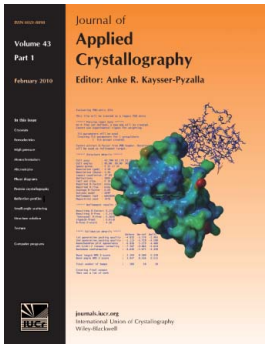
Journal of
Applied Crystallography
ISSN 0021-8898
Editor: **Anke R. Pyzalla**

EVAL15: a diffraction data integration method based on *ab initio* predicted profiles

Antoine M. M. Schreurs, Xinyi Xian and Loes M. J. Kroon-Batenburg

J. Appl. Cryst. (2010). **43**, 70–82



Journal of
Applied Crystallography
ISSN 0021-8898

Received 3 February 2009
Accepted 20 October 2009

EVAL15: a diffraction data integration method based on *ab initio* predicted profiles

Antoine M. M. Schreurs, Xinyi Xian and Loes M. J. Kroon-Batenburg*

Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. Correspondence e-mail: l.m.j.kroon-batenburg@uu.nl

A novel diffraction data integration method is presented, *EVAL15*, based upon *ab initio* calculation of three-dimensional (x, y, ω) reflection profiles from a few physical crystal and instrument parameters. Net intensities are obtained by least-squares fitting the observed profile with the calculated standard using singular value decomposition. This paper shows that profiles can be predicted satisfactorily and that accurate intensities are obtained. The detailed profile analysis has the additional advantage that specific physical properties of the crystal are revealed. The *EVAL15* method is particularly useful in circumstances where other programs fail, such as regions of reciprocal space with weak scattering, crystals with anisotropic shape or anisotropic mosaicity, $K\alpha_1/K\alpha_2$ peak splitting, interference from close neighbours, twin lattices, or satellite reflections of modulated structures, all of which may frustrate the customary profile learning and fitting procedures. *EVAL15* allows the deconvolution of overlapping reflections.

1. Introduction

Several software packages have been developed for the integration of diffraction data from area detectors. Compared with photon-counting point detectors some extra problems have to be solved to obtain accurate intensities [see Zhurov *et al.* (2008) for an error analysis of data from point detectors, CCDs and image plates]. The advantages are faster data collection and a complete picture of reciprocal space. A wide range of phenomena related to specific crystal properties can be visible at a glance, such as twinning, aperiodic structure, disorder and thermal diffuse scattering, and also unwanted effects such as the formation of ice at cryo-temperatures. The developments in detector technology and integration software have been triggered by macromolecular crystallography, where a large number of reflections can be collected simultaneously, many of which usually have a low signal. In small-molecule crystallography it was common practice to acquire reflection data through summation–integration. However, for weak reflections a better standard deviation can be achieved by profile fitting (Diamond, 1969; Ford, 1974). This involves a least-squares fit of the observed pixel intensities in a reflection peak to a learnt standard profile. This standard profile can be learnt from the underlying diffraction data. The profile learning process relies on two main assumptions (Pflugrath, 1999). Profiles of strong reflections are superimposed to construct an averaged standard profile. As the reflection profile varies with the position on the detector as a result of geometrical deformations, it is assumed that standard profiles can be learnt from spatially nearby reflections. The second assumption is that the reflection positions are predicted accurately. Uncertainties in reflection centroids lead to artificially broad profiles and to incorrect profile fits. An alternative method to profile learning has been developed by Ren & Moffat (1995). These authors model the profiles by analytical expressions that may vary slowly across the detector. Effects such as streaking of spots or anisotropic spot shape can be included.

Profile learning and fitting can be carried out in two dimensions on a single image, as with *DENZO*/*HKL-2000* and *MOSFLM* (Otwinowski & Minor, 1997; Leslie, 1999), or in a complete three-dimensional reflection box, as with *XDS*, *d*TREK*, *SAINT* and *CrysAlis* (Kabsch, 1988; Pflugrath, 1999; Bruker, 1998; Oxford Diffraction, 2008).

The need for yet another integration program lies in the fact that each of the existing ones lacks one of the following properties: (1) profile fitting in regions of reciprocal space where all reflections are weak – profile learning needs high I/σ reflections, usually non-existent at high resolution; (2) appropriate treatment of reflections with $K\alpha_1/K\alpha_2$ splitting, which is a prerequisite for high-resolution studies; (3) use of twin matrices; (4) deconvolution of overlapping reflections.

Kabsch (1988) developed an elegant procedure to obtain uniform three-dimensional profiles for all reflections by transformation to an undistorted reciprocal space, thereby overcoming the need for strong nearby reflections in profile learning. However, even then the results are better if the standard profiles are learnt separately from different regions on the detector, provided that suitable reflections are available. As reflection profiles are a convolution of broadening

effects, such as crystal size and shape, mosaicity, the beam focus dimensions and divergence, wavelength dispersion, experimental geometry, lattice distortions and internal structure of the crystal, detector point spread, and spatial distortion, an exact transformation to reciprocal space is impossible. This insight led us to the point where we were able to predict accurate reflection profiles, by taking into account all these effects explicitly, and to apply this standard profile in a least-squares fit.

EVAL15 is based on the concept of 'general impacts', as introduced in *EVAL14* (Duisenberg *et al.*, 2003). In that program an *ab initio* reflection boundary is calculated, within which summation–integration is performed. The method is widely used in chemical crystallography, in particular the version implemented in *COLLECT* (Nonius, 1999). Building on that experience, *EVAL15* calculates a complete standard reflection profile from general impacts, *i.e.* impacts originating from different parts of the crystal, beam focus, wavelength spectrum and crystal mosaic orientations. We will discuss the methods and algorithms of *EVAL15*, the details of its implementation, and the quality of the profiles. The improvement of the integration of weak reflections by using profile fitting with *EVAL15* data is demonstrated. In two separate papers, the *EVAL15* data quality for small-molecule and protein crystals is further explored and integration of difficult cases as well as overlap deconvolution will be addressed.

2. The *EVAL15* method

In this section, all steps in the *EVAL15* data integration method are explained. General impacts are generated by sampling from distributions of physical parameters. These impacts have to be convoluted with a detector point-spread function, in order to obtain a realistic predicted profile. For each individual reflection such a profile is used in a least-squares minimization using singular value decomposition (SVD) to obtain the integrated intensity and its standard deviation. Contributions to the standard deviations are discussed.

2.1. The concept of general impacts

A method for tracing X-rays in the diffraction process is explained in detail by Duisenberg *et al.* (2003); here we give only the principles.

Consider a diffraction experiment with one rotation axis and an area detector. The reflection normal $\mathbf{S}_0$ for reflection *hkl* in the zero position of the goniometer is

$$\mathbf{S}_0 = \begin{pmatrix} S_{0x} \\ S_{0y} \\ S_{0z} \end{pmatrix} = \begin{pmatrix} a_x^* & b_x^* & c_x^* \\ a_y^* & b_y^* & c_y^* \\ a_z^* & b_z^* & c_z^* \end{pmatrix} \begin{pmatrix} h \\ k \\ l \end{pmatrix}. \quad (1)$$

The matrix containing the reciprocal cell axes in the laboratory axis system is called the **R** matrix. If $\mathbf{S}_0$ can be rotated over some angle ω to a position $\mathbf{S}_\omega$ such that the angle between $\mathbf{S}_\omega$ and the primary beam equals $90^\circ - \theta$ then and only then *hkl* will reflect. The diffracted ray departs from the crystal along a direction

$$\mathbf{r} = \mathbf{S}_\omega - \mathbf{X}/\lambda, \quad (2)$$

where **X** is a unit vector along the primary beam pointing to the beam focus centre. Equation (2) follows from $\mathbf{S}_\omega$ bisecting $\angle(\mathbf{X}, \mathbf{r})$. We denote the 'central impact' coordinates by (x, y, ω), with x, y the impact position on the detector plane and ω the rotation angle at which *hkl* is brought to reflection. This point in (x, y, ω) space represents the complete reflection that would be obtained from a point source, a point crystal with no mosaicity and pure, nondiverging, monochromatic radiation.

In practice a reflection results from radiation of different wavelengths, coming from different parts of the focus and scattered by different parts of the crystal having different orientations of the mosaic blocks. Each combination of these parameters may yield a general impact, with coordinates (x, y, ω), as follows.

Consider one point **K** of the crystal, one mosaic vector $\mathbf{S}_\mathrm{m}$, one possible focal point **F** and one wavelength λ; this combination will reflect if, by ω rotation, the angle between $\mathbf{S}_{\mathrm{m},\omega}$ and $\mathbf{F} - \mathbf{K}_\omega$ can be made to be $90^\circ - \theta$. The outgoing direction **r** for a general impact is given by

$$\mathbf{r} = \mathbf{S}_{\mathrm{m},\omega} - [(\mathbf{F} - \mathbf{K}_\omega)/|\mathbf{F} - \mathbf{K}_\omega|]/\lambda. \quad (3)$$

The origin of **r** and $\mathbf{S}_{\mathrm{m},\omega}$ is not (0, 0, 0) but $\mathbf{K}_\omega$. The formula follows from $\mathbf{S}_{\mathrm{m},\omega}$ being the bisector of $\angle[(\mathbf{F} - \mathbf{K}_\omega), \mathbf{r}]$, the incoming and reflected ray, respectively. The subscript ω denotes ω-rotated vectors.

