Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

Computing Challenges and Advancements in HEP

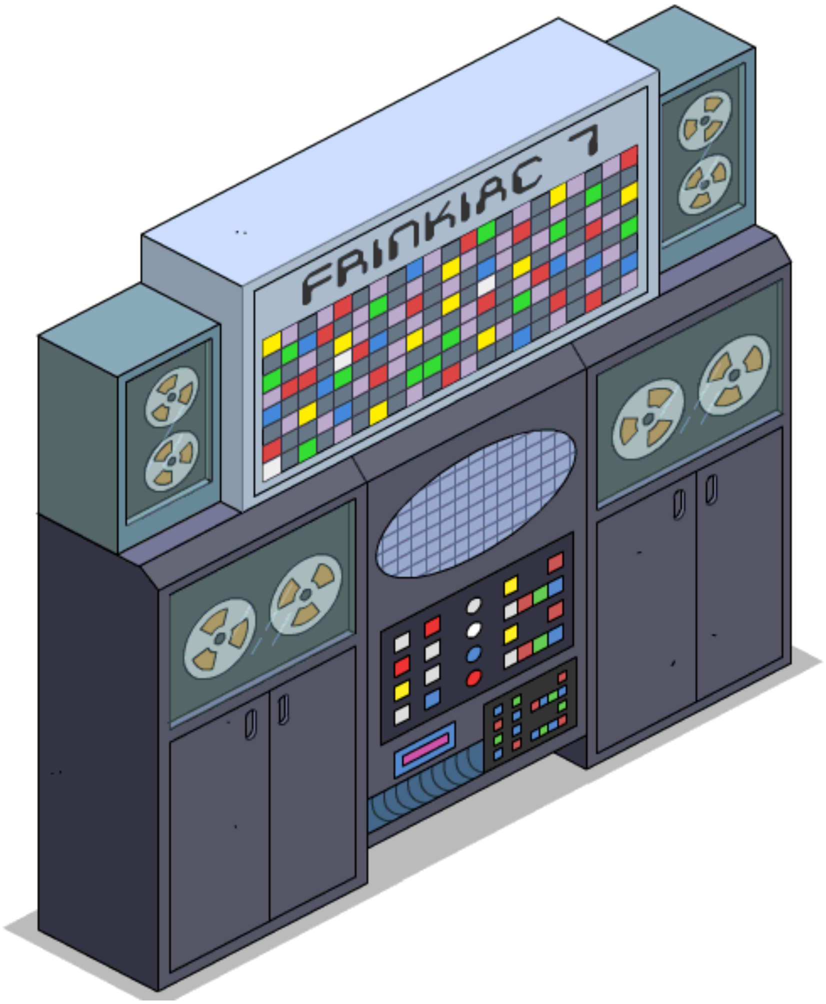
Michael Kirby (BNL)

25th International Workshop on Neutrinos from Accelerators

Argonne National Laboratory

September 17, 2024

Some Computing Challenges in HEP



- topic of extremely broad scope and significant impact
- optimize all aspects of computing - algorithms, storage, and resource utilization
- development of algorithms with awareness of architecture is an important aspect
- focus on some of the common challenges faced across the field
- a few unique to neutrino physics
- my personal biases are in collider and US neutrino physics

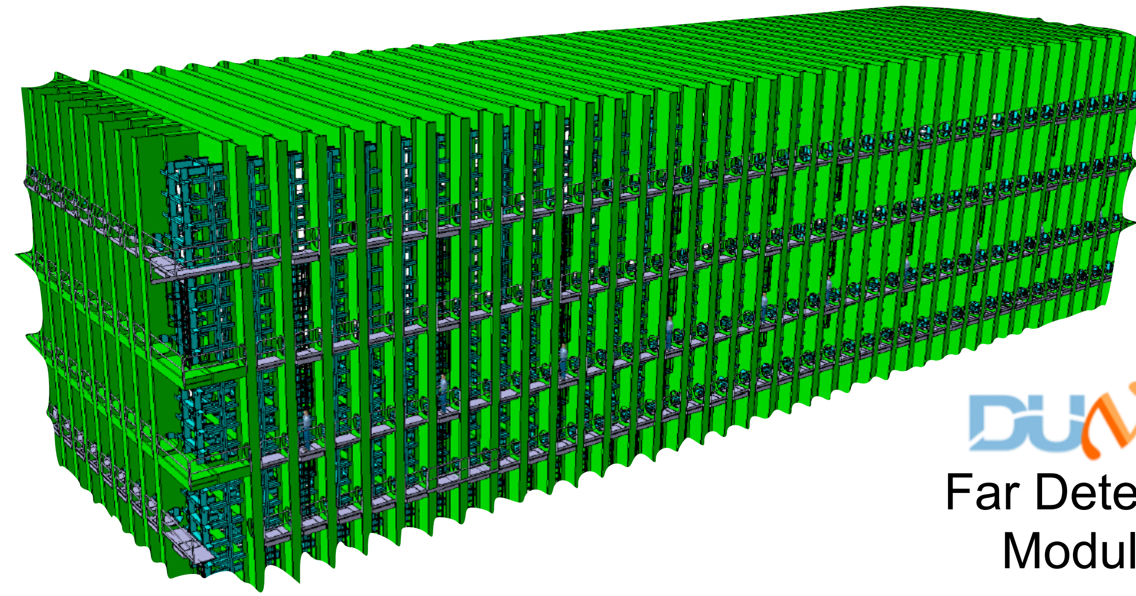
Firehose of high-resolution data

due to increased capabilities of detectors and accelerators incredible amounts of data

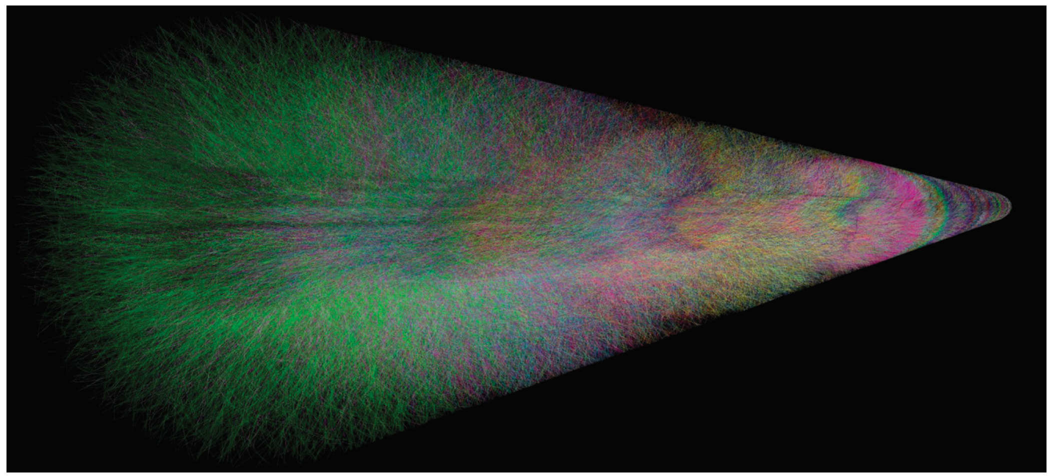
vastly different detector designs and interactions rates, but result is the same

LAr TPC detectors produce large data records based upon drift time and mm resolution

LHC experiments high event rates even with complex trigger algorithms



DUNE
Far Detector Module



2 ms time frame of ALICE Pb-Pb data

Brookhaven National Laboratory

High Luminosity LHC Data Volumes

After Long Shutdown 3, LHC Run 4 luminosity will take a significant jump

$2x10^{34}\ s^{-1}cm^{-2} \rightarrow 7.5x10^{34}\ s^{-1}cm^{-2}$

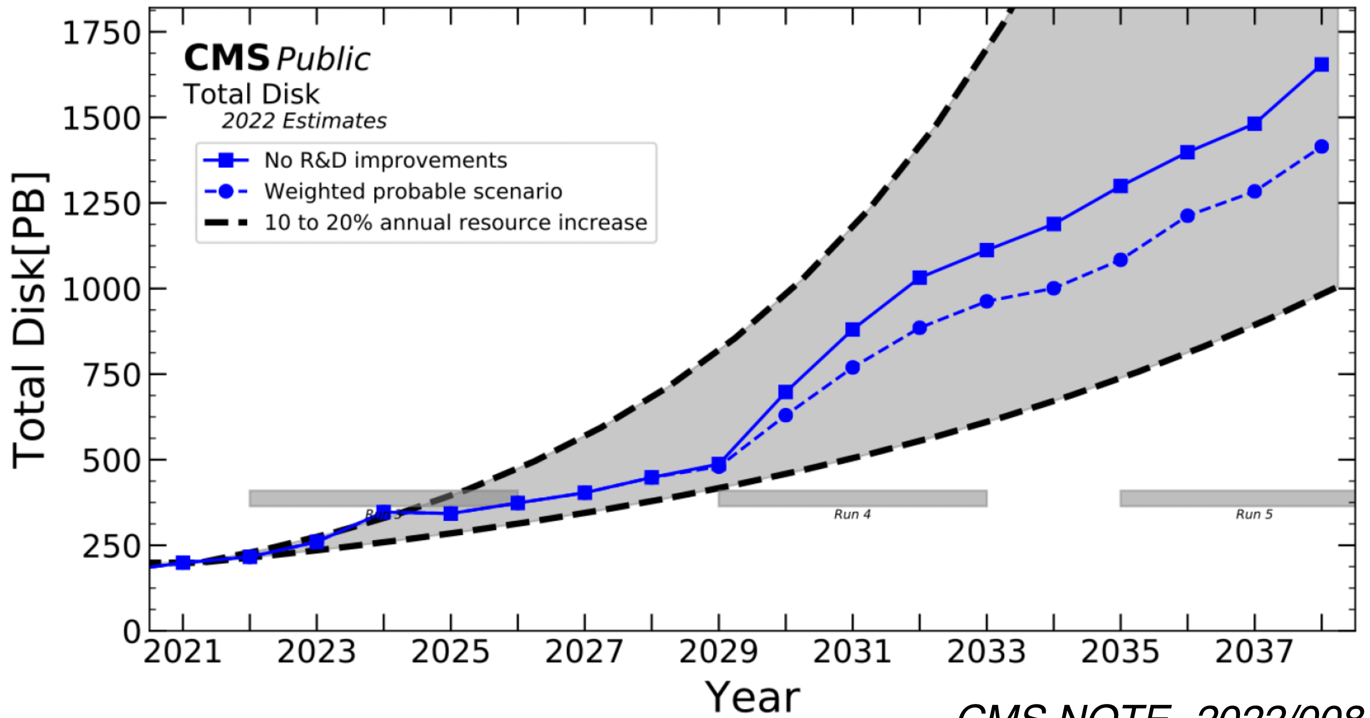
Detector upgrades produce finer granularity

Projections show that data taken in Run 4 presents significant challenge

improved triggering and compression are essential to reducing data size

exploring data reduction through use of RNtuple

development of data carousels - rotating data off of tape onto disk



CMS NOTE -2022/008

High Luminosity LHC Data Volumes

After Long Shutdown 3, LHC Run 4 luminosity will take a significant jump

$2 \times 10^{34}\ s^{-1} cm^{-2} \rightarrow 7.5 \times 10^{34}\ s^{-1} cm^{-2}$

