

Reduction of two-dimensional small- and wide-angle X-ray scattering data

Peter Boesecke

Received 16 August 2006

Accepted 9 January 2007

ESRF, 6 rue Jules Horowitz, BP220, 38043 Grenoble Cedex, France. Correspondence e-mail: boesecke@esrf.fr

At the beamlines ID01 and ID02 of the European Synchrotron Radiation Facility in Grenoble, France, position-sensitive detectors for time-resolved small- and wide-angle X-ray scattering experiments are in use. The applied data reduction method has never been described comprehensively. This article outlines the parametrization of the raw data and introduces the programs developed for this purpose. Data reduction in the sense of this article means all steps between detector readout and normalization to absolute scattering intensities. This includes all corrections that can be made without any specific knowledge of the sample, *e.g.* detector dark-image correction, division by a flat-field and intensity normalization. Processed data are either two- or one-dimensional. Optionally, statistical errors can be propagated through the calculations.

© 2007 International Union of Crystallography
Printed in Singapore – all rights reserved

1. Introduction

The small-angle X-ray scattering (SAXS) geometry on pinhole cameras like the beamlines ID01 and ID02 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, is relatively simple. A two-dimensional detector is placed perpendicular to the primary beam and detects the scattered radiation from a sample positioned in front of the detector. The geometrical description becomes more complex for wide-angle X-ray scattering (WAXS) experiments, when several detectors are used at the same time or when they are inclined. Quantitative analysis needs additional data, so called metadata. In addition, a method is desired to estimate and propagate statistical errors. This article explains the scattering geometry used at the ESRF beamlines ID01 and ID02 and introduces the program package *Saxs* that has been developed for quantitative analysis of SAXS and WAXS data.

The program package allows processing of sequences of one- or two-dimensional small- and wide-angle scattering data. It is available on request from the author and from the scientific software site of the ESRF (ESRF, 2006). The package includes a manual that describes the data format and the use of the routines.

The routines were originally designed to process time-resolved two-dimensional scattering data from gas-filled detectors at ID02 (Bösecke & Diat, 1997). Since this time they have been used and regularly updated.

The SAXS routines are command-line oriented and can be included into shell scripts for batch processing or can be included as subroutines into other programs. A command line history facilitates the creation of scripts. The package does not contain a graphical interface. To display data or to create masks graphically, other programs can be used, *e.g.* *EDFPLOT* or *FIT2D* (see ESRF, 2006). *EDFPLOT* is a display program that includes most of the features described below, *e.g.* the use of reference systems.

The SAXS routines perform tasks that typically arise during or after data acquisition. Similar to the CCP13 program *BSL* (Daresbury SRS, 2006), they allow mathematical operations on scattering patterns or between scattering patterns. The SAXS routines can read

BSL files without prior conversion. In addition, they support the storage of metadata like pixel size, wavelength, history *etc.* This facilitates automatic data processing.

For this purpose a file format has been chosen that can store additional information together with the image data. At the time of creation, the ESRF data format (EDF) was introduced and had the required features. Metadata are written as keyword value pairs into an ASCII header. It can be viewed with common text editors. The binary image data are appended to the end of the ASCII header. A commonly accepted definition of this format is missing. Details are given in the supporting documents available with the package.

It is not intended to propose a storage format for small-angle scattering data. The concept described here is independent of a specific format. A comparison of file formats has been made by Homan *et al.* (2001).

2. Description

A hierarchical set of parameters describes a well defined scattering geometry. Optionally, the routines allow the propagation of statistical errors using variances, only start values need to be given. Typical applications are detector corrections and intensity normalizations that have been described in several publications (Bösecke & Diat, 1997; Barna *et al.*, 1999; Narayanan *et al.*, 2001), *e.g.* dark-image subtraction, division by a flat-field image, normalization to absolute scattering intensities expressed in number of scattered photons per steradian to number of incident photons (1/sterad), projections in reciprocal space and azimuthal integrations. Each routine allows affine coordinate transformations to map input and output images. A set of predefined reference systems is available that can be used to select the most appropriate coordinate system for calculation. For instance, reference systems that use the beam center as the origin are useful for processing scattering data, while for detector specific corrections it is adequate to use the pixel coordinates of the original detector image. With the choice of an appropriate reference system it is possible to use the same routine to subtract detector dark patterns

(reference system Region) or to subtract two scattering patterns taken at slightly different wavelengths, e.g. during an anomalous scattering experiment (reference system Saxes). The reference system Saxes is also useful for combining scattering patterns taken at different sample-to-detector distances.

