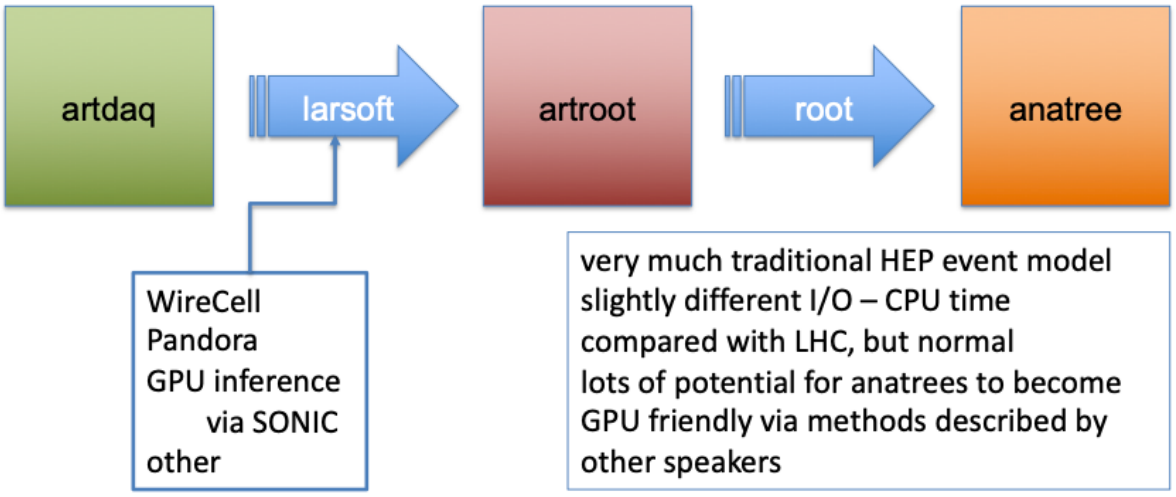
GPU Friendly Data model for the DUNE

Link to the actual talk: https://indico.fnal.gov/event/58260/

Current Status of the Experiment

- Actual Data taking from 2031
- ProtoDUNE II anticipated to operate in CY 2023
    - Scaled down version of DUNE FD
    - Eventually will be used in the actual FD
- Computation works and framework development happening with proto detectors
- Pro: Scaled down version gives time to develop framework for the actual detector
- Con:
    - Formats and frameworks could change when the detector comes online
    - Things may not work as expected

ProtoDUNE Single Phase Data Model



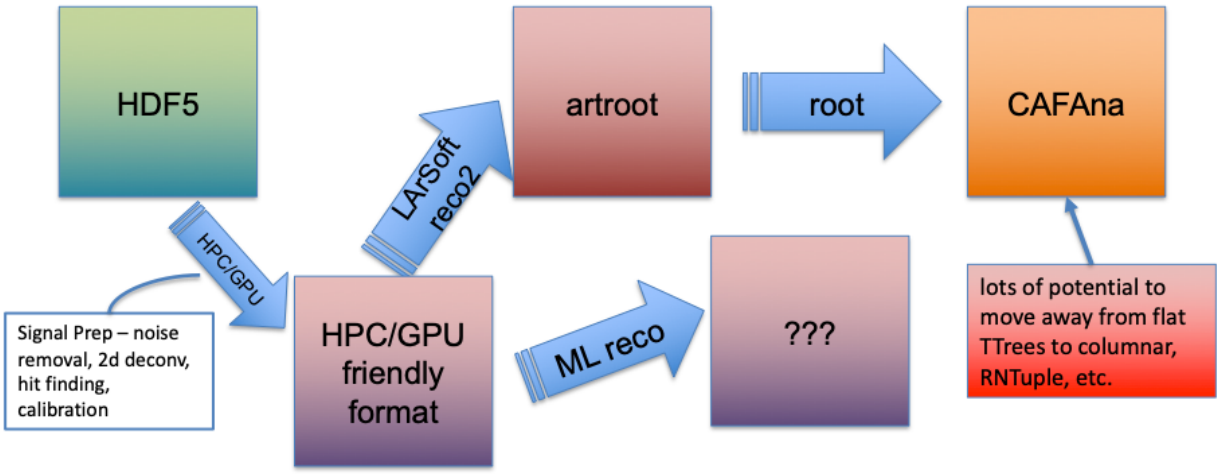
Current workflow uses GPUs (need to figure out what artdaq object types are here?)

Artdaq (has ROOT Object inside)
Larsoft (includes algorithms from Wirecell/Pandora etc) and output to ROOT format.
Then Run ROOT macros to transform into anatree (simple tree)

