DUNE Computing Overview and Perspective on HEP-CCE

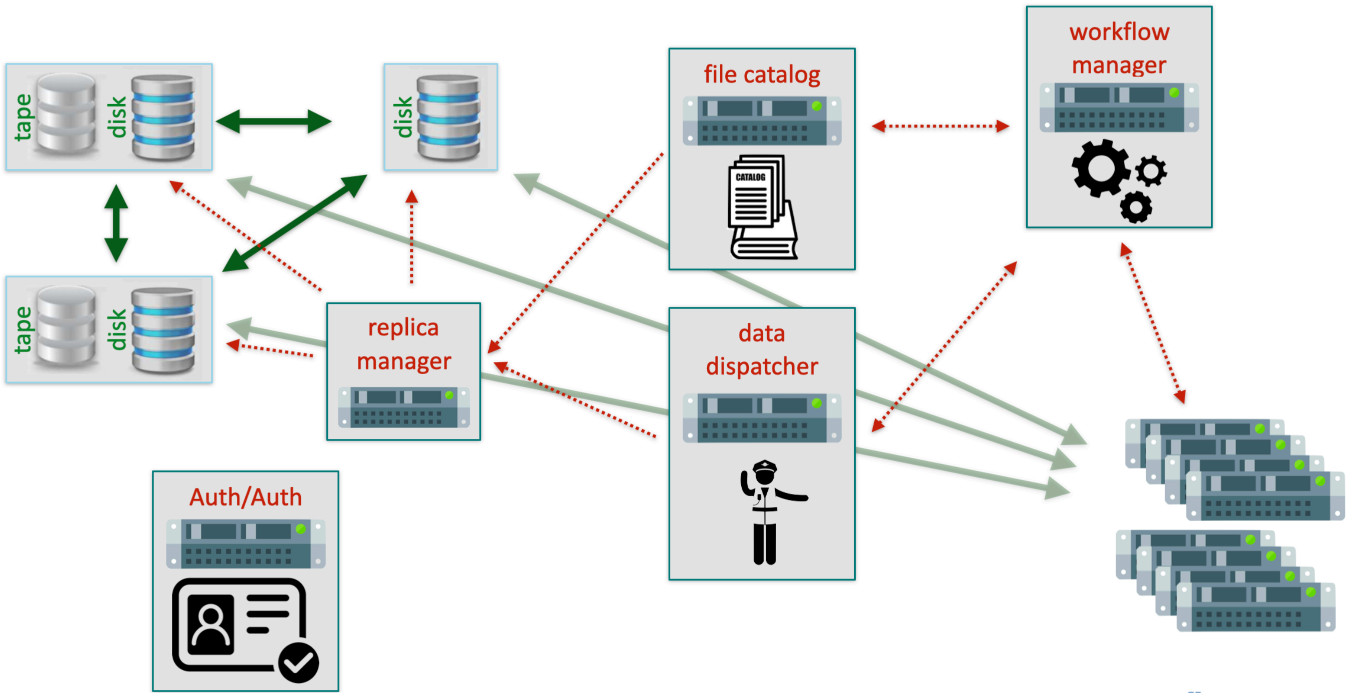
Mike Kirby

HEP-CCE All Hands Meeting

April 20 2022

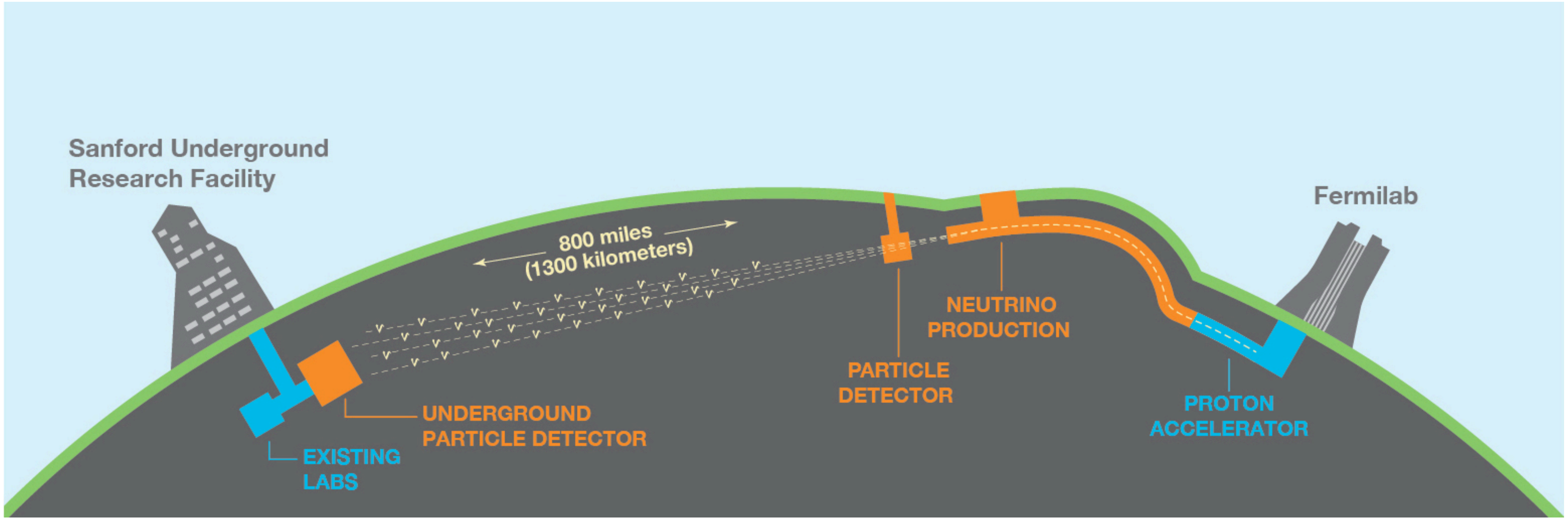
Outline

- quick introduction to DUNE and experiment timeline
- overview of the raw data format, data volumes, and projections
- overview of known use cases and processing models
- some of the unique aspects of DUNE computing requirements
- future plans with HEP-CCE
- summary