The Saxes package does not include a spatial distortion correction for arbitrarily shaped input pixels. A fast distortion correction that is based on the method described by Hammersley *et al.* (1994) is implemented in a separate program package.

2.1. Scattering geometry

Fig. 1 shows a general SAXS/WAXS scattering configuration with a rotated detector that is indicated by the rectangular frame on the left side. The sample is positioned in the origin of the laboratory coordinate system with axes x_1 , x_2 and x_3 , where x_1 is usually horizontal and x_2 is usually vertical. Axis x_3 is always pointing against the primary beam. The axes form a right-handed orthogonal coordinate system. The primary beam comes from the right. It is indicated with an undulated curve surrounded by a rectangular parallelepiped to indicate its cross section A. For simplicity, the sample is plate-like and is measured in transmission.

The scattering geometry is constructed stepwise from a standard configuration, in which the detector is placed at a distance L from the sample and oriented perpendicular to the primary beam. In this case the primary beam hits the detector in the point of normal incidence (C) and the axes f_1 and f_2 of the detector coincide with the geometrical axes x_1 and x_2 in the laboratory system (raster configuration 1). If the detector axes do not match an adequate raster orientation needs to be chosen (see below).

The center (C) is defined as the point from where the detector normal points to the sample. It is therefore also called point of normal incidence (PoNI). Attention, C is generally different from the position of the primary beam on the detector. Both points are only identical if the detector is perpendicular to the primary beam.

The geometry of rotated detectors is constructed from the standard configuration by three subsequent rotations in the laboratory coordinate system: first rotation ρ_1 (DetectorRotation_1) around axis x_1 , second rotation ρ_2 (DetectorRotation_2) around axis x_2 and third rotation ρ_3 (DetectorRotation_3) around axis x_3 . The point of normal

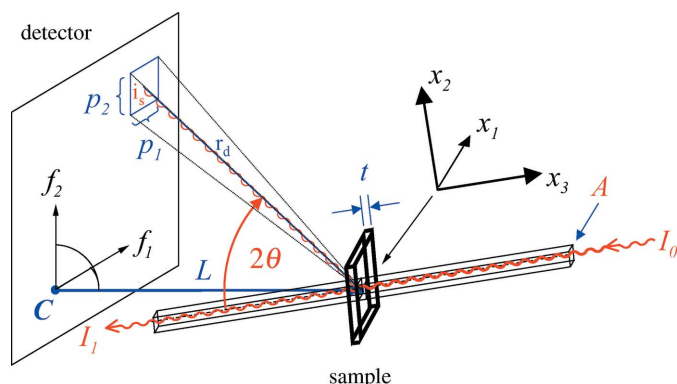


Figure 1 Scattering geometry (metadata keywords in parentheses): axes of the laboratory coordinate system x_1 , x_2 , x_3 , with origin at the sample position, axes of the detector coordinate system f_1 , f_2 , pixel sizes p_1 (PSize_1), p_2 (PSize_2), point of normal incidence C (Center_1, Center_2), sample to detector distance L (SampleDistance), number of photons in the incident beam I_0 (Intensity0), number of photons in the transmitted beam I_1 (Intensity1), sample thickness t , cross section of the beam A, distance between a detector pixel and the scattering volume r_d , number of photons scattered into a detector pixel i_s , scattering angle 2θ .