AnaTrees to become GPU Friendly using methods adopted by CMS/ATLAS

Transformation to make GPU Friendly

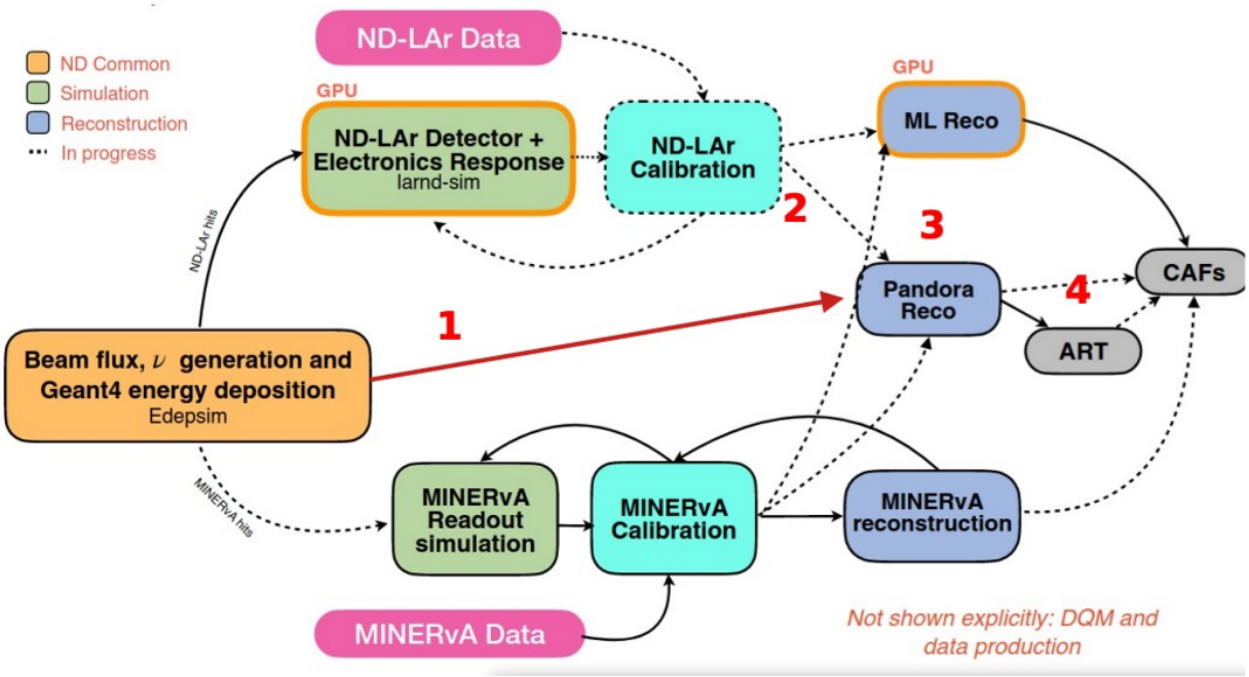
Use of HDF5 in the DAQ level files



Using HDF5 and the data format to be HPC friendly for some of the tasks.

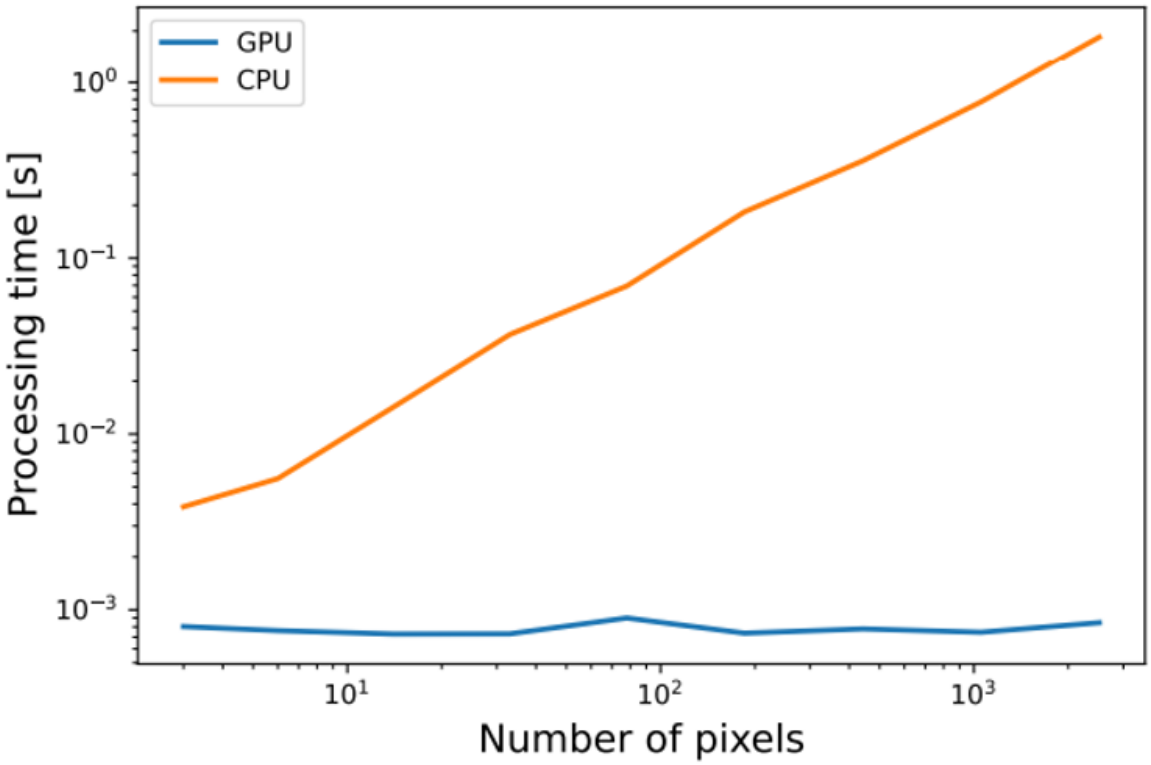
Data Structure in HDF5

- 32 bits allocated for the Header Information
- 14 bits of actual Channel Data
    - Note that GPUs generally have a memory cache of 16 bits.
    - 14 →16 bits could make the contiguous copy of data between host and memory and while using shared memory in the GPUs (Added by Amit)



Workflow chain of the DUNE

The final analysis level ntpules are produced in CAFAna. Currently 3 different methods of workflow seem to be implemented out of which one of them uses GPUs during the detector + electronics response simulation (which are usually compute intensive) and reconstruction using the ML.



Electronic Response simulated in the CPU and the GPUs (CORI?)

Current Status

- DUNE still on discussion stage regarding use of HDF5 vs. ROOT based format data model
    - Motivation being ML/AI can benefit using HDF5 at ntuple and reconstruction label (see the workflow above)
    - Need to do more studies on optimizations
    - Benefit from the studies done by CMS, ATLAS and other projects.