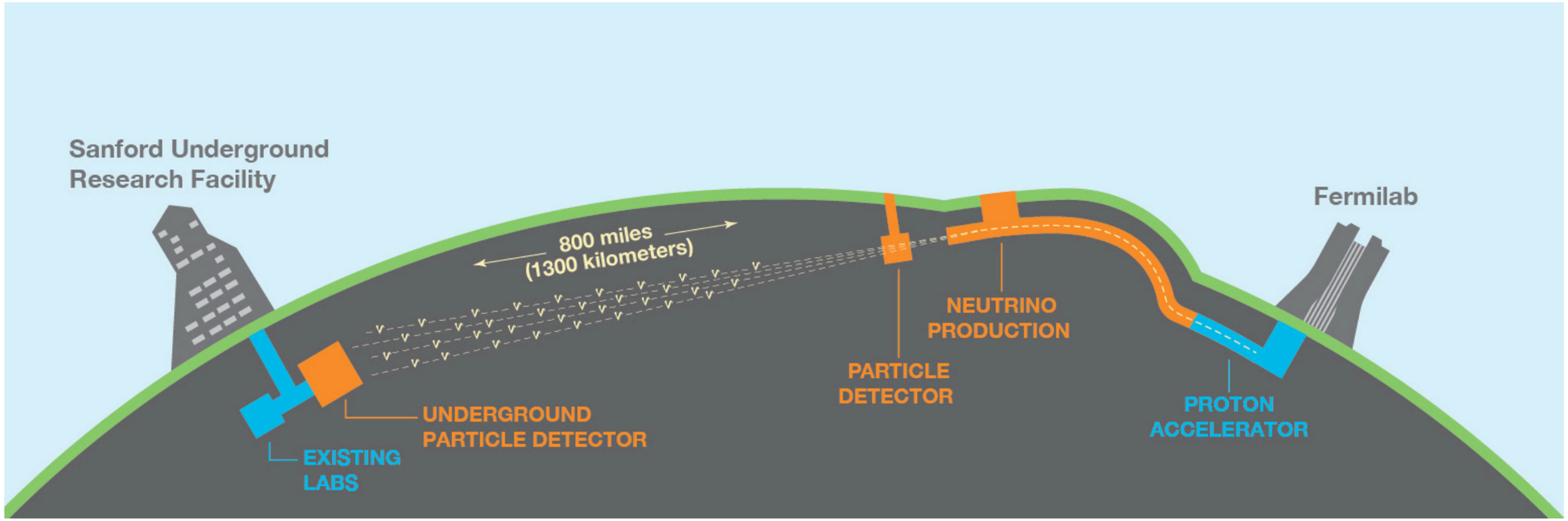
DUNE Offline Workflows

Michael Kirby, Fermilab/Scientific Computing Division

Oct 23, 2019

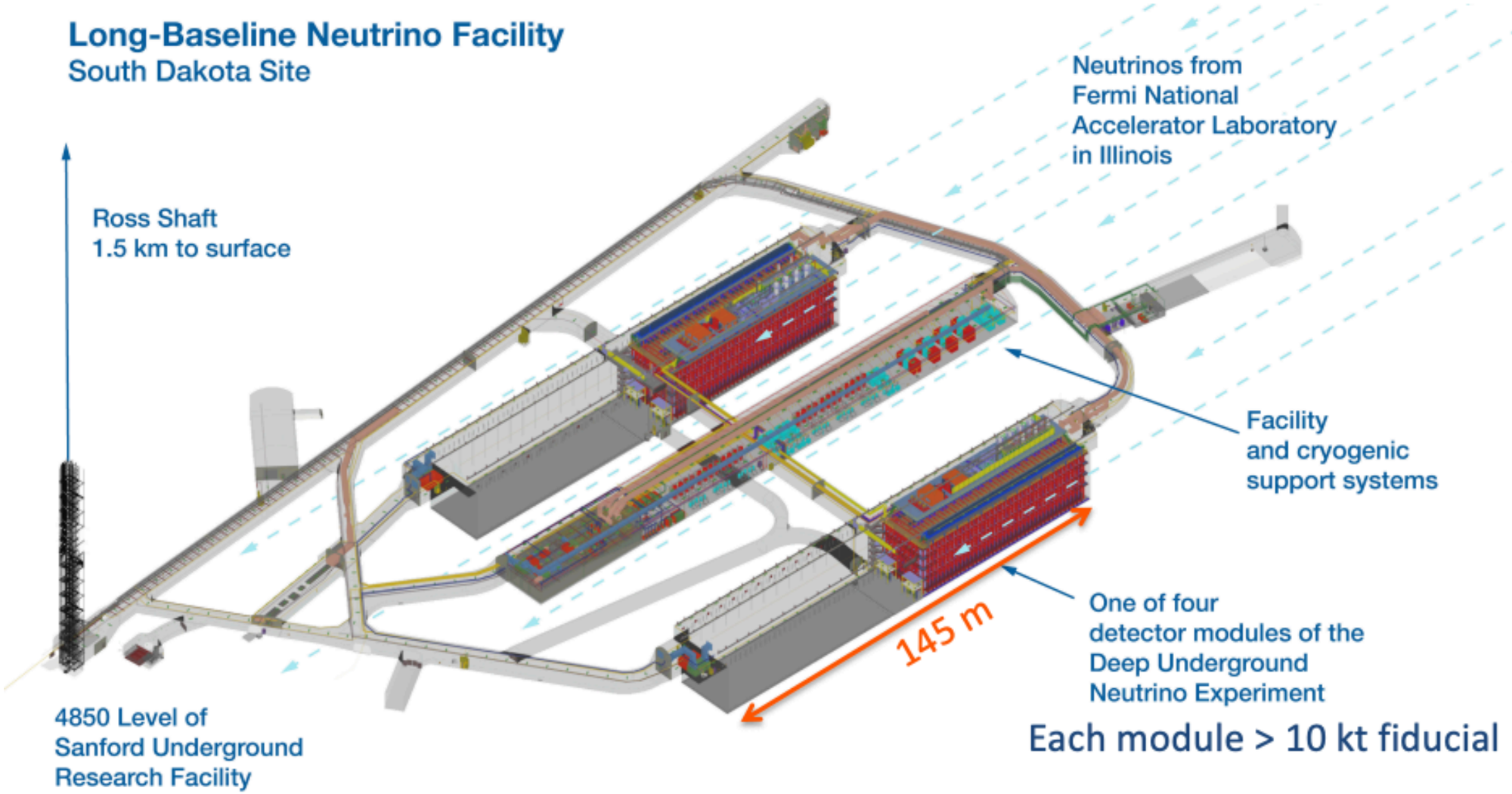
Quick reminder about DUNE



- neutrino experiment looking for neutrino oscillation parameters (mass ordering, matter vs antimatter asymmetry, unitarity), proton decay, supernova neutrinos, and more.
- 40 kT LAr TPC detectors at 4850 ft underground in Lead, SD (Homestake Mine)
- Near Detector (still in design) at Fermilab near the neutrino production
- Two prototypes at CERN - (ProtoDUNE Single Phase - ProtoDUNE Dual Phase)