incidence (C) and the distance L between sample and detector are kept constant during the rotations. The corresponding rotation matrices are

$$\begin{aligned} \mathbf{R}_1 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \rho_1 & -\sin \rho_1 \\ 0 & \sin \rho_1 & \cos \rho_1 \end{pmatrix}, \\ \mathbf{R}_2 &= \begin{pmatrix} \cos \rho_2 & 0 & \sin \rho_2 \\ 0 & 1 & 0 \\ -\sin \rho_2 & 0 & \cos \rho_2 \end{pmatrix}, \\ \mathbf{R}_3 &= \begin{pmatrix} \cos \rho_3 & -\sin \rho_3 & 0 \\ \sin \rho_3 & \cos \rho_3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned} \quad (1)$$

In the standard configuration the laboratory coordinate x of a detector pixel is given by

$$x = \begin{pmatrix} f_{N1} \\ f_{N2} \\ -L \end{pmatrix} \quad (2)$$

The detector coordinates f_{N1} , f_{N2} are the distances on the detector relative to the point of normal incidence (C) (reference system Normal, see below). The third component $-L$ is the position of C relative to the sample. Thus, the laboratory coordinate x' of a rotated detector is

$$x' = \mathbf{R}_3 \mathbf{R}_2 \mathbf{R}_1 x \quad (3)$$

The scattering angle 2θ and the azimuth of the scattered beam can be found by expressing x' in polar coordinates.

2.2. Metadata describing the experiment

The standard metadata keywords used for the description of the geometry are listed in Table 1. Standard keywords are automatically updated during all calculations once they have been set.

The keyword raster orientation describes the orientation of the image axes with respect to the directions in the laboratory space. It becomes necessary when scattering patterns of several detectors are combined. For mathematical convenience, the origin of an image in standard configuration (raster orientation 1) is in the lower left corner. Here, axis 1 is horizontal and axis 2 vertical. In all other orientations the image axes are either flipped (indicated by a minus sign) or the horizontal (h) and vertical (v) directions are exchanged. Table 2 lists all eight orientations of two-dimensional images. The

Table 1 Standard keywords describing the scattering geometry.

Default values are given in square brackets.

Keyword	Unit	Comment
Dim_1, Dim_2	pixel	Number of pixels along axis 1 [0] and axis 2 [1]
RasterOrientation	-	Raster orientation number [1]
Offset_1, Offset_2	pixel	Spatial array offsets [0]
BSize_1, BSize_2	pixel	Pixel size relative to size of unbinned pixel [1] (binning size)
Center_1, Center_2	pixel	Point of normal incidence (PoNI)
PSize_1, PSize_2	m	Pixel size in metres
SampleDistance	m	Distance between sample and point of normal incidence
WaveLength	m	X-ray wavelength
DetectorRotation_1, DetectorRotation_2, DetectorRotation_3	rad	Rotations of the detector plane in the laboratory system [0]
ProjectionType	-	Saxs or Waxs [Saxs]

Table 2

Raster orientation numbers (h = horizontal, v = vertical, - indicates antiparallel orientation).

Orientation of image		
Axis 1	Axis 2	Raster orientation number
h	v	1
-h	v	2
h	-v	3
-h	-v	4
v	h	5
v	-h	6
-v	h	7
-v	-h	8

abbreviation (h, -v) after raster orientation 3 means that axis 1 is horizontal and that axis 2 is vertically flipped, *i.e.* the displayed image would appear upside down when the raster configuration is ignored. This numbering can be extended recursively to higher-dimensional data.

Table 3 lists additional standard keywords for title, start time and a dummy value. Undefined or masked pixels are set to the dummy value, *e.g.* the area of the beamstop or the result of an invalid mathematical operation like a division by zero. It should be a number outside the value range of the image and must be different from zero. Table 4 lists keywords that are needed for absolute calibration of scattering intensities.

Keyword values can be set or changed whenever it is appropriate. For most of them default values are used. The SAXS programs accept mathematical expressions and units can be appended with an underscore. However, an early definition together with an early (online) processing helps to avoid experimental errors.

