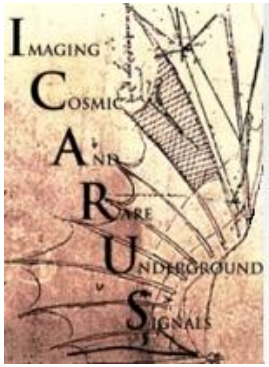


ICARUS Software

Imaging Cosmic And Rare Underground Signals

Preparing For Commissioning

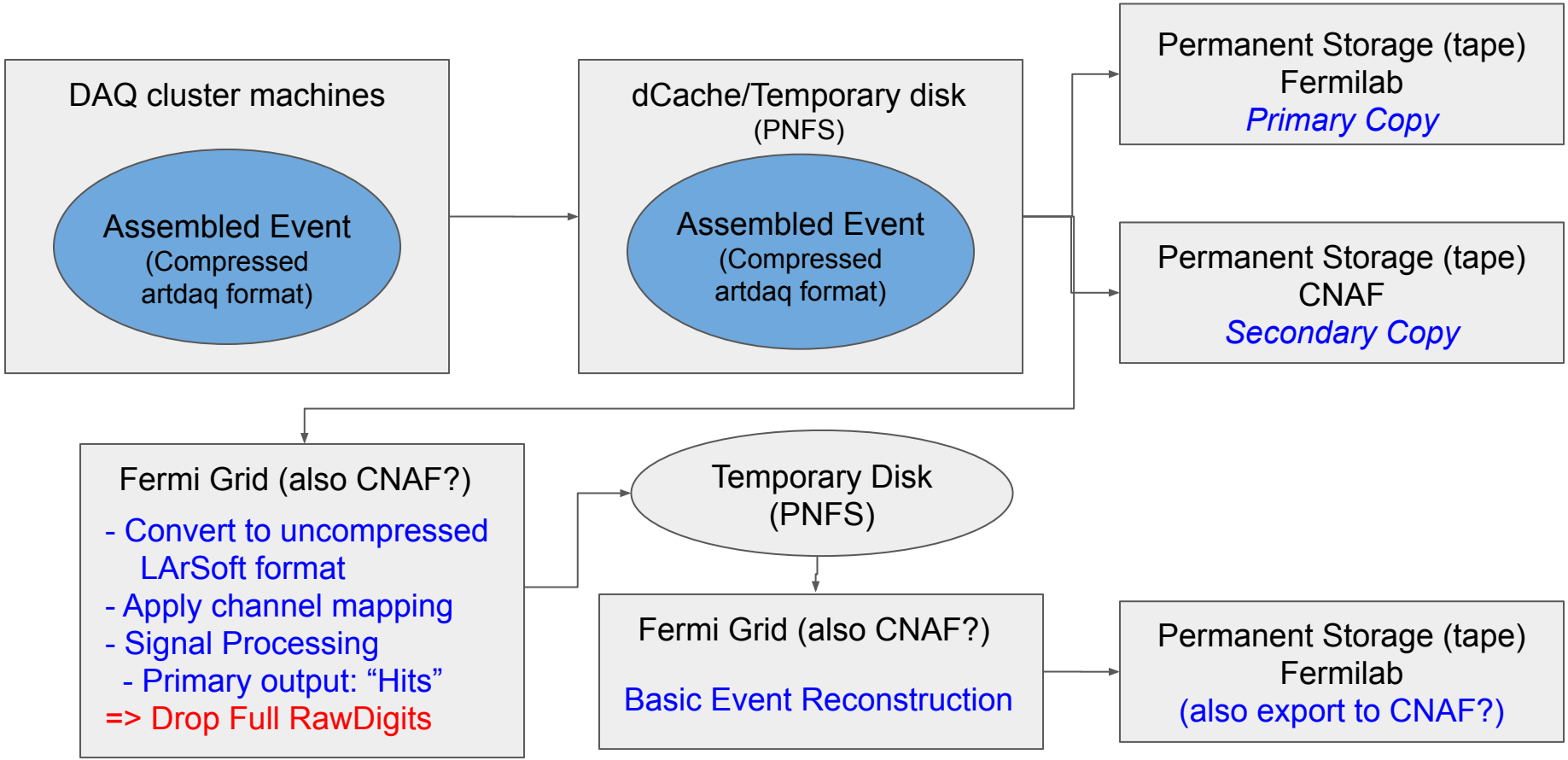
ICARUS Run Readiness Review - January 16, 2020

Tracy Usher (SLAC), Daniele Gibin (INFN Pd)
(for the ICARUS software group)

Outline

- Overview of Data Flow
- Data Transfer and Storage
- Signal Processing
  - Noise Filter, Deconvolution/Waveform Processing, Hit Finding
- Basic Event Reconstruction
- Event Displays and plans
- ICARUS Software Group - who we are, what we do and how you can help
- We want you!

Data Flow - High Level Overview



Data Transfer and Storage

- Data from the daq system in compressed artdaq format
  - Goal to package data in as concise format as possible to minimize storage footprint
- Copied from daq machines to dCache temporary storage in PNFS area
  - Note that daq machines have enough storage for several months of data at nominal rate
- From dCache
  - File written to persistent storage on tape at FNAL as primary copy
  - File copied to CNAF for persistent storage in Italy as secondary copy
  - Pass file to the first stage of the data processing chain
- First point at which data volume can be a concern
  - Estimate for commissioning data volume ~1.5 pB of data stored (no trigger, full rate, 1 month)
  - Estimates for steady operations ~2 pB/year
  - These based on assumption of TPC data compression factor ~8x
    - Currently achieving ~5x compression with test data from October
  - Issue is less about storage, more the time to recover from tape for reprocessing

Signal Processing - Commissioning Plan

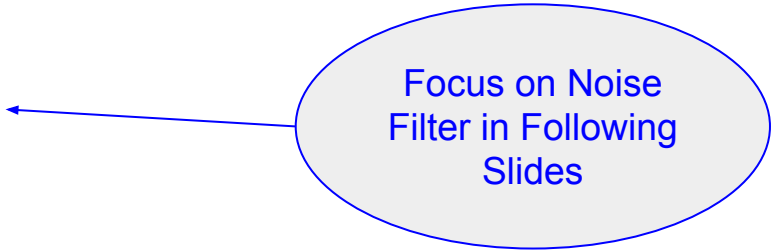
- Need to go from “artdaq” format to LArSoft format
  - Decoder job to convert formats and apply channel mappings
- At this point have uncompressed TPC waveforms
  - ~430 MB/event for the raw waveforms
  - Noise Filtering will make a denoised copy of these waveforms
    - Add another ~430 MB to the resident memory size
  - Deconvolution/waveform processing stage makes another copy
    - Yet another ~430 MB to the resident memory size *per path*
- During commissioning we need to see both the raw and noise filtered data
  - Data processing chain will have to handle the extra volume
- Must note that data volume is not sustainable post commissioning

Signal Processing - Post Commissioning Plan

- The deconvolution/waveform processing can be sparsified by finding “Regions of Interest” (ROI’s)
  - Search waveforms for candidate peaks and block out regions around these
- The downstream processing (pattern recognition, track/shower reconstruction) relies entirely on “Hits”
  - HIts are the reconstruction of the deposited charged in ROI’s found above
    - Peak time, pulse height, total charge, etc.
- Post commissioning plan will be to drop the sets of full waveforms after the signal processing is complete
  - If we need waveforms, we can turn back on
  - We can also write sparse waveforms based on the ROI’s found

