BERKELEY LAB

DEEP UNDERGROUND NEUTRINO EXPERIMENT

2x2 Calibration Status and Plans

Kevin Wood, kwood@lbl.gov

January 19, 2023

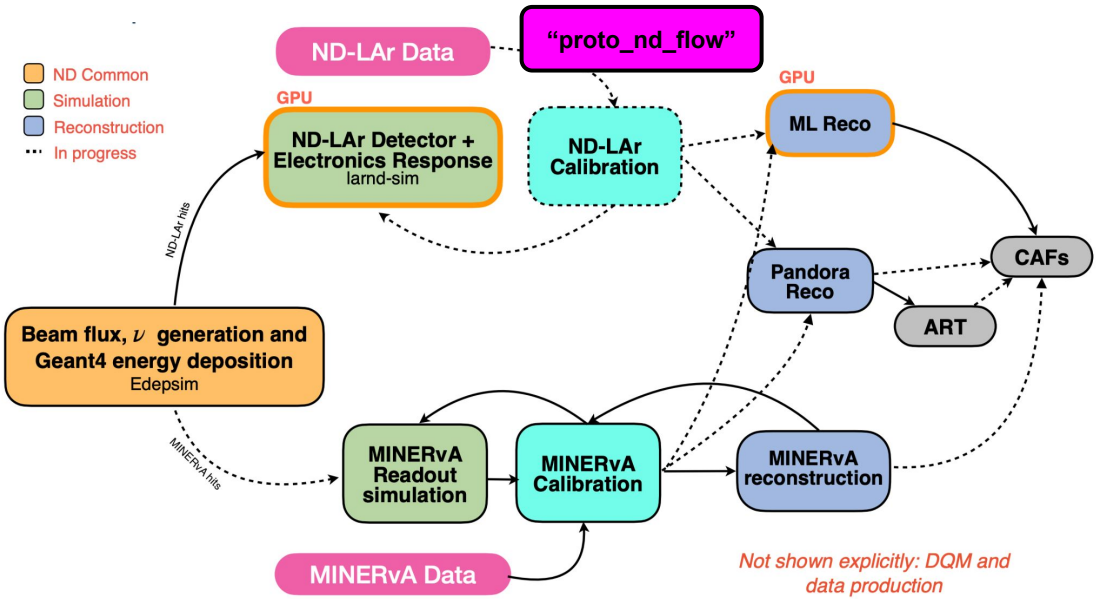
2x2 Analysis Workshop, University of Bern

Outline

- The “flow framework” for processing and calibrating 2x2 data
  - Focus on charge readout
- Data formats
- What effects do we aim to calibrate out?
- What samples can we use to this end?

2x2 Calibration Overview

- 2x2 calibration routines will be handled by a dedicated “flow” framework
- Based off module0_flow, developed by P. Madigan
- Currently working on branch: refactor/restructure-for-ndlar-flow that will extend the repo for use on different geometries/detector setups
- 2x2 code is found under proto_nd_flow subdirectories



Flow Overview

- Built on top of h5flow (also developed by Peter) for efficient handling of data
- Read and writes to HDF5 files
- Various groups for different types of information
- proto_nd_flow modules are written to process and calibrate the raw (h5) packets

```
In [1]: import h5py
        flow_out_h5 = h5py.File('/home/kwood/research/dune/2x2/data/simulation_ch

In [2]: flow_out_h5.keys()

Out[2]: <KeysViewHDF5 ['charge', 'combined', 'geometry_info', 'lar_info', 'mc_tr
        uth', 'run_info']>

In [3]: flow_out_h5['charge'].keys()

Out[3]: <KeysViewHDF5 ['calib_prompt_hits', 'events', 'ext_trigs', 'packets', 'p
        ackets_corr_ts', 'raw_events', 'raw_hits']>

In [4]: flow_out_h5['/charge/calib_prompt_hits'].keys()

Out[4]: <KeysViewHDF5 ['data', 'ref']>

In [5]: flow_out_h5['/charge/calib_prompt_hits/data'].dtype

Out[5]: dtype([('id', '<u4'), ('x', '<f8'), ('y', '<f8'), ('z', '<f8'), ('t_drif
        t', '<f8'), ('ts_pps', '<u8'), ('Q', '<f8'), ('E', '<f8')])
```

