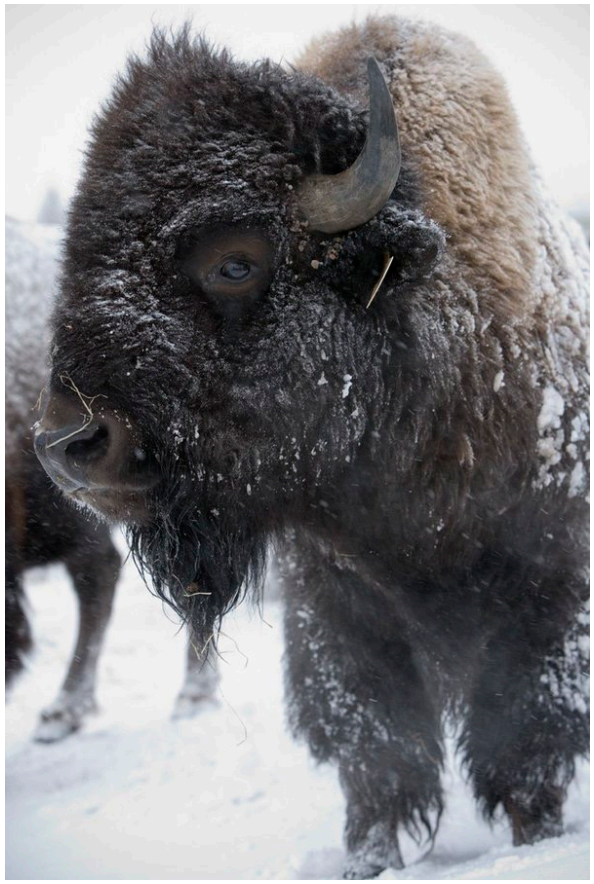
DUNE Software Tools

Thomas R. Junk

IceDUNE Workshop

June 17, 2021



Fermilab DUNE

Outline

- Challenges
- Software infrastructure
- Specific tools
- Computing facilities

The Challenge

- DUNE has many detector components:
  - Near Detector: ND-LAr, ND-GAr, TMS, SAND
  - Far Detector: FD-SP, FD-VD, others. These are very large detectors, and for convenience, we also support "workspace" versions with smaller volume (same size components though).
  - Prototypes
    - ProtoDUNE-SP
    - ProtoDUNE-DP
    - ICEBERG
    - 35t (software no longer supported)
- DUNE has many collaborators (>1000) spread across space and time.

