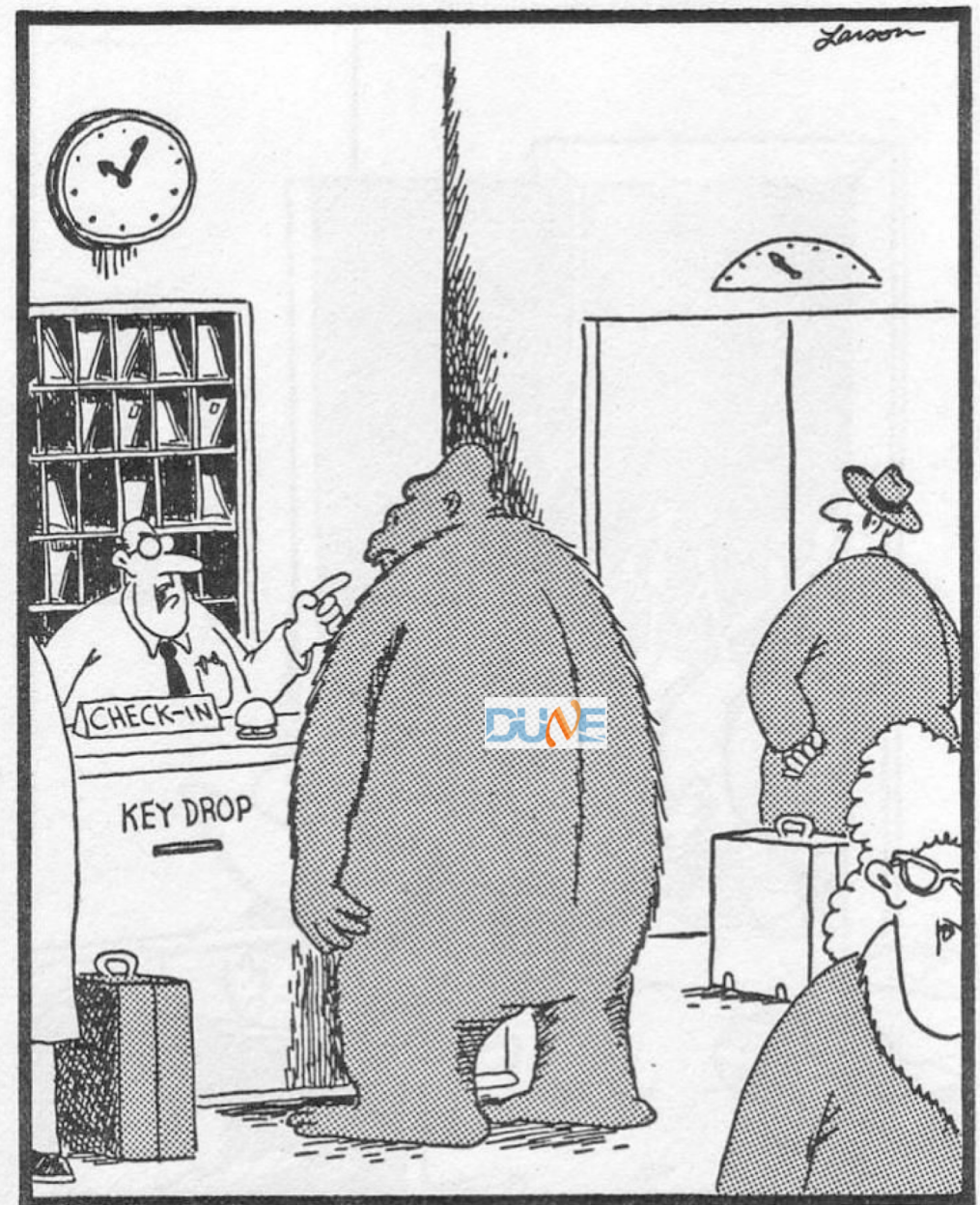# DUNE DATA OVERVIEW

Heidi Schellman, Oregon State University

9/24/19

DUNE is (semi)big

- We'd like to compare to WLCG - LHC computing structure

  - Common infrastructure

  - Technical decision process

  - Allocation of resources

  - Central operation teams funded by collaboration resources (security/networks …)
