Simulation and Reconstruction

Alex Himmel, Fermilab

LBNC Meeting

June 22nd, 2017

Introduction

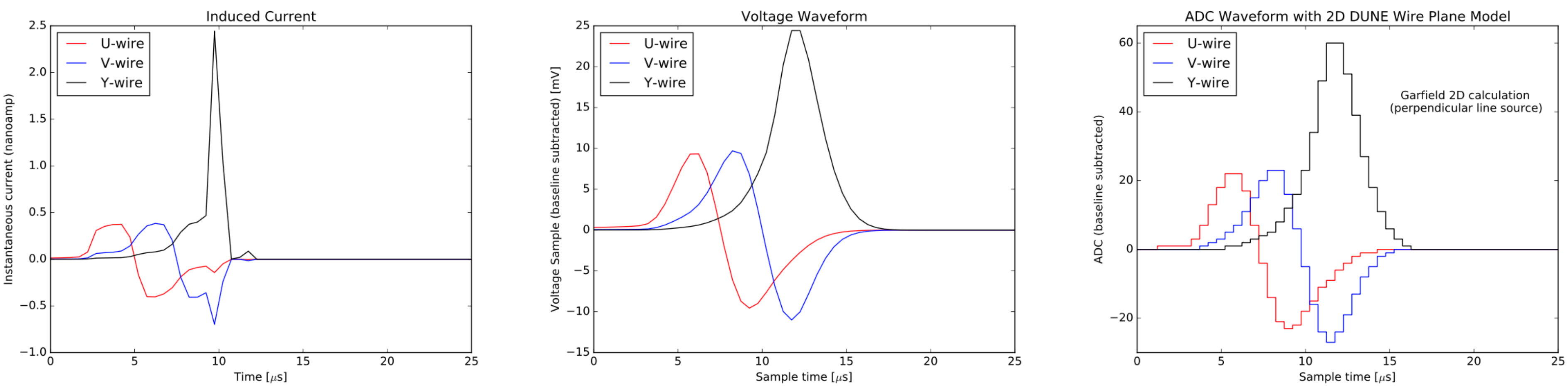
- Covers a broad range of detectors:
  - Single and dual phase detectors
  - Prototype and full-sized far detectors
- Generally 3 areas of responsibility:
  - Detector simulation (TPC and photon detectors)
  - Low-level reconstruction (hits and tracks)
  - Support for high-level reconstruction (PID and energy)
- This talk:
  - Primarily an updated on recent progress in the 3 areas above.
  - End with a look-ahead at plans in the next 6 months.

Detector Simulation

TPC and Photon Detectors

TPC Simulation

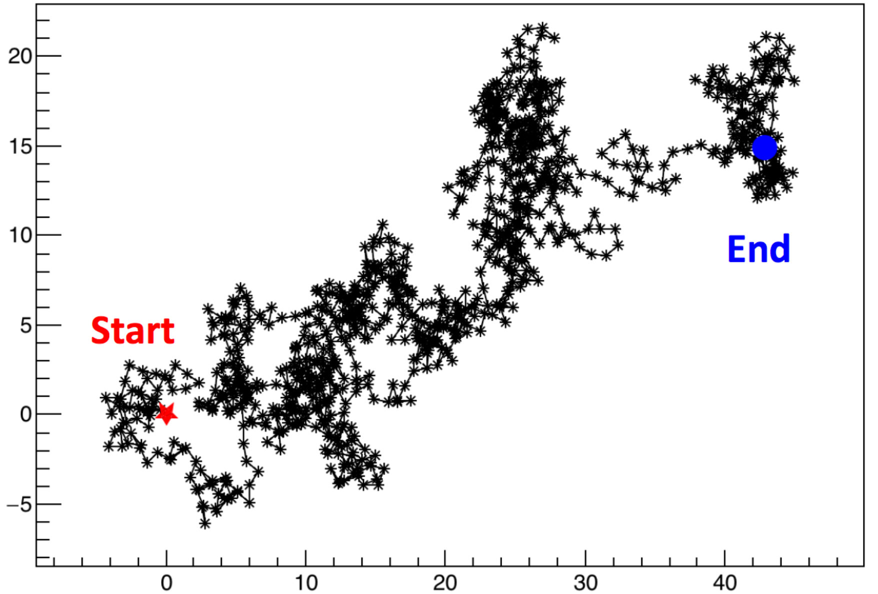
- Recent work has generally been focused on improved modeling of noise and electronics response.
  - This has been guided by experience with data in MicroBooNE and other running TPCs.



Induced current, amplified and shaped voltage and digitized ADC due to ideal isochronous, MIP track (used: 16k e^-/pitch).

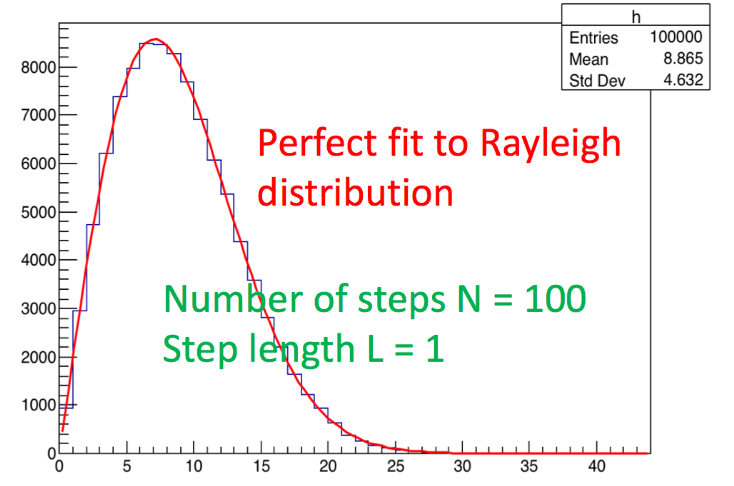
TPC Simulation

- White noise frequency distribution can be modeled using a random walk in the complex plane.
  - Random Walk (2D 2π directions, fixed step length)



