ProtoDUNE-II Offline Data Processing Strategy

Barnali Chowdhury

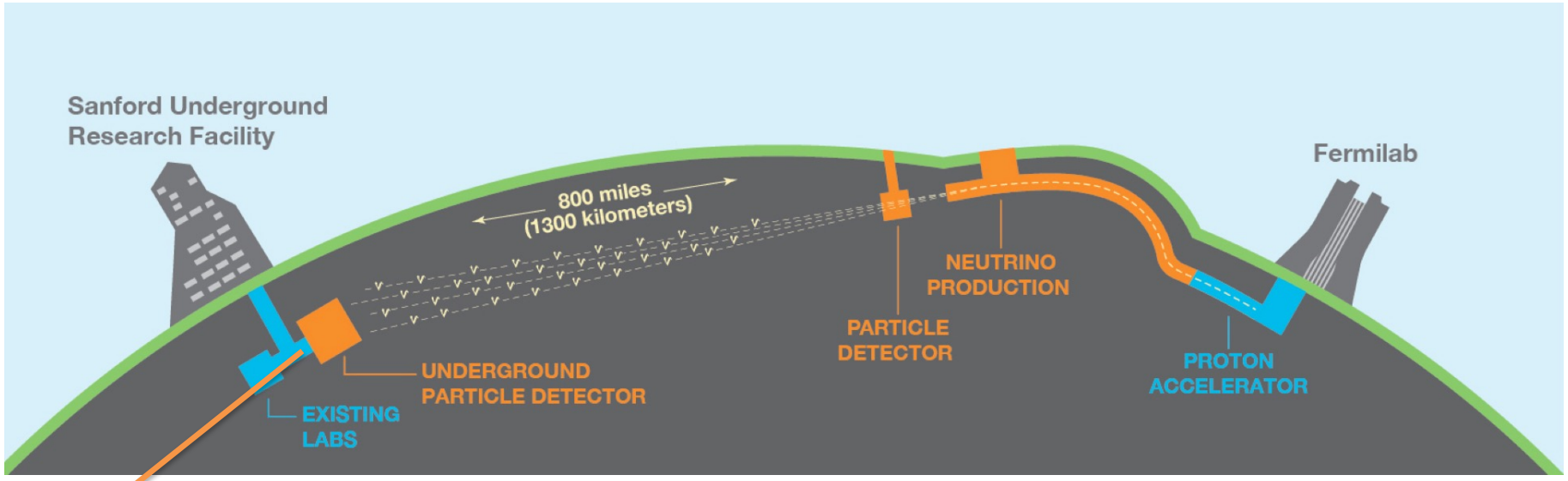
On behalf of the DUNE Collaboration

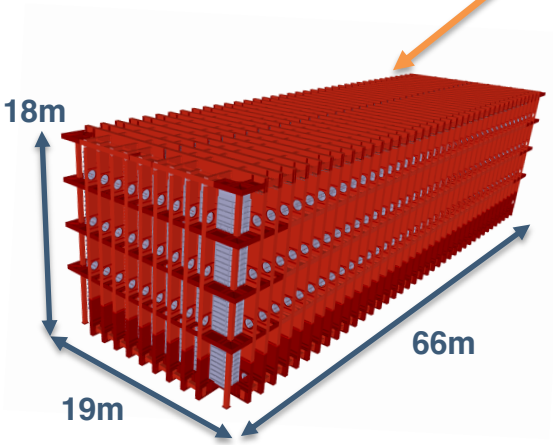
DPF-PHENO, 2024

May 16th, 2024

Main Physics Goals of DUNE

(picture not to scale)

