

Scaling Parallel I/O and Analysis to a Trillion Particles

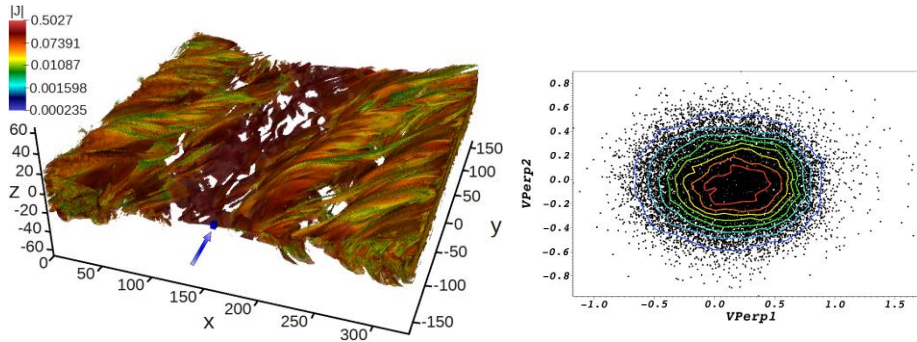
Prabhat (PI), Byna, Rubel, Wu, LBNL

Objectives

- Ability to analyze very large datasets quickly to enhance scientific understanding and discovery
- Enhance I/O on HDF5, a popular file format used by many application domains
- Demonstrate these capabilities on Trillion particle plasma physics simulation

Accomplishments

- Trillion particle plasma physics simulation conducted on 120,000 cores @NERSC
- Enhanced Parallel HDF5 obtained peak 35GB/s, and 80% sustained I/O rate
- FastBit was used to index 30TB timestep in 10 minutes and query in 3 seconds



Magnetic reconnection from a plasma physics simulation (Left). Scientists were able to query and find an asymmetric distribution of particles near the reconnection event (Right) using our software tools.

Impact

- Software enabled scientists to search and gain insights from the trillion particle dataset for the first time:
 - Confinement of energetic particles by the flux ropes
 - Asymmetric distribution of particles near the reconnection hot-spot

Surendra Byna et al, Parallel I/O, Analysis, and Visualization of a Trillion Particle Simulation. SuperComputing conference, SC'12, November 2012.