Signal Processing - Overview of Current Chain

1. Noise Filtering
    - Remove noise with no/minimum impact to signal
2. Waveform Processing
    - Deconvolution
        - Apply 1D deconvolution to the waveforms
            - Deconvolve waveforms with field and electronics responses
        - Goals: unipolar waveforms with gaussian shaped charge deposits, normalize charge response across planes
    - “Raw”
        - Emulates previous ICARUS processing with software integration of induction plane waveforms to return ROI’s with unipolar pulses
    - Sparsify waveforms - both methods return Regions of Interest rather than full waveforms
3. Hit Reconstruction
    - Deconvolution path - “gaushit” finder
    - “Raw” path - fit candidate peaks to an assymetric shape



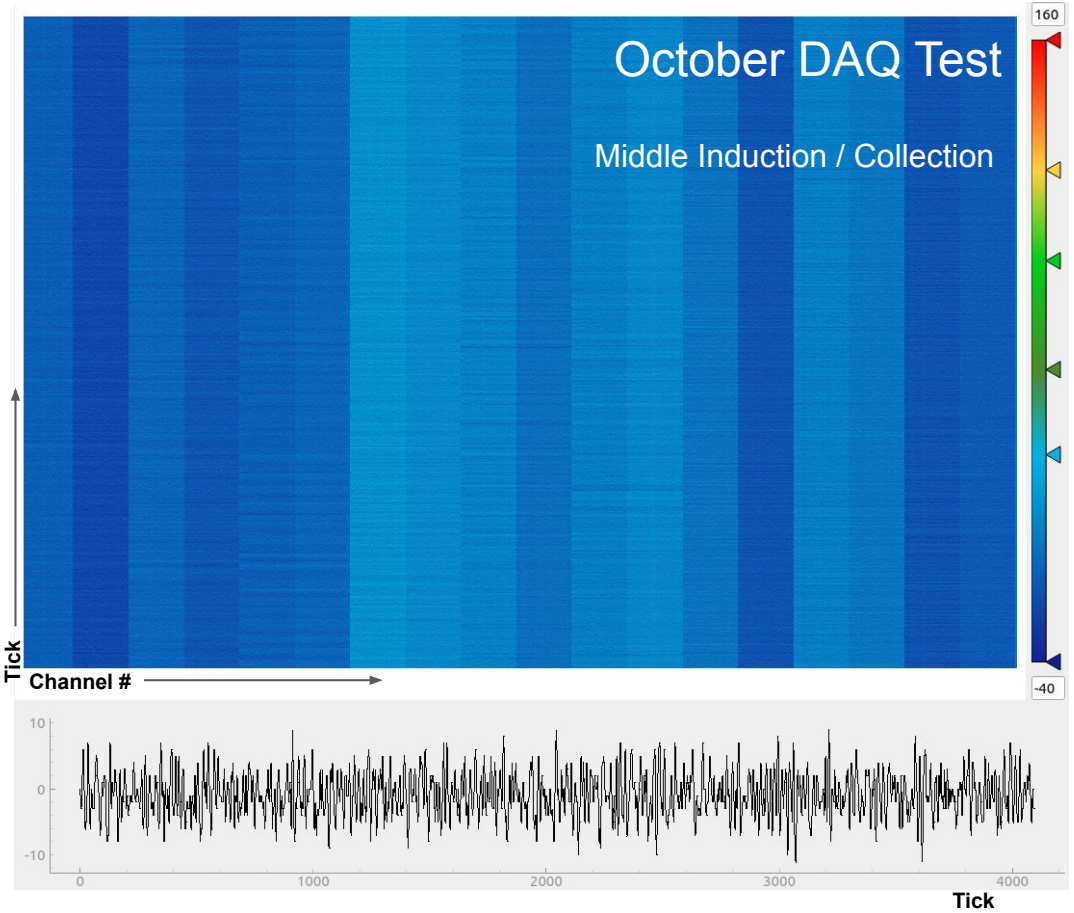
Noise Filtering - Plan of Attack

- Now:
  - Get DAQ test data whenever possible - see next slides
- Startup:
  - Have ported over LArSoft based noise filtering code from MicroBooNE
    - Used before MicroBooNE converted to Wirecell
    - Resident in icaruscode repository - can quickly modify during initial data taking and update reconstruction process without needing full LArSoft release
    - This code has been modified to be thread-safe - can determine if multi-threading will improve overall throughput
  - Augmenting the noise filtering code with ICARUS specific algorithms developed on available test data sets - algorithms in C++ shared between icaruscode and analysis platform
  - During startup will will output the full set of waveforms for continued analysis/development
- Longer term (~Summer):
  - BNL team are working to interface the Wirecell toolkit to ICARUS
    - Used by MicroBooNE and ProtoDUNE

Note: This also brings in 2D Deconvolution

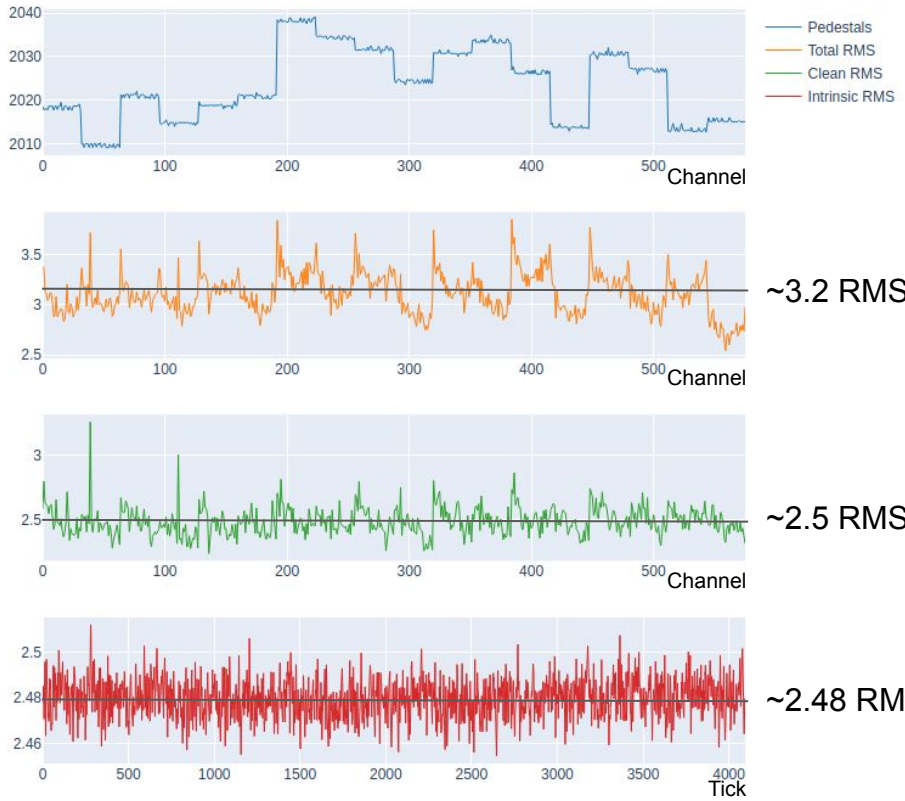
Noise Filtering - First Test Data

- Two files from October DAQ test:
  - 100 events reading 1 mini-crate
    - 576 Channels
  - 1 file middle induction and collection
  - 1 file first induction
- Lots of features!
  - Vertical bands indicating large pedestal offsets
  - Horizontal bands showing coherent noise component
  - etc.
- Can already start to do quite a bit with just this data set!

