematic errors is erroneous classification of reflections predicted as partially or fully recorded. This is a severe problem for those reflections lying close to the blind region. A second example is the presence of so-called 'zingers' in individual CCD detector pixels caused by scintillations from trace radioactivity of the taper glass. Other problems such as shadowed or inactive regions of the detector window give rise to a range of such systematic errors.

A small number of outliers may be expected from such causes. However, the total fraction of reflections flagged as outliers and rejected from the merging process should be small, certainly much less than 1%. Larger fractions indicate serious deficiencies in the hardware or the software and suggest something is very wrong with the experiment. There should always be a physical reason for rejecting outliers, other than just a need to reject those agreeing poorly with their symmetry-equivalent intensities in order to drive down R_{merge}. It is always possible to reduce R_{merge} and to provide an apparent 'improvement' in the data by rejecting a large percentage of measurements, but this is extremely bad practice.

Good crystallographic data depend strongly on an appropriate statistical procedure. It is also inappropriate to exclude those reflections with intensities lower than a cutoff limit, such as 1σ, before or during the process of data merging. Weak intensities also carry information and their neglect introduces bias into the measured intensity distribution, affecting, for example, the overall or individual atomic temperature factors.

The true outer resolution limit of the diffraction pattern is not trivial to define and indeed depends to some extent on the application. If $I/\sigma(I)$ is higher than 1.0, then a resolution shell of data indeed contains some information in a statistical sense – provided of course that $\sigma(I)$ has been correctly estimated. However, as $I/\sigma(I)$ falls close to unity there will in practice be very few significant observations amongst a great deal of noise. It is necessary to make some decision about where to cut the effective resolution. For the application of direct methods, for example using *SHELXS* (Sheldrick, 1990), the cutoff is often defined as the resolution shell where $I/\sigma(I)$ falls to 2.0, when R_{merge} usually reaches 20–40% depending on the symmetry and redundancy. Cruickshank (1999*a*,*b*) has provided a formula for a data precision indicator (DPI) which includes the effect of falling $I/\sigma(I)$ ratio.

For other applications it may be advisable to accept even very weak data. Direct methods use only a subset of the most meaningful reflections but these should extend to as high a resolution as possible. In addition, when the data are sparse from crystals that only diffract to very limited resolution, perhaps around 3 Å, then it is essential to retain all the experimental data, even if they are weak.

9.1.12. Radiation damage

9.1.12.1. *Historical perspective*

All crystals irradiated with X-rays absorb at least a fraction of the radiation, resulting in damage to the sample (Henderson, 1990). The