These lists of keywords are not exhaustive. New keywords can be introduced on an *ad-hoc* basis as long as they do not collide with standard keywords and as long as they can be clearly identified, *e.g.* ESRF_ID01_PHI.

2.3. Coordinates and reference systems

For a unique description it is necessary to use well defined pixel coordinates f . Pixel coordinates are real numbers. As convention the left (lower) edge of an array has the coordinate 0.0 and the right (upper) edge of the n th pixel the coordinate n (see Fig. 2). For clarity, all axis indices are omitted in this section.

Table 5 lists the available reference systems. Each reference system defines an affine transformation of a pixel coordinate f to a reference system coordinate f' . The transformations are calculated by using the geometrical metadata described in Table 2. All transformations can be written as

$$f' = fPs + Off, \quad (4)$$

Ps and Off are internal variables that depend on the metadata of each image. They can be calculated by rearranging the equations in Table 5. The intensities of the transformed pixels are mapped to each other according to their overlap. The intensities are either integrated or averaged.

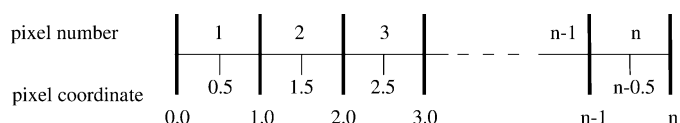


Figure 2

Definition of the pixel coordinate f along a single array axis.

Table 3

Additional standard keywords.

Keyword	Unit	Comment
Dummy	-	Dummy value
DDummy	-	Range around Dummy
Time	-	Start time of exposure with fractions of s , ISO 8601 type
Title	-	Title string

Table 4

Keywords describing beam intensity and exposure time.

The intensities should ideally be measured in number of photons integrated during the exposure time, but relative numbers are also useful if the monitors are not calibrated.

Keyword	Unit	Comment
Intensity0	(see caption)	Integrated beam intensity before the sample during the exposure time
Intensity1	(see caption)	Integrated beam intensity after the sample during the exposure time
ExposureTime	s	Exposure time

Table 5

Reference systems: metadata keywords are shown in bold.

The indices of the keywords **Offset**, **BSize**, **PSize** and **Center** are omitted. The transformations are applied to all axes independently. f is the pixel coordinate, the transformed coordinates have a subscript according to the reference system.

Reference system	Applied affine coordinate transformation
Array	$f_A = f$
Image	$f_I = \text{Offset} + f$
Region	$f_B = (\text{Offset} + f) \times \text{BSize}$
Real	$f_R = (\text{Offset} + f) \times \text{PSize}$
Center	$f_C = \text{Offset} - \text{Center} + f$
Normal	$f_N = (\text{Offset} - \text{Center} + f) \times \text{PSize}$
Saxs	$f_S = \text{Normal}(\text{Offset}, \text{Center}, \text{PSize}, f) / \text{SampleDistance} / (\text{Wavelength} / \text{nm})$

The basic detector corrections, like background subtraction and flat-field division, are pixel by pixel operations. They can be applied without knowledge about the geometrical setup of the experiment. In this case the reference system **Region** should be used.

For scattering experiments real distances from the point of normal incidence C are required. In the **Normal** reference this distance is given in metres. A useful extension is the **Saxs** coordinate f_S . It is the **Normal** coordinate f_N divided by the sample distance L and the wavelength λ

$$f_S = \frac{f_N}{L\lambda} \text{ nm}. \quad (5)$$

To convert the scale into a convenient range, it is multiplied with nanometre (10^{-9} m). The modulus f_S of the **Saxs** coordinate is related to the modulus s of the scattering vector which is defined as

$$s = \frac{2 \sin \theta}{\lambda}, \quad (6)$$

where θ is half the scattering angle and λ the wavelength of the radiation. By replacing f_N/L by $\tan 2\theta$ equation (5) can be rewritten as

$$f_S = \frac{\text{nm}}{\lambda} \tan 2\theta. \quad (7)$$

From equations (6) and (7) the relation between f_s and s can be calculated:

$$f_s = \frac{nm}{\lambda} \tan \left[2 \arcsin \left(\frac{\lambda s}{2} \right) \right]. \quad (8)$$