Flow Overview - references

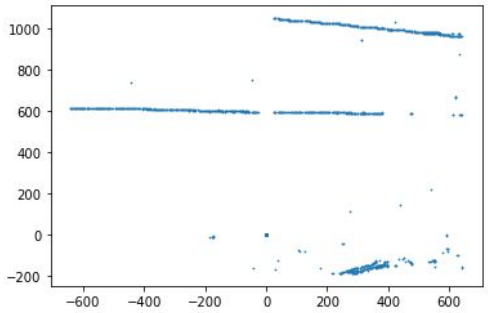
- h5flow also has mechanisms in place to establish references between datasets
  - E.g. between charge/calib_prompt_hits and charge/events *
  - E.g. for truth matching in simulation files
  - E.g. for associating hit level information back to the full packet
- (See h5flow README.md for more details)

```
In [1]: import h5flow
        import numpy as np
        import matplotlib.pyplot as plt
        flow_out_h5flow = h5flow.data.H5FlowDataManager('sim2x2_challenge_5xNuMIs
                                                        ,mode='r', mpi=False)

        WARNING:root:Running without mpi4py because No module named 'mpi4py'

In [2]: event_hits = flow_out_h5flow['charge/events','charge/calib_prompt_hits']

In [3]: plt.scatter(np.ma.getdata(event_hits['z'][3]),
                    np.ma.getdata(event_hits['y'][3]), s=0.5)
        plt.show()
```



Data Processing - Charge

1. LArPix packets from Pacman DAQ — datafile['charge/packets']
   - 1.1. Map software channel to detector location
   - 1.2. Subtract predetermined pedestals

   — datafile['charge/raw_hits']
   - 1.3. ADC → ke^- calibration assuming uniform gain,
   - 1.4. ke^- → MeV calibration assuming fixed dx
   - 1.5. Reconstruct drift coordinate

   — datafile['charge/calib_prompt_hits']
   - 1.6. Correct ADC → ke^- for gain variations
   - 1.7. ke^- → MeV calibration (refined)
   - 1.8. Calibrate for detector distortion, electric field, etc.

   — datafile['charge/calib_final_hits']

Calibrated Dataset Format

- There will be designated ‘calib_hits’ dataset from running low level reconstruction and full the suite of calibration on raw data
- A generic calibration dataset has the same hit-level datatype structure:
  - x [mm] - reconstructed/calibrated x position
  - y [mm] - reconstructed/calibrated y position
  - z [mm] - reconstructed/calibrated z position
  - Q [ke-] - reconstructed/calibrated integrated charge
  - E [MeV] - reconstructed/calibrated energy
  - t_drift [ns] - reconstructed drift time
  - t_pps [ticks] - when the charge arrived on the pixel

Light Information

- Not currently in the simulation challenge file
- module0_flow has modules for handling light information that we can use as a starting point
  - Raw waveforms
    - ~90% of the file size from single module operations (?)
  - Hits
    - SiPM level vs. detector level
    - Currently the latter where sipms on the same light detector are summed (after noise filtering, signal deconvolution) before hit finding
    - Keeping the same structure as the charge hits is possible, but x,y,z information has a different interpretation

Truth Information (simulation)

- At the moment there are handles for associating packets and hits to the GEANT truth information
  - Segments - information on the energy depositions from final state particles
  - Trajectories - description of the final state particles themselves
- Still need to put in the association back to the GENIE event record
  - Trivial to include, but there is a question about how much information we want to carry along in these files
  - Option 1) copy the entire GENIE stack in these files
  - Option 2) use metadata cataloguing to retain the association between these files and the corresponding GENIE files
  - Personal opinion: something in between - keep enough information for event reconstruction to benchmark their algorithms, but maybe not enough, e.g., for eventual systematic development

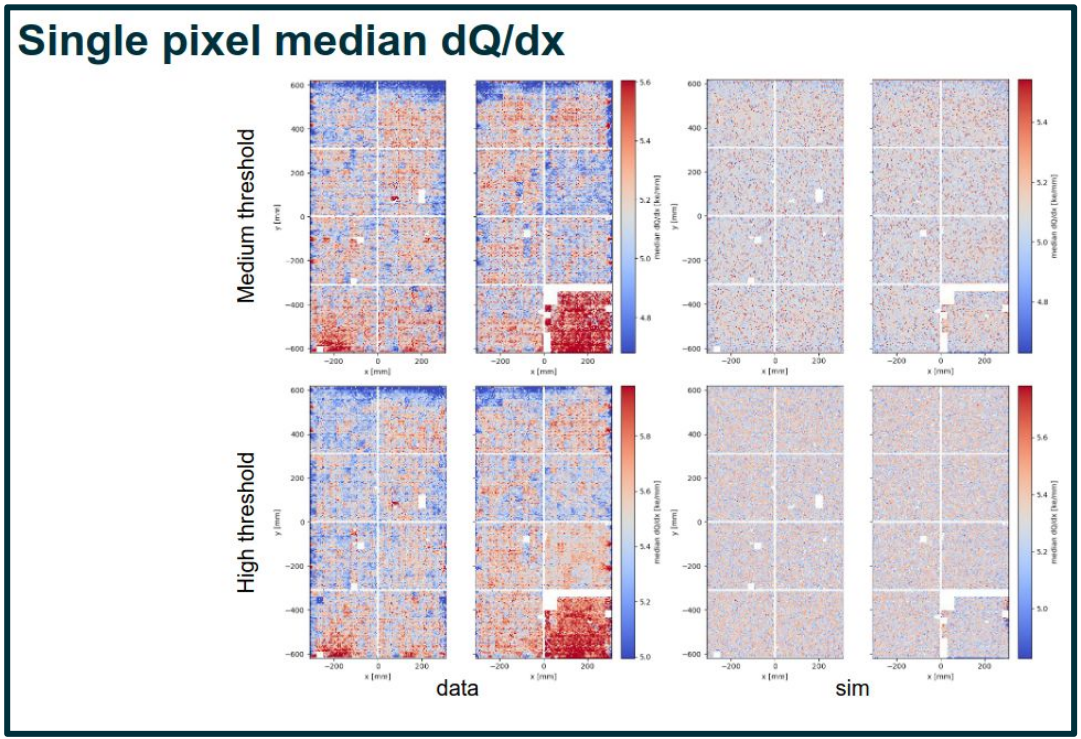
What to Calibrate for?

