20 µs-resolved high-throughput X-ray photon correlation spectroscopy on a 500k pixel detector enabled by data-management workflow

Qingteng Zhang,a Eric M. Dufresne,a Yasukazu Nakaye,b Pete R. Jemian,a Takuto Sakumura,b Yasutaka Sakuma,b Joseph D. Ferrara,b Piotr Maj,c Asra Hassan,d Divya Bahadur,d Subramanian Ramakrishnan,d Faisal Khan,a Sinisa Veseli,a Alec R. Sandy,a Nicholas Schwarz,a and Suresh Narayanan*a

X-ray Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA, bXRD Design and Engineering Department, Rigaku Corporation, 3-9-12 Matsushara-cho, Akishima-shi, Tokyo, Japan, cAGH University of Science and Technology, av. Mickiewicza 30, Krakow 30-059, Poland, and dChemical and Biomedical Engineering, FAMU-FSU College of Engineering, Tallahassee, FL 32310, USA. *Correspondence e-mail: sureshn@anl.gov

The performance of the new 52 kHz frame rate Rigaku XSPA-500k detector was characterized on beamline 8-ID-I at the Advanced Photon Source at Argonne for X-ray photon correlation spectroscopy (XPCS) applications. Due to the large data flow produced by this detector (0.2 PB of data per 24 h of continuous operation), a workflow system was deployed that uses the Advanced Photon Source data-management (DM) system and high-performance software to rapidly reduce area-detector data to multi-tau and two-time correlation functions in near real time, providing human-in-the-loop feedback to experimenters. The utility and performance of the workflow system are demonstrated via its application to a variety of small-angle XPCS measurements acquired from different detectors in different XPCS measurement modalities. The XSPA-500k detector, the software and the DM workflow system allow for the efficient acquisition and reduction of up to ~10^9 area-detector data frames per day, facilitating the application of XPCS to measuring samples with weak scattering and fast dynamics.

1. Introduction

X-ray photon correlation spectroscopy (XPCS) provides information about the dynamic properties of condensed matter across a range of time scales that spans length scales from the atomic in metallic glasses to near micrometres at fluid interfaces. Many of these applications are summarized in recent reviews (Leheny, 2012; Shpyrko, 2014; Sinha et al., 2014; Madsen et al., 2016; Sandy et al., 2018). Partially driven by XPCS-applicable area detectors with increasingly high frame rates (Trapani et al., 2018; Zhang et al., 2020, 2018; Zinn et al., 2018; Pennicard et al., 2018; Redford et al., 2018; Maj et al., 2020; Allahgholi et al., 2019; Poikela et al., 2014) but also by emerging near-diffraction-limited storage rings (DLSRs) (Eriksson et al., 2014), a growing area of interest is XPCS applied to rapidly fluctuating and often weakly scattering soft materials (Vodnala et al., 2018; Möller et al., 2019; Chung et al., 2019; Zhang et al., 2019; Dallari et al., 2020; Nigro et al., 2020; Frenzel et al., 2019; Yavitt et al., 2020).

In the case of measurements exhibiting stationary dynamics, the signal-to-noise ratio (SNR) of an equilibrium XPCS measurement is proportional to \( I_0 t_e (N_p N_t)^{1/2} \) (Falus et al., 2006; Shpyrko, 2014), with \( I_0 \), \( t_e \), \( N_p \) and \( N_t \) corresponding to the scattering rate of the sample (photons per pixel per second), the exposure time of an individual frame, the number
of nominally equivalent pixels in a frame, i.e. within a narrow band of $Q$, and the number of frames in an acquisition sequence, respectively. Area detectors allow multiple intensity–time series to be measured simultaneously, improving the equilibrium $g_2$ SNR by $(N_p)^{1/2}$, while also providing concurrent measurements of the $Q$ dependence of $g_2$, thereby improving the efficiency of measurements. In the case of two-time measurements where the dynamics evolve over time (Fluerasu et al., 2005; Madsen et al., 2010; Das et al., 2019; Mokhtarzadeh et al., 2019; Hoshino et al., 2020; Jain et al., 2020), the SNR does not improve by increasing $N_t$, since the correlation is calculated between two frames, making the large $N_p$ provided by an area detector the only way to maintain the SNR of the correlation without sacrificing the time resolution indicated by $t_0$.

