Data Processing for the SNS EQ-SANS Diffractometer

Jinkui Zhao

Spallation Neutron Source, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

zhaoj@ornl.gov, 865 574-0411

Abstract

The scattering data on the Extended Q-Range Small Angle Scattering Diffractometer at the spallation neutron source are stored as neutron events. Each event has two parts: the time of fight of the detected neutron, and an identifier for the neutron's location on the detector. In addition, each event is associated with the source pulse from which the neutron is originated. Due to Spallation Neutron Source's (SNS) high neutron flux, the size of the stored data files can easily exceed several Gigabytes. Processing and reducing these data can be very time consuming. In the present work, we implement a data processing scheme for the EQ-SANS diffractometer that is fast, versatile, and highly automated. The data are processed directly from the event files into neutron scattering intensity versus momentum transfer data sets, or other desired formats. Speed gains are obtained by the implementation of parallel computing under the Message Passing Interface framework. In addition, the implementation allows for time-slicing of the scattering data which will enable fast time-dependency studies on the EQ-SANS instrument, such as temperature jump and fast shears. Finally, our work implements the handling of scattering data from the innovative frame-skipping operation of the EQ-SANS instrument.

1. Introduction

On a steady-state source based Small Angle Neutron Scattering (SANS) diffractometer, where the scattering data are stored as two-dimensional maps of neutron counts versus detector *x*- and *y*-axis, data processing typically involves azimuthal integration of the scattering data to produce $I(Q)$, *i.e.* scattering intensity (I) versus neutron momentum transfer (Q) [1,2]. On a time-of-flight (TOF) SANS instrument, a third, TOF-dimension is added to the scattering data. Each TOF histogram corresponds to a two-dimensional data map with a certain neutron wavelength. These histograms are processed separately and their outputs are combined to produce the final result [3-6]. At the Spallation Neutron Source [7] (SNS) at the Oak Ridge National Laboratory, a new, event-based format is introduced for storing the scattering data: for every detected neutron, it's TOF and detector pixel location are stored in an event data file. Additionally, for each neutron pulse generated at the SNS target station, its absolute time stamp, total accelerator proton charge, and correlation to the detected neutron events are stored as well. Other instrumental and experimental configurations are collected and saved in XML-formatted files [8]. In comparison to TOF-histograms, the SNS event files store more comprehensive information on scattering experiments. They allow for detailed and flexible data processing after the experiment is conducted. For example, it is possible to reprocess the scattering data with different TOF resolutions. It is also possible to split the data into time slices, facilitating the studies of time-dependent phenomena. The drawback of the event format is that processing these neutron events is less straightforward than handling histograms. In addition, due to SNS's high neutron beam flux, the event files can easily exceed many Gigabytes per experiment, which slows the data analysis process.

On the Extended Q-range Small Angle Scattering Diffractometer (EQ-SANS) [9,10], most experiments require the scattering data be reduced to $I(Q)$, similar to the needs on other SANS instruments. Initial effort to produce $I(Q)$ was to transform the neutron events to TOF-histograms [11]. These resulting intermediate histograms are then further processed in a similar way as on other TOF SANS instruments [3-6]. The whole process can be slow, inflexible, and has limited usage in time-dependent studies. In this work, we present the implementation of a new data-processing scheme that is both fast and versatile. The software is implemented in C++. It takes full advantage of the SNS event data format and processes the event files directly. Further speed gain is obtained from parallel computing implemented under the Message Passing Interface architecture [12]. The software allows full user-control of the data-handling process, which includes data correction, normalization, reduction, and slicing etc.

2. Data processing on the EQ-SANS

The data processing tasks that are implemented in the current work are discussed in this section. The algorithm of processing neutron events is presented first, followed by data slicing and frame-skipping data treatment. Corrections and normalizations for the EQ-SANS data are discussed as well.

