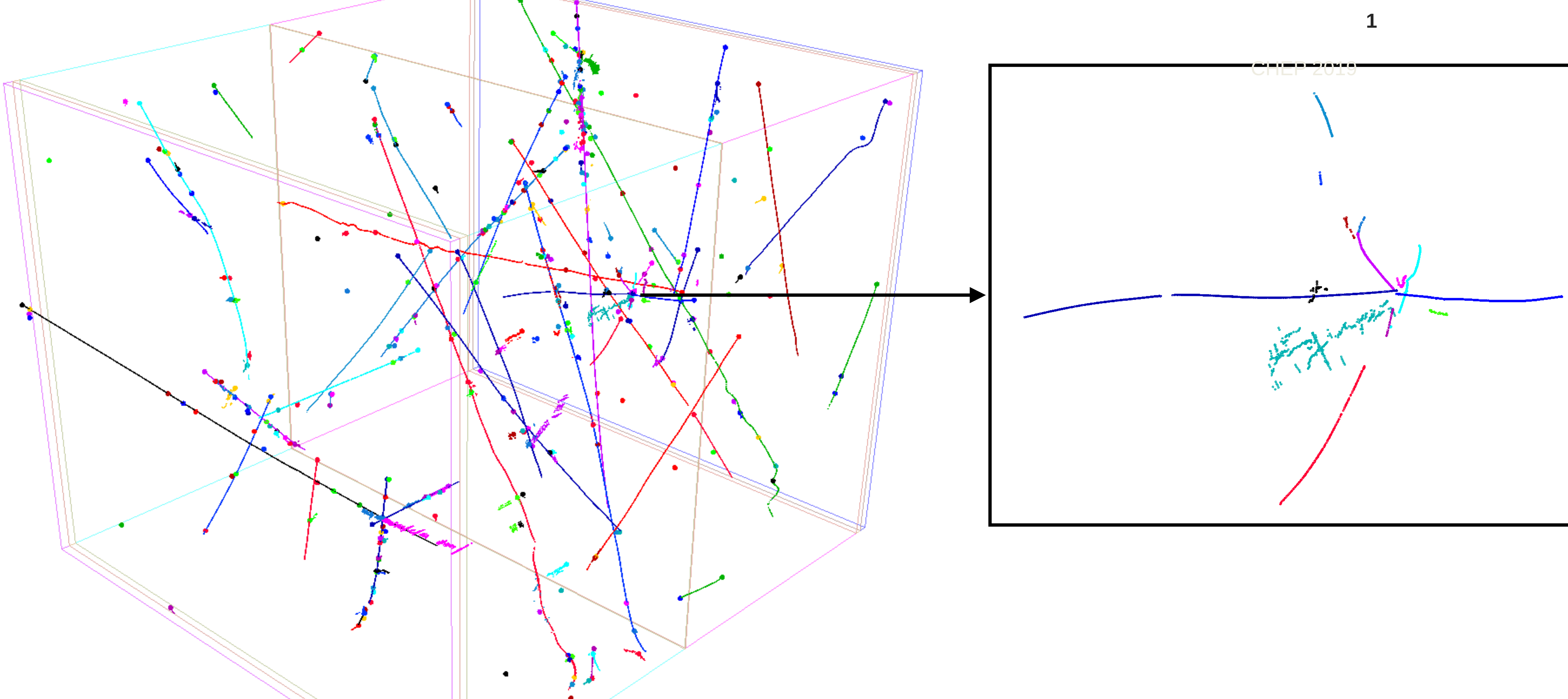
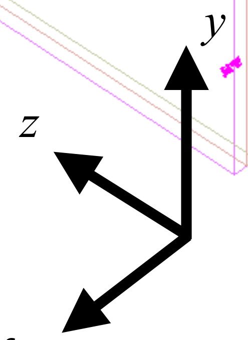




DUNE COMPUTING

Heidi Schellman, Oregon State University
For the collaboration

CHEP 1994 - San Francisco

- Lots of discussion of WWW in the parallels
- I brought my first laptop
- I gave a poster on the Linux Port of FNAL-E665 code

CHEP 1994 - San Francisco

- Lots of discussion of WWW in the parallels
- I brought my first laptop
- I gave a poster on the Linux Port of FNAL-E665 code
- Tom Nash gave a conference summary saying HEP computing was becoming irrelevant.

DUNE Computing

- The experiment
- Computational Challenges
- Results from prototypes
- Towards common solutions

DUNE's main purpose is to understand neutrino properties

ν_e ν_μ ν_τ

Flavor Basis
(Interactions)

ν_1 ν_2 ν_3

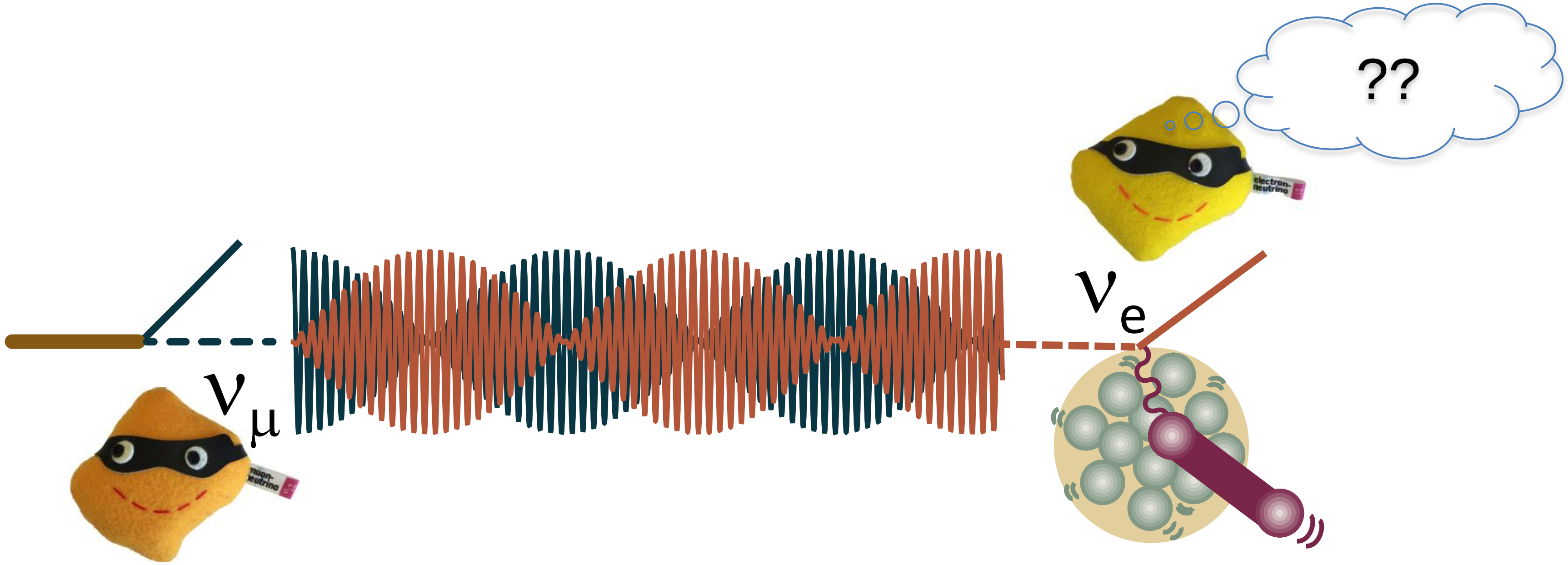
Mass Basis
(Motion)

2 different views of the same neutrinos

Particlezoo.net

The quantum wavelength of a 2 GeV muon neutrino is ~ 10^{-16} m
But it is actually a superposition of the 3 mass types of neutrinos which have slightly different wavelengths – the beat wavelength between the types is about 2000 km.

Bottom line – propagation can change a muon type neutrino 
into an electron type neutrino 

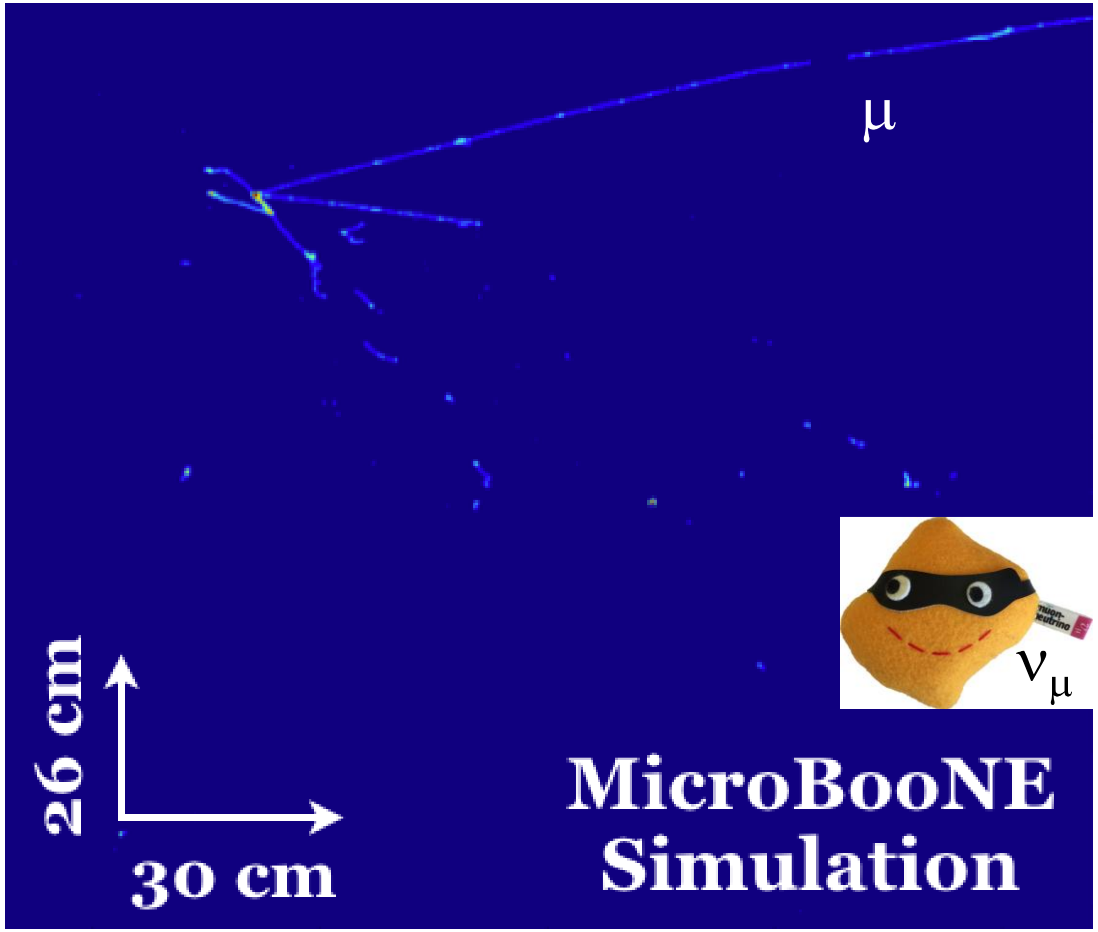


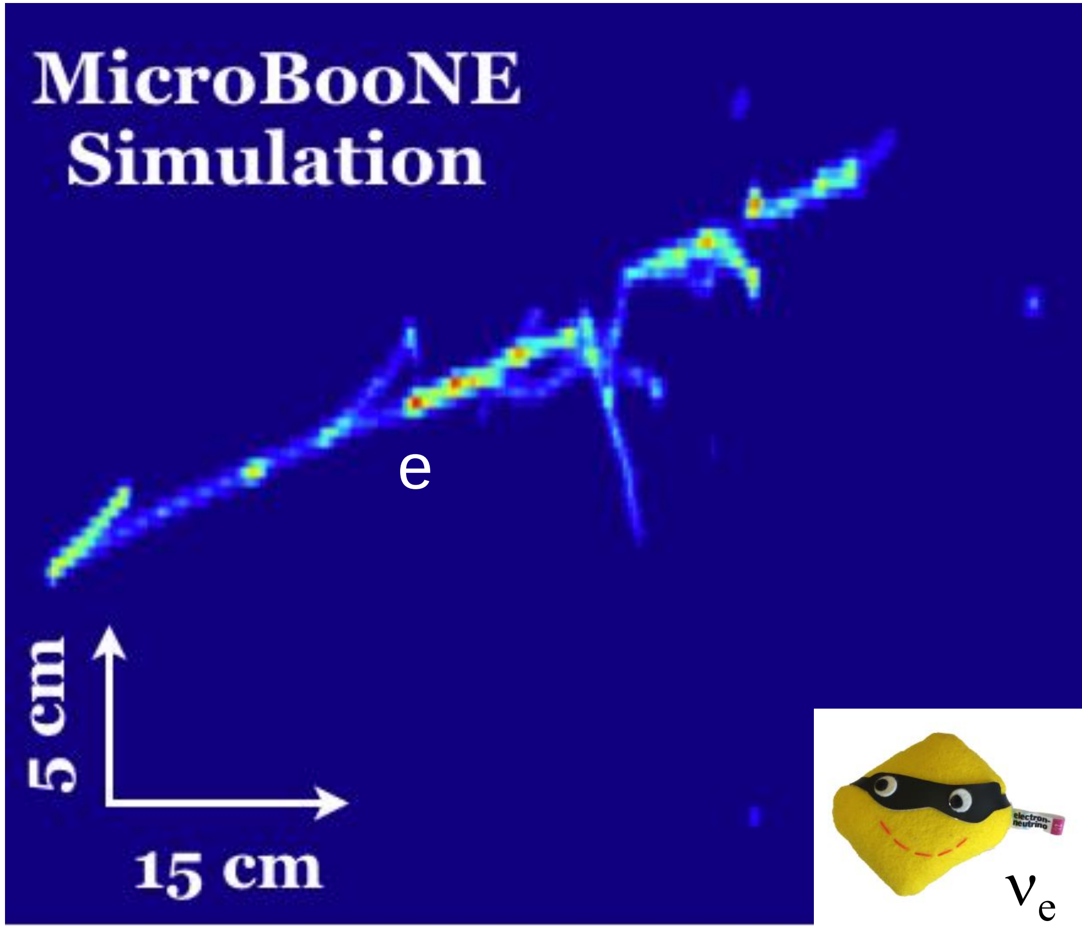
Put a huge LAr detector “DUNE” in the Homestake Gold Mine

Make a very powerful neutrino beam

Run for 10 yrs.

