Heterogeneous Prompt Processing Challenges in DUNE

Steven Timm for the DUNE collaboration

Throughput Computing Week, UW Madison

10 June 2026

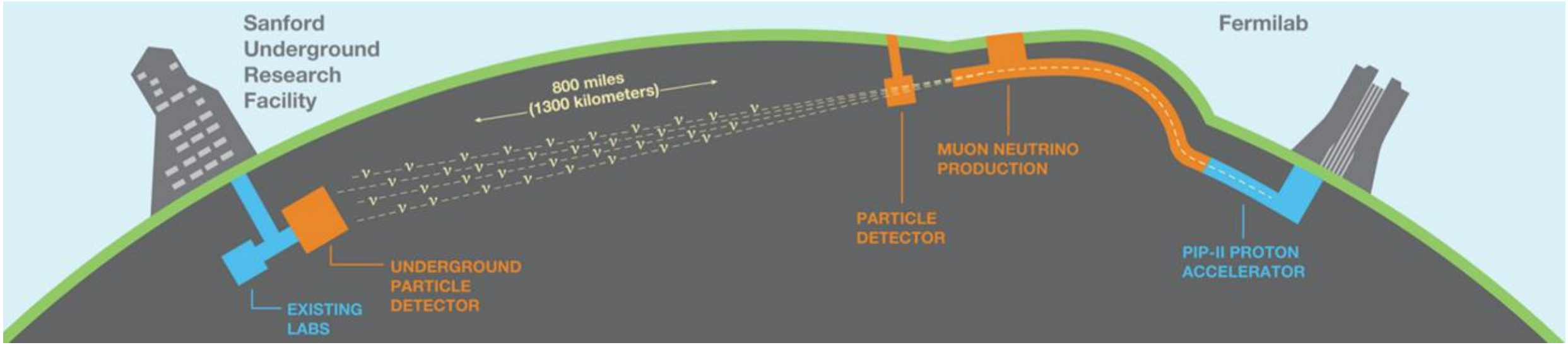
DisclAImer:

- This is not an AI talk like rest of the talks in this session
  - I originally submitted it as part of the “Multi Messenger Astronomy” category.
- DUNE is robustly using many types of AI in our reconstruction and analysis chains
- I am not qualified to speak on the merits or drawbacks of any of them.
- I will mention the various implications for our distributed computing and distributed data storage systems that arise from them, however.

A Bit of My Throughput Computing History:

- First Condor (not yet HTCondor) at Fermilab in 2003*, I was involved.
- My first Condor Week in 2005
- At the Open Science Grid opening ceremony in 2005 in Milwaukee, Wisconsin (as staff)
- Have been involved with the DUNE collaboration since its formation in 2015.
  - Lead or co-lead of Data Management for DUNE since 2017
  - And we have been accruing data all that time from 6 test beam runs thus far. (42PB)
- Most of my previous talks at this conference have either been asking some new exotic feature of HTCondor, or showing how said new feature works so well.
  - **Not this time!** HTCondor is behaving just fine and is expected to scale out to full DUNE scale without incident. Thanks to the dozens of team members over the years who helped that happen.