DUNE Far Detector Design

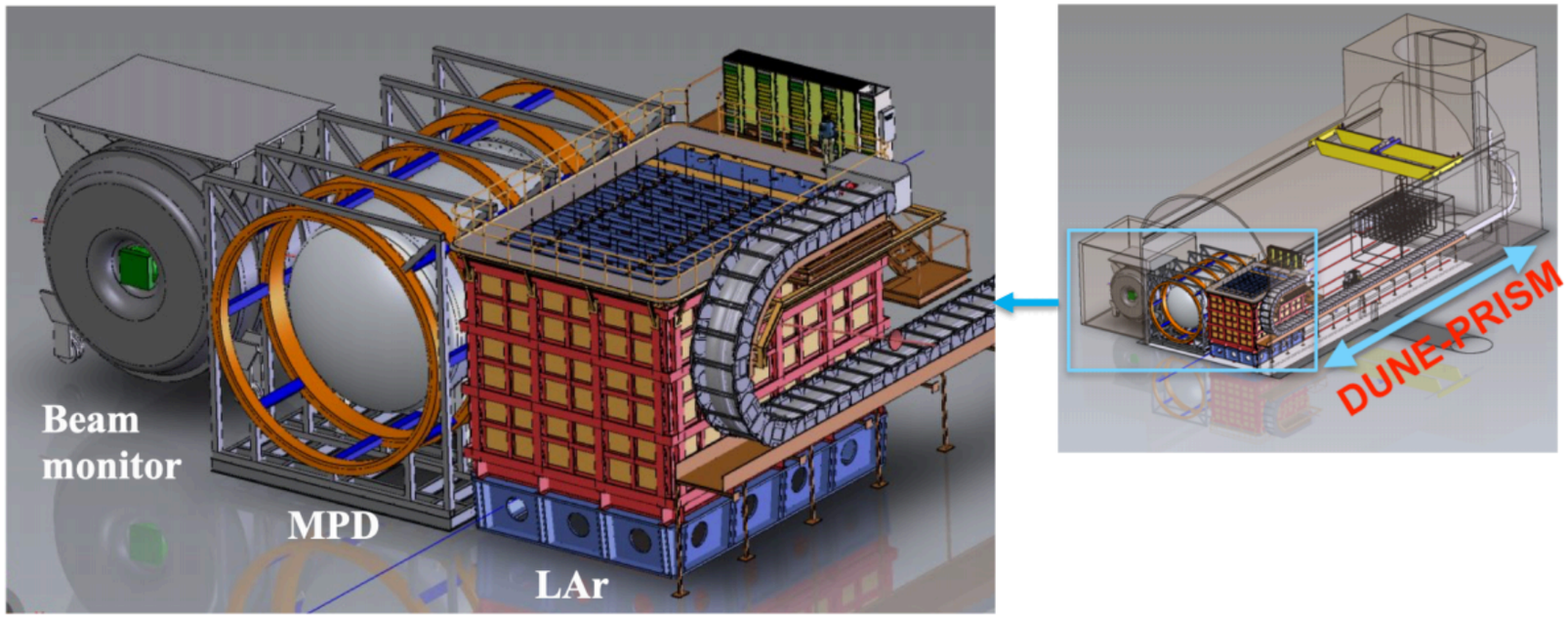
Slide: Ed Blucher



Fermilab

(Proposed) Near Detector Design

Slide: Ed Blucher

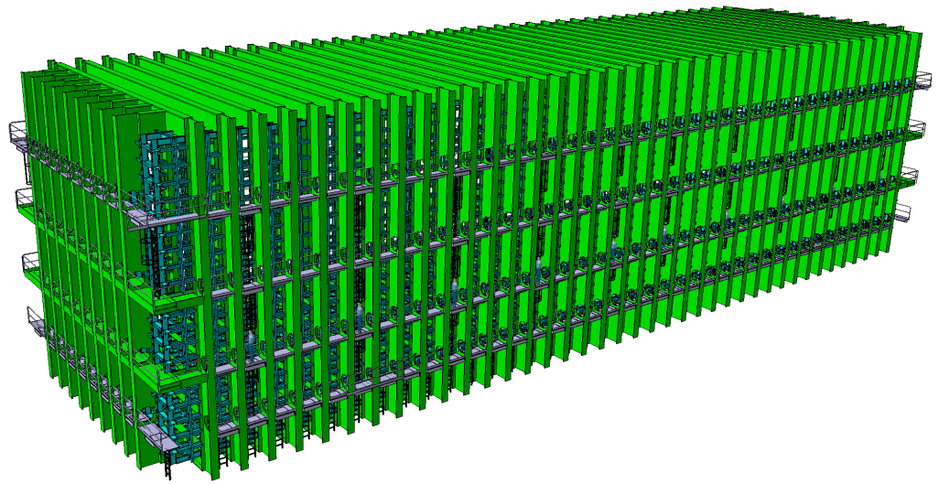


- LAr: Highly segmented LAr TPC (ArgonCube)
- MPD (Multi-purpose detector): High Pressure Gas Argon TPC, Calorimeter, and muon system magnetized by superconducting coils
- Beam monitor: High density plastic scintillator detector with tracking chambers and calorimetry in KLOE magnet
- DUNE-PRISM: Movement of LAr+MPD transverse to the beam, sampling different E_ν

Fermilab

DUNE Computing Challenge

- Actively running two experiments
  - ProtoDUNE-SP - beam (2018) and cosmic-ray operations
  - ProtoDUNE-DP - cosmic-ray operations
  - more than 3 PB of raw data for SP and DP
- Far Detector frontend is capable of producing incredible data rates
  - reduction through trigger, zero suppression, and compression of data - all areas of active research
  - DAQ constrained to 30 PB/year from far detector
  - beam trigger, calibrations, etc
- Design studies of Near Detector are very active
  - large uncertainty in final design and data structure
  - event rate expected to be >> 1 neutrino interaction per spill
- Now utilizing art/LArSoft frameworks to publish ProtoDUNE data



