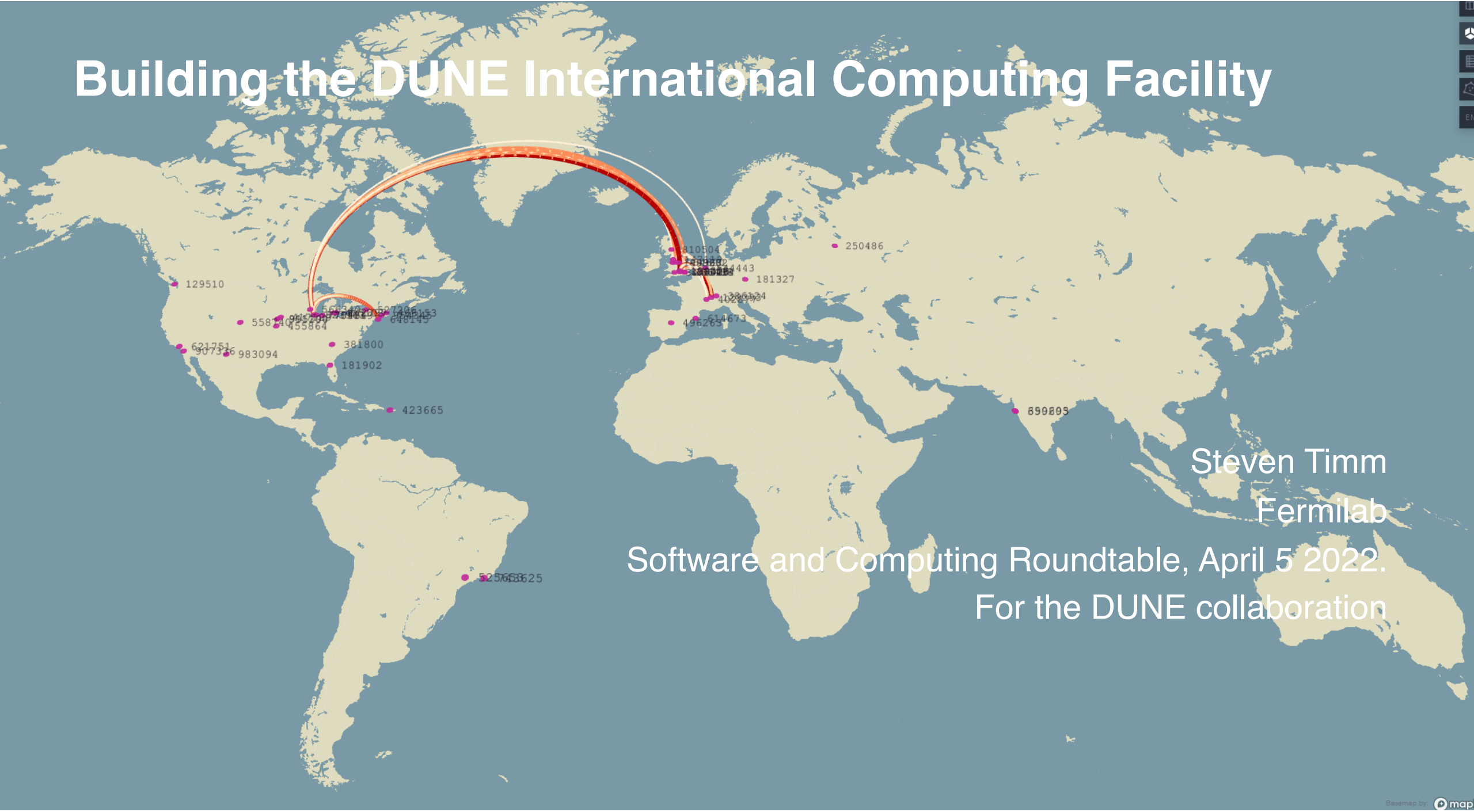
Building the DUNE International Computing Facility



Steven Timm

Fermilab

Software and Computing Roundtable, April 5 2022.

For the DUNE collaboration

DUNE's main purpose is to understand neutrino properties

ν_e

ν_μ

ν_τ

Flavor Basis
(Interactions)

ν_1

ν_2

ν_3

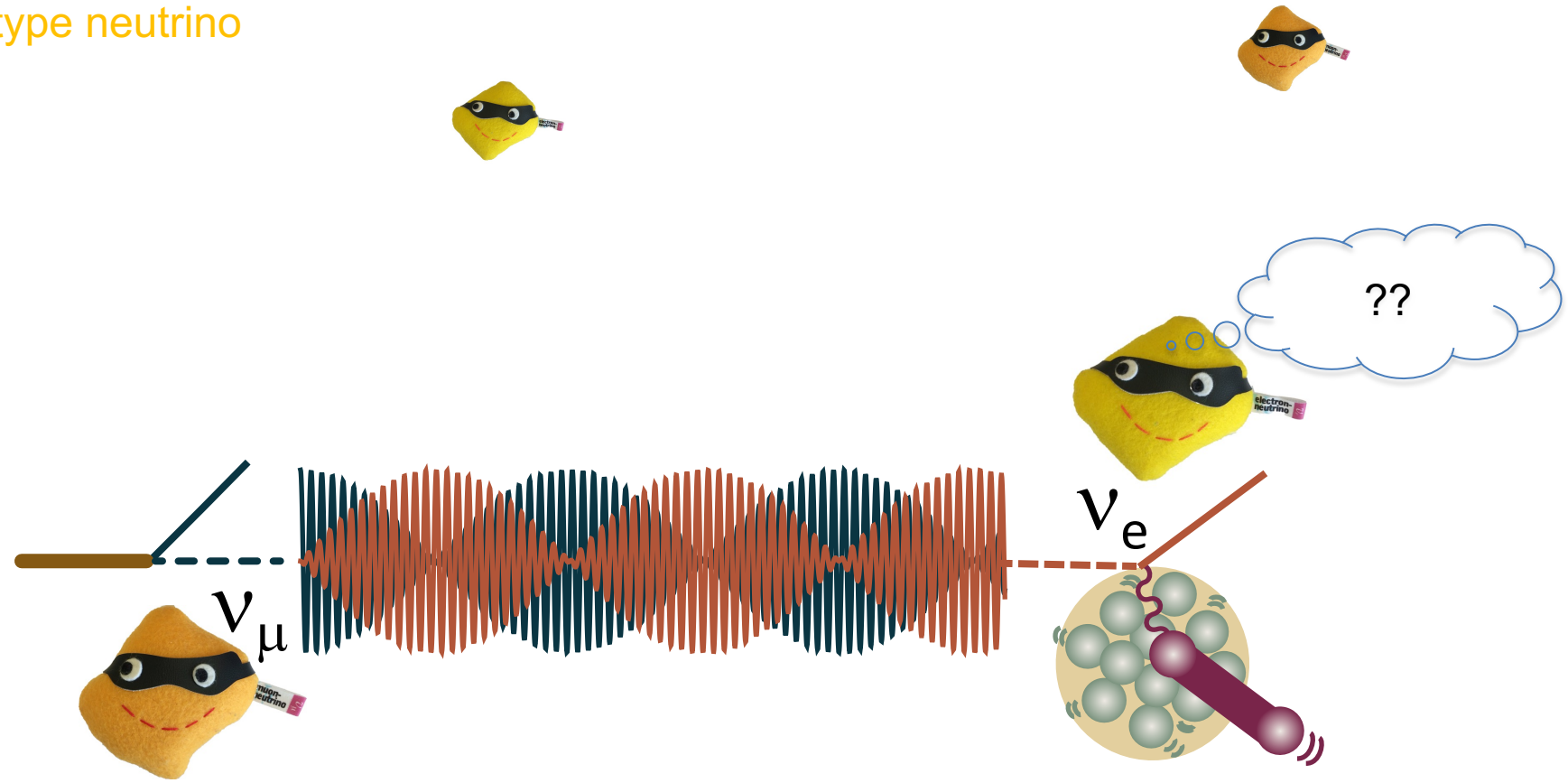
Mass Basis
(Motion)