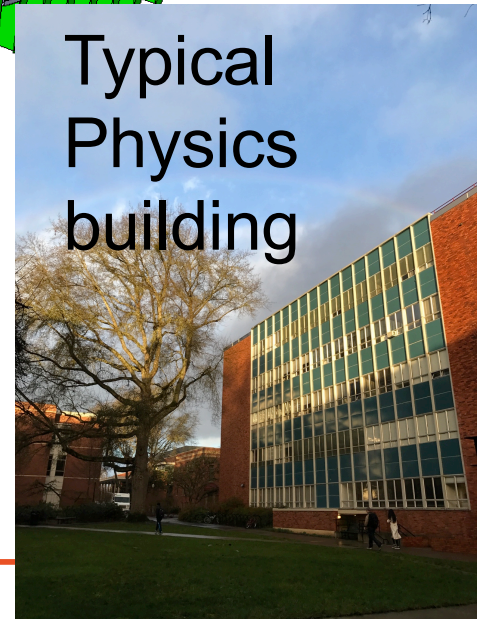
- Works well but the LHC is the 500 pound gorilla!

"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."

# 4 of these – filled with liquid Argon

Typical Physics building
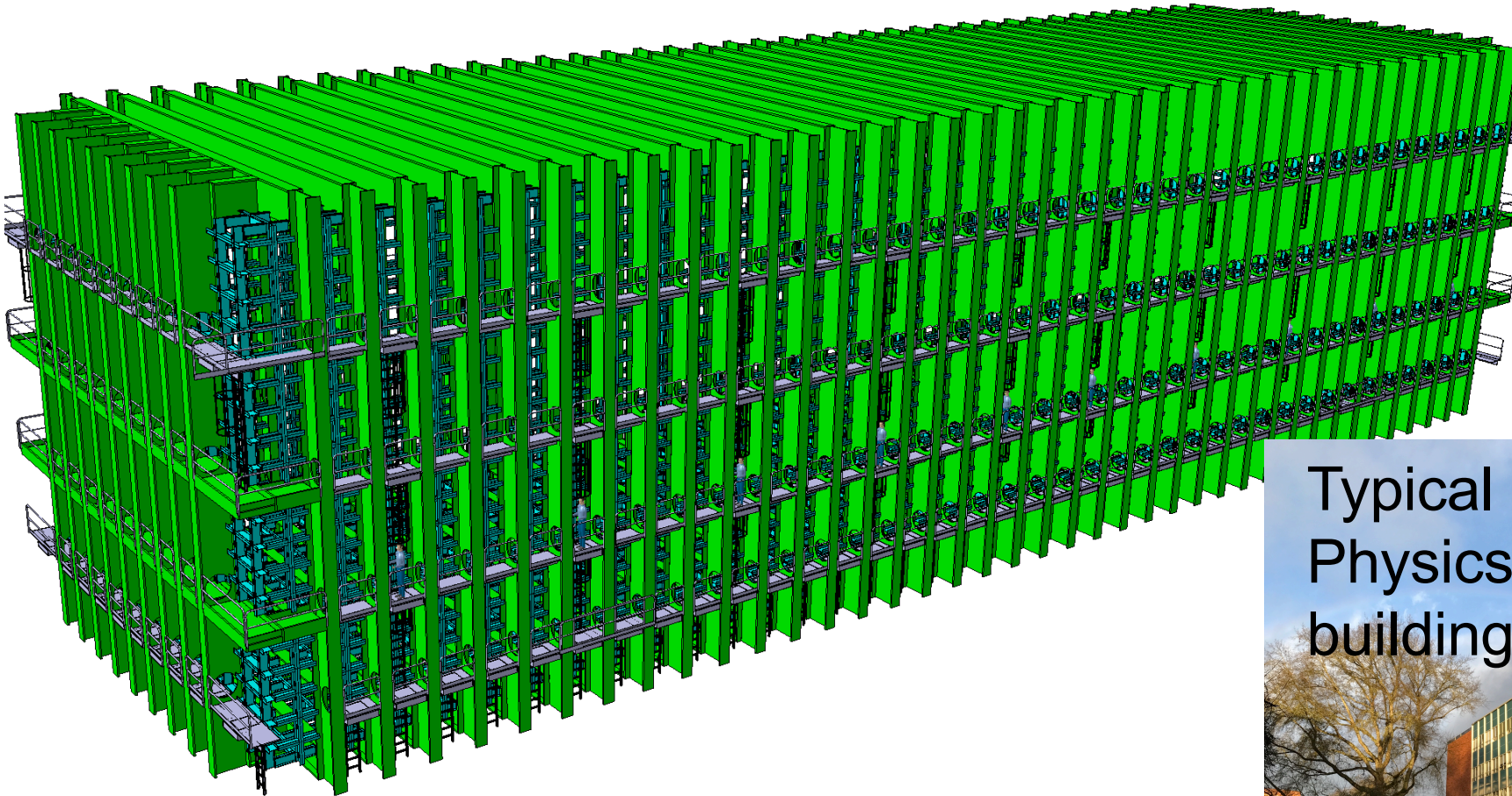
External Dimensions: 19.1m x 18.0m x 66.0m