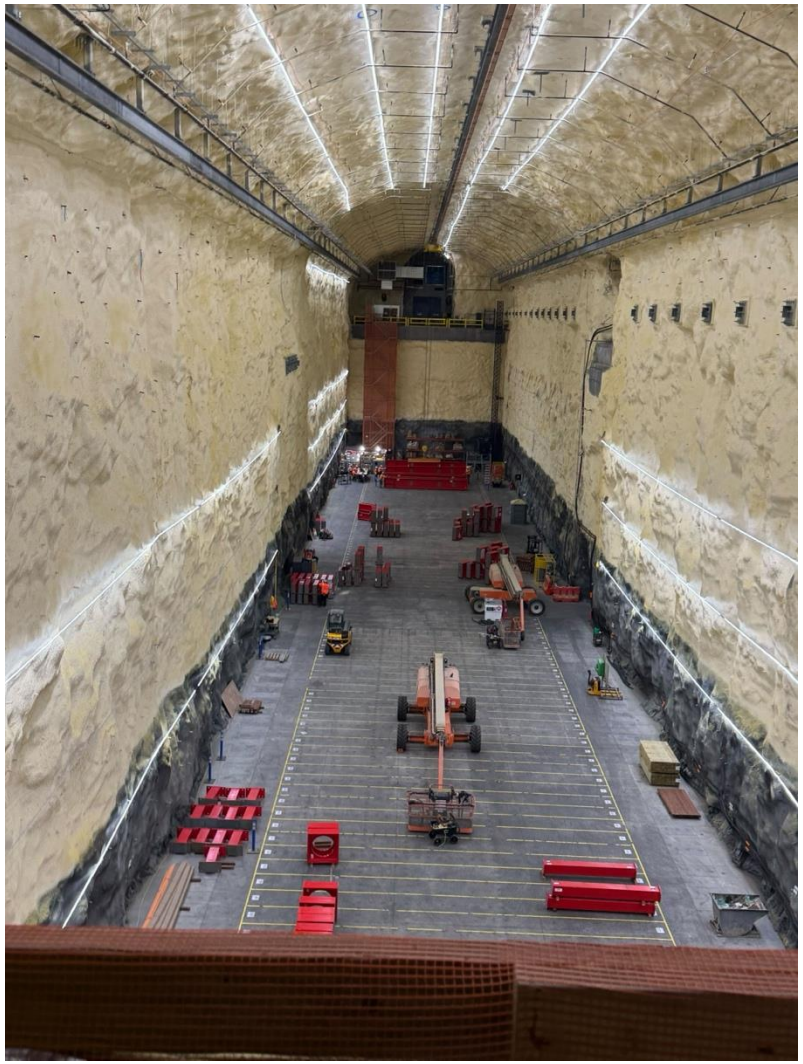
Deep Underground Neutrino Experiment (DUNE)



Neutrino oscillation experiment currently under construction

- 4 liquid argon far detector modules 1 mi underground detect beam neutrinos after oscillations at SURF in Lead, South Dakota (1300 km baseline)
- Near detector complex at 560 m from beam measures pre-oscillation flux & constrains systematics

DUNE in the Present:



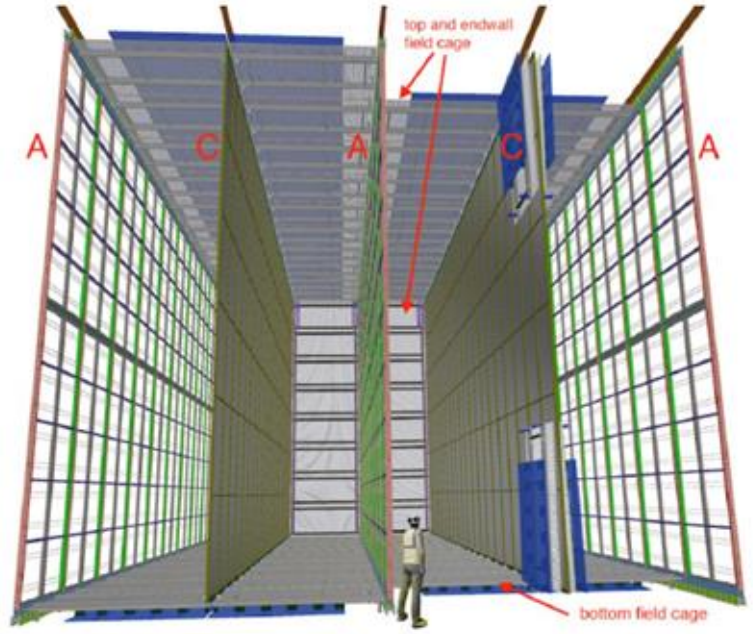
- The South Cavern of DUNE as it appeared on May 23, 2026
- 500 feet long, 90 feet tall, 4850 feet underground.
  - 2 such caverns
- Rapidly filling up with support steel for the cryostat
- Will start with 2 cryostats each with 10 kT fiducial of liquid argon
- Start of science expected early 2030.
- Very limited power to put computers in the cavern, or for that matter above ground.