2.1 Intensity versus momentum transfer *I*(*Q*)

We first examine the task of handling histograms similar to those used on existing TOF SANS instruments [3-6]. Figure 1 shows a schematic of the scattering geometry on the EQ-SANS diffractometer. The measured data can be viewed as a three-dimensional array of neutron counts, $N(x,y,tof)$, where x and y represent the location on the two-dimensional detector and tof is the neutron's time of flight from the neutron source (*i.e.* moderator) to the detector. Since tof (in seconds) is related to wavelength λ (in Ångstroms) by $tof = m_N L\lambda/h = L\lambda/3956$, the data array can be regarded as $N(x,y,\lambda)$ as well. h is Planck's constant, m_N is neutron's mass, and L(in meters) is the distance between the detector and the source (see Fig. 1). For most small angle scattering experiments, the scattering pattern for each TOF-histogram is symmetric around its center and azimuthal integrations are performed to reduce the data into $I(Q)$. Azimuthal integration is carried out by calculating a pixel's corresponding Q value and adding the pixel's neutron count $N(x,y)$ to $I(Q)$. The neutron's momentum transfer Q is calculated according to $Q = 4\pi/\lambda\cdot\sin\theta$, where θ is *half* the scattering angle. The resultant $I(Q)$ is further divided by the total solid angle of all contributing pixels. This normalization is required because $I(Q)$ is proportional to the sample's total scattering cross section, $d\sigma/d\omega$, and is in regard to *per unit solid angle*. Therefore, azimuthal integration in two dimensions can be written as $I(Q) = \Sigma N(x,y)/n\omega$, where n is the number of contributing pixels to the intensity at Q and ω is the per-pixel solid angle (see Fig. 1). Note that even though all pixels on the EQ-SANS detector are of the same size, ω is different for pixels at the center and on the edge of the detector, especially when the detector is near the sample. Therefore, ω should be regarded as $\omega(x,y)$ in general. Error estimate for $I(Q)$ follows that described in chapter 5 of Bevington [13] and the standard deviation σ of $I(Q)$ is calculated using $\sigma^2 = \Sigma[N(x,y)/\omega - I(Q)]^2 / (n\text{-}1)$.

The thus reduced $I(Q)$ - Q data are stored in a three-column format of (I Q σ). On a time-of-flight SANS instruments, the uncertainty in neutron momentum transfer (Q) varies with Q-values. Therefore, the reduced $I(Q)$ - Q data can also includes a fourth column δQ, which is the standard deviation of neutron's momentum transfer Q.

In the event mode, to transform the data directly into $I(Q)$, a solid angle normalization factor Ω has to be pre-calculated and stored for each detector pixel. Similar to that in azimuthal integrations discussed above, $\Omega = n\omega$, which is the total solid angle of all pixels on an azimuthal ring (see Fig. 1). For each neutron that falls onto this ring, the value $1/\Omega$ is added to the intensity at the corresponding Q. The resultant $I(Q)$ is thus the total neutron counts per unit solid angle. For error handling, the square root of the total neutron counts at this Q-value is taken as the uncertainty of $I(Q)$. For software implementation, the total neutron counts at Q are stored as C during neutron event processing. The standard deviation of I is calculated at each Q using $\sigma = C^{1/2}\,(I/C) = I/C^{1/2}$. The scaling factor I/C arises from the fact that the intensity obtained above is the total number of neutrons per solid angle, i.e. $I = C/\Omega$. Therefore, I/C is the solid angle normalization factor $1/\Omega$ that was applied to the intensity during data processing. This error-handling scheme is somewhat different from the azimuthal integration discussed above in that all the detector pixels contributing to the intensity at Q are now viewed as one single detector.

The event data processing scheme described above leaves one artificial effect in the reduced data. Because neutrons with different times of flight, or wavelengths, cover a different Q-range, the reduced data will undercount the $I(Q)$ at both the low and high Q ends. To eliminate this effect, the neutron event data can be reduced to $I(Q,tof)$. In another word, $I(Q)$ is stored for each TOF bin separately during data processing. At the end, $I(Q,tof)$ is averaged over TOF bins to produce the final $I(Q)$. This is a similar to the histogramming approach, but without the actual three-dimensional histogramming of the scattering data. The TOF bin size is chosen to correspond to neutron's pulse width on the EQ-SANS instrument, which is about 20μs/Å [10,14]. Therefore, ~500 TOF-channels is adequate for the 16.7ms wide frame on the 60Hz SNS source.

