INTRODUCING AMIT BASHYAL

By Amit Bashyal

Research Associate

Brookhaven National Laboratory

INTRODUCTION

- Recently joined the Brookhaven NPPS group as Postdoctoral Research Associate
- This talk about myself, past works (in Argonne National Laboratory) and projects in Brookhaven National Laboratory

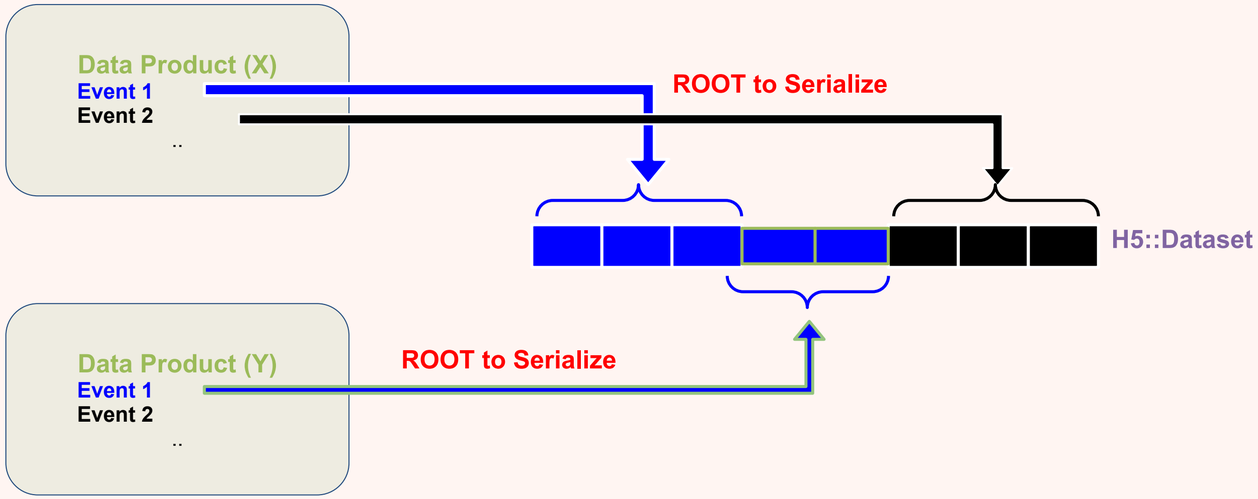
BEFORE ARGONNE NATIONAL LABORATORY

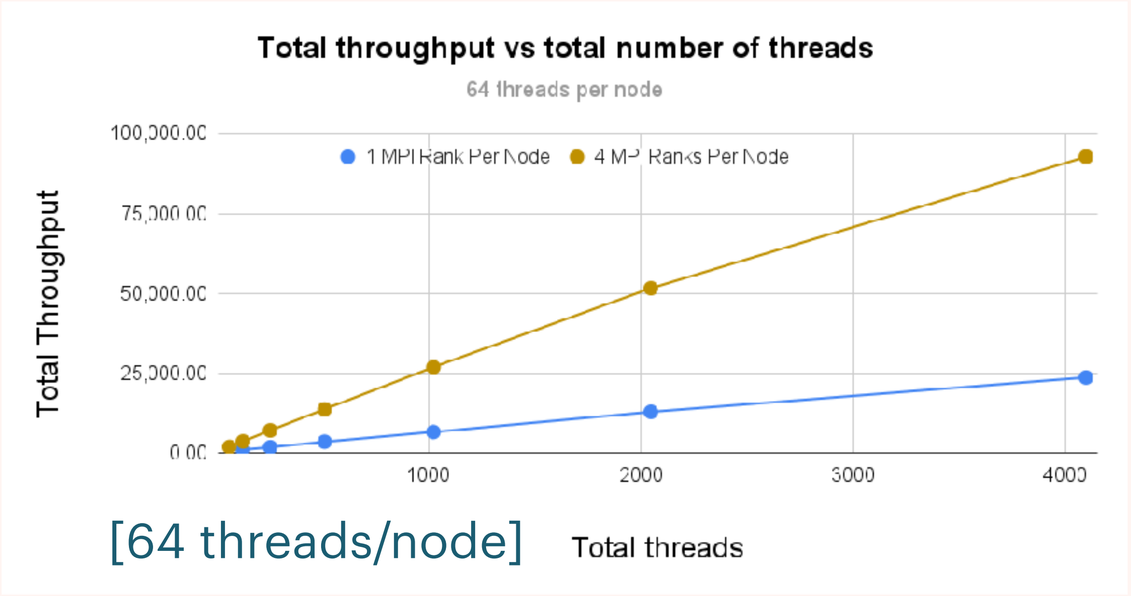
- Completed by undergraduate in 2014
  - Interest in Particle Physics
- Ph.D at Oregon State University (2015 - 2021)
  - Supervisor : Heidi M. Schellman
  - Thesis Topics : Neutrino Fluxes in DUNE and Cross-section measurement in MINERvA experiment (FNAL Seminar)
    - 2 journal papers (here and here)
  - Internship at Argonne National Laboratory (2018)
    - Implemention of Message Parsing Interface (MPI) libraries for parallel I/O in ROOT Framework (Link)