2.2. Modelling the profile

For each reflection, *EVAL15* general impacts are calculated for randomly selected $(\mathbf{F}, \mathbf{K}, \mathbf{S}_\mathrm{m}, \lambda)$ combinations, chosen from the sets of all focal points, crystal points, mosaic vectors and wavelengths, respectively. A sufficiently large number of selections from realistic distributions will eventually generate a true reflection profile. (See Appendix *A* for sampling methods of the various distributions.) The simulated profile is used as a standard profile in a least-squares fit. As the simulation is carried out for each individual reflection, specific reflection geometries are automatically accounted for.

The focus is modelled by a rectangular surface with realistic dimensions (*e.g.* 0.3 × 0.3 mm) consisting of a grid of point sources that scatter in all directions. It is assumed that from each point source a ray can hit any point in the crystal. By changing the distance of this virtual focus from the crystal, the divergence of the beam can be changed. A small distance corresponds to a larger divergence. The points on the focus are sampled uniformly, although a Gaussian distribution of intensities around the focal centre could be more realistic when certain optical elements are used. This procedure delivers a collection of vectors **F**.

The crystal shape can be described by face indexing, or, alternatively, by one of five basic shapes built into the program (a pie segment, a box, a sphere approximated by a dodecahedron or an icosahedron, or a cylinder based on an octagon). The crystal is treated as a fine cubic grid and each of the grid

points inside the crystal can be selected, thus obtaining a collection of vectors **K** defined relative to the origin.

Three distributions can be used to describe the mosaic spread. Random polar angles are sampled according to a block-shaped function within the range given by the mosaicity μ, or by a Gaussian or a Lorentzian distribution of width σ_m. For the latter two, $3\sigma_m$ corresponds to the mosaicity μ. Each vector $\mathbf{S}_m$ is then obtained by rotating $\mathbf{S}_0$ over the sampled polar and random azimuthal angles (see Appendix *A* for details). Anisotropic mosaicity is defined as an additional mosaicity μ_{aniso} around an anisotropic mosaic axis **A**. Both the size and the direction are determined by inspection of reflection profiles using the *EVAL15* display window. The graphical display of the orientation of the reciprocal axes of reflection *hkl* in reflecting position (see §3.1) is instrumental in finding the correct anisotropic mosaic axis.

The wavelength of the rays is described by a spectrum built from several Gaussians or Lorentzians, each having a central λ value, width σ_λ and a defined relative integrated ratio. Characteristic radiation from a sealed tube is described by a pair $K\alpha_1$, $K\alpha_2$ with an intensity ratio 2:1. Algorithms for sampling the various distributions are explained in Appendix *A*. The produced rays hit the detector at impact positions (x, y) and will each be collected in one associated pixel.

Position-sensitive detectors convert X-ray photons into an electronic signal. This process involves several steps, including absorption of the photons by a phosphor layer, photon storage or conversion to visible light photons, laser readout (in the case of image plates) or light transportation through a fibre optic taper to a CCD chip, and analogue-to-digital conversion (Arndt, 1986). This cascade causes the X-ray signal to spread out over several pixels, although it hits the detector at a single point. The main source of the point spread is usually the phosphor layer (Bourgeois *et al.*, 1994), and its broadening effect increases with layer thickness and with incidence angle of the impact. We found that, when simulating impacts for realistic dimensions of the crystal, focus and mosaic spread, the resulting profiles were too narrow when the point spread was neglected. We have introduced a two-dimensional pseudo-Lorentzian as the point-spread function (PSF) and took care that the integral over space to infinity, in terms of polar coordinates measured from the centre, converges to 1.0. A symmetric function is currently implemented in *EVAL15* (Fig. 1),

$$\mathrm{PSF}(x, y) = \frac{\gamma}{4\pi\left[(x^2 + y^2) + (\gamma/2)^2\right]^{3/2}}, \qquad (4)$$

where γ denotes the width of the function. See Appendix *B* for details of the implementation. Every simulated impact is spread out over neighbouring pixels using this PSF. We determined, by comparing with many observed reflections, that γ = 0.6 pixels gave realistic profiles on our Nonius KappaCCD detector. This corresponds to a full width at half-maximum, at 1% and at 0.1% of the PSF of 50, 300 and 650 µm, respectively, where the size of one pixel is 110 µm. Fig. 2 shows the effect of including the point spread for a strong reflection.

Some detectors may need γ values that depend on the incidence angle or different γ values for different parts of the detector, for example, in the case of mosaic detectors. This functionality is not yet implemented in *EVAL15*. Fig. 3 demonstrates the effect of a change in one of the parameters determining the distributions (**F**, **K**, $\mathbf{S}_m$, λ) and the point-spread width γ. To predict a reflection profile, it is necessary to find good parameters for each of the instrumental constants and crystal properties. The number of reflections (observations) is large enough to allow optimization of all necessary parameters. To that end a selection of reflections with varying impact coordinates on the detector, θ values and duration (see §3.1 for the definition) is used, all with medium or large I/σ (see Xian *et al.*, 2010*b*), for which the figures of merit of the reflection boxes are minimized. Often the number of parameters that have to be determined is limited. The dimensions of the focus and the wavelength spectrum (*e.g.* $K\alpha_1$, $K\alpha_2$ for a home source or a monochromatic wavelength for synchrotron radiation, each possessing a small dispersion) are more or less known beforehand. The divergence of the primary beam depends on the instrumental setup (Greenhough & Helliwell, 1982*a*,*b*). The focus distance and the point spread of the detector have to be determined once for a particular instrument. We often have a microscope image of the crystal or even a face-indexed description, obtained with *COLLECT* (Nonius, 1999), that we can use in *EVAL15*. This leaves only the mosaicity of the crystal as the parameter to be established. For protein crystals one would typically know only roughly the size of the crystal and, in general, not how it is oriented in the beam. *EVAL15* has the ability to optimize relevant parameters in an automatic fashion on a selection of reflections (some 10–50 in number). These may vary in resolution, position on the detector, I/σ or duration. The best procedure for doing this is described by Xian *et al.* (2010*b*).

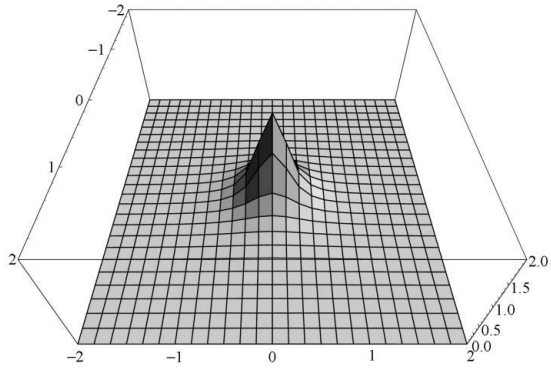


Figure 1
Two-dimensional pseudo-Lorentzian representing the point spread, here corresponding to 2 × 2 pixels. This graphic was made using logarithmic function values with *Mathematica* (Wolfram Research, 2005).

2.3. Parameter optimization

The predicted profile is taken as a normalized standard profile, used in minimization of the residual:

$$\chi^2 = \sum_{i=1}^{N} w_i \left(\rho_i - JP_i - \sum_{m}^{M} J_m P_{im} - ax_i - by_i - c \right)^2, \quad (5)$$

where N is the total number of pixels in the reflection box, ρ_i is the observed photon count, P_i is the normalized predicted profile value at pixel i, x_i and y_i are the horizontal and vertical pixel coordinates, J is the scale factor between the standard and the observed profile such that the integrated intensity $I = \sum_i JP_i$, and a, b and c define a plane describing the local background. The weights w_i are the inverse of σ_i^2 (Leslie, 1999). Assuming a Poisson distribution of counting errors, the standard deviation of the pixel intensity $\sigma_i = \rho_i^{1/2}$. M neighbouring reflections in the reflection box have their own profile P_m and scale factor J_m; some of these reflections may be significantly overlapping the main reflection. In this procedure, overlapping neighbour reflections are automatically deconvoluted from the main reflection. The parameters can be found by solving an overdetermined set of normal equations following from equation (5): $(\mathbf{A}^{\mathrm{T}}\mathbf{A}) \cdot \mathbf{c} = \mathbf{A}^{\mathrm{T}} \cdot \mathbf{t}$, where $\mathbf{A}$ is an $N \times (M + 4)$ matrix given by

$$\mathbf{A} = \begin{pmatrix} P_1/\sigma_1 & P_{11}/\sigma_1 & \cdots & x_1/\sigma_1 & y_1/\sigma_1 & 1/\sigma_1 \\ P_2/\sigma_2 & P_{21}/\sigma_2 & & & & \\ . & & & & & \\ . & & & & & \\ P_N/\sigma_N & P_{N1}/\sigma_N & \cdots & & & \end{pmatrix}, \quad (6)$$

$\mathbf{c}$, a vector of dimension $(M + 4)$, represents the fitting coefficients and $\mathbf{t}$, a vector of dimension N, contains the observations:

$$\mathbf{c} = \begin{pmatrix} J \\ J_1 \\ . \\ a \\ b \\ c \end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix} \rho_1/\sigma_1 \\ . \\ . \\ . \\ . \\ \rho_N/\sigma_N \end{pmatrix}. \quad (7)$$

N is the number of pixels in the three-dimensional reflection box (see §3.1) and T denotes the transpose of the matrix.