The distance between 'Start' and 'End' follows

Distance between 'Start' and 'End'



$$\sigma^2 = 0.5 \cdot N \cdot L^2$$

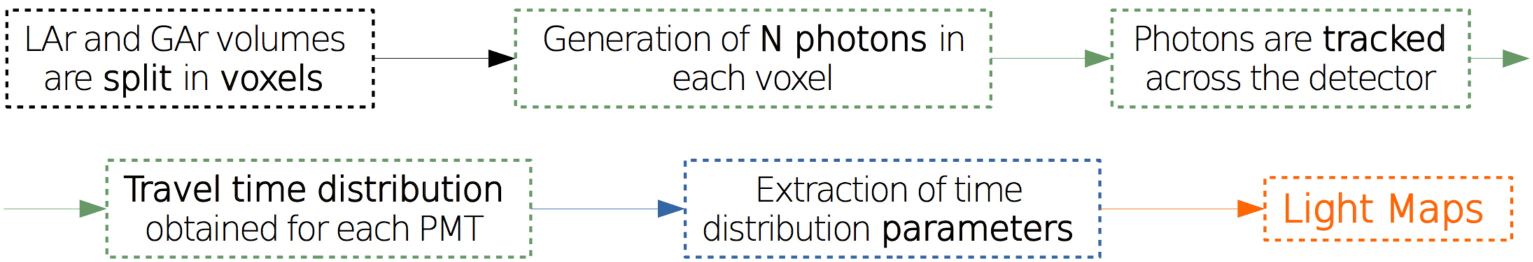
$$f(x;\sigma) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \text{ (Rayleigh Distribution)}$$

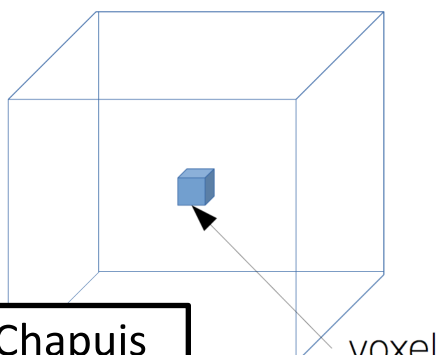
$$\text{Mean} = \sigma\sqrt{\pi/2}, \text{ Mode} = \sigma$$

Hanyu Wei

Photon Simulation

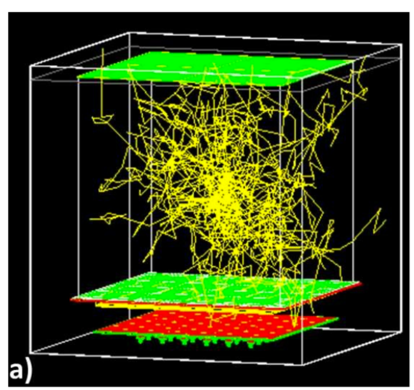
- Exploring cooperation with the dual phase simulation effort.
  - We have independently developed similar techniques.
  - Some ideas we can adopt in LArSoft:
    - e.g. Interpolation of voxels



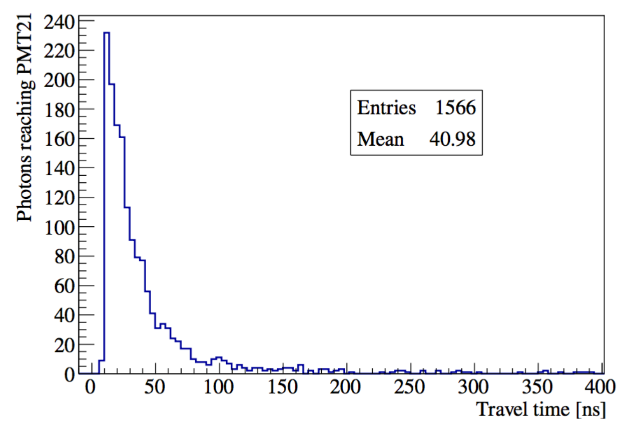


Anne Chapuis

Production point



Tracking

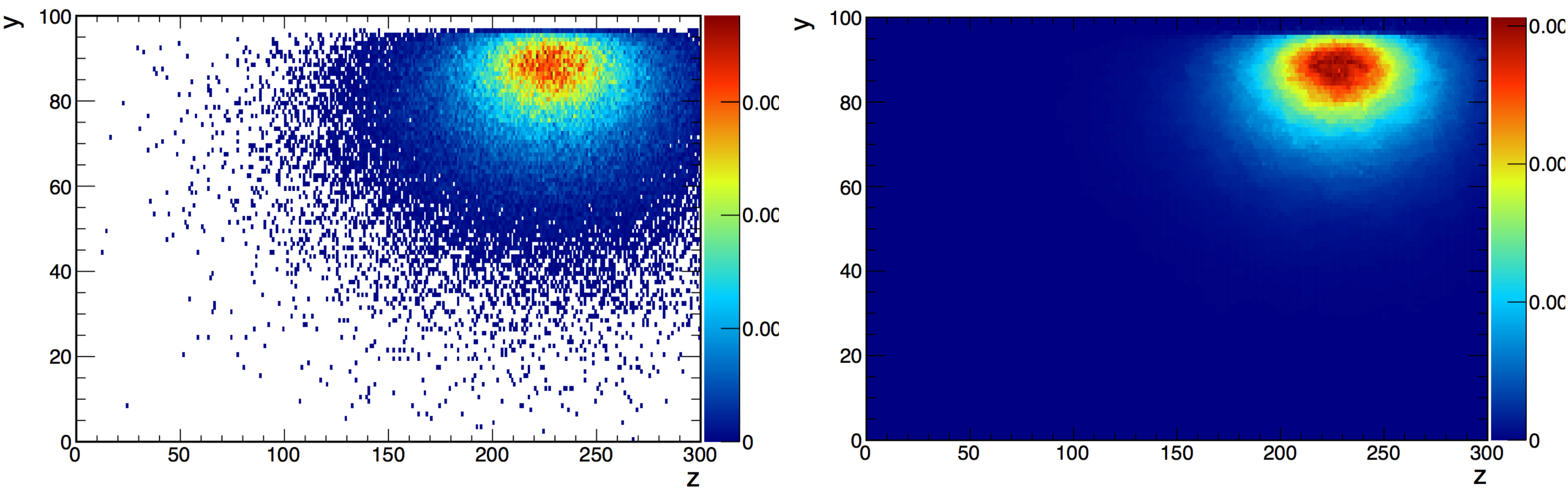


Travel time distribution

Photon Simulation

- Photon libraries cannot scale to the full-sized far detector.
  - They take up too much memory for any geometry larger than the 1x2x6 workspace.