University

ProtoDUNE is running @CERN

- Two walls of the cryostat are covered by **6 Anode Plane Assemblies (APAs)** which have 3 planes of wires with 0.5 cm spacing. Total of **6x2560 =15,360** wires

- The electrons take ~ 3msec to drift across and you need to detect and time them.

- Each wire is read out by 12-bit ADC's every 0.5 microsecond for 3 msec. Total of **6,000** samples/wire/readout.

- Around **30 MB/readout/APA uncompressed with overheads**

- **12 MB compressed.**

- ProtoDUNE has run since Oct 2018 and written ~ 6 PB of data+ reconstructed output
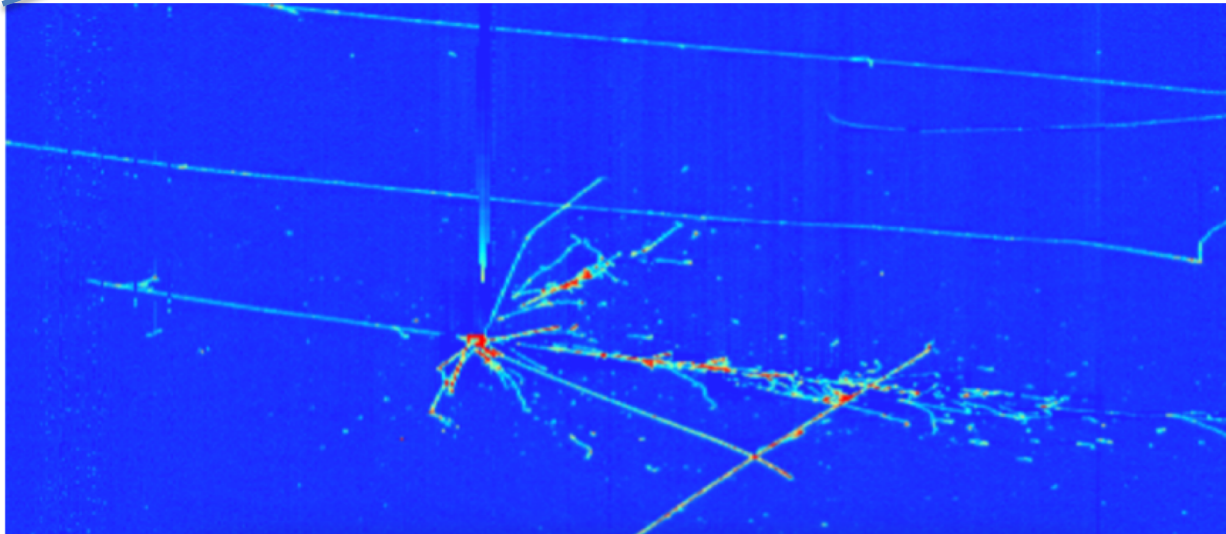