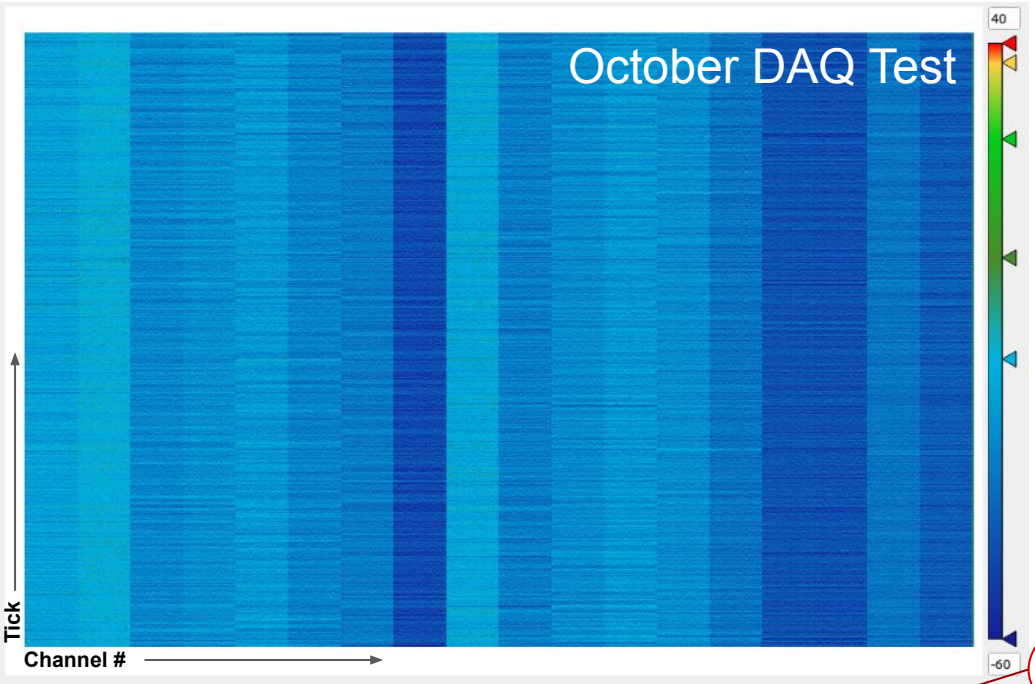


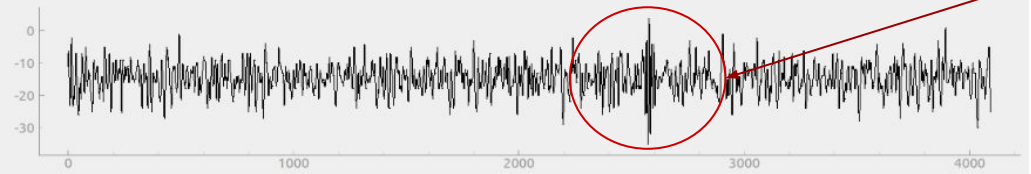
Noise Filter - Middle Induction / Collection Planes

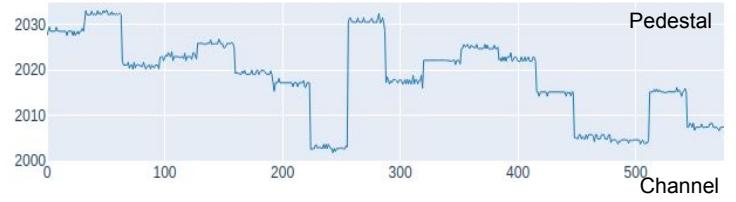
- Average pedestals by channel
  - Averaged over 100 events
- Average RMS by channel
  - Averaged over 100 events
- Average RMS after coherent noise removal (see later)
- Average RMS per tick in blocks of 64 channels - “Intrinsic RMS”