ARGONNE NATIONAL LABORATORY

- Postdoctoral Research Associate at Argonne National Laboratory (2021 - 2025)
  - Involved in High Energy Physics - Center for Computational Excellence (HEP-CCE)
  - Leveraging HPC resources for future generation HEP experiments (HL-LHC, DUNE)
  - Development of tools for I/O the HPC envronment, RNTuple investigation
    - Use of HDF5 libraries for parallel writing of HEP data
    - Accelerator Friendly Data model for the HEP experiments
    - Persistence of DUNE data models in RNTuple
  - Received LDRD seed grant to investigate modern lossy compression algorithms for DUNE data
  - Level 3 co-manager for the Accelerator Integration Efforts in the DUNE experiment

PARALLEL I/O OF HEP DATA





HEP data products are serialized before writing them into HDF5. Process is parallelized such that multiple parallel processes can write into one HDF5 Dataset simultaneously.

Throughput as a function of number of threads with 64 threads per compute node. Throughput is shown for 1 parallel process per node and 4 parallel process per node configuration. Test was done in CORI@NERSC.

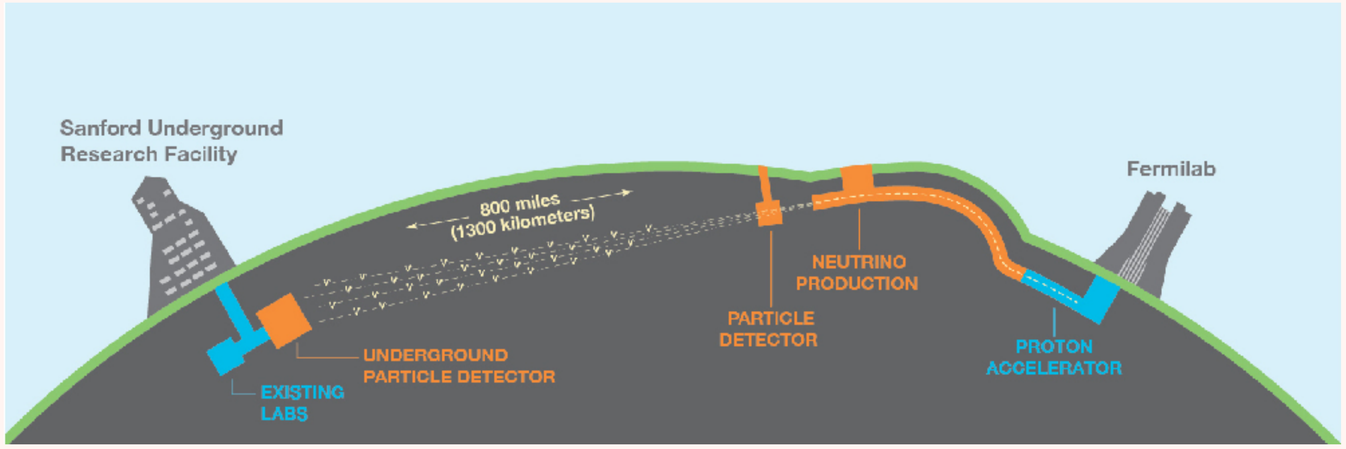
GPU FRIENDLY DATA MODEL: SURVEY

- Led the survey of efforts made by experiments to make their data models GPU friendly
- Summary
- Complex Event Data Model (EDM) → Challenge to adopt GPUs
  - Most of the data models designed for ROOT storage (supports persistence of complex C++ objects)
  - Most of the efforts on Data model transformation (AoS →SoA)

Experiment	ATLAS	CMS	DUNE	EDM4hep	NOvA*
General Overview					
Speaker	Scott Snyder	Matti Kortelainen	Mike Kirby	Benedikt Hegner	Marc Paterno
Talk Link (Indico)	https://indico.fnal.gov/event/57595/contributions/256583/attachments/162731/215145/2023-01-10-edm.pdf	https://indico.fnal.gov/event/55536/	https://indico.fnal.gov/event/58260/	https://indico.fnal.gov/e/55542	https://indico.fnal.gov/event/58962/contributions/262454/attachments/165673/220182/DataOrganizationForParallelProcessing.pdf
Github Link:	https://gitlab.cern.ch/akraszna/asyncgaudi.git		–	https://github.com/key4hep/EDM4hep	https://github.com/art-framework-suite/hep-hpc
Languages	C++/CUDA	C++/CUDA/ALPAKA	C++/CUDA	C++/python	python (PandAna*)
Storage Support	ROOT	ROOT	ROOT/HDF5	ROOT/LCIO	HDF5
Data Model	xAOD	Arrays	art	POD (Plain	CAF