DUNE Distributed Computing

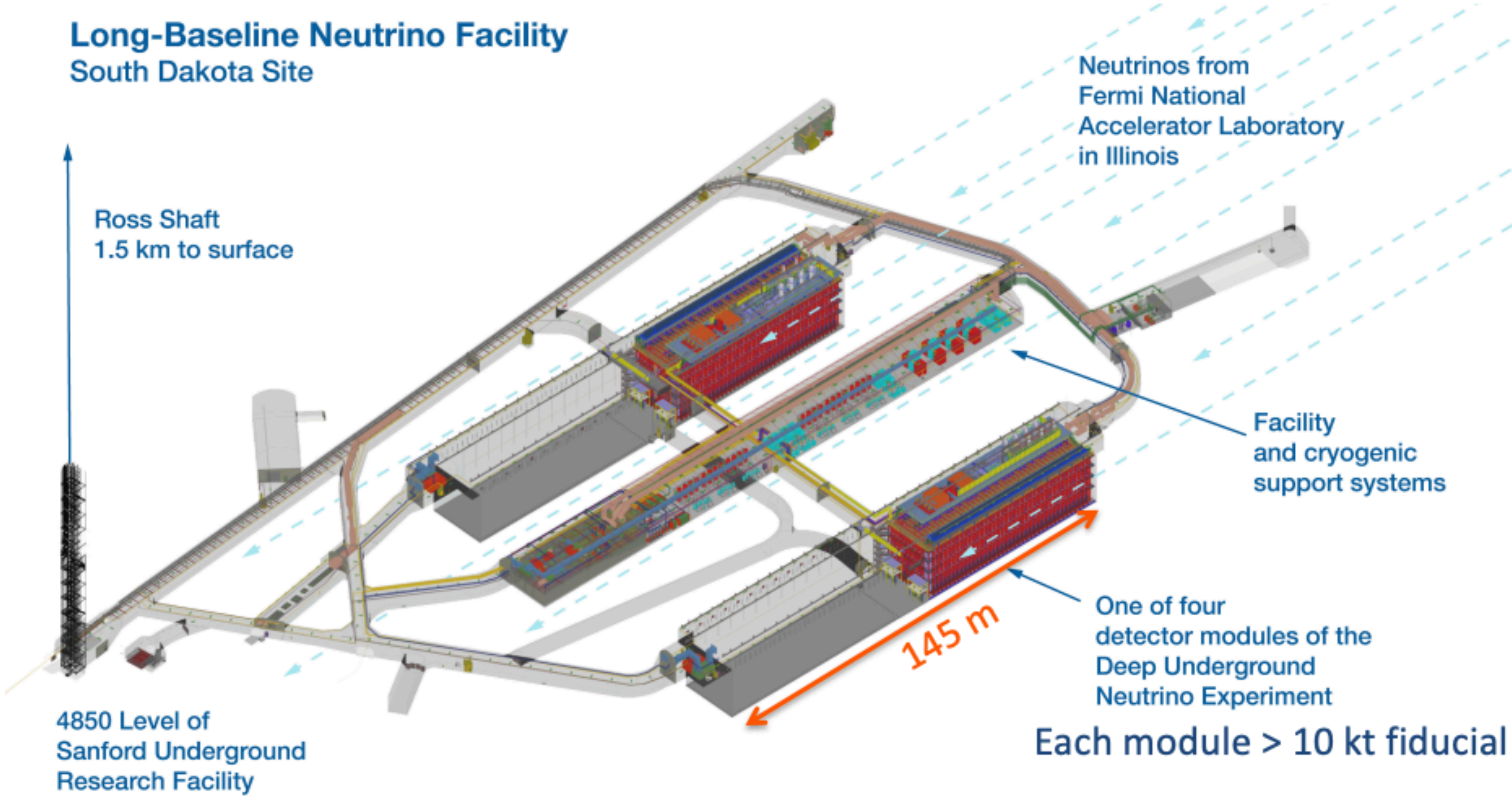
Quick reminder about DUNE



- 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 in design (3 sub-detectors, two that move)at Fermilab near neutrino production
- Two prototypes at CERN - (ProtoDUNE II Horizontal Drift - ProtoDUNE II Vertical Drift)