When calculating the scattering angle, a neutron's fall due to gravity after it is scattered by the sample needs to be taken into account for long wavelength neutrons and large sample-to-detector distances (SDD). The gravity drop is readily determined using $\frac{1}{2}g(SDD\cdot\lambda/3956)^2$ and the resultant value is added to neutron's vertical position y before the scattering angle 2θ is calculated. $g = 9.8\text{m}\cdot\text{s}^{-2}$ is earth's gravity constant. λ is in Angstroms. SDD and the resultant drop are in meters. On the EQ-SANS, the maximum sample-to-detector distance is ~10m. At this distance, the gravity drop is ~3mm for λ = 10Å and ~7mm for λ = 15Å neutrons respectively. These values are comparable to the pixel size of the EQ-SANS detector and can no longer be ignored.

2.2. Intensity versus *Q* in two dimensions $I(Q_x,Q_y)$

For samples with asymmetric scattering patterns, such as magnetic samples or samples with preferred alignments, neutron events need to be converted to *I vs. Q* in two dimensions, $I(Q_x,Q_y)$. On a steady-state source based SANS instrument [1,2], The two-dimensional data map from the detector $N(x,y)$ directly corresponds to $I(Q_x,Q_y)$. On time-of-flight instruments, the three dimensional data $N(x,y,tof)$ need to be integrated along the

Q_x- and Q_y-axes. Conceptually, the procedure for producing $I(Q_x,Q_y)$ is the same as that for $I(Q)$ discussed above. The difference lies with implementation details. If Q_x and Q_y bin sizes are not chosen appropriately, artificial effects such as streaks will show up in the reduced data. To avoid such binning streaks, it is also necessary to use fractional bins: since a neutron event's Q value will most likely fall in between two Q bins, a fraction of that event is assigned to each of these two Q bins. Fractional binning does not have noticeable effects on one-dimensional $I(Q)$ unless there are sharp features like diffraction peaks in the data.

2.3. Additional data formats

Beside the intensity versus neutron momentum transfer data sets, the implementation also allows data output in various other formats. Even though the data are processed directly from neutron events, it is nonetheless useful to be able to produce TOF histograms. Combining these histograms together will result into a two-dimensional map of intensity versus detector pixel, $I(x,y)$. $I(x,y)$ can be used to examine the detector performance, or to find the center of the scattering pattern. Integrating the scattering data over the detector for each TOF-histogram yields the total intensity *vs*. time of flight, $I(tof)$. For samples such as vanadium, which scatters neutrons incoherently for the most part, $I(tof)$ can be regarded as the spectrum distribution of the neutron beam at the sample position.

2.4. Data slicing for time-dependent studies

As discussed in section 1, the SNS event format allows EQ-SANS data to be split into time slices. Implementing the data slicing ability is an important and necessary step to enable time-dependent studies on the EQ-SANS instrument. The EQ-SANS event data themselves store only the time-of-flight information for the detected neutrons, which is referenced to the moment when the source pulse is produced. The absolute time stamp for the source pulse is stored in an accompanying pulse id file [8]. The pulse id file also stores the starting neutron event number for each pulse. Therefore, by correlating neutron events to their originating source pulses, the absolute time for all detected neutrons are known and the experimental data can be split into subsets. Because the proton current of SNS accelerator, and hence the neutron beam flux, varies with time, the time-sliced scattering data need to be normalized by the source flux. The accelerator proton charge for each pulse that is stored in the pulse id file is used for this purpose. Even though the temperature of the neutron moderator also affects the neutron beam flux, it is considered stable during typical neutron scattering experiments. Currently, beam monitors on SNS instruments only work in histogram mode and their data cannot be sliced for such normalization purpose.