PERSISTENCE OF DUNE ANALYSIS DATA MODEL IN RNTUPLE

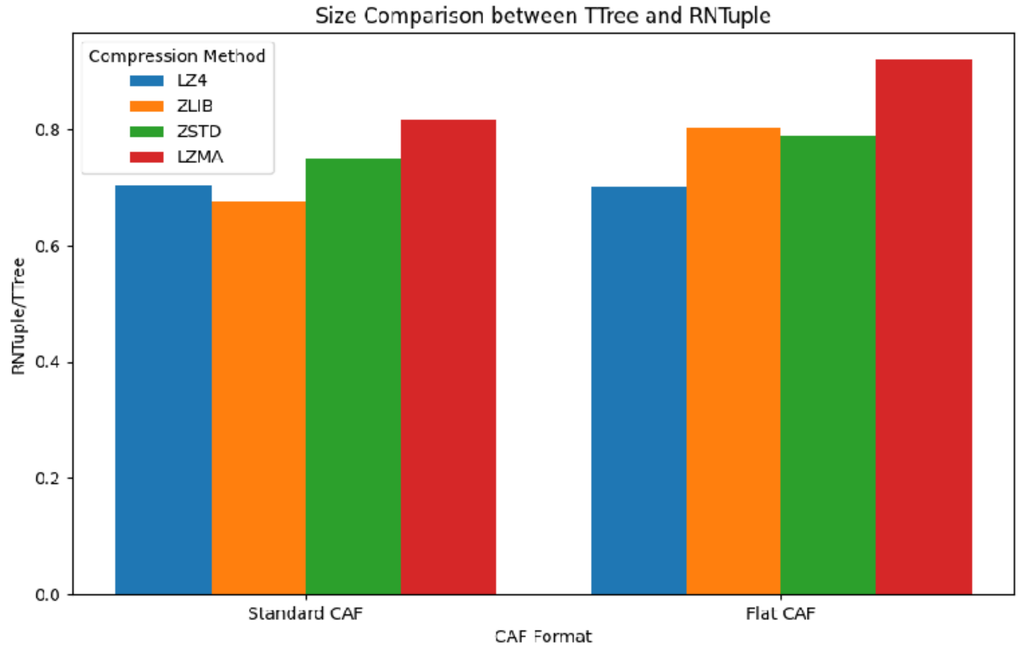
Neutrino oscillation experiments measure neutrino events in two detectors.
Near the neutrino source to profile the neutrino beam.
Far (~1200 km in the case of DUNE) to record (un)oscillated neutrino events



- DUNE analysis data model →CAF (Common Analysis Format) Data model
- Records events in near and far detectors (often with different detector designs/technologies) in same data model
- Events are recorded as tracks and showers
- Generator level information for simulated events

CAF : TTREE TO RNTUPE

- In the DUNE/HL-LHC Era, RNTuple will replace TTree as the primary I/O subsystem and storage backend
  - Many details in many talks (including this) but future experiments have to adopt RNTuple to stay current with the ROOT Eco-System
- Wrote standalone framework that does I/O of CAF obects using RNTuple API and stores in RNTuple formats
  - Toy framework to generate CAF objects
  - Replacement of raw pointers with smart pointers.
  - Data model by itself is (mostly) compatible with RNTuple
  - Examples (in both TTree and RNTuple) to show the usage RNTuple for U/O



File size comparison of CAF objects persisted in RNTuple to TTree format with different compression algorithms provided by ROOT.

More info here

Brookhaven National Laboratory

INTELLIGENT LOSSY COMPRESSION OF DUNE DATA

- Use of SZ3 algorithm to compress the DUNE raw waveform data
- LDRD Seed proposal to mentor a student (Final presentation here, technote here)
- Test framework that allows (de)compression of DUNE raw data using SZ3 compression algorithm

