

New two-dimensional data treatment software for
small-angle scattering

G rard P py

Received 16 August 2006

Accepted 26 February 2007

Institute of Solid State Physics, SzFKI, POB 49, H1525-Budapest, Hungary, and Laboratoire L on Brillouin (LLB), CEA Saclay, 91191 Gif sur Yvette CEDEX, France. Correspondence e-mail: gerard.pepy@cea.fr

A new program is presented which performs data treatment in one, two or three dimensions. While one-dimensional data treatment can be performed easily by many programs (canSAS, EMBL, ISIS, ILL, NIST), few deal efficiently with more dimensions. Indeed, specific attention has to be paid to the selection of the relevant data. Their display and models are relatively complex. This new program has been developed according to the needs of small-angle scattering users, but is not limited to these fields. Its original purpose was to model forward anisotropic scattering and diffuse scattering typically produced by large structures such as polymers, aggregates, self-assembly systems or micellar solutions. It is also suited to modelling Bragg scattering. With time, many filter configurations (rectangles or sectors, with possible symmetries and display along various coordinates) and many model functions (centred or not, possibly with cofactors, with Cartesian, polar, or three-dimensional coordinates) have been added. Models are fitted by the steepest descent model with χ^2 as the minimization function. The software is written in Fortran with the *PGPLOT* graphics package. It runs with the Windows operating system.

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

1. Introduction

Sometimes complex pictures are obtained while performing small-angle (and large-angle) scattering, for instance if the sample shows anisotropic or diffuse scattering or Bragg peaks: the data-treatment problem becomes two-dimensional; meanwhile most available programs [canSAS (Ghosh, 2006), EMBL (Svergun, 2006), ISIS (Heenan, 2004), ILL (Ghosh & Rennie, 2007), NIST (Munter, 2004)] are one-dimensional. Moreover, it may be of interest to consider simultaneously files obtained while varying the temperature or rotating an angle: a third dimension comes in. Such problems are very often solved by the restriction of the fit to narrow bands, each of which may be considered as one-dimensional. This is reasonable only if the measured intensity changes slowly along the neglected dimension (Len *et al.*, 2007); in any case most pixels are just thrown away even though they have valuable information. Correlation problems may appear for parameters appearing in distinct fits [for instance, different fits along different directions of reciprocal space usually do not produce the same value for $I(0)$, the intensity at zero scattering vector]. Multidimensional fitting is needed.

The representation of the data, the way they are extracted by filters, how the fit is handled and finally how the results are saved are all complex. For the new program presented here, named *PXY*, a lot of attention has been given to these matters and several solutions have been tried by users before the most efficient has been selected. Alternative ways are often kept in order to test various user feelings. Two features, simplicity and unicity, are specific to this program:

Simplicity: from the start we assumed that a single program with a central menu with many functionalities was easiest to handle with a minimum of learning by a casual user.

Unicity: it was a crucial requirement that every action leading to the fit results should be displayed in the working window and that this same window could be saved and printed.

This software is written in Fortran, with the *PGPLOT* graphics package from Caltech (Pearson, 2002), and runs with the Windows operating system. The *PXY* program displays only two windows on the computer screen: one is a *PGPLOT* graphics window where the mouse is active, the other one provides short guidelines for keyboard input.

Data treatment consists of five steps: reading the data files, displaying the data, applying filters, fitting models and recording the results. We shall review these steps and get a general feeling for the program by processing a case study. Then we shall look at some examples that exhibit some interesting characteristics.