2.5. Treatment of frame-skipping data

One of the distinctive features of the EQ-SANS instrument is that it can be operated in the innovative frame-skipping mode [10]. An example setup is the simultaneous usage of two non-continuous frames with the wavelength of 2-5.4Å and 8.9-12.3Å, which is employed for the experiment shown in Figures 2. Frame skipping gives the instrument three times the wavelength bandwidth as that of the normal, 60Hz operation. Neutrons from the two selected frames arrive at the detector adjacent to each other. However, the two frames originate from two different source pulses. To deal with the skipped frame, we first recognize the frame-skipping mode by analyzing the speeds and phases of the EQ-SANS bandwidth choppers. The time gap of the skipped frame is then added back to the second frame before data are processed further. Because the SNS moderator brightness varies strongly with neutron wavelength, the scattering data from the two frames have different count rates and need to be normalized by the incident beam flux, as discussed below.

2.6. Data correction

Two of the data correction tasks that we have implemented are dark current subtraction and detector efficiency calibration. On the EQ-SANS instrument, dark current is measured with the incident beam shutter closed while the SNS source is running. It contains the electronic noises in the detector and the ambient background around the instrument. While the ambient background is a function of the SNS source power, their exact correlation is not easy to quantify. The dark current subtraction in the current implementation is therefore normalized by the measuring time, rather than by the incident neutron beam flux. In place of or in addition to dark current, an empty beam data can be used as well. An empty beam measurement is conducted with neutron beam turned on but without a sample. Its subtraction is normalized by both the beam flux and the sample transmission, since it mainly corrects the background created by the imperfect primary beam interception by the beam stop. This background is proportional to the number of neutrons that reach the beam stop. The correction for the detector efficiency on the EQ-SANS instrument is carried out by dividing the scattering data with a detector efficiency data set generated from an isotropic scatter.

2.7. Data normalization

Incident beam normalization for EQ-SANS data implemented in the present work covers two parts: the spectrum distribution and the total flux of the neutron beam. Though they can be performed in a single step with the absolute spectrum of the incident beam, they are discussed separately here. Three different types of spectrum measurements are used on the EQ-SANS instrument. The first is from the low efficiency ^{3}He transmission beam monitor [8]. The TOF data from the monitor are first converted into flux vs. neutron wavelength and are then used to normalize the scattering data. We note that the monitor data cannot be used when the EQ-SANS is operated in the frame-skipping mode as discussed in section 2.5. The two frames overlap in their times of flight when they pass through the monitor and cannot be distinguished. The second type of spectrum measurement is the integrated data from a vanadium sample as discussed in section 2.6. The last one is to use the *in situ* detector counts behind a pinhole on the beam stop at the center of the detector [10].

For normalizing the scattering data against the total incident beam flux, either the integrated monitor counts or the SNS accelerator current is used. After these normalizations, the scattering data are then ready for further processing, such as absolute scaling, or buffer subtraction. Absolute scaling is achieved by comparing the scattering intensity with that of a standard scatter under the same measurement conditions [15].

3. Software implementation

The data processing tasks described in section 3 are implemented in C++ on the EQ-SANS Linux-cluster at the SNS. The main implementation is a C++ class, *EXPERIMENT*, which both stores and manipulates the experimental data. In the current version, the class also stores two pointers referencing to a background and a detector efficiency measurement. These additional measurements, when supplied, are automatically applied as corrections to the experimental data. Upon start, the software automatically searches for the folder where the experiments are stored. Depending on user inputs, neutron events, beam monitor and the pulse id files are all read in and processed automatically. The default output is the $I(Q)$ data. More output as discussed in section 3 can be enabled through user inputs. To allow for maximum flexibility, the program parses through different levels of configuration files and user inputs in the following order: a) Data acquisition system (DAS) configurations; b) instrument configurations; c) experiment configurations; and d) command line inputs. Overlapping input entries that are read in later will overwrite earlier ones. DAS configurations are created during the experiment by the SNS DAS. They are found in the same folder as the experimental data and have the suffixes of cvinfo.xml and runinfo.xml [8]. Some of the main entries in them are sample descriptions, run time, accelerator status, detector location, bandwidth chopper status (which defines the neutron wavelength band), and slit selections etc. The instrument configuration file stores common data reduction parameters for a batch of user experiments, such as detector resolution and masking, sample location relative to the neutron moderator (which may change depending on the sample environment used), slit installations on the slit wheels etc. By convention, it is named as *eqsans_configuration.n* in a system folder, where the extension *n* is the smallest experimental run number from which onward the stored configurations are valid. All experiments on the EQ-SANS are assigned a unique and increasing run number. The experimental configuration file, saved as *eqsans_configuration.ini* in user's current working folder, stores the same information as that in the instrument configuration file. It is used to fine tune the processing of particular experimental data sets. Finally, user inputs in the form of command line switches can be supplied to the program. The reduced data can be further processed by a data operator (*eqsans_dataop*), which performs arithmetic operations (+,-,×,and /) between $I(Q)$ data sets with appropriate error propagations according to that outlined in chapter 4 of Bevington[13].