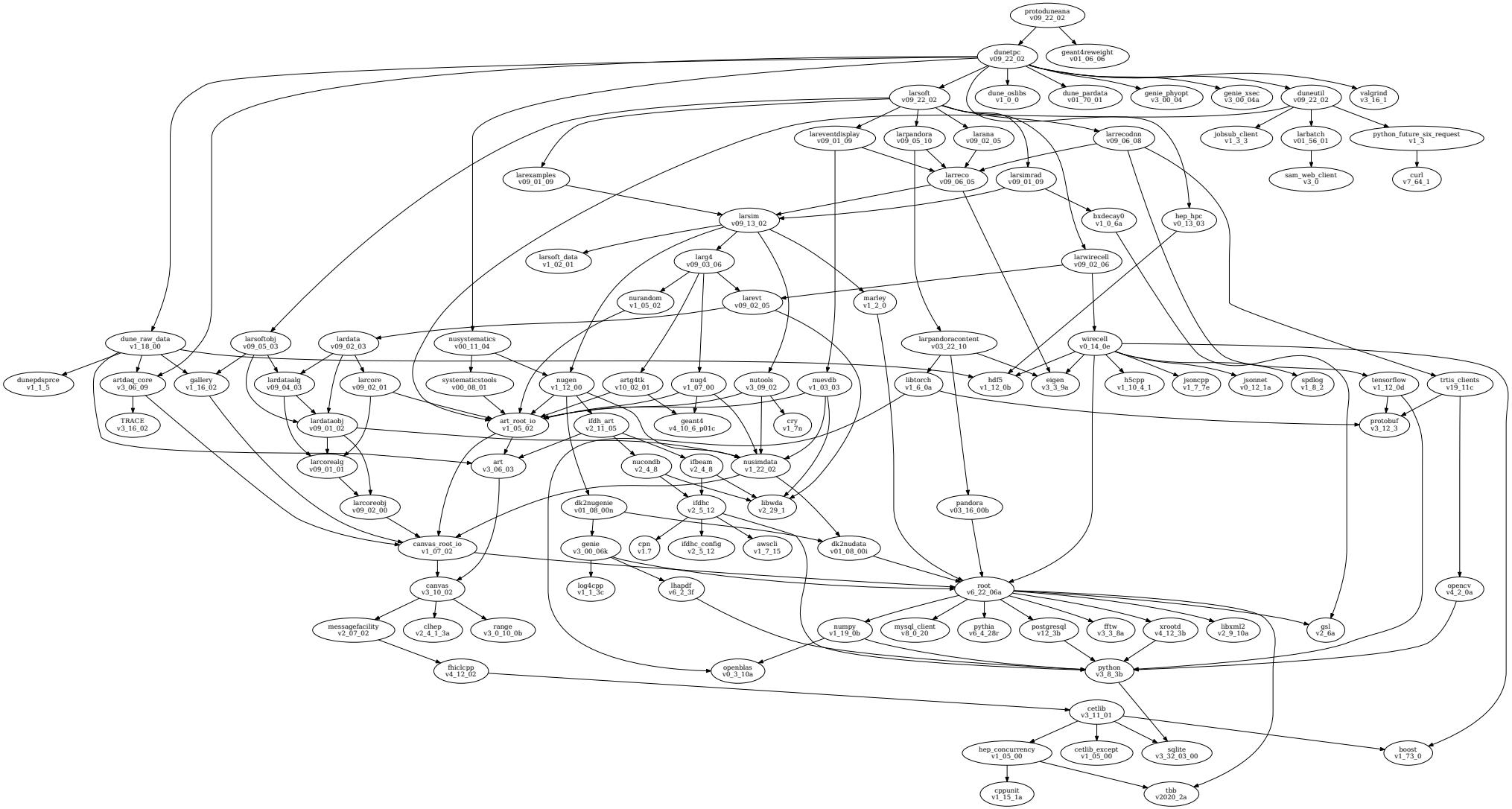
Challenges

- LArTPC data do not look like collider data
  - non-zero-suppressed waveforms: multiple GB per FD module per trigger
  - SNB trigger: ~150 TB per FD module. And a desire to analyze it quickly
- LArTPC reco is difficult.
  - somewhere between being a tracker and a calorimeter
  - 2D wire readout of 3D events requires special care
- Other LArTPC experiments have paved the way, and will come online:
  - ICARUS, ArgoNeuT, MicroBooNE, LArIAT, CAPTAIN, SBND and ICARUS again.

Software Infrastructure

- Strategy is to share as much infrastructure and algorithms as possible
- LArSoft, a software toolkit based on the *art* event-processing framework, for LArTPC simulation and reconstruction
  - https://larsoft.org
  - https://art.fnal.gov
  - LArSoft provides interfaces to GEANT4, ROOT, GENIE, and other components.
  - See the Event Generator Tools Workshop in 2020: https://indico.fnal.gov/event/22294/timetable/?view=standard
- Experiment code is built on top of LArSoft
- Code repositories are in GitHub and Fermilab's Redmine
- Software is distributed via CVMFS and downloadable tarballs on scisoft.fnal.gov, and is set up using UPS.

DUNE Software Stack: UPS Dependency Tree



Tool Highlights

- Data preparation
  - Unpacking: ROOT and HDF5 (in development).
    Already a DUNE-specific problem – cannot load all the raw data in memory for a trigger at once – need to stream it.
  - ADC mitigation and noise filtering
  - Pedestal finding and AC-coupling correction
- Wire-Cell Toolkit https://lar.bnl.gov/wire-cell/
  - 2D Deconvolution – wire vs. time
  - Produces 3D image of charge deposits
- Hit Finder – fits pulses to deconvolved waveforms
- SpacePoint Solver – 3D hypothesis for events. Disambiguation of data from wrapped wires.

Tool Highlights

- PANDORA, a multi-algorithm framework originally designed for ILC reconstruction.