1 APA – 2,560 channels

Oregon State University
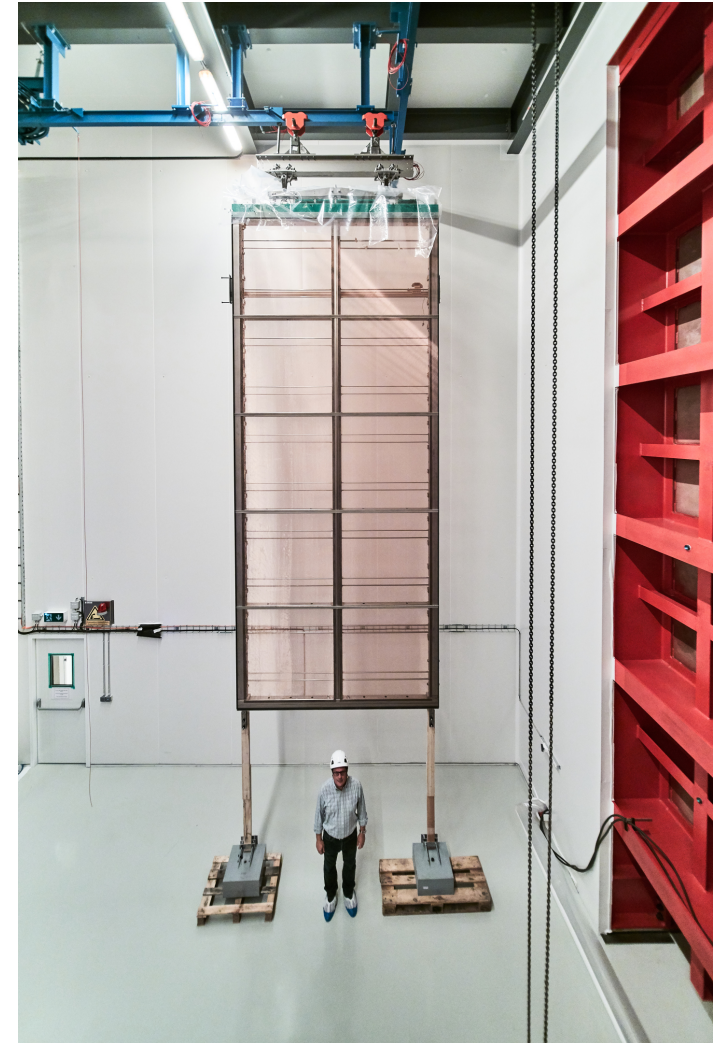
DUNE
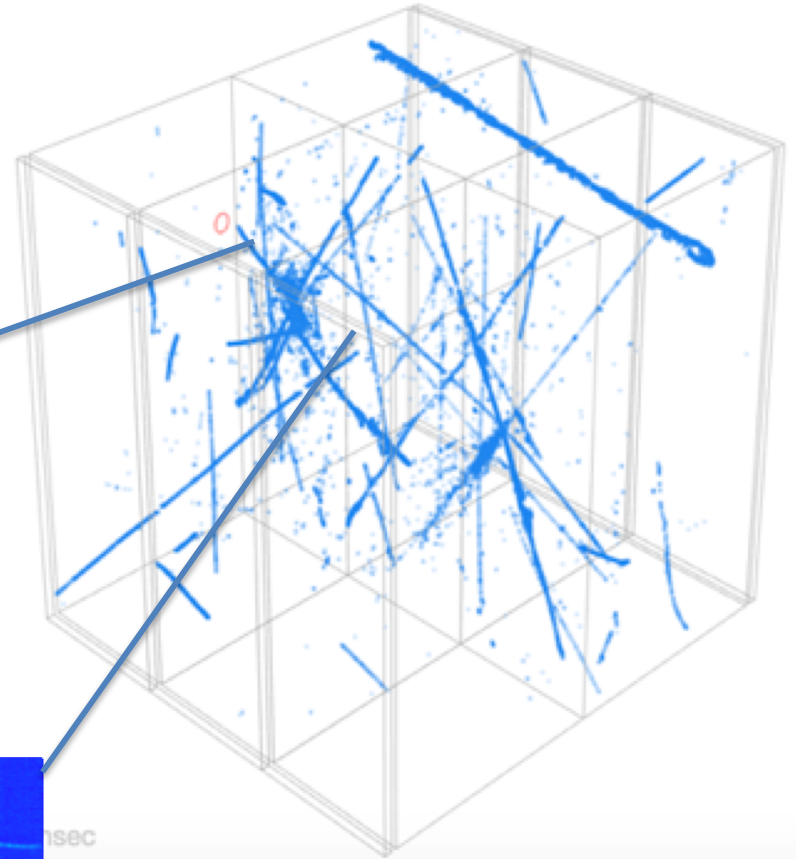
ProtoDUNE-SP Event sizes