The normal matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ can become singular or almost singular when the standard profiles P and P_m are almost linearly dependent. A general approach to solve the numerical instability and to choose a reasonable solution is SVD (Nash, 1990). In SVD the matrix $\mathbf{A}$ is decomposed into $\mathbf{U}\cdot\mathbf{W}\cdot\mathbf{V}^{\mathrm{T}}$, where $\mathbf{W}$ is a diagonal matrix made up of the square roots of the eigenvalues of the normal matrix. The best solution to the normal equations is $\mathbf{c} = \mathbf{V}\cdot[\mathrm{diag}(1/W_j)]\cdot\mathbf{U}^{\mathrm{T}}\cdot\mathbf{t}$. It follows that the coefficients c_j are given by

$$c_j = \sum_{i=1}^{M} (\mathbf{U}_i \cdot \mathbf{t}/W_i) V_{ji}. \quad (8)$$

If the singular value W_i is (close to) zero, the normal matrix is singular. The corresponding $1/W_i$ has to be set to zero (Press *et*

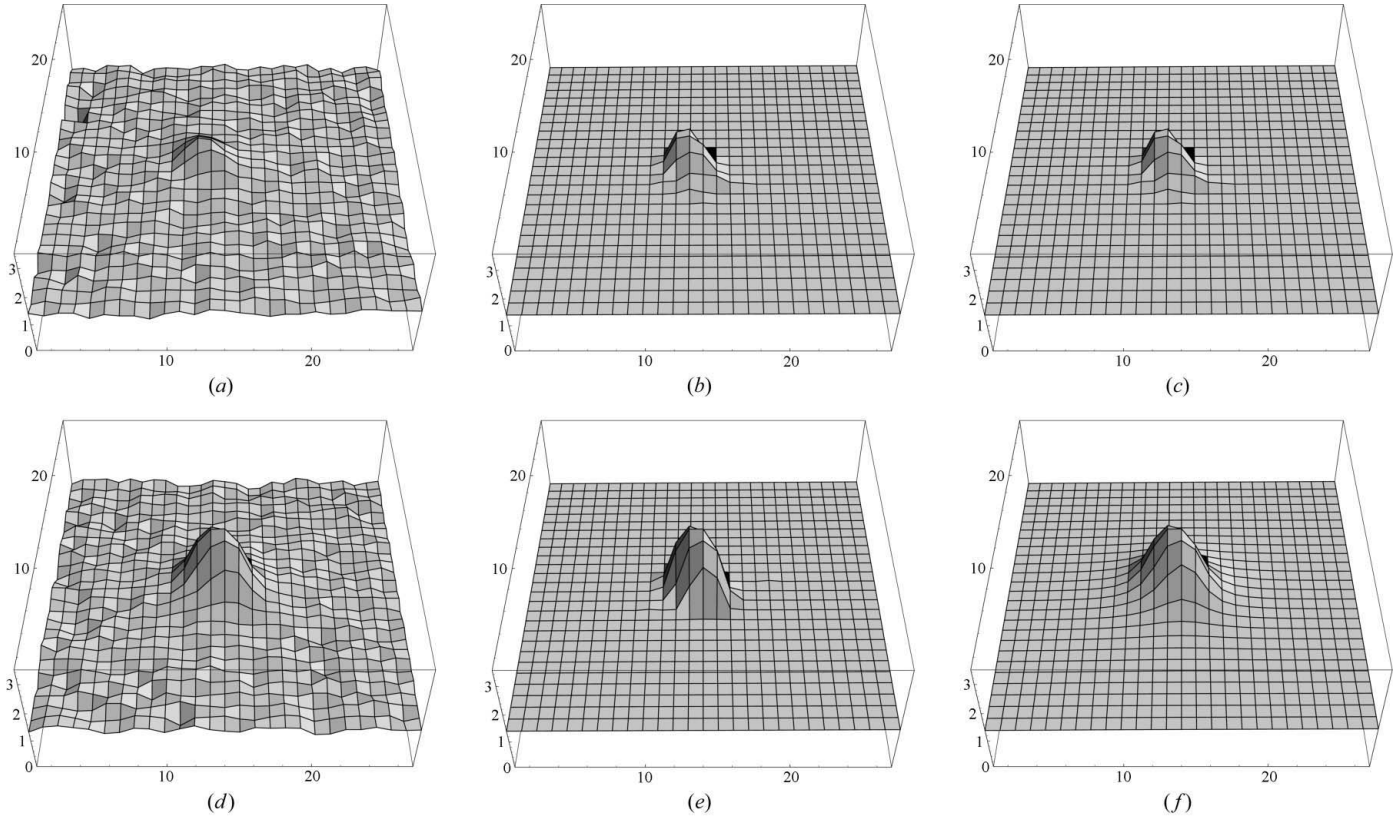


Figure 2
Profiles in two consecutive frames. Frame 1: (*a*) observed profile, (*b*) simulated profile and (*c*) simulated profile + point spread. Frame 2: (*d*) observed profile, (*e*) simulated profile and (*f*) simulated profile + point spread. This graphic was made using logarithmic function values with *Mathematica*.

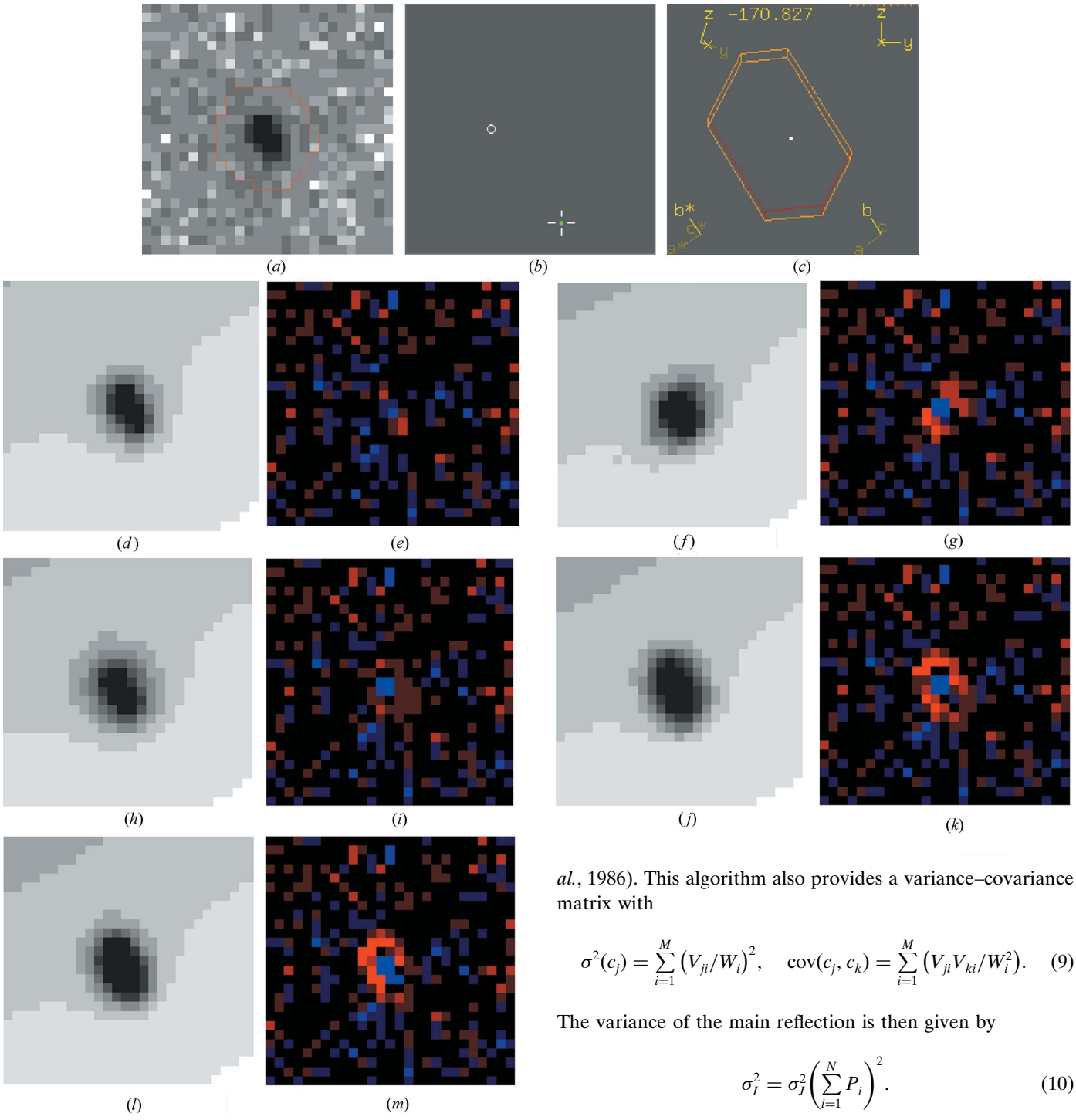


Figure 3
Effect of parameter choice on reflection profiles. (*a*) Observed profile of a reflection with $I/\sigma \simeq 50$; (*b*) position on the detector; (*c*) face-indexed crystal model viewed from the direction of the X-ray beam; (*d*) model profile with optimal parameters: mosaicity $\mu = 0.2^\circ$, point spread $\gamma = 0.5$ pixels, focus distance = 150 mm, resulting in $\text{fom}_{\text{peak}} = 1.05$; (*e*) difference (*a*) − (*d*); (*f*) $\mu = 1.0^\circ$ ($\text{fom}_{\text{peak}} = 1.43$); (*g*) difference (*a*) − (*f*); (*h*) $\gamma = 1.0$ pixels ($\text{fom}_{\text{peak}} = 1.20$); (*i*) difference (*a*) − (*h*); (*j*) focus distance = 50 mm ($\text{fom}_{\text{peak}} = 2.27$); (*k*) difference (*a*) − (*j*); (*l*) crystal size multiplied by a factor 2.0 ($\text{fom}_{\text{peak}} = 2.51$); (*m*) difference (*a*) − (*l*). The observed and model profiles are shown on a grey scale using logarithmic intensities; the difference profiles are coloured blue and red for positive and negative differences on a linear scale by Δ/σ units. For the definition of fom_{peak} see §2.3.