  Eur.Phys.J.C 75 (2015) 9, 439 e-Print: 1506.05348

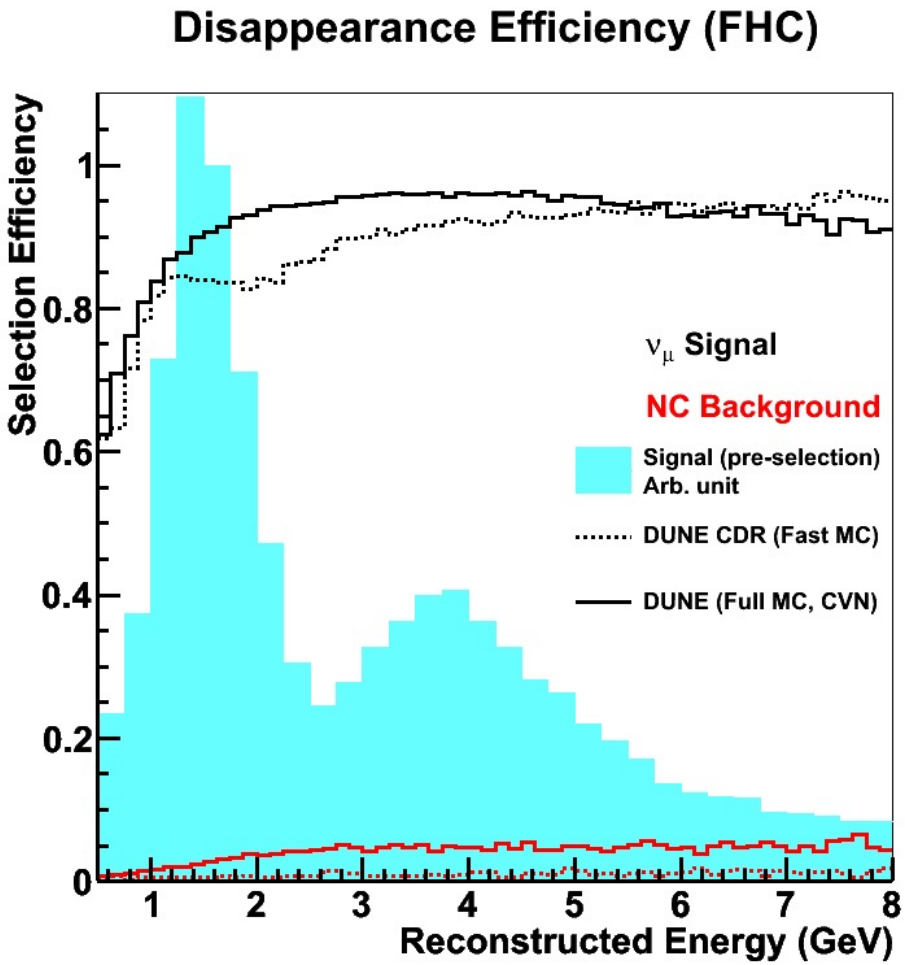
  - Pattern recognition: clustering, tracking and shower finding in 2D and 3D
  - Vertexing
  - Calorimetry
  - Particle ID
  - Output: particle hypotheses