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

- Compare ~2 MB for ATLAS/CMS p-p

- And ~8 MB for ALICE Pb-Pb

~5 mm resolution over 7x7x7m

Oregon State University

DUNE

"Challenge"

- Far detector data comes in **very** large chunks – 25xProtoDUNE

- Beam and cosmic interactions are 1-6 GB each per 10kt Module
  - Rates are ~ 5000/day/module dominated by cosmics
  - Need to read out 3-10 ms of data to get a full drift

- One 5.4 ms readout means
  - 1 tick = 12 bits
  - 1 channel = 10,800 ticks -> 16 KB
  - 1 APA = 2,560 channels = 41 MB uncompressed
  - 1 module = 150 APA's = 6.2 GB uncompressed

- All data types add up to about:
  - **~12 PB/year/module (uncompressed)** x 4 modules
  - ~1.6 GB/sec for 4 modules, DC…

- ProtoDUNE-SP already ran at this rate, but for only 6 weeks.

Oregon State University

DUNE

To make it more interesting

- DUNE should be sensitive to nearby (Milky Way and friends) supernovae. Real ones are every 30-200 years but radioactive decays can make false alarms
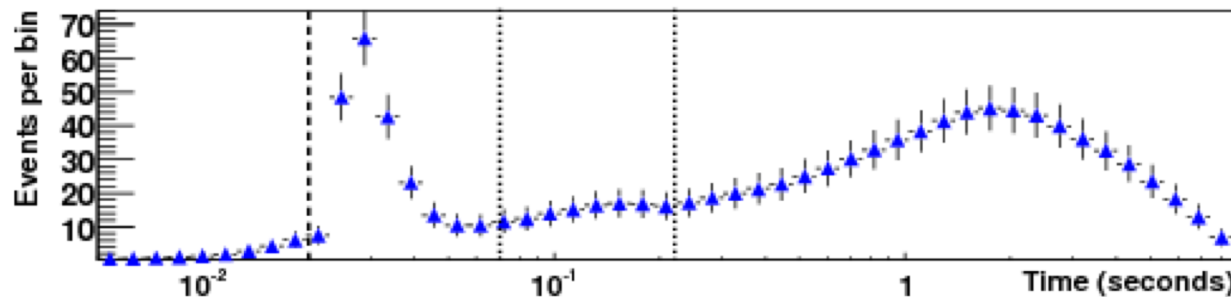
- Supernova readout = 100 sec, one trigger/month

- 100 sec readout implies
    - 1 channel = 300 MB uncompressed
    - 1 APA = 768 GB uncompressed
    - 1 module = 115 TB uncompressed
    - 4 modules = **460 TB** … takes 10 hrs to read at 100 Gbs
- Some calibration runs will be similar in scope….