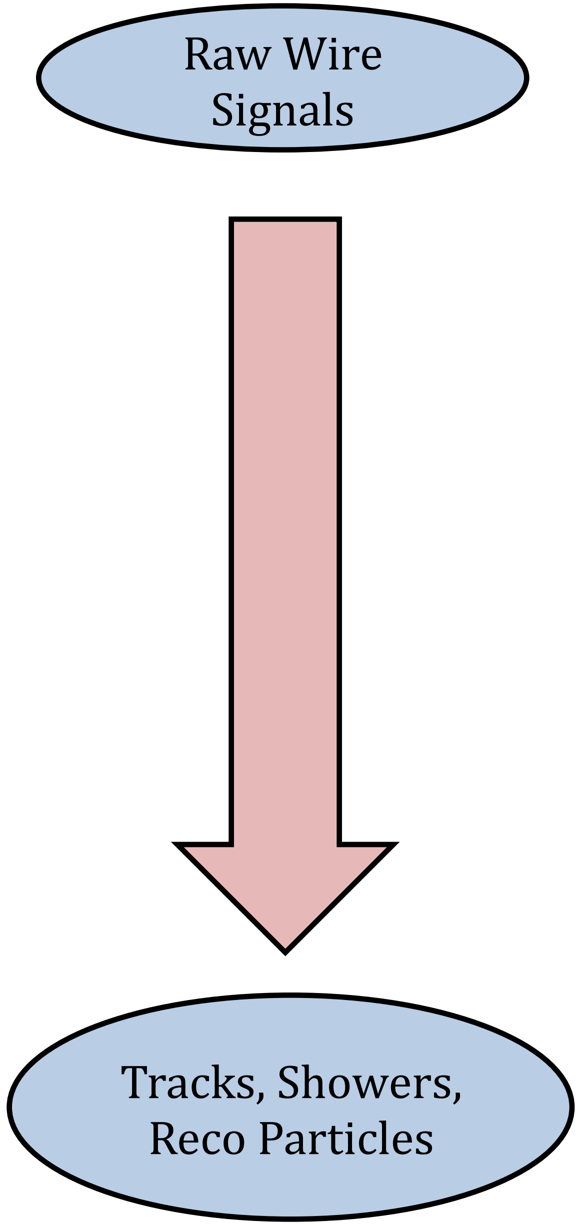
- Exploring an alternative with lower memory use and a smoother response – decision trees.
  - First iteration reduces memory use by 1/3, but many new ideas can be explored.