The main speed hurdle for working with neutron events is the shear amount of data that have to be read in and processed. Incidentally, such tasks are well suited for parallel computing. The software is parallelized under the Message Passing Interface architecture [12]. Each MPI process reads in a portion of the even data using MPI's high performance parallel file IO. After the data are processed, the results from all MPI processes are then combined together. Running on 6 CPUs (2.8GHz AMD Opterons) on the EQ-SANS instrument computer, the program processes ~1 GB of neutron event data into $I(Q)$ in ~5sec. Using more CPUs in this case will not reduce the data processing time. The apparent bottleneck here is the speed of the computer hardware that moves the data from harddisk into main computer memory. Hardware upgrades to high performance distributed storage can alleviate or remove this data congestion. In the mean time, if more tasks are carried out by the software during data processing, for example producing $I(Q)$,

$I(Q_x,Q_y)$, and $I(x,y)$ etc in the same time, running on more CPUs will obviously reduce the total processing time.

4. Example usage

The program is started on a command line either by *fatcat* or by *fatcats*, the latter being the parallel version of same software. Inputs to the program are controlled by switches, each proceeded by a negative sign (-). For the simplest task of reducing a data set to $I(Q)$, only the run number switch (-*r*) is needed. Each measurement on the EQ-SANS instrument is assigned a unique run number.

Figures 2 and 3 show two EQ-SANS data sets [10] processed to $I(Q)$ and $I(Q_\mathrm{x},Q_y)$ respectively using the software presented in this work. Figure 2 illustrates the handling of frame skipping data and the time-slicing ability of the program. The experiment was collected over a 25 hour period. However, only the first hour of the data are cut out and shown. The color map in Figure 3 is produced by the co-supplied *Java*-based program, *jzg*. It demonstrates the difference between a steady-state source based SANS data set and the reduced $I(Q_\mathrm{x},Q_y)$ data on the EQ-SANS instrument. In the former case, the eight non-functioning detector tubes would have shown up as missing data on the right side of the detector. With the EQ-SANS, this missing data gap is bridged over. Data correction will smooth out the arc-like feature in the figure but is not carried out in order to illustrate this difference.

5. Summary

The fast data processing implemented on the EQ-SANS will greatly facilitate the scientific productivity of the EQ-SANS diffractometer. The programs are versatile, fast, and easy to use. Some of the main features of the implementation include the ability to enable time-dependent studies and to analyze and include instrument resolutions into the reduced data to facilitate further data analysis [17].

6. Acknowledgement

This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains, and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

References

[1] K. Ibel, (1976) *J. Appl. Crystallogr.* **9**, 296-309.

[2] C. J. Glinka, J. G. Barker, B. Hammouda, S. Krueger, J. J. Moyer, W. J. Orts, (1998) *J. Appl. Crystallogr.* **31**, 430-445.

[3] R. P. Hjelm Jr, P. A. Seeger, (1989) *Advanced Neutron Sources 1998 Proceeding*, Inst. Phys. Conf. Series. **97**,367.

[4] P. A. Seeger, R. P. Hjelm Jr, (1991) *J. Appl. Crystallogr*. **24**, 467-478.

[5] R. K. Heenan, J. Penfold, S.M. King, (1997) *J. Appl. Crystallogr.*, **30**, 1140-1147.

