movements of the physical goniostat are converted into appropriate changes in the diffraction pattern. The physical goniostat appears only to describe the data collection and, optionally, to calculate the physical goniostat angles that achieve particular crystal alignments.

The *DENZO* coordinate system (Gewirth, 1996) is used in the definition of crystal goniostats, 2θ goniostat, Weissenberg coupling and polarization.

This discussion of the coordinate systems shows that the conceptual complexity of the program description does not result in complexity of the actual use of the program. The success of data analysis does not require a full understanding of the relations between internal *DENZO* goniostats and the coordinate systems. The reason for this complexity was to create a simple pattern of correlations between crystal and detector parameters in *DENZO* refinement. This in turn allows for simple and easy-to-understand control of the refinement process and simplifies problem diagnostics. For example: the definition of refined *crystal rotx* as rotation around the data-collection axis makes hardware problems when driving the spindle and shutter result only in fluctuations of *crystal rotx*. The constant nonzero value of the refined shifts between frames of *crystal roty* and *rotz* is a sign of misalignment of the data-collection axis. Although the program compensates for this misalignment with changes in crystal orientation, this introduces a small error in the Lorentz factor. The nature of these problems is such that they do not result in a complete failure of the experiment, but they do have an impact on the quality of the result. It is up to the experimenter and the instrument manager to assess the significance of these indications.

11.4.5. Experimental assumptions

To achieve the main target of a diffraction experiment – the estimation of structure factors – three components need to be determined, with maximum possible precision:

(1) the crystal response function (the relationship between the crystal structure factor and the number of diffracted X-ray photons, which depends also on the X-ray source characteristics);

(2) the detector response function; and

(3) the geometrical description of the detector relative to the directions of the X-ray beam and crystal goniostat axes.

The main difficulty of data analysis in protein crystallography is the complexity of the process that determines these components. *HKL* can determine all three directly from the data produced by the analogue-to-digital converter (ADC). The only extra program needed is one that sends the raw ADC signal to the computer disk. For charge-coupled-device (CCD) detectors, spatial detector distortion and sensitivity per pixel functions need to be established in a separate experiment. Usually it is worthwhile to establish a geometrical description of the detector in a separate diffraction experiment. A precise determination requires a well diffracting, high symmetry, non-slipping crystal and a special data-collection procedure.

11.4.5.1. *Crystal diffraction*

The crystal response function consists of two types of factors included in the analysis: additive factors, which are represented by the background, and a number of multiplicative factors, such as exposed crystal volume, overall and resolution-dependent decay, Lorentz factor, flux variation, polarization, *etc.* Other factors, like extinction and non-decay radiation damage (radiation damage can result not only in decay, but also in a change in the crystal lattice, often a main source of error in an experiment), are ignored by *HKL*, except for their contribution to error estimates.

11.4.5.2. *Data model*

The detector response function is the main component for the data model. *HKL* supports

(1) data stored in 8 or 16 bit fields;

(2) overflow table;

(3) linear, bilinear, polynomial and exponential response, with the error model represented by an arbitrary scale;

(4) saturation limit;

(5) value representing lack of data;

(6) constant offsets per read-out channel;

(7) pattern noise;

(8) lossless compression;

(9) flood-field response; and

(10) sensitivity response.

HKL supports most data formats, which represent particular combinations of the above features. The formats define the coordinate system, the pixel size, the detector size, the active area and the fundamental shape (cylindrical, spherical, flat rectangular or circular, single or multi-module) of the detector.

The main complexity of the data-analysis program and the difficulties in using it are not in application of the data model but rather in the determination of the unknown data-model parameters. The refinement of the data-model parameters is an order of magnitude more complex (in terms of the computer code) than the integration of the Bragg peaks when the parameters are known.

The data model is a compromise between an attempt to describe the measurement process precisely and the ability to find parameters describing this process. For example, the overlap between the Bragg peaks is typically ignored due to the complexity of spot-shape determination when reflections overlap. The issue is not only to implement the parameterization, but also to do it with acceptable speed and stability of the numerical algorithms. A more complex data model can be more precise (realistic) under specific circumstances, but can result in a less stable refinement and produce less precise final results in most cases. An apparently more realistic (complex) data model may end up being inferior to a simpler and more robust approach. The complexity of model-quality analysis is due to the fact that some types of errors may be much less significant than others. In particular, an error that changes the intensities of all reflections by the same factor only changes the overall scale factor between the data and the atomic model. Truncation of the integration area results in a systematic reduction of calculated reflection intensities. A variable integration area may result in a different fraction of a reflection being omitted for different reflections. The goal of an integration method is to minimize the variation in the omitted fraction, rather than its magnitude. Similarly, if there is an error in predicting reflection-profile shape, this constant error has a smaller impact than a variable error of the same magnitude.

The magnitude and types of errors are very different in different experiments. The compensation of errors also differs between experiments, making it hard to generalize about an optimal approach to data analysis when the data do not fully satisfy the assumptions of the data model. For intense reflections, when counting statistics are not a limiting factor, none of the current data models accounts for all reproducible errors in experiments. This issue is critical in measuring small differences originating from dispersive effects.

11.4.5.3. *Data-model refinement*

The parameters of the data model can be classified into four groups:

(1) Those refinable from self-consistency of the data by a (nonlinear) least-squares method.

(2) Parameters that can be determined from internal self-consistency of the data, but for which least squares is not implemented. For example, error-estimate parameters are in this category.