2. Processing a case

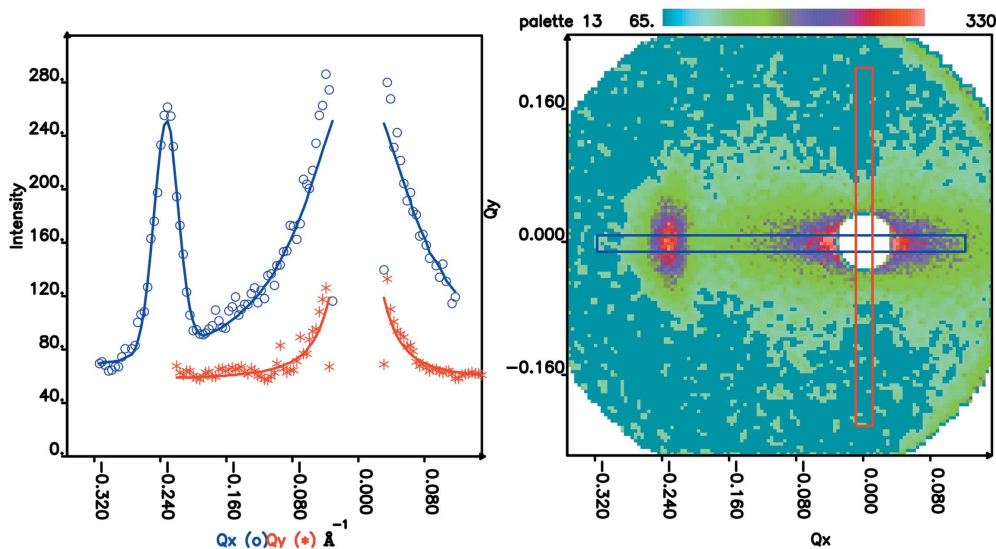
The *PGPLOT* graphics window (Fig. 1) is divided into four parts: the upper left (Fig. 1a) is devoted to the data file(s) input and gather its characteristics, the lower right (Fig. 1b) maps the data, the lower left (Fig. 1b) displays the intensity from filtered pixels and the upper right (Fig. 1a) contains information about the fit. Submenus appear according to the functionality selected in the general menu. The fifth step, recording the results, consists of a copy of the graphics window as a PostScript or GIF file (without the menu).

In all the subsequent paper it will be assumed that the multi-detector is flat and perpendicular to the incoming beam, with wavevector \mathbf{k}_i whose modulus is $2\pi/\lambda$, where λ is the incoming radiation wavelength. Every detector pixel determines a scattered beam direction with wavevector \mathbf{k}_f and scattering angle θ between \mathbf{k}_i and \mathbf{k}_f . As the scattering is assumed to be elastic, \mathbf{k}_i and \mathbf{k}_f have the same magnitude. The scattering vector \mathbf{Q} is defined as $\mathbf{Q} = \mathbf{k}_f - \mathbf{k}_i$. Its modulus is $Q = (4\pi/\lambda) \sin(\theta/2)$. Its coordinates Q_x , Q_y , Q_z are best defined in an orthogonal frame with the X and Y directions in the

PXY7 written by G Pépy release Dec 2006 2/2/2007 12h 22min 16s image /tmp/0802.d-122216.ps history /tmp/start121352

File	xy0802.dat	wavelength	4.0	Mask +	10 120 63 67	Start	End	$\text{\AA}^{-1}/\text{Intensity}$	\AA																														
Label	xy0802	Distance	2000.0		88 92 10 118	I_B	66.00	57.347	272.4																														
General title	LCPolym	X origin	89.9	1 smoothing < 100																																			
Sample	H/D PA-OCH3	Y origin	65.5	cell by cell calibration																																			
Date	18/02/1999	beam catcher	9.0	<table border="1"> <tr> <td>Function 1</td> <td>type</td> <td>Gauss/Lorentz</td> </tr> <tr> <td>I</td> <td>180.00</td> <td>179.366 852.0</td> </tr> <tr> <td>X</td> <td>31.00</td> <td>30.693 -0.23251</td> </tr> <tr> <td>Y</td> <td>65.00</td> <td>65.927 0.00168</td> </tr> <tr> <td>width</td> <td>5.00</td> <td>5.315 0.02087 47.91</td> </tr> <tr> <td>width</td> <td>9.00</td> <td>8.344 0.03277 30.52</td> </tr> <tr> <td>Function 2</td> <td>type</td> <td>Lorentz/Lorentz</td> </tr> <tr> <td>I</td> <td>350.00</td> <td>245.547 1166.3</td> </tr> <tr> <td>width</td> <td>15.00</td> <td>18.605 0.07306 13.69</td> </tr> <tr> <td>width</td> <td>5.00</td> <td>4.923 0.01933 51.73</td> </tr> </table>						Function 1	type	Gauss/Lorentz	I	180.00	179.366 852.0	X	31.00	30.693 -0.23251	Y	65.00	65.927 0.00168	width	5.00	5.315 0.02087 47.91	width	9.00	8.344 0.03277 30.52	Function 2	type	Lorentz/Lorentz	I	350.00	245.547 1166.3	width	15.00	18.605 0.07306 13.69	width	5.00	4.923 0.01933 51.73
Function 1	type	Gauss/Lorentz																																					
I	180.00	179.366 852.0																																					
X	31.00	30.693 -0.23251																																					
Y	65.00	65.927 0.00168																																					
width	5.00	5.315 0.02087 47.91																																					
width	9.00	8.344 0.03277 30.52																																					
Function 2	type	Lorentz/Lorentz																																					
I	350.00	245.547 1166.3																																					
width	15.00	18.605 0.07306 13.69																																					
width	5.00	4.923 0.01933 51.73																																					
Time	15:14:40	detector cell	5.0																																				
Acquisition time	752	maximum	369	Chi2	0.3761	Dim 2 variables X,Y																																	
Monitor	21053	average	70	Total	+ 502 0.1773	FIT: cycles 7 Npar 9																																	
		Minimum	16		- 443 0.1988	well finished																																	
		Transmission	1.00																																				
		Subtract	1.00 xy0804.dat																																				
		Calibration	1.00 xy0805.dat																																				