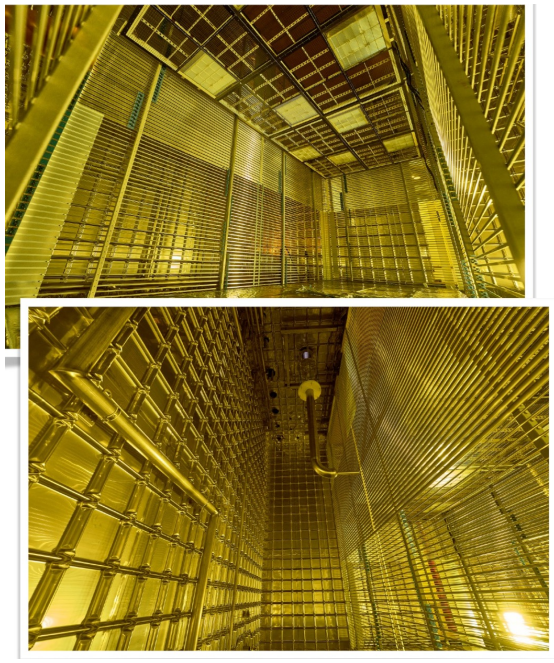


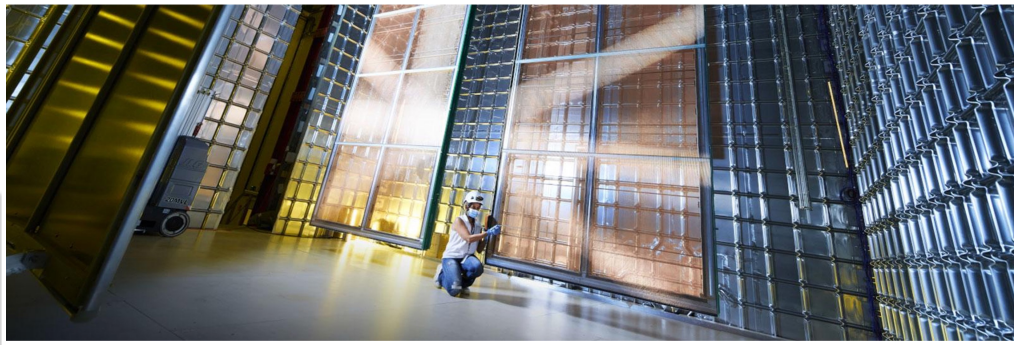


One of the 4 FD modules

- Determine neutrino mass ordering, observe and measure CP violation (if it is present) in the neutrino sector.
- Sensitivity to neutrinos from astrophysical sources (solar, atmospheric, supernova burst) and BSM physics.
- Far Detector (FD) located 1.5 km underground and designed to be four 17 kT Liquid Argon TPC at Sanford Underground Research Facility.
- Two LArTPC designs for two FD modules: Horizontal drift (HD) and Vertical drift (VD)

DUNE Far Detector (FD) Prototypes





Interior view of prototypes

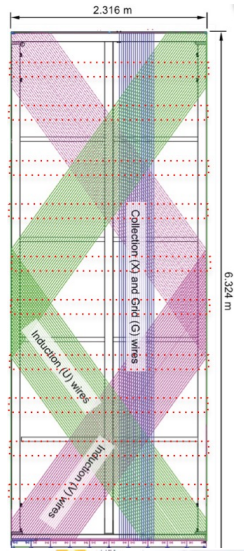
- DUNE has constructed two 800-ton FD prototypes at CERN to demonstrate proof-of-principle.
  - Horizontal Drift (HD) : ionization charges are drifted horizontally in liquid argon (LAr)
  - Vertical Drift (VD) : ionization charges are drifted vertically in liquid argon (LAr)

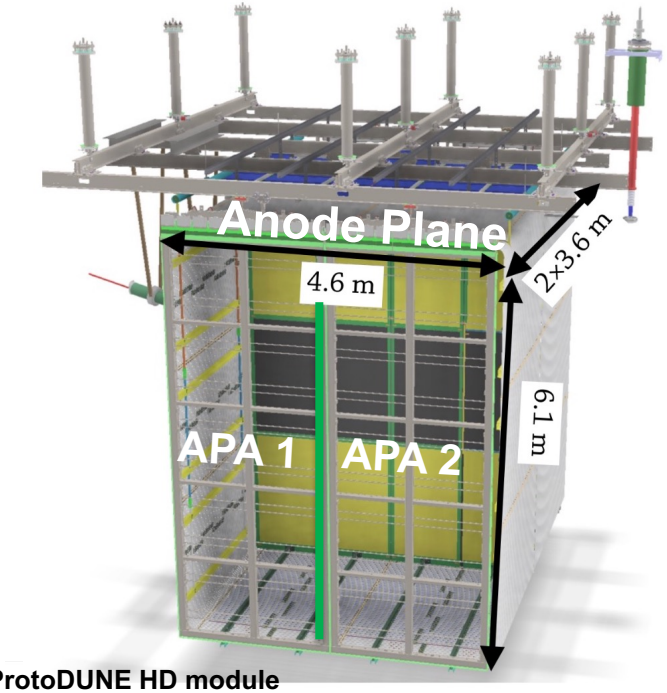
- Goal : ProtoDUNE HD and ProtoDUNE VD are the prototypes for full-scale elements of the FD HD and FD VD modules in DUNE.

ProtoDUNE HD : Overview

Time Projection Chamber (TPC):

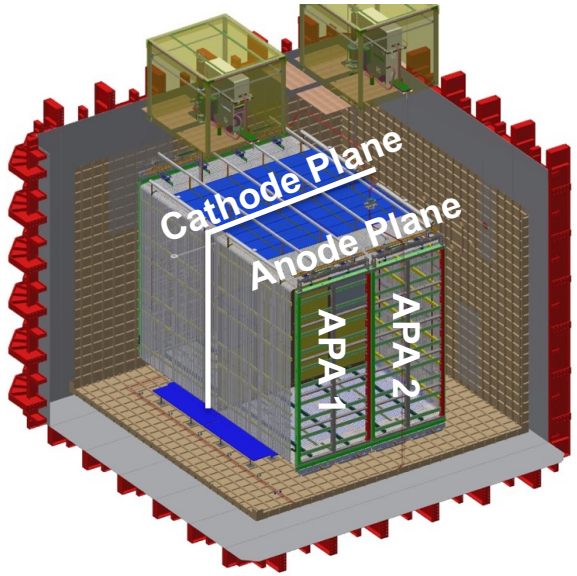
- The modular anode (A) and cathode (C) planes are constructed of "units", called anode plane assemblies (APAs) and cathode plane assemblies (CPAs)
- ProtoDUNE HD is instrumented with
  - 2 anode planes on sides – 2 APA / anode plane – total 4 APA units
  - 1 cathode plane in the middle – total 12 CPA units
- An individual APA has 2560 readout channel and each channel reads out a 14-bit analog-to-digital converter (ADC) every 0.5 µsec
- CPA is Held at 180 kV providing an E field of 500 V/cm





ProtoDUNE HD module

An Anode Plane Assembly (APA)



