Using a supercomputer for massive parallel merging of XFEL reflections

Robert Bolotovsky[a], Aaron S. Brewster[a], Asmit Bhowmick[a], Derek Mendez[a] and Nicholas K. Sauter[a]

[a]Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Bldg. 33R0345, Berkeley, CA 94720, USA
rbolotovsky@lbl.gov

The vast amount of data produced by XFEL experiments requires high performance computing for merging highly redundant measurements of Bragg intensities in real time. We have developed algorithms in the XFEL data processing package cctbx.xfel to distribute all steps of the merging process (data loading, filtering, scaling, postrefinement, error analysis, and averaging) between multiple compute nodes and CPU cores. This software, utilizing Message Passing Interface (MPI), has been tested on the Cray XC40 Cori supercomputer at the National Energy Research Supercomputing Center. We processed two data sets acquired at the Linac Coherent Light Source. A lysozyme data set comprising 1.5×10^7 intensity measurements was loaded and averaged using 32 Cori Knights Landing nodes (2,176 CPU cores) in less than 5 minutes. A Photosystem II dataset with 7.5×10^9 intensity measurements was processed using 82 nodes (5,576 cores) in 20 minutes.