(a)



(b)

Figure 1

The four parts of data treatment. The sample was a side-chain liquid-crystal polymer oriented by a horizontal magnetic field (PAXY spectrometer, LLB, CEA Saclay; Noirez, 1999). After a fit all four parts are gathered in a single A4 image, stored in a file. On the left of part (a) one can find the characteristics of the data and correction files. On the right of part (b) the intensity mapping around the beam catcher hole exhibits anisotropic forward scattering by the main chain (half of the polymer main chains were deuterated); on the left side is a Bragg peak, the signature of the sample smectic phase. Two rectangles which represent filters can also be seen. The projections of the pixels inside these filters on the relevant axes are shown on the left of part (b) (circles and stars for the horizontal and vertical rectangles, respectively). The solid line corresponds to a fit by two-dimensional functions, a centred Lorentz–Lorentz function for the forward scattering and a non-centred Gauss–Lorentz function for the Bragg peak. The characteristics of the fit are displayed on the right of part (a).

detector plane, X along the horizontal direction and Y along the vertical direction (Fig. 1b).

2.1. Reading data files

Currently it is possible to read data files with formats from various laboratories: small-angle neutron scattering (SANS) spectrometers from Institut Laue Langevin (ILL), Laboratoire Léon Brillouin (LLB) and Budapest Neutron Center (BNC), Hahn Meitner Institut (HMI), Forschungszentrum Jülich (including USANS), SAND at the Argonne National Laboratory, YuMO at the Frank Laboratory for Neutron Physics, small-angle X-ray scattering (SAXS) beamlines ID01 and ID02 at the European Synchrotron Research Facility (ESRF), and some Bruker machines. We also implemented general formats like FITS (an astronomy format for charge-coupled-device cameras) and any image format (like BMP or TIFF) after conversion to ASCII. The *PXY* program was initially developed for SAXS, therefore the data file input may include the classical corrections files for calibration, reference and background. Then *PXY* immediately calculates the uncertainties. Alternatively, the sample data file may be provided with its own uncertainty table (SAND at Argonne). *PXY*

may handle several files for comparison, display and calculation; from four for 1250×1250 pixel files up to 16 if their size is less than 128×128 pixels.

The main menu (Fig. 2a) is divided into chapters corresponding to the above four parts. From it one may navigate either directly to simple functionalities (reset picture . . .) or to submenus (pick a filter, image intensity range, make a fit . . .).

2.2. Mapping the data

The intensity (I) map associates a colour palette (among 25 predefined palettes) to 84 intensity levels. The number and range of these levels may be adjusted and one may associate the palette to functions of I , $\ln(I)$ or \sqrt{I} in order to enhance details. It is possible to bore holes (rectangular, circular, elliptic), to remove or to restore given data.