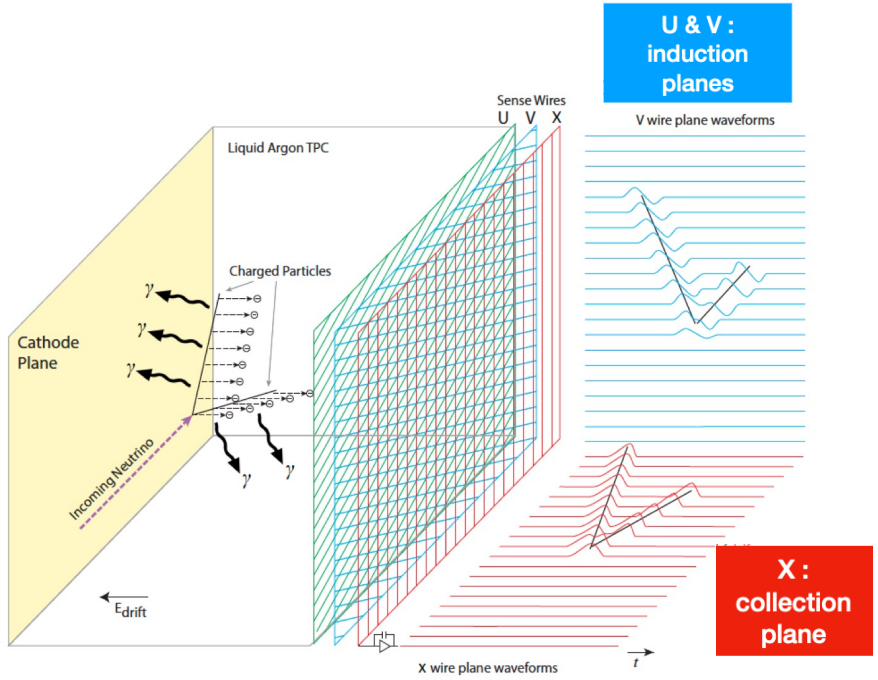
ProtoDUNE HD module

Light Collection:

- X-ARAPUCA photon detector modules mounted inside APAs
  - 10 photon detector modules per APA
  - Total 40 modules embedded

ProtoDUNE HD : Operation

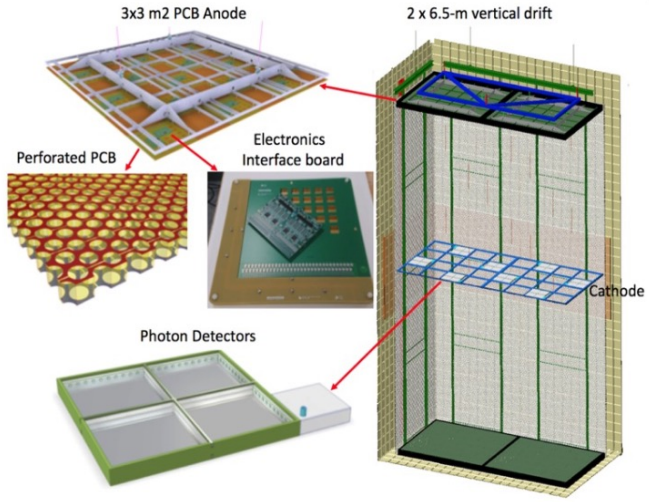
- Fully active interaction medium of liquid argon.
- Negatively charged ionization electrons from the neutrino interaction drift towards the APAs horizontally.
- Drift charges are collected on the anode plane, made up of the U, V and X wires and recorded by 3 readout wire planes.
- Photons are emitted through recombination between Ar^+ and e^- and detected by photon sensors.
- **Operational status of ProtoDUNE HD**
  - Detector: Construction and installation is complete
  - Filling with LAr: Filling process ended on 05/03, purification started 05/04
  - Detector Studies: DAQ continues collecting noise data and cosmic activity data
  - Test Calibration techniques: Planning to collect laser track calibration data, radioactive source data, etc.
  - Beam Run: First week of beam approximately starts June 19th for one week
    - 8 weeks of beam in July-August
  - Beam Characterization: Will be exposed to charge particle beam (0.3 -7 GeV/c) of both polarity
    - Similar momenta to those of particles produced in neutrino interactions at DUNE



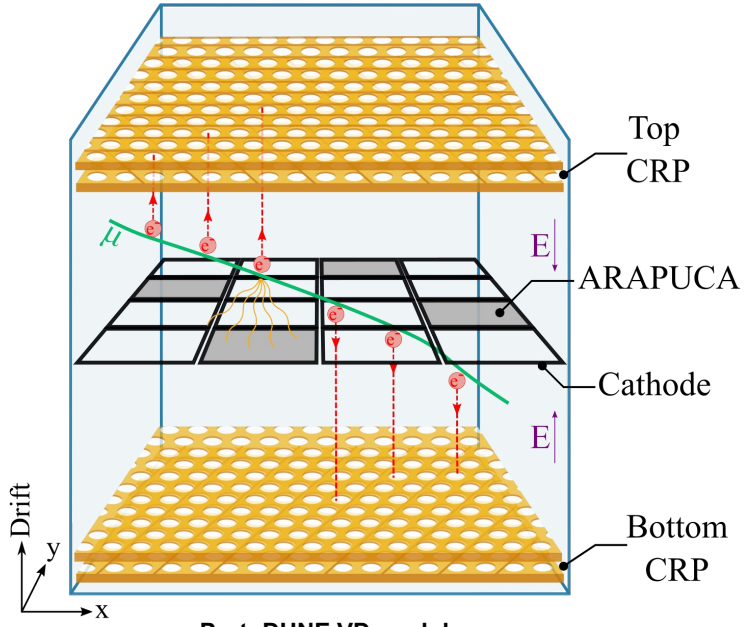
The general operating principle of the ProtoDUNE HD LArTPC