al., 1986). This algorithm also provides a variance–covariance matrix with

$$\sigma^2(c_j) = \sum_{i=1}^{M} \left(V_{ji}/W_i\right)^2, \quad \text{cov}(c_j, c_k) = \sum_{i=1}^{M} \left(V_{ji}V_{ki}/W_i^2\right). \quad (9)$$

The variance of the main reflection is then given by

$$\sigma_I^2 = \sigma_J^2 \left(\sum_{i=1}^{N} P_i\right)^2. \quad (10)$$

In this procedure, we automatically obtain the intensities and variances of neighbouring (overlapping) reflections in the box too. Even if the overlap with neighbouring reflections is insignificant, the profiles of the neighbours are still important to calculate an appropriate background. The covariance of the main reflection and an overlapping neighbour tells us if we can reliably split the two or if we can only determine the sum intensity of the overlapping reflections. If the value of χ^2 in equation (5) is large then the standard profile does not give a good fit. In *EVAL15* we use the following figure-of-merit to indicate the quality of the fit:

$$\mathrm{fom}_{\mathrm{box}} = \left[\frac{\sum_{i=1}^{N} w_i(\rho_i - \rho_i^{\mathrm{calc}})^2}{N - N_{\mathrm{p}}}\right]^{1/2}, \qquad (11)$$

where N_{p} is the number of fitting parameters, usually $(M + 4)$, and $\rho_i^{\mathrm{calc}} = JP_i + \sum_m J_m P_{im} + ax_i + by_i + c$. In a similar way we calculate $\mathrm{fom}_{\mathrm{peak}}$ and $\mathrm{fom}_{\mathrm{bg}}$, where the summation runs over the pixels in the peak and those in the background, respectively. For this purpose the peak area is defined by those pixels that receive at least 0.3% of the total number of calculated impacts. We have chosen to optimize J, J_m, a, b and c simultaneously for the whole reflection box.

An approach to include neighbours in the least-squares fit in the case of overlapping reflections, as described by equation (5), was also followed by Bourgeois *et al.* (1998) and Ren & Moffat (1995). However, the selection of pixels that are included is different. Bourgeois *et al.* (1998) developed the PROW algorithm, which optimizes the profile fitting area such that I/σ is maximized. The background parameters are fitted separately. Ren & Moffat (1995) created a dynamic mask based on the relative profile value in a learnt analytical profile. Both procedures included all pixels of a neighbour reflection if it is predicted to overlap. In *EVAL15*, neighbouring reflections are taken into account if they produce impacts in a significant number of pixels in the reflection box, but might not be included entirely. We verified that the intensity and standard deviation of a main reflection are not sensitive to the number of pixels that are included for the neighbour, unless they strongly overlap. However, in the latter case all pixels of the neighbour are included automatically because it is also positioned near the centre of the reflection box.

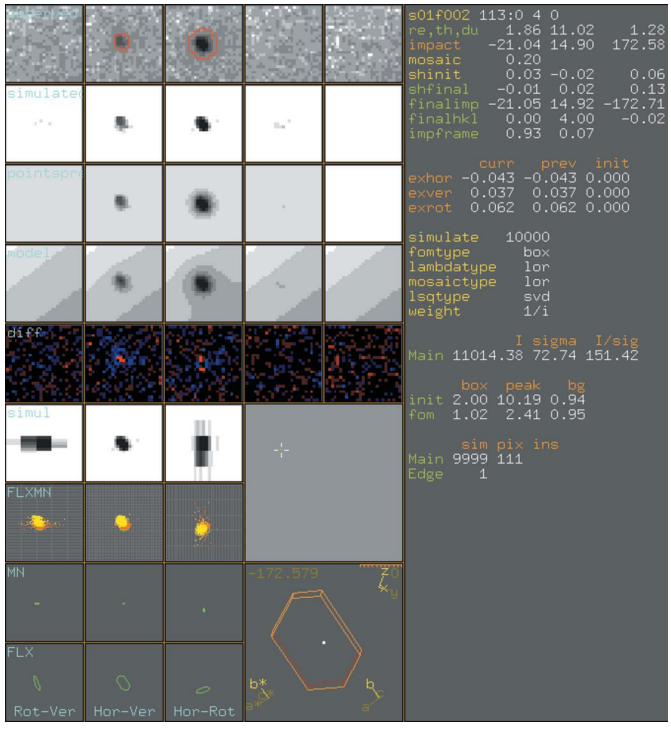


Figure 4
The *EVAL15* graphical display.

All papers describing profile-fitting algorithms conclude that the reflection positions should be known accurately, both for the learning and for the fitting procedure. *EVAL15* will optimize the position (horizontal, vertical and rotational impact coordinates) of the reflection, by minimizing $\mathrm{fom}_{\mathrm{box}}$ using the simplex method (Nelder & Mead, 1965; Press *et al.*, 1986). This can only be done reliably for reflections that are strong enough. If shifts of the impact positions of a collection of reflections with similar ω values are larger than is acceptable, a post-refinement is carried out and the procedure is started all over again. In practice the final shifts are typically one-third to one-half of the size of a pixel and one-third of the rotation range ω. Since the profile is recalculated for each position no extrapolation of the profiles has to be carried out.

2.4. Standard deviations and gain of the detector

Every detector converts X-ray photons into an electronic signal that is read out and stored in an image file. The detective quantum efficiency (DQE) is a measure of the efficiency with which photons are detected and of the noise performance of the detector. It is defined as the signal-to-noise ratio of the output signal divided by that of the input signal. For an ideal detector this ratio would be 1.0. In practice many factors reduce this number, such as phosphor absorption efficiency, window transmission, phosphor noise factor, readout noise, dark current and detector gain (Phillips *et al.*, 2002). The DQE is a determinant factor for the data quality, *i.e.* the output signal/noise ratio. For data integration our concern is the correct estimation of I/σ. The definition of gain varies in the literature. In the rest of the paper we will use gain as the number of ADUs (analogue-to-digital units) per X-ray photon. In *EVAL15* all pixel intensities are divided by the gain, if this number is available from the header of the image files; otherwise it can be input manually. A whole cascade of processes is responsible for the gain value. In the ideal case, the manufacturer determines this number such that, after dividing by gain, an estimate of the standard deviation for each pixel intensity can be obtained using Poisson statistics. It is specific for the wavelength used. The background intensity of a reflection box is represented by a plane with parameters a, b and c (see above). Noise causes deviations between fitted and observed background pixel intensities that are measured by $\mathrm{fom}_{\mathrm{bg}}$. These deviations are expected to follow a Gaussian distribution such that $\mathrm{fom}_{\mathrm{bg}}$ should be near 1.0 provided the correct value for gain is used. This reasoning assumes that analog-to-digital converter noise and/or dark current are effectively removed from the background so that it only consists of X-ray scattering. A large deviation of $\mathrm{fom}_{\mathrm{bg}}$ from 1.0, in the various reflection boxes, indicates an incorrect gain value. The value of gain can then be adapted. However, in §3.3 we show that the use of the correct gain value is not all that important, when the standard deviations are multiplied by $\mathrm{fom}_{\mathrm{peak}}$.

Popov & Bourenkov (2003) elaborate on the various contributions to standard deviations from summation–integration. These can be described by a second-order polynomial

in I. The zeroth-order coefficient is related to the incoherent background scattering, dark current and readout noise. In *EVAL15* the noise originating from dark current and readout is included *via* the parameter bgnoise, which is estimated from the dark images. The first-order term in Popov & Bourenkov's approach is the standard deviation due to Poisson counting statistics of the integrated intensity. Both effects are introduced as weights in the least-squares fit by using $\sigma_i = (\rho_i + \text{bgnoise}^2)^{1/2}$, resulting in the standard deviation σ_I [equation (10)] of the integrated intensity I. The second-order term is the contribution of instrument errors. Both systematic measurement errors and errors in the profile will unavoidably lead to misfits of profiles, especially at large I/σ. If the fom_{peak} value is larger than 1.0 the deviation between model and observation is larger than expected and obviously the model is not completely correct; somehow this should be expressed in the estimated standard deviation of the intensity. Two approaches seem justified to adapt the standard deviations σ_I of the integrated intensities. (1) Multiply σ_I by fom_{peak}. A similar approach is followed by Leslie (1999). This approach accounts for both the profile and the instrument errors. (2) Leave it to the scaling program, in our case *SADABS* (Sheldrick, 1996), to find an error model for the standard deviations from the internal r.m.s. deviations $\sigma_{\text{int}} = [\sum_i (I_i - \langle I \rangle)^2/(N-1)]^{1/2}$ of equivalent reflections. In this case *EVAL15* only outputs σ_I (Poisson + bgnoise) and *SADABS* provides the second-order term representing instrument errors.