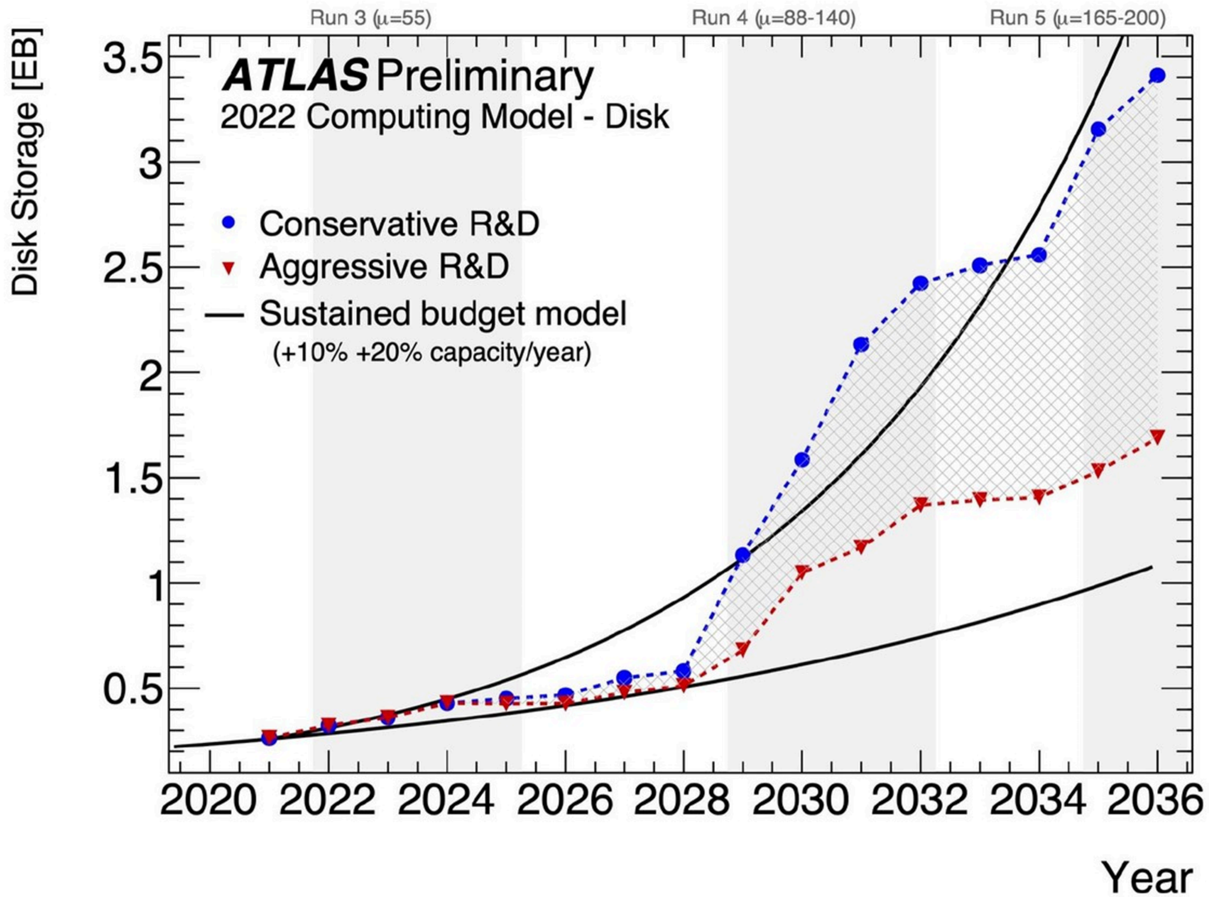
Detector upgrades produce finer granularity

Projections show that data taken in Run 4 presents significant challenge

improved triggering and compression are essential to reducing data size

exploring data reduction through use of RNtuple

development of data carousels - rotating data off of tape onto disk

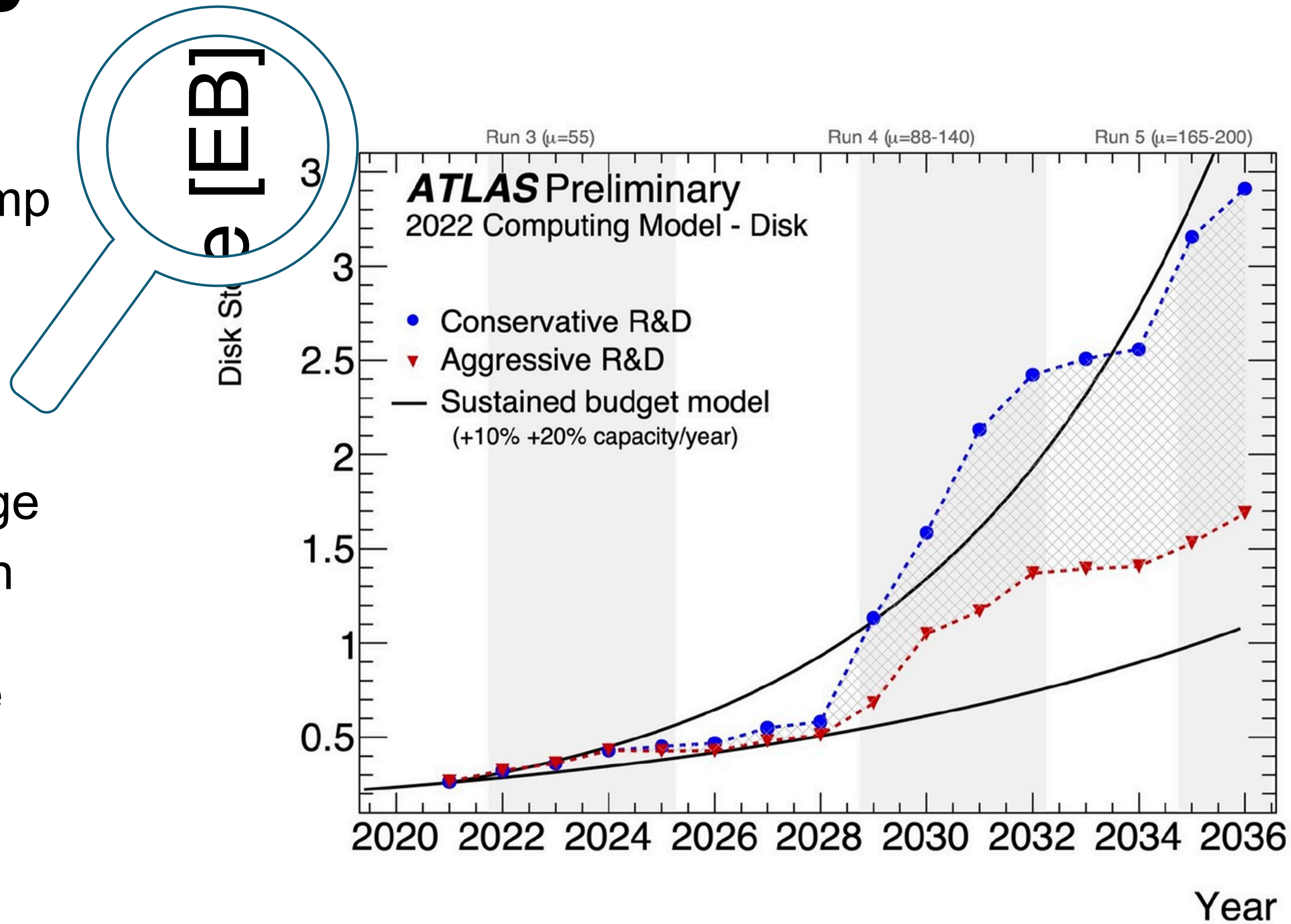


ATLAS Software and Computing HL-LHC Roadmap - CERN-LHCC-2022-005

Brookhaven National Laboratory

High Luminosity LHC Data Volumes

- After Long Shutdown 3, LHC Run 4 luminosity will take a significant jump
  - $2x10^{34}\ s^{-1}cm^{-2} \rightarrow 7.5x10^{34}\ s^{-1}cm^{-2}$
- Detector upgrades produce finer granularity
- Projections show that data taken in Run 4 presents significant challenge
- improved triggering and compression are essential to reducing data size
- exploring data reduction through use of RNtuple
- development of data carousels - rotating data off of tape onto disk



[ATLAS Software and Computing HL-LHC Roadmap](#) - CERN-LHCC-2022-005

Brookhaven National Laboratory

DUNE Data Volumes present different challenge

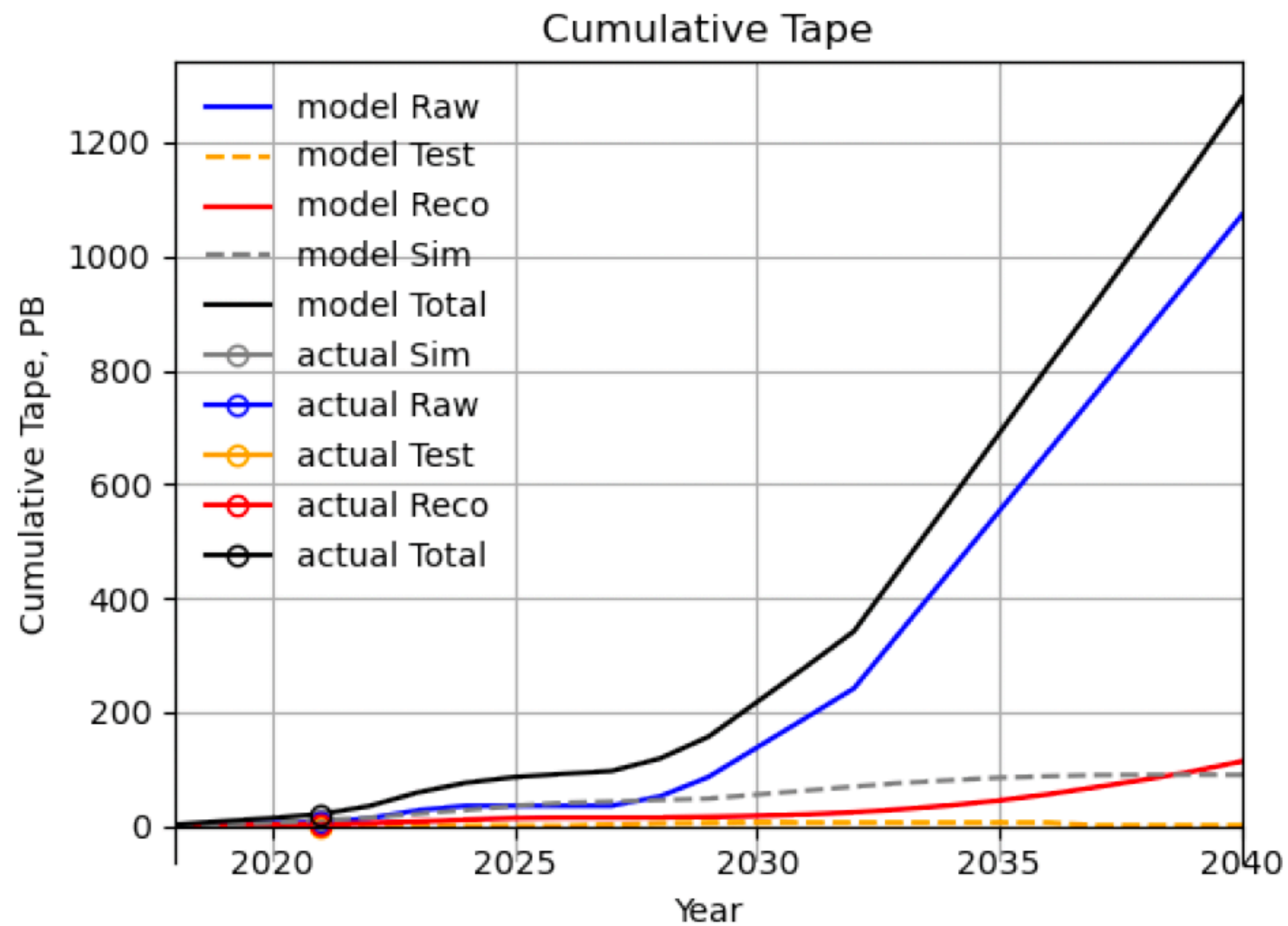
single Far Detector module can produce 8 GB for one drift window