Source	Annual Data Volume	Assumptions
Beam interactions	27 TB	10 MeV threshold in coincidence with beam time, including cosmic coincidence; 5.4 ms readout
^{39}Ar, cosmics and atmospheric neutrinos	10 PB	5.4 ms readout
Radiological backgrounds	< 2 PB	< 1 per month fake rate for SNB trigger
Cold electronics calibration	200 TB	
Radioactive source calibration	100 TB	< 10 Hz source rate; single APA readout; 5.4 ms readout
Laser calibration	200 TB	10^6 total laser pulses; half the TPC channels illuminated per pulse; lossy compression (zero-suppression) on all channels
Random triggers	60 TB	45 per day

Estimated Data Volume from one FD Single Phase Module

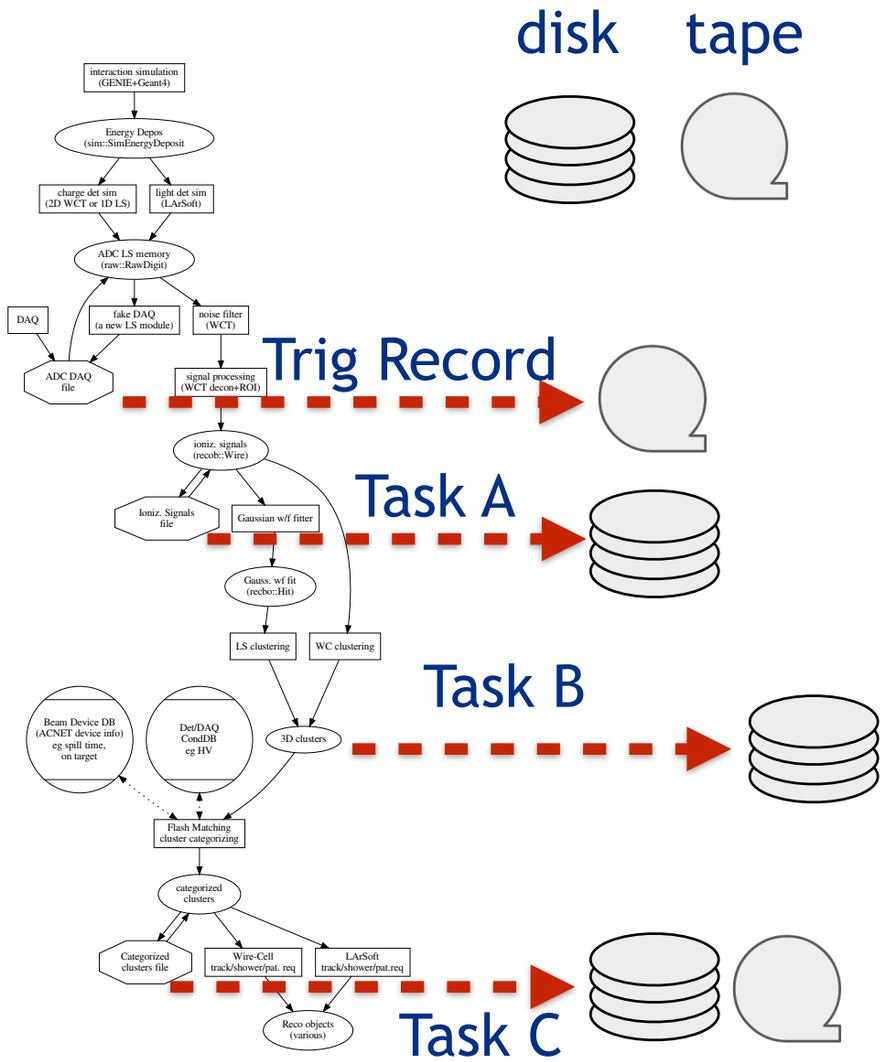
DUNE Event Data Model and Offline Processing

- As the Near Detector is not yet fully defined, have focused on the Far Detector for Event Data Model
- Readout of the FD is considered a “Trigger Record”
- Two dominant types of trigger records are considered:
  - time extent of readout is on the order of N times the TPC drift time (N ≲ 5) (neutrino beam, non-beam neutrinos, calibrations, proton decay)
  - extended time readout - 100 seconds of continuous readout of FD (supernova trigger proton decay)
- processing of a single trigger record can generate multiple “events” - consider these events to be causally separable regions of interest
- creation of events is done to minimize data volume and facilitate additional processing
- DUNE’s goal is to have a single framework that can be used for processing all trigger records and events