Oregon State University

DUNE

DUNE FD-Data for Supernova

4 150-APA Modules
460 TB over 100 s

Our processing systems
can hold a couple of APA's

8
8
9/24/19
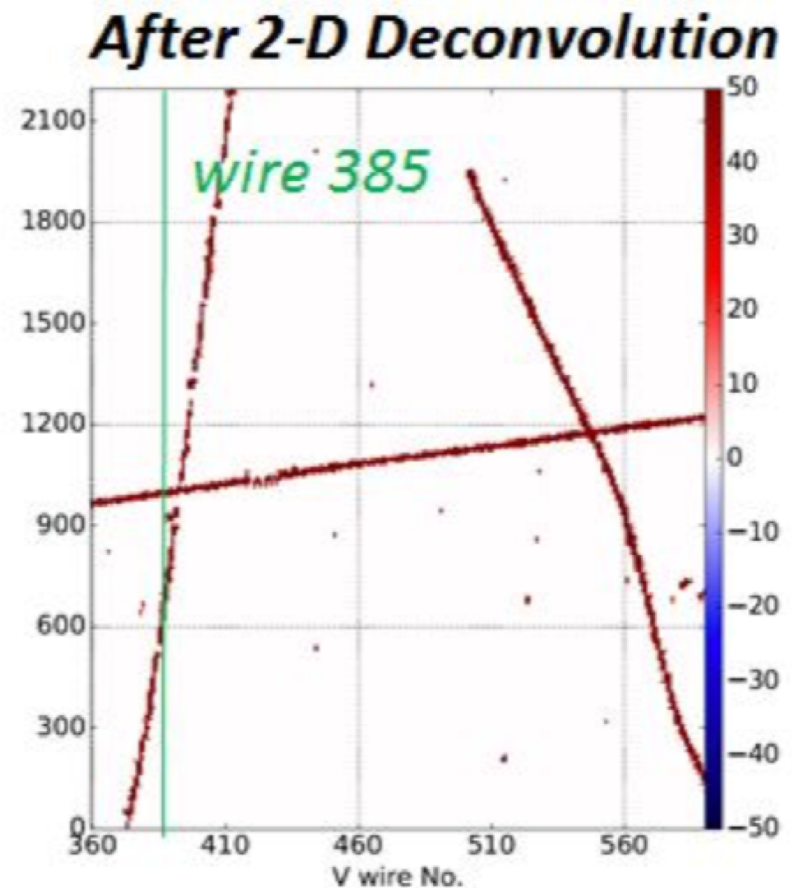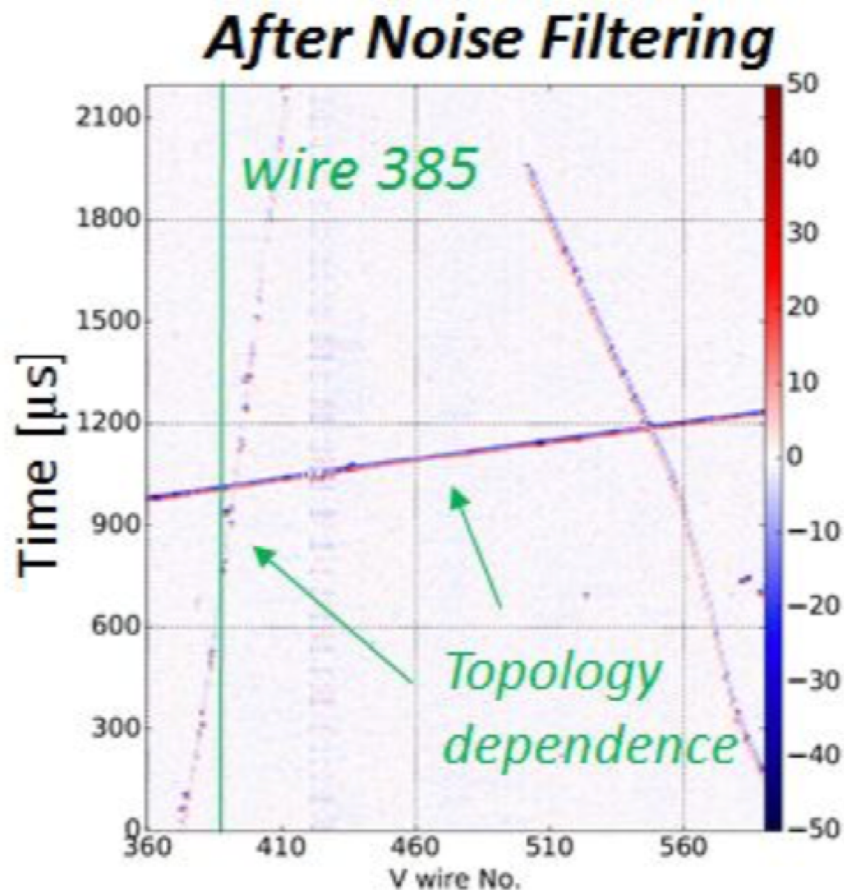
DUNE FD-Data