- significant limitations on the computing resources available at the far site for triggering and compression
- potential physics signals extend across at least 5 orders of magnitude in energy scale

large calibration samples spanning the full detector

reprocessing of DUNE raw data in out years considerable challenge - build upon experience of LHC data carousels

raw data this large doesn't fit nicely on a WLCG worker node

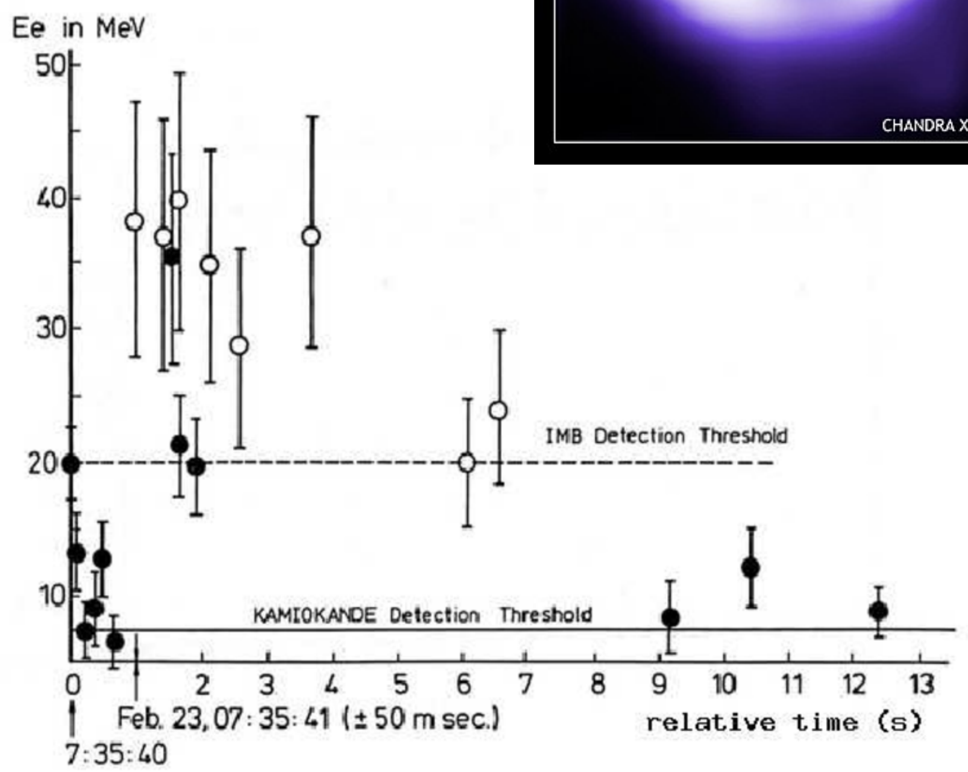


DUNE Computing CDR

DUNE Supernova Readout

- Potentially the most time critical data from DUNE Far Detector
- record data continuously for ~100 seconds
- ~600 TB of data from 4 Far Detector modules
- pointing for optical follow up needs to be ~10 minutes
- extremely limited computing resources in the detector caverns
- need to find creative algorithms/ resources to reconstruct events with pointing information

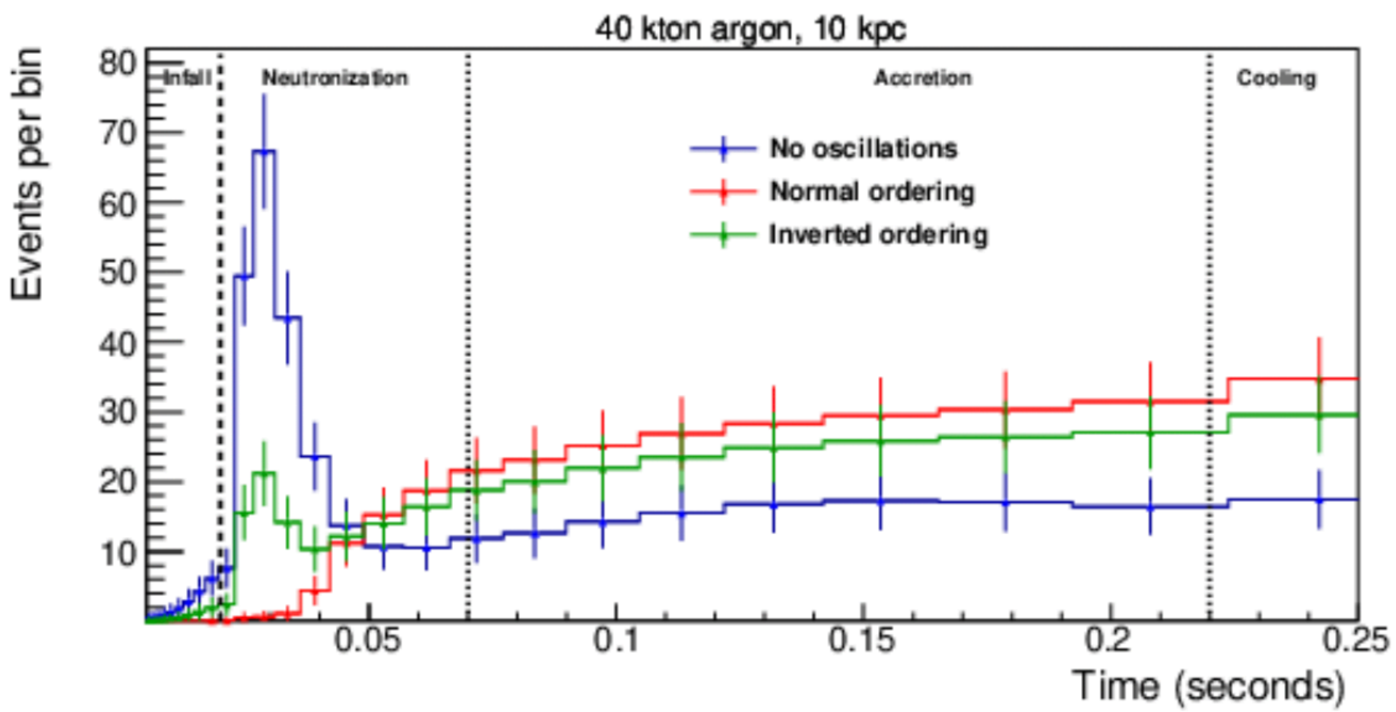
SN1987a



Class.Quant.Grav.27:084019,2010

DUNE Supernova Readout

- Potentially the most time critical data from DUNE Far Detector
- record data continuously for ~100 seconds
- ~600 TB of data from 4 Far Detector modules
- pointing for optical follow up needs to be ~10 minutes
- extremely limited computing resources in the detector caverns
- need to find creative algorithms/ resources to reconstruct events with pointing information

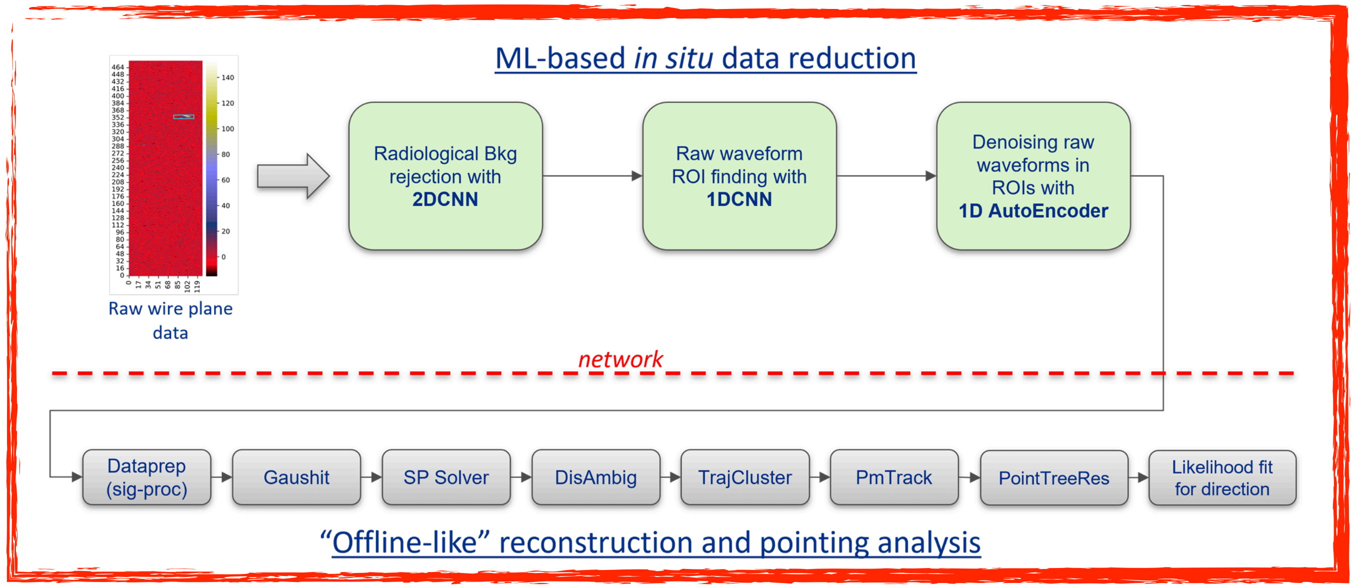


https://arxiv.org/pdf/2008.06647.pdf

Brookhaven National Laboratory

DUNE Supernova Readout

- Potentially the most time critical data from DUNE Far Detector
- record data continuously for ~100 seconds
- ~600 TB of data from 4 Far Detector modules
- pointing for optical follow up needs to be ~10 minutes
- extremely limited computing resources in the detector caverns
- need to find creative algorithms/ resources to reconstruct events with pointing information