DUNE Far Detector Design

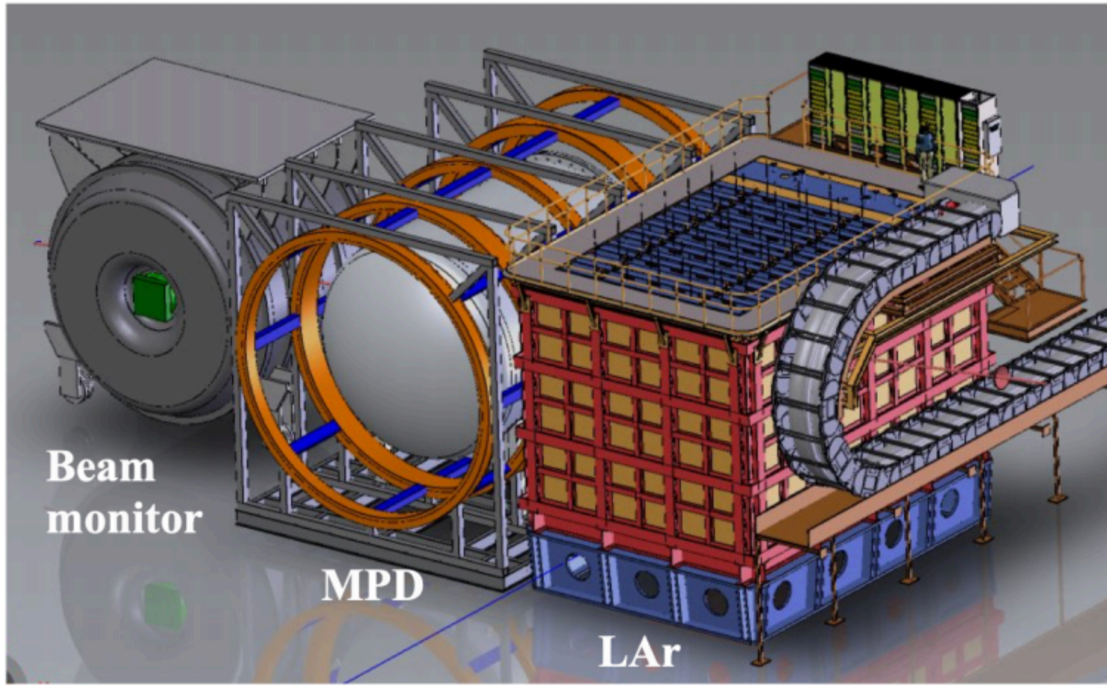
Slide: Ed Blucher

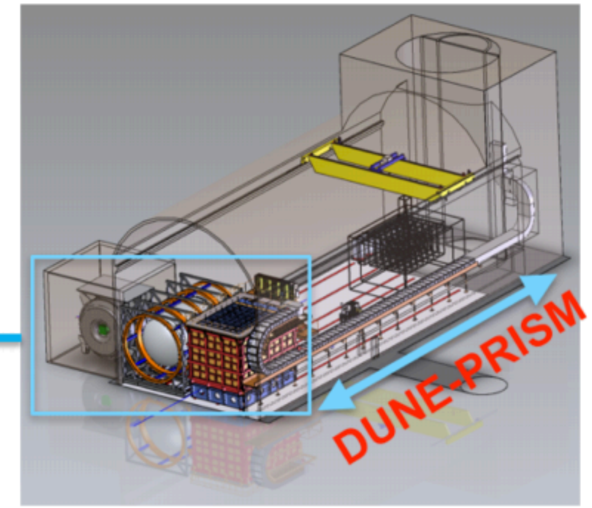


Each module > 10 kt fiducial

(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}

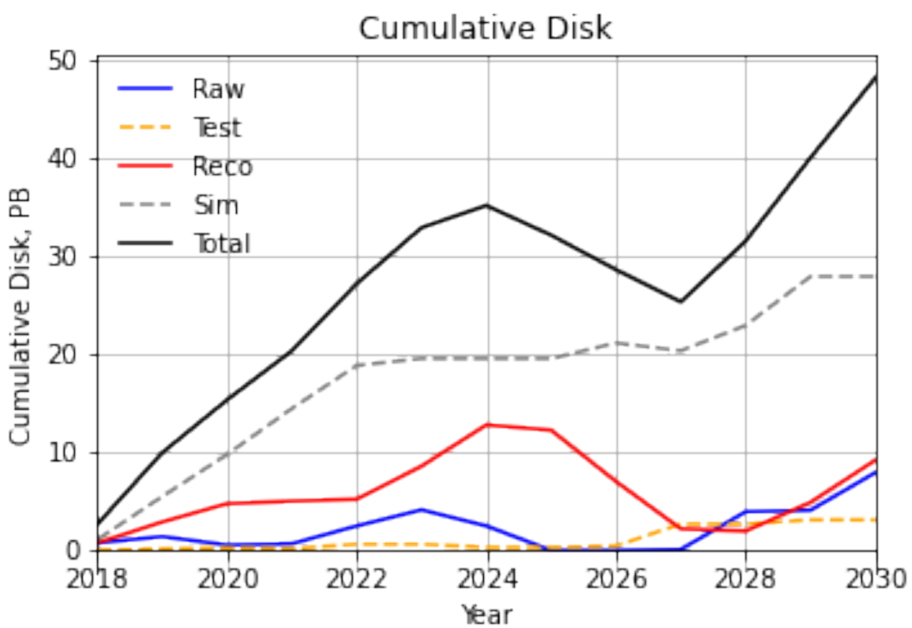
DUNE Introduction and Timeline

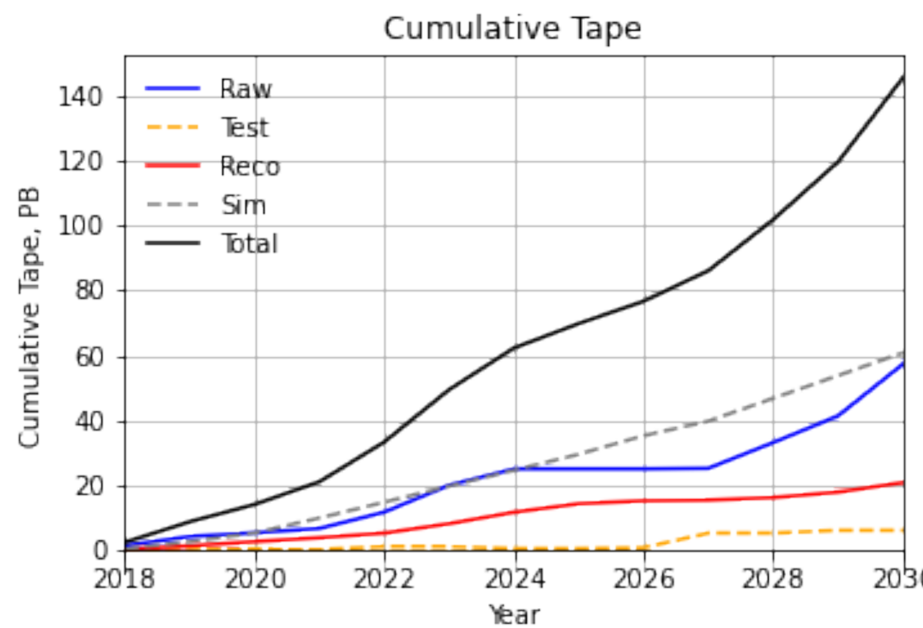
- Staged approach with Far Detector modules, beam, and then Near Detector commissioned in successive years
- raw data from FD dominates tape needs, simulation and derived datasets combined to be similar scale

DUNE FD WAN Bandwidth Timeline Projections:

Date	*Stage of the experiment*	*Primary Path*	*Secondary Path*	*Tertiary Path*
Now	Cavern excavation	10GE	< 1GE via SURF	none
2025	Detector construction	10GE	< 1GE via SURF	none
2027	Computing/DAQ deployment	100GE	10GE	< 1GE via SURF
2028	Cryo deployment completed	100GE	10Gb/s+	10GE
2029	Start of science	100GE	10Gb/s+	10GE

vLAN service provided by REED/GPN (shared)
Dedicated circuit Ross Dry Bldg. to Chicago
Dedicated circuit Yates Complex to Denver (10GE or 100GE)

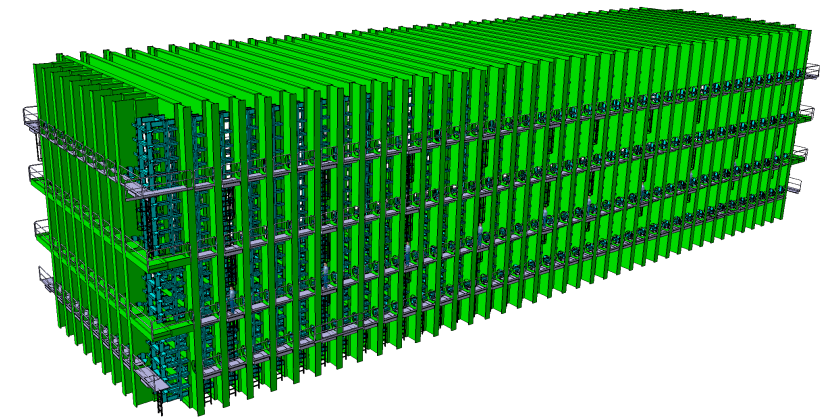




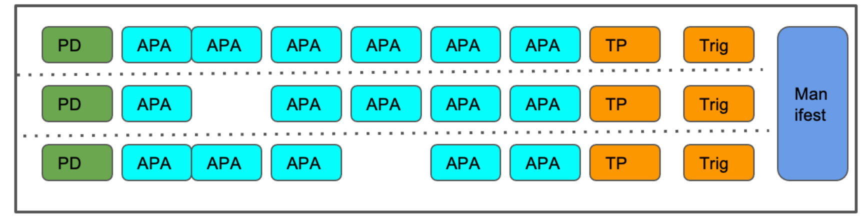
Fermilab

DUNE Far Detector raw data

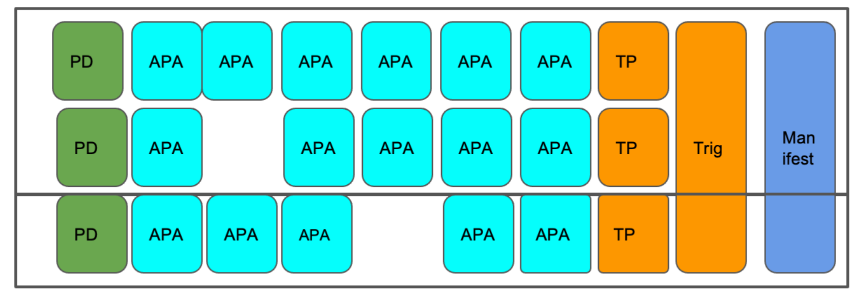
- Far Detector frontend is capable of producing incredible data rates
  - reduction through trigger, zero suppression, and compression of data - all areas of active research
  - DAQ constrained to 30 PB/year from far detector
  - beam trigger, calibrations, supernova time-extended trigger records
  - DAQ has chosen HDF5 for the raw data output format for ProtoDUNE II operations
    - simultaneous write from multiple processes is seen as an advantage given distributed nature of FD
    - currently in discussion/design phase of the serialization of raw data for output
    - HDF5 datasets consist of headers + binary fragments defined by FE Readout, but also additional complex objects created by DAQ and trigger