ProtoDUNE VD : Overview & Operation

- Anode planes are horizontal, cathode plane is suspended at mid-height.
- Ionization electrons move vertically.
- Anodes consist of Charge Readout Planes (CRPs)
- ProtoDUNE VD is instrumented with
  - 2 anode planes on top and bottom – 2 CRP / anode plane – total 4 CRP units
  - 1 cathode plane in the middle
- Total 16 Photon detectors (8 on cathode and 8 behind field cage) based on the X-ARAPUCA technology
- **Operational status of ProtoDUNE VD**
  - Detector: Vertical Drift main TPC components, for example, top and bottom CRPs, cathode and PD modules are installed
  - Filling with LAr: LAr will be transferred from HD in 2024
  - Detector Studies : DAQ continues collecting noise data
  - Beam Run : Expect beam in 2025
  - Beam Characterization: Will be exposed to charge particle beam (0.3 -7 GeV/c)
    - Similar momenta to those of particles produced in neutrino interactions at DUNE

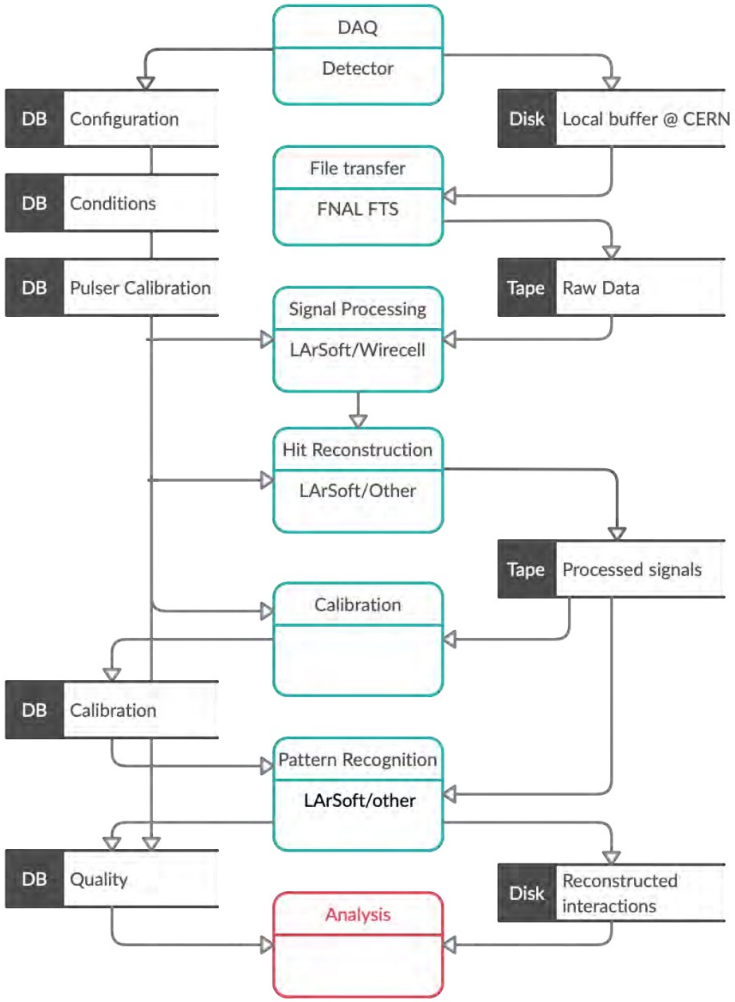


A cross-section of a single vertical drift module



ProtoDUNE VD module

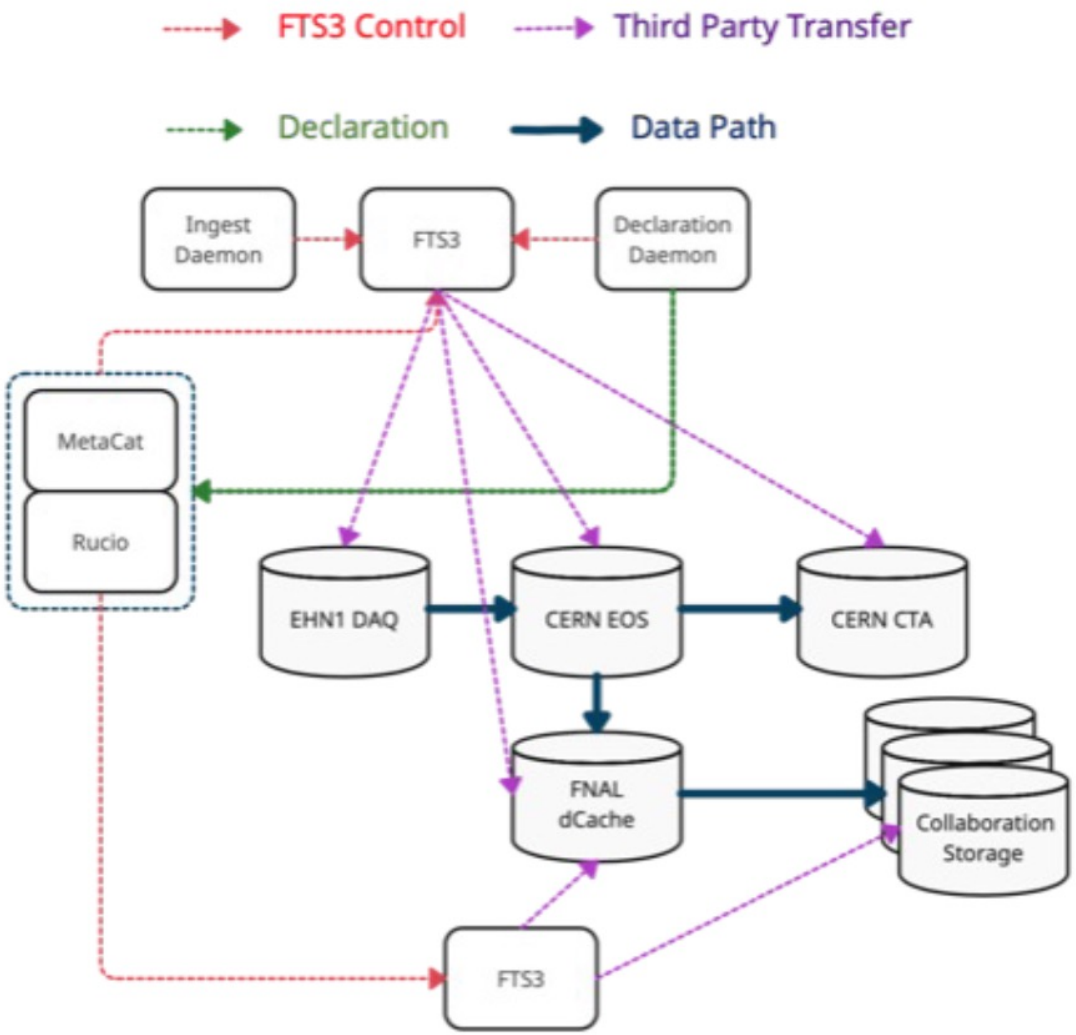
ProtoDUNE HD & VD : Offline Processing of Raw Data



Outline of the offline processing flow for raw data in ProtoDUNE-II

- The central boxes show the processing steps
- Offline processing starts with the transfer of data from neutrino facility at CERN
- Offline processing stage, labeled “Signal Processing”
  - Noise reduction, Deconvolution etc.
  - **x 5 (ProtoDUNE)** data reduction
- Hit finding
- Pattern recognition (Tensorflow, Pandora, WireCell)
- Analysis sample creation
- The software algorithms are suitable for different computing architecture

ProtoDUNE HD & VD : Data Pipeline Diagram