2 different views of the same neutrinos

Particlezoo.net

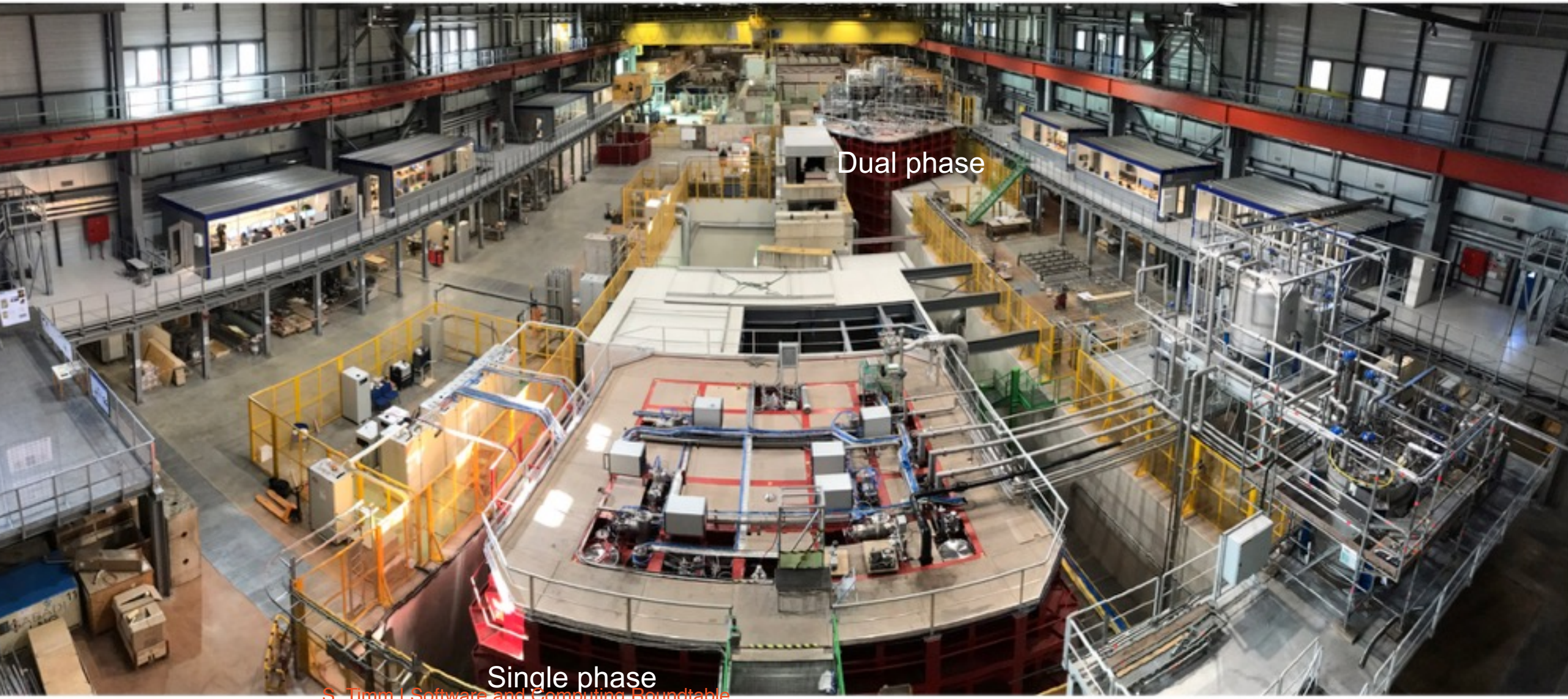
The quantum wavelength of a 2 GeV muon neutrino is ~ 10^{-16} m
But it is actually a superposition of the 3 mass types of neutrinos which have slightly different wavelengths – the beat wavelength between the types is about 2000 km.

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.

~1kt LAr-TPC's at CERN



Build “small” prototypes @CERN

DUNE Computing—Now and Future

- Collaboration formed in 2015
- ~1400 collaborators, 5 continents
- 100 compute elements, 15 storage elements
- 5% scale prototypes at CERN taking beam (plus 1 starting at FNAL this year).
- 18 PB on tape already > (CDF + D0)
- Already some legacy software—analysis preservation already an issue
- Have to make full computing plan now for an experiment that has long time to beam.
- DUNE Experiment will still be running 20 years from now.. Maybe longer.
- Past end of Unix epoch!
- Past useful life of x86_64 architecture
- Can't assume Linux will survive that long
- Can't assume grid computing will survive that long.
- Heterogeneous CPU/GPU/FPGA architecture is part of present and essential to future.
- Will quantum computing be widely available by then?