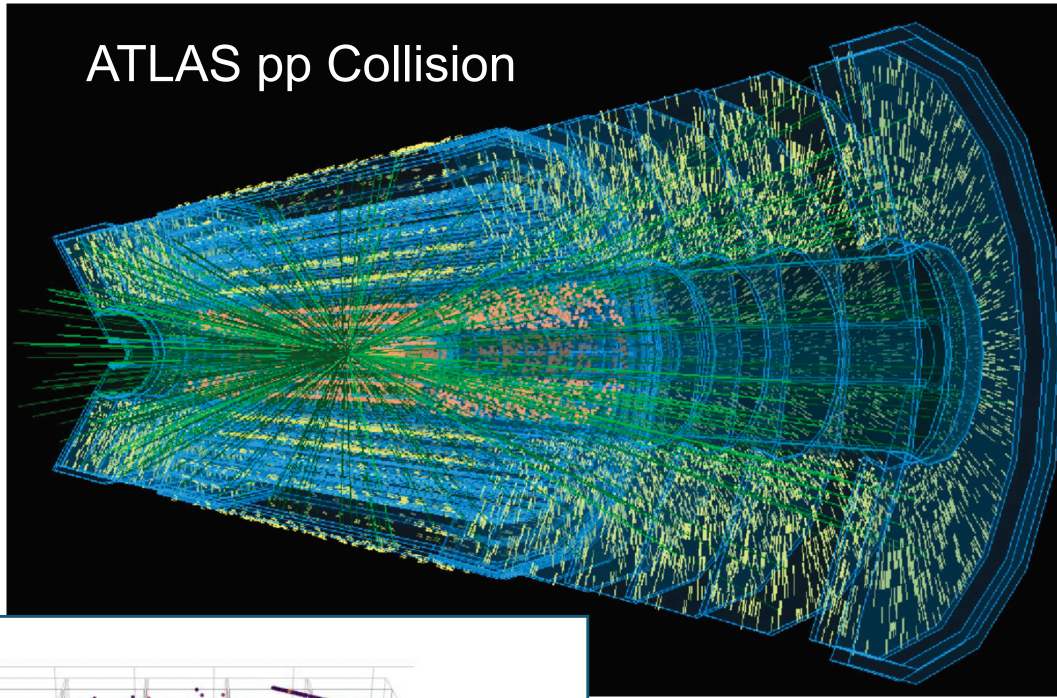


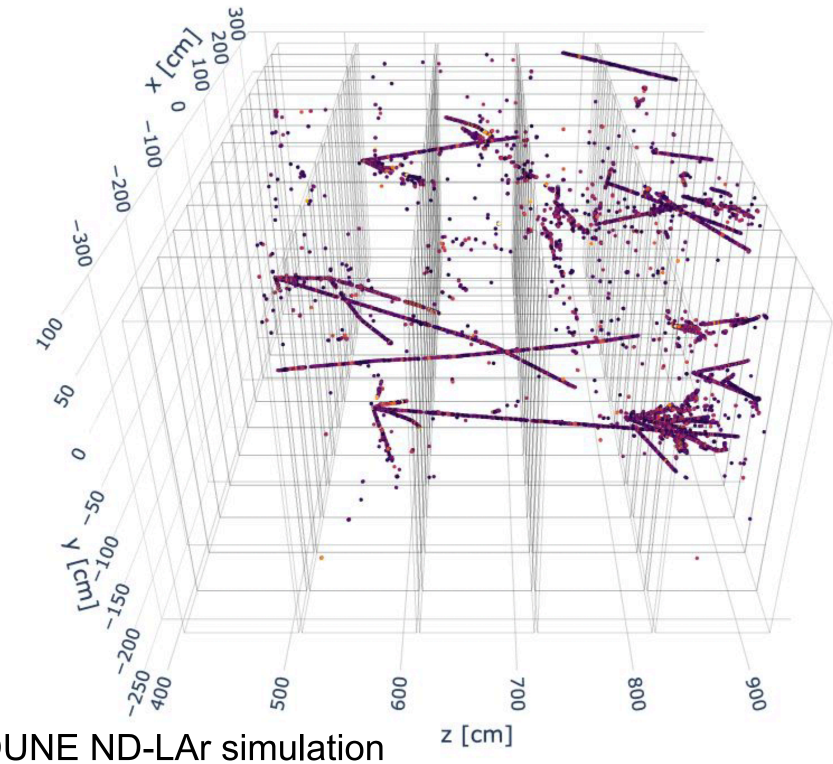
M. Wang (FNAL), et al., 24th IEEE Real Time Conference

Brookhaven National Laboratory

Computational Challenge of the data

- overlap of interactions increases the computation complexity of reconstruction
  - deconvolve hits
  - tracking combinatorics
  - vertexing
- HL-HLC expect μ > 200
  (currently μ = 64)
- DUNE ND-LAr overlap of neutrino interactions as high as 10-20 interactions in a 10 μs spill
- simulation of these complex events and large open detectors drives increased in computational resource needs
- projected compute need could be met by HPC and Accelerators resources
  - these solutions also present new challenges


ATLAS pp Collision

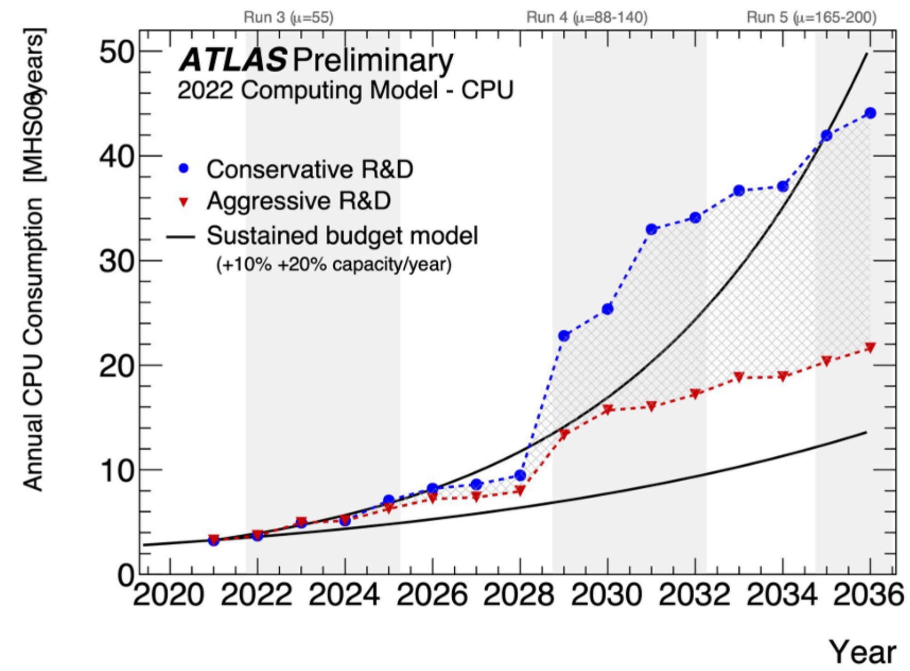


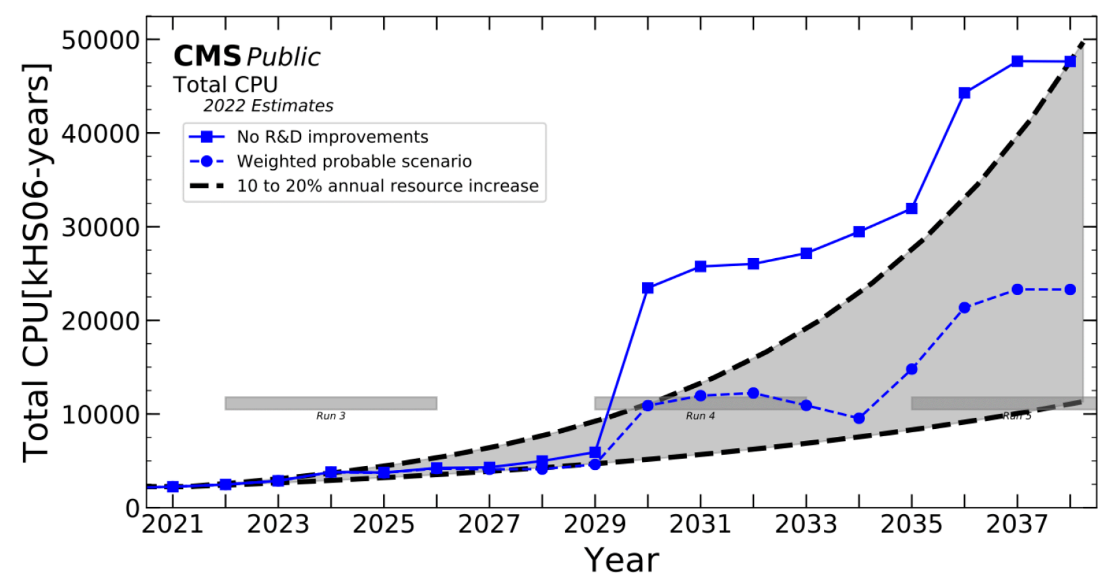
DUNE ND-LAr simulation

Brookhaven National Laboratory

Computational Challenge of the data

- overlap of interactions increases the computation complexity of reconstruction
  - deconvolve hits
  - tracking combinatorics
  - vertexing
- HL-HLC expect μ > 200
  (currently μ = 64)
- DUNE ND-LAr overlap of neutrino interactions as high as 10-20 interactions in a 10 μs spill
- simulation of these complex events and large open detectors drives increased in computational resource needs
- projected compute need could be met by HPC and Accelerators resources
  - these platforms also present new challenges