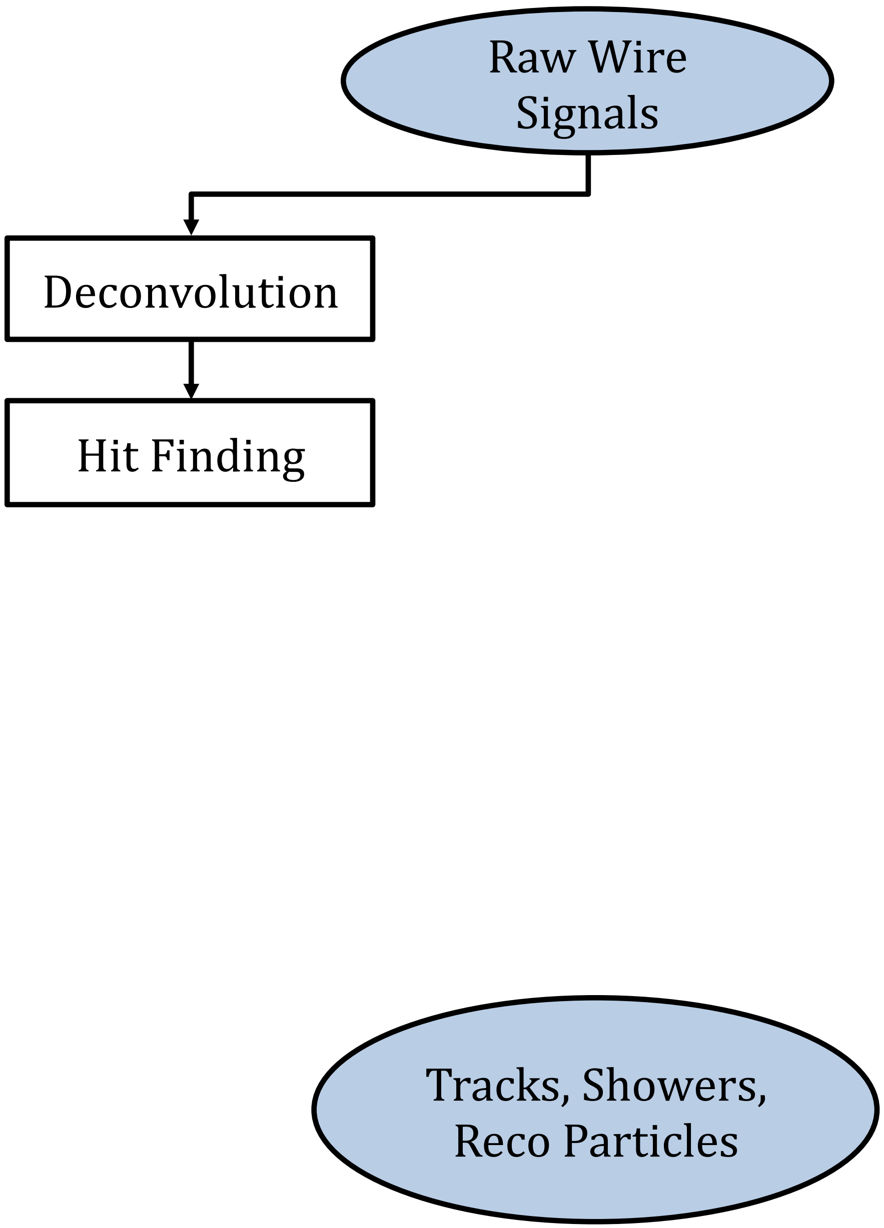
Chris Backhouse

Low-Level Reconstruction

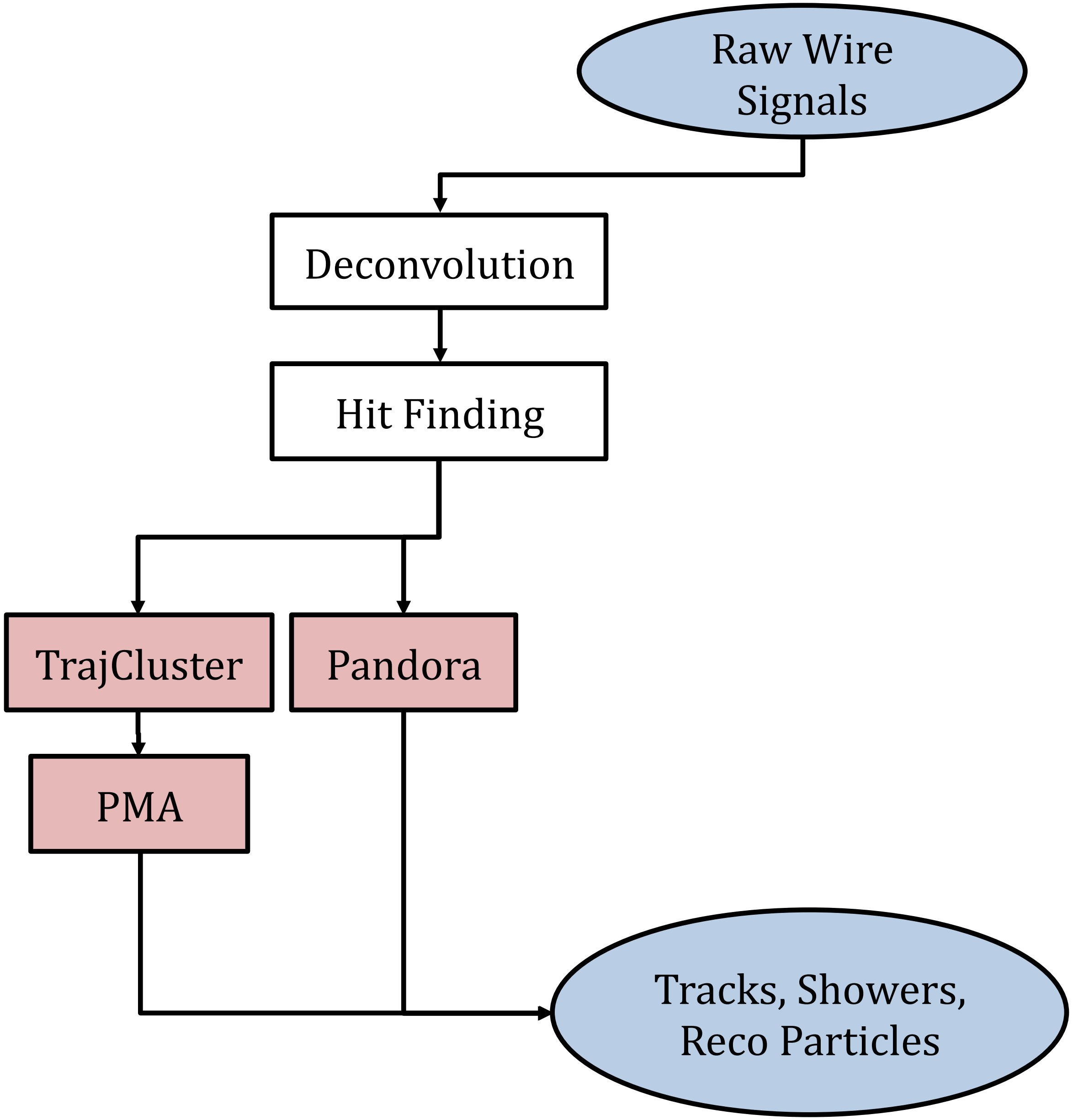
Reconstruction Algorithms



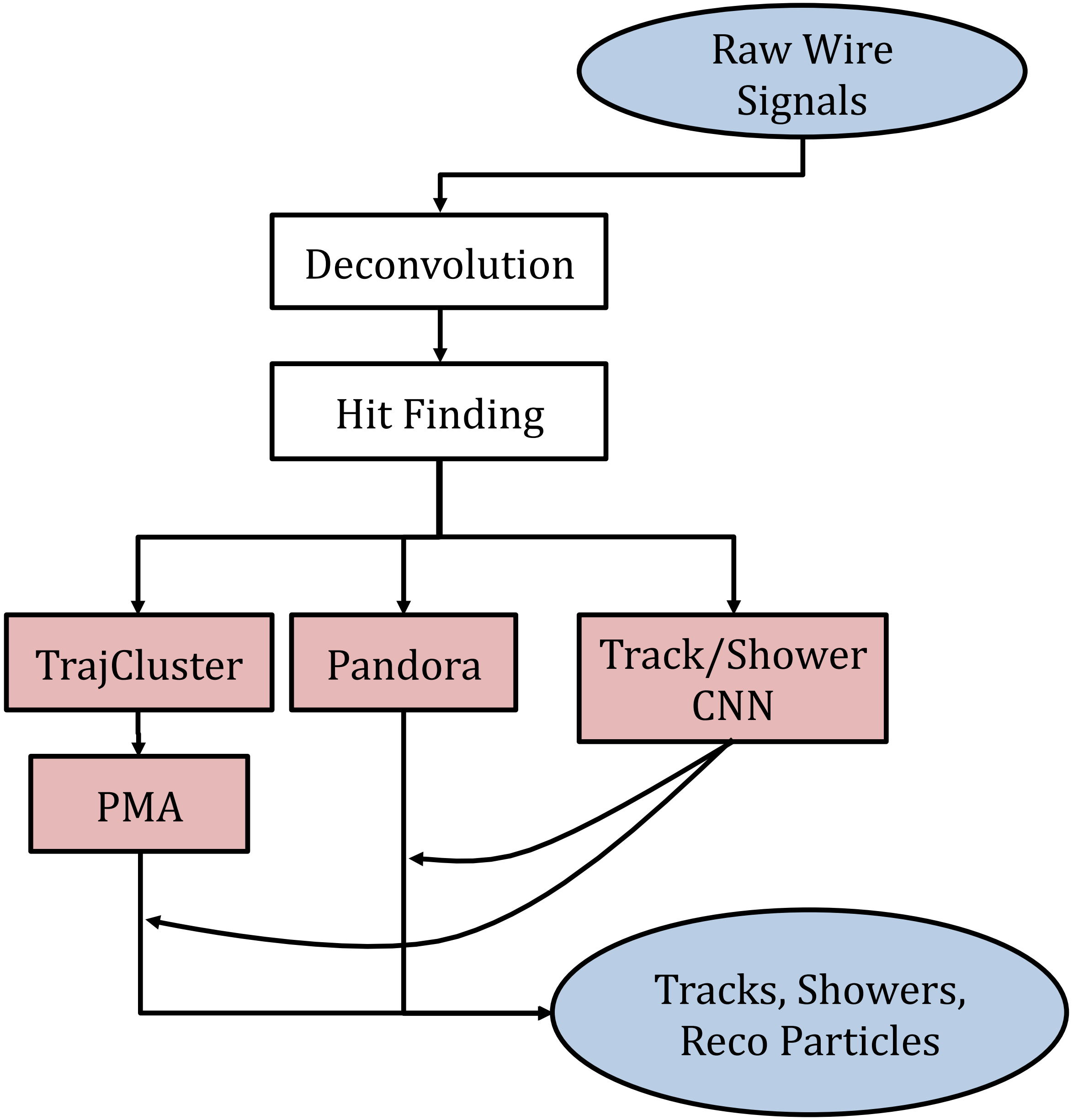
Reconstruction Algorithms



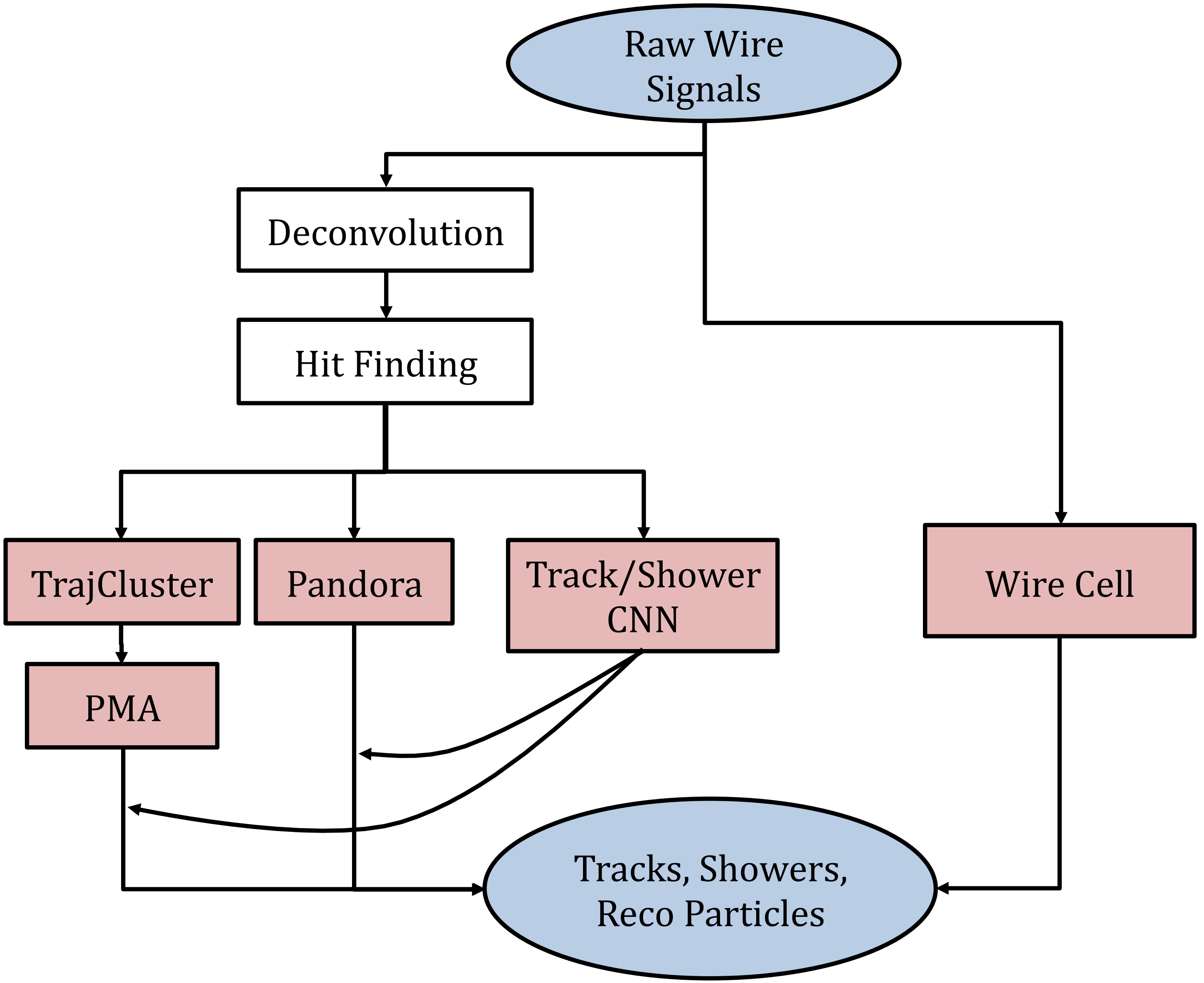
Reconstruction Algorithms



Reconstruction Algorithms



Reconstruction Algorithms



What is Pandora

Pandora Team incl. Nick Grant, Lorena Escudero, Andy Blake, John Marshall

- Multi-algorithm, particle-flow approach.