Similarities with LHC

- Use many of the same underlying software tools and packages
- Have global collaboration with many different ideas on how to do things.
- Participate in organizations like OSG, WLCG, HEP-CCE, IRIS-HEP, DOMA, FIM4R
- Similar data volume in DUNE to what WLCG is already doing
- Have distributed computing and storage model

Scaling

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

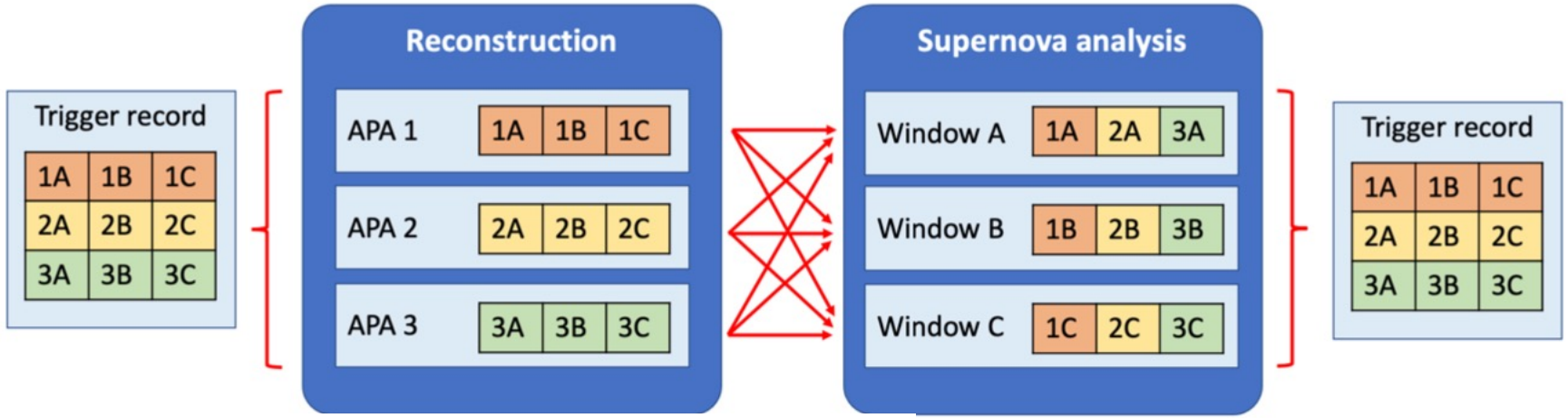
~~2025: Beam/cosmic ray event~~
in 1 FD module -- 150 APA ~ 6GB at < 0.1 Hz

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

DUNE Unique Framework Requirements

Three APAs: (1, 2, 3)
A full module has 150

Three time windows: (A, B, C)
Time windows may overlap and may be of different duration

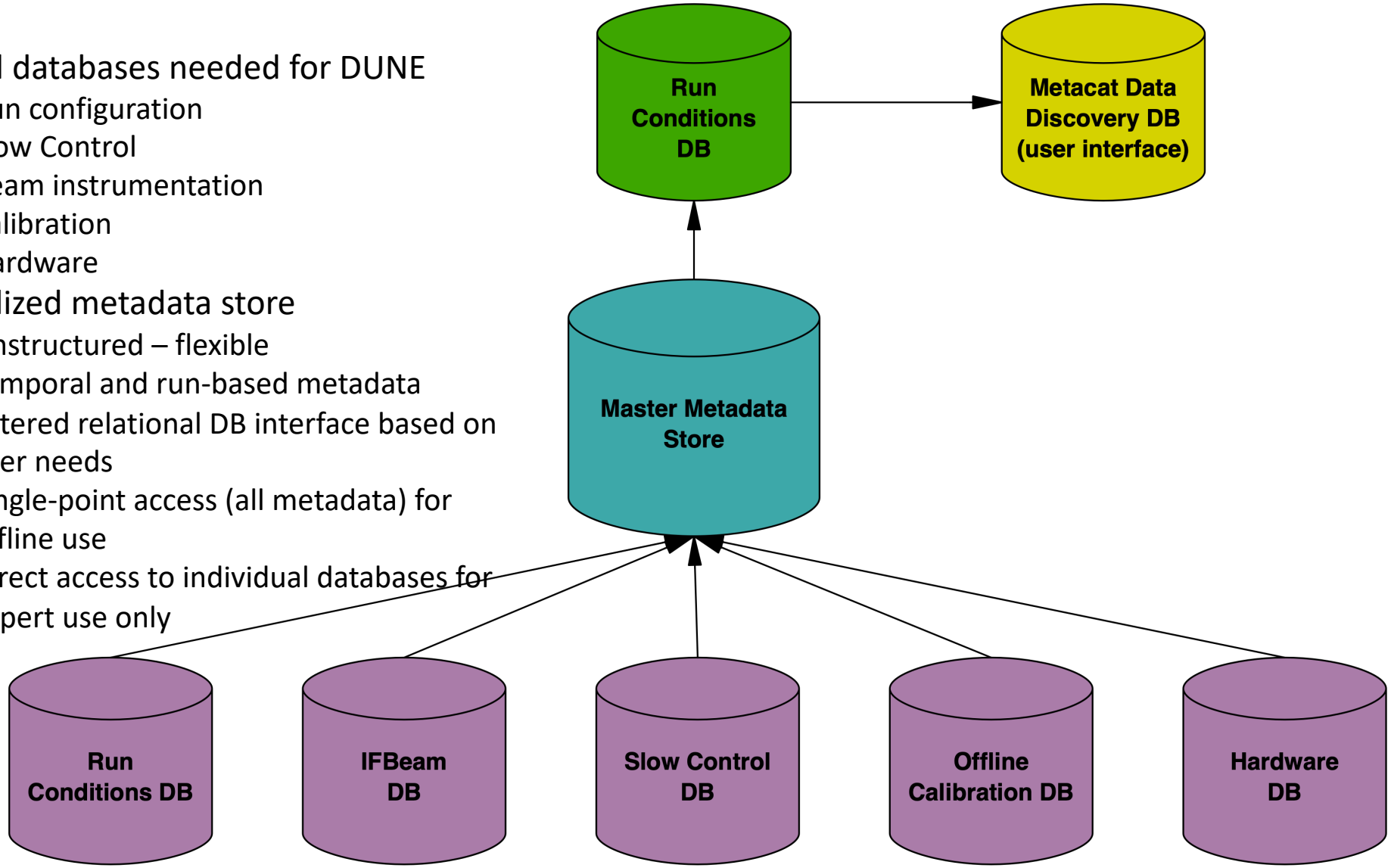


5.3.3 Unique Software Framework Requirements for DUNE

- Far detector partial region simulations and subsetting of temporal and spacial data.
- Far detector partial region reconstruction and subsetting trigger records for data processing
- Temporal and spatial "stitching" of readout trigger records into new extended windows
- Contextual switching of primary data atom types for driving event loops