2.3. Filtering the data

Selecting pixels in filters has a double role:

- (i) only pixels inside the filters will be fitted to a model;

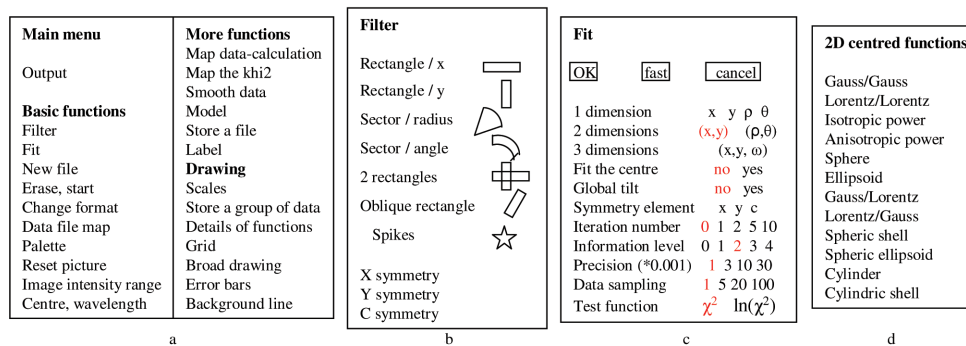


Figure 2

After loading the data files, a first map of the intensity (Fig. 1b) is automatically displayed and the main menu appears (a). From it one may modify the intensity map or go to the filter selection. The filter menu is shown in (b). Several geometries are available: horizontal, vertical or oblique rectangles, with display *versus* the X and Y axis, or the wavevector, respectively, or sectors with display *versus* the radius, ρ , or rings *versus* the angle, θ . For the fit procedure it is possible to define the fit characteristics (dimension number, iteration number, information level *etc.*) from a general menu (c). Model functions, centred or not centred on the main beam, may be chosen from a submenu (d).

(ii) the intensity of the pixels inside a filter will be displayed through a projection on an axis corresponding to the filter (X , Y , ρ , θ).

Indeed, not all pixels are always relevant to the problem. Most of all, when starting to analyse data it is useful to limit the number of data to accelerate the process and to get a fast understanding of it, therefore narrow filters are useful at the beginning while larger filters including all significant data are used at a later stage. It is equally important to use filters with adequate geometry. For this purpose various filters are available (Fig. 2b): horizontal and vertical rectangles, sectors with display *versus* the radius, crowns with display *versus* the angle, oblique rectangle or cross, as in Fig. 1. A 'spike' filter is under development to match star-like scattering occurring in foams or polycrystals. All these filters may benefit from symmetry operations *versus* the X or Y axis, or central symmetry (Fig. 2b). In the case of sectors (or rings) as many as six may be used, with or without central symmetry, periodically spaced or individually defined.

It is easier to consider a graph of a one-dimensional function than any other representation. Moreover, it is nearly impossible to compare calculated curves with data points otherwise. Therefore the program includes a one-dimensional representation of the pixels inside a filter. This is obtained through a projection of the pixels on the relevant coordinate axis. For instance, the pixels inside a rectangle elongated along X will be displayed *versus* X , and the pixels inside a sector will be displayed along ρ (or θ according to the filter type). The pixels inside each of the two rectangles of a 'cross' filter will be displayed *versus* X and Y , respectively.

It should be stressed that this one-dimensional representation is a mere aid to the user. In any case, each pixel is individually fitted, most of the time by a two-dimensional function. After a fit the calculated points are projected and displayed in the same way as the data points in order to ease the evaluation of the fit. However, especially at a preliminary stage of a fit process, one may use a one-dimensional function in a narrow filter, but this should not be the normal procedure for treating data exhibiting rapidly varying scattering functions (Len *et al.*, 2007). On the other hand, one may use three-dimensional models to fit data in a set of two-dimensional data files depending on a parameter like temperature or pressure.

It is also possible to store in a file the intensities of the projected pixels seen in the display as a one-dimensional set. Users can then use any data treatment or display program (*e.g.* *ORIGIN*, *EXCEL*, *KALEIDAGRAPH* and so on).

Many other functions are available. For instance, it is possible to write a title or a comment on the graphics page. It is also possible with the mouse to select a pixel and get its position in reciprocal -space

coordinates, with its intensity, and the average intensity over its next neighbours, or over the next and the second nearest neighbours. It is also possible to view the contents of a filter with complementary images: one resulting from a projection along X and the other *versus* Y (or ρ and θ).

2.4. Fitting the data

The fitting process is the well known 'steepest descent' of a minimization function (Bevington, 1969; Press *et al.*, 1992). The most common function is

$$\chi^2 = \frac{1}{N-p} \sum_{i=1}^{i=N} \left(\frac{I_i - Y_i\{P\}}{\Delta I_i} \right)^2, \quad (1)$$

where I_i is the intensity in pixel i , N is the number of data points, p is the number of free parameters; $\{P\}$ is the set of parameters, Y_i is the calculated intensity in pixel i and ΔI_i is the uncertainty over I_i . If the random variable I_i follows a normal probability law, $\chi^2 = 1$, which is a good test for the fit.