FD Beam Data Processing

- main production processing of LBNF beam trigger records for reconstruction of far detector interactions targeting physics measurements.
- Three tasks for processing:
  - signal processing and data reduction to include the noise filtering, 2D deconvolution, and hit finding.
  - Task A - hits and waveforms from Regions of Interest
  - Task B - input into processing which applies calibrations, and runs 3D reconstruction algorithms (e.g. WireCell, Pandora, etc)
  - Task C - categorized clusters/evts analysis or pattern recog

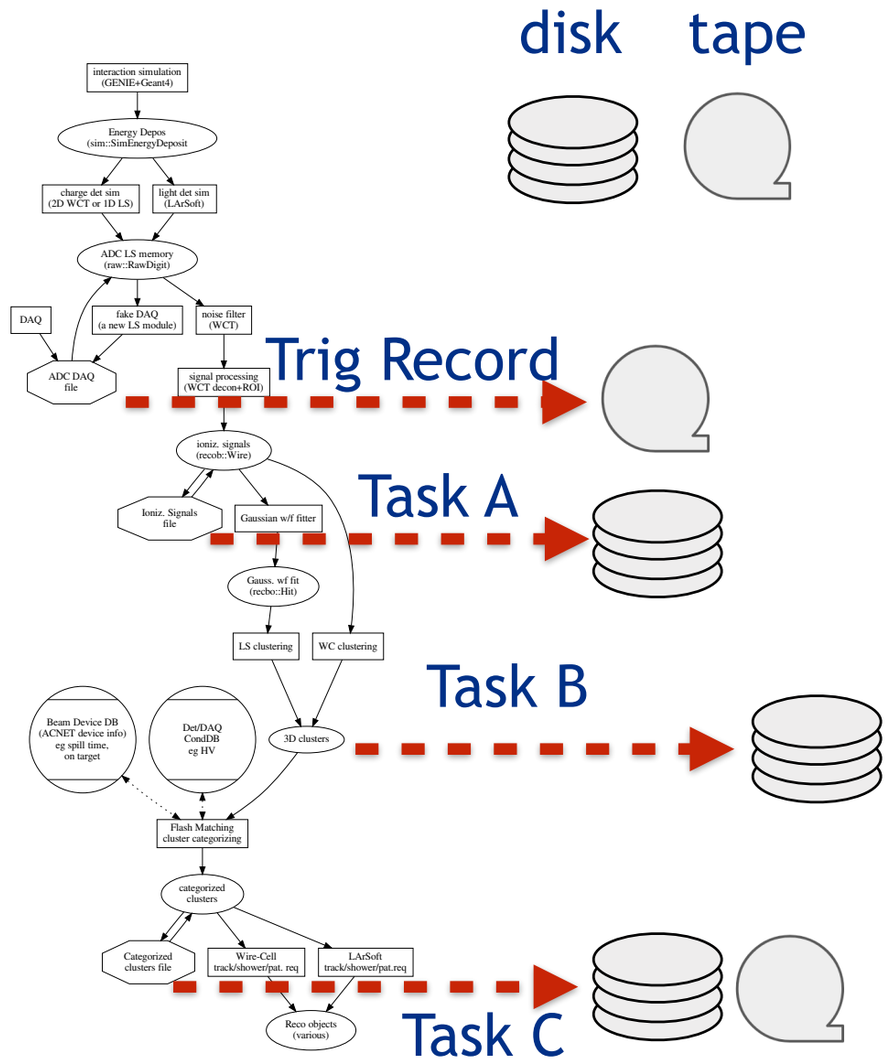
Data reduction estimates	MB/evt	MB/evt	
	Rawdigit	Wirehits	reduction
ProtoDUNE data	53.2	4.6	11.6
ProtoDUNE MC	46.1	11.9	3.9
FD nue	2073.6	11.6	179.1
FD bkg	2073.6	7.1	291.1