Final state – muon or electron?





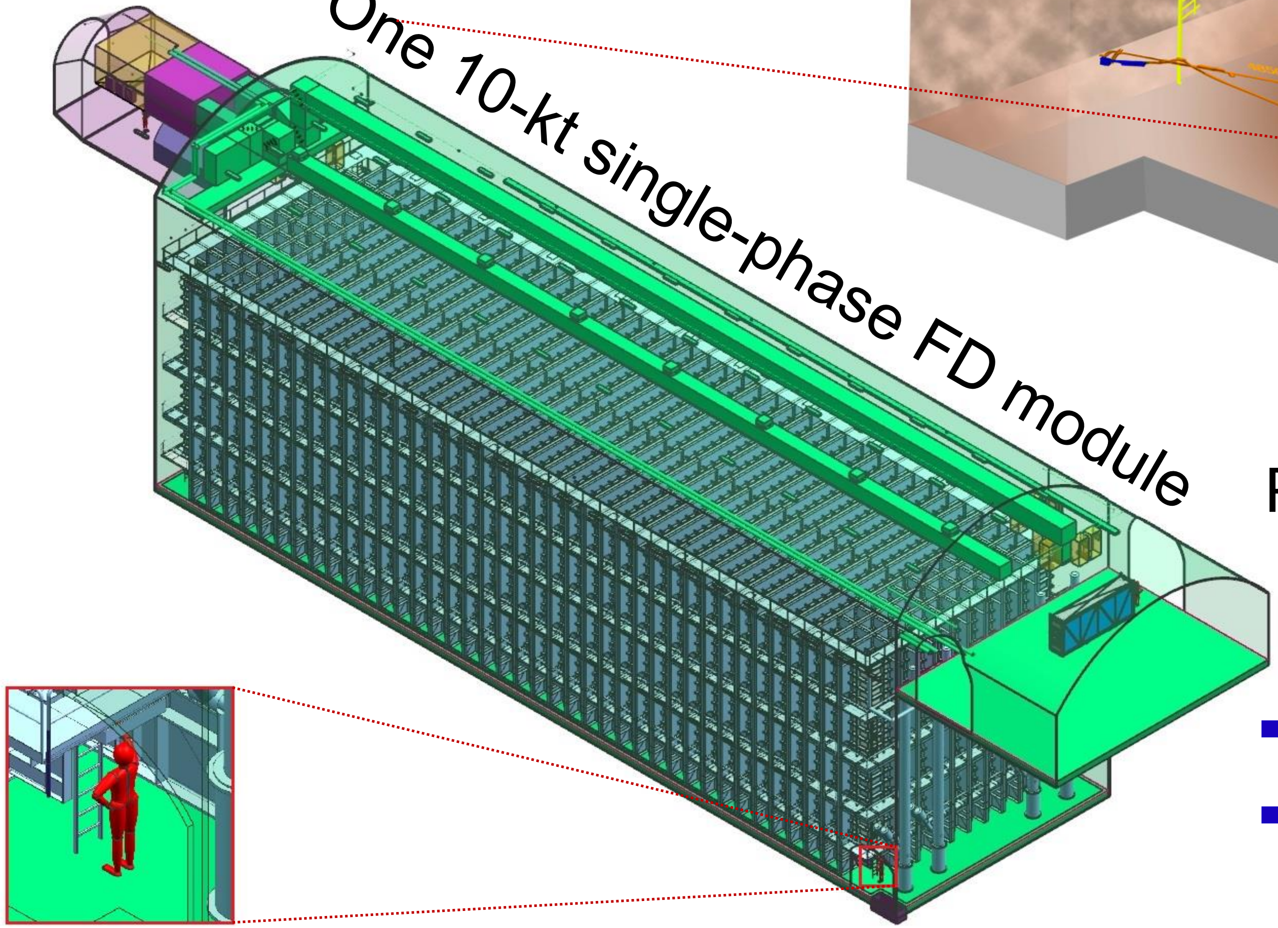
Problem is you need to instrument ~50,000 m^3 with cm granularity and no dead material

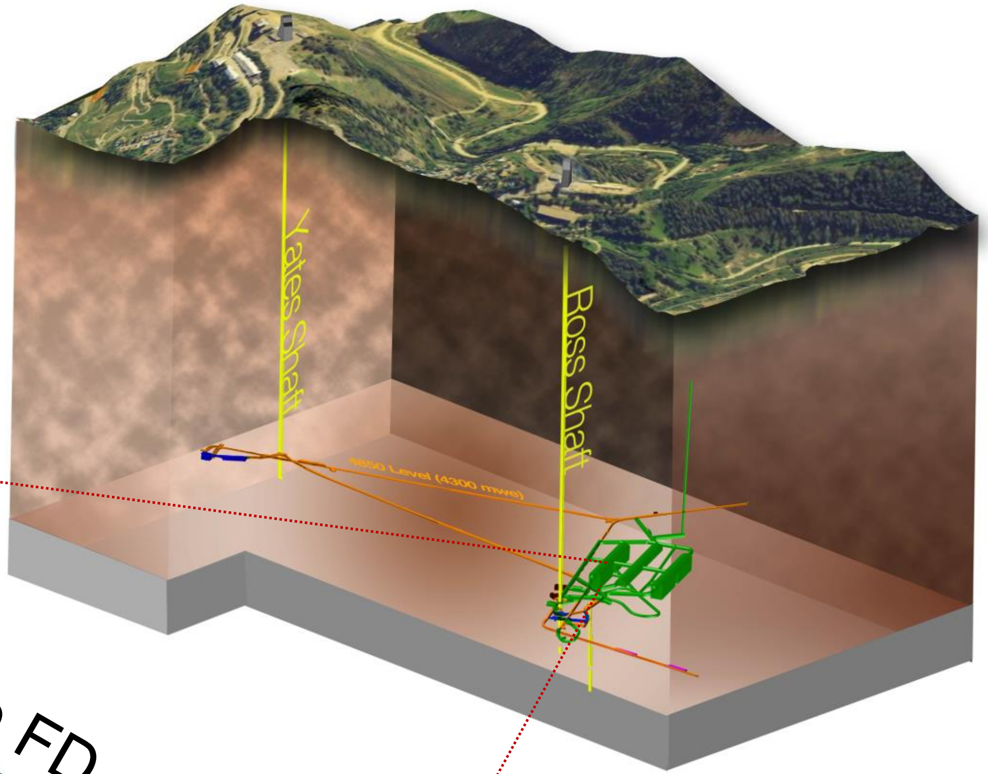
Far Detector

40-kt (fiducial) liquid argon time projection chambers

- *Installed as four 10-kt modules*

One 10-kt single-phase FD module

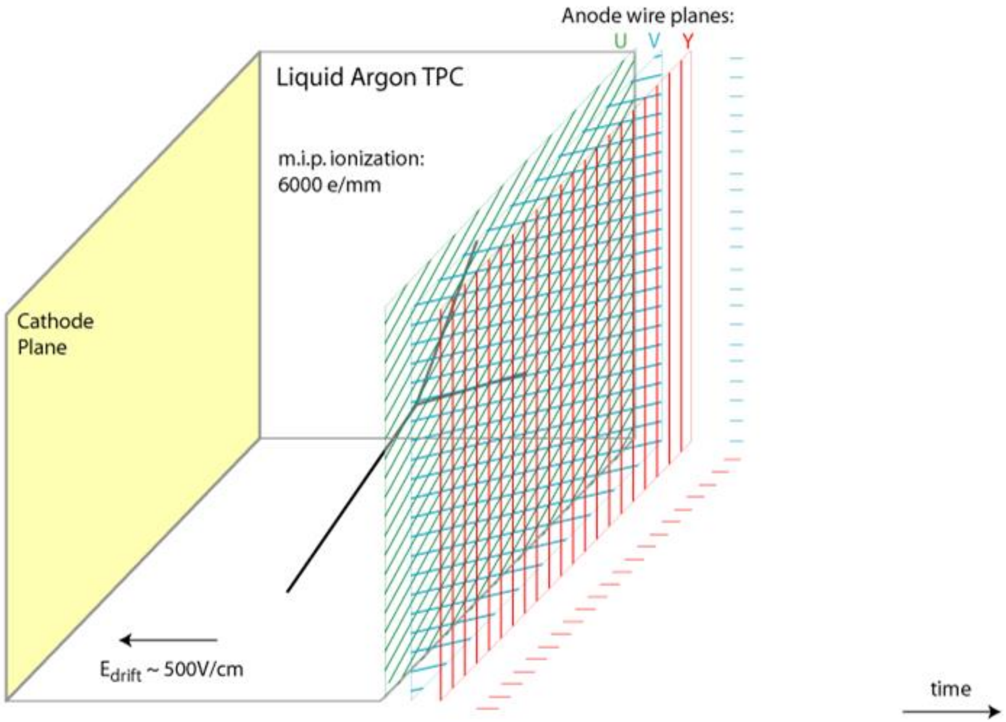


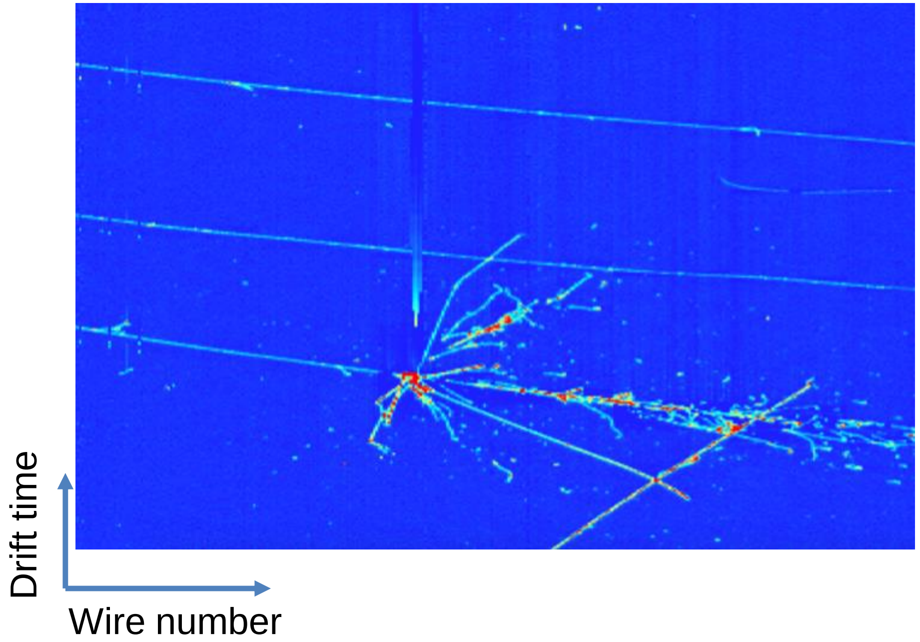


Sanford Underground Research Facility (SURF)

- **4850' level at SURF**
- First module will be a **single phase LAr TPC**

Liquid Argon Time Projection Chamber (LArTPC)