Recent developments in fast XPCS-suitable area detectors provide access to increasingly short delay times and a large dynamic range that is relevant to measuring dynamics in soft materials but also exacerbates the challenge of quickly and straightforwardly managing and reducing large amounts of time-series data to correlation functions. For instance, Fig. 1 shows $g_2$ measured from 20 nm-diameter octadecyl-grafted silica nanoparticles suspended in decalin at a volume fraction of 20% (hereinafter referred to as D20) for varying numbers of frames. Fig. 1(a) shows $g_2$ determined from a single sequence of 100 000 frames acquired at 52 kHz using the Rigaku XSPA-500k detector (described further below) while Fig. 1(b) is the average of 300 repeats or ‘batches’ of this same measurement. Evidently, the quality of the result in Fig. 1(b) is much better but achieving this efficiently requires a high-throughput workflow that can keep up with the unprecedented data accumulation rate of the XSPA-500k to provide near real-time human-in-the-loop experiment feedback. In particular, for the example shown Fig. 1(b), near real-time feedback amounts to our system acquiring, reducing and providing access to increasingly short delay times and a large dynamic range.

In the remainder of this article we describe the performance of the Rigaku XSPA-500k detector and then document the architecture and performance of the workflow and data-reduction tools on representative small-angle XPCS measurements using different detectors.

**Figure 1**
(a) $g_2$ measured from D20 via a single batch of 100 000 frames. The contour of the ROI is shown in Fig. 3 (black dashed lines) and corresponds to $Q = 0.0026$ Å⁻¹ with a total of 643 pixels. (b) $g_2$ averaged from 300 batches.

2. XSPA-500k detector

The XSPA-500k (X-ray Seamless Pixel Array) detector is a photon-counting detector with a pixel array size of $1024 \times 512$, a pixel size of 76 μm and a maximum frame rate of 52 kHz (signal depth is 2 bits in this mode of operation). In the context of XPCS workflow and data management and reduction, we present one extreme of XSPA-500k operation, namely the detector running at its highest frame rate where acquisition is limited to short measurements of 2 s separated by 2 s of sparsification and data movement off the detector control computer.

The XSPA-500k comprises 16 UFXC modules (Grybos et al., 2016; Zhang et al., 2016; Kleczek et al., 2019) each with 256 × 128 pixels. A single 320 μm-thick silicon sensor spans all 16 modules. As illustrated schematically in Fig. 2, a unique feature of this detector is that the sensor is bonded to the readout module with the sensor pixels displaced from the readout modules’ pixels in a manner that results in seamless readout of the entire sensor area of 78 mm × 39 mm. Specifically, Fig. 2(a) shows a typical sensor and tiled readout pixel-array detector (PAD) design and Fig. 2(b) the XSPA-500k design with displaced sensor and readout pixels. The red rectangles and squares are sensor pixels, the black squares are readout module pixels, the black circles are the connections.
between the sensor and readout module pixels, and the cross-hatched blue squares in Fig. 2(a) represent the gaps between the readout modules. Notice that the sensor pixels at the gaps in Fig. 2(a), a typical PAD design, are rectangles, while the sensor pixels in Fig. 2(b) are all squares and are displaced from the readout module pixels to cover the gap regions.

Example data from the XSPA-500k are presented in Figs. 1 and 3. Fig. 3(a) is the 2D time-averaged small-angle X-ray scattering (SAXS) pattern of D20 determined from a single data-acquisition sequence of 100 000 frames, while Fig. 3(b) is the azimuthally averaged scattering pattern obtained by averaging the 2D pattern around the origin of reciprocal space. Fig. 1(a) is $g_2$ calculated from this same sequence and Fig. 1(b) is the average of 300 such sequences or batches. The SAXS and XPCS analyses were performed using the same method as documented by Zhang et al. (2018), except that the $Q$ range within which the SAXS intensity is averaged is logarithmically spaced instead of linearly spaced. Note that, due to the limitation of the 19 µs time resolution, $g_2$ from D20 can only be accurately evaluated for $Q < 0.008$ Å$^{-1}$, although Fig. 3 shows coverage up to $Q = 0.016$ Å$^{-1}$ thanks to the large field of view of the XSPA-500k. More efficient use of the large field of view of the XSPA-500k for XPCS studies can therefore be achieved by operating the XSPA-500k at a detector

distance much larger than the current 4 m on beamline 8-ID-I at the Advanced Photon Source (APS), Argonne. An example is the up to 16 m setup on the 11-ID beamline (coherent hard X-ray scattering, CHX) at NSLS-II (Brookhaven, New York, USA) (Yavitt et al., 2020; Johnson et al., 2019; Headrick et al., 2019).