Fermilab

Dataset Volume from FD Processing

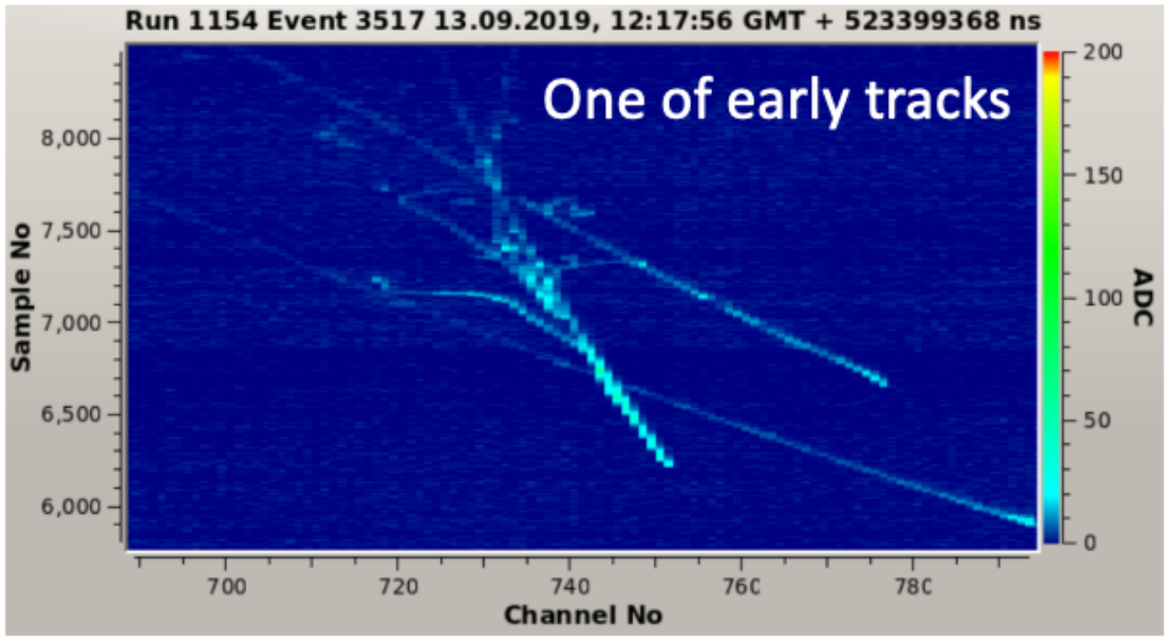
Far Detector								
Data Type	Amount/yr	Copies	Num of versions	Total/yr	Disk	Tape	Lifetime on disk	Note
Raw	10 PB	2	1	20 PB		100%	infinite	
Reduced Signal (A)	50 TB	3	2	300 TB	100%	100%	5 year	
Reco Path B	60 TB	3	2	360 TB	100%	0%	1 year	
Reco Path C	65 TB	3	2	390 TB	100%	0%	1 year	
MC	100 TB	3	2	600 TB				10:1 Ratio with raw?
Raw like	1 PB	1	1	1 PB	100%		2 year	~10% Raw for ML Training



Fermilab

FD Beam Processing Framework

ProtoDUNE - DP



- 2D deconvolution of signals, hit finding, ROI finding now parallel algorithms
- Work to make signal processing optimized on GPUs/Accelerators is already underway
- ability to have a trigger record input, but multiple events output seems different than collider event data models
- energy deposition can be separated by > 50 meters and causally separated
- flexibility to adapt to presence or absence of detector elements
- exploring ideas of having a collection of events within a data record
- Multi-threaded, parallel processing in active development for simulation and signal processing with the WireCell Toolkit
- great deal of interest and progress made in image processing algorithms applied to Liquid Argon TPC data - ability to utilize those tools is important part of DUNE future frameworks

FD Beam Processing Framework

- 2D deconvolution of signals, hit finding, ROI finding now parallel algorithms
- Work to make signal processing optimized on GPUs/Accelerators is already underway
- ability to have a trigger record input, but multiple events output seems different than collider event data models