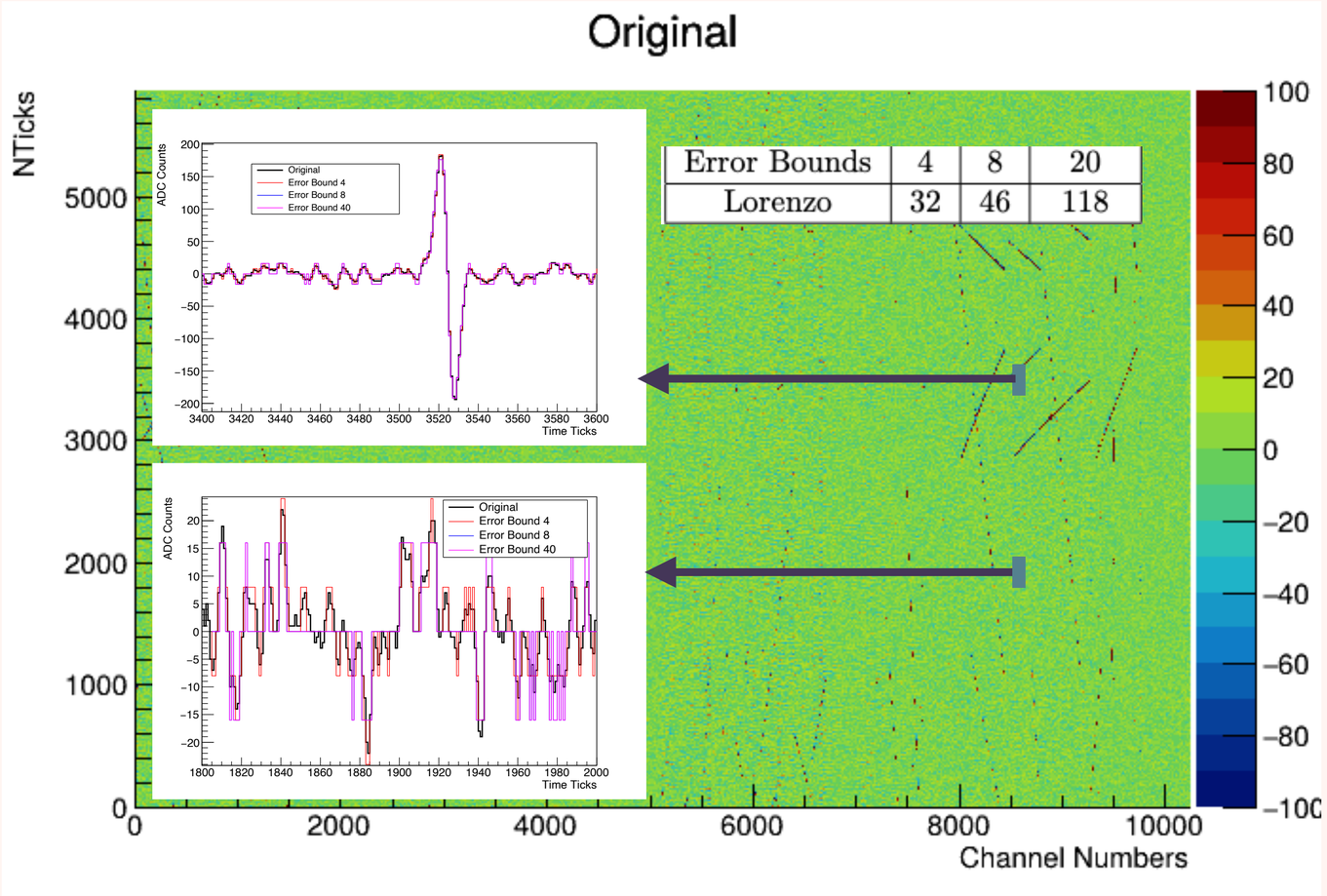


Figure shows the Proto-DUNE wave form data as a function of channel numbers and time ticks. Color scale is the ADC count recorded by the channels. 1 D histograms shows the 1 D projection of readouts in different regions of channel 8500 with different resolution decompressed data. Compression with Lorenzo Prediction of SZ3 algorithm.

TRANSITION TO BNL

- Joined NPPS group as postdoctoral Research Associate in March 3, 2025.
  - DUNE and AID(2)E projects
- **DUNE**
  - LDRD-D Project on the research and development of LAr TPC raw data processing utilizing heterogeneous compute resources for DUNE Far detector
  - Optimize and develop payload and DP algorithms to utilize GPUs efficiently
  - Synergy with current role as L3 manager for acceleration integration in DUNE
- **AID(2)E (AI-Assisted Detector Design for EIC)**
  - Design of second EIC detector using AI with Multiobjective Optimization (MOOP)
  - Optimization using ax libraries
  - Scaling of compute intensive optimization using PanDA and iDDS system (my current focus)
- Many thanks to NPPS group members to help me with onboarding (access to machines, repositories, tools and answering my questions)

TRANSITION TO BNL

- Moving to NY before August.
- Hope to see (most of) you soon!!