At small scattering angles 2θ the Saxe coordinate f_s can be interpreted as a commonly used approximation for s

$$f_s = \frac{\tan 2\theta}{\lambda/nm} \approx \frac{2\theta}{\lambda/nm} \approx \frac{2 \sin \theta}{\lambda/nm} = s/nm^{-1}. \quad (9)$$

The Saxe reference system maps exactly scattering patterns to each other that have been measured at different distances L . It is an approximation for the comparison of scattering patterns measured at slightly different wavelengths λ .

2.4. Projection types

To display WAXS and SAXS patterns together in a common plane, e.g. on the beamline computer, a projection of the Ewald sphere is required. A useful projection has been described earlier (Urban *et al.*, 2003). It consists of a rotation of the scattering vector s from the Ewald sphere to the tangential plane that crosses the origin perpendicular to the incoming wave (see Fig. 3). The azimuthal angle and the length of the scattering vector are conserved during the transformation. This projection is called Waxe projection. It corresponds to a virtual detector plane in reciprocal space. The distortions that are introduced into the detector scattering pattern by this transformation are relatively small, up to scattering angles of 90° .

In the Waxe projection, the point C marks the origin of the reciprocal space and the wavevector k_0 is perpendicular to it. Practically, the detector pattern is mapped in such a way that the distance from C is proportional to the modulus of s when the Saxe reference system (Table 5) is used:

$$f_s^{\text{Waxe}} = s/nm^{-1}. \quad (10)$$

The superscript Waxe indicates that the reference system Saxe is applied to a Waxe projected image. Apart from losses due to the mapping of pixel intensities both presentations are equivalent and can be mutually converted. To distinguish both representations the keyword ProjectionType is used. It has the value Waxe for the projected pattern and Saxe for the original pattern. At small scattering angles both projections look nearly identical due to the insignificant curvature of the Ewald sphere. The Waxe projection

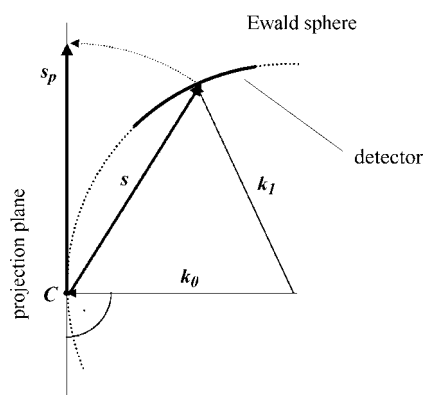


Figure 3 Projection of scattering vector s from the Ewald sphere (dotted) to s_p in the projection plane. The projection plane is perpendicular to k_0 . C is used as the origin of the reciprocal space. The detector is indicated as a continuous line on the Ewald sphere. k_0 is the wavevector of the primary beam, k_1 the wavevector of the elastically scattered beam. The modulus of the wavevectors is $1/\lambda$.

allows a representation of WAXS and SAXS data in the same plane and an analysis of the azimuthal intensity distributions of WAXS data like SAXS data.

It should be taken into account that the Waxe projection is only an aid for visualization and that it only represents a real cut through the reciprocal space for isotropically scattering samples.

3. Propagation of statistical errors

The error propagation calculation is based on variance calculations. The general procedure is described in standard text books, e.g. Blobel & Lohrmann (1998). The basic relations are repeated here

$$1 = \int f(\mathbf{x}) d\mathbf{x} \quad (11)$$

$$E_i = \langle x_i \rangle = \int x_i f(\mathbf{x}) d\mathbf{x} \quad (12)$$

$$V_{ii} = \text{var}(x_i) = \int (x_i - \langle x_i \rangle)^2 f(\mathbf{x}) d\mathbf{x} \quad (13)$$

$$V_{ij} = \text{cov}(x_i, x_j) = \int (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) f(\mathbf{x}) d\mathbf{x}. \quad (14)$$