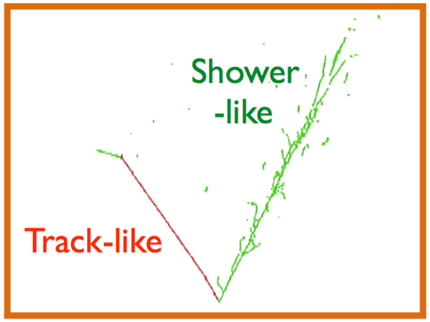
Basic flow of PandoraNu reconstruction, skipping over some subtleties:

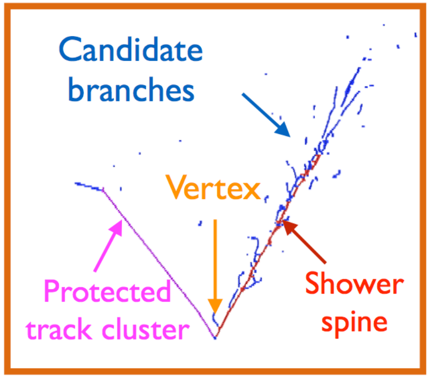
1. Track-oriented 2D clustering
2. 2D topological association
3. 3D track matching
4. Track vs. shower id
5. 2D shower growing
6. 3D shower matching
7. Shower refinement
8. Particle recovery
9. 3D hit creation
10. Event building, characterisation

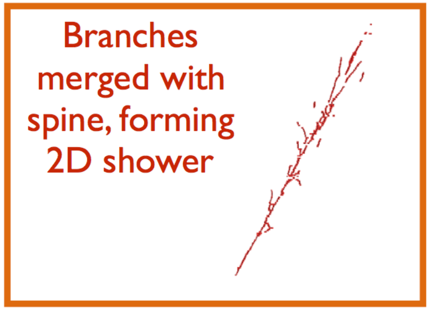
Track oriented

Try to address track vs. shower "tension"