The first step in the fit general menu (Fig. 2c) is to choose the dimensionality. The main other selections are the number of iterations, the precision (for the test ending the fit) and the level of information during the fit. Once the general characteristics of the fit are defined, another menu allows the introduction of the background and model functions which can be combined linearly or with cofactors (two cofactors are available: magnetic modulation in the case of a saturated ferromagnetic sample and a correlation factor). Selecting the function box opens a specific menu (Fig. 2d). The model functions may be centred on the main beam or not centred; in this case they include two more parameters for their origin. Many functions are predefined, such as Gauss–Gauss, Gauss–Lorentz, the scattering by a sphere, an ellipsoid, a cylinder and so on. The mixed function Gauss–Lorentz means that it is a Gaussian *versus* the first variable and a Lorentzian *versus* the second variable. Most of these functions include the possibility to apply a dispersion to one parameter. The built-in dispersions are Gaussian, square, triangle, log-normal and Schultz–Zimm. Finally, one runs the fit and hopefully reaches a good result. If this is not the case, several functionalities may help. For instance, it is possible to draw a map of the minimized function, χ^2 , over two variables. Probably the most useful functionality is the map of the difference between the calculation and the data. The differences are calibrated for each pixel by the mean-square deviation in order to build a meaningful picture.

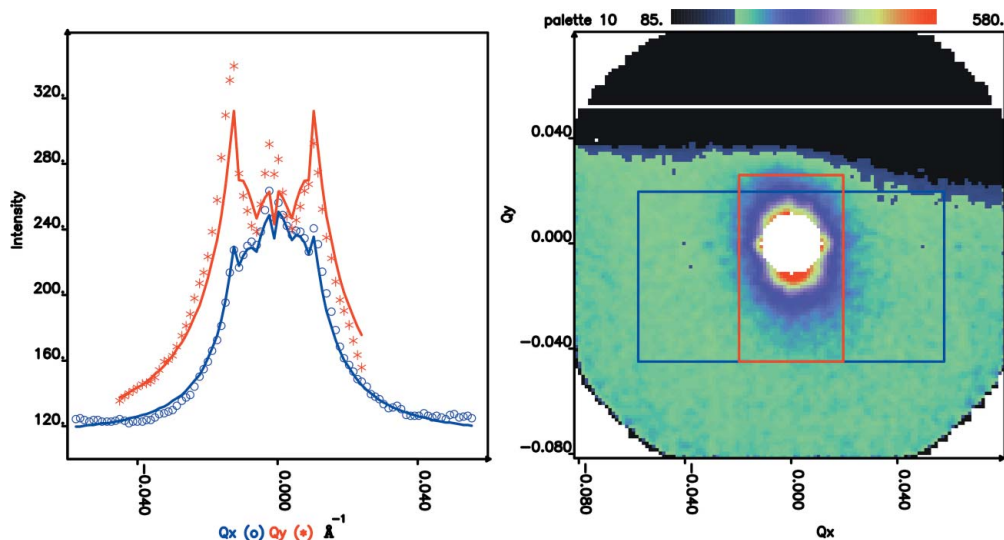


Figure 3
 This measurement concerns a side-chain liquid-crystal polymer (PAXY spectrometer, LLB, CEA Saclay; Noirez & Lapp, 1997). The sample is sheared in a device which produces the shadow in the upper part. Because of the shearing the scattering is tilted. Therefore a general tilt has been applied to the functions in the model, a Lorentz-Lorentz function for the polymer and a power law to take into account the scattering from a few catalytic particles. However, this fit was not satisfactory. A two-dimensional representation of the difference between the data and the model (Fig. 4) helped in understanding the problem.

2.5. Recording the results

Once the fitting is satisfactory, the ‘output’ functionality in the main menu allows the graphics page with all its information (without the menus) to be recorded as a PostScript or a GIF file (Fig. 1). Along with this, a text file is created with the full history of how this page was reached. It is possible to run the ‘history’ file up to the step where the page was recorded at a later time, and to resume the data treatment from this point. Note that the ‘output’ includes a short menu allowing some of the four elements of the page to be selected and the typography of the lines and symbols to be changed.