A data acquisition sequence of 100 000 frames can be triggered from the beamline control system as frequently as every 4 s. Two seconds are used for continuous data acquisition, with the remaining time used for sparsification on the detector control computer, transfer and augmentation of the detector data with metadata describing the beamline and experiment configuration (De Carlo et al., 2014) to the beamline storage system, and purging the detector memory and readying it for the next acquisition sequence. Sparsification is performed by reading sequences of detector frames into RAM and converting them into sparse data using parallel C# code running on all 20 threads of an i9-9900 CPU with 128 GB RAM. Each photon event is recorded as a 64 bit little-endian number with bits 64–41, 36–17 and 11–1 corresponding to the frame number, pixel index and counts, respectively. Unused bits are reserved to accommodate future detector developments, such as identifying multiple XSPA-500k modules, identifying the acquisition mode or accommodating larger dynamic ranges. Running at full speed, this operation mode of the XSPA-500k can record 0.2 PB of data every 24 h. This should be compared with other state-of-the-art detectors such as 0.03 PB per 24 h for the two-million-pixel ePix10k (van Driel et al., 2020) or 2.1 PB per 24 h for the one-million-pixel AGIPD (Allahgholi et al., 2019; Lehmkuhler et al., 2020) [both primarily used for free-electron laser (FEL) applications; data rate estimated assuming 2 bytes per pixel and continuous operation over 24 h], while flagship products from Dectris and X-Spectrum are in the petabyte per 24 h range. The large data rate from detectors at next-generation X-ray sources therefore makes automated high-throughput workflow an operation-critical component for effective beamline operation.

3. Automated high-throughput data-management workflow

A data-management (DM) workflow has been developed at APS and implemented on many beamlines there, including those dedicated to tomography where it was first deployed (Veseli et al., 2018). The DM system is a software suite that provides a common framework and a set of services and tools that beamlines can use to tailor their data acquisition, cataloging, storage, distribution and reduction processes. Of particular relevance for this work is the DM processing service that provides support for managing user-defined workflows and for submitting, running and monitoring processing jobs based on those workflows. A workflow definition is kept in the beamline workflow database and is represented by a set of processing steps (stages) that are executed in a defined order. Each stage is associated with a command or script with an arbitrary number of input arguments and output variables. Input arguments are provided by the user and/or assembled by

![Figure 3](image-url)

(a) Time-averaged 2D SAXS pattern from D20 acquired by the XSPA-500k detector (100 000 frames at 52 kHz). (b) Azimuthally averaged scattering pattern obtained from (a). Black dashed lines show the ROI where $g_2$ from Fig. 1 is calculated.
The major steps in the XPCS workflow and data reduction are illustrated in Fig. 4 and are as follows:

1. Data are acquired by the detector and stored in local memory.
2. Raw detector data are sparsified in the memory of the detector control computer.
3. Sparse data are transferred from storage on the detector control computer to beamline-local networked storage.
4. Sparse data are combined with metadata and transferred from beamline-local storage to an HPC cluster or workstation via high-bandwidth Globus transfer tools (https://www.globus.org).
5. Data are reduced (turned into correlation functions) using newly developed OpenMP-enabled correlation code.
6. Calculated results and metadata are assembled in an HDF5 file and copied to local networked storage and a Globus endpoint for examination and visualization by and distribution to users.

![Figure 4](image)

(a) Data-management workflow for high-throughput XPCS on APS beamline 8-ID-I. The labeled steps are explained in the main text.

The metadata capture parameters required for XPCS analysis and also information about the measurement to facilitate interpretation of the measurements (De Carlo et al., 2014). They are assembled from beamline EPICS (Experimental Physics and Industrial Control System, https://epics-controls.org/) process variables and information provided by the user, such as sample identification, and stored in an HDF5 file that accompanies the raw data (compressed detector data are currently kept in a separate but associated binary file so that they can be used in a sparse format throughout processing, a feature not currently supported in HDF5). One example of metadata is the region of interest (ROI) on the detector. Another example can be indices attached to detector pixel maps that define groups of pixels with nominally equivalent $Q$ values for azimuthally averaged SAXS and XPCS analysis (Khan et al., 2018).