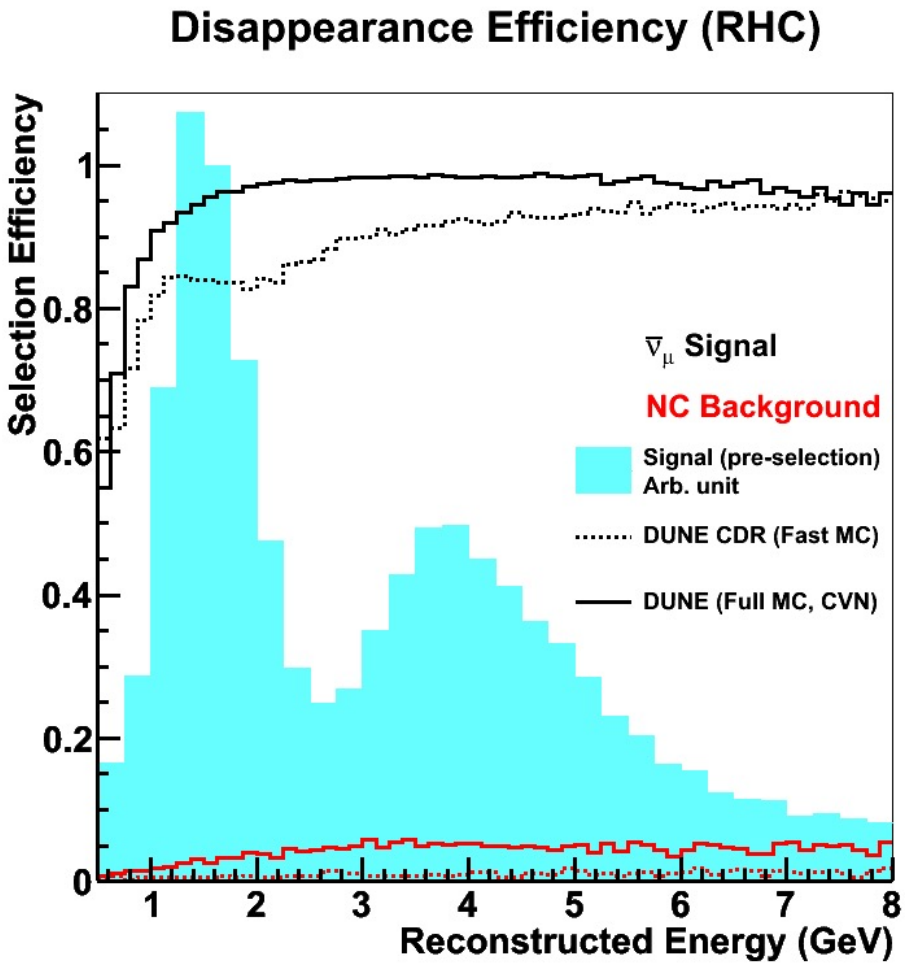
- Kalman-filter track fitter in LArSoft

Tool Highlights

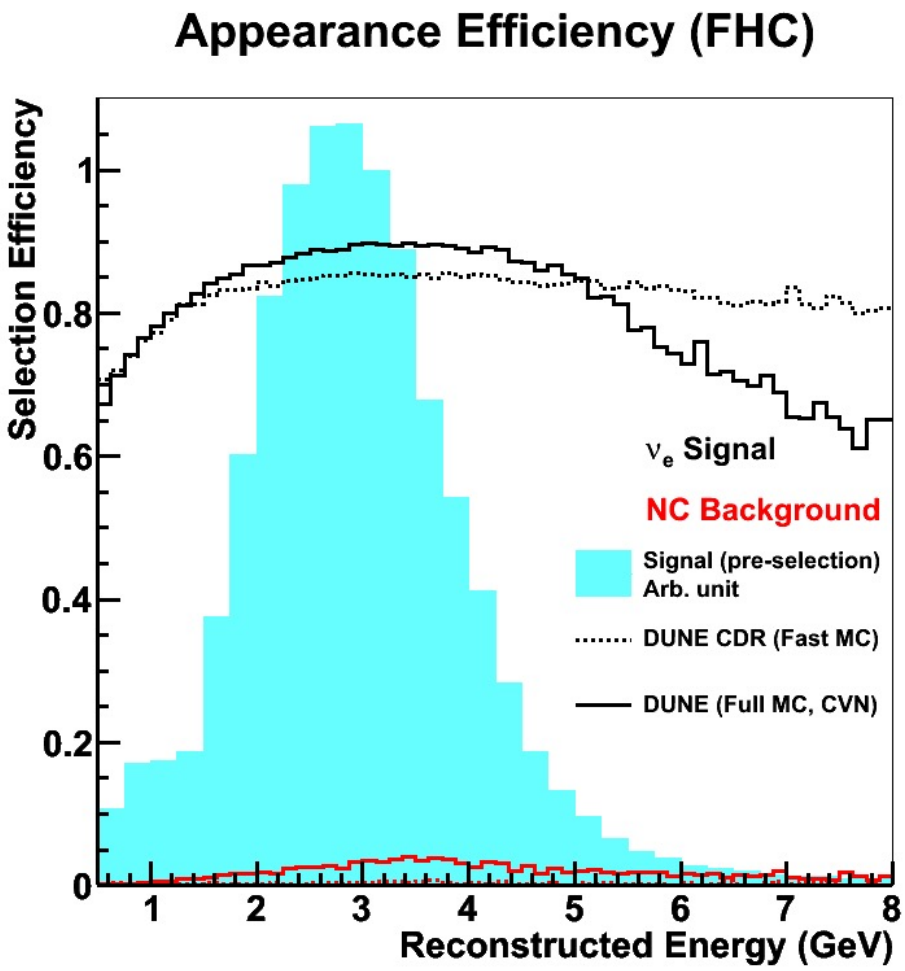
- Deep-Learning tools
  - Convolutional neural-networks used to classify pixels (time vs. wire) as track-like or shower-like
  - CVN event selection for ν_eCC events is the best performer
  - Semantic segmentation algorithms developed for MicroBooNE: https://arxiv.org/abs/2012.08513
  - I have seen starts on using such tools for DUNE, mostly in the ND efforts.
  - Pixel-based LAr-ND reco has started off with deep-learning tools first
  - SAND, TMS, and ND-GAr have more traditional pattern recognition and fitting tools.