Shower oriented

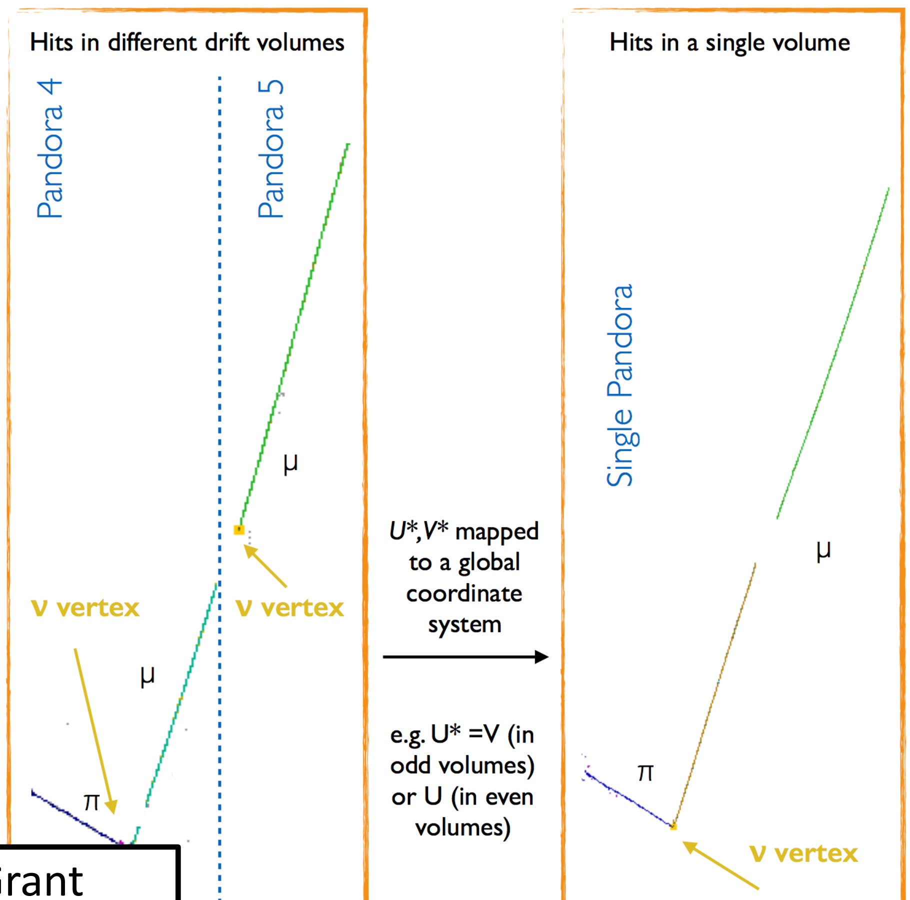






Recent Pandora Updates

- Work underway to adapt Pandora to unique features of DUNE.
  - Ex. handle multiple drift regions with unified coordinates.



- Exploit common wire angles in every TPC in DUNE FD or ProtoDUNE to create global coordinate system (X, U*, V*, Z)
- Events can be reconstructed as if there is a single drift volume:
  - Can use MicroBooNE-style algorithms w/o stitching*
  - But, do still need to address gaps between drift volumes
- Framework to support this approach has been implemented.

*Note there will be complicating factors, such as space charge.

Nick Grant

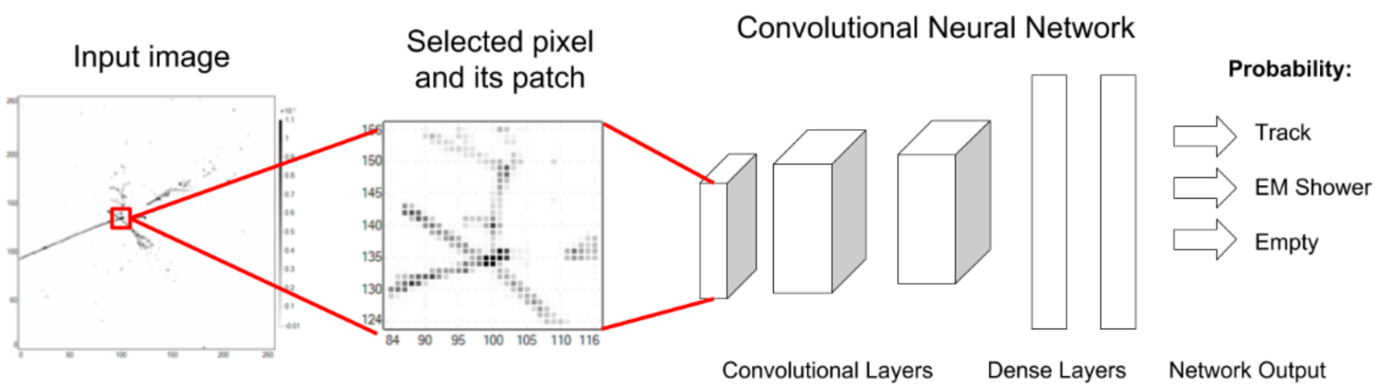
What is a CNN?

- Machine learning technique
  - Developed in the field of computer vision
- This implementation classifies individual hits as: **Track** or **EM shower** or **Michel Electron**

Robert Sulej
Dorota Stefan

MC-Trained CNN to classify each hit created as shower-like or track-like

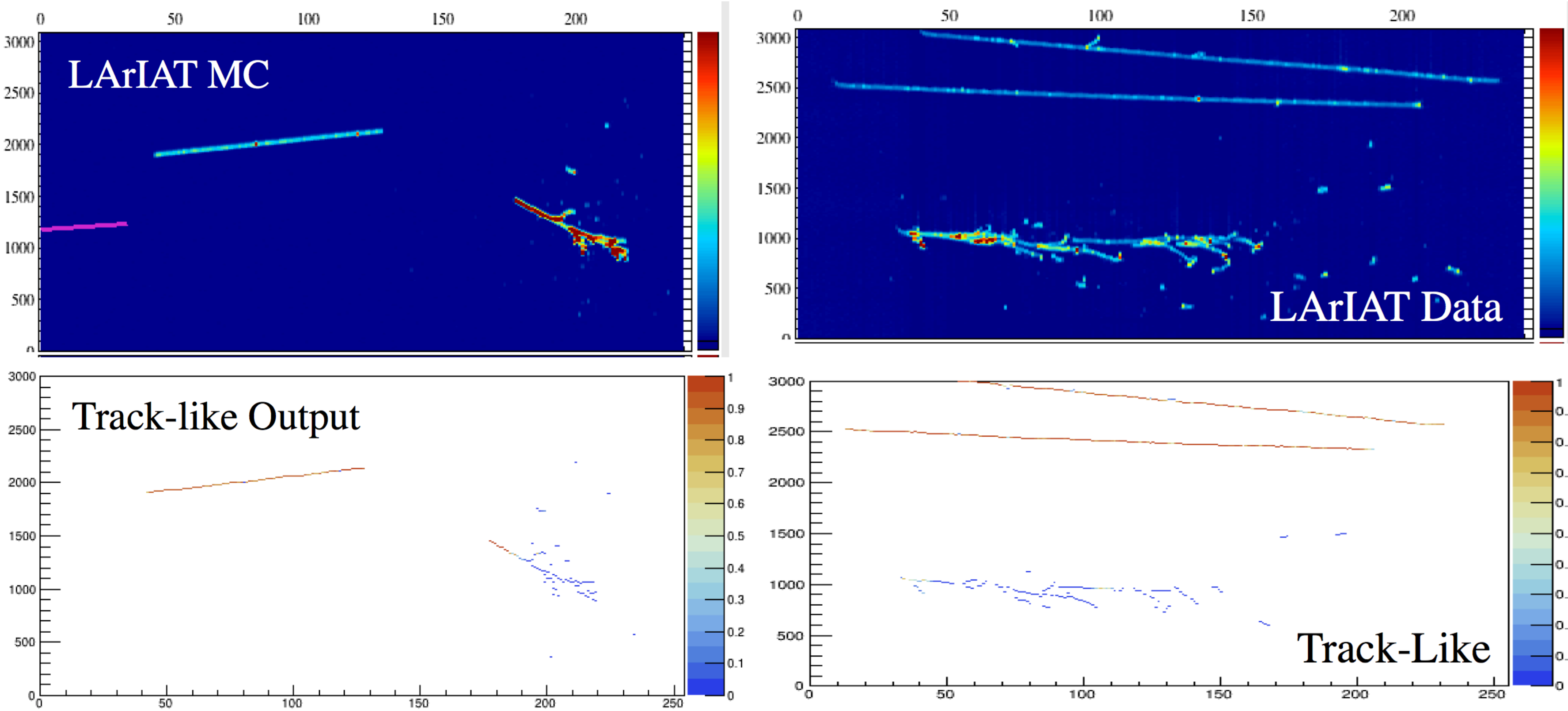
Performed on noise-filtered ADC values and
only after hit finding, making it one of the first reconstruction steps