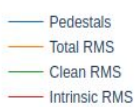


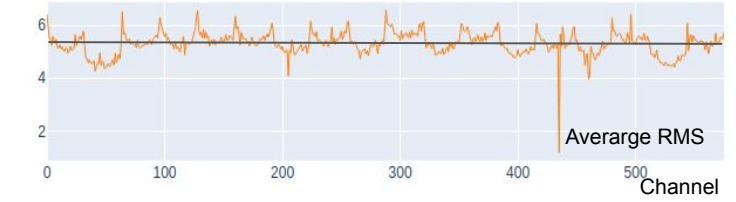
Noise Filter - First Induction





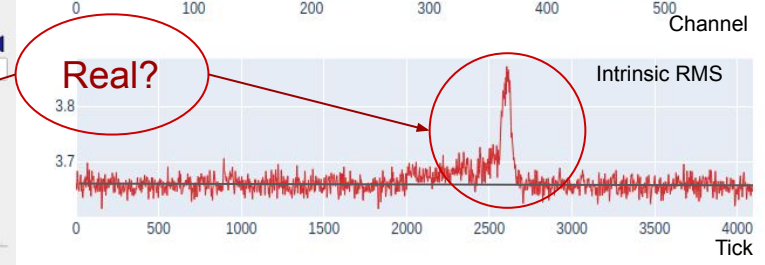






~4.8

~3.6

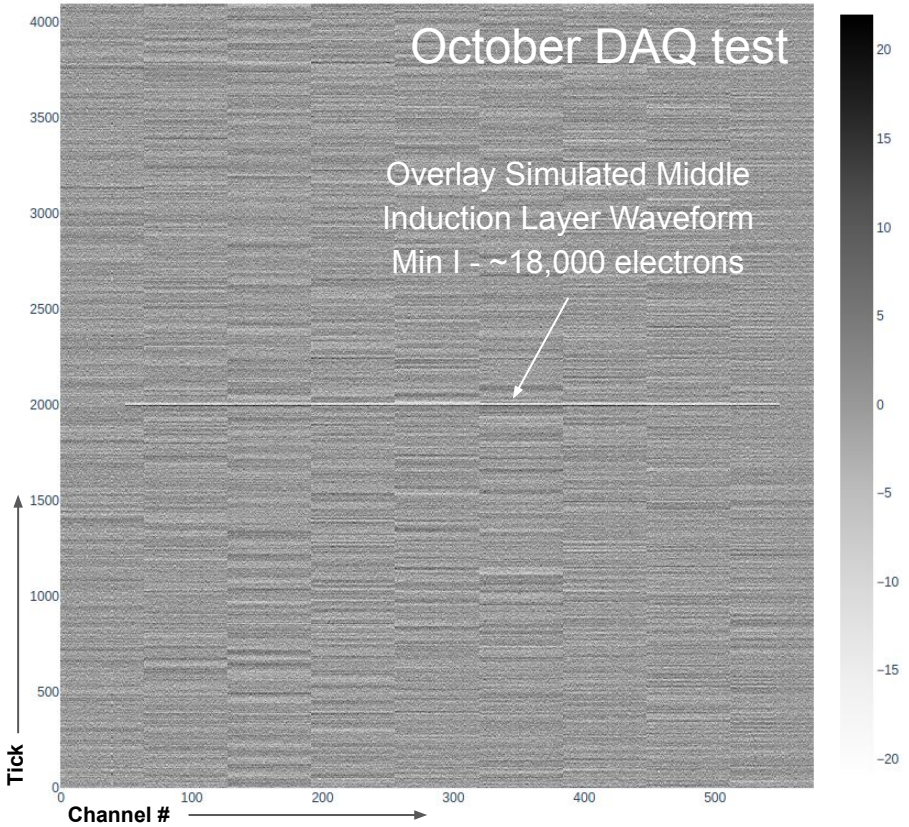


~3.6

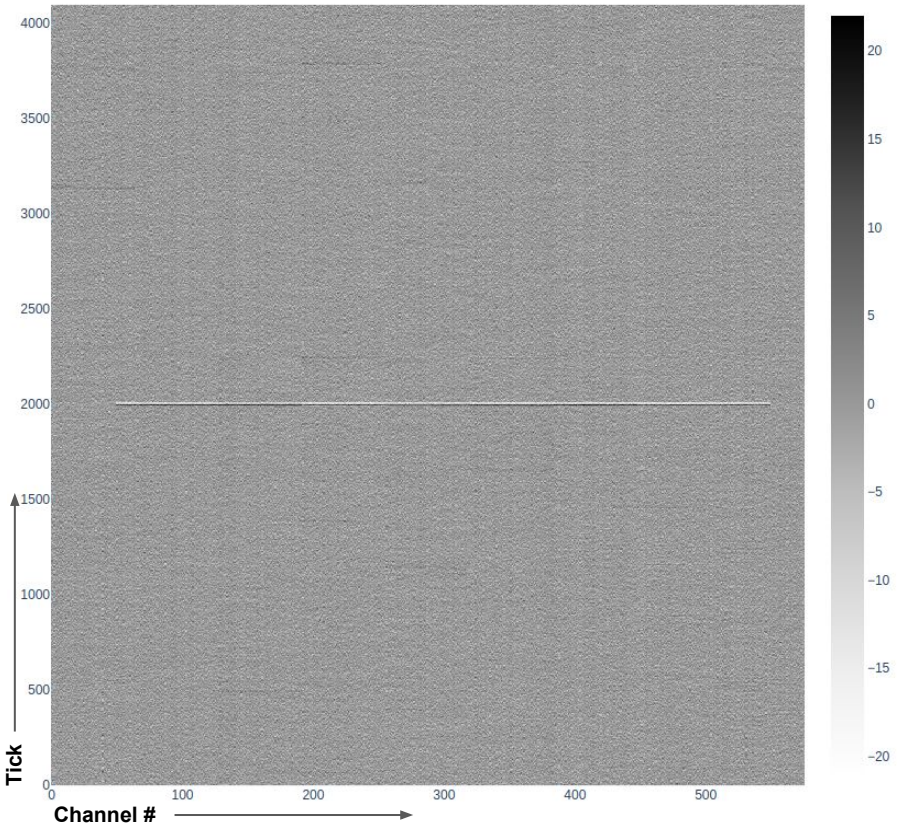
Noise Filtering - Coherent Noise Subtraction

- First perform channel-by-channel pedestal subtraction
- Coherent noise removal:
  - Channels display coherence across blocks of 64 channels
  - Loop over channels in each block and for each waveform tick compute the median
  - Subtract the resulting “median waveform” from each waveform in the block
- Obvious issue:
  - If a track trajectory is running parallel to the wire plane (so at a constant set of ticks in all the waveforms of a block) the above procedure will also subtract out the track!
- How to Handle?
  - Run algorithm to find and “protect” signal regions in the waveforms
  - Study with test data by “overlaying” simulated track on data waveforms
    - Simulate track using convolution of field and electronics responses
    - Choose middle induction response since most challenging to “protect”

Pedestal Corrected Waveforms



Pedestal & Coherent Noise Corrected



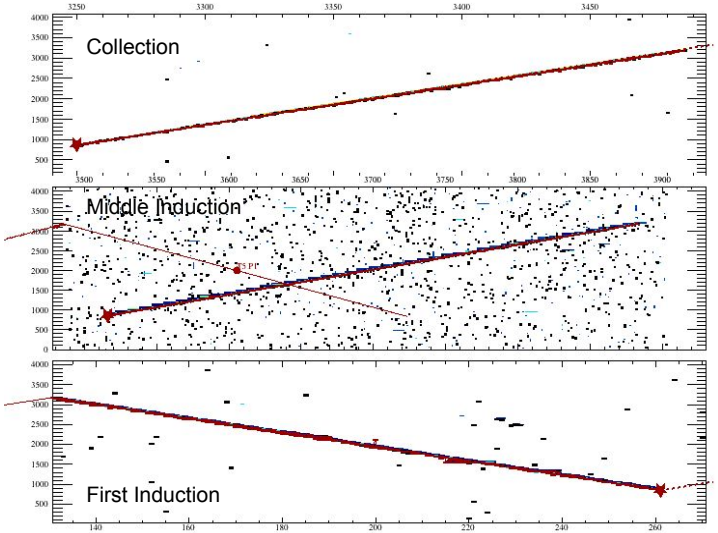
Initial python based 1D algorithm for signal protection. Dae Heun Koh has made significant progress in this area utilizing 2D techniques implemented in C++ - direct transfer to LArSoft code.

Noise Filtering - Initial Conclusions

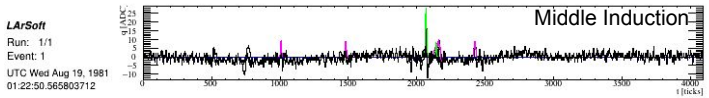
- October test data - take with a grain of salt?
  - Middle Induction / Collection Plane data
    - Pedestals not correct on collection plane data
    - Significant (~30% increase) of rms noise due to coherent noise
    - Simple method for subtracting coherent noise gets close to “intrinsic” limit
    - “Intrinsic” RMS (~2.5) at a bit higher than expected but can live with?
  - First Induction
    - Concern that run conditions not well defined due to significant effect ~2600 ticks
    - “Instrinsic” RMS (~3.6) is much higher than expected and will present challenging environment for good hit efficiency for low pulse height hits.
  - Overall RMS for all planes (with coherent noise) impacts on compression algorithms
    - Is it possible to address this in hardware?
- Now have access to larger data sample with 8 mini-crates, 4608 channels
  - Plan to convert to LArSoft format and then repeat previous analyses
- Significant noise analysis toolset developed/available in jupyter notebooks
  - In github repository - available to anyone

Basic Event Reconstruction

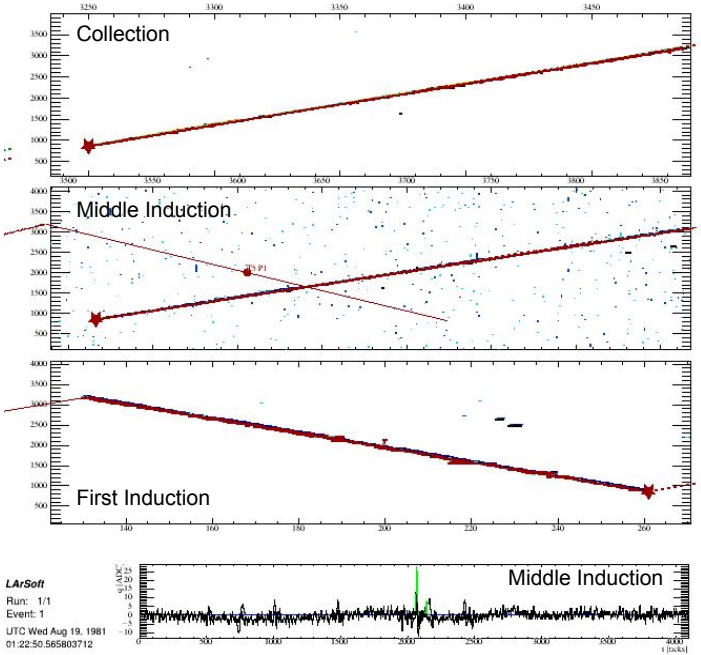
- Data Driven Noise Model in Simulation
  - Based originally on LNGS data
  - Significantly enhanced with data from CERN tests of new electronics (Summer 2018)
  - Includes coherent noise component
  - Developed by Filippo Varanini



- Interesting Feature
  - Model also contains a significant incoherent low frequency oscillation component
    - Unfortunately, cannot run the coherent noise subtraction with the simulated event
  - For deconvolution/gauss hit finding path, extra level of noise leads to spurious hit issues
    - Particular issue in middle induction where we need low thresholds to maintain efficiency
  - Can mitigate this issue post hit finding by forming 3D space points from hits matched in each of the three planes, pass only the hits forming 3D space points to downstream reconstruction
    - Added benefit that can also pass 3D space points to the Machine Learning group...