- Data generated at CERN needs to be buffered locally and then transferred to permanent storage at the host lab(s)
- Ingest Daemon and Declaration Daemon
  - Ingest daemon brings files from experimental systems to dropbox
  - Declaration daemon declares them to MetaCat and Rucio and makes rules to get them to final destinations.
- All transfers done via FTS3, a CERN product
- There are 2 buffers at CERN.. one on the local DAQ machines, the second one on EOSPUBLIC
- The files are available at CERN and at Fermilab as well.

ProtoDUNE HD Raw Data Volumes

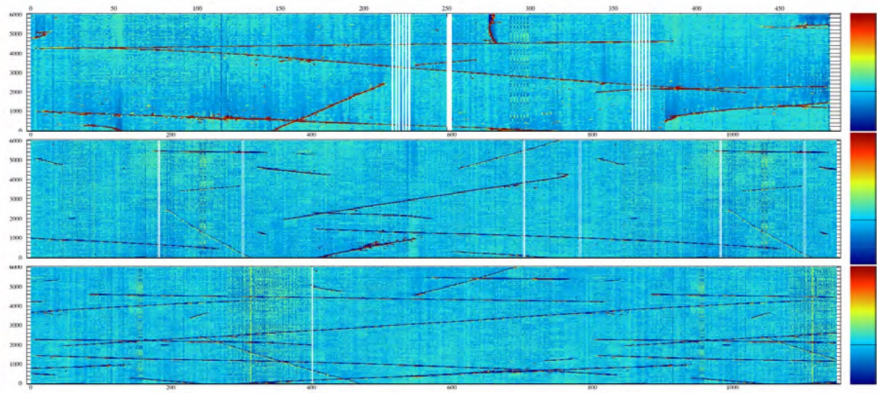
- For the initial ProtoDUNE HD data volumes, we use our ProtoDUNE SP experience and assume that raw data sizes and hit-finding CPU times scale with the number of APAs
- Rate of beam data from ProtoDUNE HD will be similar to rate of events from the full Far Detector.
- ~ 25 MB of uncompressed raw data from a single APA

ProtoDUNE HD Raw Data Volume Estimate

Quantity	Value	Explanation
Number of APAs	4	DAQ spec.
TPC channels/APA	2560	DAQ spec.
TPC ADC sampling time	512 ns	DAQ spec.
TPC ADC dynamic range	14 bits	DAQ spec.
Readout time	3 ms	DAQ spec.
Single APA readout	~25 MB	Uncompressed estimate
Full detector readout	~120 MB	Uncompressed
Effective compression factor	1	

ProtoDUNE SP Raw Data Volume

Quantity	Value	Explanation
Number of APAs	6	
Number of channels/APA	2,560	
Readout time	3 ms	
# of time slices	6000	
Single APA readout	23 MB	Uncompressed estimate
Full detector readout	178 MB	Uncompressed real
Full detector readout	70 MB	Compressed real
Effective compression factor	2.5	



Event display from 1 APA's worth of Raw Data is shown for ProtoDUNE-I trigger record collected in October 2018. ~25 MB of data.