Greatly speeds up tracking and makes shower clustering possible

Dan Smith

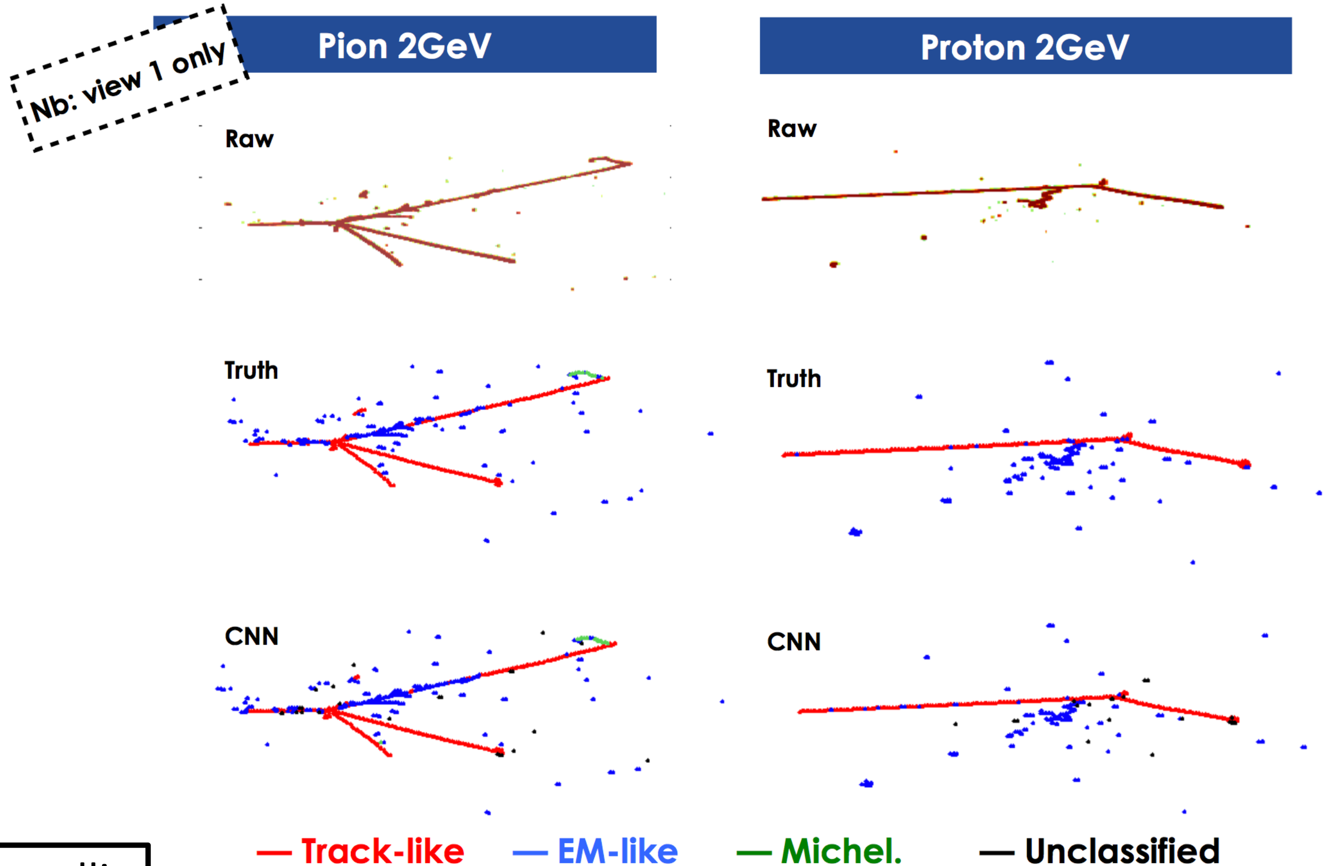
CNN Applied to LArIAT Data



- CNN technique can clearly separate track-like and shower-like hits in simulation and data.
- Some challenges related that data-MC difference being addressed with a *Generative Adversarial Network*

Dan Smith

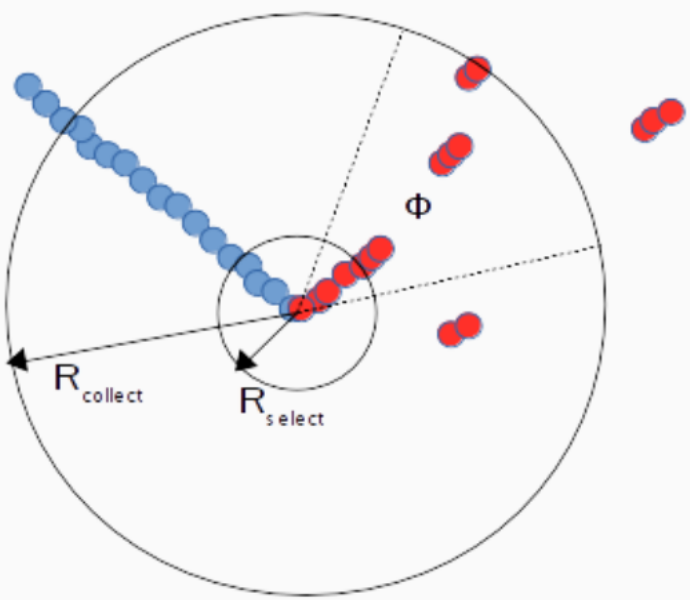
CNN Applied to Dual Phase



Andrea Scarpelli

- Work is just beginning, but the same network design used in single phase is working well for dual phase.
  - Retrained on dual-phase data.