Future solutions

Logistical problem

- We can split the data up into 40 MB 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 together.
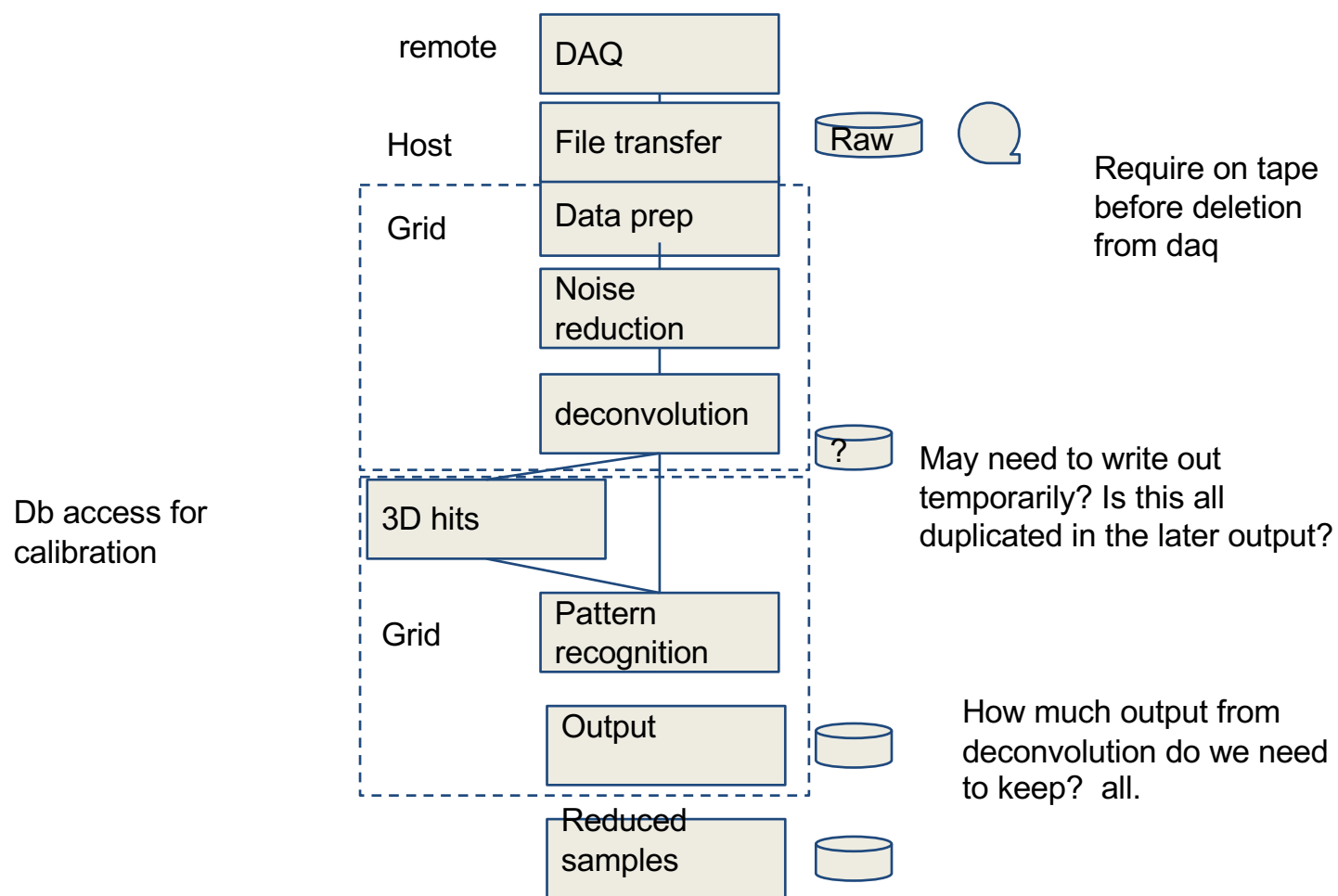
Part size

Noise filtering requires > 100 samples for Fourier analysis
2-D deconvolution requires many adjacent wires

9/24/19

Oregon State
University

Reconstruction workflow

remote DAQ

Host File transfer Raw

Require on tape
before deletion
from daq

Grid Data prep

Noise
reduction

deconvolution ?

May need to write out
temporarily? Is this all
duplicated in the later output?