Localized readout aggregate (cosmics/beam)



Extended (SNB) readout aggregate



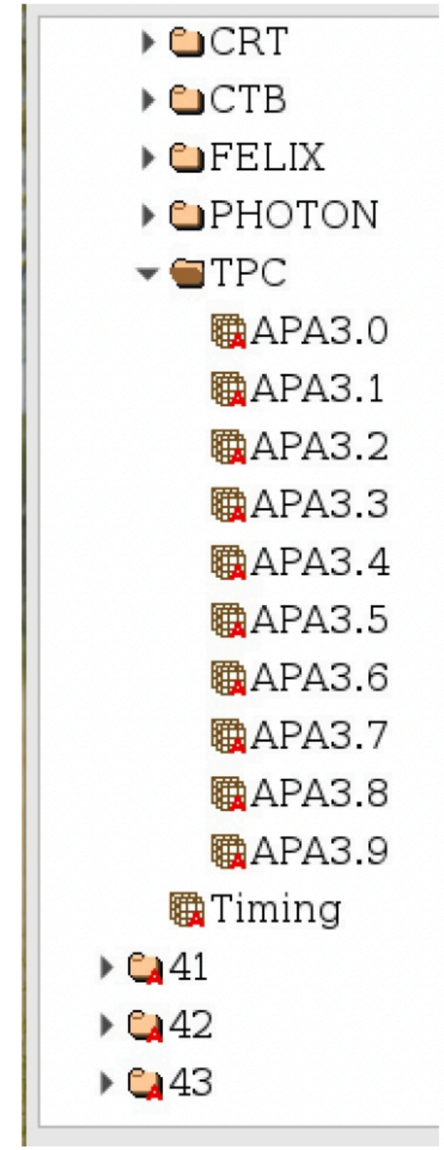
Process	Rate/module	size/instance	size/module/year
Beam event	41/day	3.8 GB	30 TB/year
Cosmic rays	4,500/day	3.8 GB	6.2 PB/year
Supernova trigger	1/month	140 TB	1.7 PB/year
Solar neutrinos	10,000/year	35 TB/year	
Calibrations	2/year	750 TB	1.5 PB/year
Total			9.4 PB/year

Fermilab

Michael Kirby, HEP-CCE AHM 2022

Discussion of IOS needs and HDF5

- Current DAQ for ProtoDUNE II (and designed/anticipated for DUNE) outputs HDF5 raw data files
  - choice of HDF5 was made to address data packaging requirements and DAQ design issues
  - ‘chunking’ or ‘slicing’ or ‘deciding which pieces of a Trigger Record or Stream go into which file’)
  - with distributed FD DAQ ability to have multiple threads writing in parallel to single file was very beneficial
  - ability of configurable packaging based on Trigger or Stream
- Active discussions about data format within HDF5
  - current implementation based on C++ standard vs serialization package vs HDF5 compound datasets vs removing padding vs ???
  - understanding performance benefits are definitely of interest
  - support model is important given lifetime



Offline workflow use cases

- Currently working on CDR for DUNE Computing
- includes an extensive list of use cases that are helping define our resource needs and work packages for the coming years
- varied across physics, production, analysis, event classification, ML/AI, simulation, etc
- in the world of HTC, we are not yet drastically resource limited

Table 3.3: Summary of resources needed per file for compute intensive tasks. The total CPU needed for a task can be determined from the total size of the sample and these numbers.

Use case	memory	input file size	output file size	CPU time	input	cores/job
units	GB	MB	MB	sec	MB/s	
Simulation+reco	6	100	2000	27000	0.00	1
data reco	4	8000	4000	60000	0.13	1
tuple creation	3	4000	1	100	40.00	1
calibration	3	4000	1	100	40.00	1

i

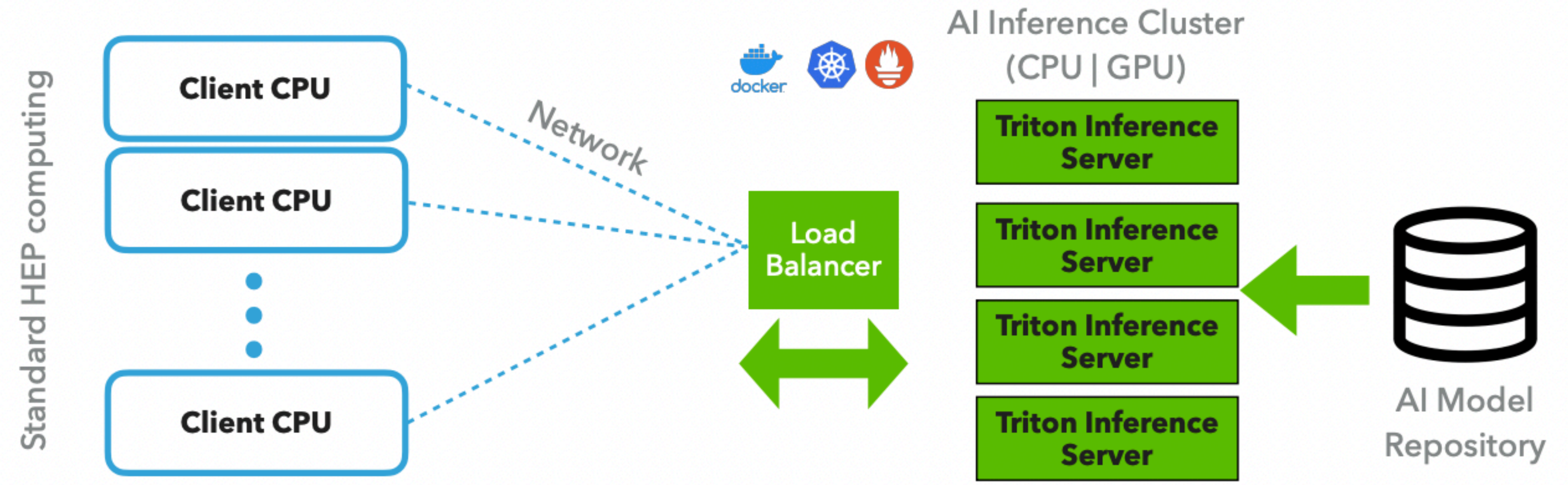
What are the tasks where we need to utilize non-CPU architectures to enables out physics goals?

