Launch of the I13-2 data beamline at the Diamond Light Source synchrotron

A J Bodey[1] and C Rau[1,2,3]

[1] Diamond Light Source, Oxfordshire, OX11 0DE, UK
[2] School of Materials, University of Manchester, Manchester, M13 9PL, UK
[3] Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA

andrew.bodey@diamond.ac.uk

Abstract. Users of the Diamond-Manchester Imaging Branchline I13-2 commonly spend many months analysing the large volumes of tomographic data generated in a single beamtime. This is due to the difficulties inherent in performing complicated, computationally-expensive analyses on large datasets with workstations of limited computing power. To improve productivity, a 'data beamline' was launched in January 2016. Users are scheduled for visits to the data beamline in the same way as for regular beamlines, with bookings made via the User Administration System and provision of financial support for travel and subsistence. Two high-performance graphics workstations were acquired, with sufficient RAM to enable simultaneous analysis of several tomographic volumes. Users are given high priority on Diamond's central computing cluster for the duration of their visit, and if necessary, archived data are restored to a high-performance disk array. Within the first six months of operation, thirteen user visits were made, lasting an average of 4.5 days each. The I13-2 data beamline was the first to be launched at Diamond Light Source and, to the authors' knowledge, the first to be formalised in this way at any synchrotron.

1. Motivation

The Diamond-Manchester Imaging Branchline I13-2 at Diamond Light Source (DLS) is dedicated to imaging and tomography experiments that span numerous disciplines including biology, materials science, geology and palaeontology. The branchline provides partially-coherent 5-35keV X-rays and variable in-line phase contrast. The high flux of the beam enables datasets to be recorded rapidly, and the branchline is commonly used for high-throughput tomography and time-resolved radiography (eg. [1]) and tomography (eg. [2]), leading to very large data volumes.

Tomographic reconstructions can have very large file sizes. 32-bit reconstructions computed from images collected at I13-2 by a pco.edge 5.5 (PCO AG, Germany) detector operating in full-frame mode (2560 x 2160 pixels) are 57GB in size, while a pco.4000 detector can generate 172GB tomographic volumes. While these can be compressed by a factor of 2 or 3 by conversion to 16- or 8-bit respectively, loading such volumes into RAM (while leaving enough for image processing) still requires considerable computing power.

A similar problem exists with raw data. Inspecting raw datasets (sample, flat-field and dark-field images) allows users to check for experimental problems and is most efficiently achieved by loading complete datasets into RAM. A single-file raw dataset of 4,000 sample images from a pco.edge 5.5

detector is 42GB in size. While only 4,022 images are necessary for complete rotational sampling with this detector's full array, 12,000 were collected during a recent experiment at I13-2 to optimise the signal:noise ratio, yielding a raw dataset of 127GB.

Beamtimes at I13-2 often produce hundreds of tomographic scans and generate tens of terabytes of data. Beamtimes involving time-resolved studies can produce considerably more than 100TB of data. I13-2 currently produces approximately one third of all data generated at DLS.

In order to extract useful information from tomograms, visualisation and analysis are required. The latter commonly involves filtering, segmentation (partitioning a tomographic volume into multiple segments) and measurements of particle sizes and morphologies, porosity and connectedness. Some image processing operations are very computationally expensive. The non-local means filter for example, can be very effective in denoising tomograms (figure 1), but commonly requires several hours to process a complete tomographic volume – even with a powerful workstation.

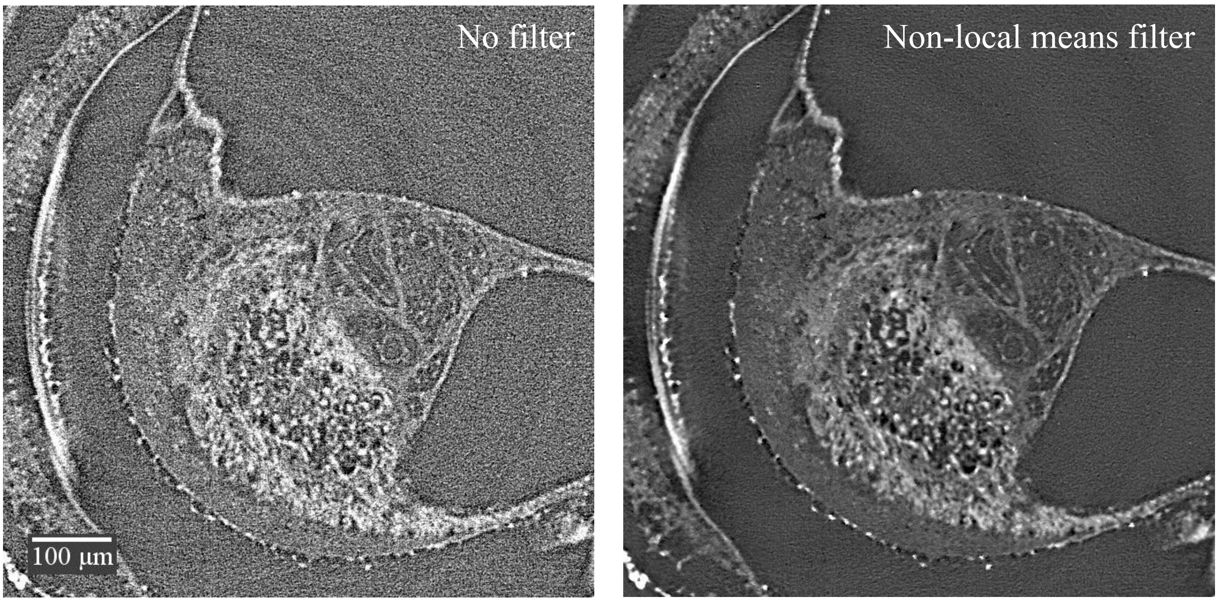

Figure 1. As this tomographic slice through a cochlea [3] demonstrates, the computationally-intensive non-local means filter can be very effective at removing noise from tomograms. Such denoising is particularly useful for samples which must be imaged with a limited dose, such as biological specimens and those involved in rapid time-resolved tomography.

