Microsecond time-resolved serial crystallography data pipeline for Swiss Light Source 2.0

F. Leonarski\textsuperscript{1}, G. Assmann\textsuperscript{1}, K. Smith\textsuperscript{1}, A. Mozzanica\textsuperscript{1}, H.-C. Stadler\textsuperscript{1}, P. Gasparotto\textsuperscript{1}, L. Barba\textsuperscript{1,2}, B. Béjar\textsuperscript{1,2}, H. Mendonça\textsuperscript{3}, F. Dworkowski\textsuperscript{1}, M. Janousch\textsuperscript{1}, M. Burian\textsuperscript{4}, M. Wang\textsuperscript{1}

\textsuperscript{1}Paul Scherrer Institute (PSI), 5232 Villigen, Switzerland \textsuperscript{2}Swiss Data Science Center (SDSC), 1015 Lausanne, Switzerland \textsuperscript{3}Swiss National Supercomputing Centre (CSCS), 6900 Lugano, Switzerland \textsuperscript{4}DECTRIS AG, 5405 Baden-Dättwil, Switzerland

filip.leonarski@psi.ch

**Keywords:** X-ray data analysis, data acquisition, X-ray detectors, FPGA, GPU, FAIR data

The years to come for macromolecular crystallography at the Swiss Light Source (SLS) are bright. The SLS will undergo a major upgrade in years 2023-2025 that will increase X-ray brilliance by more than an order of magnitude. This, paired with beamline improvements and developments in charge-integrating X-ray detector technology [1], opens new opportunities for time-resolved synchrotron serial crystallography [2]. The new source will allow for 10 microsecond measurements, while the current source limits time resolution to the millisecond range. However, handling many GBs of X-ray images per second is a big data challenge [3] that must be overcome in order to make collecting and processing diffraction images at rates up to 10 kHz feasible in the future (see Fig. 1). In this presentation, I will highlight a few aspects of the development at the PSI to handle big data in the future.

![Figure 1: A. Data rates of detectors used at PSI macromolecular crystallography beamlines and planned for the future [4]. B. Graphical presentation of main features of the new SLS 2.0 pipeline for the time-resolved serial synchrotron crystallography.](image)

I will first present Jungfraujoch, a detector read-out system with FPGAs and GPUs capable of handling 30 GB/s data rates within a single server. The system can run real-time image analysis to provide fast experimental feedback, including spot finding and radial integration. After the Jungfraujoch system writes images to the facility file system, the next step is an automated data analysis pipeline currently under development. The pipeline is designed to produce electron density maps and to inform the experiment team during beamtime about the progress.

I will present PSI and SDSC efforts in improving spot-finding and indexing algorithms using modern machine-learning frameworks optimized to run efficiently on distributed CPUs and GPUs. I will also give an outlook for operating our image processing pipeline at the high-performance computing cluster Alps, allowing online data processing with massive data rates.

I will show how the FAIR principles will be applied to the data that will be written by the pipeline. The images, metadata, and analysis results produced by the pipeline are written according to community standards [5]. All collected datasets are ingested into the SciCat data catalogue, saved on tapes for long-term storage, and accessed according to the PSI Data Policy.

Finally, I will present a collaboration with DECTRIS to include hardware-accelerated real-time data acquisition and analysis features in the DECTRIS Next Generation Detector Control Unit, allowing more facilities to use these features.


F.L. acknowledges support for part of this work by the Open Research Data Program of the ETH Board. P.G., L.B., B.B., H.M. acknowledge Swiss Data Science Center for funding RED-ML project under grant no. C19-03.