Basic Event Reconstruction



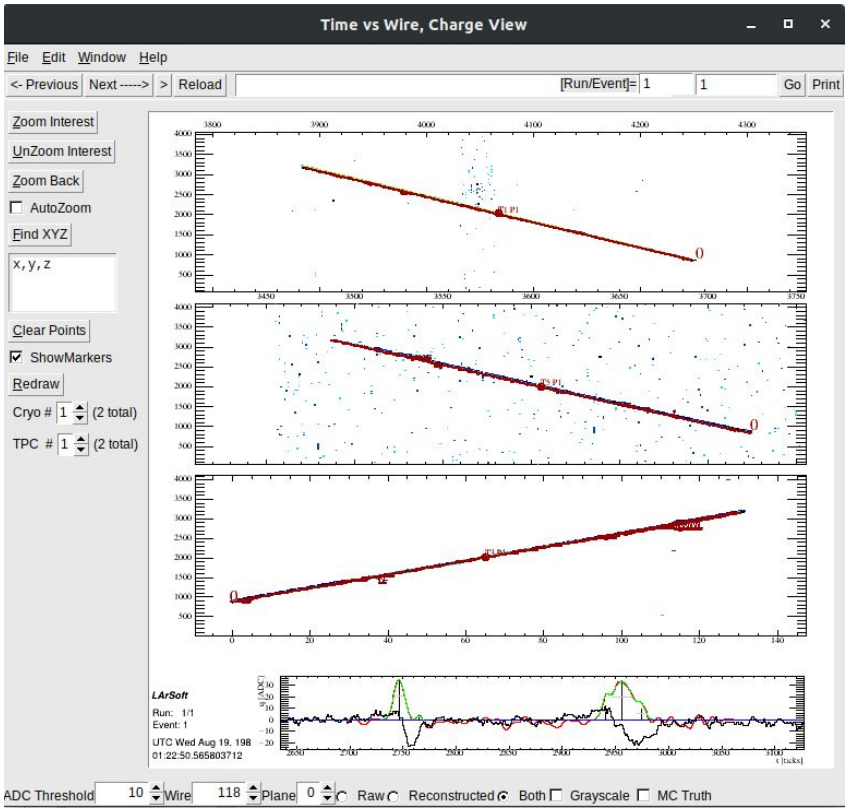
- Basic Path in Common with SBND
  - Input to this stage are reconstructed hits
  - Using the Pandora framework for Pattern Recognition
    - Returns candidate tracks, showers and vertices
  - Track fits with standard LArSoft tools (Kalman Filter)
  - Shower reconstruction with standard “SBN Shower” reconstruction modudle
- Currently Running Along Two Paths:
  - Hits reconstructed from the “Raw” (emulating original ICARUS signal processing) path
  - Hits reconstruction from the 1D deconvolution and “gaushit” finding path
- Notes:
  - We are running Pandora “out-of-the-box” with only cursory tuning of parameters for ICARUS!
    - Need to develop ICARUS specific experts to start fine tuning to maximize performance
  - At this point we are not running calorimetery and/or particle ID as part of standard recon

Event Displays

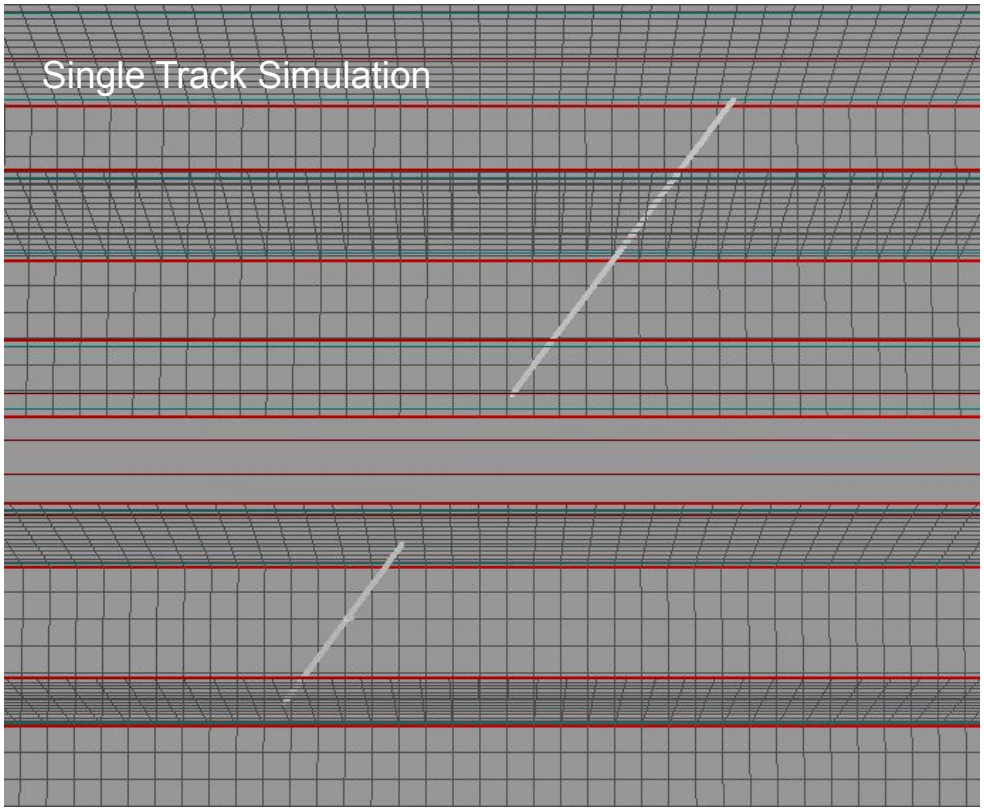
- Needs
  - Data taking: Want fast high resolution 2D display for studying waveform data
  - Data Analysis:
    - High resolution 2D display with capability to display simulated and reconstructed objects
    - 3D display with capability to display simulated and reconstructed objects
- What is available
  - LArSoft Event Display (root based)
    - In principle can display everything both in 2D and 3D
    - Very slow, almost unusable over network, even fgz from GPVM at FNAL!
  - QT based event display
    - Developed originally by Corey Adams for MicroBooNE, recently significantly updated in the context of ICARUS and SBND by Marco Del Tutto with help from Gianluca Petrillo
    - Very fast, easy to use and will be good tool for commissioning and data taking
  - Others available for ICARUS soon
    - Wirecell “Bee” event display (Chao Zhang)
    - Eve based event display (Umut Kose)

Event Display - LArSoft Example

2D Display (Single TPC)