Discussion of parallelism and portability

- Important to understand where DUNE stands right now
  - offline processing is based upon dunetpc -> larsoft -> art
  - implementations of external reco packages (WireCell, Pandora, Deep Learning/AI, etc.)
  - currently in a single event processing model that is not threadsafe
- Impressive amount of work on WireCell implementation within Kokkos
  - there had previously been work on CUDA implementation
  - making WireCell multithreaded likely hampered by framework that it is wrapped inside of for DUNE
- DUNE is not currently resource limited for processing on new architectures
- we are resource limited in transitioning to a multithreaded paradigm (years behind CMS)
- for HEP-CCE, we anticipate adding a new effort more directly possible to evaluate portability

Discussion of parallelism in particle ID (MichelEMID and GPUaaS)

- There has been work in AI/ ML/image recognition algorithms within the LArTPC community
- development (training) has mostly been done on dedicated hardware with inference on CPU
- GPUaaS research shows impressive results
- but not part of mainstream processing and resource limits are network not GPUs

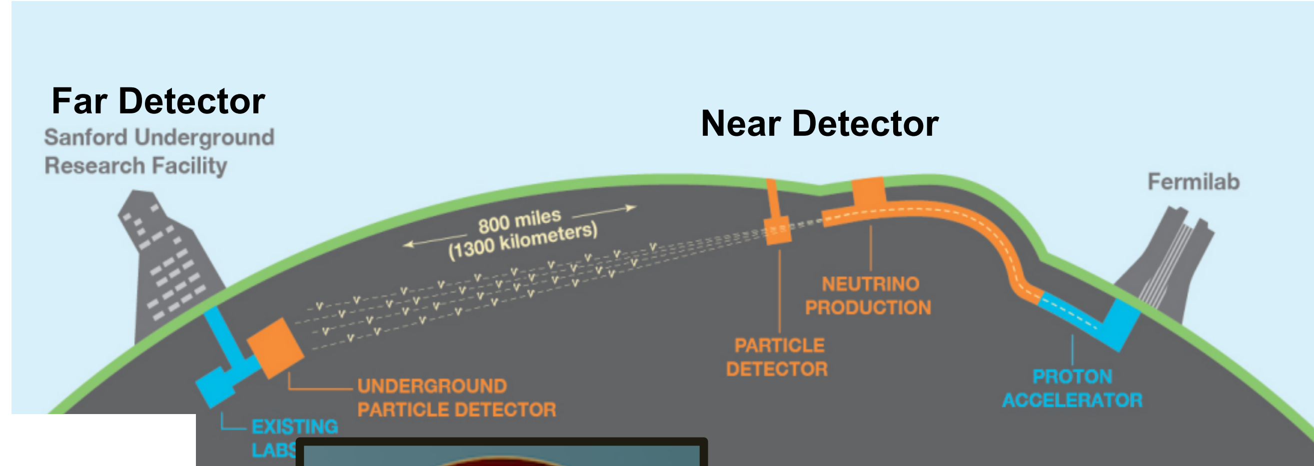


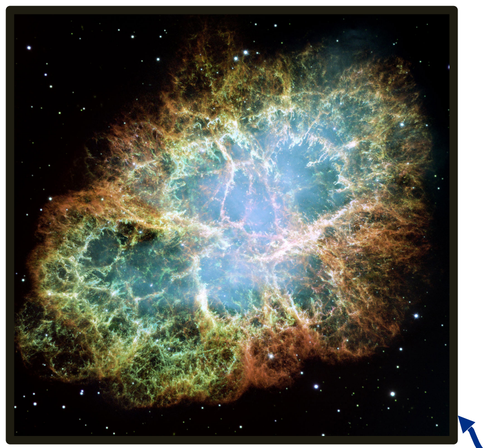
Wang, et al., FERMILAB-PUB-20-428-ND-SCD

	CPU time/event	SONIC
Non-ML module	110s	110s
ML module (EmTrkMichelId)	 220s	~12s (~2.7s on GPU)
Total	330s	122s

DUNE - Future Flagship Experiment @ Fermilab

- **World's largest multi kiloton LArTPC detector**
- **One of the major physics goals**
  - **study rare (off-beam) events at Far Detector**





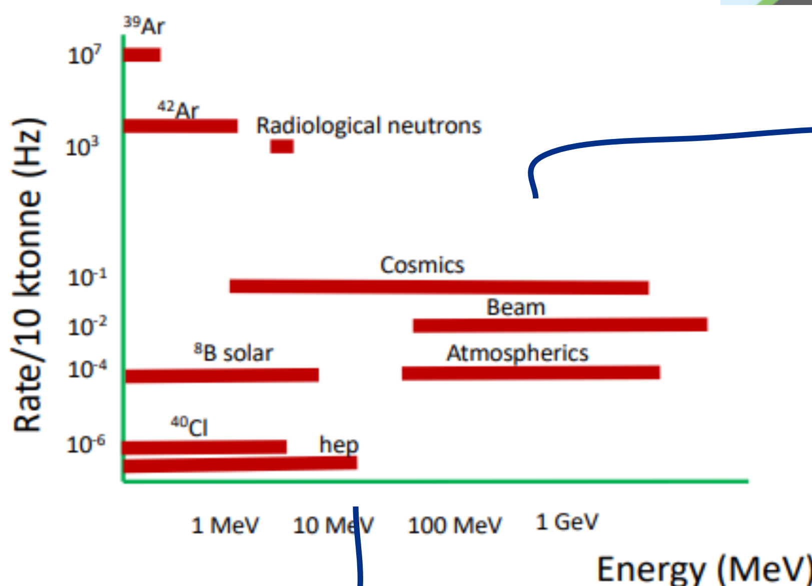
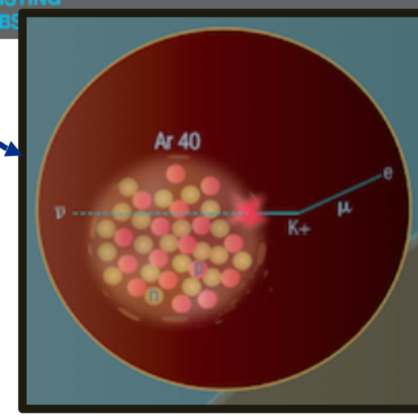


Figure 7.2: Expected physics-related activity rates in a single 10 kt module.