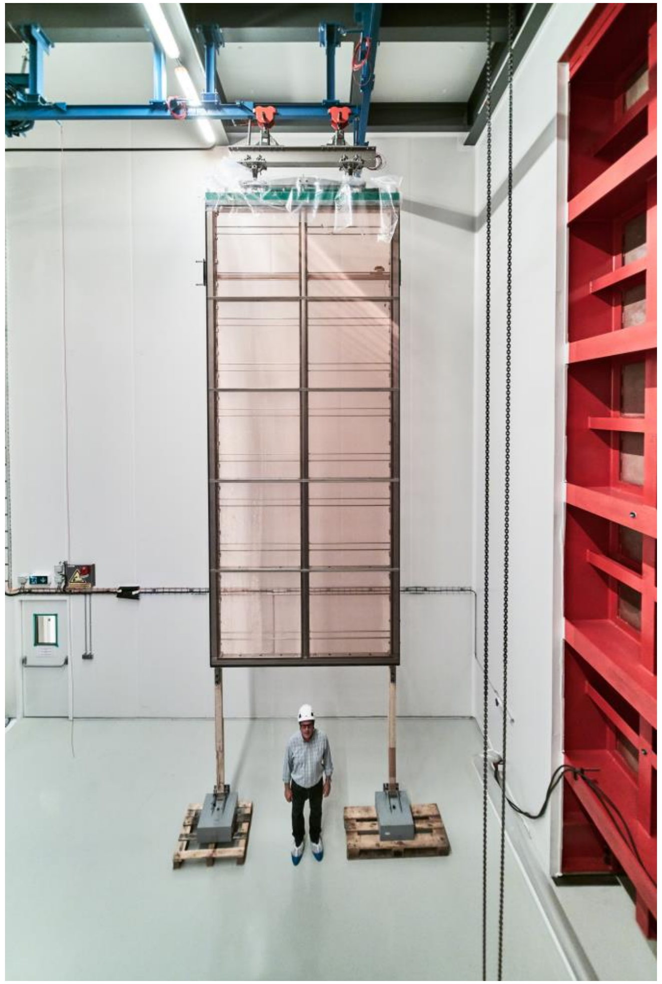


- **The DUNE far detector will consist of four LArTPC detector modules**
  - High spatial and calorimetric resolutions
  - Each module has a total mass of 17 kton, located 1.5 km underground
  - Prototyping is critical for such a big detector --> ProtoDUNE SP and DP

LAr TPC data volumes

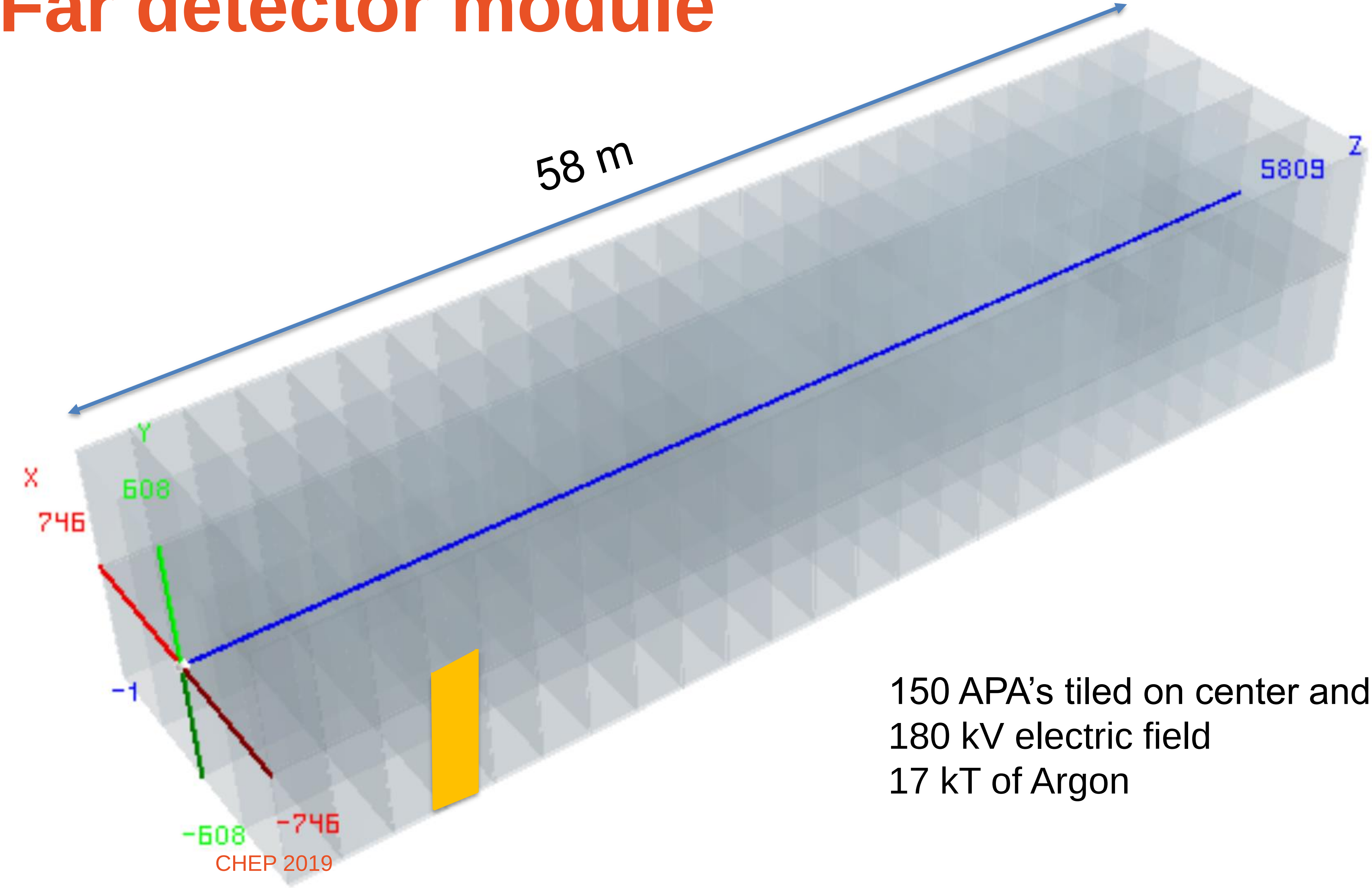
- The first far detector module will consist of 150 **Anode Plane Assemblies (APAs)** which have 3 planes of wires with 0.5 cm spacing. Total of **2,560** wires per APA
- Each wire is read out by 12-bit ADC's every 0.5 microsecond for 3-6 msec. Total of **6-12k** samples/wire/readout.
- Around 40 MB/readout/APA uncompressed with overheads → **6 GB/module/readout**
- 15-20 MB compressed/APA → **2-3 GB/module/readout**
- Read it out ~5,000 times/day for cosmic rays/calibration → **3-4PB/year/module (compressed)**

(x 4 modules x stuff happens x decade) = ….

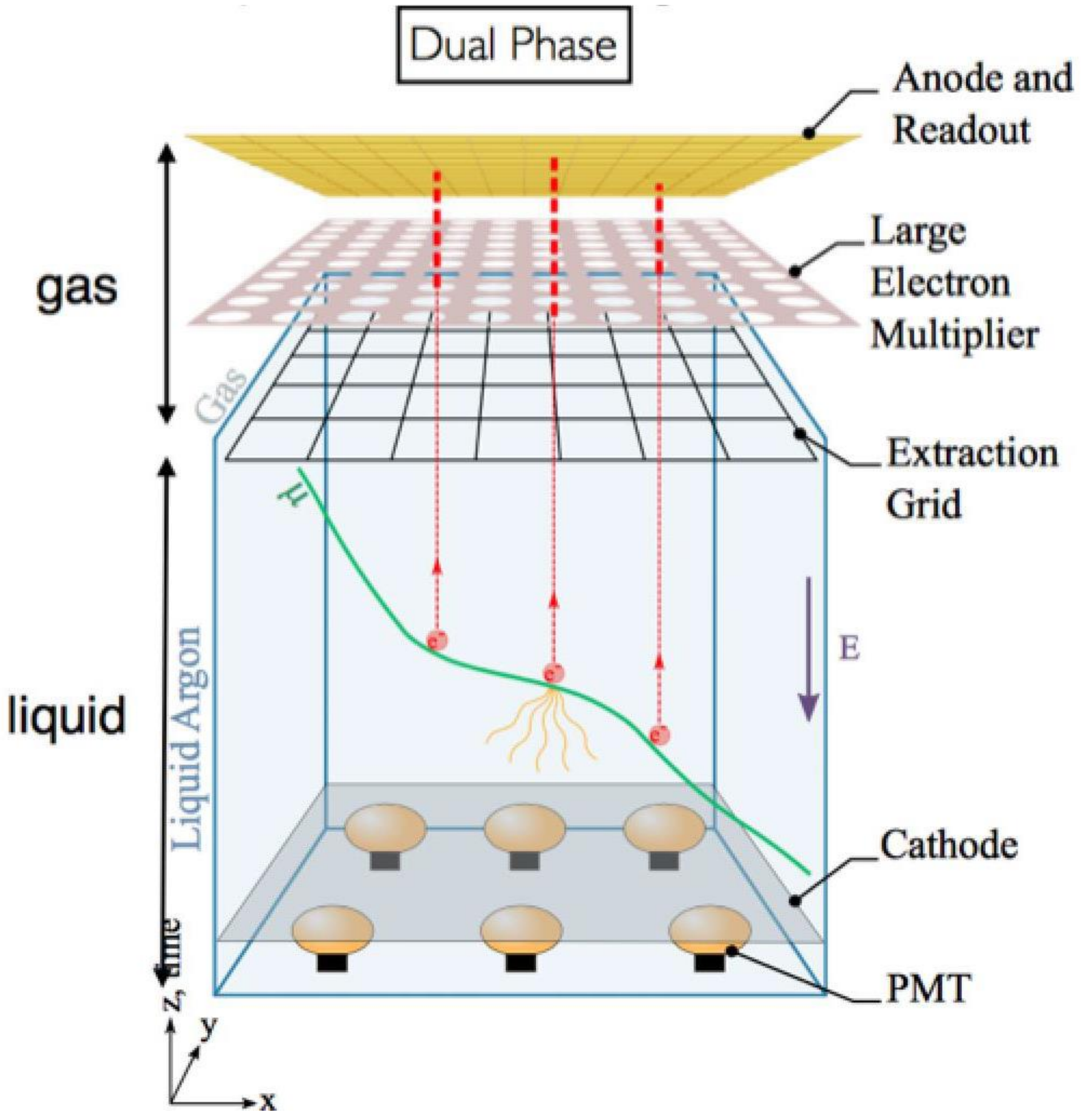


1 APA – 2,560 channels
150 of these per FD module

1 Far detector module



150 APA's tiled on center and walls
180 kV electric field
17 kT of Argon



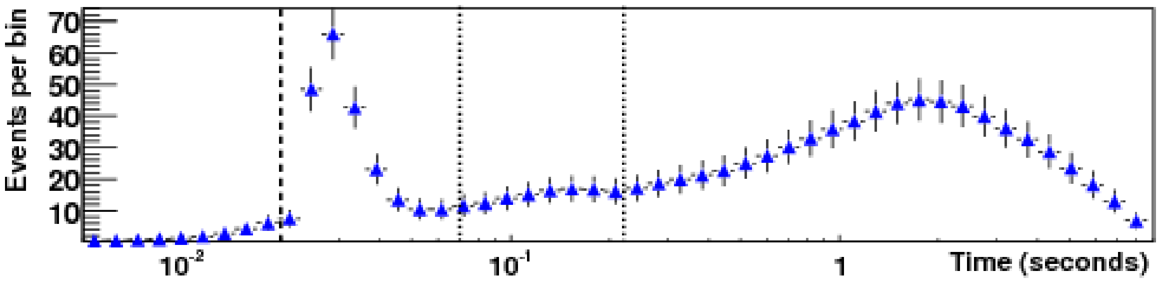
Dual Phase Design

Dual-phase Design: long drift (up to 12 m), high S/N:

- Vertical drift → electrons leave the liquid and are amplified by avalanches in micro-pattern detectors LEM (Large Electron Multipliers) operating in pure argon gas.
- Light is readout performed with an array of cryogenic photomultipliers below the cathode

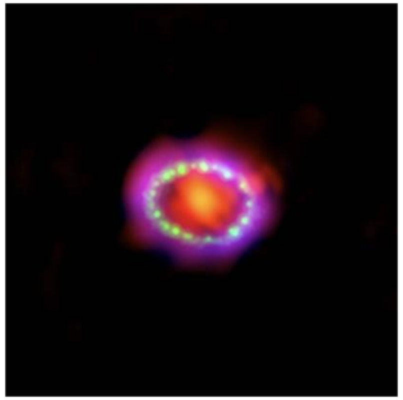
To make it more interesting

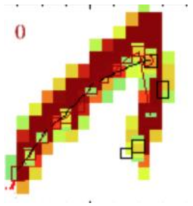
- DUNE should be sensitive to nearby (Milky Way and friends) supernovae. Real ones are every 30-200 years but we expect 1 false alarm/month



- Supernova readout = 100 sec, one trigger/month
- 100 sec readout implies
  - 1 channel = 300 MB uncompressed
  - 1 APA = 768 GB uncompressed
  - 1 module = 115 TB uncompressed
  - 4 SP modules = **460 TB** … takes 10 hrs to read at 100 Gb/s
  - Dual Phase technology has higher S/N → smaller per module
- Some calibration runs will be similar in scope….