Database Systems Design

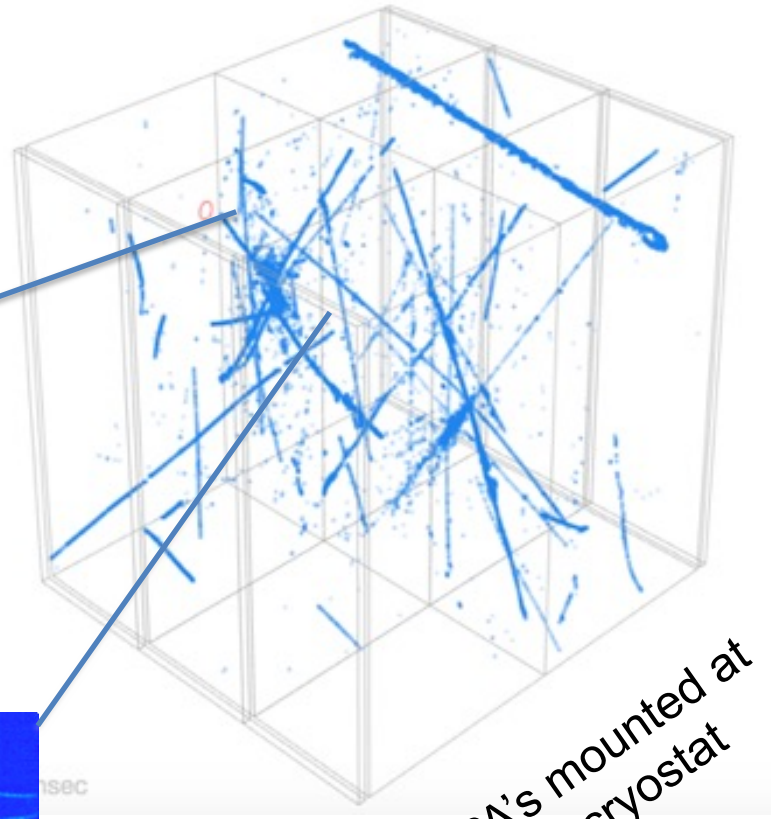
- Several databases needed for DUNE
  - Run configuration
  - Slow Control
  - Beam instrumentation
  - Calibration
  - Hardware
- Centralized metadata store
  - Unstructured – flexible
  - Temporal and run-based metadata
  - Filtered relational DB interface based on user needs
  - Single-point access (all metadata) for offline use
  - Direct access to individual databases for expert use only



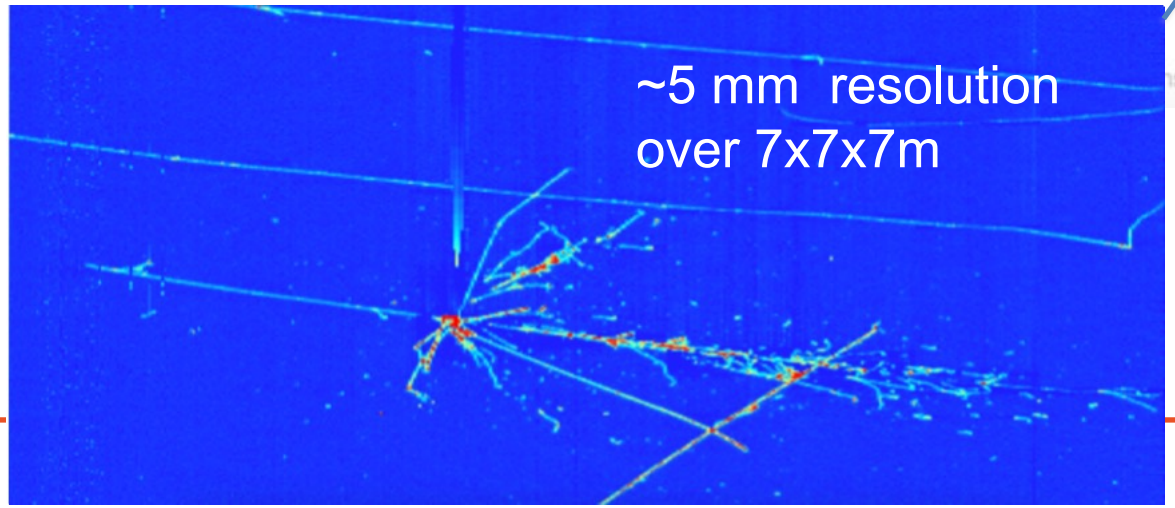
ProtoDUNE-SP Event sizes

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

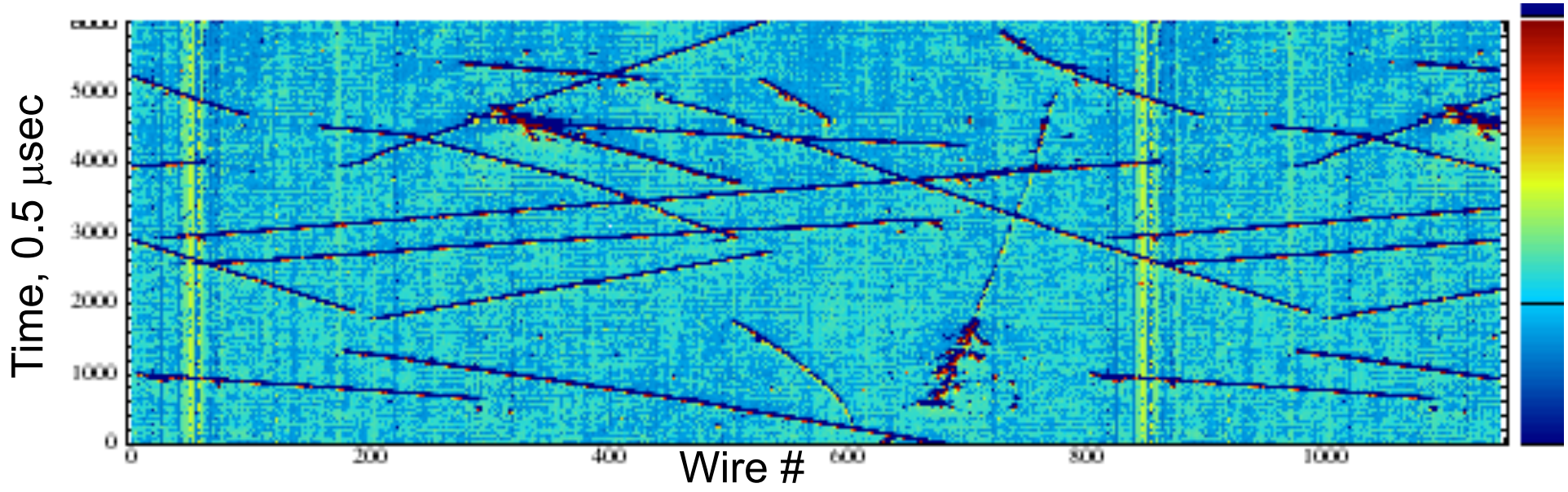
- Compare ~2 MB for ATLAS/CMS p-p
- And ~8 MB for ALICE Pb-Pb