For XSPA-500k measurements, the typical data-reduction time is approximately 5 s per acquisition, although the exact time can vary depending on the size of the file generated by the measurement. Since the data reduction for each acquisition is distributed and performed in parallel across many nodes, overall data reduction roughly keeps up with measurement time so users can examine the calculated and averaged correlation functions and initiate further local processing of the batch results during the experiment. At the end of the experiment, all user data, including sparsified detector frames, correlation functions and user-defined local processing routines, are transferred to a user-specific Globus endpoint located outside the APS firewall for distribution. Of the DM workflow steps, Step (2) typically leads to a hundred-fold reduction in data volume due to the low photon occupancy of individual frames. Steps (3) and (4) are enabled using Globus high-bandwidth command-line interface (CLI) data-transfer tools. The maximum network transfer speeds for Steps (3) and (4) are 10 and 40 Gb s$^{-1}$, respectively, but the actual speed is limited by the speed of writing to the beamline storage hard disk and this is currently 5 Gb s$^{-1}$.

Step (5) is performed in newly developed multi-threaded software that expands on the capabilities previously developed (Khan et al., 2018). This software reads the sparsified data directly and converts them to a sparsified matrix format without expansion, saving considerable processor memory. The matrix format and the sparse operations have been implemented in C++ to improve performance. The final operation in the $g_2$ calculation, namely the division of the numerator by the denominator, is performed using the vectorization method from the Eigen library (Guennebaud & Jacob, 2010) for optimal speed and memory use. Step (6) assembles the calculated results and extensive metadata aggregated from the experiment control and DM systems in an HDF5 file and sends this back to local networked storage for further examination and visualization by users using a graphical user interface (GUI) developed in MATLAB (Sikorski et al., 2011). The GUI allows the time-averaged and time-dependent batch results to be examined individually, overlayed or merged, for example, to improve SNR.
4. XPCS data reduction

Two of the most commonly used XPCS reduction algorithms are multi-tau (Schatzel, 1983; Schatzel, 1987) and two-time (Ju et al., 2019; Chesnel et al., 2016; Perakis et al., 2017; Bikondoa, 2017; Sutton et al., 2003; Cipelletti et al., 2003). The multi-tau reduction is suitable for XPCS studies where the time scale of the dynamics does not vary significantly during the measurement. In multi-tau, correlations within a time series are averaged over pairs of frames with the same time spacing, resulting in a 1D array. A key feature of the multi-tau correlation routine is using progressively collapsed averages of frames when correlating across successively larger time separations. This feature improves the SNR in calculated correlation functions at long delay times, reducing the computational complexity ($N \log N$ versus $N^2$) and providing a practical way of calculating correlation functions over a large dynamic range. The result from a multi-tau reduction of a single XSPA-500k acquisition sequence from a stationary sample (D20) is shown in Fig. 1(a), where a batch of measurements is performed and reduced as illustrated in Fig. 4. Users can then load XPCS results from individual batches into the MATLAB-based XPCS GUI (Sikorski et al., 2011) and generate the averaged XPCS result shown in Fig. 1(b) for further analysis.

For dynamics where the time scales vary significantly over time, such as dynamics driven by external stimuli or dynamics displayed during relaxation towards equilibrium, it is necessary to express the correlation as a function of both the measurement time since the impulse and the time separation between pairs of frames. This results in a 2D map of correlation where the x and y axes are the measurement times of the frames and the value of the element at those indices corresponds to the correlation. This is the so-called two-time correlation.

Fig. 5 is an example of a two-time correlation function at $Q = 0.026 \text{ Å}^{-1}$ showing the aging of a ‘soft’ glass after a rheological shear event. The glass is formed of a concentrated suspension of charged silica nanoparticles in water (Chen et al., 2020). The measurements were performed using the Lambda detector running at 10 Hz. Details regarding the rheo-XPCS experimental setup and beam parameters can be found in the work of Chen et al. (2020). As illustrated in Fig. 5(a), elapsed time since the shear event is obtained from the lower-left to upper-right diagonal in the figure, while vertical or horizontal cuts extending from this diagonal show the decay of the correlation as the elapsed time increases. The aging or slowing down of dynamics after the rheological shear event is indicated by the widening of the correlation with increased time after the shear event. The same information can be obtained from Fig. 5(b), which shows selected cuts of Fig. 5(a) along the delay-time cut directions as a function of time after shear. With the DM workflow, different types of XPCS data reduction on data collected from different detectors can be performed quickly and automatically via commands in an execution script, allowing for near real-time human-in-the-loop feedback on the experiment.