3. Examples

3.1. Difficulty with power law

On Fig. 1b we had selected narrow rectangles in order that the projected display would not be too different to the well known shape of Lorentzian (scattering by a polymer coil) and Gaussian (resolution-widened Bragg peak) curves. For the case in Fig. 3, the rectangles have been tailored to include as many of the relevant pixels as possible, while avoiding a perturbed part of the picture. The fit was not considered to be good enough. Indeed, the difference calculation data in Fig. 4 revealed a mis-setting of the central position. This effect was particularly strong because of the presence in the model of a power law; close to the centre it is very steep and sensitive to a centre-position error. In this case, it was enough to use functions with fitted centre parameters, instead of the functions centred on the assumed beam position, to reach an excellent fit.

3.2. A complex model

The data file displayed in Fig. 5 shows the SAXS by nanochannels in a track-etched membrane. It has more than 1.5 million pixels. However, thanks to the sample orientation, the Ewald sphere cut of the scattering function reduces the number of interesting pixels. It was mandatory to perform a fit including the two dimensions in order

to verify the model. Some samples required a small dispersion over the channel radius along the beamline resolution. Let us mention that the model depends strongly on the angle of the channel axis *versus* the main beam. It has been possible to fit simultaneously two files with very different pictures corresponding to two different angular positions of the sample: a first step to three-dimensional fitting.

3.3. A convenient display

The picture in Fig. 6 is complex because it is the superposition of two anisotropic functions. In this case, the best approach to the problem implied a polar representation. Most of the pixels are selected for the fit. They belong to four sectors (with their centrosymmetric partners) which allows a pertinent representation of the anisotropy and an appreciation of the fit quality.

3.4. Data evolution

The case considered in Fig. 7 was designed to explore the evidence for small changes (like anisotropy) by fitting several files simultaneously. It is possible to provide each file with the same set of model functions. The parameters of subsequent files may be linked by a linear relation to the parameters of the previous file. Thus, it is possible to fit some parameters independently for each file and to fit others with linear constraints between subsequent files.

4. Developments

Throughout the development of this program, it was a constant concern to keep it user-friendly. Therefore the immediate projects are to write menus as Visual Basic objects to improve their convenience and to give users the possibility to link their own entry format and model function.

Further improvements will include the implementation of the MINUIT fitting kernel from the European Organization for Nuclear Research (CERN) and the extension of the functionalities to grazing-

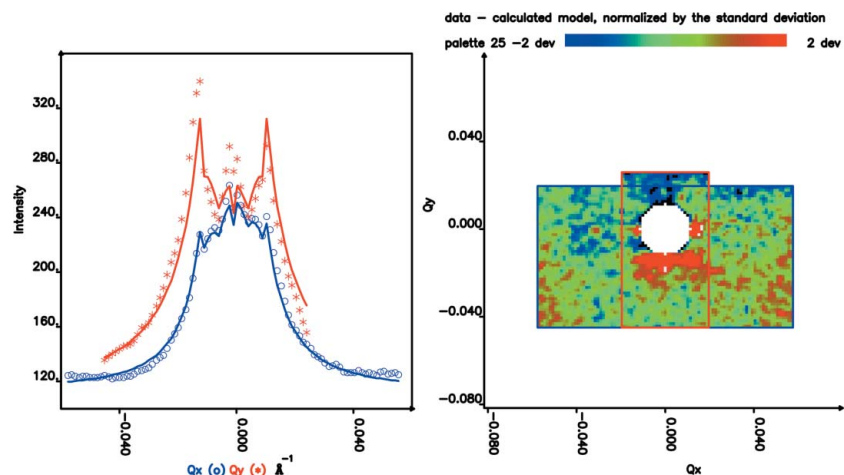


Figure 4

The difference map (data – model) is calculated for the pixels inside the filters shown in Fig. 3. It is calibrated by the local mean-square deviation in order to allow comparison between pixels. Here a vertical mis-setting appears clearly. Its origin is a small error in the centre position. Using non-centred functions (with the centre coordinates as variables) instead of the centred ones allowed an excellent fit to be obtained.