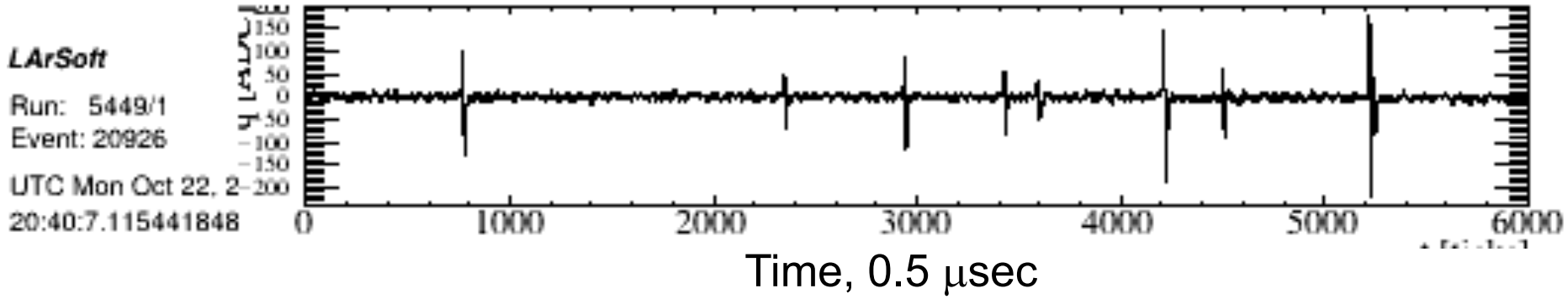


6 APA's mounted at sides of cryostat



Signal processing for 1 APA

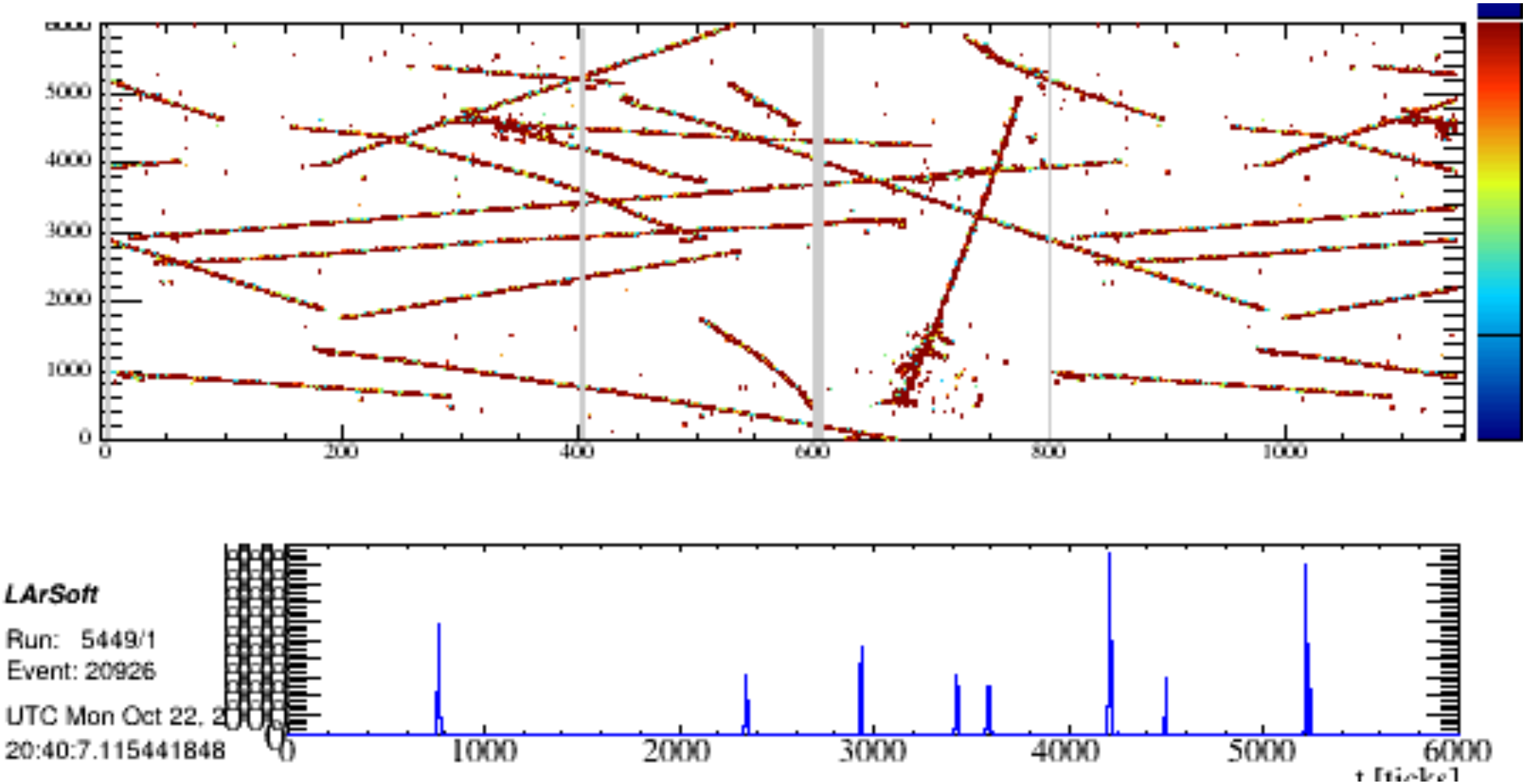