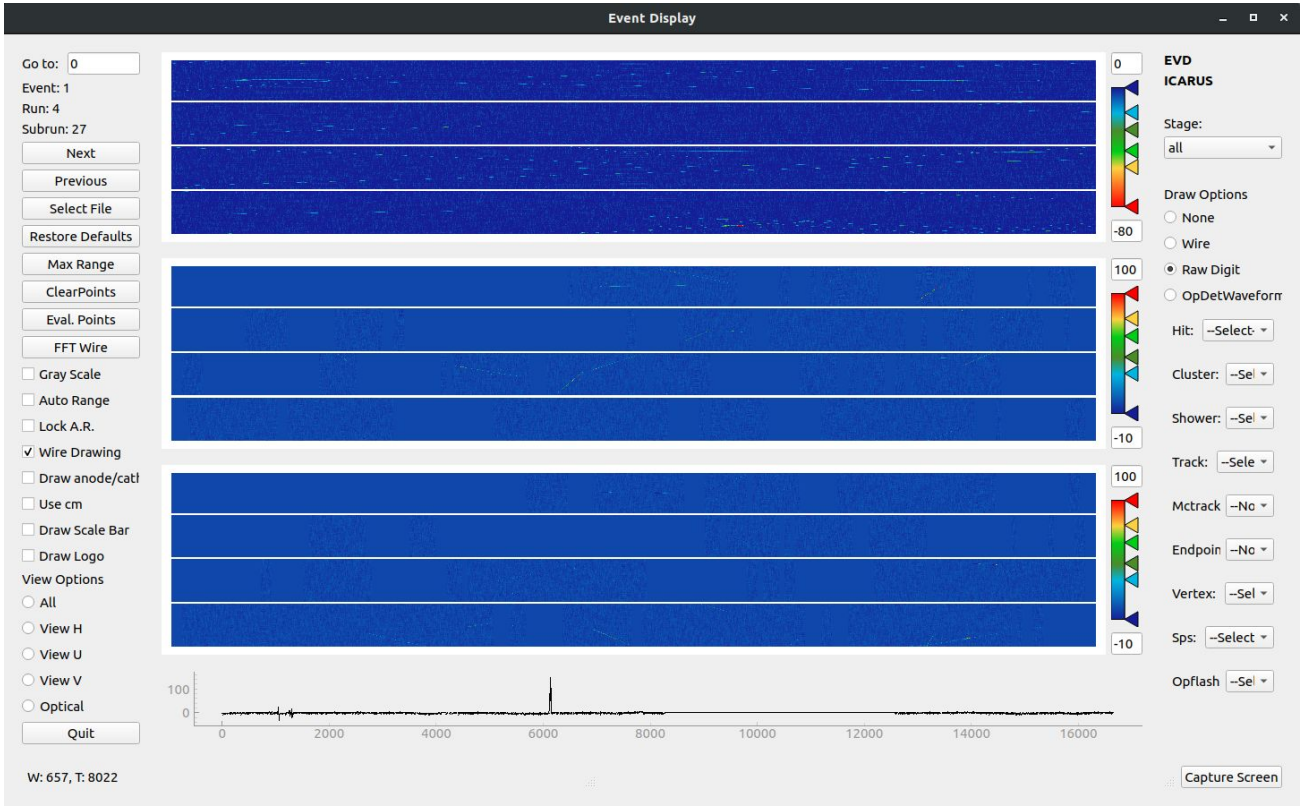


3D Display



Event Display: QT Display

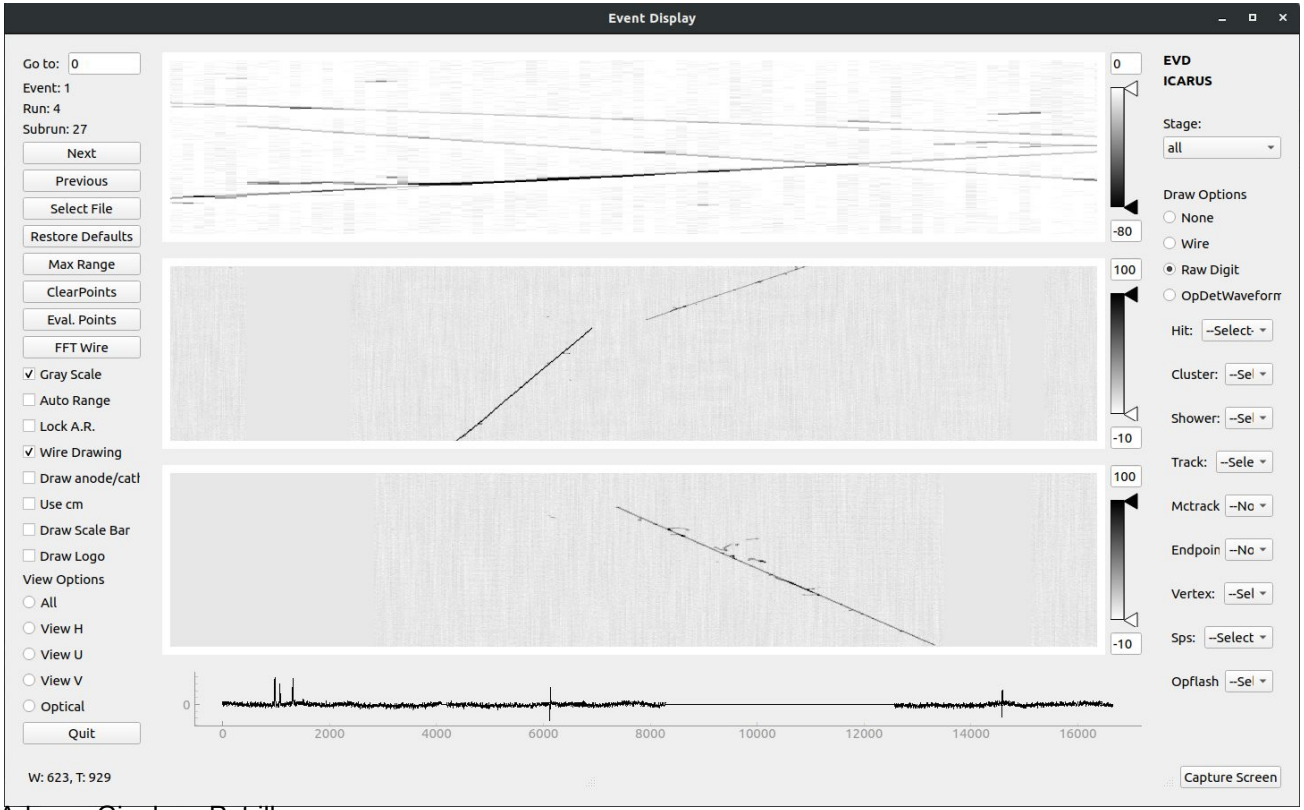
2D Display - All Four TPCs



Marco Del Tutto, Corey Adams, Gianluca Petrillo

Event Display: QT Display

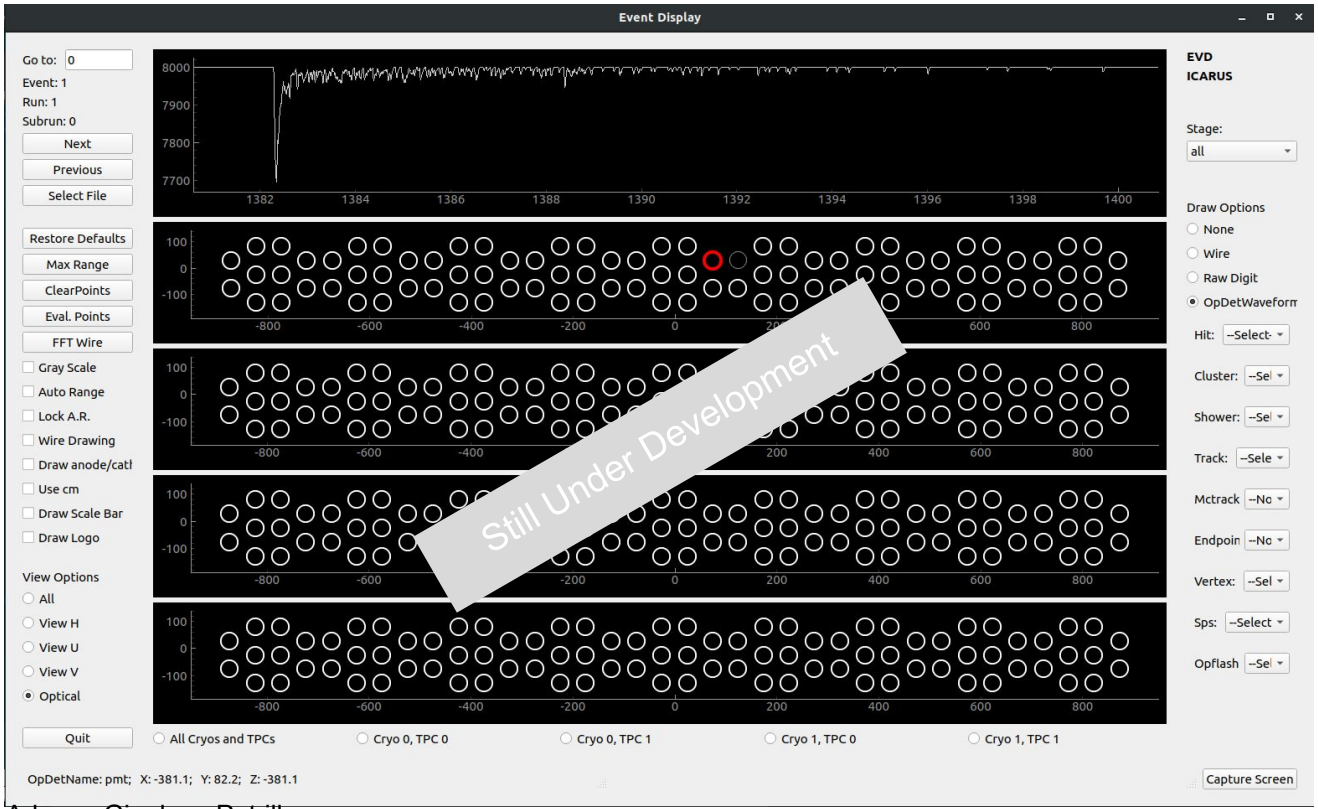
2D Display - Zoomed to Single TPC



Marco Del Tutto, Corey Adams, Gianluca Petrillo

Event Display: QT Display - Optical

Optical Display



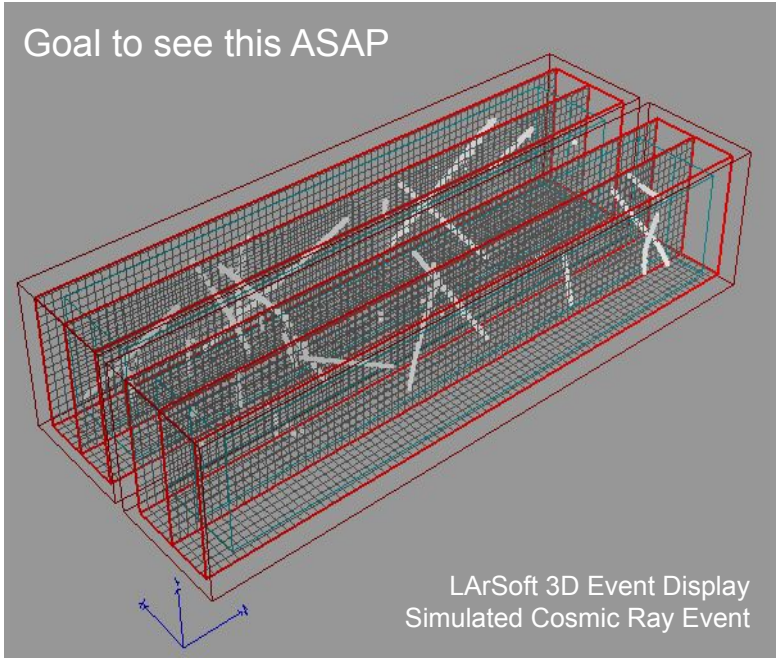
Marco Del Tutto, Corey Adams, Gianluca Petrillo

Event Display - QT Display Plan for Commissioning

- Snapshots on preceding pages don’t do usefulness of this display justice
  - Those who have used with MicroBooNE know it works well
  - For others happy to arrange live demo

- Plan to have this set up in ROC West
  - Idea is to have process running on local machine to access incoming DAQ data
    - Decoder/Signal Processing process running to convert to LArSoft format and provide noise filtered data
  - Output of process will be accessed by event display as available
    - New event ~10 minutes?
    - Can be faster if we don’t do noise filtering first…
  - Display will have pause option so people can focus on interesting events
  - Marco Del Tutto taking lead on implementing this

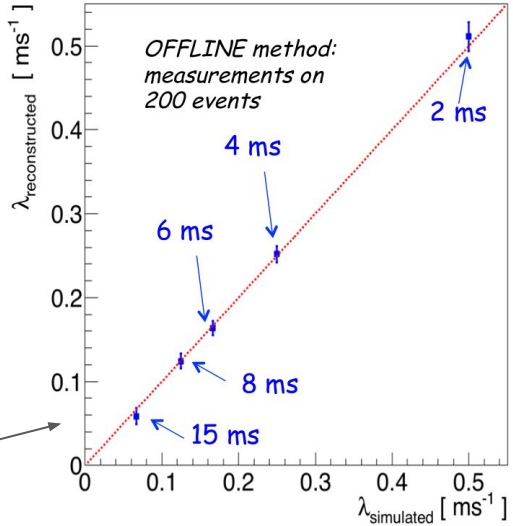
TPC Commissioning