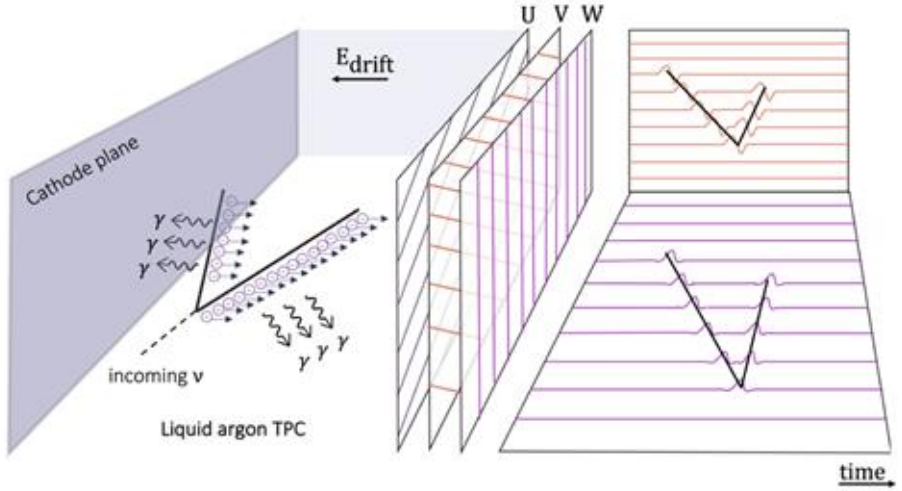
DUNE Detectors – Far Detectors

Horizontal-drift module (FD-HD)

- Wire-based readout: 150 Anode Plane Assemblies (APAs)
  - 2560 channels / APA → 384k total
- 14 bit readout
- 512 ns sampling period
- 2.6 ms readout window (beam/atmospheric triggers)

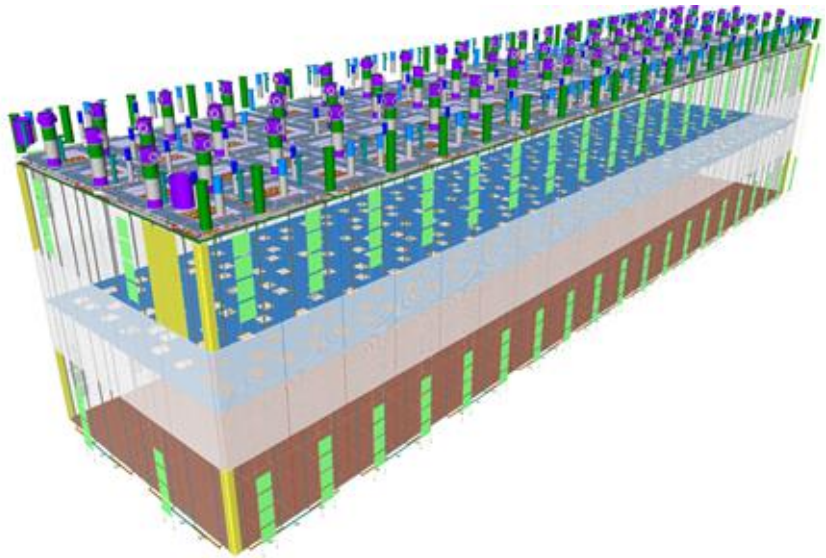
~3.4 GB / trigger record





Fermilab DUNE

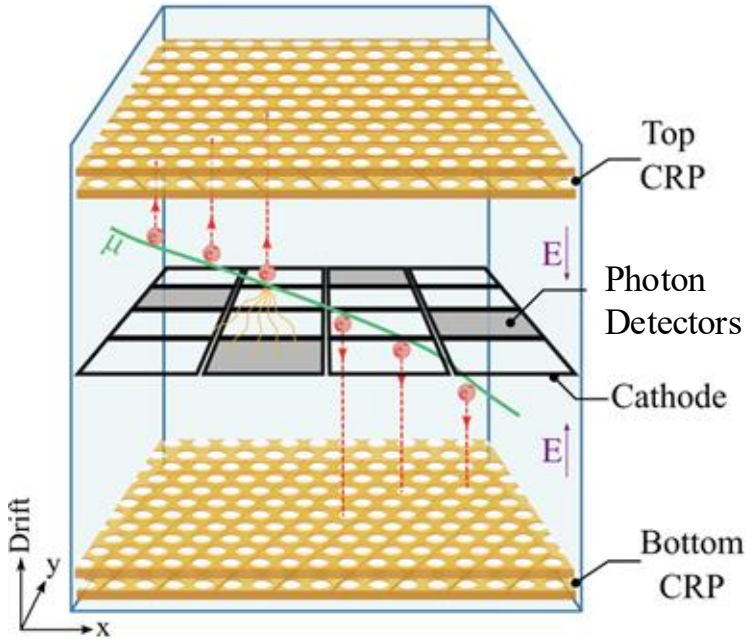
DUNE Detectors – Far Detectors



Vertical-drift module (FD-VD)

- Strip-based readout: 160 Charge Readout Planes (CRPs)
  - 3200 channels / CRP → 512k total
- 14 bit readout
- 512 ns sampling period
- 4.25 ms readout window (beam/atmospheric triggers)

~8 GB / trigger record



Huge event size. Can't hold it all in memory with legacy "ART" structures. New framework "PHLEX" will handle memory management much better.