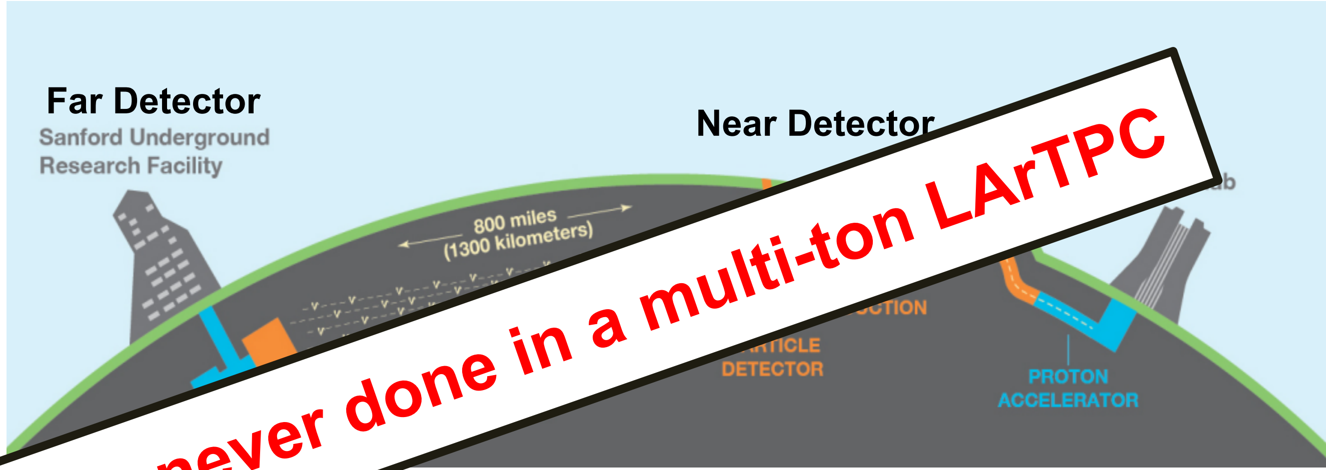


Expected data rate ~1.15 TB/s/ 10kT

Requirement : efficient and continuous data processing

DUNE - Future Flagship Experiment @ Fermilab

- **World's largest multi kiloton LArTPC detector**
- **One of the major physics goals**
  - **study rare (off-beam) events at Far Detector**
- **Requirement : efficient and continuous data processing**



Implement a TPC data-driven trigger system - never done in a multi-ton LArTPC

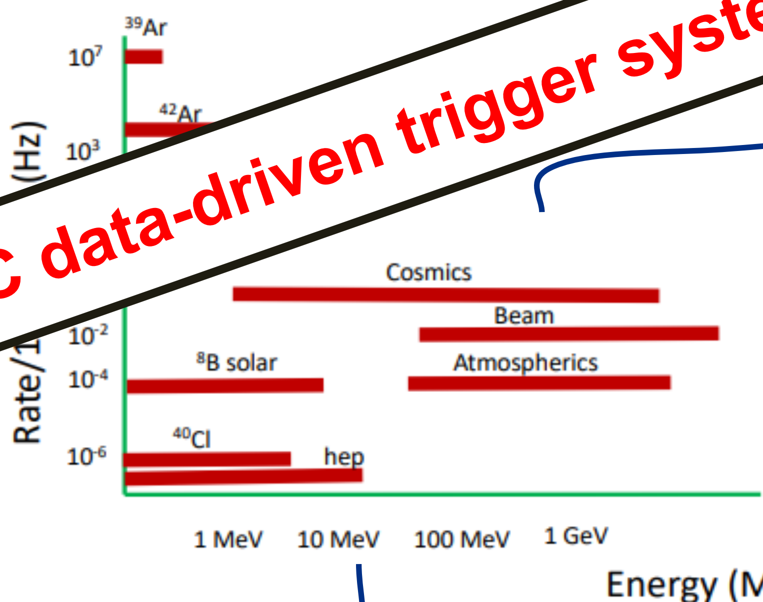
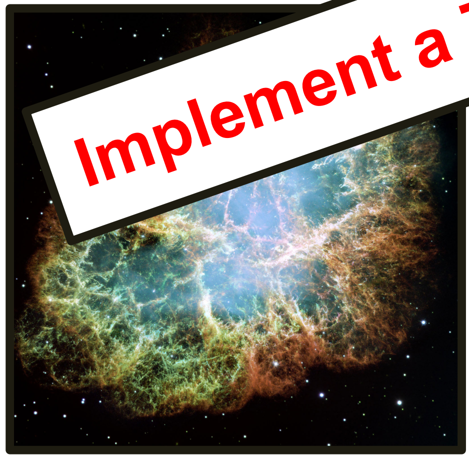
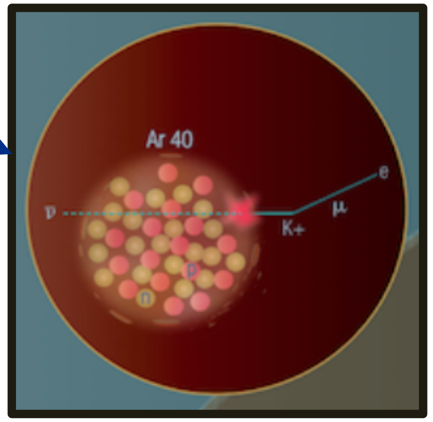
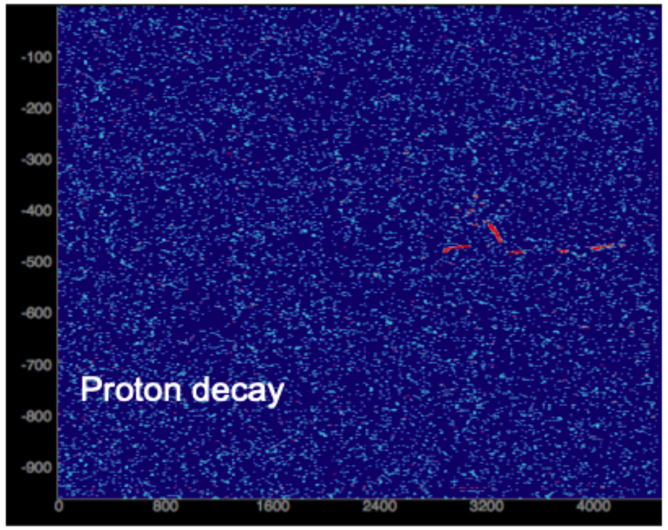


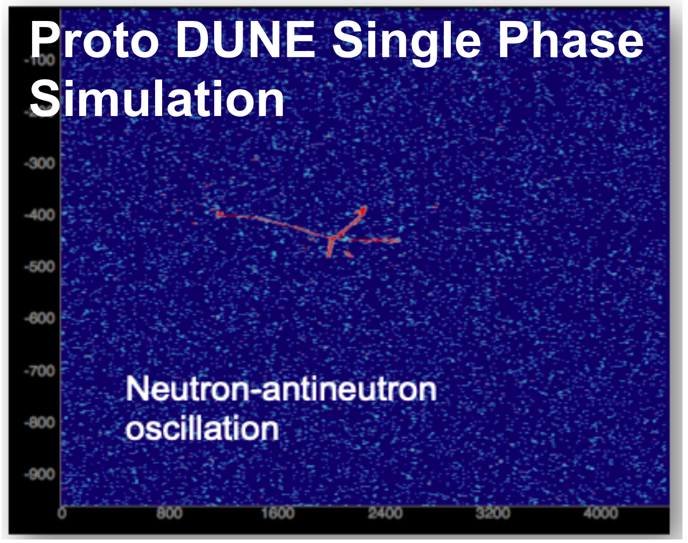
Figure 7.2: Expected physics-related activity rates in a single 10 kt module.

