

Using the Gold Standard for data archival at kilohertz speeds

Aaron Brewster¹, Herbert Bernstein², Andreas Förster³, Graeme Winter⁴

¹*Lawrence Berkeley National Lab, Albany, United States of America;*

²*Ronin Institute for Independent Scholarship;*

³*DECTRIS Ltd;*

⁴*Diamond Light Source Ltd;*

asbrewster@lbl.gov

Serial femtosecond X-ray crystallography (SFX) involves the collection of thousands to up to millions of images in a few minutes. Being able to process these data at speeds that match the data collection rate is critical for scientists who need fast feedback on their data quality. Doing this while simultaneously creating data that fits FAIR standards (Findability, Accessibility, Interoperability, and Reusability) is challenging, and has been the focus of the High Data-Rate Macromolecular Crystallography (HDRMX) working group. We have recently published a consensus best-practice NeXus representation of a complex, multi-panel detector, the Jungfrau 16M from the SwissFEL Bernina endstation [1-2]. 256 individual panels are described and positioned in real space using a vector transformation system that is standard in NeXus, is machine readable, and completely specifies the experimental geometry.

We have implemented this approach during a subsequent data collection at the EuXFEL on the AGIPD detector, similar in geometry to the Jungfrau 16M [3]. Here we collected data at 2kHz and demonstrated the ability to process these data at these speeds with the software package DIALS on the Maxwell computing cluster at DESY, using 96 nodes, 80 cores per node. This required careful attention to how the data were laid out on disc. These methods and the NeXus framework for SFX will be presented.

[1] Bernstein, H.J., et. al. (2020). Gold Standard for macromolecular crystallography diffraction data. *IUCrJ*, 7(5) 784–792.

[2] Ingold, G., et.al. (2019). Experimental station Bernina at SwissFEL: condensed matter physics on femtosecond time scales investigated by X-ray diffraction and spectroscopic methods. *J. Synchrotron Rad.* **26**, 874–886.

[3] Allahgholi, A., et. al. (2015). AGIPD, a high dynamic range fast detector for the European XFEL. *JINST* 10 C01023.

Keywords: XFEL, FAIR, MX data, Gold Standard, Archiving

The work was supported in part by funding from Dectris Ltd., from the U. S. Department of Energy (BES KP1605010, KP1607011, DESC0012704), from the U. S. National Institutes of Health (NIGMS P30GM133893, R01GM117126).