The function f is the probability density of the vector x and is normalized to 1 [equation (11)]. The n components of x are the statistical variables of the experiment. The integrations are taken over the whole range of all variables. In the specific case of two-dimensional scattering data, the statistical variables x_i are pixels. The expectation value of x_i is written $\langle x_i \rangle$ [equation (12)], the variance of x_i is written $\text{var}(x_i)$ [equation (13)] and the covariance between two statistical variables x_i and x_j is written $\text{cov}(x_i, x_j)$ [equation (14)]. Variance and covariance are the diagonal and non-diagonal elements V_{ij} of the covariance matrix $V[x]$. In linear approximation, the basic rule of error propagation for a function $y(x)$ with m components can be written as:

$$\mathbf{V}[y] = \mathbf{B}\mathbf{V}[x]\mathbf{B}^T, \quad (15)$$

$\mathbf{V}[y]$ is the covariance matrix of the m -dimensional output variable y . The matrix \mathbf{B} contains the partial derivatives of the output variables y_i with respect to the input variables x_j . For the exact definition of \mathbf{B} and \mathbf{V} see Blobel & Lohrmann (1998). The size of the covariance matrix $\mathbf{V}[x]$ scales with n^2 and the derivative matrix \mathbf{B} scales with $n \times m$. Because of the required huge matrix dimensions in case of images a straightforward application of the error propagation rule [equation (15)] seems to be impossible.

As simplification it is assumed that all images have the same size ($n \times n$) and that all input images are uncorrelated with respect to each other. In the case of pixel by pixel operations equation (15) can be split into the following sum over N input images:

$$\mathbf{V}[y] = \sum_{k=1}^N \mathbf{B}_k \mathbf{V}_k [\mathbf{x}_k] \mathbf{B}_k^T. \quad (16)$$

All sub-matrices with the subscript k have the same dimension as the output image and the \mathbf{B}_k matrices are diagonal. However, the variance matrices \mathbf{V}_k with the statistical variables x_k are not necessarily diagonal because of pixel by pixel correlations in a single image. Owing to the diagonality of the \mathbf{B}_k matrices, the non-diagonal elements of \mathbf{V}_k the covariance values, do not mix during pixel by pixel operations.

It is assumed that the detector noise consists of the statistical noise of the detected photons plus an electronic noise of the detector. The

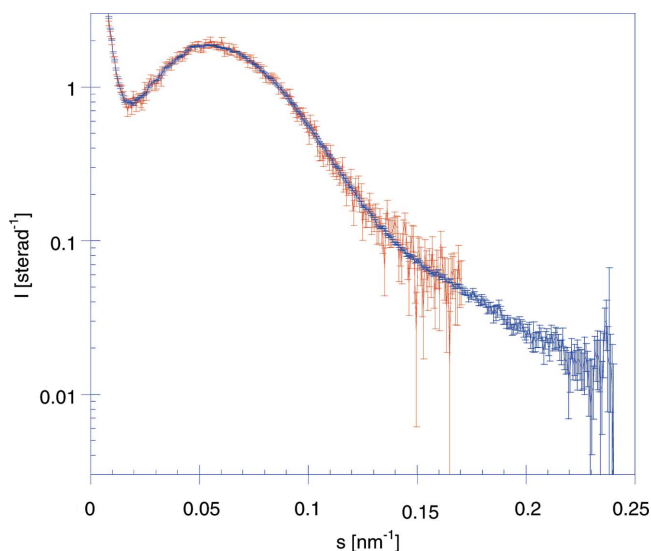


Figure 4
Comparison of an azimuthal integration of a normalized scattering pattern (in units of sterad^{-1}) with a 10 pixel wide projection (shorter curve). The length of the error bars is 2σ . The curves have been calculated from a 512×512 large simulated scattering curve (3 mm thick crystallized polyethylene). The following parameters have been used: pixel size $200 \times 200 \mu\text{m}^2$, $I_0 = 10^{10}$ photons, transmission 0.75, wavelength 0.1 nm, distance 3 m, detector quantum efficiency of 1. Poisson statistics have been used for the scattered photons and an r.m.s. detector noise of 1 analog to digital unit (ADU).