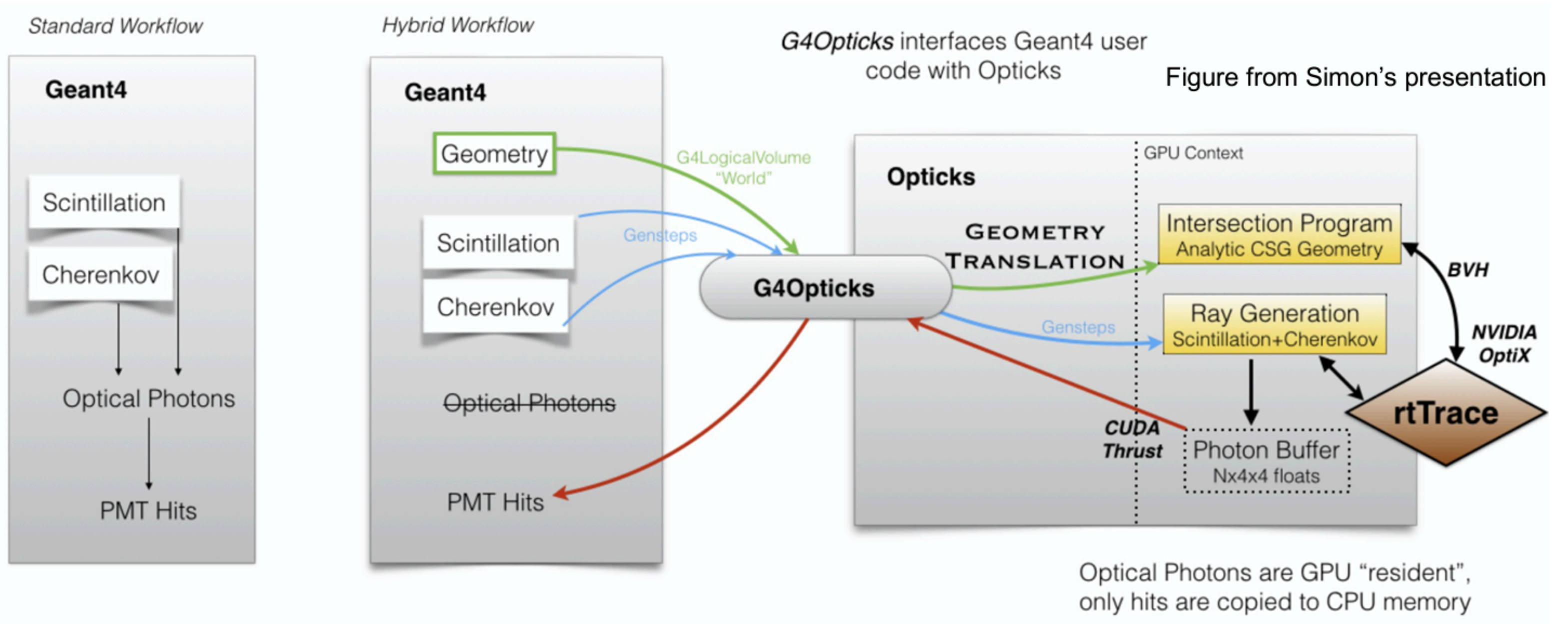




Brookhaven National Laboratory

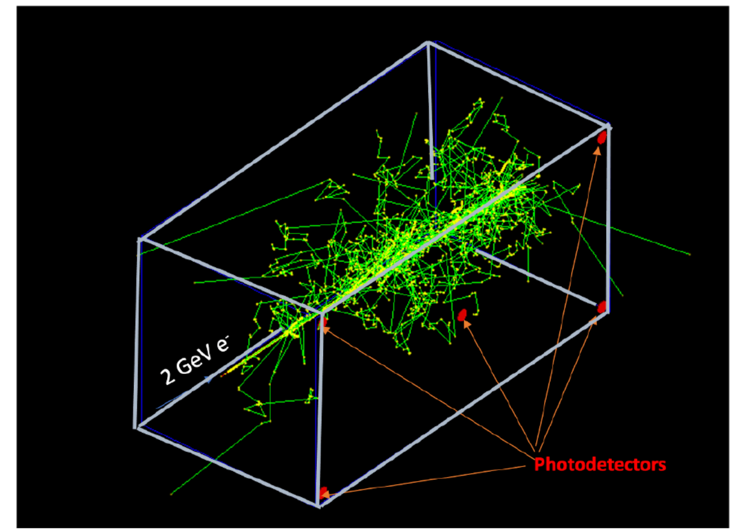
DUNE ND-LAr simulation

GPU based photon simulation



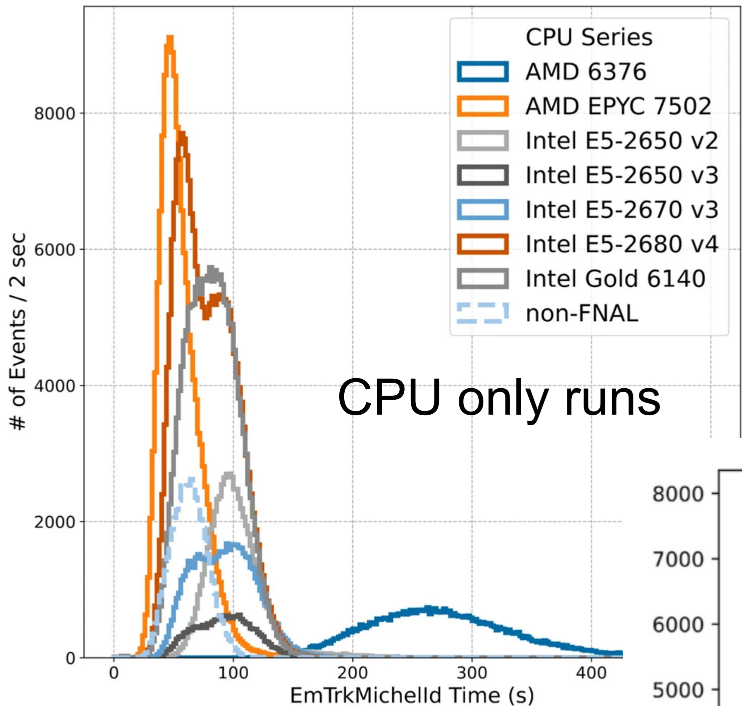
- simulation of photons within open detectors can be incredibly CPU intensive and limit event by event simulation
- often solved with voxel-ized lookup tables
- Opticks developed by Simon Blyth for JUNO, now part of GEANT4 releases

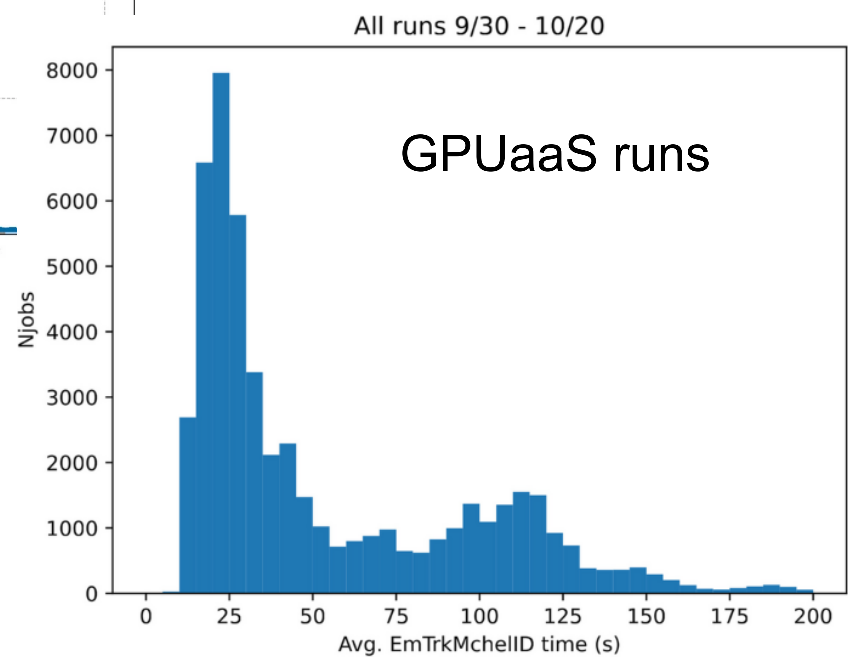
- CaTS: integrate GEANT4+Opticks in LArSoft for simplified LArTPC
- speedup of several 100 times for photon simulation
- also report that 1 core could saturate the GPU



Brookhaven National Laboratory

a cautionary tale - GPUaaS Inference





ProtoDUNE utilized CNN to classify reconstructed hits as track or shower like - EmTrkMichelID

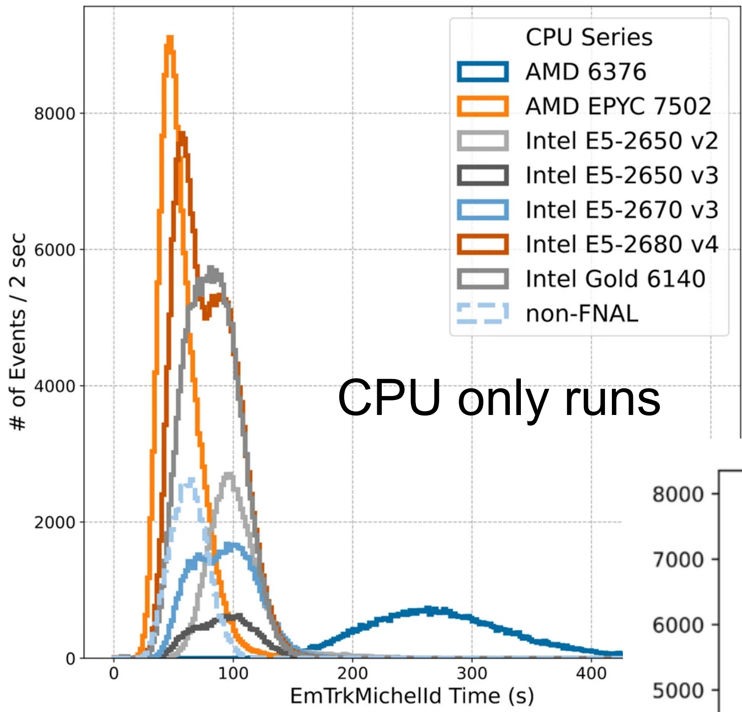
utilize NVIDIA Triton Inference Service to offload via network CCN inference

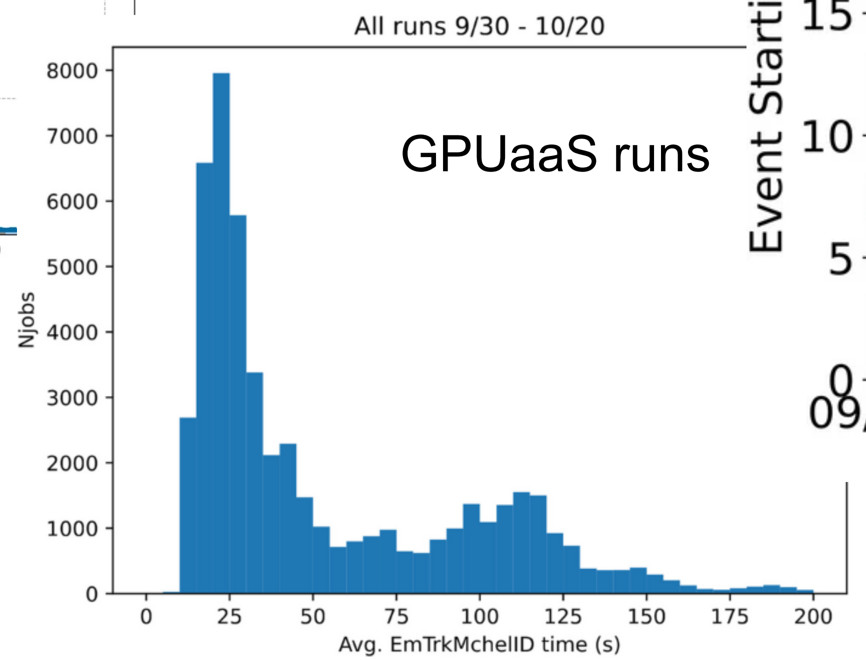
distributed jobs across FermiGrid and few OSG sites across North America