Supernova 1987A





30 MeV ν_eCC

10 MeV NC

$\nu + A \rightarrow \nu + A^*$

DUNE FD-Data for Supernova

Pack 150 5 ms APA readouts into a 6 GB file

Ship 20,000 time slices (x 4 modules)

Logistical problem

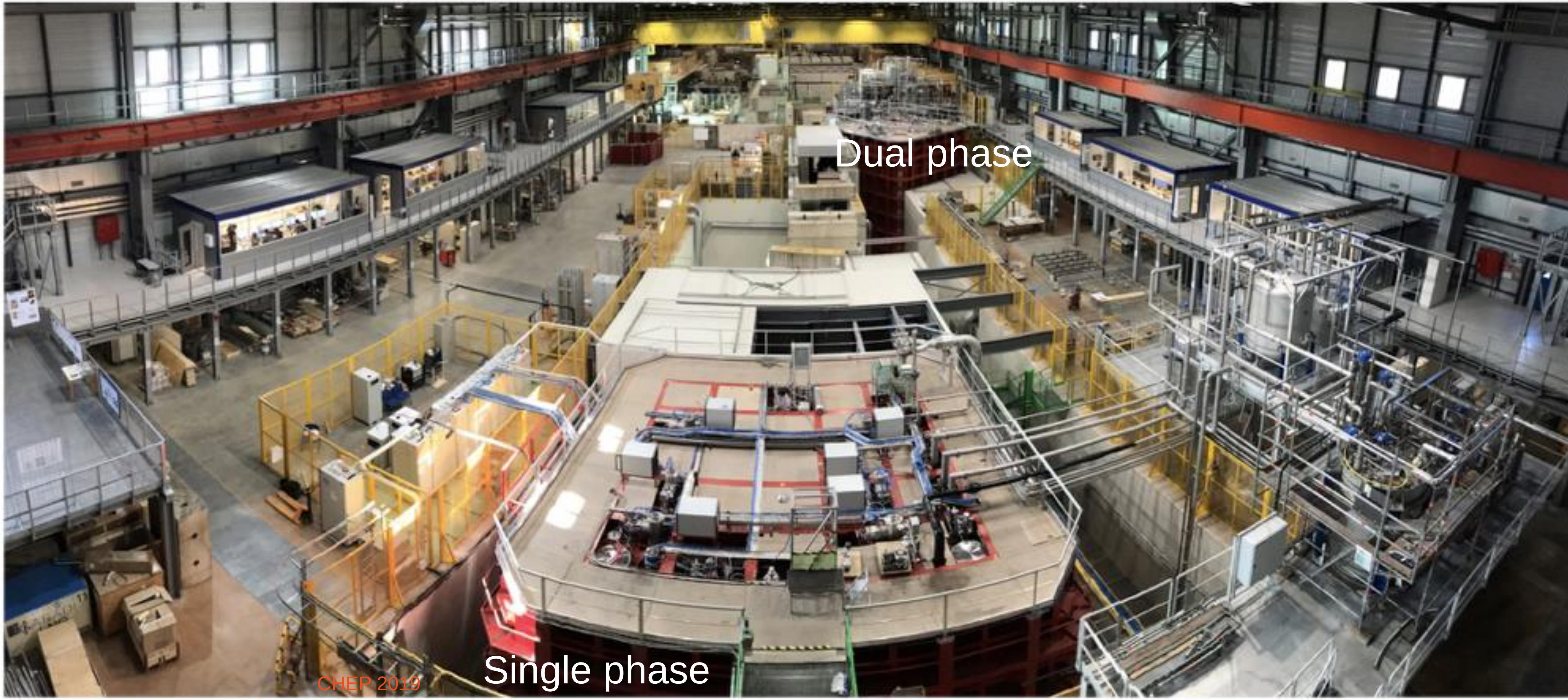
- A ”normal” HEP CPU has ~ 2GB of memory
  - Enough for 1-2 APA
  - Need to split things up to process
- We can split the data up into 1,000,000 40MB APA chunks but to understand an interaction, we have to be able to put them back together again.
- If we split things up, we need to find all the containers to put the car back together.

Solutions

- ProtoDUNE tests
  - Infrastructure
  - Algorithms
- Future

Build “small” prototypes @CERN

~1kt LAr-TPC’s at CERN



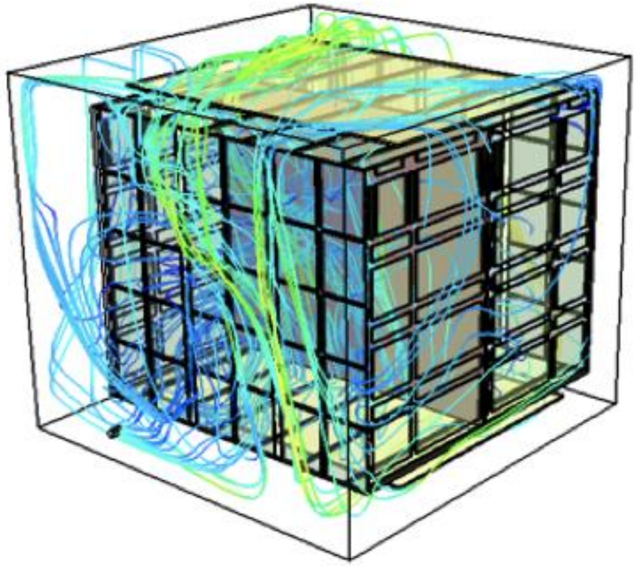
This is not your dad's LHC expt.

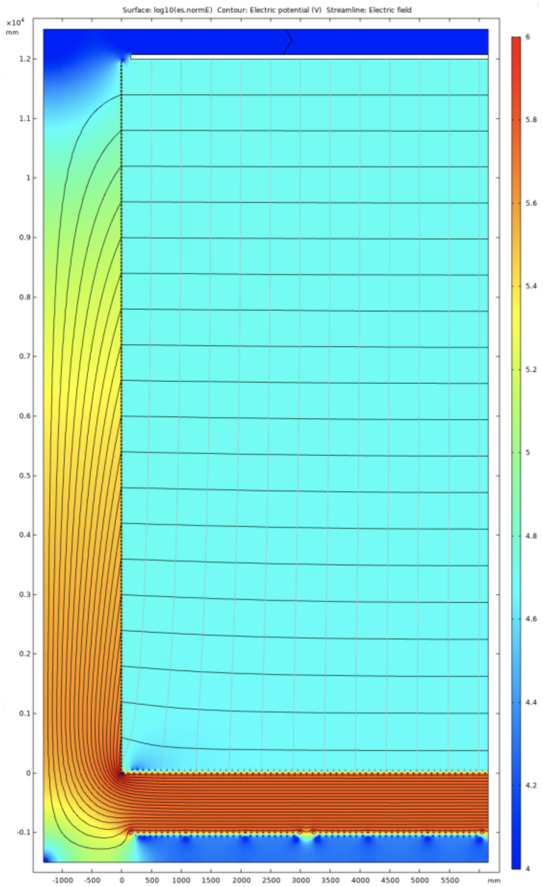
Good news:

- Volume filled with uniform material
- `geant4` really likes this

Bad news:

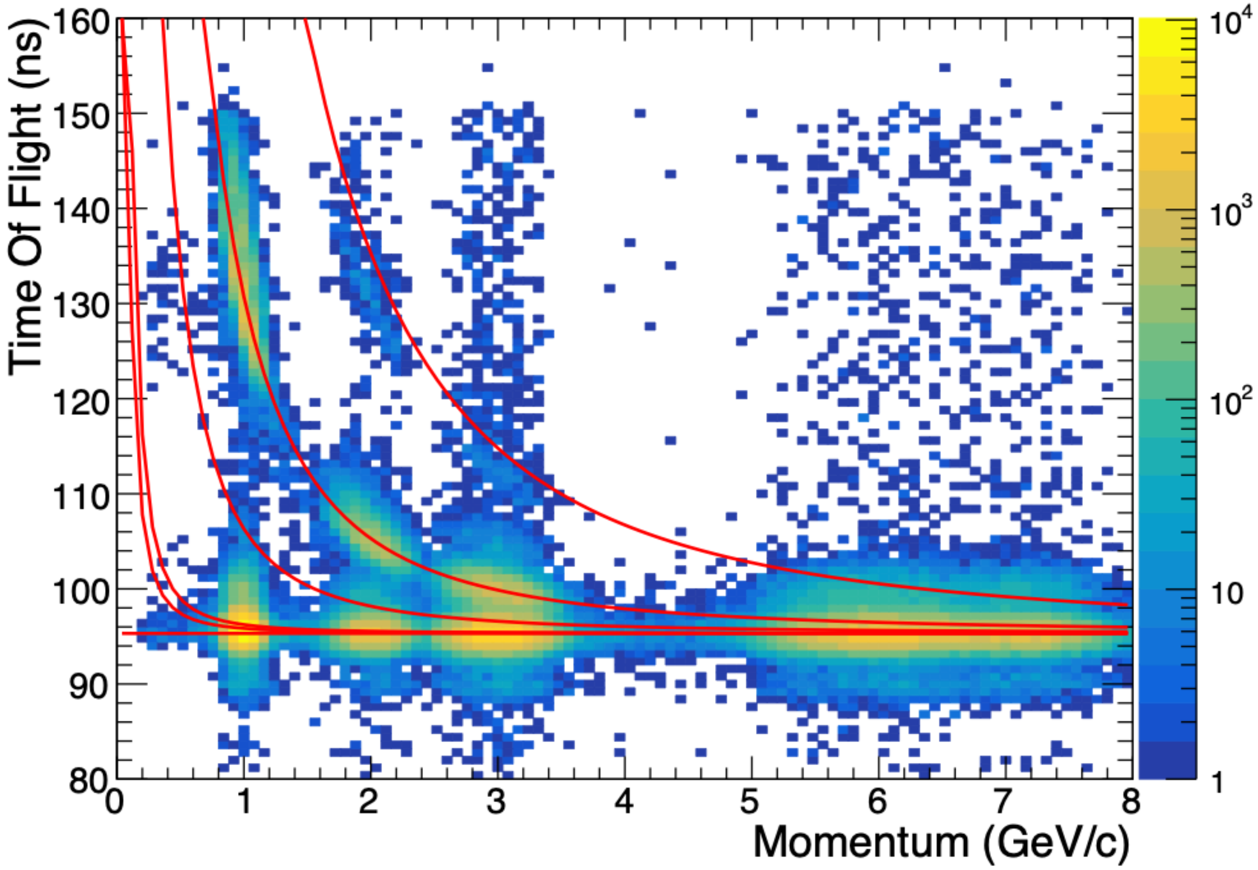
- Field non-uniformities
- Liquid flow
- Impurities





Beam events: Oct-Nov 2018

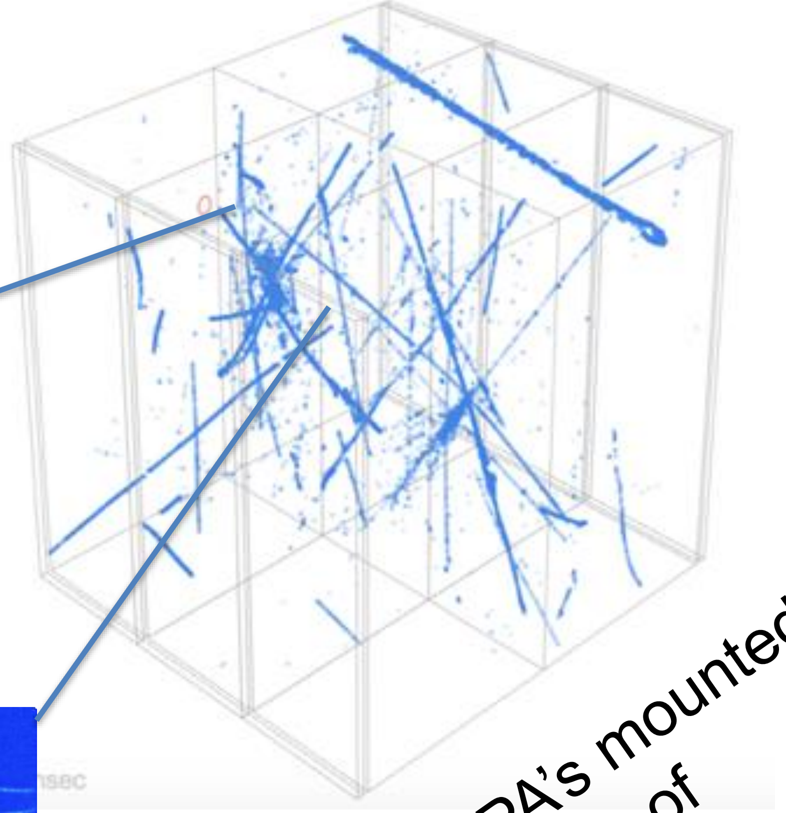
- 8M events taken with beam
- Beam tagged:
  - 300 k pion events each at 1, 2, 3, 6, 7 GeV/c.
- Large statistics proton and electron data. Some high energy kaon data.
- Since then > 10M Cosmic gates (> 40 tracks/event) with varying:
  - Purity
  - HV settings