3. Results and discussion

3.1. The *EVAL15* graphical display

After finding the **R** matrix with *DIRAX* (Duisenberg, 1992) and refinement of variables determining the reflection positions with *PEAKREF* (Schreurs, 1999), integration boxes are extracted, one for each separate reflection (of typically 27 pixels × 27 pixels × 5 frames), from the images using the 'datcol' procedure in *VIEW* (Schreurs, 1998). The size of the boxes is chosen by visual inspection and should normally be sufficiently large to contain the complete reflection and a fair portion of background, though this is less critical in *EVAL15*, as it can also integrate incomplete reflections. Fig. 4 shows the graphical display of *EVAL15* for one reflection. The top-left panel shows successive observed ω slices. In the second row, the resulting profile from a sample of 10 000 impacts calculated using equation (3) is shown. In the third row, the point-spread function is applied; then the scale factor for the profile and the background parameters are determined and in the fourth row the resulting model is displayed. Finally, the difference between observation and model is shown in a red/blue colour scale. The least-squares procedure is applied to all slices simultaneously. The right side of the window contains information on the position of the reflection: the resolution, θ and relative duration (*i.e.* the rotation range compared with a reflection with the same $2\sin\theta$ passing through the Ewald sphere at the equator) as well as the central horizontal, vertical and rotational impact coordinates. The difference between the original and final impact is also shown on the right panel (`impact` *versus* `finalimp`). The type of distributions (Lorentzian, Gaussian or block) and the numerical values for relevant parameters are shown as well. The values I, σ and I/σ corrected for the Lorentz (Milch & Minor, 1974) and polarization factors and the figures of merit can be found a few lines lower. As mentioned in §2.4, a (local) planar background in the reflection box is assumed. Occasionally, zingers may occur in the peak or background. In *EVAL15* a pixel rejection procedure is implemented, based on $(\rho_i^{\text{obs}} - \rho_i^{\text{calc}})/\sigma_i > 5$–10. The user can decide to avoid areas with ice rings through the peak or in the background in the *VIEW* datcol procedure. The lower-left part of the window shows *EVAL14* contours. In the case shown, the shape of the crystal was obtained through face indexing and it is shown by default in the orientation at the diffracting position seen from the X-ray source

Table 1
Data collection for sucrose.

Crystal data	
$2C_{12}H_{22}O_{11}$	V = 715.47 (9) Å^3
Mr = 342.30	Z = 2
Monoclinic, $P2_1$	Mo $K\alpha$ radiation
a = 7.7613 (7) Å	μ = 0.144 mm^{-1}
b = 8.7061 (7) Å	T = 298 K
c = 10.8652 (4) Å	0.24 × 0.12 × 0.21 mm
β = 102.960 (4)°	
Data collection	
Nonius KappaCCD diffractometer	Absorption correction: face indexed (*SADABS*)
Detector distance: 40 mm	φ and ω scans with 1° increments

3.2. Predicted profiles for standard and notoriously difficult cases

In this section, we demonstrate the performance of *EVAL15* for a high-resolution data set of sucrose. Next we will show examples of profile predictions for cases that present most integration software packages with significant difficulties. In the two following papers such cases will be treated extensively (Xian *et al.*, 2010*a*,*b*).

3.2.1. Comparison of *EVAL14* and *EVAL15* for high-resolution data with $K\alpha_1/K\alpha_2$ splitting. Crystal data for sucrose are listed in Table 1. The performance of *EVAL14* and *EVAL15* for data of a sucrose crystal is compared in Table 2. The crystal was face indexed using *COLLECT* (Nonius, 1999) and the description was used in *SADABS* for absorption correction. Refinement was carried out with *SHELXL* (Sheldrick, 2008). Data were integrated to a resolution of 0.46 Å, at which resolution reflections are clearly split owing to the use of $K\alpha_1/K\alpha_2$ radiation. The basic difference between *EVAL14* and *EVAL15* is the use of summation–integration *versus* profile fitting. In *EVAL14* the pixel intensities within a reflection contour are summed while subtracting the average background intensity determined from an area surrounding the reflection. We noticed, while studying the point-spread function of a detector, that reflection intensity is spread over the complete area of the integration box. Thus there is in fact no way to determine the true background scattering close to

Table 2
Data for sucrose after scaling and absorption correction with *SADABS*, and refinement with *SHELXL*.

	Max. resolution†	No. measured No. unique No. > 2σ	I/σ‡	R_1(strong) R_1(all)	wR_2	S	K_{min} K_{max}	$\Delta\rho_{max}$ $\Delta\rho_{min}$	Weights: a b	⟨s.u.⟩ C—C§	⟨s.u.⟩ U_{eq}(C)§
EVAL14	0.77	13157 3281 3107	30.9 (12.5)	0.0266 0.0291	0.0665	1.062	0.959 1.067	0.27 −0.21	0.0389 0.1161	0.0019	0.00027
	0.46	36170 14400 8362	16.0 (0.6)	0.0567 0.1221	0.1238	1.019	0.970 1.253	0.47 −0.38	0.0549 0.0	0.0013	0.00015
EVAL14-like	0.77	13344 3280 3164	33.7 (13.6)	0.0269 0.0285	0.0683	1.072	0.958 1.068	0.28 −0.20	0.0395 0.1210	0.0019	0.00027
	0.46	35874 14335 9381	20.3 (1.0)	0.0578 0.1062	0.1325	1.116	0.970 1.184	0.45 −0.33	0.0508 0.0292	0.0013	0.00014
EVAL15	0.77	13605 3284 3229	35.6 (19.7)	0.0256 0.0261	0.0667	1.055	0.956 1.056	0.28 −0.19	0.0429 0.1061	0.0018	0.00023
	0.46	37850 14421 10521	24.4 (1.4)	0.0420 0.0690	0.1072	0.998	0.976 1.029	0.38 −0.26	0.0573 0.0	0.0009	0.00010

† Maximum resolution in Å. ‡ Number in parentheses for highest-resolution shell. § s.u. for distances in Å and atomic displacement parameters in Å^2. K = minimum and maximum resolution-dependent scale factor K for F^2_{obs} *versus* F^2_{calc}; $R_1 = \sum ||F_{obs}| - |F_{calc}|| / \sum |F_{obs}|$ for $F_{obs} > 4\sigma(F_{obs})$, and for all structure factors; $wR_2 = \{\sum[w(F^2_{obs} - F^2_{calc})^2] / \sum[w(F^2_{obs})^2]\}^{1/2}$; $S = \{\sum[w(F^2_{obs} - F^2_{calc})^2]/(n-p)\}^{1/2}$, where n = number of reflections, p = number of refined parameters; $\Delta\rho_{max,min}$ = maximum and minimum difference density (e Å^{-3}); weights $w = 1/[\sigma^2(F^2_{obs}) + (aP)^2 + bP]$, where $P = (F^2_{obs} + 2F^2_{calc})/3$.

the reflection in a summation–integration procedure. In *EVAL15* the intensity and background parameters are determined in one single least-squares fit and we believe therefore that the background is appropriately determined. We have implemented an *EVAL14*-like summation–integration in the *EVAL15* software, where the pixel intensities are summed within the area confined by 0.3% of the maximum of the model profile and the background is obtained from the *EVAL15* least-squares fit. We found that on average the integrated intensities in *EVAL15* are 7% larger than in the *EVAL14*-like procedure. This is caused by the fact that, because of the point spread, some reflection intensity is outside of the contour and by applying a reflection boundary some 7% of the intensity is lost (bold hatched area in Fig. 5). The *EVAL14*-like intensities approach those of *EVAL15* when the contour size (boundary) is increased. In the normal *EVAL14* method summation–integration is also performed within a predicted contour, while the background plane is determined separately from pixels in the background area. Since in fact all pixels in the reflection box are affected by reflection intensity spread through the point-spread function, the background level becomes too high and the net intensity becomes too small. Therefore, in *EVAL14* an additional 1% of the intensity is lost.

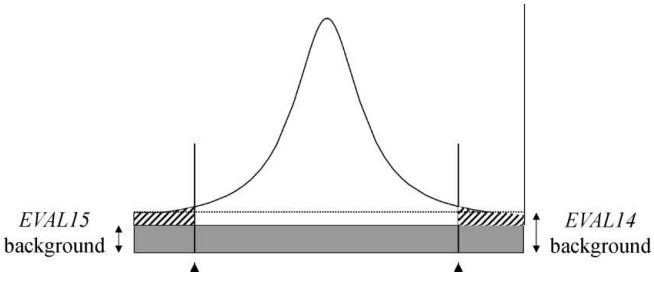


Figure 5
Difference in background and reflection intensities for the two methods. The grey box indicates the true background as determined by *EVAL15*. The *EVAL14*-like integration, while using the *EVAL15* background, fails to include the hatched area as a result of a boundary cut-off (boundary indicated by the arrow heads). The normal *EVAL14* method has a higher background (dotted line) because that includes reflection intensity due to point spread.

After post-refinement it was no longer necessary to shift the positions of reflections. The results of a complete integration with and without applying shifts are indistinguishable. The average accuracy of the reflection positions is one-third of the pixel size and 0.1° in the rotation direction.

The refinement results for the lower-resolution data sets are similar for all three methods, although *EVAL15* data have larger average I/σ values. The difference between the methods becomes visible at high resolution (0.46 Å). *EVAL15* clearly performs better for low-intensity reflections. It has fewer weak data points ($I < 2\sigma$) and I/σ is larger over the whole resolution range.

The R values in the refinement are significantly lower, as are the residual densities. The estimated standard deviations for the coordinates, expressed in terms of C—C covalent bond distance standard deviations, and of the ADUs, expressed as the s.u. values of U_{eq}, are notably smaller than for *EVAL14*. Integration of high-resolution reflections with home X-ray sources requires appropriate treatment of $K\alpha_1/K\alpha_2$ splitting. *EVAL14* proved over the years to predict the corresponding contours correctly and delivered good data. However, we have now shown that *EVAL15* performs even better in the case of weak data.