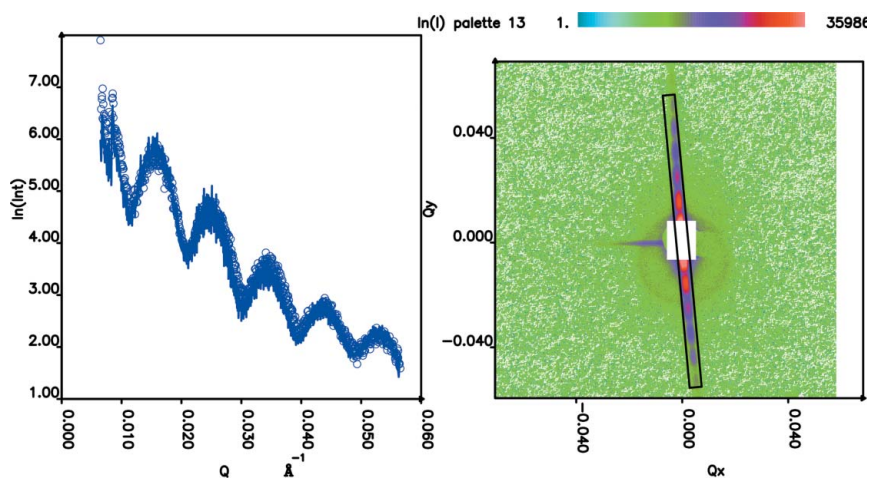


Figure 5

SAXS by nanochannels in a track-etched amorphous polycarbonate membrane (Pépy *et al.*, 2007). The polymer foil was 10 μm thick. It had been irradiated with 3×10^8 ions cm^{-2} and etched for 3 min in 5 M NaOH at 333 K. The channel axis is off by about 13° with respect to the X-ray beam (ID01 beamline, ESRF, Grenoble). The significant scattering (and filter) is oblique because it was not possible to orient the channel axis inside the horizontal plane. The model is the form factor of a cylinder with radius $R = 33.8$ nm (note that the scattering is varying very fast in both directions, which requires a two-dimensional fit; the display on the left is a projection of the pixel intensity *versus* the scattering vector modulus Q , the best possible representation, while somewhat blurred because no one-dimensional representation exists).

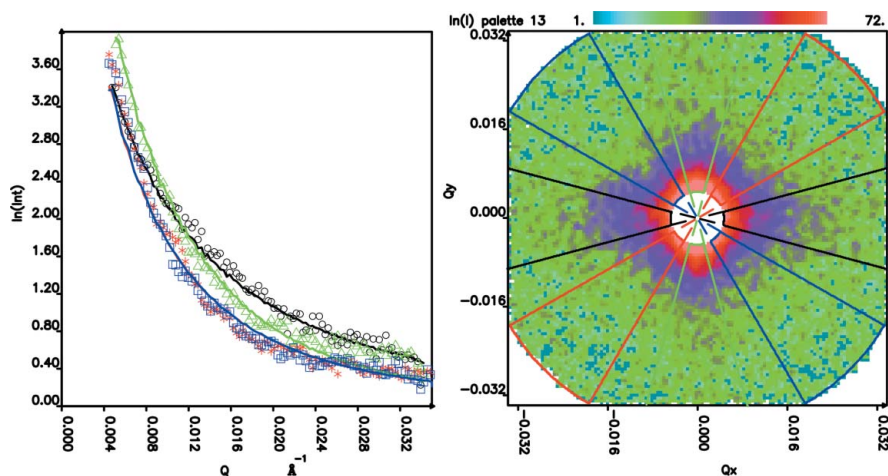


Figure 6

The scattering function of a mixture of a nematic liquid crystal and a reticulated polyacrylate (PAXY spectrometer, LLB, CEA Saclay; Gautier *et al.*, 2003) is a Lorentz–Lorentz function. A few impurities remained in the sample and worsened the experiment because they are anisotropic and oriented by the mesogenic phase. An anisotropic power law was thus necessary to take into account their contribution. The display of azimuthal averages in four different centrosymmetric sectors is very useful for checking the anisotropy and the fit quality even when the intensity is poor.