[6] T. Worlton, J. Hammonds, D. Mikkelson, R. Mikkelson, R. Porter, J. Tao, A. Chatterjee, (2006) *Physica B* **385–386**, 1340-1342.

[7] www.sns.gov.

[8] R. Riedel, (2007) SNS Document: 107030200-TD005-R04 (Oak Ridge, TN, USA).

[9] J.K. Zhao (2000) SNS Document: 107080100-TD0001-R00, (Oak Ridge, TN, USA).

[10] J.K. Zhao, C. Y. Gao, & D. Liu (2010) *J. Appl. Crystallogr.* **43** 1068-1077.

[11] P. F. Peterson, M. A. Reuter, (2008) SNS Document: 107030214-TD0001-R00 (Oak Ridge, TN, USA).

[12] MPICH2, www.mcs.anl.gov/research/projects/mpich2/

[13] P. R. Bevington (1969) *Data reduction and error analysis for the physical sciences*, (McGraw-Hill, New York).

[14] E. B. Iverson, P. D. Ferguson, F. X. Gallmeier, and I. I. Popova, (2002) SNS Document: 110040300-DA0001-R00 (Oak Ridge, TN, USA).

[15] G. Wignall & F. Bates, (1987) *J. Appl. Crystallogr*. **20**, 28-40.

[16] R. Gilles, U. Keiderling, A. Wiedenmann, (1998) *J. Appl. Crystallogr.* **31**, 957-959.

[17] All software programs presented in this work are available on the EQ-SANS linux cluster and will be available through the SNS neutron portal at www.sns.gov. Source codes are available upon request.

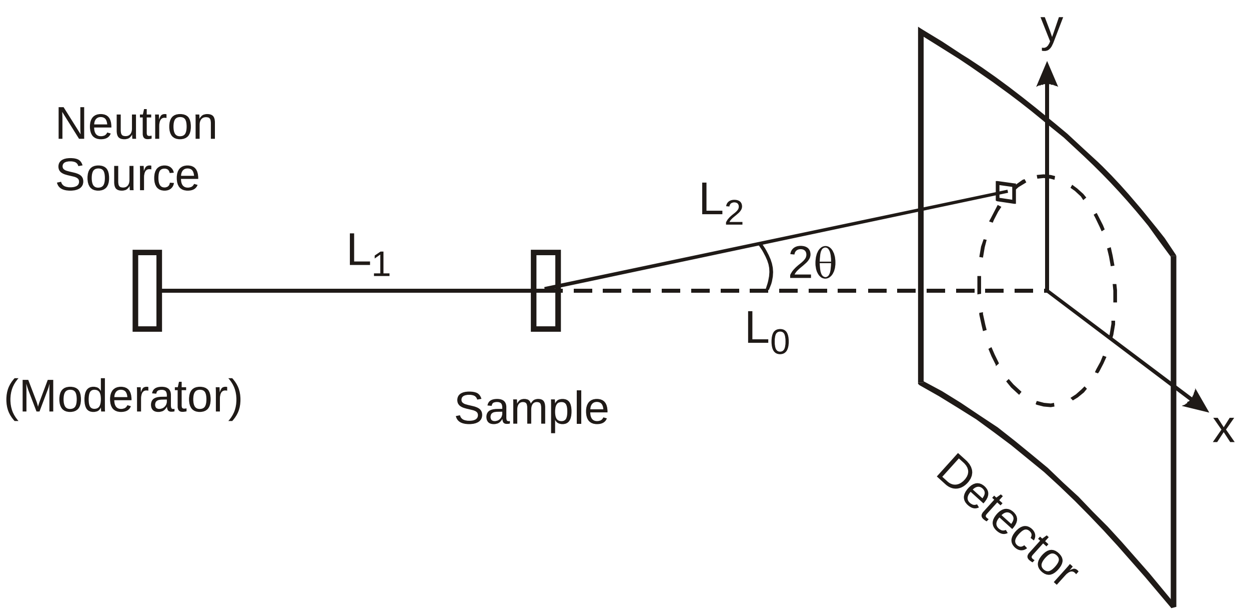

FIG 1. Schematic representation of the scattering geometry on the EQ-SANS diffractometer [10]. $L=L_1+L_0$ is the total instrument length. $L=L_1+L_2$ is the actual flight path of a scattered neutron. For simplicity, they are not distinguished in the discussions but their differences are accounted for in the software implementation. The EQ-SANS detector is slightly curved as indicated. It is made of 1m long, 8mm in diameter 1D position sensitive 3He tubes arranged on the detector arc. The scattering angle (2θ), an azimuthal ring, and a detector pixel on the ring are depicted.