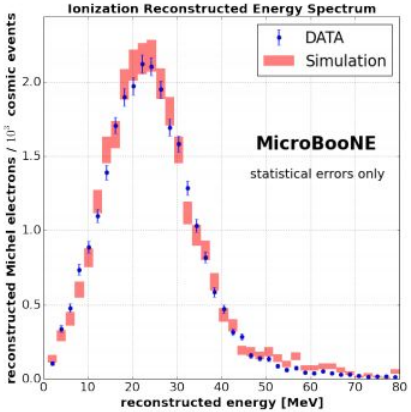
- Before HV on
  - Study noise data, verify algorithms operational
  - Monitor purity (DQM)
- After HV on
  - Study noise data, verify algorithms operational
    - Hopefully no surprises!
  - Adjust bias voltages?
    - Maximize transparency of induction planes
    - Cross check interplane drift time (necessary for high efficiency 3D tracking)
  - Start looking at tracking performance
    - Finding hits with good efficiency, understanding interplane timing, etc… tracking should “just work”

With Tracking - Next Steps

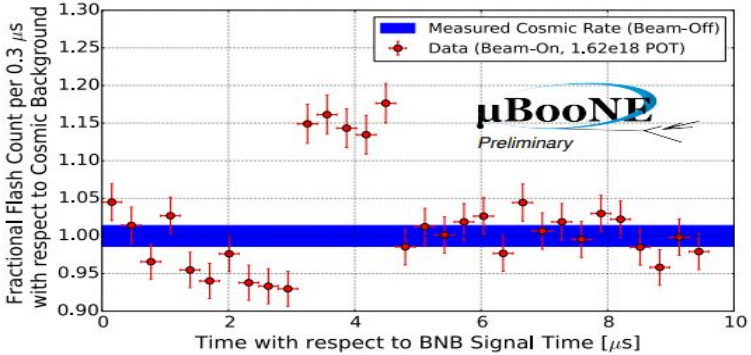
- Offline measurement of argon purity
  - First get calibration constants for each TPC channel
    - Track the dE/dx for cosmic ray tracks for each channel
    - Can probably reach sufficient resolution with ~5k CR tracks
  - Develop a sample of anode-to-cathode crossing tracks
    - Know t0 unambiguously in this sample
  - Goal to see where we are on this plot
  - Christian Farnese/Calibration Group (see this docdb presentation)



- Michel Electron spectrum
  - Tag Michel electrons from stopped muon decays
  - Get electron energy spectrum with comparison to simulation
  - Goal to be able to produce this plot
  - Kazu Terao/Laura Domine/Francois Drielsma (ML group)



PMT Group Commissioning Plan



Goal is to make this plot!

