

Addressing protein serial crystallography 36 GB/s data-rate challenge with FPGAs and GPUs

Saturday 7 October 2023 13:30 (20 minutes)

Serial crystallography [1] is a technique used at synchrotrons and X-ray free electron lasers to solve protein structures from random still diffraction images of thousands of small crystals. The technique is one of the most data intensive techniques at X-ray facilities. With novel detectors, like the 9 MPixel JUNGFR AU [2] currently commissioned at the Paul Scherrer Institute, it is possible to acquire a continuous stream of images at 36 GB/s.

Such large data rates challenge the current way images are handled in crystallography, i.e., it is no longer possible to save to disk storage every image and every pixel irrespective of their value for the scientific question [3]. On-the-fly data analysis and compression become a key to sustainable operations of high data rate detectors. Given a very high data throughput, such analysis requires computing accelerators, like field programmable gate arrays (FPGAs) and general-purpose graphical processing units (GPUs).

In this presentation, I will talk about our practical experience from implementing data science methods for on-the-fly analysis on computing accelerators [4]. I will give a practical example of spot finding algorithms that we implemented on GPUs and FPGAs (with high-level synthesis), highlighting differences in both approaches. I will also give an outlook of our early-stage developments in image analysis with machine learning methods.

[1] T. Weinert et al. (2019). *Science*, 365, 61-65.

[2] F. Leonarski et al. (2018). *Nat. Methods*, 15, 799–804.

[3] F. Leonarski et al. (2020). *Struct. Dyn.*, 7, 014305.

[4] F. Leonarski et al. (2023). *J. Synchrotron Rad.*, 30, 227.

Author: LEONARSKI, Filip

Presenter: LEONARSKI, Filip

Session Classification: Project presentations/demos