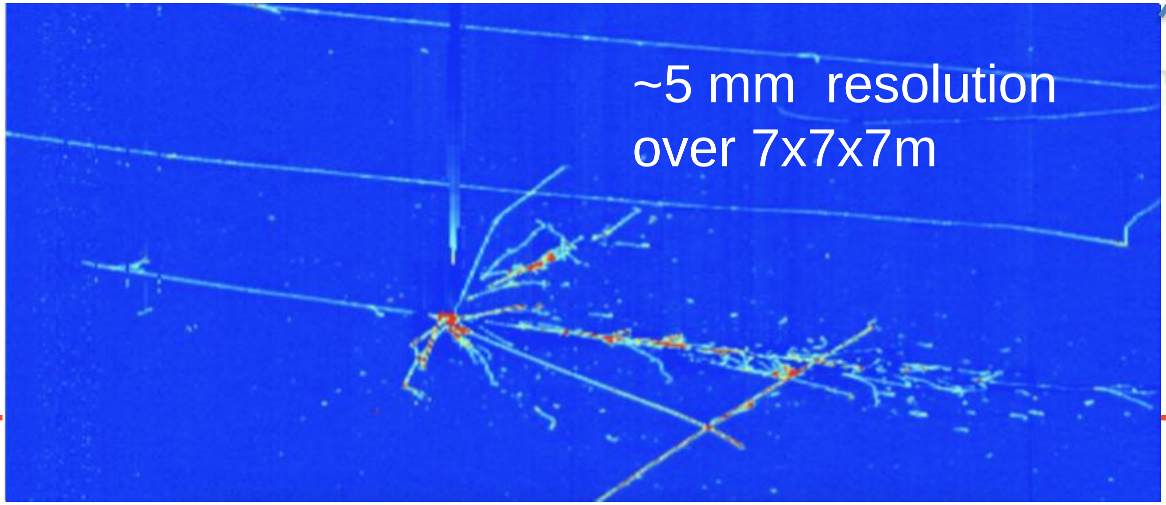


ProtoDUNE-SP Event sizes

protoDUNE raw events are each about 75 MB (compressed), at 10-25Hz

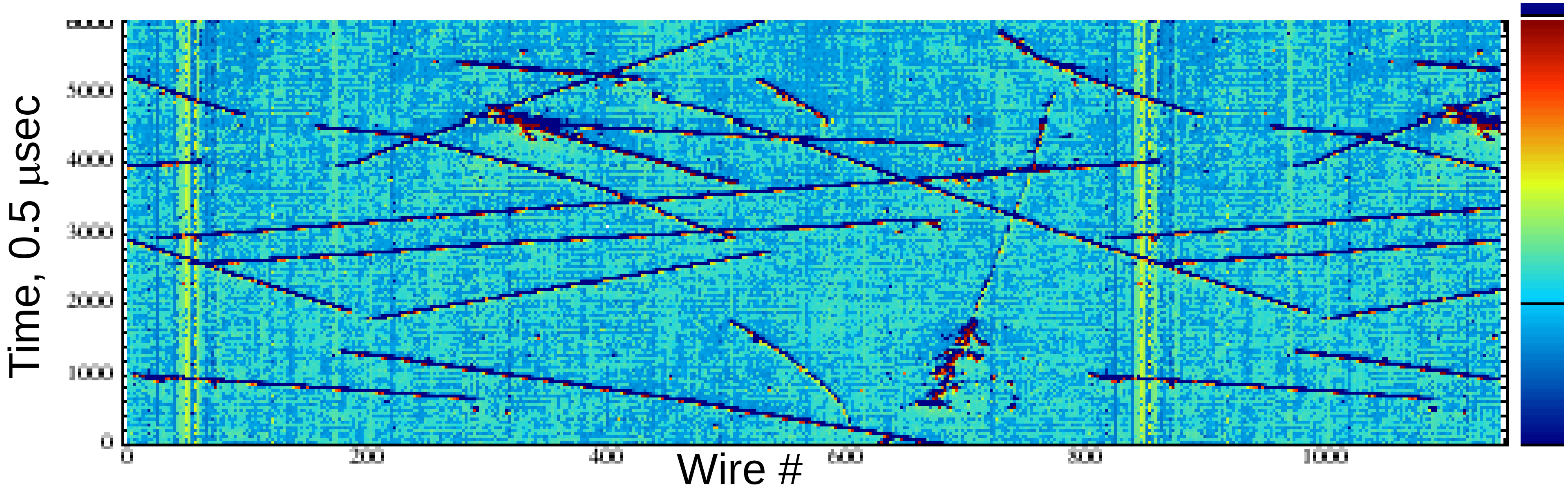
- Compare ~2 MB for ATLAS/CMS p-p
- And ~8 MB for ALICE Pb-Pb