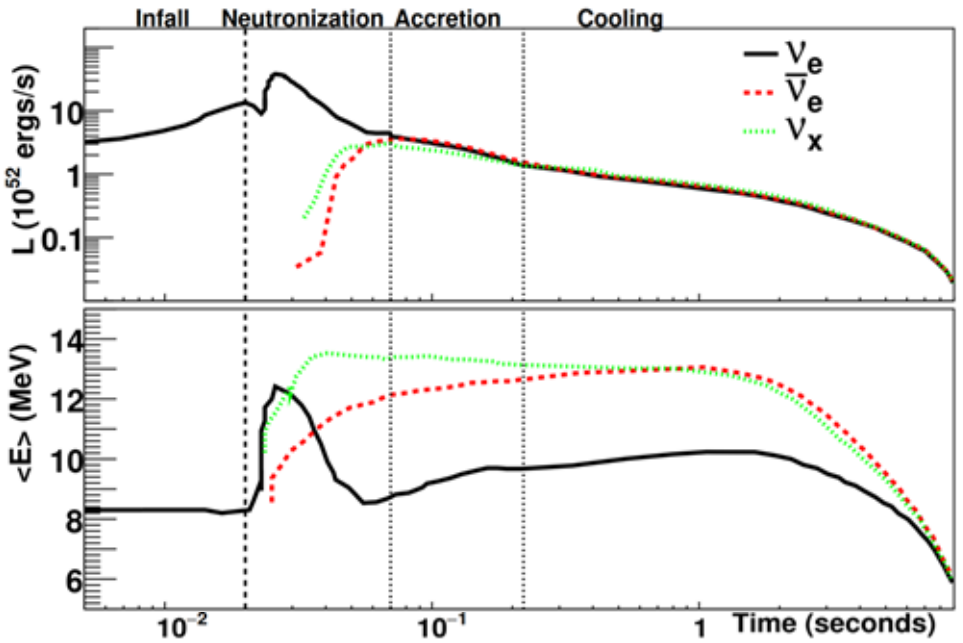
Special Challenge: SNB Readouts

SN1987A only neutrino-producer measured thus far

Time profile of neutrinos from supernova bursts provide a lot of information

- SNB trigger → read out detectors for 100 s → Per trigger record:
  - 140 TB from FD-HD
  - 180 TB from FD-VD
  - Record will be split into 1000s of files
  - 1 SN ”trigger” a month on average
- Unique challenge to move data from SURF to FNAL (or elsewhere), then process in a timely manner