ProtoDUNE HD & VD Data Representation

- DAQ writes data in HDF5 format for raw data storage
  - Technical reasons – no need for ROOT data model support at raw data level, lower overhead, higher performance
  - Multiple threads can write to the same file
  - HDF5 is used commonly in ML applications and in HPC workflows
- HDF5 will be used as raw data file format in ProtoDUNE HD and VD
  - Have successfully taken data in HDF5 from HD, VD Cold Box testing
- ProtoDUNE HD and VD will support multiple data representation
- Offline Read-in Software for HDF5
  - Delayed Reading: Implemented for HDF5. Put proxies in the art event which lets downstream tools access the input file and deserialize it.
  - Input source just opens and closes the file(s) and leaves a file handle in the *art* event memory.
  - Decoder tools do the actual I/O and transform data into useable formats.
- Existing Input source and decoder tools for each detector/prototype
  - Vertical Drift Coldbox, Horizontal Drift Coldbox, ICEBERG
  - Supports per-APA/CRP reading

ProtoDUNE HD & VD TPC Signal Processing

Noise Filter:

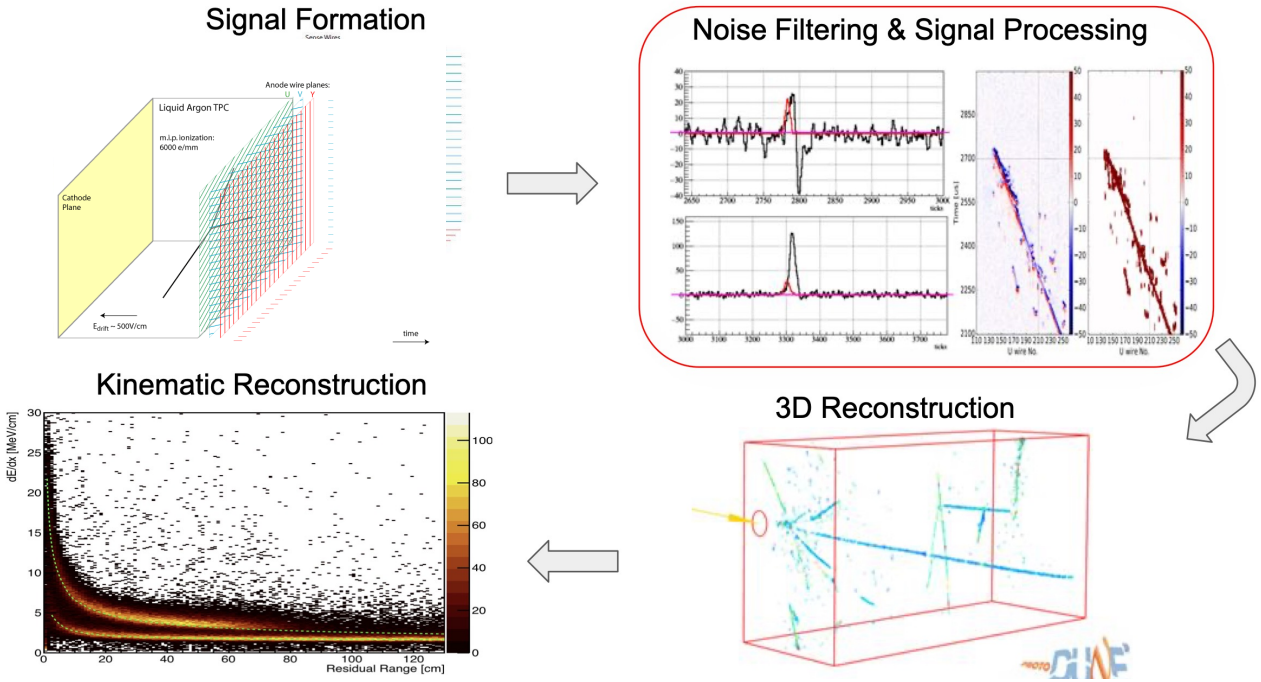
- Noise filtering is a key step towards a high-quality signal processing
- Different types of noise were found from HD & VD coldbox data
  - Single-channel : RC undershoot correction (aka “tail removal”) and the pedestal removal.
  - Group-channel : Coherent noise
  - Microphonic noise : Slow pedestal variation in space & time

- Currently analyzing noise run from ProtoDUNE HD to investigate various noise and to mitigate them

Signal Processing:

- Signal extraction is a key step for 3D event reconstruction
- The goal of signal processing is to convert the recorded waveform into ionization electron distribution
  - Accurate calculation of the field response – by Garfield software
  - Appropriate signal extraction techniques – by 2D Deconvolution method:

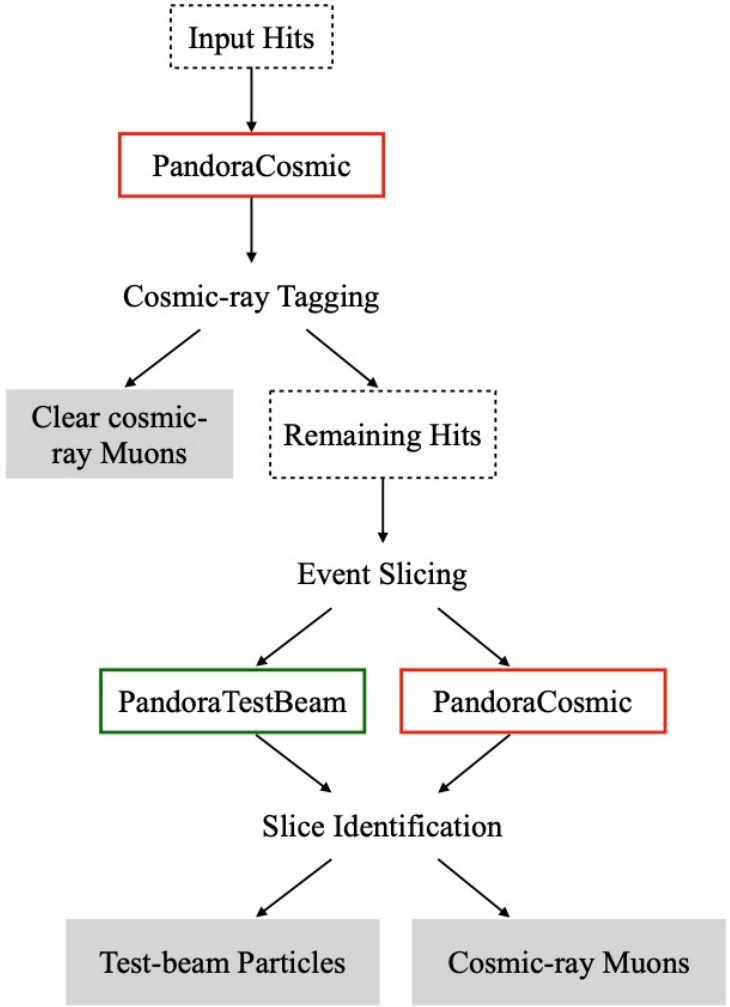
$$M_i(t_0) = \int_{-\infty}^{\infty} (\ldots + R_1(t_0 - t) \cdot S_{i-1}(t) + R_0(t_0 - t) \cdot S_i(t) + R_1(t_0 - t) \cdot S_{i+1}(t) + \ldots) \cdot dt$$



Overview of LArTPC Reconstruction

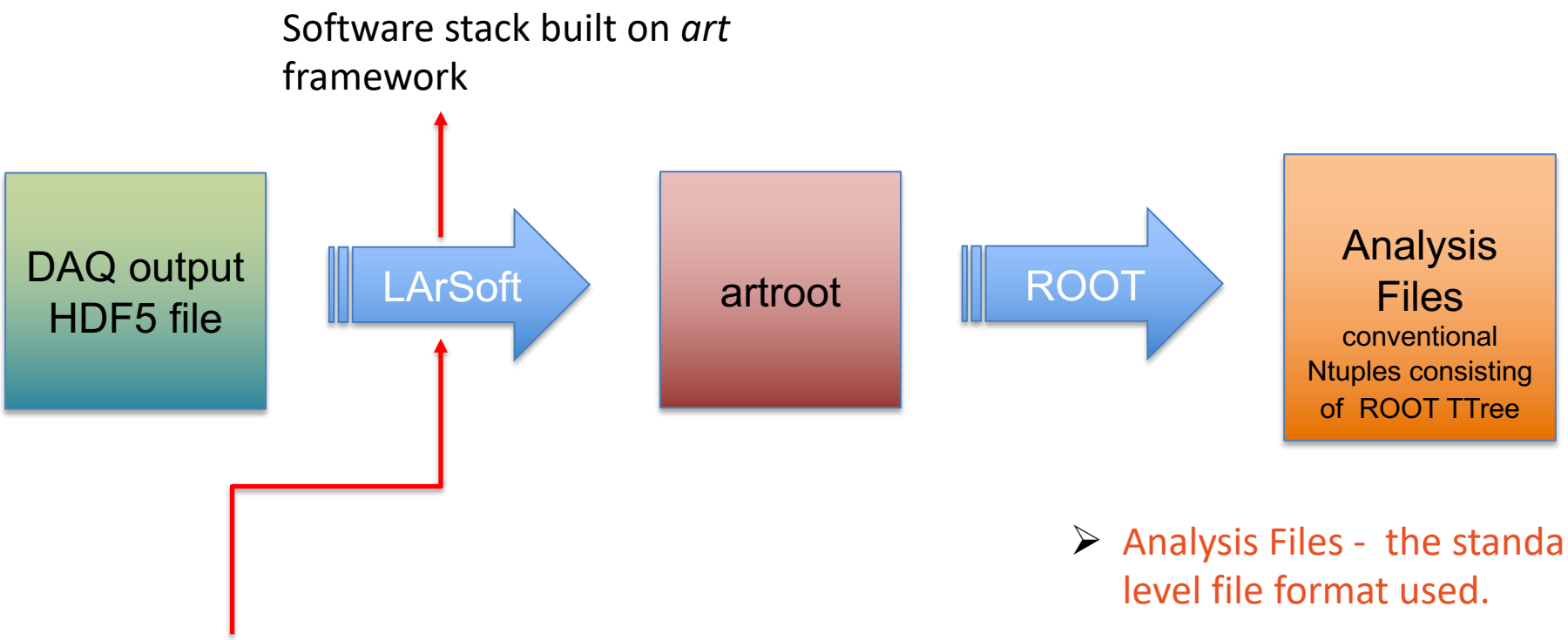
ProtoDUNE Event Reconstruction (with PANDORA)