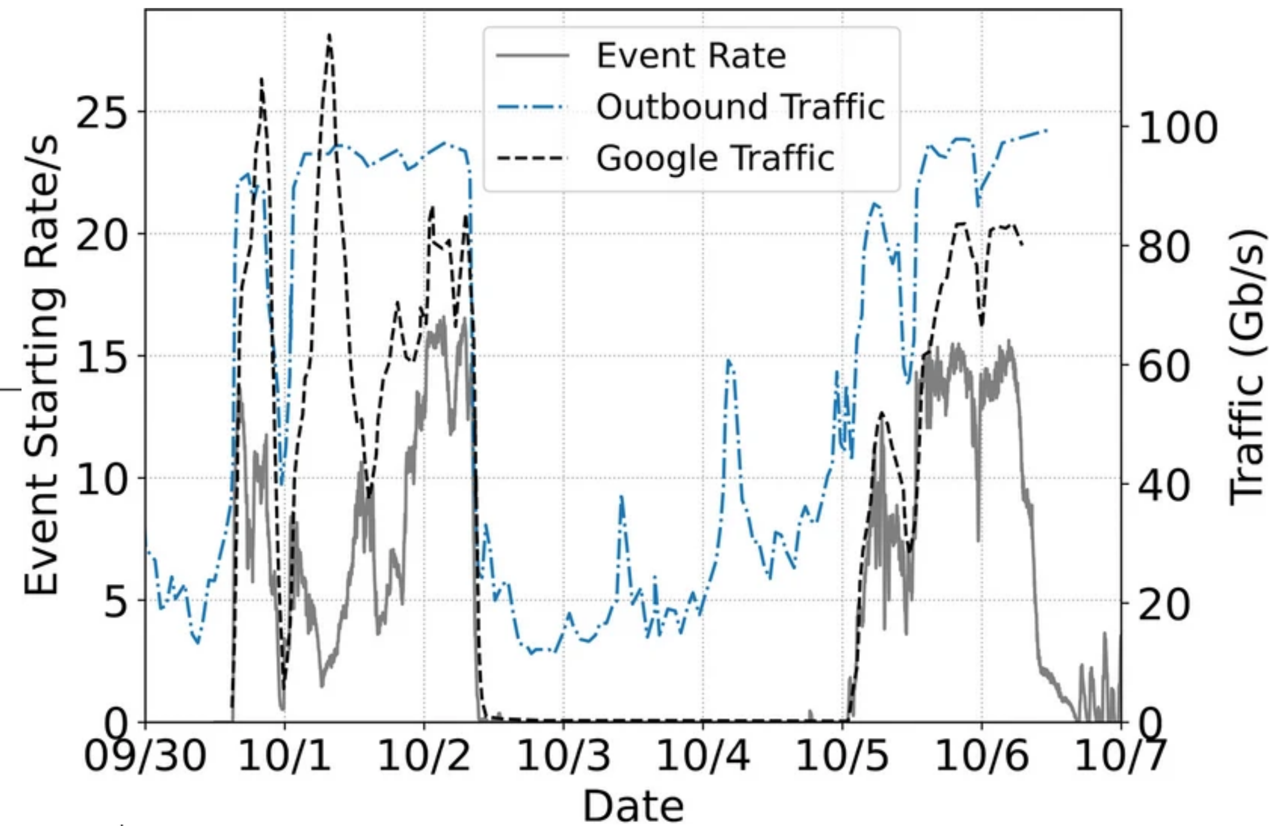
about that second bump…

Brookhaven National Laboratory

a cautionary tale - GPUaaS Inference

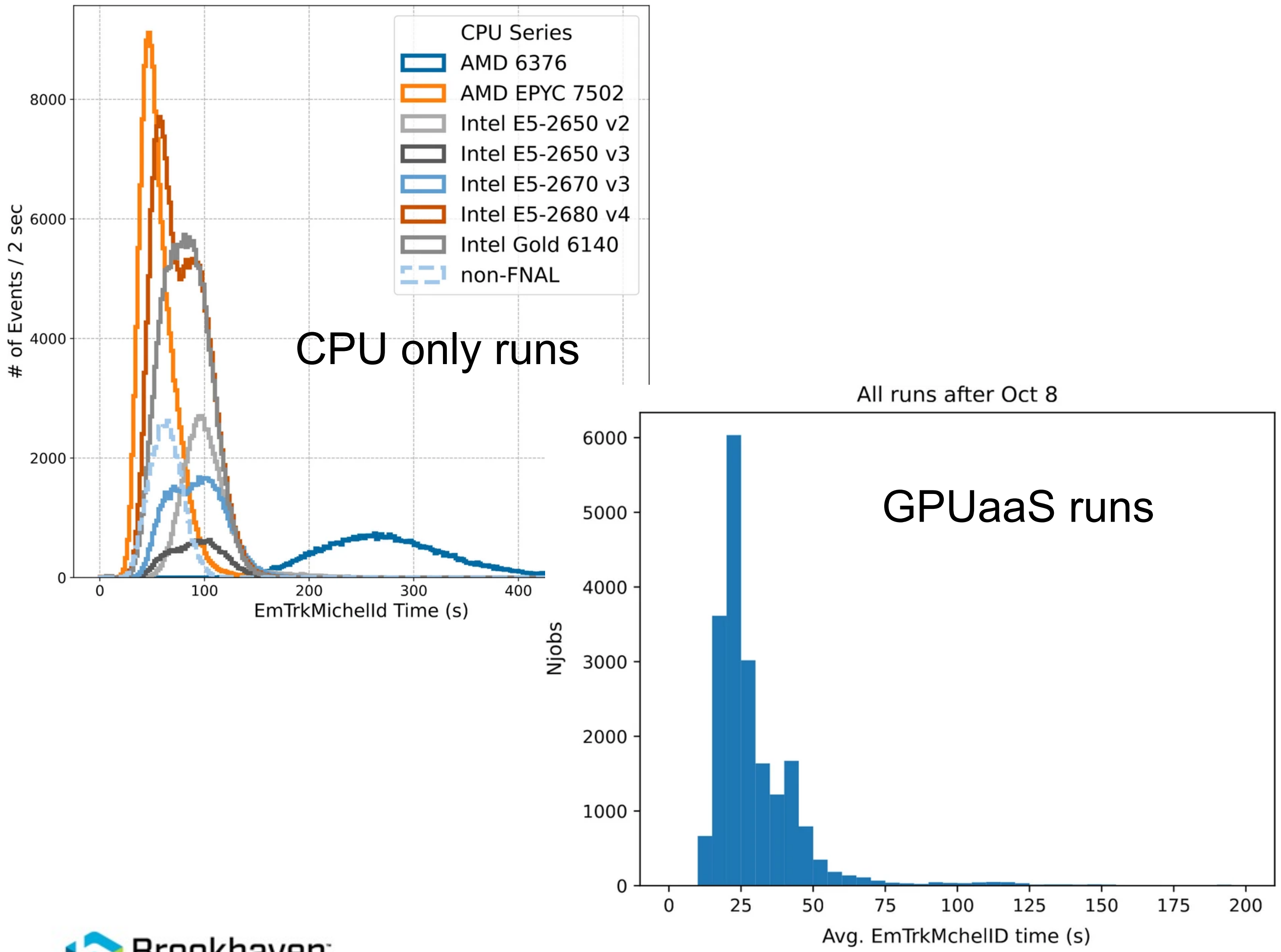






about that second bump…

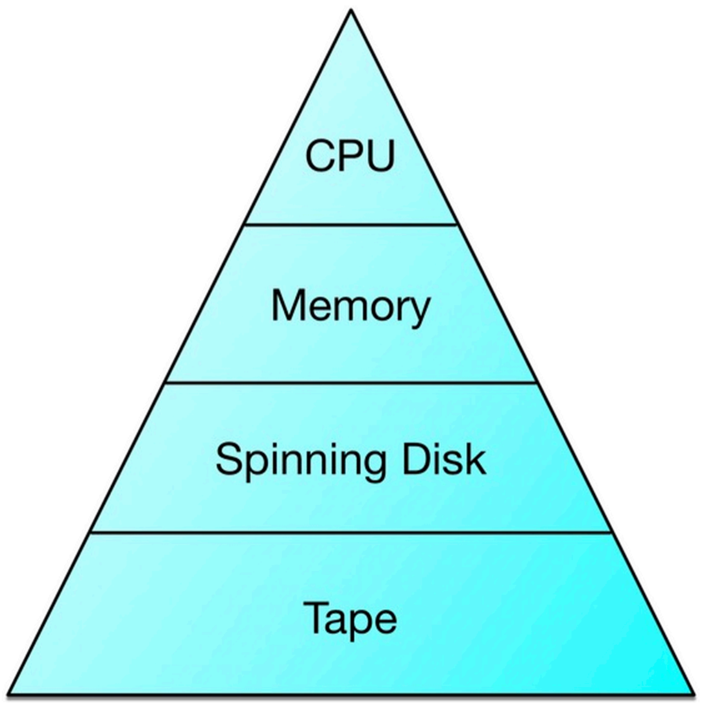
a cautionary tale - GPUaaS Inference



- ProtoDUNE utilized CNN to classify reconstructed hits as track or shower like - EmTrkMichelID
- utilize NVIDIA Triton Inference Service to offload via network CCN inference
- distributed jobs across FermiGrid and few OSG sites across North America
- about that second bump…

Brookhaven National Laboratory

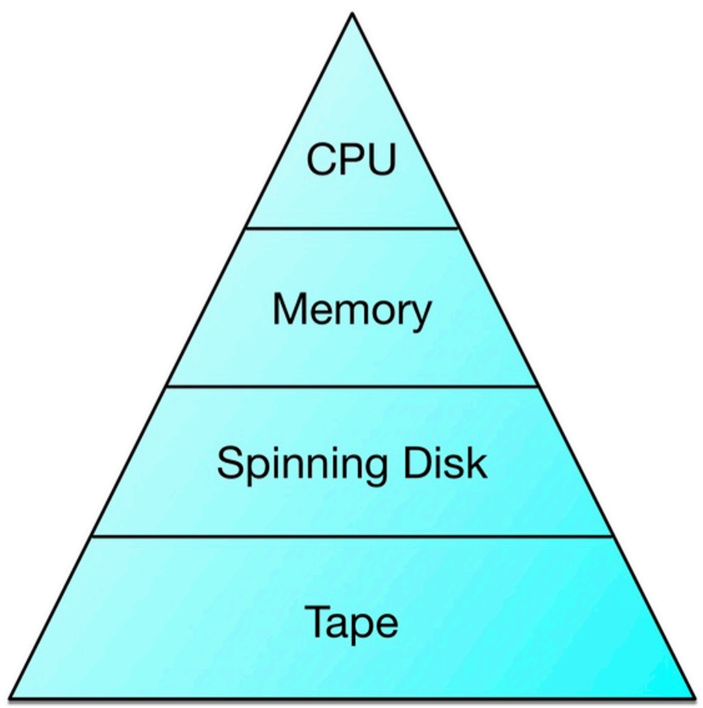
Challenge presented by HPCs and GPUs



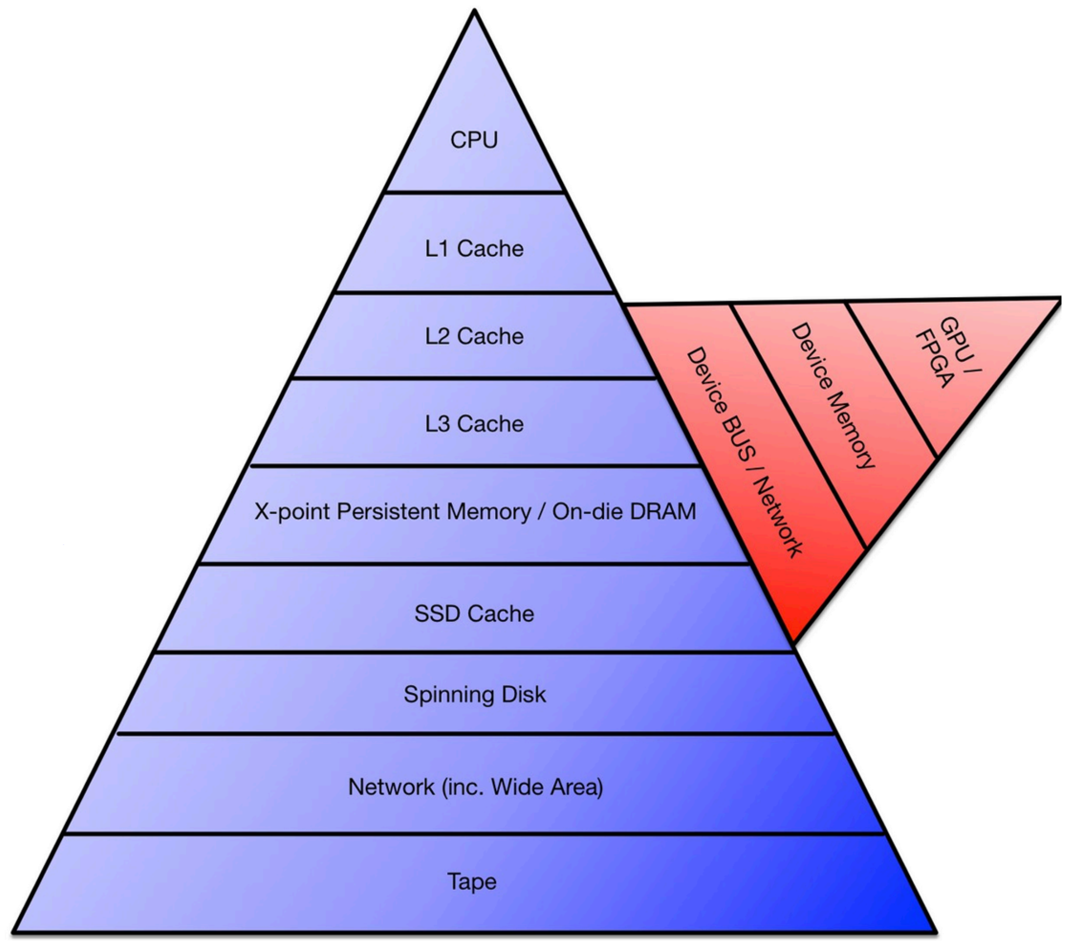
[Frinkiac 7]