Neutrinos from SNB 10 kpc from Earth



https://doi.org/10.1140/epjc/s10052-021-09166-w

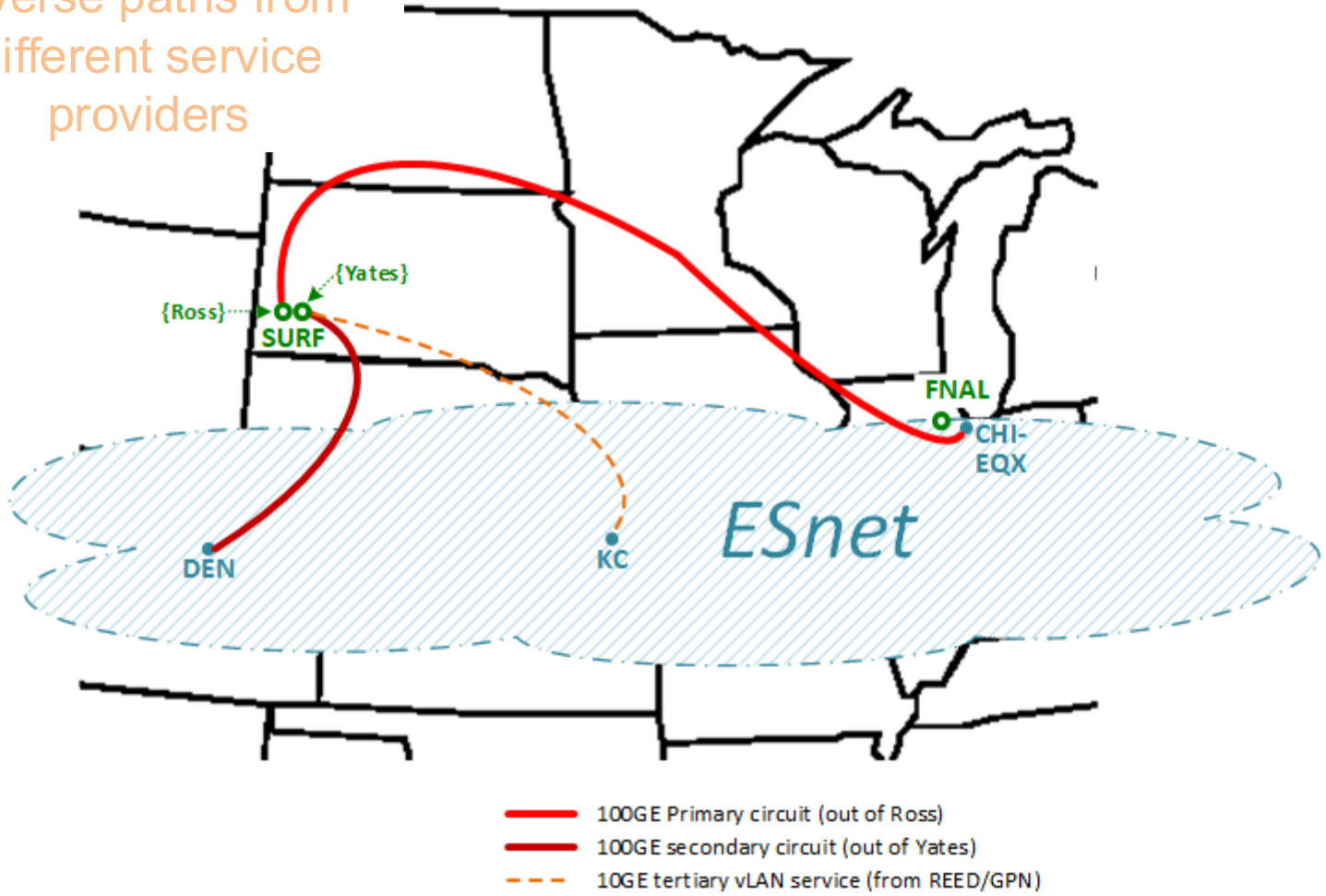
Note: many small ‘blips’ distributed across detector & occurring over a long time rather than a single, localized interaction from beam events

Studies currently under way to use AI/ML to improve signal isolation over background (radioactive decays).