Signal for 1 channel

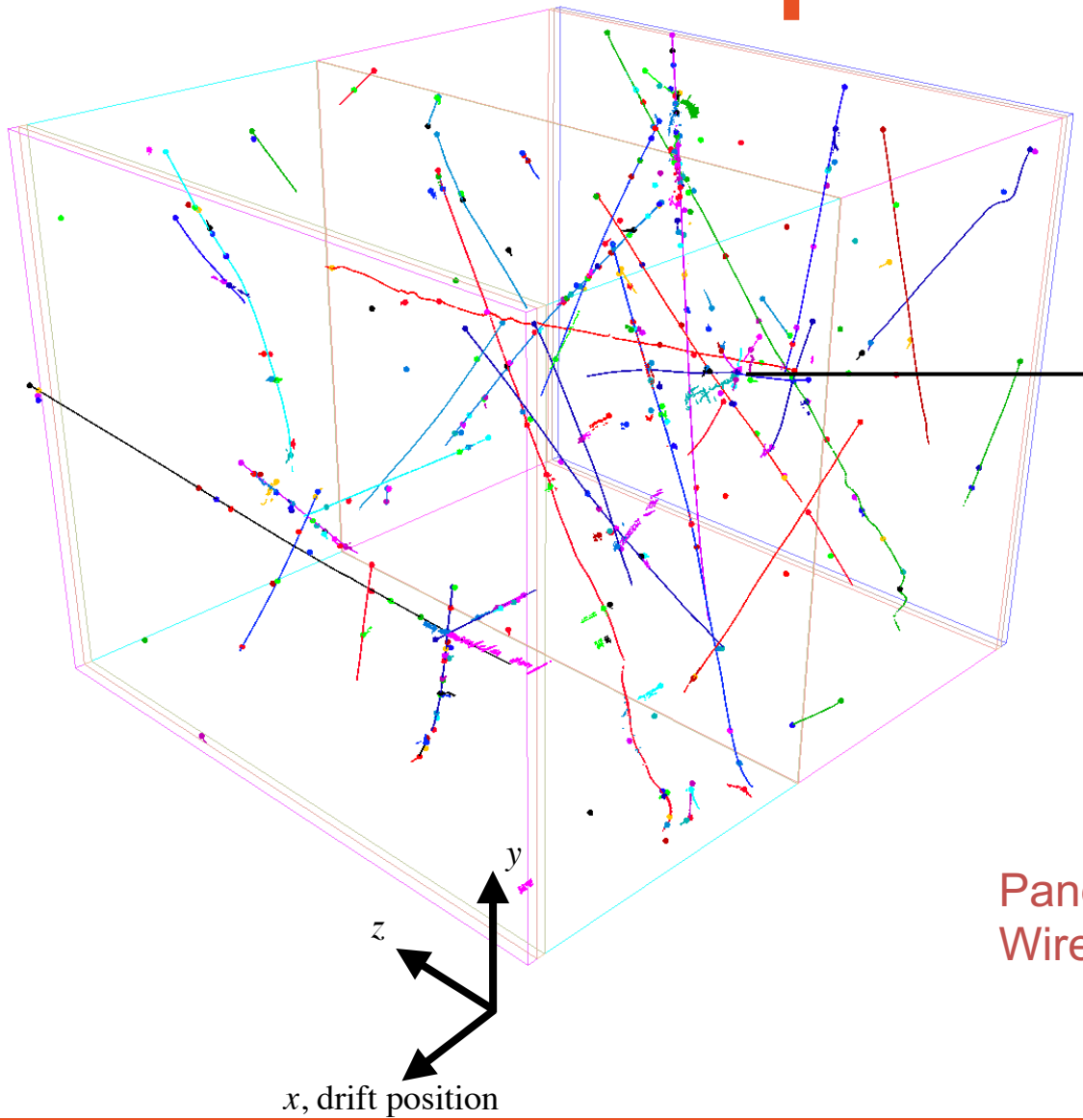
JINST 13 (2018) no.07, P07006 arXiv:1802.08709