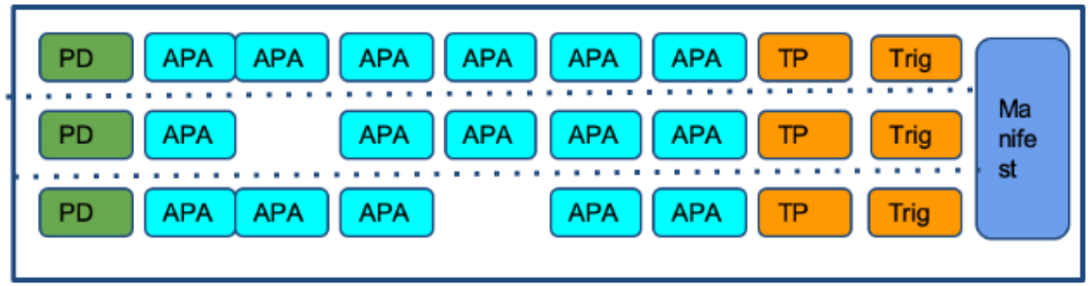
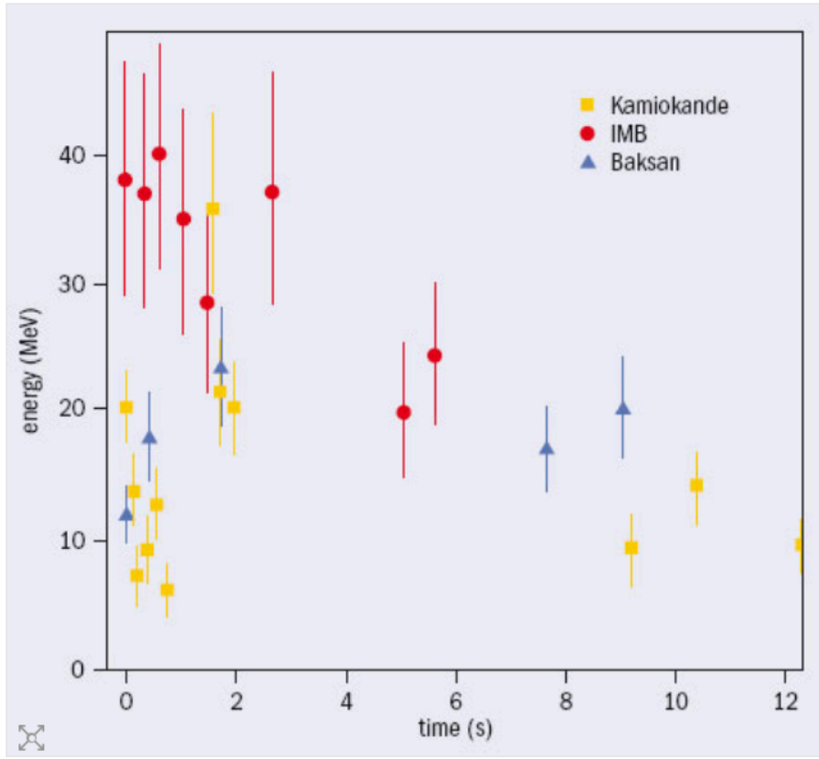


Figure 1: A file (dark box) containing multiple trigger records (separated by dashed lines). Not all APA's are read out in each trigger record. The smaller boxes represent "cells" of data with defined time and spatial extent.

- energy deposition can be separated by > 50 meters and causally separated
- flexibility to adapt to presence or absence of detector elements
- exploring ideas of having a collection of events within a data record
- Multi-threaded, parallel processing in active development for simulation and signal processing with the WireCell Toolkit
- great deal of interest and progress made in image processing algorithms applied to Liquid Argon TPC data - ability to utilize those tools is important part of DUNE future frameworks

Extended Readout Processing

- DUNE plans a continuous readout trigger stream
- nominal DAQ design is for 100 seconds of continuous readout for every supernova trigger record
- estimates are that extended readout trigger record volume will be 100 TB/module
- need to stitch together different time windows for 2D deconvolution of signals and create events
- potentially stitch together data from different files
- unlikely that SN neutrino interactions will need to be processed collectively in reconstruction



SN1987A neutrino events observed by Kamiokande, IMB and Baksan showed that the neutrino burst lasted about 13 s.