(3) Parameters that have to be established in a separate experiment, *e.g.* pixel sensitivity from flood-field exposure.

(4) Parameters that are obtained from hardware description.

The least-squares method is based on minimization of a function that is a sum of contributors of the following type:

$$(\text{pred} - \text{obs})^2/\sigma^2 = \chi^2, \qquad (11.4.5.1)$$

where pred is a prediction based on some parameterized model, obs is the value of this prediction's measurement and σ^2 is an estimate of the measurement and the prediction uncertainty. *DENZO* has the following least-squares refinements:

(1) refinement of unit-cell vectors in autoindexing;

(2) refinement of background and background slope; and

(3) refinement of crystal orientation, unit cell, mosaicity, beam focus and position, detector orientation and position, and geometrical distortions that are parameterized differently for different detectors.

SCALEPACK can refine the following parameters by least-squares methods:

(1) unit cell, crystal orientation and mosaicity, including changes of these parameters during an experiment;

(2) goniostat internal alignment angles;

(3) crystal absorption, using spherical harmonics (Katayama, 1986; Blessing, 1995) expansion of the absorption surface;

(4) uniformity of exposure, including shutter timing error;

(5) correction to the Lorentz factor resulting from a misalignment of the spindle axis;

(6) reproducible wobble of the rotation axis resulting from a misalignment of gears in a spindle assembly;

(7) non-uniform smooth detector response, for example, resulting from decay of the image-plate signal during scanning; and

(8) other factors contributing to scaling resulting from a slow fluctuation of beam intensity, change in exposed volume, overall crystal decay and resolution-dependent crystal decay.

11.4.5.4. *Correlation between parameters*

Occasionally, the refinement can be unstable due to high correlation between some parameters. High correlation results in the errors in one parameter compensating for the errors in other parameters. In the case where compensation is 100%, the parameter would be undefined, but the error compensation by other parameters would make the predicted pattern correct. In such cases, eigenvalue filtering [related to singular value decomposition, described by Press *et al.* (1989) in *Numerical Recipes*] is employed to remove the most correlated components from the refinement to make it more stable. Eigenvalue filtering works reliably when starting parameters are close to the correct values, but may fail to correct large errors in the input parameters if the correlation is close to, but not exactly, 100%. Once the whole data set is integrated, global refinement [also called post refinement: Rossmann *et al.* (1979); Winkler *et al.* (1979); Evans (1987); Greenhough (1987); Evans (1993); Kabsch (1993)] can refine crystal parameters (unit cell and orientation) more precisely and without correlation with detector parameters. The unit cell used in structure-determination calculations should come from the global refinement (in *SCALEPACK*) and not from *DENZO* refinement.

11.4.5.5. *Single- and multiframe refinement*

The crystal and detector orientation parameters can be refined for each group of images or for each processed image separately. Refinement performed separately for each image allows for robust data processing, even when the crystal slips considerably during data collection.

11.4.5.6. *Active area*

Not every pixel represents a valid measurement. Specification of the active detector area in *DENZO* is derived from the format and the definition of the detector size. Detector calibration with flood-field exposure will calculate the sensitivity for each pixel and will also determine which pixels should be ignored. The input command can additionally label some areas of the detector to be ignored, most frequently the shadow caused by the beam stop and its support. To define the shape of the area shadowed by the beam stop, the useful commands are *ignore circle* and *ignore quadrilateral.* There are also commands to ignore triangular shapes, margins of the detector and a particular line or pixel.

11.4.5.7. *Flood field*

The basic method for calibration of the spatial dependence of detector sensitivity is to measure the response to a flood-field exposure. The amount of relative exposure per pixel needs to be known. *DENZO* allows for either a uniform or an isotropic source. If the source is at the crystal position, *DENZO* refinement (with a separate crystal exposure) can be used to define the geometry of the source relative to the detector. To calculate the flood-field response, an earlier determination of the detector distortion is required. The flood-field response is converted to a sensitivity function. Large deviations from the local average are used to define inactive pixels. The edge of the active area needs special treatment, depending on the method of phosphorus deposition.

11.4.5.8. *Absolute configuration*

Absolute configuration is defined relative to the data-coordinate system and is only affected by the sign of the parameter *y scale.* A mirror transformation of the data does not affect the self-consistency of the data. Thus, the correctness of the absolute configuration cannot be verified by data-reduction programs.

11.4.5.9. *Correcting diffraction images*

HKL can also generate data corrected for the above factors and/or for geometrical conversion and distortion in uncompressed, lossless compressed and lossy (non-reversible to the last digit) compressed modes in linear or 16 bit floating-point encoded format. Fig. 11.4.5.1 shows data from the APS-1 detector in (*a*) uncorrected mode, (*b*) transformed to an ideal rectangular detector and (*c*) transformed to a spherical detector.

11.4.5.10. *Detector goniostat*

The detector goniostat in *DENZO* can have only one rotation axis – 2θ. In the complex transformations described in equation (11.4.2.8), the geometrical scale is affected by pixel-to-millimetre conversion and distortion. For different instruments, the scale is defined differently. For detectors without distortion, the scale is defined by the value of the pixel size in the 'slow' direction. For detectors with distortion characterized by polynomials (*e.g.* CCD detectors), the scale is also defined by the way the distortion was determined. In such a case, the source of scale is the separation between holes in the reference grid mask or, alternatively, the goniostat translation. As the distance of the detector active surface from the crystal cannot be measured precisely, the difference between the two distances is the ultimate source of the scale reference. The angle between the detector distance translation and