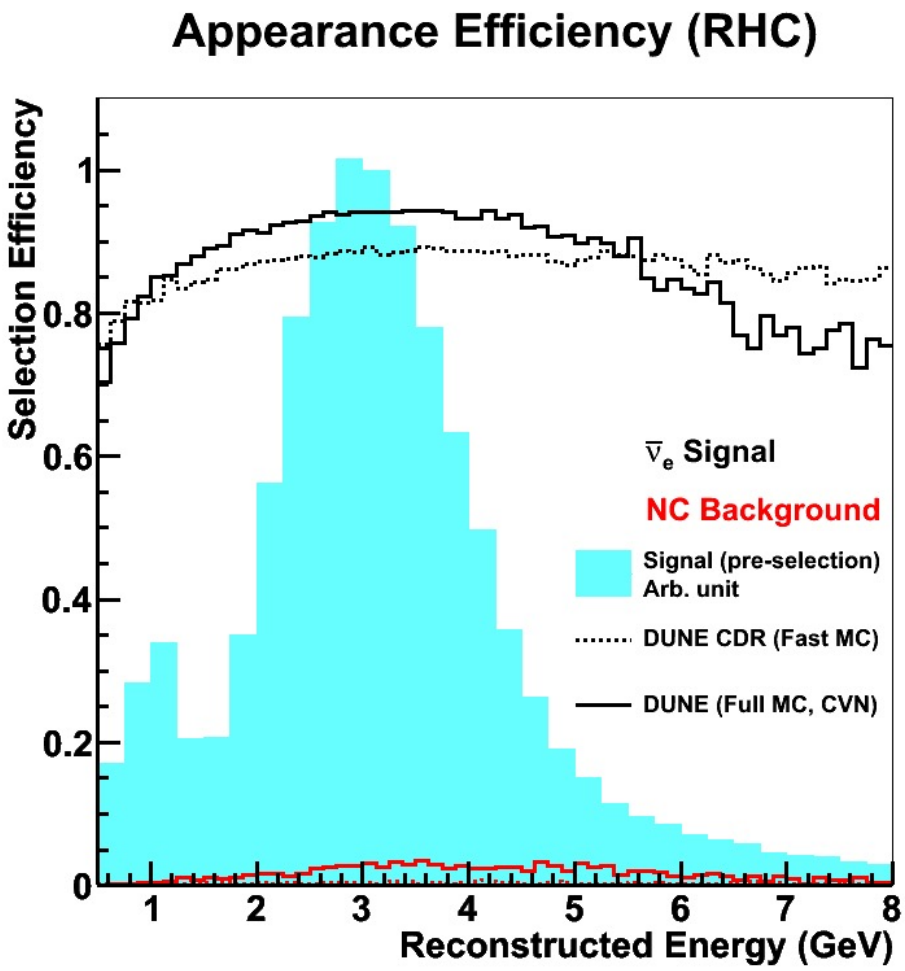
$\nu\mu$CC Performance with CVN





ν_eCC Performance with CVN





Geometry and Simulation Tools

- GGD: General geometry description tool

  https://github.com/brettviren/gegede
- Creates GDML files from python source
- Used extensively for the near detector

- edep-sim
  https://github.com/ClarkMcGrew/edep-sim
- A more lightweight interface to GEANT4 than the one in LArSoft
- Needed for a heterogeneous ND complex with loosely-coupled software.
- ND: SAND also uses FLUKA

Oscillation fits

- DUNE's Long-Baseline Physics Working Group uses CAFAna, which has been ported from NOvA
- CAF files have TTrees that are used to make histograms, used in binned likelihood fits.
- Generally, Feldman-Cousins is used to produce multi-dimensional confidence regions
- Can take quite a lot of CPU time! NOvA used 20M CPU hours to make delta-CP vs sin^2(theta_23) plots

https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiWq6-14JzxAhXKX80KHRhdBnEQFjAAegQIBBAD&url=https%3A%2F%2Findico.cern.ch%2Fevent%2F727646%2Fcontributions%2F3038971%2Fattachments%2F1683107%2F2705058%2FSousa_SciDac4_NOvA_HPC_BOFWorkshop.pdf&usg=AOvVaw04B3BcL-8ckX5Ogtr-jwxW

Computing Needs and Facilities

- We expect the DUNE experiment to accumulate ~30 PB/year of raw data.
- Much of it is from cosmic-ray triggers and SNB readouts.
- Additional data storage from processing, MC and analysis.
  - factor of 10-20 on the beam neutrino samples for all data requirements.
  - Still being designed
- CPU usage is currently in the tens of millions of hour/year, just working on prototypes and simulations.
- DUNE uses OSG and NERSC resources. Data storage at Fermilab, CERN, and contributing computing sites.