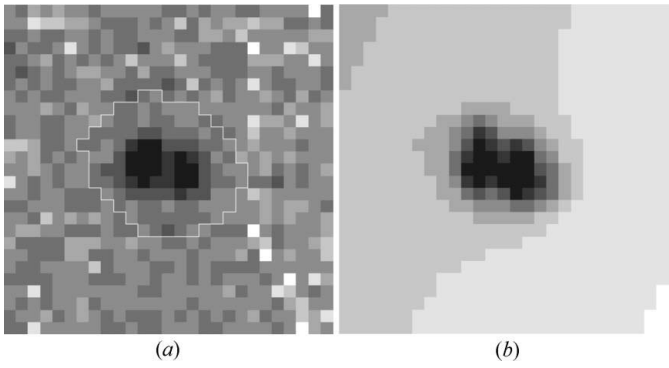


Figure 6
$K\alpha_1/K\alpha_2$ splitting for reflection $\overline{12}$ 0 0, as occurring at an example resolution 0.63 Å in the data of sucrose. (*a*) Observed and (*b*) simulated profiles.

Reflections are notably split at higher resolution as a result of the use of $K\alpha_1$, $K\alpha_2$ radiation from sealed tubes or rotating anode sources. Since *EVAL15* uses both wavelengths in the simulation, with a ratio of 2:1, accurate profiles are obtained, as can be seen in Fig. 6.

3.2.2. Fine slicing. Pflugrath (1999) discusses the possible advantages of fine slicing, *i.e.* images recorded over a rotation-angle range significantly smaller than the effective mosaic spread.[1] The advantages could be a lower X-ray background per image, fewer saturated pixels, fewer spatial overlaps and better positional accuracy of the reflection after post-refinement. Inherent disadvantages are that the intensity of a reflection is spread over a larger number of pixels (reducing the signal/noise), readout noise is accumulated over several images, and the process is more demanding in terms of disk space, goniometer hardware, shutter synchronization and the scaling procedure. As *EVAL15* integrates three-dimensional reflection boxes the treatment of fine sliced data is straightforward. Fig. 7 shows that the shape of partial reflections can be different from frame to frame; similar behaviour was observed by Pflugrath (1999). The profile-fitting algorithm in *EVAL15* is in no way hampered by these differing shapes; in fact the profiles are predicted accordingly.

3.2.3. Overlapping reflections. Overlapping reflections due to long cell axes or twin lattices can be deconvoluted even if up to 90% of their pixels overlap (Xian *et al.*, 2010*a*). The intensity and standard deviation for all reflections in the reflection box are estimated, though we only use that of the main reflection for which the box is made. The neighbour's intensity will be integrated in a separate neighbour reflection box. The profile of the neighbour is obtained using the same sets of sampled rays that impact on the pixels centred near the predicted position of the neighbour (Fig. 8). The indexing programs *DIRAX* (Duisenberg, 1992) or *CELL_NOW*

[1] The rotation range of a reflection is determined by the size and mosaic spread of the crystal, the wavelength dispersion, the beam divergence, and the Lorentz factor (Helliwell *et al.*, 1993). The relative duration used in *EVAL15* is defined as the Lorentz factor divided by $2\sin\theta$ and thus is the duration relative to a reflection passing through the Ewald sphere in the equatorial plane when the rotation axis is perpendicular to the primary beam.

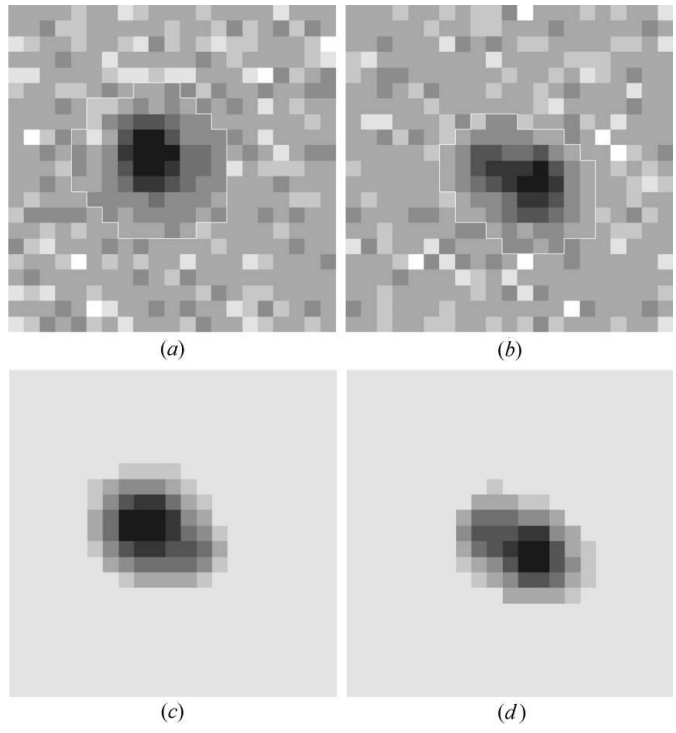


Figure 7
(*a*), (*b*) Observed and (*c*), (*d*) simulated profiles of a reflection on two successive frames.

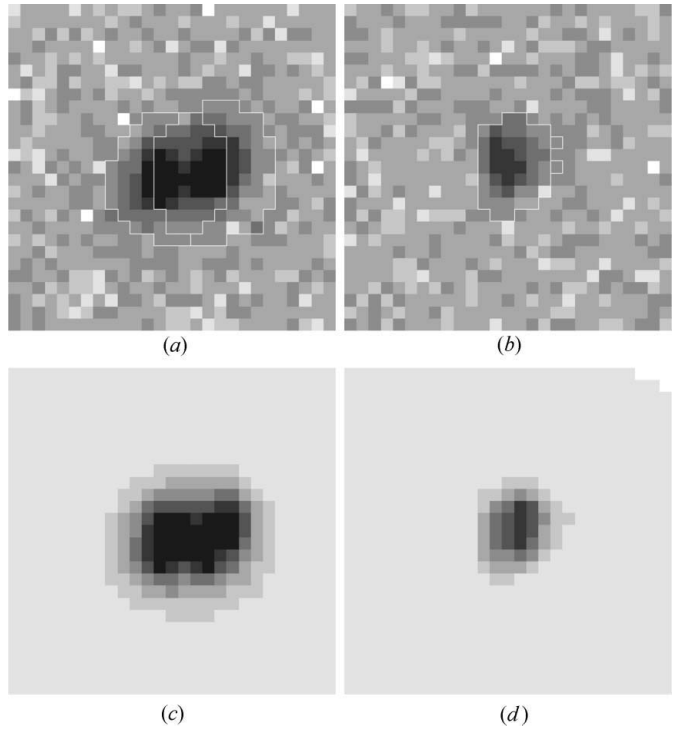


Figure 8
Overlapping reflections from twin lattices can be simulated and their relative intensities are obtained through the SVD algorithm. (*a*), (*b*) Two consecutive frames of the observed reflection box are shown. (*c*), (*d*) The model profiles are made using the fitted relative intensities (I/σ = 77.6 for *hkl* main = 147 and I/σ = 67.9 for *hkl* neighbour = $14\overline{7}$).

(Sheldrick, 2005) are particularly suited to find interfering lattices.

3.3. Standard deviations

The error model used in *SADABS* for the standard deviations is $\sigma_c = K[\sigma_I^2 + (g\langle I\rangle)^2]^{1/2}$, where g (typically ~0.02 for KappaCCD/*EVAL* data) accounts for (systematic) instrument errors (McCandlish *et al.*, 1975). Using this expression and the fact that we expect σ_{true} to be equal to $\mathrm{fom}_{peak}\sigma_I$ we can write $\mathrm{fom}_{peak} = \sigma_{true}/\sigma_I \simeq K[1 + (gI/\sigma_I)^2]^{1/2}$. Fig. 9 shows that the *EVAL15* fom_{peak} values and the standard deviation multiplication factor of *SADABS* are strongly correlated for a test data set (for details see Xian *et al.*, 2010*b*) and that fom_{peak} accounts for a large portion of the instrument errors indicated by *SADABS*. Minimization of fom_{peak} for a selected set of strong reflections ($I/\sigma > 20$) turns out to be a good guide in finding the optimal profile prediction parameters and reduces the contribution of the profile part to the value of g in *SADABS* [see a separate paper (Xian *et al.*, 2010*b*) for a recipe to find the best profile prediction parameters].

The values of gain and bgnoise may not be known exactly. We have examined the consequence of the choice of these values on the estimation of the standard deviations σ and on the refinement. KappaCCD test data of a crystal of an organometallic compound were integrated and the gain was initially set to 1.5, the value given in the header. However, from the average fom_{bg} we estimated it to be 1.2. Table 3 shows that the *SADABS* error model parameters change on changing either gain or bgnoise, but the *SHELXL* refinement results were not significantly different except for the weighting scheme. This implies, that although the estimation of standard deviations varies even after going through *SADABS*, the refinement results are similar after applying the weighting scheme.

We also investigated the effect of multiplying the standard deviations with fom_{peak}. Since every reflection has its own fom_{peak}, the I/σ values are sometimes changed considerably (Fig. 10). This has little consequence on the refinement. However, the error model parameters and the weights in the refinement become similar for the different gain and bgnoise values (Table 3) and the value for g is near 0.0. It can be concluded that multiplying the standard deviation with fom_{peak} reduces the sensitivity to the values of gain and bgnoise and gives better estimates of the true standard deviations.

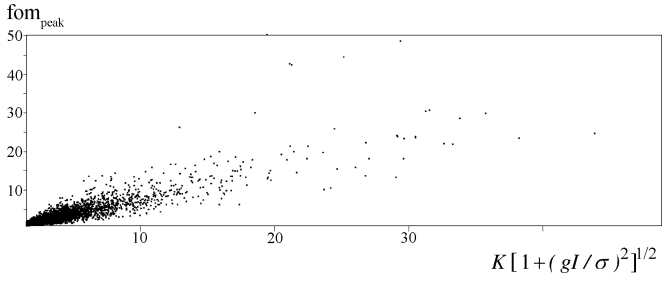


Figure 9
Linear relationship between fom_{peak} and the correction factor for standard deviations obtained with *SADABS*. Graphics produced using *ANY* (Schreurs, 2007).