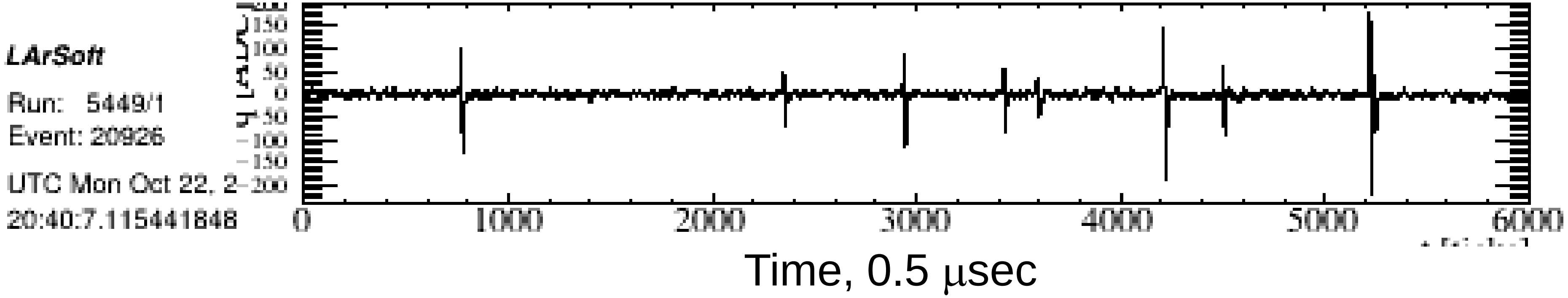




6 APA's mounted at sides of cryostat

Signal processing for 1 APA

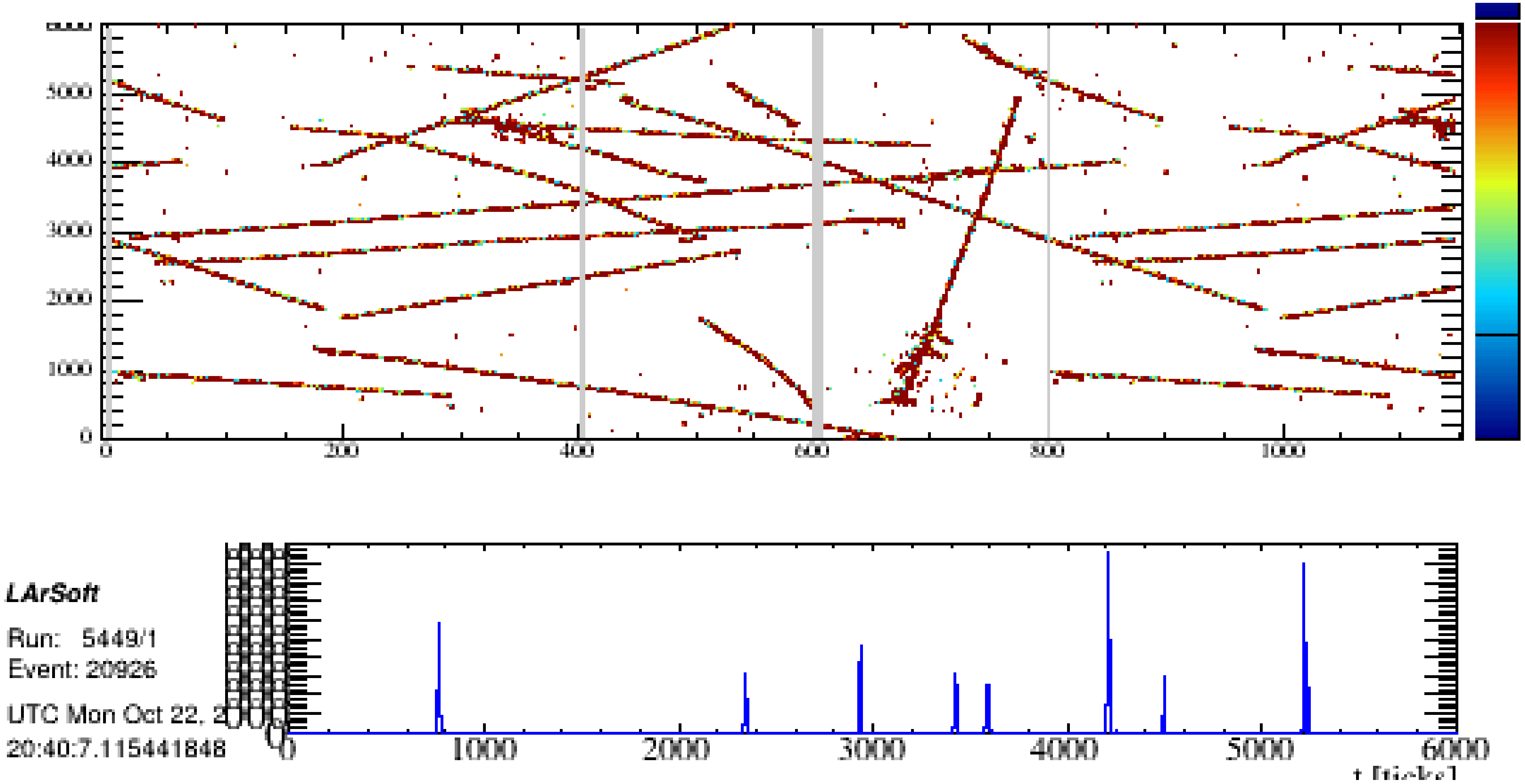




Signal for 1 channel

JINST 13 (2018) no.07, P07006 arXiv:1802.08709

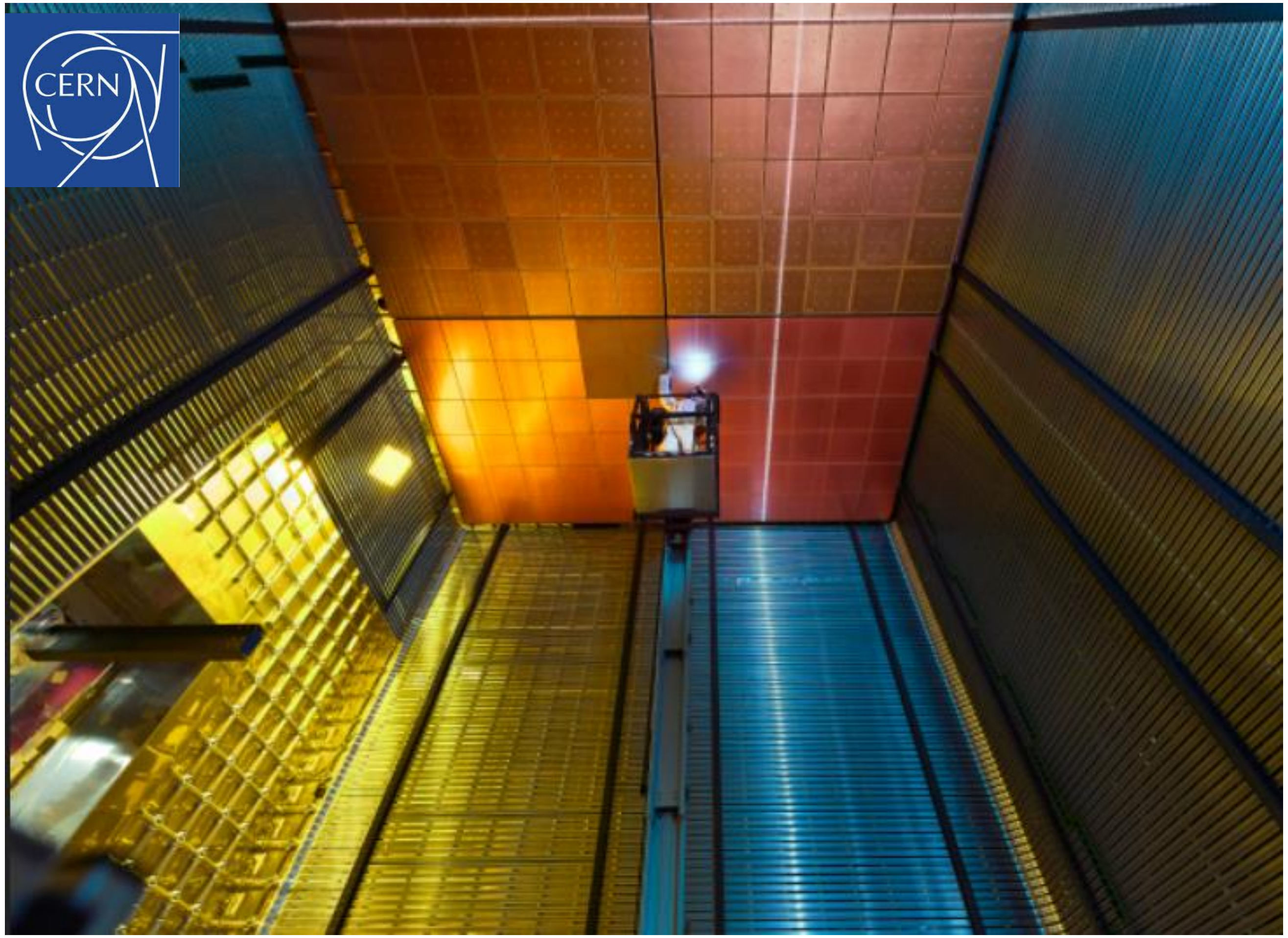
Signal processing for 1 APA

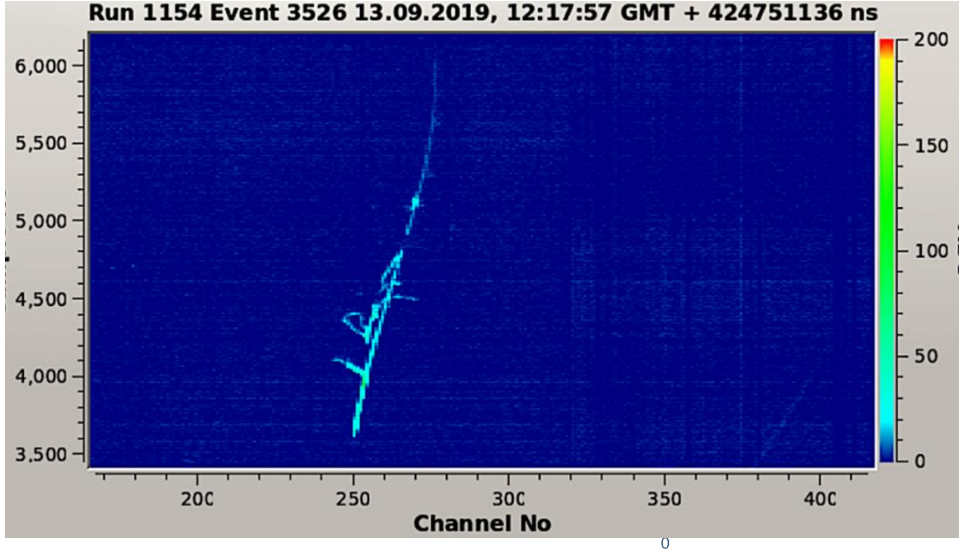


Remove bad hits, coherent noise, deconvolute, 2560x6000 12 bit

ProtoDUNE Dual-Phase

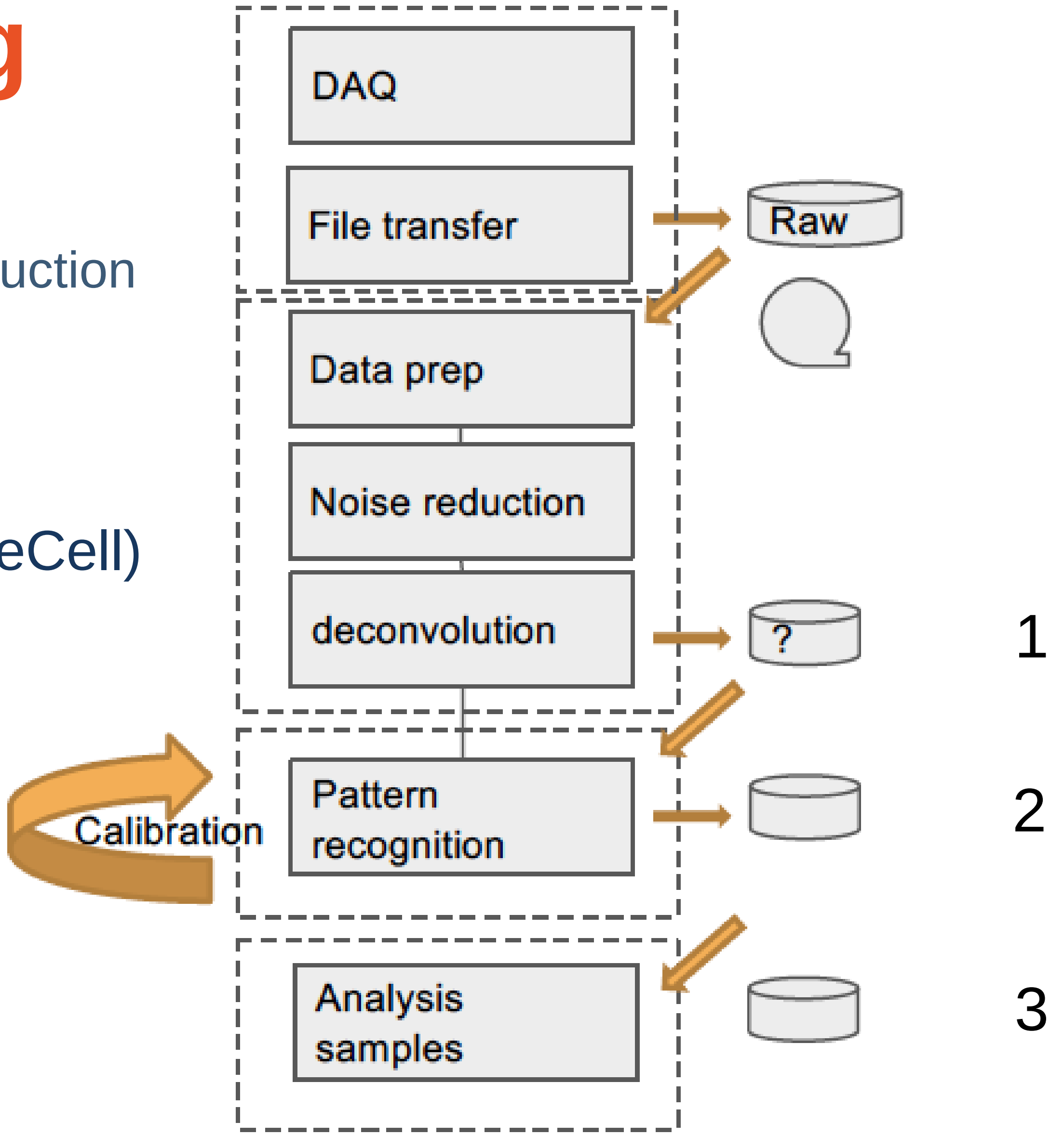
- Gas amplification raises S/N
- Data taking started late Aug 2019
- 157 TB of raw data so far
- 110 MB/event
- First reconstruction pass coming in November





LAr TPC data processing

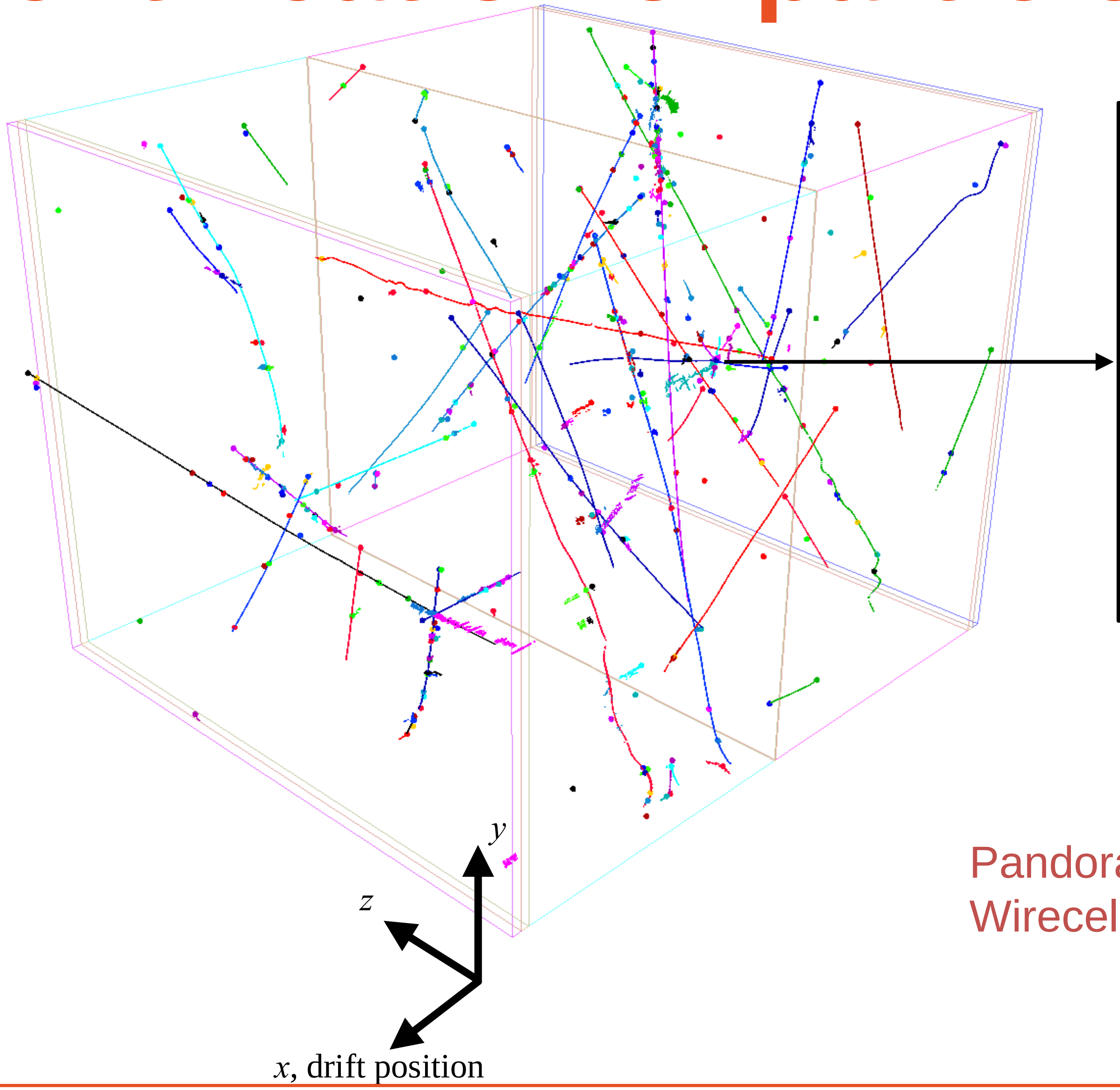
- hit finding and deconvolution
  - **x5 (ProtoDUNE) -100 (Far Detector)** data reduction
  - Takes 30 sec/APA
  - Do it 1-2 times over expt. lifetime
- Pattern recognition (Tensorflow, Pandora, WireCell)
  - Some data expansion
  - Takes ~30-50 sec/APA now
  - Do it ? times over expt.
- Analysis sample creation and use
  - multiple[2] iterations
  - Chaos (users) and/or order (HPC)