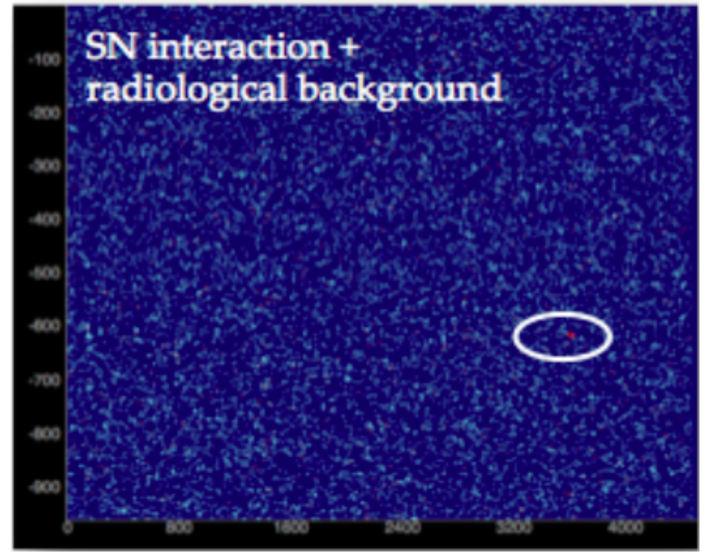


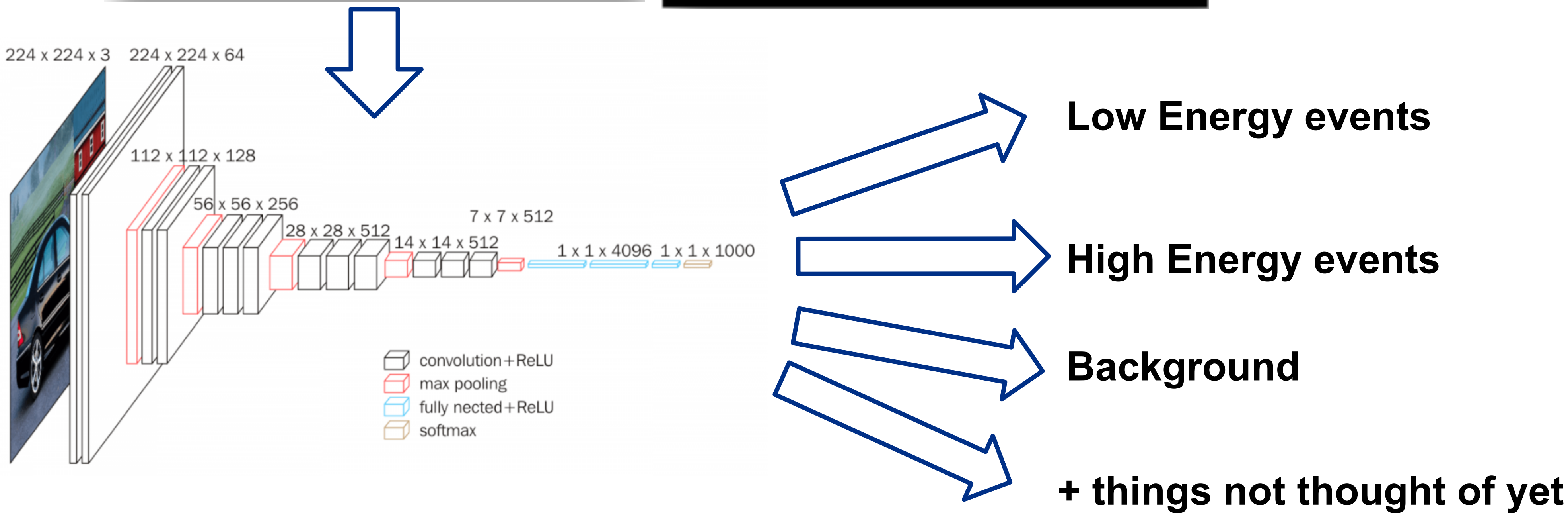


Machine Learning Based Trigger Algorithm









Fermilab

Potential Networks

Submanifold Sparse CN

- LArTPC images are sparse - except for the region near tł interaction vertex

- Employ a sparse network as opposed to dense CNN -

  - Pros :
    - Significant reduction in time and memory consumptic (tests done in MicroBooNE)-

      - Time for processing a 512 X 512 image - ~ 5s → ~ 0.5s (Intel core i7-8750H CPU 2.2 GHz)
      - Memory usage - ~ 5GB → ~ 1GB
  - Cons :
    - Shower accuracy falls significantly from test samples while using full BNB
    - PyTorch not Tensorflow