DEEP UNDERGROUND NEUTRINO EXPERIMENT

P. Madigan

- (Discussion)
- Beyond naive expectation from simulation, data driven

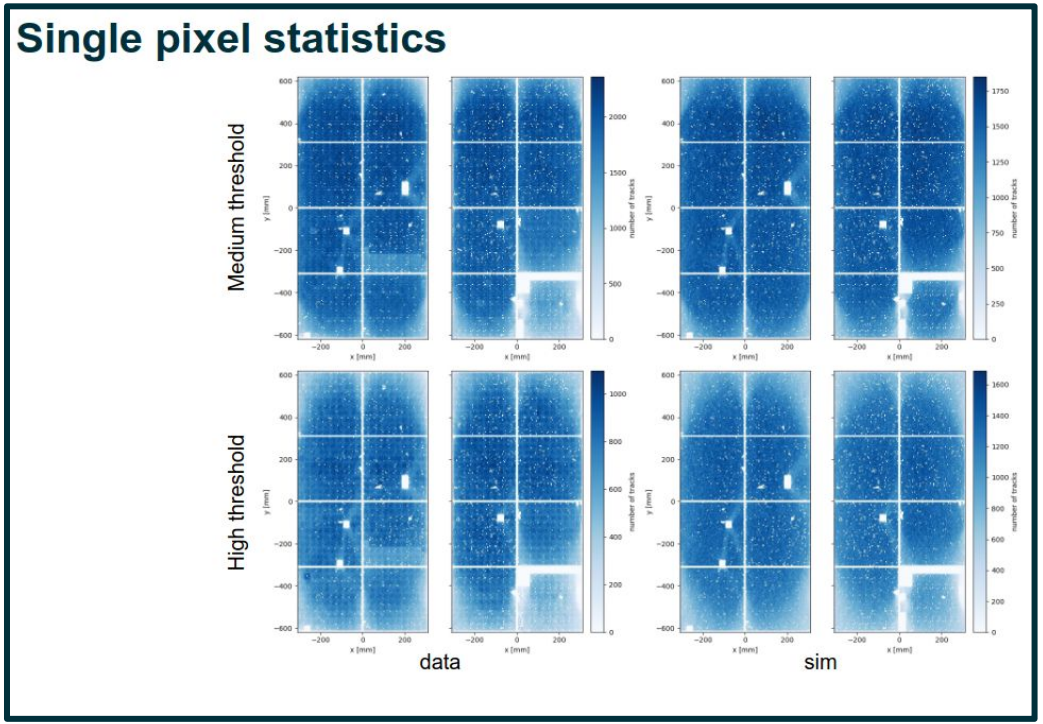
1. Channel-to-channel gain variation
2. Electric field non-uniformity
3. Space charge (?)
4. Detector distortion
5. Run conditions
6. …



What samples to use?

P. Madigan

- (Discussion)
- Rock muons will be plenty
  - Subsamples from selecting, e.g., anode-cathode crossing tracks
- Fewer cosmics, but will be useful to have vertical oriented tracks in addition to the predominantly horizontal rock muons
- Stopping muons will be fewer still, but Michel spectra would be a good validation, especially if we could bin by position in the detector
- Ar39? Requires very low thresholds



Sample Collection

- How often do we collect these samples?
- Self triggering system
  - If we have activity in the detector, we will read it out and save it
- How long does it take to acquire the requisite statistics?
- Matt Kramer is standing up an efficient spill simulation that will be informative here
- Connections with DQM

LAr Calibration Summary

- Status
  - proto_nd_flow for processing and calibrating low level data from 2x2
  - Naive ADC→MeV calibration based on simulation in place
  - Data driven calibration to be developed
  - Sample simulation file with 5 NuMI spills available
- Plans
  - Finalize light and truth datasets
  - Efficient spill simulation with rock muons included
  - Develop data driven calibration modules
  - Collect data!