Figure by Graeme Steward

Challenge presented by HPCs and GPUs



Frinkiac 7



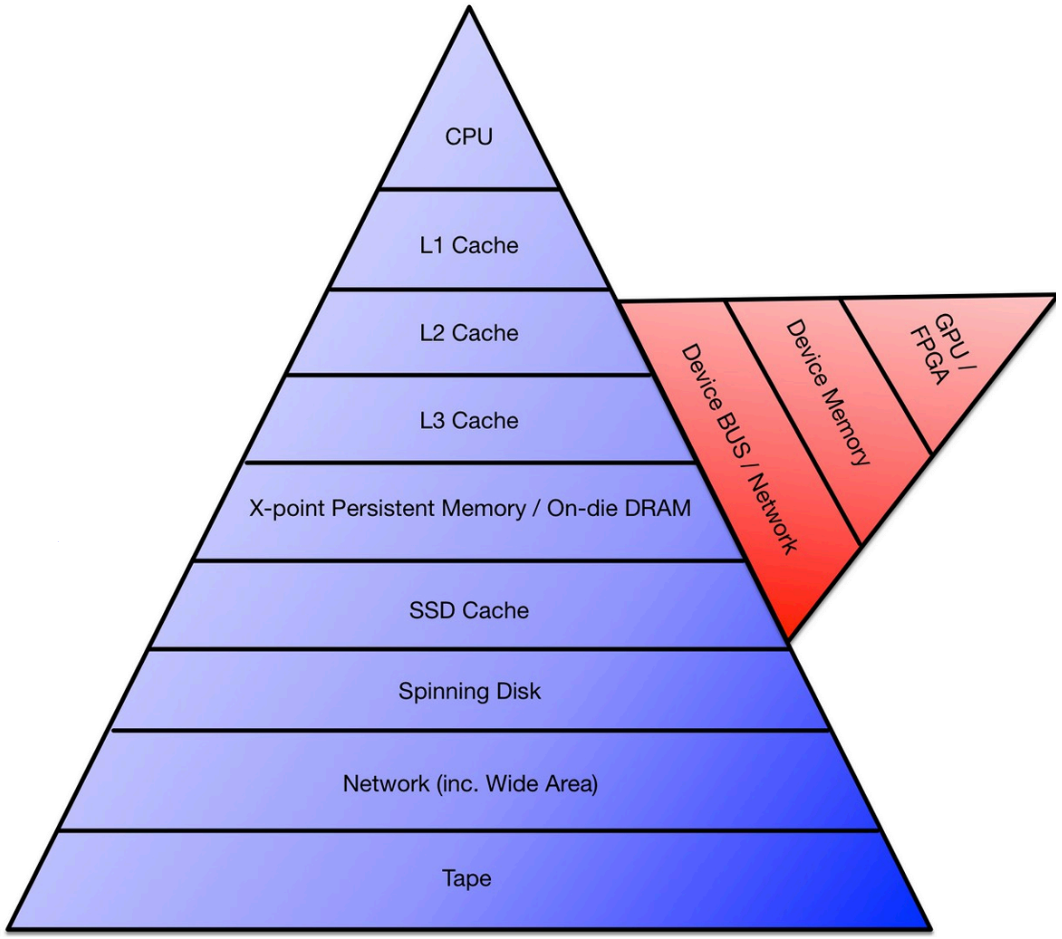
Aurora

Figure by Graeme Steward

Brookhaven National Laboratory

Challenge presented by HPCs and GPUs

- multithreaded/multiprocess workflows to optimize throughput and memory
  - LHC experiments have done extensive work on MT/MP
  - existing large code bases present huge lift for LArTPC experiments
  - scheduling many core architectures is complex
- moving from CPU code to GPU code can take signifiant restructuring of algorithm
  - not every task suited for GPU
  - optimizing for multilayer access is difficult
  - hardware/tools are rapidly evolving
- resources are (and will continue to be) heterogenous



Aurora

Figure by Graeme Steward

High Energy Physics - Center Computing Excellence

- Started in 2020 to address challenges posed by accessing US Leadership Class Facilities
- Combined effort across 5 US National Labs
- Experiments from Energy, Intensity, and Cosmic frontier
- Four main thrusts of reseach
  - Portable parallelization strategies
  - HEP I/O and HPC Storage
  - Event Generation
  - Complex Workflows

High Energy Physics - Center Computing Excellence

- Started in 2020 to address challenges posed by accessing US Leadership Class Facilities
- Combined effort across 5 US National Labs
- Experiments from Energy, Intensity, and Cosmic frontier
- Four main thrusts of research
  - Portable parallelization strategies
  - HEP I/O and HPC Storage
  - Event Generation
  - Complex Workflows

HEP-CCE Portable Parallelization

- With heterogeneous architectures, the desire for software to automatically adapt to hardware
- port existing algorithms into different portability layers to see performance
- evaluate on more than just objective metrics (how easy it was to port)
- recommended solution ended up being very dependent on application
    - simple kernels all performed at a similar level
    - complex algorithms/multiple kernels started to show limitations
- did not eliminate the need to adapt to the conceptual differences CPU →GPU

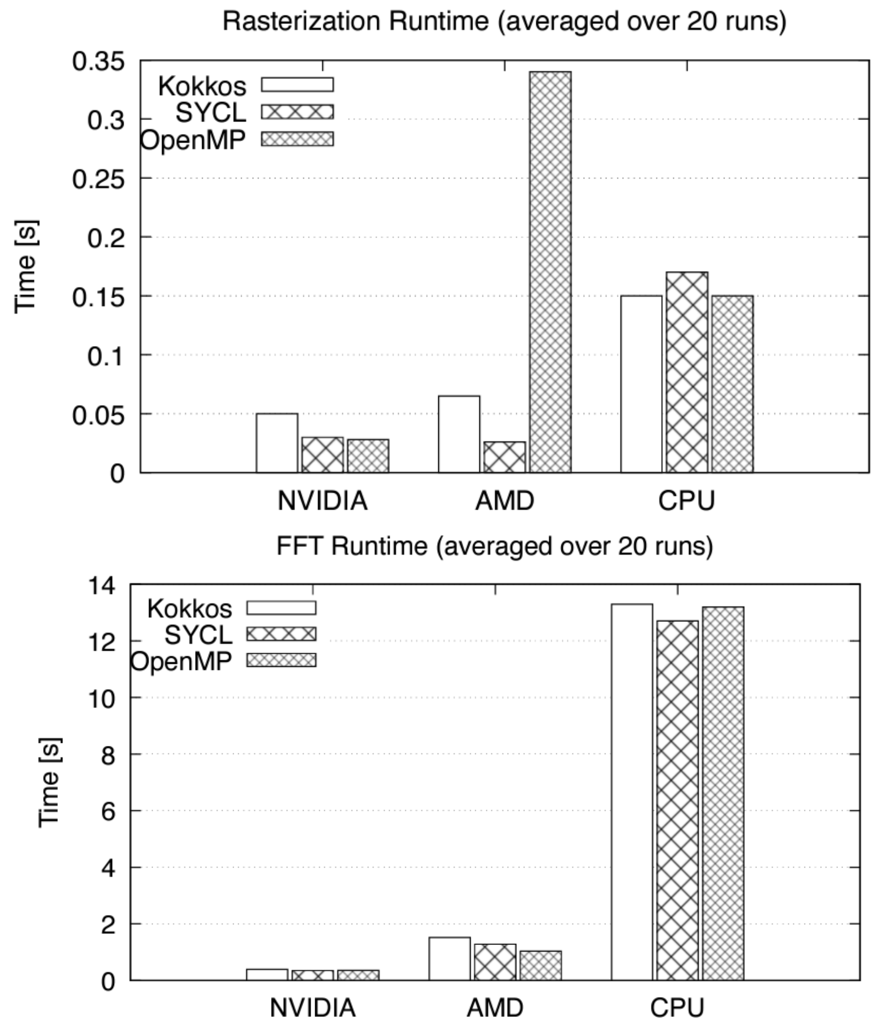
	Kokkos	SYCL	OpenMP	Alpaka	std::par
Patatrack	Done	Done*	WIP	Done*	Done compiler bugs
Wirecell	Done	Done	Done	no	Done
FastCaloSim	Done	Done	Done	Done	Done
P2R	done	Done	OpenACC	Done	Done

Charles Leggett
CHEP 2023

Brookhaven National Laboratory

HEP-CCE Portable Parallelization

- With heterogeneous architectures, the desire for software to automatically adapt to hardware
- port existing algorithms into different portability layers to see performance
- evaluate on more than just objective metrics (how easy it was to port)
- recommended solution ended up being very dependent on application
  - simple kernels all performed at a similar level
  - complex algorithms/multiple kernels started to show limitations
- did not eliminate the need to adapt to the conceptual differences CPU →GPU



WireCell Simulation
arXiv:2304.01841

Brookhaven National Laboratory

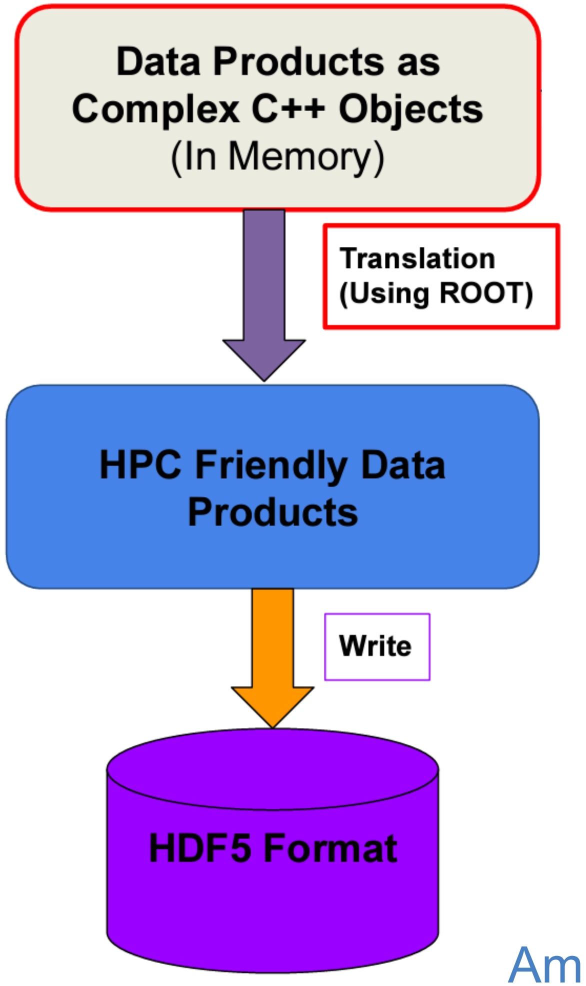
HEP-CCE Fine Grained I/O and Storage

Darshan I/O monitoring libraries applied to HEP applications

- track I/O activities
- data movement by task
- access patterns and volume
- analyzed ATLAS, CMS, DUNE
- identified ROOT serialization bottleneck