variances of both contributions are added. The main assumptions are that cross correlations between pixels are caused by the spatial resolution of the detector and distant pixels and pixels of different images are uncorrelated. Under these assumptions it should be possible to bin images in such a way that the pixel sizes become larger than the spatial resolution of the detector and the cross correlations between neighbouring pixels become negligible. In this case all non-diagonal terms of the variance matrices \mathbf{V}_k should vanish. Equation (15) reduces to a calculation of variances:

$$\text{var}(y_i) = \sum_{k=1}^n \left(\frac{\partial y_i}{\partial x_k} \right)^2 \text{var}(x_k). \quad (17)$$

For this case the variances of the input variables x_i should be determined, e.g. by analyzing a time series of images or by statistical analysis of a single image. If, afterwards, the pixel sizes are changed their variances are recalculated as if they were uncorrelated. However, it should be taken into account that the variances are only valid for pixel sizes where cross correlations with neighbouring pixels are negligible.

In this way statistical error propagation is implemented in the *Saxs* program package. Practically, the variance values are stored in a

separate variance array that contains the variance of each pixel. The file size is usually doubled.

Fig. 4 illustrates the error calculation capabilities. The input is a simulated two-dimensional scattering pattern with simulated noise. Two curves are shown in the figure, the longer curve with less noise is an azimuthal average of the whole pattern, the shorter curve is a projection of 10 pixels.

4. System requirements

The *Saxs* package can be compiled with the GNU C compiler GCC (GCC, 2006) on UNIX, LINUX and on MacOS systems. It can also be compiled on Windows systems by using *CYGWIN* (Cygwin, 2006).

5. Conclusion

The *Saxs* program package is in use at beamline ID01 and ID02 at the ESRF, at ID02 since 1995. The publication lists of these beamlines are available online. Practically all SAXS and anomalous SAXS experiments with two-dimensional detectors that have been done at these beamlines since 2002 have at least taken advantage of this package. It has been integrated into an online correction program that runs automatically during data acquisition. The user can select the required corrections before exposure. The program package uses the ESRF data format as a practical method to store series of two-dimensional X-ray scattering patterns together with metadata. The geometrical description, however, is independent of the data format.

I thank all colleagues and users of the beamlines ID01 and ID02 that have given to me a positive feedback and who have helped to develop and improve the computer programs.

References

- Barna, S. L., Tate, M. W., Gruner, S. M. & Eikenberry, E. F. (1999). *Rev. Sci. Instrum.* **70**, 2927–2934.
- Blobel, V. & Lohrmann, E. (1998). *Statistische und numerische Methoden der Datenanalyse*. Wiesbaden: B. G. Teubner Verlag.
- Bösecke, P. & Diat, O. (1997). *J. Appl. Cryst.* **30**, 867–871.
- Cygwin (2006). *CYGWIN*, <http://www.cygwin.com/>.
- Daresbury SRS (2006). *BSL, Scientific Software*, <http://srs.dl.ac.uk/ncd/computing>.
- ESRF (2006). *Scientific Software*, <http://www.esrf.eu/UsersAndScience/Experiments/TBS/SciSoft>.
- GCC (2006). *GNU C compiler*, <http://gcc.gnu.org/>.
- Hammersley, A. P., Svensson, S. O. & Thompson, A. (1994). *Nucl. Instrum. Methods Phys. Res. A*, **346**, 312–321.
- Homan, E., Konijnenburg, M., Ferrero, C., Ghosh, R. E., Dolbnya, I. P. & Bras, W. (2001). *J. Appl. Cryst.* **34**, 519–522.
- Narayanan, T., Diat, O. & Bösecke, P. (2001). *Nucl. Instrum. Methods A*, **467**, 1005–1009.
- Urban, V., Panine, P., Ponchut, C., Boesecke, P. & Narayanan, T. (2003). *J. Appl. Cryst.* **36**, 809–811.