- Signal processing is done.
- Hits are formed from the collected or induced charge waveforms by fitting Gaussian functions to peaks in the waveforms.
- The inputs to the Pandora pattern recognition are hits, and each hit represents a signal detected on a specific wire at a specific time.
- The performance of Pandora has been extensively evaluated for simulated charged test-beam and cosmic-ray interactions in the ProtoDUNE-SP detector.
- ProtoDUNE is going to use Pandora for 3D event reconstruction and classification.
- Ref. https://arxiv.org/pdf/2206.14521



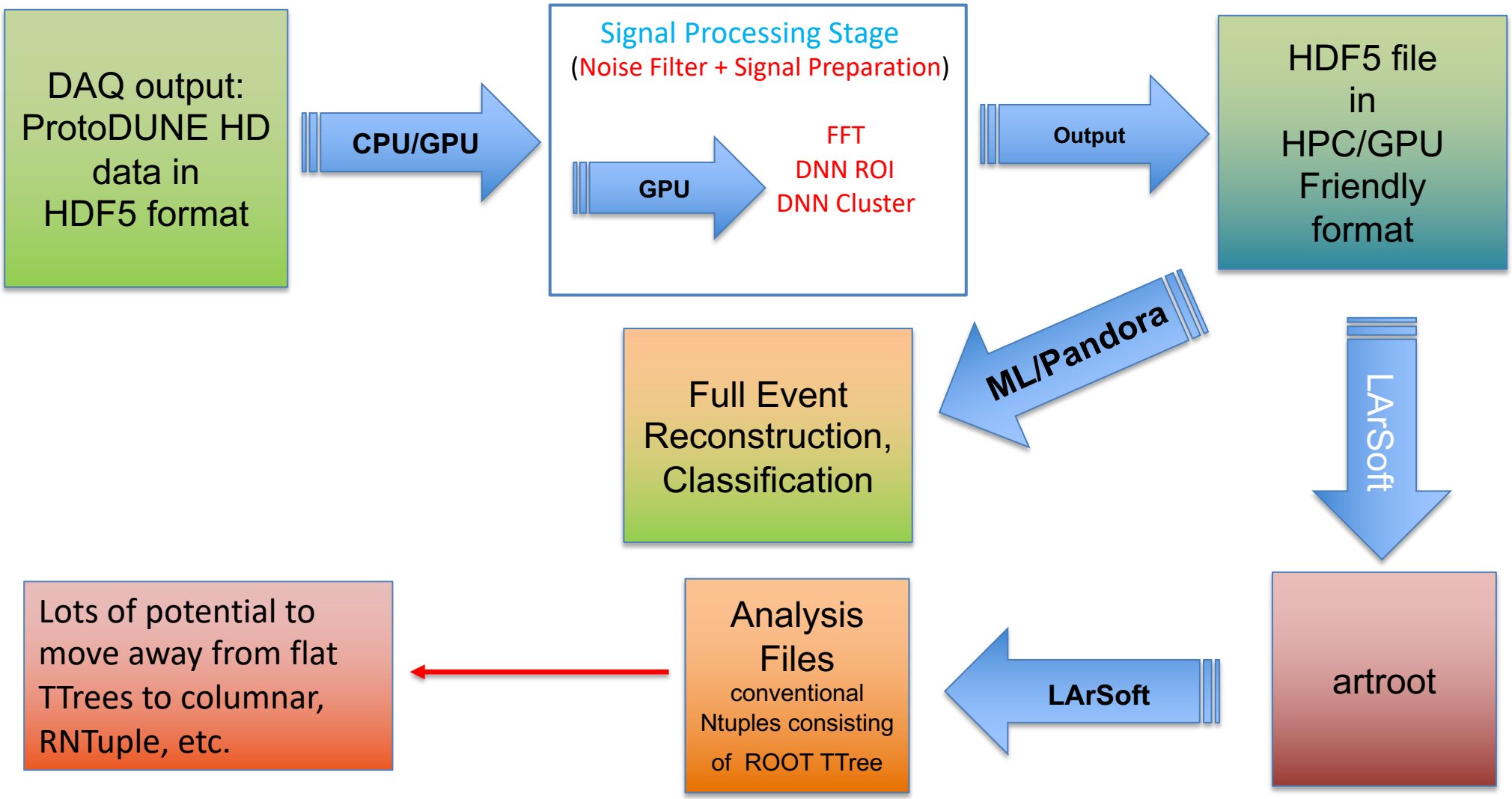
Outline of the Pandora consolidated reconstruction

Offline Workflow in Current ProtoDUNE Data Model



- Analysis Files - the standard analysis-level file format used.
- Conventional ntuples, consisting of a ROOT TTree, in which each record represents a single physics interaction or “slice”.

Offline Workflow in Potential ProtoDUNE Data Model



Summary

- ProtoDUNE HD is collecting noise run and cosmic activity.
- ProtoDUNE HD will start receiving beam in the week of June 19th.

- ProtoDUNE VD is collecting noise data.
- ProtoDUNE VD will start receiving beam in 2025.

- For both detectors, DUNE is currently analyzing noise and cosmic data.
- Finalizing the details of our data processing strategy.

Thank You!

DUNE Collaboration