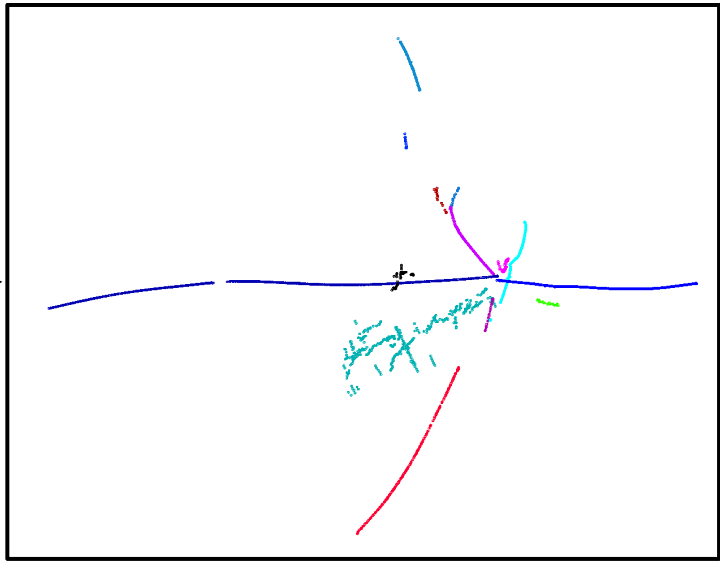
Oregon State University

DUNE

Identification of particles in a beam event

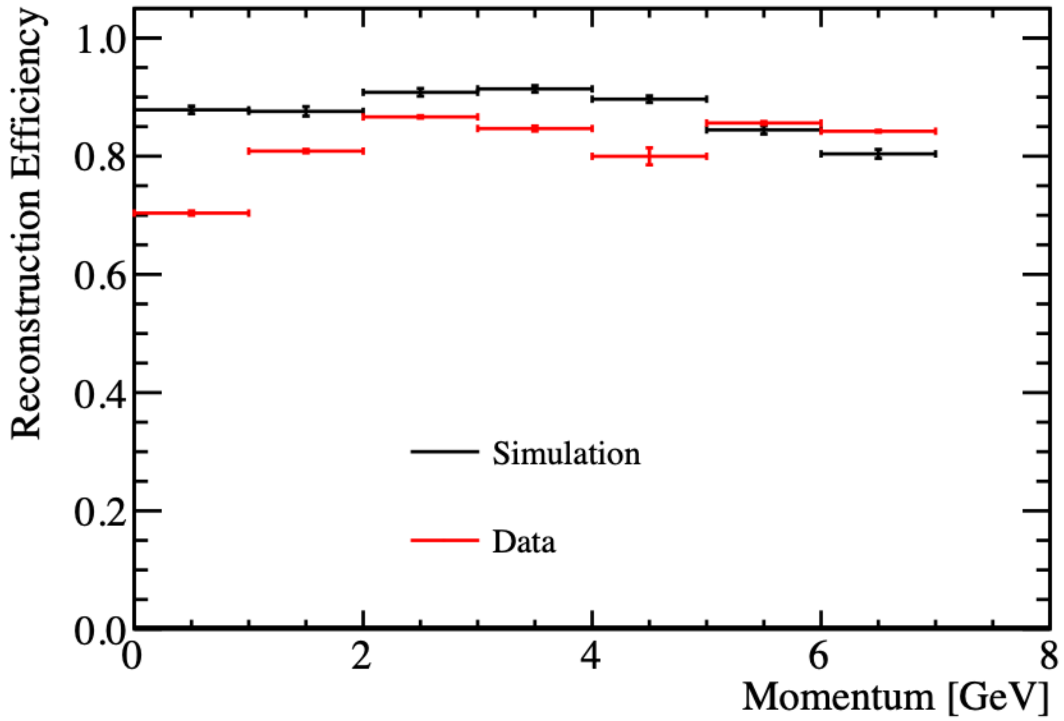


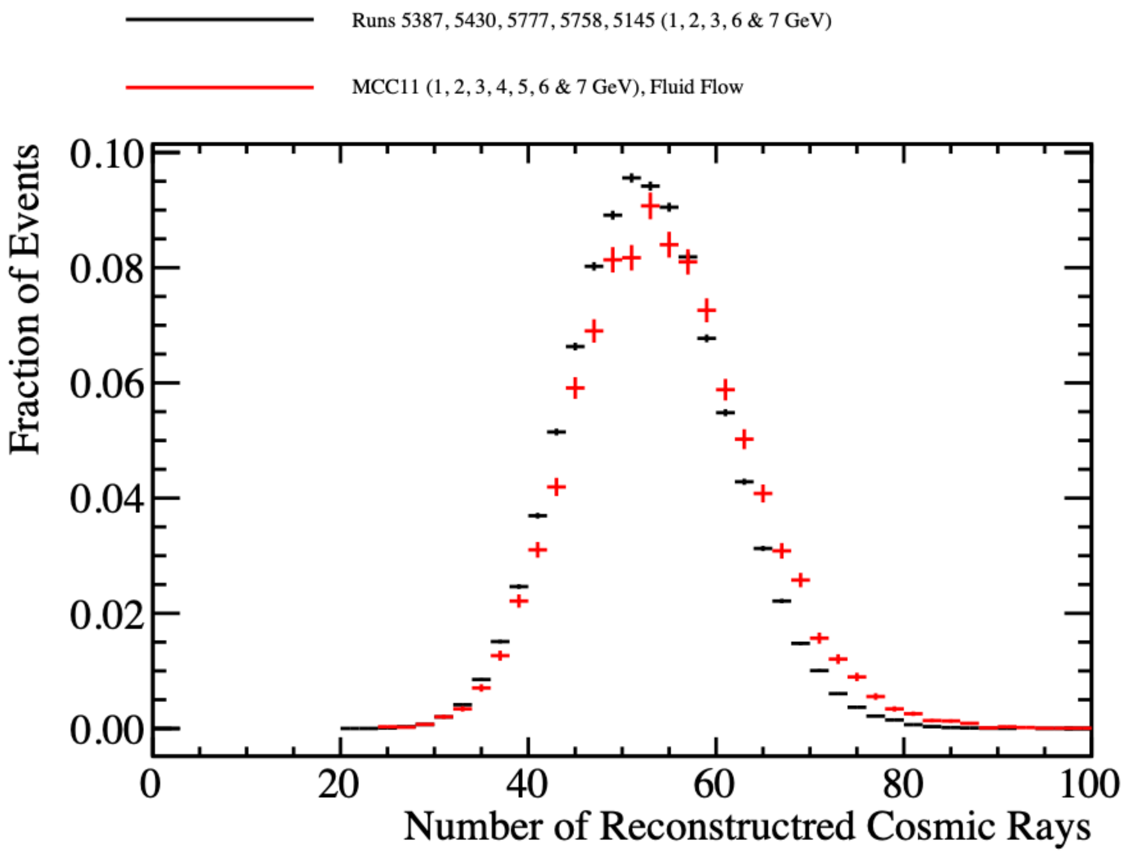
PANDORA



Pandora: Eur.Phys.J. C78 (2018) no.1, 82
Wirecell: J.Phys.Conf.Ser. 762 (2016) no.1, 012033

Reconstruction Quality





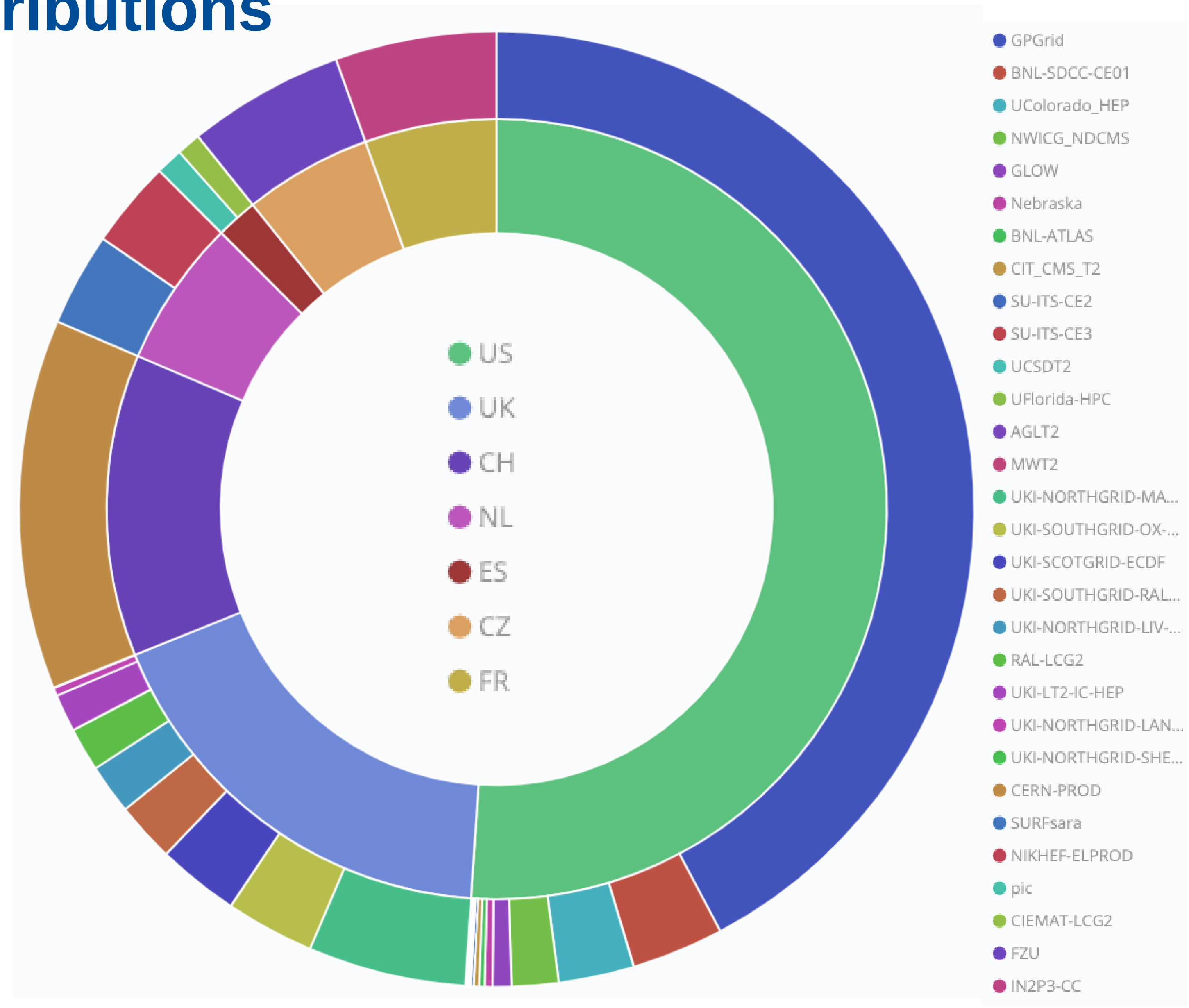
International Contributions

PDUNE-SP data took 6 weeks to collect

Reprocessing passes are generally 4-6 weeks on ~8000 cores

In 2019 so far, **49% of production wall hours are from outside USA**

Actively working to add more sites and countries



Current status

- Processing chain exists and works for protoDUNE-SP
  - Data stored on **tape** at FNAL and CERN, staged to dCache in 100 event 8GB files
  - Use **xrootd** to stream data to jobs
  - Processing a 100 event 8 GB file takes ~**500 sec/event** (80 sec/APA)
    - Signal processing is < 2 GB of memory
    - Pattern recognition is 2-3 GB
  - Copy 2 GB output back as a single transfer.
  - TensorFlow pattern recognition likes to grab extra CPU's (fun discussion)
- Note: ProtoDUNE-SP data **rates** at 25 Hz are equivalent to the 30 PB/year expected for the full DUNE detector. (Just for 6 weeks instead of 10 years)
- ProtoDUNE-DP
  - Data transfer and storage chain operational since August – up to 2GB/s transfer to FNAL/IN2P3
  - Reconstruction about to start

Scaling

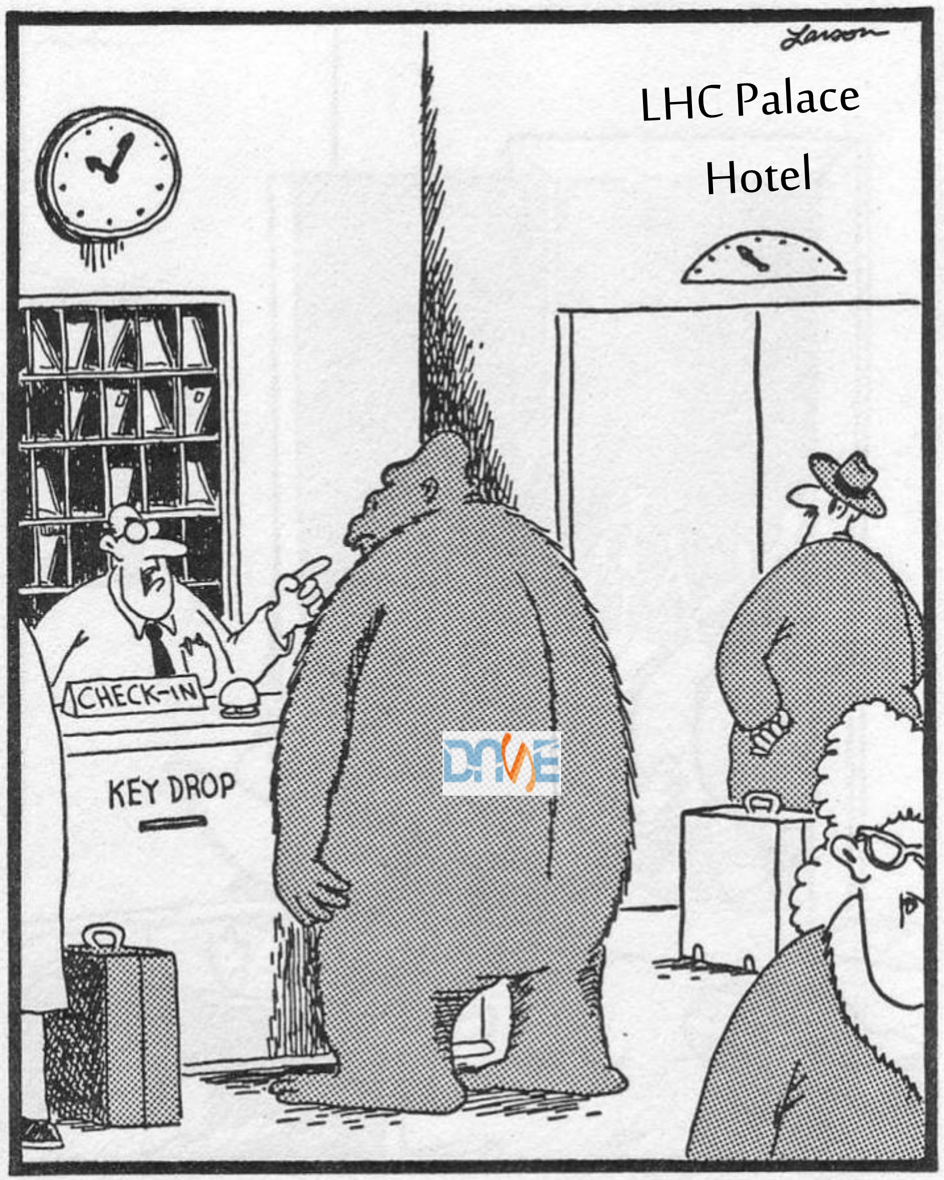
2018: ProtoDUNE event
6 APA ~ 130 MB
At 25 Hz

2025: Beam/cosmic ray event
in 1 FD module -- 150 APA ~ 6GB at < 0.1 Hz

Someday: Supernova
150x4x20,000 5 ms APA
~400 TB. 1/month

Where do we go from here?