Db access for
calibration

3D hits

Grid Pattern
recognition

Output

How much output from
deconvolution do we need
to keep? all.

Reduced
samples

Data layout requirements

- Treat data as cells = 1 APA x 5-10 ms = 40-80 MB compressed

  - APA level ensures full information for deconvolution is present

- Signal processing time is ~72 sec/APA

- Beam/cosmic trigger readouts deliver up to 150 APAs together – 1-3 GB

  - These can be processed together on the same "track"

- SNB readouts will span multiple (like 10,000) files and take ~10 hrs to transfer but only happen ~1/month.

  - Requires special treatment

Oregon State University

Data layouts

Time Localized readout
(cosmics/beam/calibration)
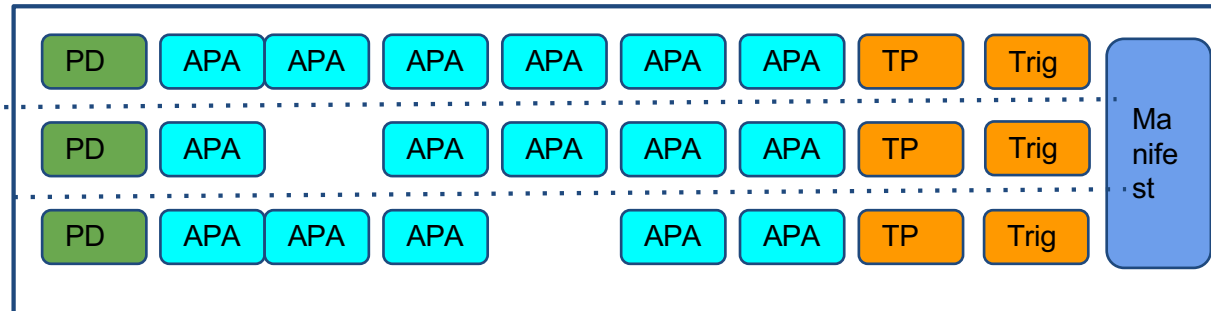
Time Extended (SNB) readout
aggregate
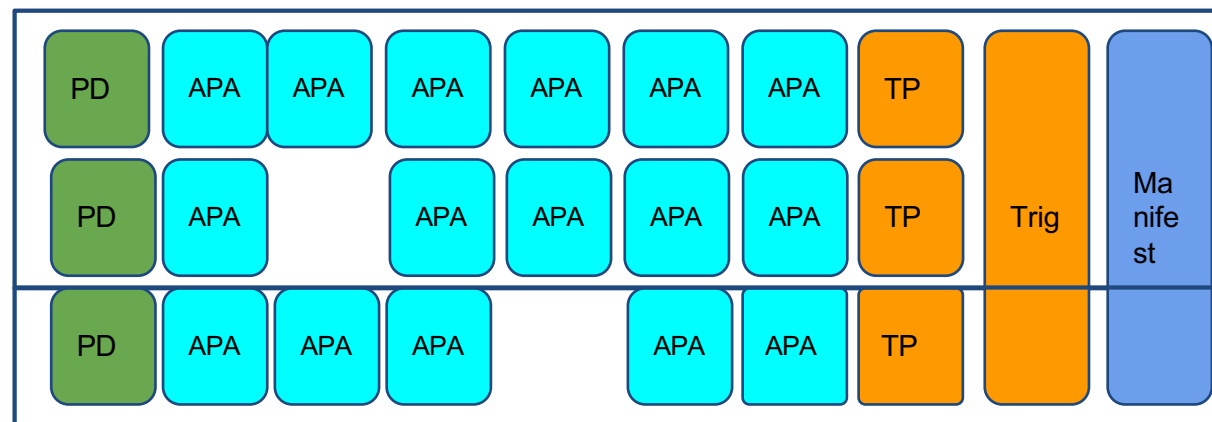
Issues

- What is an event?

  – For normal operations - it's the usual

  – For ND – need to stitch 2-3 detectors together – MINOS/MINERvA?

- Parallel processing

  – As APA's are separable, can they be processed in parallel?

  – Multistep processing – temp or permanent storage of intermediate steps?

- SNB issues

  – Need to have 2 time slices in memory together to do transform without edge effects.

Oregon State
University