Signal processing for 1 APA

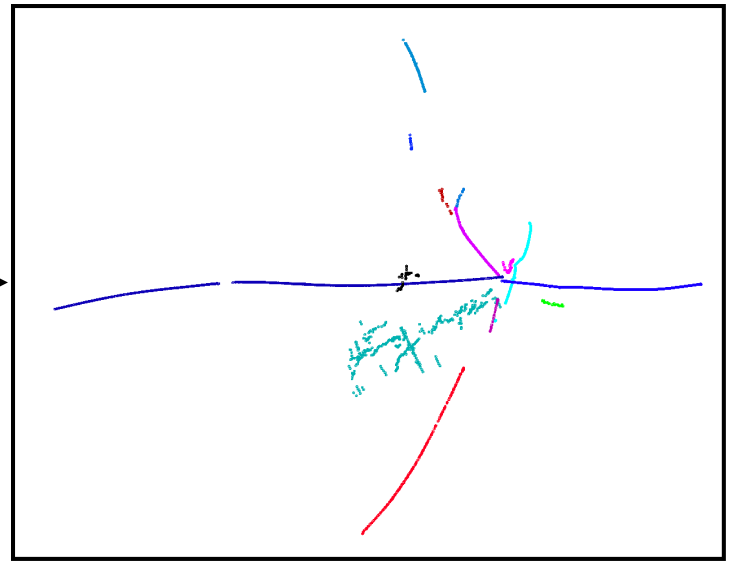


Remove bad hits, coherent noise, deconvolute, 2560x6000 12 bit

Identification of particles in a beam event

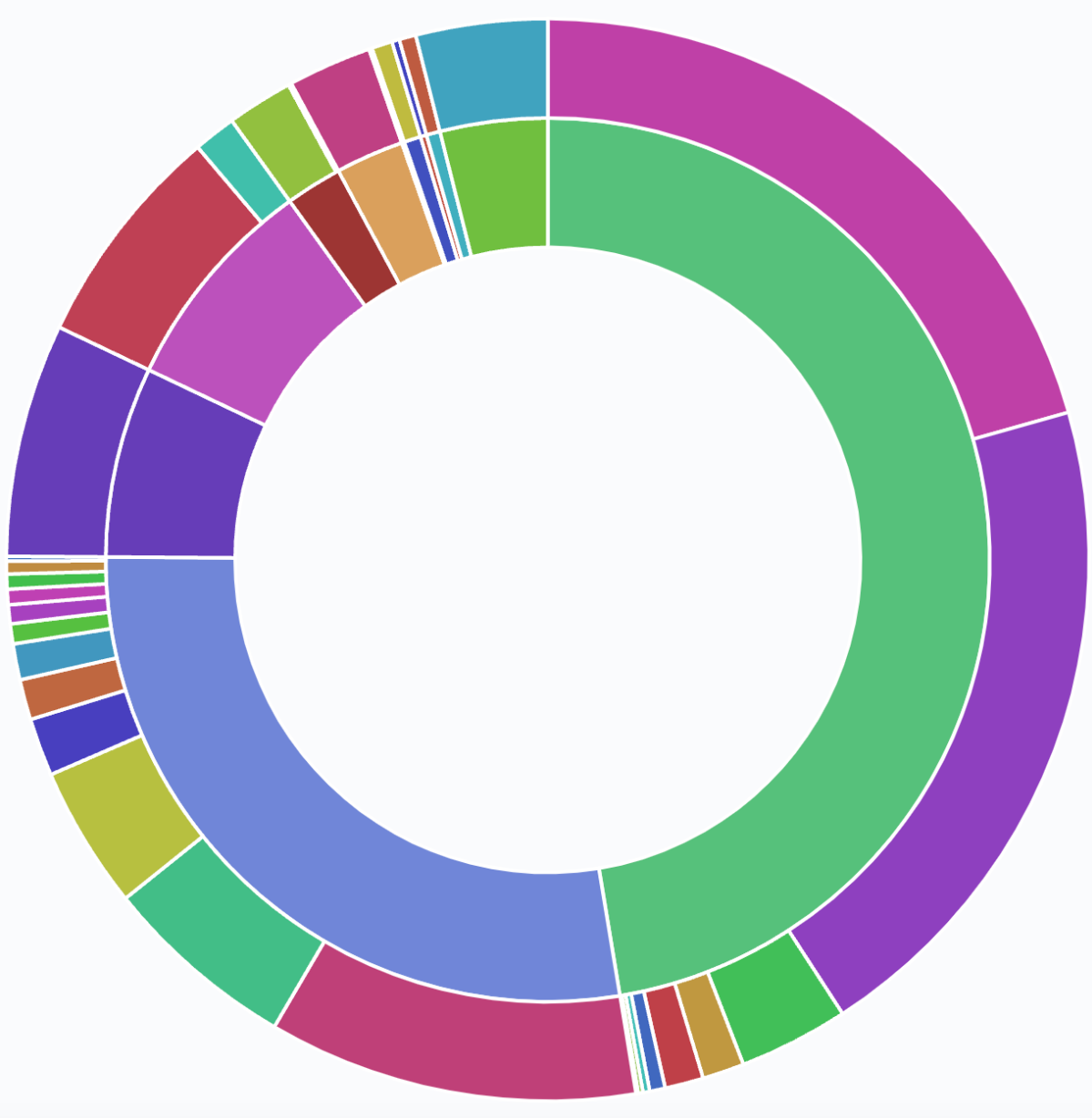


PANDORA



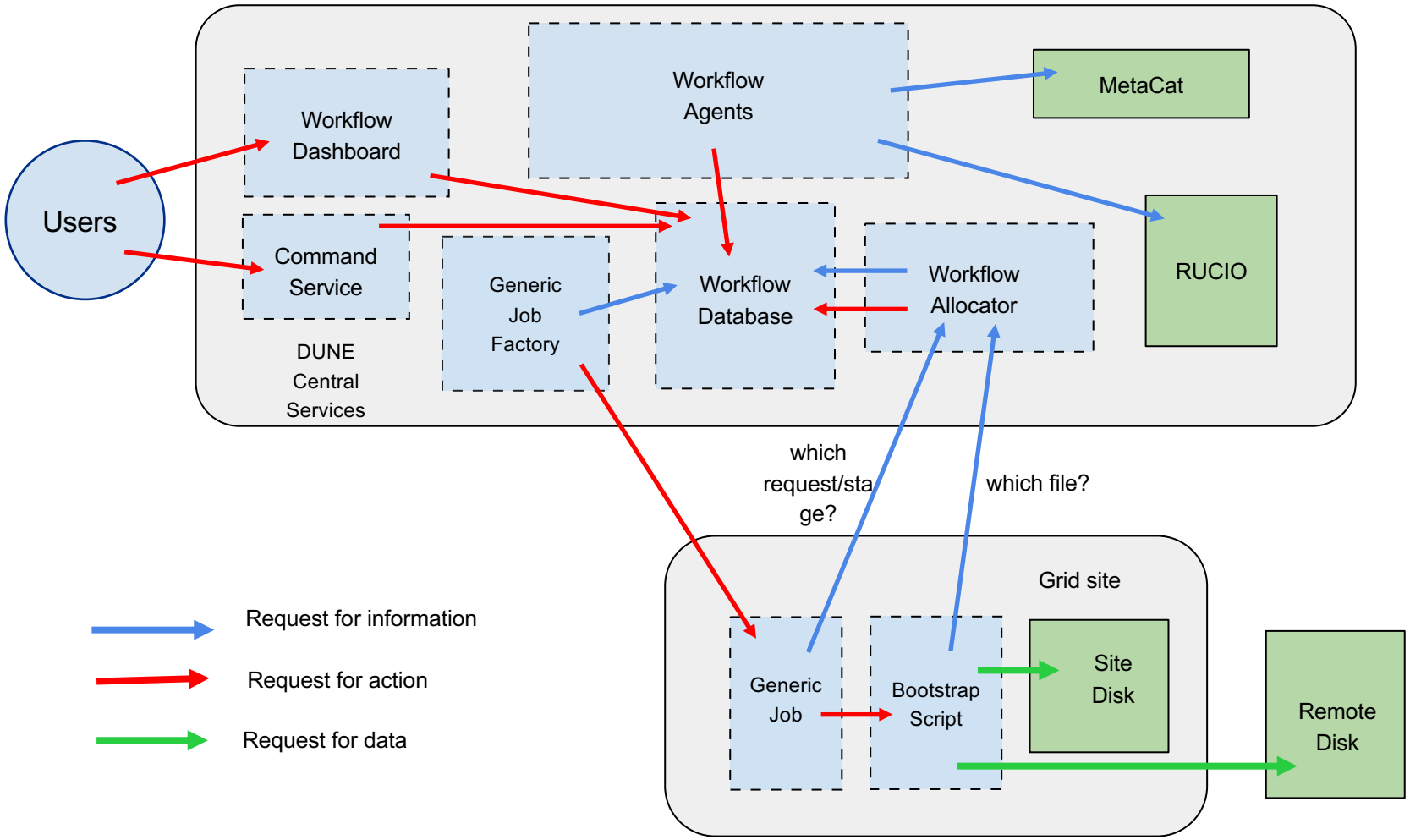
Pandora: Eur.Phys.J. C78 (2018) no.1, 82
Wirecell: J.Phys.Conf.Ser. 762 (2016) no.1, 012033