5. Performance

The performance of the DM workflow system and data-reduction system for the two use cases described above is documented in Table 1. A significant difference between the two detectors is the ability to sparsify data concurrently with the Lambda detector as opposed to the XSPA-500k detector. This is largely a consequence of the far higher peak frame rate provided by the XSPA-500k. Concurrent sparsification with the Lambda detector is accomplished using the EPICS area-Detector plugin (Rivers, 2010).

For a more equivalent comparison of workflow and data reduction between the two detectors, the two rightmost columns of Table 1 also document the performance when measuring a 1 mm-thick aerogel static reference. With a 4 m sample-to-detector distance, the Lambda sees a scattering rate of 10 photons per pixel per second with a 55 $\mu$m pixel size, while the XSPA-500k sees a scattering rate of 20 photons per pixel per second with a 76 $\mu$m pixel size. The table documents

![Figure 5](image)
performance with the Lambda operating at 1 kHz and the XSPA-500k at 52 kHz.

We note that the run time of the data-reduction step [Step (5)] is typically more than ten times longer than the rest of the steps combined, indicating that the biggest throughput gains can be achieved by significantly improving the performance of this step. In practice, however, since the reduction occurs on a single node of an HPC cluster and can happen concurrently with the other steps, the effective average time per job can be reduced proportionately by running multiple jobs on different nodes of the cluster. Also, since the bottleneck of the correlation calculation is the arithmetic operation between two equal-length slices of the 1D array recording the scattering intensity at different frames in time, which is evaluated independently at each pixel (Zhang et al., 2018; Khan et al., 2018), significant speed-up can be achieved by parallelizing the pixel-wise computation on a GPU, which typically has a hundred times more cores than a generic CPU.

### 6. Conclusions

We have demonstrated the new Rigaku XSPA-500k detector that acquires data at 52 kHz from a 500k pixel gapless sensor. The data output rate of 0.2 PB per 24 h of continuous operation motivated the application and further optimization of an automated and efficient workflow system coupled to XPCS data reduction on an HPC cluster. We have also demonstrated successful integration of the workflow with various state-of-the-art XPCS-suitable area detectors. The capabilities we have described ease the operational burden of performing XPCS measurements and facilitate human-in-the-loop experiment feedback. Our work has found immediate applications in soft-matter studies on beamline 8-ID-I, typically examining fast dynamics from weakly scattering materials, where automation and high-repetition short-burst measurements are often key to progress.

Looking forward, we see greatly increased data flow arising from XPCS end stations either being discussed or planned for near-diffraction-limited synchrotron sources like MAX-IV (Martensson & Eriksson, 2018; Plivelic et al., 2019), ESRF-EBS (Sanchez Del Rio et al., 2019; Tsigaridas & Ponchut, 2019), APS Upgrade (Shi et al., 2017) and PETRA-IV (Schroer et al., 2018). We believe that the infrastructure we have described here will become an essential component for XPCS experiments at future light sources.

### Acknowledgements

We are grateful for the expert technical assistance of Don Jensen Jr and Ray Ziegler. The authors thank Kacper Switalski (DOE Science Undergraduate Laboratory Internships) for data acquisition and commissioning of beamline control software, and Robert Leheny, Jim Harden and Yihao Chen for providing us with their Rheo-XPCS data to test system performance.

### Funding information

This research was performed on beamline 8-ID-I of the APS, a DOE Office of Science User Facility operated for the DOE Office of Science by ANL supported under Contract No. DE-AC02-06CH11357. AH, DB and SR would like to acknowledge funding from NSF FAMU CREST center award No. 1735968. UFXC32k detector development, measurements and study at the AGH University of Science and Technology were supported by the National Centre for Research and Development, Poland (PBS1/A3/12/2012) and by NAWA – Polish National Agency for Academic Exchange (contract No. PPI/ APM/2018/1/00049/U/001).

### References