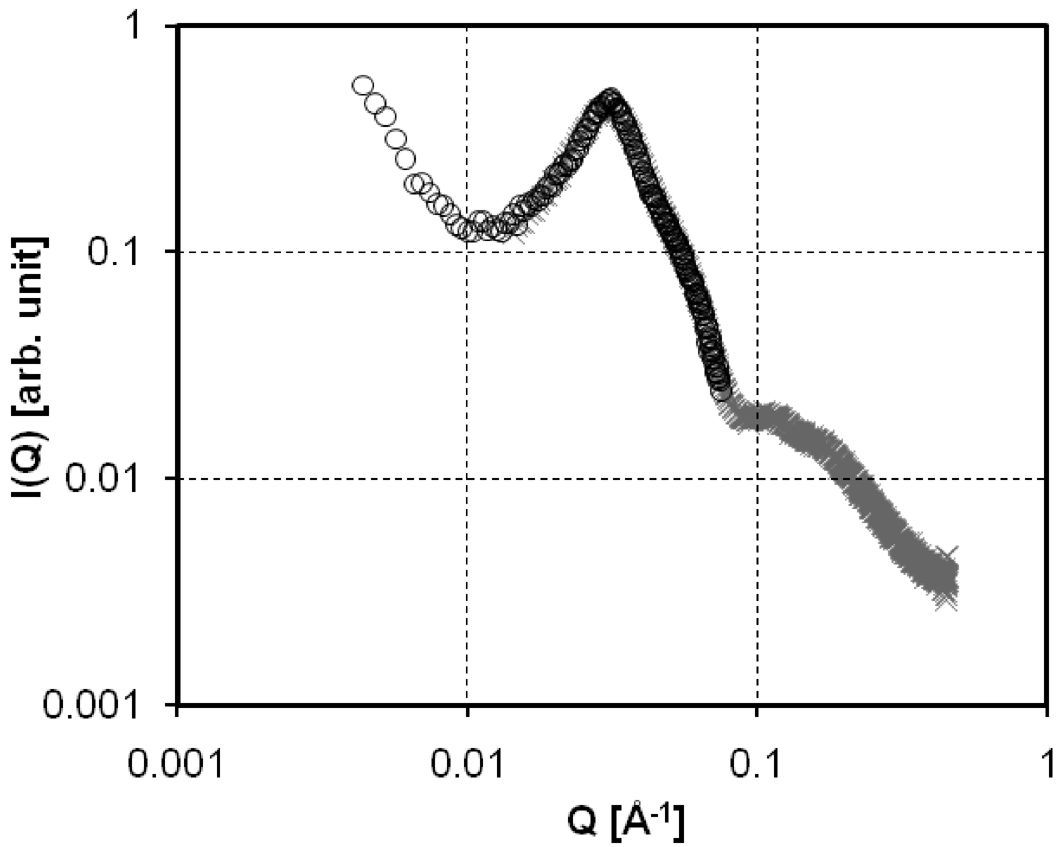

FIG 2. Scattering data of a dendrimer sample [10] processed to $I(Q)$ using the *fatcat* software. The sample was measured in the frame-skipping mode with frames at 2-5.4Å and 8.9-12.3Å and the sample-to-detector distance of 5m. Data from the two frames are reduced separately to illustrate their Q-ranges. The overlapping grey crosses are from the frame at 2-5.4Å and the black circles are from 8.9-12.3Å. Their region of overlap lies between Q~0.02-0.08Å^{-1}.