DUNE Global Computing Resources

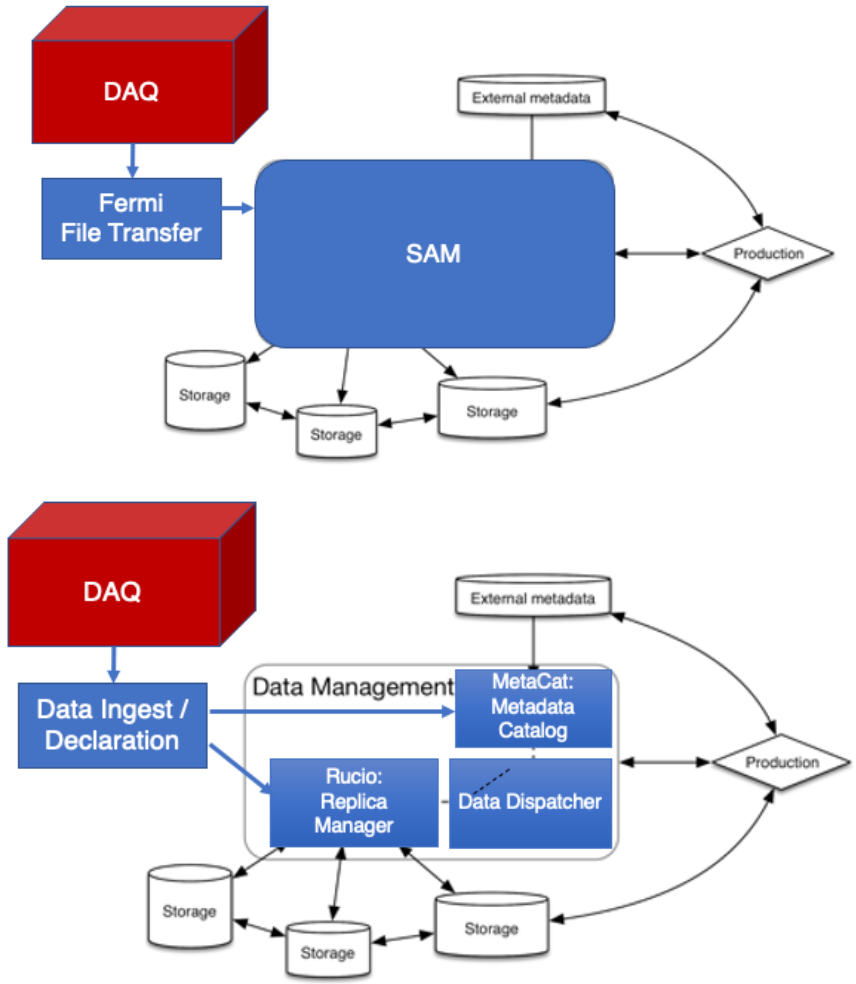


US
UK
CERN
NL
ES
CZ
FR
RU
BR
CA
IN
US_NERSC
US_FermiGrid
US_BNL
US_Colorado
US_Wisconsin
US_WSU
US_UCSD
US_NotreDame
US_PuertoRico
US_MWT2
US_SU-ITS
UK_RAL-Tier1
UK_RAL-PPD
UK_Bristol
UK_Oxford
UK_QMUL
UK_Brunel
UK_Lancaster
UK_Sheffield
UK_Liverpool
UK_Imperial
UK_Edinburgh
UK_Manchester
NL_SURFsara
NL_NIKHEF
ES_PIC
ES_CIEMAT
CZ_FZU
FR_CCIN2P3
RU_JINR
BR_CBPF
CA_Victoria
IN_TIFR

Workflow System



New Data Management System to replace SAM



- SAM=Sequential Access With Metadata
- Developed for D0, used by most Fermilab experiments since then.
- Unified system to locate data, keep track of metadata, and keep track of projects.
- Need a replica management system that's better equipped for a global distributed data system. ***Rucio***
- Also need a hierarchical metadata system for more flexibility. ***MetaCat***
- And need to preserve the project/URL delivery and staging features we have in SAM now. ***Data Dispatcher***

New Components: MetaCat

- New Metadata Catalog
- Written by Igor Mandrichenko @ FNAL
- Code base stable for more than a year, beginning integration testing at scale now.
- Provides all current features of SAM metadata, plus
  - Options for user-defined scopes
  - Options for reserved fields
  - Options for hierarchical metadata attached to a data set and not to each individual file.
  - Option to query files based on MetaCat plus external conditions databases

New Components: Data Dispatcher

- Data Dispatcher
  - Written by Igor Mandrichenko
  - In Beta testing now
- Provides server to keep track of files in a project
  - And when they are available for use
- Client calls server to get next file to process,
  - Server returns URL of next file which can be streamed or copied to the worker node.
- Client notifies server when file is either complete or has failed

Challenges Beyond High-Throughput Computing

- Semi-infinite appetite for GPU and machine learning
  - Several stages of DUNE software good fit for GPU but where do we get them from.
- High appetite for container/Jupyter based analysis
  - Column-based store analysis very helpful to some DUNE analyzers already but how do you keep it reproducible/loggable/etc.
- How to handle the extreme I/O intensive merges.
  - Disk is getting bigger but not faster. Even SSD's won't deliver IOPS we need.
  - Tape robots already a bottleneck
- Heterogeneous platform execution and validation
  - Field has gotten soft from 20 years of RedHat Linux and x86. *Pax Redhat* is over.
  - Need ways to execute and validate the workflow on number of platforms.
- Data and Execution Environment preservation—already have published papers.
- How to make all this friendly to mobile environments

Summary

- DUNE Computing has unique challenges to reconstruct, simulate, and analyze our detector
- We have built a global computing consortium and are making good progress with addressing many of those problems
- We are grateful not only to the members of our consortium but to many others from the various laboratories who have worked on solving those problems to date.
- Also thanks to all those in the various scientific software communities on whose work we are building and on whose shoulders we stand.