Fermilab DUNE

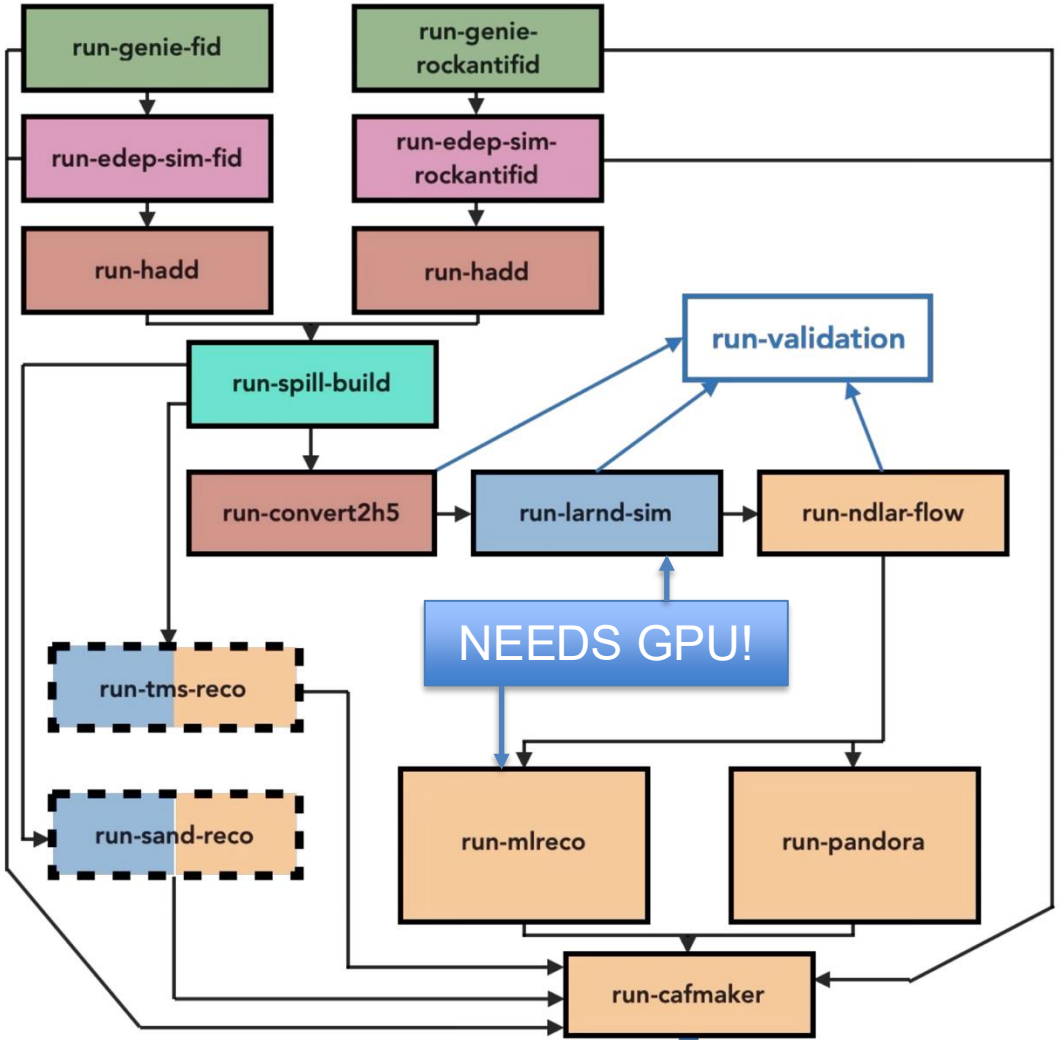
US DUNE Networking

Geographically-diverse paths from different service providers



- **Primary SURF-FNAL link currently live at 10Gbit/s**
- **100GBit/s coming in October**
- 1PB of disk cache available at FD
- 8Gbit/s steady state
- 100Gbit/s necessary to
  - Get supernova data out of mine
  - Catch up in timely manner if there's outage.
- About 4 hours to get a whole supernova trigger record to Fermilab.

Heterogeneous ML Workflows



- DUNE Near Detector WF features machine-learning based workflows for reconstruction of data “SPINE”
  - 3D space point readout with 25-100 tracks per trigger record.
  - Must be run, not just trained on GPUs
  - Already using 40K GPU node-hours at NERSC a year
  - Note the quick switch between CPU and GPU at several points of the workflow
  - This workflow hungry for both CPU and GPU memory
    - Continued struggle to make it fit in 40GB GPU RAM.

GPUs (and CPUs): how to get them fast

- Expecting GPU optimization in signal processing (Wirecell), track reconstruction (SPINE), and end-of-event inference as a service.
- Requires close collaboration with leadership class computing facilities,
  - Active NESAP program with NERSC at the moment
  - Need network routing to be able to send to those facilities depending on which one is up
  - New APIs currently in alpha for uniform submission interface—IRI (Integrated Research Infrastructure)
- Also for case of supernova readout need to get the data with most pointing info there first.
  - Some have suggested moving just the ”trigger primitives” which made the trigger.
- About 10 years ago we started looking at fast deployment.
  - HEPCloud@Fermilab and Condor Annex here both address the problem of getting a lot of resources up and commissioned fast.
  - May well need a “real time” Quality Of Service to be able to make this happen

OSG Usage of DUNE

- Dependent on the GlideinWMS Factories
  - We run our own frontends and Decision Engines
  - They have worked with us very well as we flip our remaining resources to Scitoken auth.
- Dependent on CVMFS
  - For code distribution and also for expanded container distribution
  - Also for OSDF/Pelican (what we still call Stashcache).
  - Used for files that are read by a lot of different jobs and not big enough to store in tape system that handles data files.
  - Streaming still an important use case.
- Depending on accounting system to get our info to APEL (soon to be PEAR).

Summary:

- DUNE CPU and Disk storage needs comparable to clusters that already exist now
  - But GPU acquisition will be a challenge, as will prompt access to CPU for burst processing
- Our challenge is fast movement and processing of very big events
  - Coupled with short jobs that run quickly and produce a lot of output
- Currently finding the scaling limits of all our data management services
  - Rucio, MetaCat, FTS
- After that we begin to benchmark the underlying storage
  - How many streaming connections, total I/O, etc.
- Very thankful to the OSG, its staff, and its member sites who help keep us going.
- Be watching the SNEWS network a few years from now, hopefully we will have news.