(before BNB shutdown)

Plan (once system on and data flowing):

- Check baseline mean and rms for all channels. Check stability over time (throughout 24 hours, then maybe a week)
- Calculate SPE area and amplitude gain from cosmic sample per PMT, check the mean is where expected + quantify the spread (per channel as well as channel-to-channel). Check area/amplitude ratio across PMTs to find any weird pulse shape outliers. Start with one day (or hours) of data (should be decent statistics)
- Quantify OpHit rate per PMT as well as charge spectrum (i.e. area or amp per OpHit). Maybe roughly tune OpHit reco parameter if it's terrible (like threshold too low producing too much hits or something). Identify outlier pmts if any.
- Quantify OpFlash rate and PE spectrum. Make sure re-flashing is not happening frequently (if so, increase the flash finding threshold). Make sure the rate is "reasonable": we have a guess from expected flash rate from MC for an expected rate of cosmic rays going through our detector. Should be able to observe MIP peak in flash PE spectrum for a ball-park estimation of light yield (~3m muon track quantified from simulated cosmic ray samples).
- Ready to make beam timing plot.

Software Group - ICARUS Specific Activities

- Release Management
  - -Your name here-
- MC/data Production
  - Maya Wospakrik, Francois Drielsma
- Infrastructure (Geometry, Properties, etc.)
  - Alessandro Menegolli, Gianluca Petrillo
- Signal Processing
  - LArSoft: Filippo Varanini, Dae Heun Koh, Mike Mooney
  - Wirecell: Chao Zhang, Andrea Scarpelli, Wenqiang Gu
- Decoder Process
  - Bishu Behera, Bruce Howard, Gianluca Petrillo, Wes Ketchum

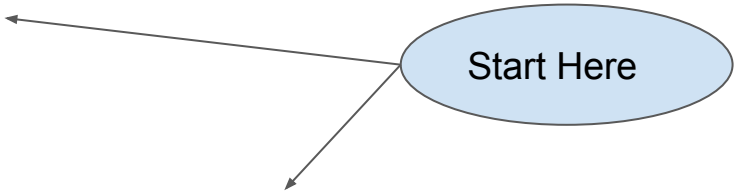
Plus additional activity in joint ICARUS/SBND groups - e.g. event selection (See Daniele's PAC presentation)

- TPC Simulation/Reconstruction
  - Filippo Varanini, Yun-Tse Tsai, Bruce Howard, Bishu Behera, Christian Farnese, Ryan Lazur, Mike Mooney
- PMT Simulation/Reconstruction
  - Kazu Terao, Laura Domine, Dae Heun Koh, Francois Drielsma, Marta Babicz, Alessandro Menegolli, Gianluca Petrillo
- CRT Simulation/Reconstruction
  - Chris Hilgenberg, Umut Kose
- ML Reconstruction Group
  - Kazu Terao, Laura Domine, ...
- Calibrations
  - Christian Farnese
- Calorimetry and Particle ID
  - Minerba Betancourt

ICARUS Software - Where to Get Information

Conveners: Daniele Gibin (daniele.gibin@pd.infn.it), Tracy Usher (usher@slac.stanford.edu)

- Mailing Lists (see https://listserv.fnal.gov/):
  - Primary - icarus_reconstruction@fnal.gov
  - Less Active - icarus-software@fnal.gov

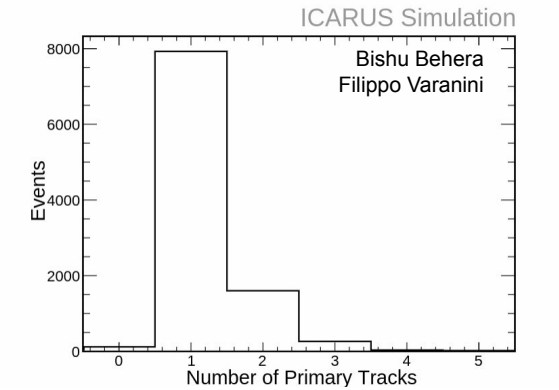


- ICARUS Specific Meetings
  - General Software meeting: Monday 9:00 am CST (WH12SE, zoom: 171495393)
  - PMT Reconstruction: Thursday 11:00 am CST (zoom only, icarus-oprec@fnal.gov)
  - Machine Learning Reconstruction: Wednesday 11:00 am CST (zoom only, icarus-ml@fnal.gov)
  - ICARUS Calibration: Tuesday 12:00 pm CST (zoom: 3288157393)

- Some active and relevant Joint ICARUS/SBND Meetings:
  - Shower Reconstruction: Wednesday 12:30 pm CST (zoom: 9799707990, sbn-shower@fnal.gov)
  - TPC Simulation/Calibration: Friday 10:00 am CST (zoom: 3288157593, sbn-tpc-sim@fnal.gov)

- SLACK chat - icarus-sbn.slack.com

Examples of Some Current Issues

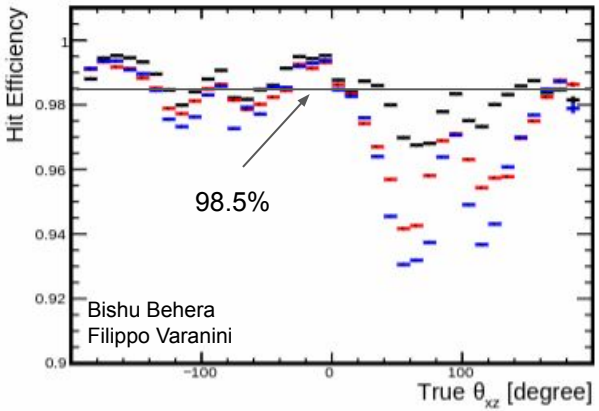


Broken tracks: Single particle simulation, Pandora returns 1 track in ~80% of events, no tracks ~1% of events, >1 track ~20% of events

- Need to understand tracking failures (phase space for this data set?)
- Some fraction of >1 tracks are crossing cryostat boundaries
  - Can these be matched to form a single track
- Some fraction are “broken”, understand how to “fix” these

Efficiency for finding hits in first, middle and collection planes

- Good: average hit efficiency in range 98-99%
- But… plot vs angle in the x-z plane demonstrates interesting behavior
  - Expect a variation as the angle changes from parallel to the z axis (length of the TPC) to parallel to the x axis (drift direction)
  - However, expect this to be symmetric but observe asymmetry!



⇒ Interesting problems like this for people to get involved with!

WE WANT YOU!