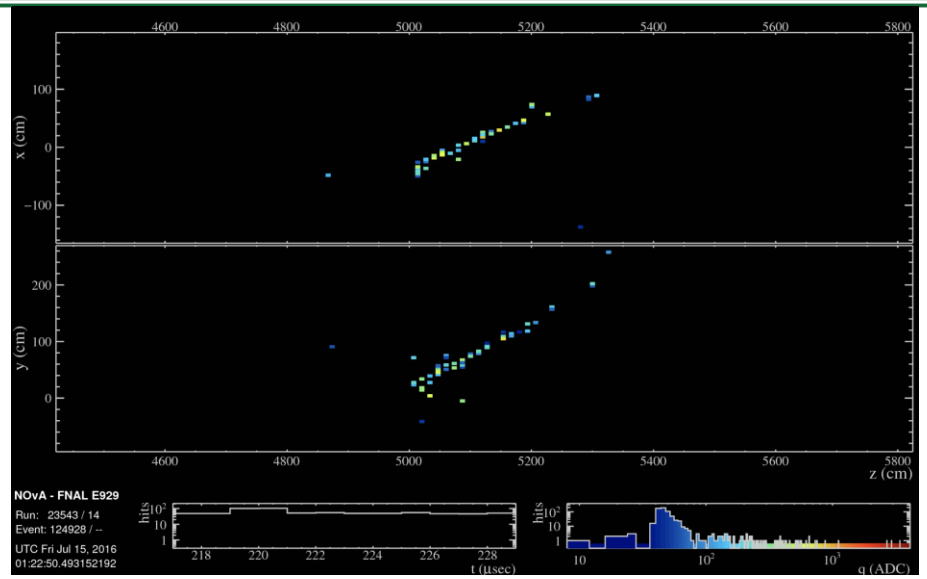
Tools for HPC-scale physics analysis

Scientific Achievement

A parallel data storage and access library for multi-terabyte physics data sets for use in HPC environments.

Significance and Impact

Obtain better understanding of neutrino oscillations by allowing improved systematic study of event-selection criteria used in analysis.



One of the 18 electron antineutrino appearance candidates selected by NOvA after analysis of 1.04 billion candidate interactions. NOvA observes a 4 sigma strong evidence for electron antineutrino appearance in a muon antineutrino beam

Research Details

- Using hephpc::hdf5 C++ library for writing HDF5 files from traditional HEP analysis programs.
- Demonstrated parallel reading speed in Python prototype code: >40 TB read in <20 seconds, using >76,000 KNL cores on Cori at NERSC.
- Demonstrated conversion of more than 4 TB of analysis data from NOvA's HEP-traditional analysis data format to our HDF tabular organization.
- The NOvA collaboration has taken ownership of the *art* framework module developed for this.
- Demonstrated ease-of-use of efficient high-level libraries (Python *pandas*), to support implicitly-parallel analysis code.

Currently published on BitBucket at https://bitbucket.org/fnalscdcomputationalscience/hep_hpc
art is described at http://art.fnal.gov
M.Paterno, *et al*. *Parallel Event Selection Performance on HPC Systems*. Paper presented at CHEP 2018, Sofia, Bulgaria. To be published in EPJ Web of Conferences (2019).