The choice of the profile model clearly matters for the fom values and the refinement residuals as is seen in Table 3. If the mosaic spread is chosen too small (0.2°), the results are significantly worse.

Most refinement programs [*e.g. SHELXL* (Sheldrick, 2008) and *Crystals* (Betteridge *et al.*, 2003)] establish a weighting scheme for the intensities or structure factors, to account not only for additional experimental errors but also for model errors. Normal probability plots (Abrahams & Keve, 1971; Fig. 11) indicate that the standard deviations are under-

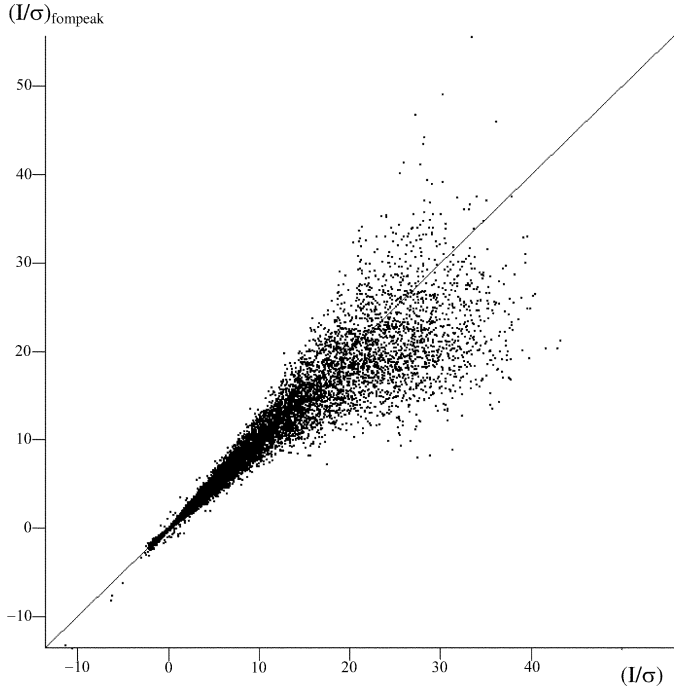


Figure 10
After applying *SADABS* the I/σ values are changed considerably by multiplying the initial σ with fom_{peak}.

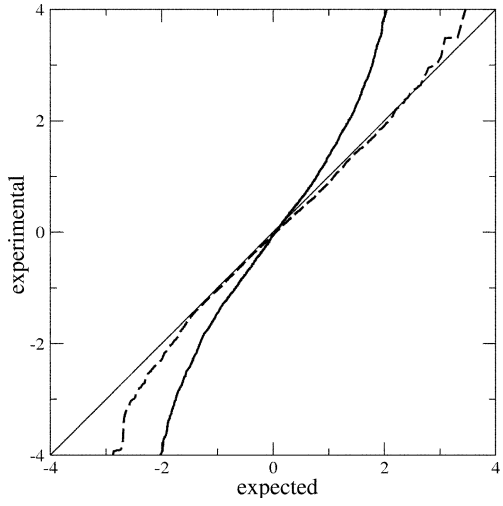


Figure 11
Normal probability plots for $(F_{obs}^2 - F_{calc}^2)/\sigma$ (solid line) and $w^{1/2}$ $(F_{obs}^2 - F_{calc}^2)$ (broken line). The standard deviations were estimated using $\sigma_I \mathrm{fom}_{peak}$.

Table 3
SHELXL refinement of data of α-tris(2,4-pentanedionato-κ^2-*O*,*O*′)cobalt(III) (von Chrzanowski *et al.*, 2007) using different estimations of σ.

Mosaic spread	Gain bgnoise	$\langle fom_{peak}\rangle$	$\langle fom_{bg}\rangle$	σ_{EVAL15}	*K*	*g*	R_1(strong) R_1(all)	wR_2	*S*	$\Delta\rho_{max}$ $\Delta\rho_{min}$	Weights: *a* *b*
0.7	1.0	1.756	1.094	σ_p	1.63	0.0201	0.0344	0.0806	1.062	0.35	0.0253
	0.83						0.0512			−0.59	1.52
0.7	1.2	1.599	0.994	σ_p	1.54	0.0205	0.0342	0.0803	1.049	0.33	0.0261
	0.83						0.0511			−0.57	1.51
0.7	1.5	1.425	0.884	σ_p	1.42	0.0213	0.0329	0.0802	1.038	0.35	0.0294
	0.83						0.0506			−0.58	1.28
0.7	1.5	1.434	0.894	σ_p	1.43	0.0214	0.0335	0.0800	1.040	0.35	0.0271
	0.67						0.0507			−0.59	1.44
0.7	1.5	1.3418	0.7962	σ_p	1.32	0.0246	0.0318	0.0785	1.028	0.36	0.0299
	1.87						0.0503			−0.59	1.13
0.7	1.5	1.2319	0.6860	σ_p	1.18	0.0318	0.0316	0.0787	1.041	0.33	0.0315
	3.0						0.0506			−0.60	0.87
0.7	1.5			σ_p^*	1.43	0.0003	0.0330	0.0803	1.051	0.34	0.0265
	0.67			fom_{peak}			0.0503			−0.58	1.32
0.7	1.5			σ_p^*	1.44	0.0003	0.0327	0.0796	1.044	0.33	0.0263
	0.83			fom_{peak}			0.0504			−0.58	1.33
0.7	1.5			σ_p^*	1.45	0.0003	0.0324	0.0804	1.032	0.34	0.0286
	1.87			fom_{peak}			0.0503			−0.60	1.30
0.7	1.5			σ_p^*	1.50	0.0003	0.0328	0.0813	1.032	0.32	0.0292
	3.0			fom_{peak}			0.0510			−0.62	1.29
0.7	1.2			σ_p^*	1.43	0.0003	0.0331	0.0795	1.050	0.31	0.0251
	0.83			fom_{peak}			0.0504			−0.56	1.39
0.2	1.2	1.778	1.008	σ_p	1.53	0.0307	0.0391	0.0982	1.046	0.64	0.0353
	0.83						0.0562			−0.66	1.79

$R_1 = \sum ||F_{obs}| - |F_{calc}|| / \sum |F_{obs}|$; $wR_2 = \{\sum[w(F_{obs}^2 - F_{calc}^2)^2] / \sum[w(F_{obs}^2)^2]\}^{1/2}$; $S = \{\sum[w(F_{obs}^2 - F_{calc}^2)^2]/(n-p)\}^{1/2}$, where *n* = number of reflections, *p* = number of refined parameters; $\Delta\rho_{max,min}$ = maximum and minimum difference density (e Å^{-3}); weights $w = 1/[\sigma^2(F_{obs}^2) + (aP)^2 + bP]$, where $P = (F_{obs}^2 + 2F_{calc}^2)/3$.

estimated for area-detector data, as was shown earlier by Zhurov *et al.* (2008). In refinement of data up to a resolution of 0.77 Å, usually the model errors are substantial so that a weighting scheme is essential. Using the square root of the refinement weights instead of the σ values, the normal probability plots behave much better. It follows that the estimated standard deviations of *EVAL15* as such contribute little to the weights. However, large values of the parameters in the weighting scheme are an indication that errors in the integrated intensities are substantially larger than what is expected from the σ values. The weights in the refinement for high-resolution structures in programs like *JANA* (Petricek *et al.*, 2000) and *XD* (Koritsanszky *et al.*, 2003), where the model errors are small, are taken to be $1/\sigma^2$. A correct estimation of the standard deviations would be profitable in such cases. We believe that the use of fom_{peak} in combination with a scaling program like *SADABS* will give reliable standard deviations. Note that fom_{peak} can be lower than one, meaning that the σ values are overestimated. This is caused by the use of too large a gain value. This can be corrected for by multiplying σ with fom_{peak} or by applying the error model in *SADABS*.

4. Conclusions

In this paper, we show that it is possible to make an *ab initio* prediction of reflection profiles as found in X-ray diffraction area-detector data. The *EVAL15* profile prediction method needs only a modest number of physically realistic parameters to simulate reflection profiles. We have shown that high-quality profiles are obtained, even in more demanding cases such as fine sliced data, $K\alpha_1/K\alpha_2$ splitting and overlapping reflections. Moreover, deviation of the profiles from what is expected on the basis of the physical parameters gives insight into unusual crystal properties or instrumental peculiarities. The simulated profiles are successfully applied in a profile fitting analysis to obtain accurate integrated intensities. It is relatively easy to include additional properties of the crystal (like anisotropic mosaic spread and lattice distortion) or of the instrument (like focusing mirrors or newly developed detectors) into the ray tracing simulation. This fully flexible approach has the potential to solve many difficult diffraction problems. *EVAL15* has the ability to work with multiple lattices (several **R** matrices) and can include overlapping neighbour reflections (from the same or from different lattices) in the least-squares procedure.