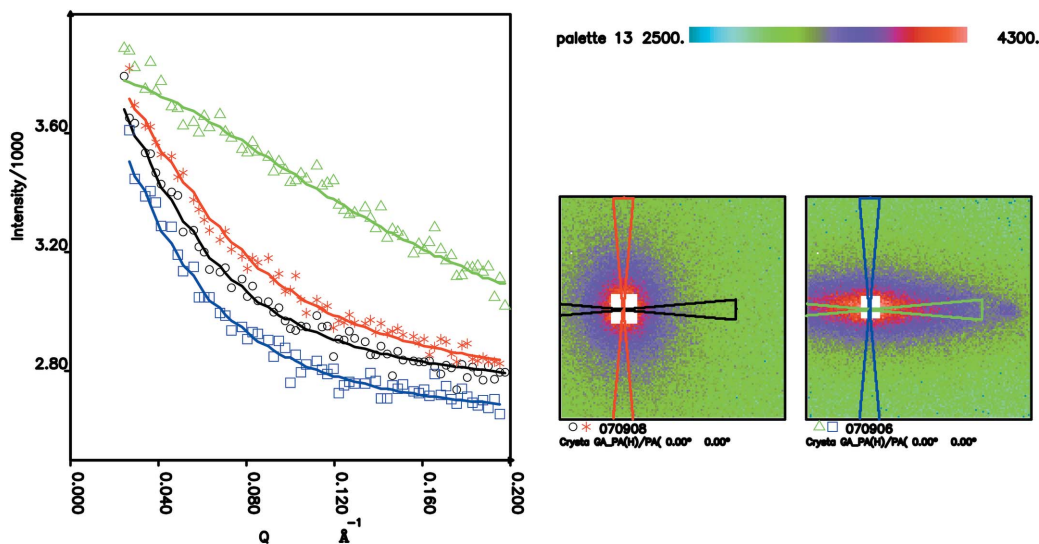


Figure 7
 The scattering function of this side-chain liquid crystal polymer (D11 spectrometer, ILL; Mendil, 2006) changes *versus* temperature. The purpose is to fit two files simultaneously with the same model and common parameters to check the sensitivity of this method to very small variations (obviously in this test example the difference was big).

incidence SAXS and reflectometry, each with its own style. Suggestions and contributions are welcome.

5. Conclusion

The *PXY* program models include many of the functions used in one-dimensional fitting, combined in such a way as to perform two-dimensional fits. Special care has been taken to help the program user with a variety of useful displays and functionalities.

The author thanks C. Trautmann and B. Schiedt (GSI Darmstadt) who provided the nanochannel sample and L. Noirez (LLB, CEA Saclay) who provided all the liquid-crystal polymer examples. Moreover, the constant interest and encouragements of L. Noirez had a fundamental influence in the development of this program.

References

Bevington, P. R. (1969). *Data Reduction and Error Analysis for the Physical Sciences*. New York: McGraw-Hill.
 Gautier, P., Brunet, M., Grupp, J., Noirez, L. & Anglaret, E. (2003). *Phys. Rev. E*, **68**, 11709-1–11709-12.
 Ghosh, R. (2006). <http://www.ill.fr/lss/canSAS/>.
 Ghosh, R. & Rennie, A. (2007). http://www.ill.fr/data_treat/newsans.html.
 Heenan, R. K. (2004). *FISH*, <http://www.isis.rl.ac.uk/largescale/>.
 Len, A., Pépy, G., Rosta, L. & Harmat, P. (2007). Private communication.
 Mendil, H., Baroni, P., Grillo, I. & Noirez, L. (2006). *Phys. Rev. Lett.* **96**(7), 077801-1–077801-12.
 Munter, A. (2004). *Simulator*, <http://www.ncnr.nist.gov/resources/simulator.html>; http://www.ncnr.nist.gov/programs/sans/data/red_anal.html.
 Noirez, L. (1999). *Europhys. Lett.* **46**(6), 728–734.
 Noirez, L. & Lapp, A. (1997). *Phys. Rev. Lett.* **78**(1), 70–73.
 Pearson, T. J. (2002). *PGPLOT*, <http://www.astro.caltech.edu/~tjp/pgplot/>.
 Pépy, G., Boesecke, P., Kuklin, A., Manceau, E., Schiedt, B., Siwy, Z., Toulemonde, M. & Trautmann, C. (2007). *J. Appl. Cryst.* **40**, s388–s392.
 Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1992). *Numerical Recipes in FORTRAN77 – The Art of Scientific Computing*. USA: Cambridge University Press.
 Svergun, D. (2006). <http://www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html>.