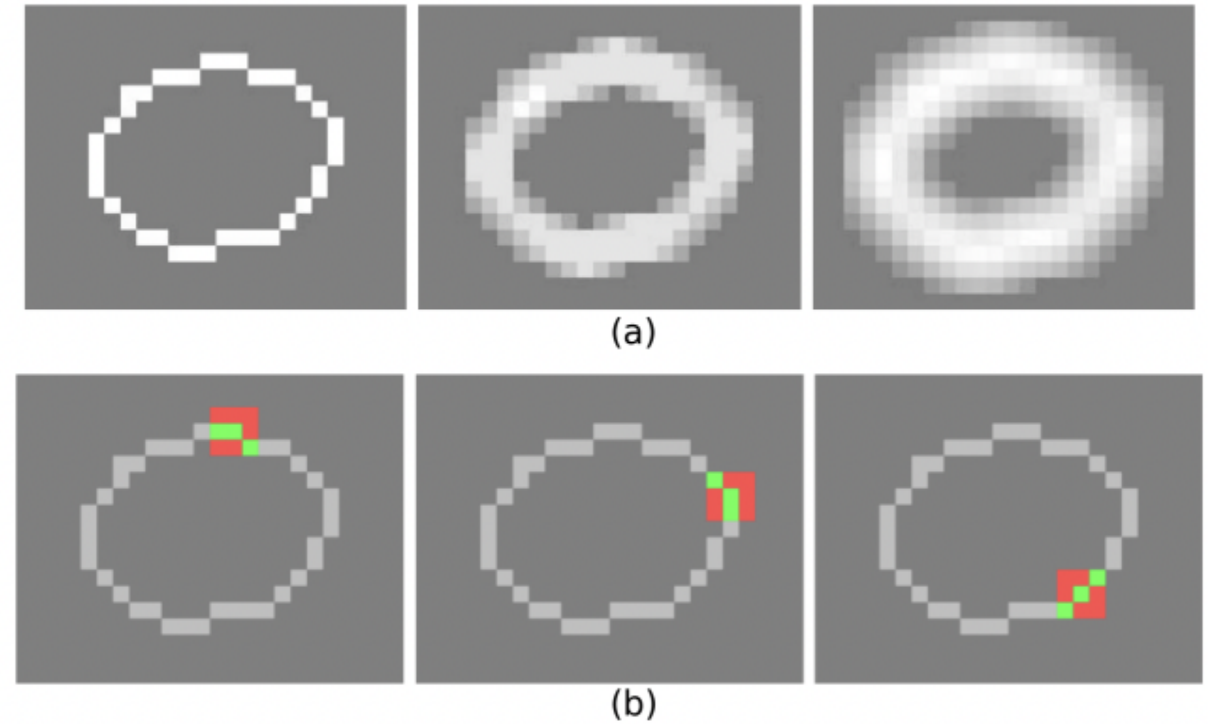
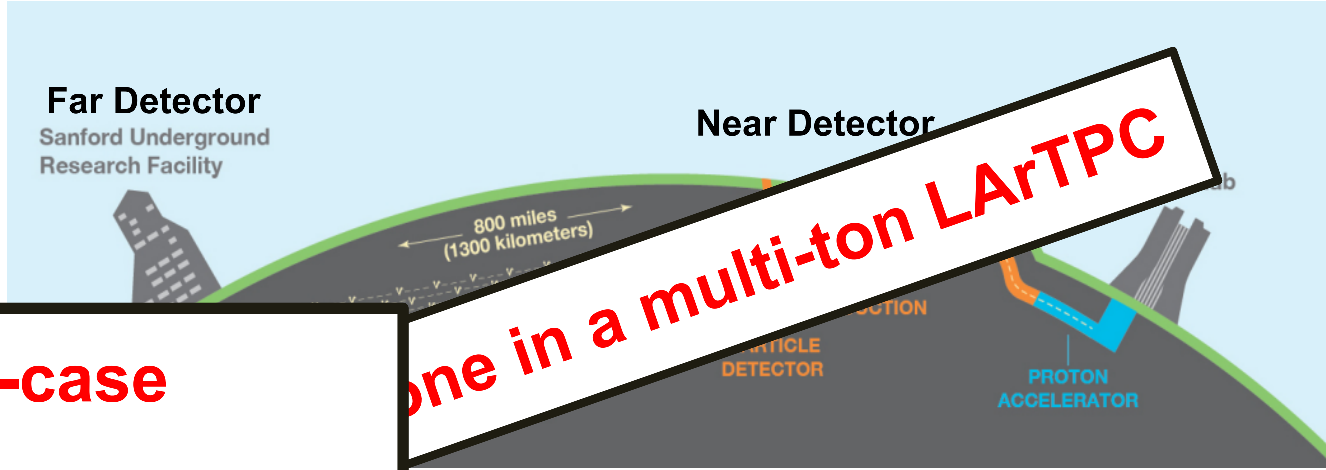


FIG. 1: (a) an example of an image being dilated after two dense convolution operations using a filter with size 3×3, weights $1/9$, and stride 1. (b) a non-dilated image using sparse convolutional layers, the green label represents pixels that are kept for consecutive layers and the red label represents pixels that would have acquired values in dense CNNs, but do not in SparseSSNet (image taken from [17]).

DUNE - Future Flagship Experiment @ Fermilab

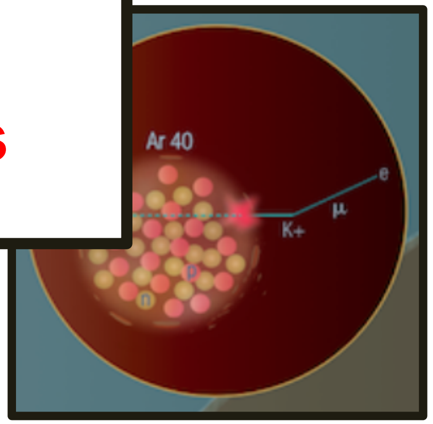
- **World's largest multi kiloton LArTPC detector**
- **One of the major physics goals**
  - **study rare (off-beam) events at Far Detector**
- **Requirement : eff**
  processing



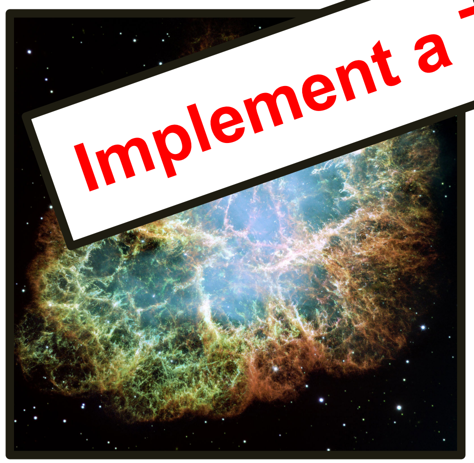
ne in a multi-ton LArTPC

Great candidate as a use-case

- **(intended to be) simple**
- **Less number of dependencies**



Implement a TPC



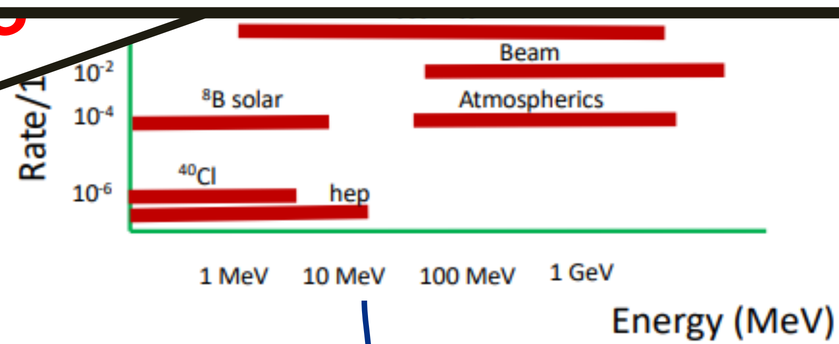


Figure 7.2: Expected physics-related activity rates in a single 10 kt module.

Summary

- HEP-CCE IOS research
  - DUNE data format/IOS does have some interesting questions that could use assistance form HEP-CCE
  - need to understand performance of HDF5 and serialization
  - streaming of HDF5 files appears to be a feature required by our workflow
  - current production workflow is dependent upon streaming files from RSEs
- HEP-CCE PPS research
  - production processing need effort and time to transition code and workflows to threadsafe/multithreaded
  - assistance on how to accomplish that may not be well suited for HEP-CCE
  - at the same time, some inference scenarios do lend themselves to understanding portability and performance
- DUNE is still 6 years from taking Far Detector physics data
- choices for ProtoDUNE II has been made but provide opportunity for R&D with IOS
- future AI LArTPC trigger development is interesting use case for PPS