For these reasons, as well as the difficulties that can otherwise be inherent in segmenting and taking measurements from tomographic data, users commonly spend many months analysing data recorded in a single beamtime. Computing power is therefore key to improving productivity and reducing the time between I13-2 beamtimes and publication. Computers that can load complete or multiple volumes into RAM are advantageous, as is sufficient processing power to accelerate computationally-intensive operations. Many of the software packages currently available for visualisation and analysis are not amenable to cluster operation and must be run on single workstations.

I13-2 users commonly lack the computing power to analyse data efficiently at their home institutions, and some even lack computers with sufficient RAM into which they can load complete tomographic volumes. Other users do not have access to expensive licenses for image-processing software. It was therefore considered useful to make powerful dedicated graphics workstations available to users. To manage access to these machines, a 'data beamline' was established.

2. Project management

Users are scheduled for visits to the data beamline in the same was as visits are made for regular beamlines, with bookings made via the User Administration System [4] and provision of financial support for travel and subsistence. Scheduling is managed by an appointed beamline staff member. Data from I13-2 beamtimes are held on a high-performance GPFS disk array for 120 days before being archived. Data beamline visits are preferentially made within this timeframe, but if visits are made after data have been archived they are restored to the high-performance array. Users are given high priority on the central DLS computing cluster for the duration of their visit.

3. Workstation specifications

Workstation specifications were determined via discussion with FEI Company about optimising performance of their software package Avizo, and are as follows: 512GB DDR4 RAM, dual Intel Xeon E5-2687W v3 CPUs (3.1GHz, 10 cores, 20 threads), dual NVIDIA Quadro K6000 GPUs with 12GB GDDR5 RAM and ECC, dual 512GB SSDs in RAID 1, $10Gbs^{-1}$ optic fibre connections and 30-inch monitors. Use of RAM three times larger than loaded data is recommended. A typical tomographic volume generated from pco.edge 5.5 images, once converted from 32- to 16- or 8-bit, is 28 or 14GB in size, respectively. 512GB RAM therefore allows for up to six 16-bit or twelve 8-bit tomographic volumes to be loaded and processed simultaneously.

4. Software

Tomographic reconstruction, via Dawn [5] and Savu [6], is performed on the central DLS computing cluster, although reconstructions are normally performed during I13-2 beamtimes. Various packages are available for visualisation and analysis of tomograms and radiographs, including Avizo (FEI Company, USA), Drishti [7], ImageJ [8] and Fiji [9]. Proprietary programming languages including MATLAB (Mathworks, USA) and Mathematica (Wolfram Research, USA) are available. Other software can be installed upon request.

5. Usage to date

The data beamline was launched in January 2016. In the six months that followed, thirteen user visits were made (some with multiple attendees), lasting 4.5 days on average. The system is due to be replicated on other beamlines at DLS.

6. Acknowledgements

The authors would like to express their thanks to various people at DLS who helped to set up the data beamline: K Poulter for project management support, M Basham for advice on computer specifications, A Fifield for IT acquisition, T Friedrich for cluster management, F Ferner for data storage management, R Lear for network administration, and A Strange for User Administration System support.

References

[1] Xu W W, Tzanakis I, Srirangam P, Mirihanage W U, Eskin D G, Bodey A J and Lee P D 2016 Synchrotron quantification of ultrasound cavitation and bubble dynamics in Al–10Cu melts *Ultrasonics Sonochemistry* **31**

[2] Menke H P, Andrew M G, Blunt M J and Bijeljic B 2016 Reservoir condition imaging of reactive transport in heterogeneous carbonates using fast synchrotron tomography — Effect of initial pore structure and flow conditions *Chemical Geology* **428**

[3] Rau C and Richter C-P 2015 Imaging cochlear soft tissue displacement with coherent x-rays *Physica Scripta* **90**

[4] Diamond User Administration System https://uas.diamond.ac.uk/uas/

[5] Basham M, Filik J, Wharmby M T, Chang P C Y, El Kassaby B, Gerring M, Aishima J, Levik K, Pulford B C A, Sikharulidze I, Sneddon D, Webber M, Dhesi S S, Maccherozzi F,

Svensson O, Brockhauser S, Naray G and Ashton A W 2015 Data Analysis WorkbeNch (DAWN) *J Synchrotron Radiat* **22** 853-8
[6] Atwood R C, Bodey A J, Price S W T, Basham M and Drakopoulos M 2015 A high-throughput system for high-quality tomographic reconstruction of large datasets at Diamond Light Source *Philos Trans A Math Phys Eng Sci* **373**
[7] Limaye A 2012 Drishti: a volume exploration and presentation tool *Proc SPIE 8506: Developments in X-Ray Tomography VIII*
[8] Schneider C A, Rasband W S and Eliceiri K W 2012 NIH Image to ImageJ: 25 years of image analysis *Nat Meth* **9** 671-5
[9] Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White D J, Hartenstein V, Eliceiri K, Tomancak P and Cardona A 2012 Fiji: an open-source platform for biological-image analysis *Nat Meth* **9** 676-82