- Bottom line:
  - Neutrino experiments are no longer small
  - Up to 30 PB/year of raw data
  - 10-15 years of running
  - 1,200 collaborators
  - Complex codes
  - Precision calibrations
- Solutions:
  - Don’t reinvent the wheel
  - HEP Software foundation
  - Neutrino community – LArSoft, generators
  - LHC tools



“Look. I’m sorry . . . If you weighed 500 pounds, we’d certainly accommodate you — but it’s simply a fact that a *400-pound gorilla* does *not* sleep anywhere he wants to.”

What's the plan?

- Form a Global Consortium
- Collaborate with other neutrino experiments (Larsoft + generators)
  - ArgoNeut
  - Lariat
  - MicroBooNE
  - NOvA
- Collaborate with other experiments on common tools
- Use standard grid tools
  - FNAL **jobsub** talking to WLCG and OSG sites
  - **Cvmfs** for file distribution
  - **http** interfaces for database communication
- In progress
  - **Rucio** for file handling
  - Tested **DIRAC+SAM**
- Future
  - Federated storage
  - Lots of R+D needed for future architectures

Global consortium – still growing

Institution	Country
CBFP	Brazil
Unicamp	Brazil
York Univ.	Canada
CERN	CERN
FZU	Czech Republic
CCIN2P3	France
Indian groups	India
KISTI	Korea
Nikhef	Netherlands
Bern	Switzerland
CIEMAT	Spain
Edinburgh	UK
GridPP	UK
Manchester	UK
Queen Mary Univ.	UK
RAL/STFC	UK

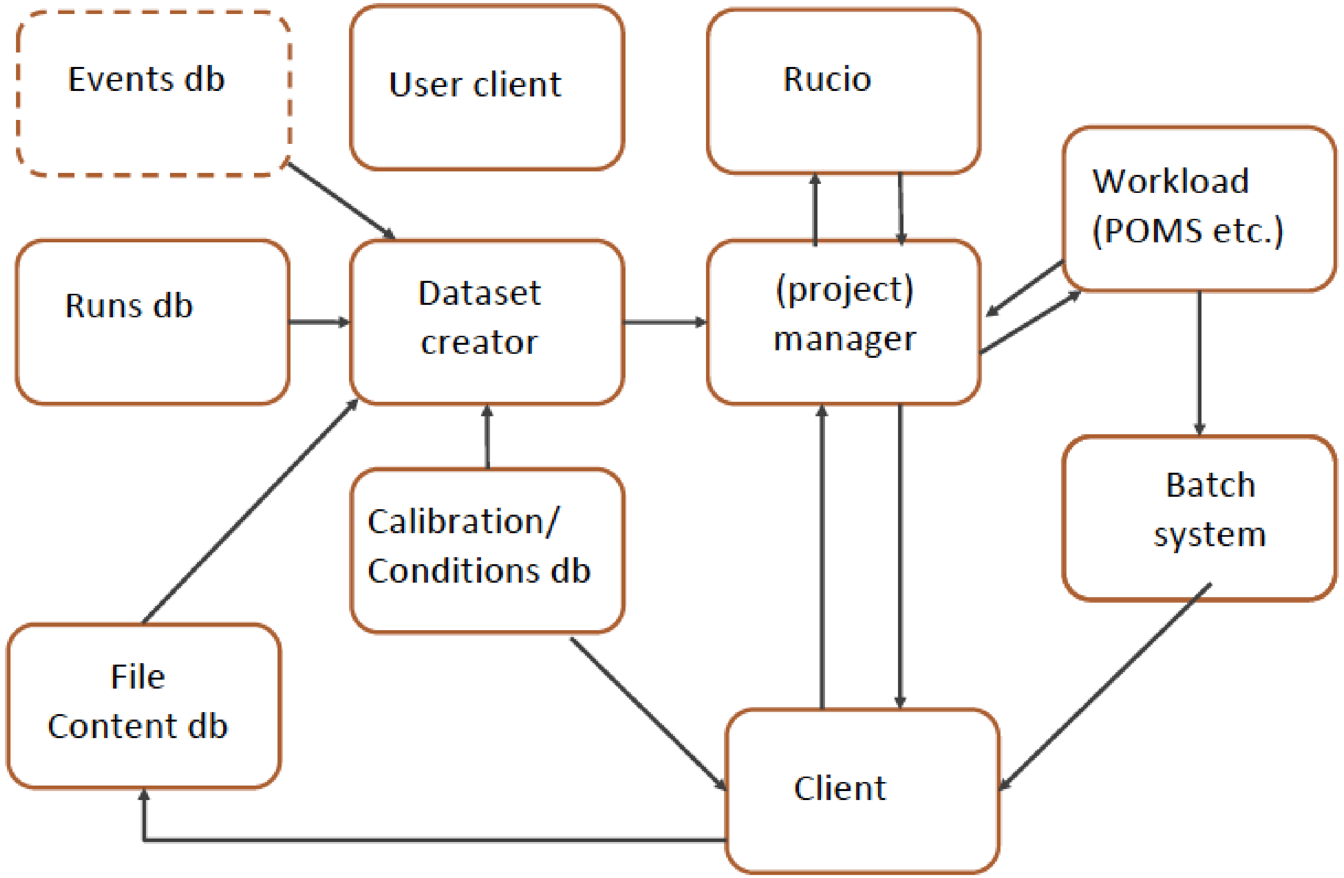
Institution	Country
Argonne	USA
Berkeley	USA
BNL	USA
Colorado State	USA
CU Boulder	USA
Fermilab	USA
Florida	USA
LBNL	USA
Minnesota	USA
Northern Illinois Univ.	USA
Notre Dame	USA
Oregon State University	USA
SLAC	USA
Texas, Austin	USA

Data layout requirements

- APA's = BOXES: Treat data as cells = 1 APA x 5-10 ms = 40-80 MB compressed
  - APA level ensures full information for deconvolution is present
- FILES = CONTAINERS: Beam/cosmic trigger readouts of each FD module deliver up to 150 APAs together – 1-3 GB compressed
  - Process together
- SHIPS: SNB readouts will span multiple (like 10,000) files and take ~10 hrs to transfer at 100Gb/s but only happen ~1/month.
  - Requires special treatment

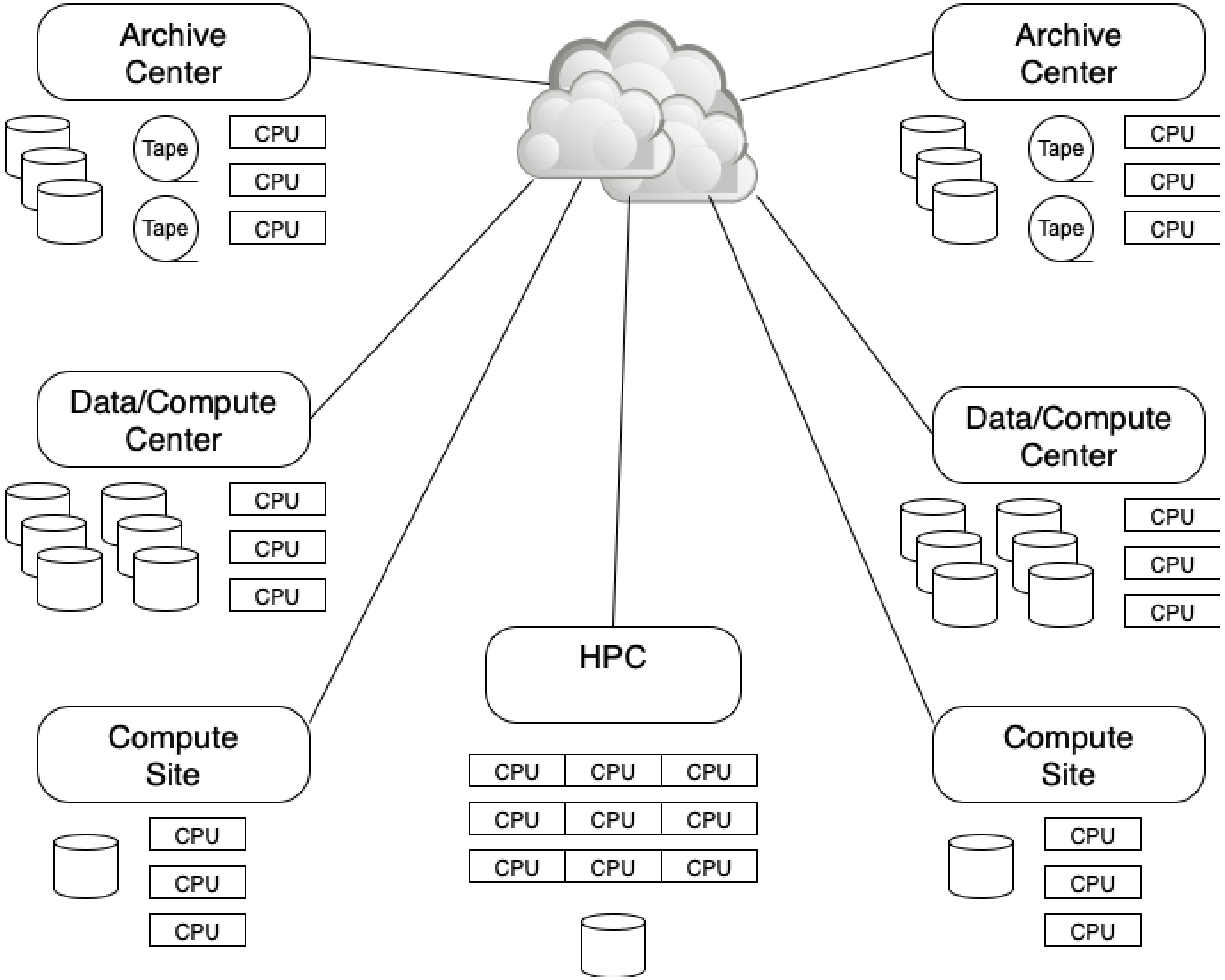
Data tracking

- FNAL neutrino experiments use an updated version of the SAM* file database from D0/CDF
  - Needs a remodel (gut renovation?)
- Develop replacement for SAM components that describe data
  - Beam/detector config
  - Processing provenance
  - Normalization
- Use Rucio for file placement and location
- * SAM first appeared at CHEP 1997



Distributed computing model

- Less "tiered" than current WLCG model → **DOMA**
- Collaborating institutions (or groups of institutions) provide significant disk resources (~1PB chunks)
- **Rucio** places multiple copies of datasets
- **We likely can use common tools:**
  - **But need our own contribution system**
  - **And may have different requirements for dataset definition and tracking**



CPU needs

RECONSTRUCTION

- ProtoDUNE events are more complex than our long term data.
  - ~**500** sec to reconstruct 75 MB compressed – 7 sec/MB
  - For FD, signal processing will dominate at about 3 sec/MB
  - < 30 PB/year of FD data translates to **~100 M CPU-hr/year**
  - That's **~ 12K cores** to keep up with data. But no downtimes to catch up.
- Near detector is unknown but likely smaller.

ANALYSIS (Here be Dragons)

- NOvA/DUNE experience is that data analysis/parameter estimation can be very large
  - ~ 50 MHrs at NERSC for NOvA fits

Unknowns for the future

- $$$
- Near detector:
  - Rate ~ 1 Hz, technology not yet decided.
  - Occupancies will be similar to ProtoDUNE at 1 Hz → O(1) PB/year?
- Processor technologies
  - HPC's
  - Less memory/more memory?
  - GPU's? << signal processing may love these!
- Storage technologies
  - Tape
  - Spinning disk
  - SSD
  - Something else?

We stand on the shoulders of giants

- **Art framework, Larsoft, Pandora and WireCell**
  - NOvA
  - ArgoNeut
  - MicroBooNE
- **Models and simulation**
  - GEANT4 and Fluka
  - GENIE, Neut, GiBUU, NuWro, …
- **Beam models**
  - G4numi -> g4lbnf
  - ppfx
- **Infrastructure**
  - Jobsub/POMS
  - WLCG and OSG
  - Enstore, dCache
  - uCondb and ifbeam
  - SAM catalog
  - Elisa logbook
  - Rucio
  - Authentication systems
- **OSG/WLCG/HSF for new ideas!**

Thank you