CERN Courier, Jan 2007

Fermilab

Extended Readout Processing

- DUNE plans a continuous readout trigger stream
- nominal DAQ design is for 100 seconds of continuous readout for every supernova trigger record
- estimates are that extended readout trigger record volume will be 100 TB/module
- need to stitch together different time windows for 2D deconvolution of signals and create events
- potentially stitch together data from different files
- unlikely that SN neutrino interactions will need to be processed collectively in reconstruction

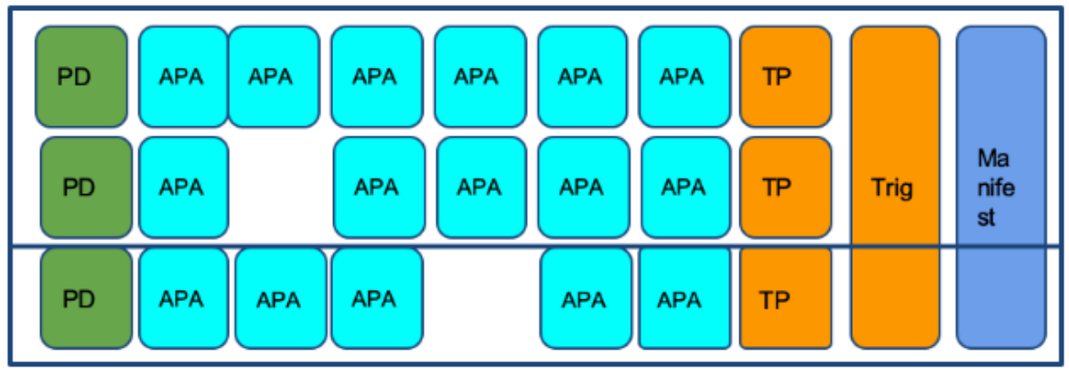
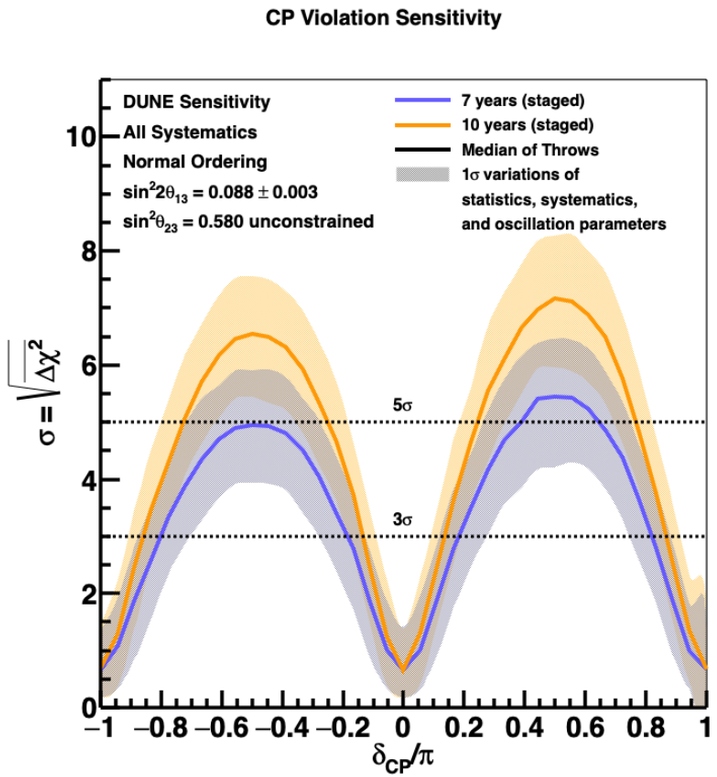


Figure 2: A multi-file aggregate corresponding to a single longer extended trigger record. Files (solid boxes) are opened and closed as output of the trigger record continues.

Fermilab

Summary

- DUNE faces a data reduction challenge in terms of raw trigger records to event reconstruction
- creation of collection of events from a trigger record poses possible challenges to framework tracking and processing
- extended readout trigger records are completely different compared with any collider offline workflow that I know of
- Past experience - overlay of zero-bias trigger data on simulated neutrino interactions has significant technical requirements
- working together with the Fermilab Future Framework Initiative to define requirements for offline workflows
- not a future problem - ProtoDUNE Run II is < 2 years away
- don’t want to handcuff ourselves from utilizing new tools and architectures developed before 2026



Fermilab

Summary

- DUNE faces a data reduction challenge in terms of raw trigger records to event reconstruction
- creation of collection of events from a trigger record poses possible challenges to framework tracking and processing
- extended readout trigger records are completely different compared with any collider offline workflow that I know of
- Past experience - overlay of zero-bias trigger data on simulated neutrino interactions has significant technical requirements
- working together with the Fermilab Future Framework Initiative to define requirements for offline workflows
- not a future problem - ProtoDUNE Run II is < 2 years away
- don't want to handcuff ourselves from utilizing new tools and architectures developed before 2026