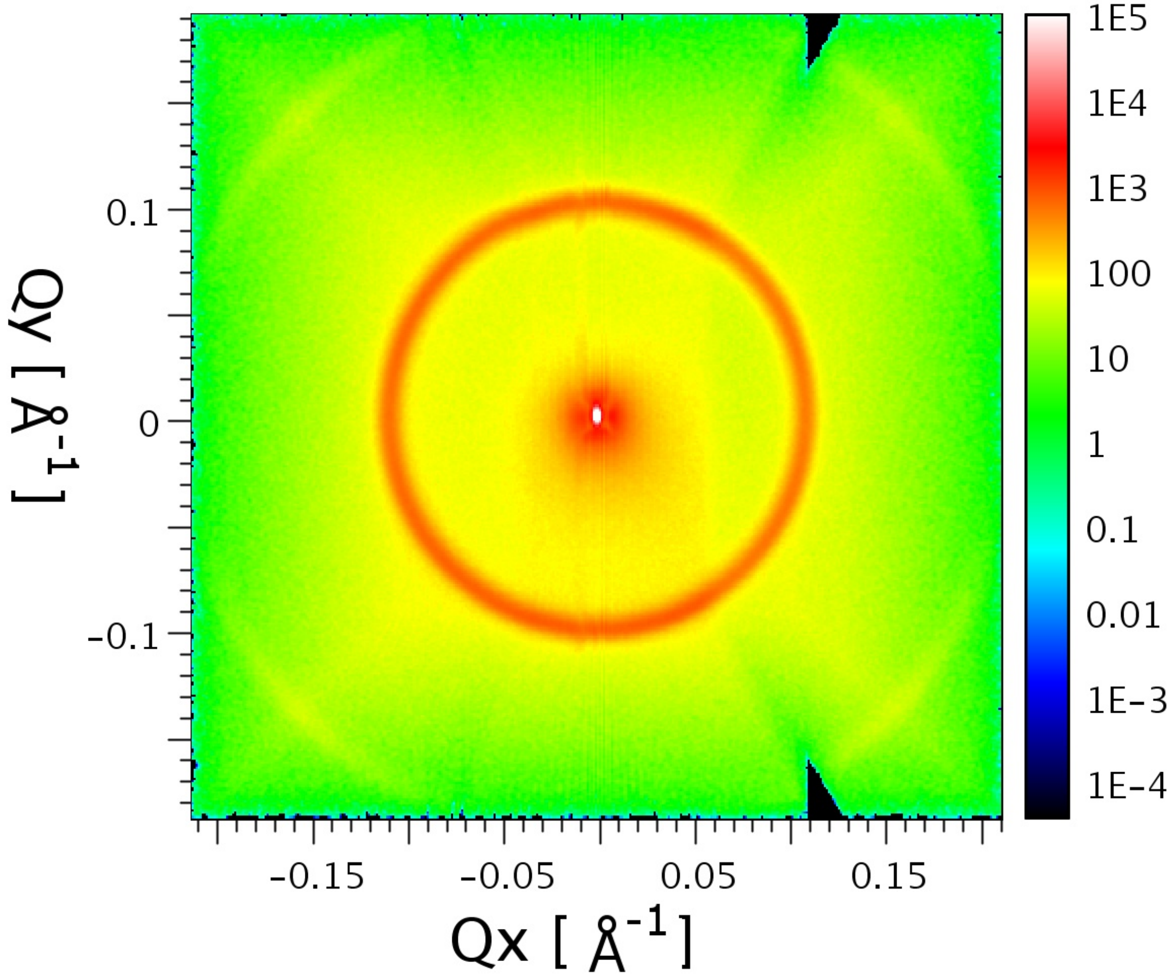

FIG 3. Neutron scattering data of a silver behenate sample [16] processed to $I(Q_x,Q_y)$ using the *fatcat* software. The sample was measured with the sample-to-detector distance of 5m and the instrument running at 60Hz. The used neutron wavelength frame was at 3-6.4Å. Eight of the detector tubes on the right side were non-functioning during the measurement. The missing detector data are reflected in Q-space as an arc-like feature around Q_x~0.6-0.8 Å^{-1}. The bright spot in the middle is the transmission pixels that are behind a pinhole on the beamstop.