APPENDIX *A*
Distributions

When certain parameters are required to have specific non-uniform distributions care has to be taken to ensure unbiased sampling. This applies, for example, to the wavelength distribution within the spectrum or to the distribution of mosaic orientations. If

$$p(y) = \frac{dx}{dy} p(x) \tag{12}$$

is the required distribution and $p(x)$ represents a uniform distribution of deviate x, it follows that

$$F(y) = \int p(y)\, dy = x. \tag{13}$$

The transformed deviate $y(x) = F^{-1}(x)$ has the required distribution (Press *et al.*, 1986). For instance, a one-dimensional Gaussian distribution can be obtained from $y = \mathrm{erf}^{-1}(x)$. Press *et al.* also describe how a two-dimensional Gaussian distribution can be obtained. Random points with Cartesian coordinates (v_1, v_2) are selected inside a circle with square radius $R = v_1^2 + v_2^2$ and transformed to normal deviates $y_1 = [-2\ln(R)/R]^{1/2} v_1$ and $y_2 = [-2\ln(R)/R]^{1/2} v_2$, which are

distributed according to a Gaussian and represent a radial coordinate along a one-dimensional section through a two-dimensional Gaussian distribution. For obtaining mosaic distributions the coordinates y_1 and y_2 are transformed to spherical coordinates giving the polar and azimuthal angles over which the central $\mathbf{S}_0$ has to be rotated to arrive at $\mathbf{S}_m$. In the case of anisotropic mosaic spread, the y_1 coordinate is multiplied by $(\mu + \mu_{aniso}\sin\delta)$, where δ is the angle between $\mathbf{S}_0$ and $\mathbf{A}$. $\mathbf{A}$ is the anisotropic mosaic axis and μ_{aniso} is the anisotropic mosaic spread. Subsequently, y_1 and y_2 are rotated around $\mathbf{S}_0$ such that y_1 lies along the normal $\mathbf{S}_0 \times \mathbf{A}$. The result is that the mosaic distribution is stretched in the direction perpendicular to the anisotropic mosaic axis.

In a similar way a two-dimensional Lorentzian distribution can be obtained. In this case taking

$$y_1 = \left| \left(\frac{1}{2R^{1/2}} \right)^2 - \frac{1}{4} \right|^{1/2} \frac{v_1}{R^{1/2}} \qquad (14)$$

and proceeding in a similar way gives a two-dimensional Lorentzian distribution (Fig. 12).

APPENDIX *B*
Point-spread function

The point spread of a detector can be conveniently described by a pseudo-two-dimensional Lorentzian. It is based on a normal one-dimensional Lorentzian where the variable is replaced by two Cartesian variables. Integration of this function to infinity only converges when the determinant of the Jacobian matrix for transformation of the Cartesian to polar coordinates is included, which is $1/(x^2 + y^2)^{1/2}$. In addition, we have included a factor $(\gamma/2)^2$ to prevent the denominator from becoming zero. The resulting point-spread function is

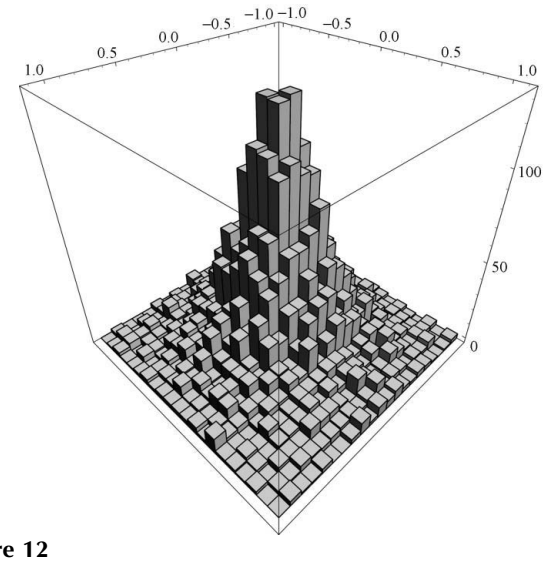


Figure 12
A two-dimensional Lorentzian distribution obtained from uniform sampling of random coordinates ($v1$, $v2$) and transformation to the corresponding coordinates ($y1$, $y2$.) A histogram of ($y1$, $y2$) is plotted. The standard deviation σ was chosen to be 1.0. Graphics produced using *Mathematica*.

$$\mathrm{PSF}(x, y) = \frac{\gamma}{4\pi\left[(x^2 + y^2) + (\gamma/2)^2\right]^{3/2}}. \qquad (15)$$

x and y are the distances in the horizontal and vertical directions measured from the centre of the impacted pixel. The intensity accumulated in a pixel is thus spread over neighbouring pixels at a point (x, y) away from its centre. It is wrong to assume that the contribution to a target pixel depends only on the distance from its centre to the centre of the source pixel. In fact, this function has to be evaluated as an integral over the surface of the whole target pixel. The integral is given by

$$\frac{1}{2\pi}\tan^{-1}\left\{ \frac{(2/\gamma)\,xy}{\left[(\gamma/2)^2 + x^2 + y^2\right]^{1/2}} \right\} \qquad (16)$$

and the four corners of the pixel are taken as the integration limits.

The authors thank Dr Rob Hooft of Bruker AXS for fruitful discussions, and Dr Albert Duisenberg and Dr Martin Lutz for critically reading the manuscript. The financial support of The Netherlands Technology Foundation STW, project number UPC 6148, is gratefully acknowledged.

References

Abrahams, S. C. & Keve, E. T. (1971). *Acta Cryst.* A**27**, 157–165.
Arndt, U. W. (1986). *J. Appl. Cryst.* **19**, 145–163.
Betteridge, P. W., Carruthers, J. R., Cooper, R. I., Prout, K. & Watkin, D. J. (2003). *J. Appl. Cryst.* **36**, 1487.
Bourgeois, D., Moy, J. P., Svensson, S. O. & Kvick, Å. (1994). *J. Appl. Cryst.* **27**, 868–877.
Bourgeois, D., Nurizzo, D., Kahn, R. & Cambillau, C. (1998). *J. Appl. Cryst.* **31**, 22–35.
Bruker (1998). *SAINT*. Version 4.0. Bruker AXS Inc., Madison, Wisconsin, USA.
Chrzanowski, L. S. von, Lutz, M. & Spek, A. L. (2007). *Acta Cryst.* C**63**, m283–m288.
Diamond, R. (1969). *Acta Cryst.* A**25**, 43–55.
Duisenberg, A. J. M. (1992). *J. Appl. Cryst.* **25**, 92–96.
Duisenberg, A. J. M., Kroon-Batenburg, L. M. J. & Schreurs, A. M. M. (2003). *J. Appl. Cryst.* **36**, 220–229.
Ford, G. C. (1974). *J. Appl. Cryst.* **7**, 555–564.
Greenhough, T. J. & Helliwell, J. R. (1982*a*). *J. Appl. Cryst.* **15**, 338–351.
Greenhough, T. J. & Helliwell, J. R. (1982*b*). *J. Appl. Cryst.* **15**, 493–508.
Helliwell, J. R., Ealick, S., Doing, P., Irving, T. & Szebenyi, M. (1993). *Acta Cryst.* D**49**, 120–128.
Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.
Koritsanszky, T., Howard, S. T., Richter, T., Macchi, P., Volkov, A., Gatti, C., Mallinsson, P. R., Farrugia, L., Su, Z. & Hansen, N. K. (2003). *XD*. Free University of Berlin, Germany.
Leslie, A. G. W. (1999). *Acta Cryst.* D**55**, 1696–1702.
McCandlish, L. E., Stout, G. H. & Andrews, L. C. (1975). *Acta Cryst.* A**31**, 245–249.
Milch, J. R. & Minor, T. C. (1974). *J. Appl. Cryst.* **7**, 502–505.
Nash, J. C. (1990). *Compact Numerical Methods for Computers: Linear Algebra and Function Minimalization*, ch. 3, 2nd ed. Bristol: Adam Hilger.
Nelder, J. A. & Mead, R. (1965). *Comput. J.* **7**, 308–313.
Nonius (1999). *COLLECT*. Nonius BV, Delft, The Netherlands.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Oxford Diffraction (2008). *CrysAlis*. Oxford Diffraction Ltd, Abingdon, Oxfordshire, UK.
Petricek, V., Dusek, M. & Palatinus, L. (2000). *JANA2000*. Institute of Physics, Czech Academy of Sciences, Prague, Czech Republic.
Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.
Phillips, W. C., Stewart, A., Stanton, M., Naday, I. & Ingersoll, C. (2002). *J. Synchrotron Rad.* **9**, 36–43.
Popov, A. N. & Bourenkov, G. P. (2003). *Acta Cryst.* D**59**, 1145–1153.
Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1986). *Numerical Recipes*. Cambridge University Press.
Ren, Z. & Moffat, K. (1995). *J. Appl. Cryst.* **28**, 461–481.
Schreurs, A. M. M. (1998). *VIEW*. Utrecht University, The Netherlands.
Schreurs, A. M. M. (1999). *PEAKREF*. Utrecht University, The Netherlands.
Schreurs, A. M. M. (2007). *ANY*. Utrecht University, The Netherlands.
Sheldrick, G. M. (1996). *SADABS*. University of Göttingen, Germany.
Sheldrick, G. M. (2005). *CELL_NOW*. University of Göttingen, Germany.
Sheldrick, G. M. (2008). *Acta Cryst*. A**64**, 112–122.
Wolfram Research (2005). *Mathematica*. Version 5.2. Wolfram Research Inc., Champaign, IL, USA.
Xian, X., Schreurs, A. M. M. & Kroon-Batenburg, L. M. J. (2010*a*). *J. Appl. Cryst.* In preparation.
Xian, X., Schreurs, A. M. M. & Kroon-Batenburg, L. M. J. (2010*b*). *J. Appl. Cryst.* In preparation.
Zhurov, V. V., Zhurova, E. A. & Pinkerton, A. A. (2008). *J. Appl. Cryst.* **41**, 340–349.