Data structure studies

- HDF5 data format as optimized for MT/MP workflows
    - widely used and supported at HPC facilities
- ROOT7 transition to RNtuple
    - updated data structures and C++ libraries
    - capable of columnar processing
    - 20-40% storage savings (ATLAS RNtuple)

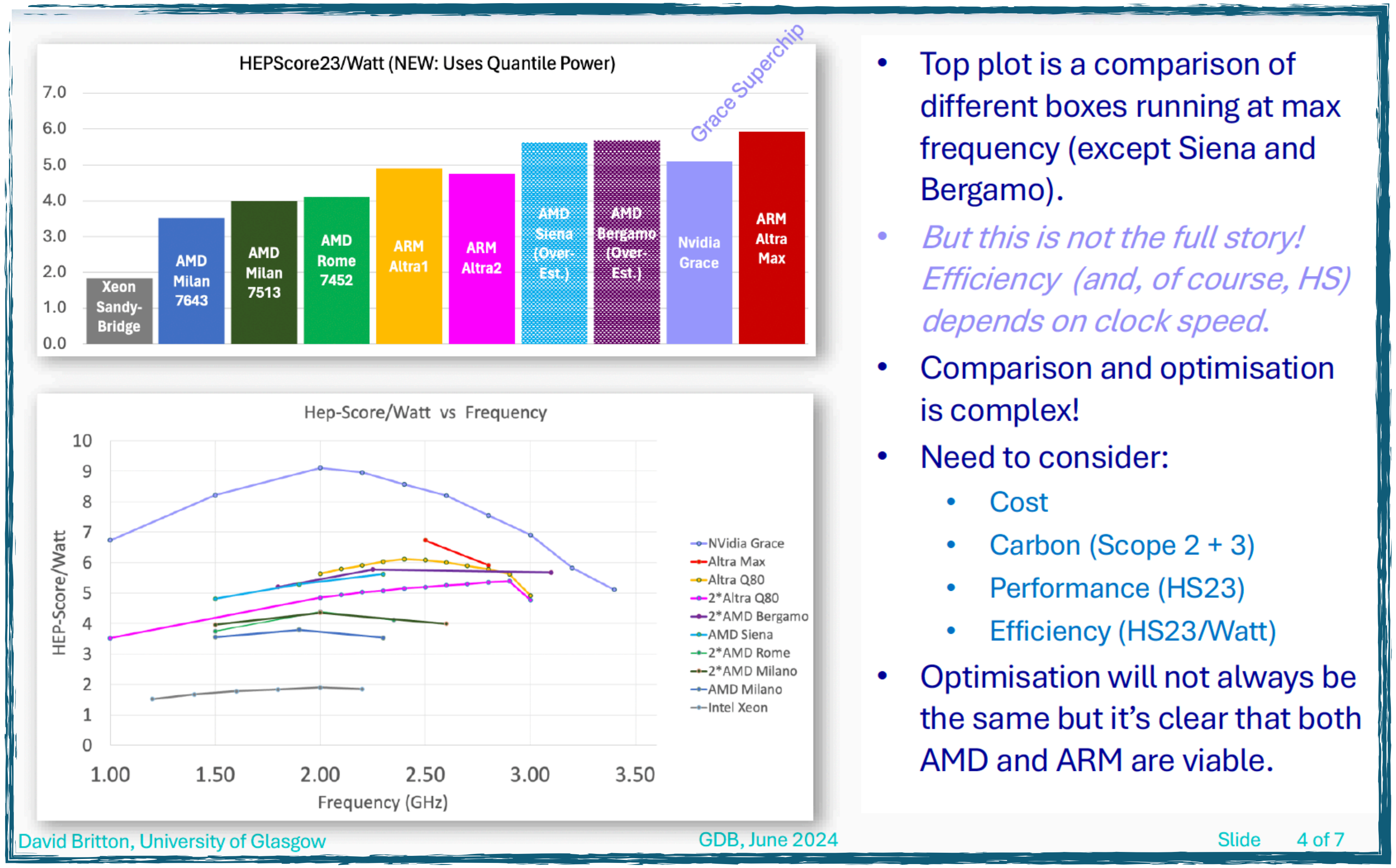


Amit Bashal, et al.

Brookhaven National Laboratory

Ecological responsibility in computing

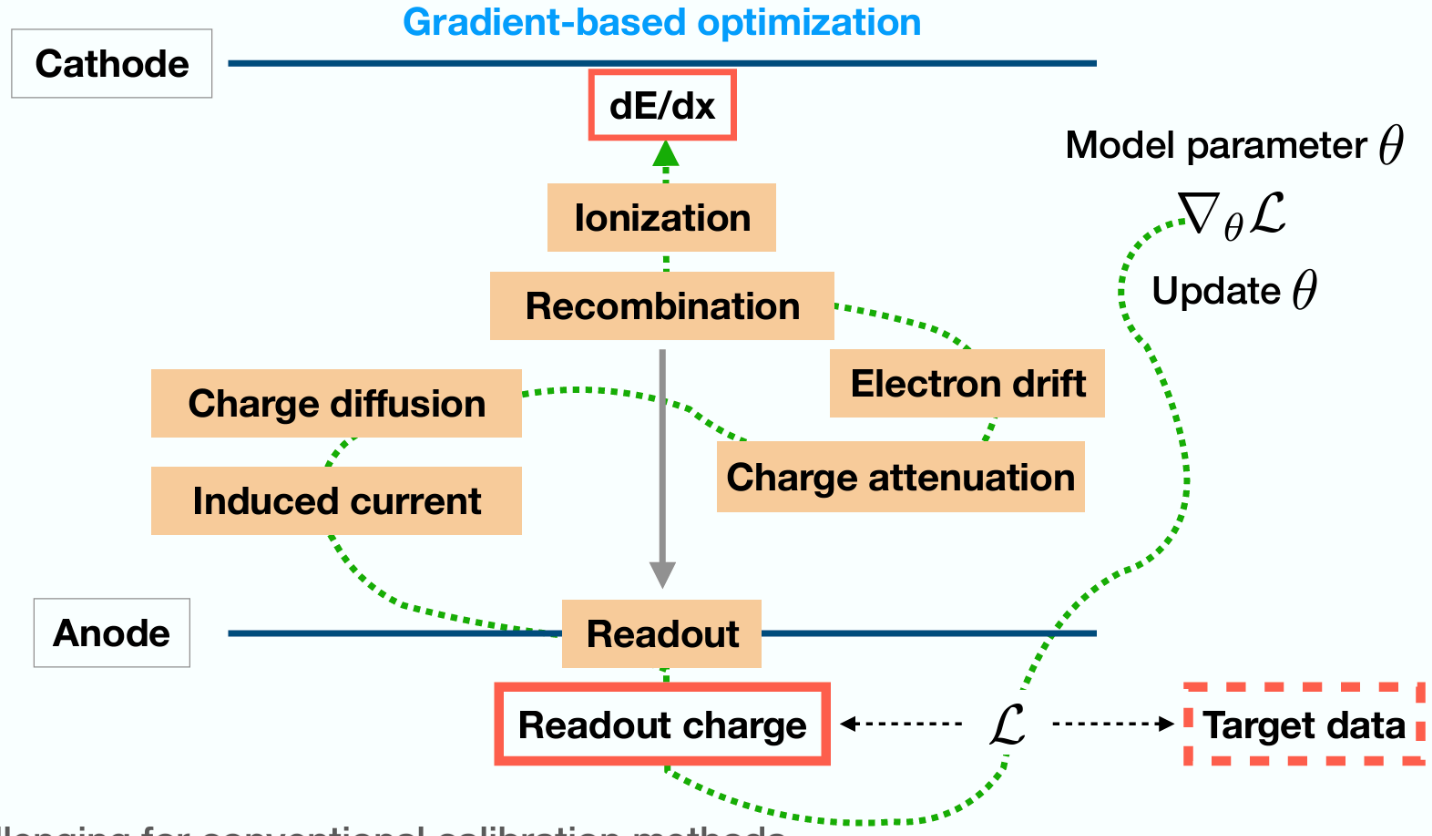
- strong encouragement for greater utilization of ARM chips
- limited instruction set, but traditionally thought of as lower power
- studies from David Britton (Glasgow) presented at WLCG GDB make optimization more complicated



Brookhaven National Laboratory

Development of Differential Simulation

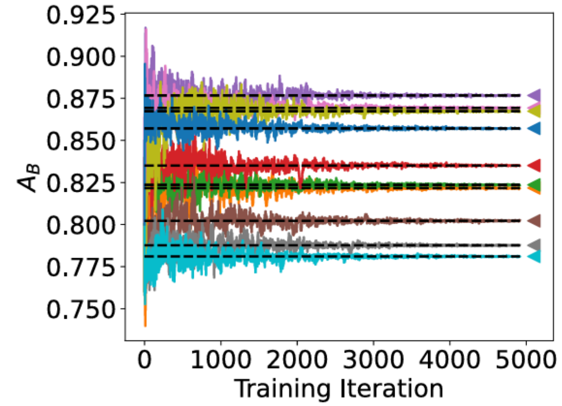
Calibration with a Differentiable Simulation



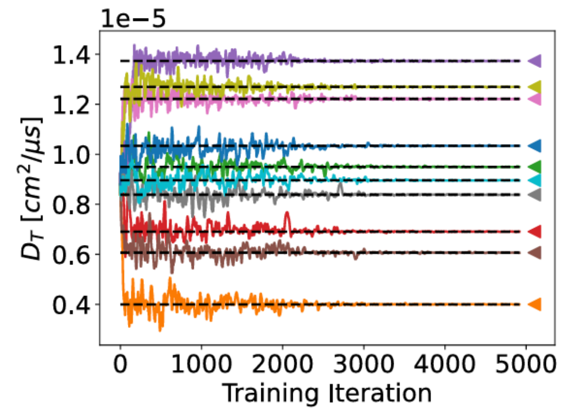
- Challenging for conventional calibration methods
- **Led to development of a differentiable simulation for high dimensional calibration**
  - Simultaneous optimization for multiple model parameters
  - Straightforward application of the calibration

Differentiable simulation of a LArTPC 4 Yifan Chen — SLAC, Stanford University

Recombination model A_B



Transverse diffusion coefficient



slides from Yifan Chen

Summary

- R&D is needed to improve storage beyond just bulk capacity
- take advantage of new facilities (ALCF, HPC) and architectures (GPU, FPGA, ARM) to meet computational needs
  - Balancing our workloads for efficiency
  - Impedance matching between resources and workflows
  - Adapting large code bases to Multithreading/Multiprocessing
- try to keep pace with advancements around us and externally developed tools
  - landscape of “Big Data” has changed around HEP
  - learn to incorporate external tools into software with extremely long lifetime
- shared developments serve essential role in this development (HEP Software Foundation, HEP-CCE, WLCG, OSG, etc)

backup