CNN Used for Michel Tagging



Event Selection

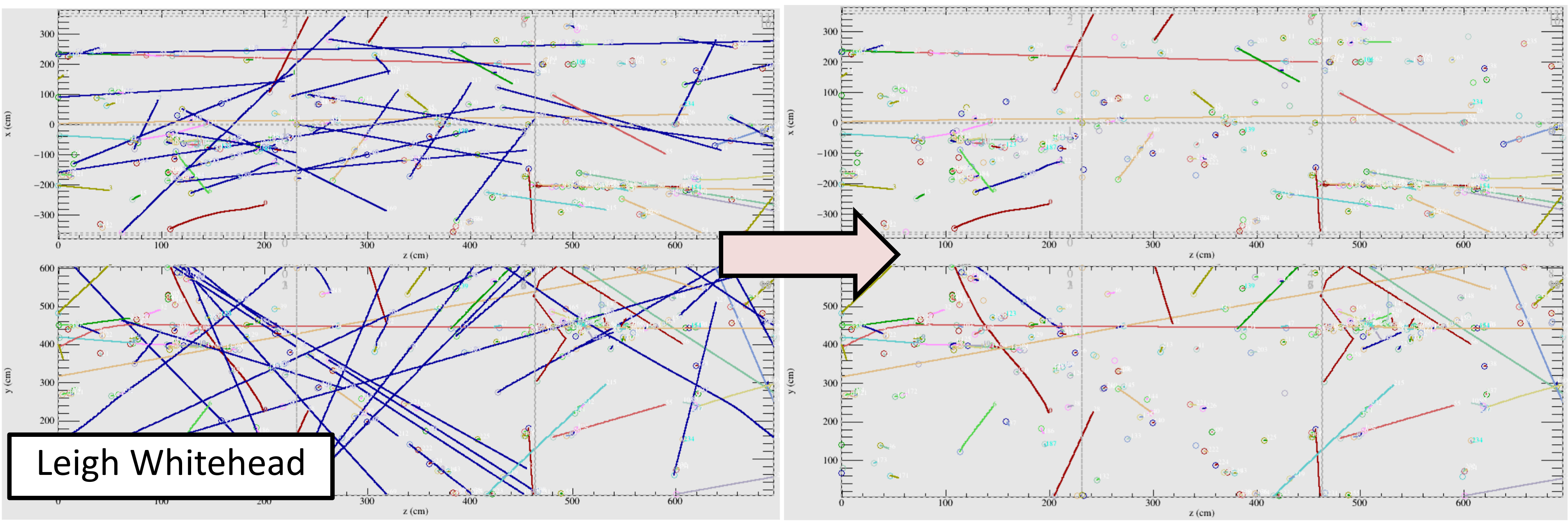
- Project track onto a single plane and find end point
- Define collection radius R_{select} around the track end point
- Count hits within this radius with a CNN output greater than A_{CNN}
- If count exceeds N_{select} select as a Michel event

- Work is also underway to use the CNN output to improve the accuracy of Michel tags.

PMA

- PMA (and TrajCluster) are the “traditional” LArSoft reconstruction chain.
  - Build up clusters, then objects in 2D, and then match across views.
  - Makes use of CNN hit tagging, too.

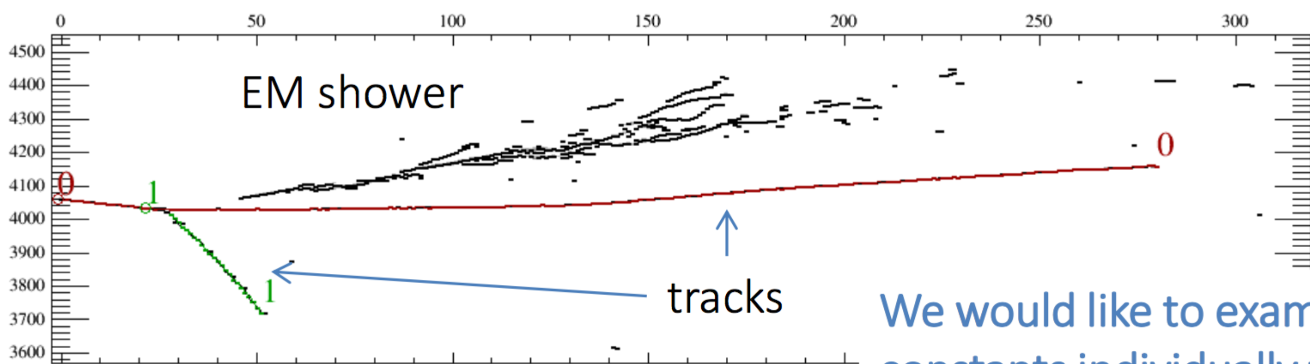
- Recent developments (focused on protoDUNE):
  - Hits tagged by CNN as track-like are reconstructed, vetoed as cosmic rays.



PMA

- PMA (and TrajCluster) are the “traditional” LArSoft reconstruction chain.
  - Build up clusters, then objects in 2D, and then match across views.
  - Makes use of CNN hit tagging, too.
- Recent developments (focused on protoDUNE):
  - Hits tagged by CNN as EM-like are built up into showers.

Illustration of the reconstructed ProtoDUNE event using CNN based EM/track separation.



Tracking only on clusters tagged as track-like.

We would like to examine energy calibration constants individually for EM showers and tracks, and potentially draw more concusions from this separation.

Dorota Stefan

Wirecell

- A reconstruction chain which works directly from the wire signals, deconvolving simultaneously in space and time.

- The toolkit is nearly ready for integration into LArSoft and wider use.
  - Driven by MicroBooNE paper schedule – ready well in time for protoDUNE.

What's left to do?

Areas where help is welcome are in blue.

- Noise filtering
  - Rework to better follow WCT interfaces/components patterns. (bv)
  - Rework LArSoft integrating to follow Art Tool paradigm (??, or bv)
- Improved detector simulation:
  - Long-range/fine-grained detector response, **done** (Yichen, bv).
  - Normalization and validation, **in progress** (Hanyu, bv).
  - Proper **noise** (Jyoti/Milind/Arbin) and **drift** (Hanyu) models: **in progress**.
  - Implement FEE "channel shuffle" and match numbering convention. (bv)
  - Integrate WCT sim components into LArSoft. **design** (Brian, bv)
- Signal processing
  - Port prototype code into toolkit: **started** (bv, Xin)
  - Validate signal processing with simulation, develop "truth metrics" (Brooke), understand SP under different signal and noise assumptions.
  - Integrate WCT sigproc components into LArSoft. **design** (Brian, bv)
  - Finish MicroBooNE paper (whole SP team)
- Toolkit infrastructure miscellany
  - Various improvements/cleanups in configuration layer and build/source management. (bv)

Brett Viren

Performance Evaluation: Vertexing

- Pandora sometimes gives multiple vertices.
- Evaluated the choice of vertex
  - The best performance comes from the most upstream vertex when looking at beam events.

- Using the upstream vtx from Pandora gives the best reco vertex, which has been used in the MVA tree v3.1.

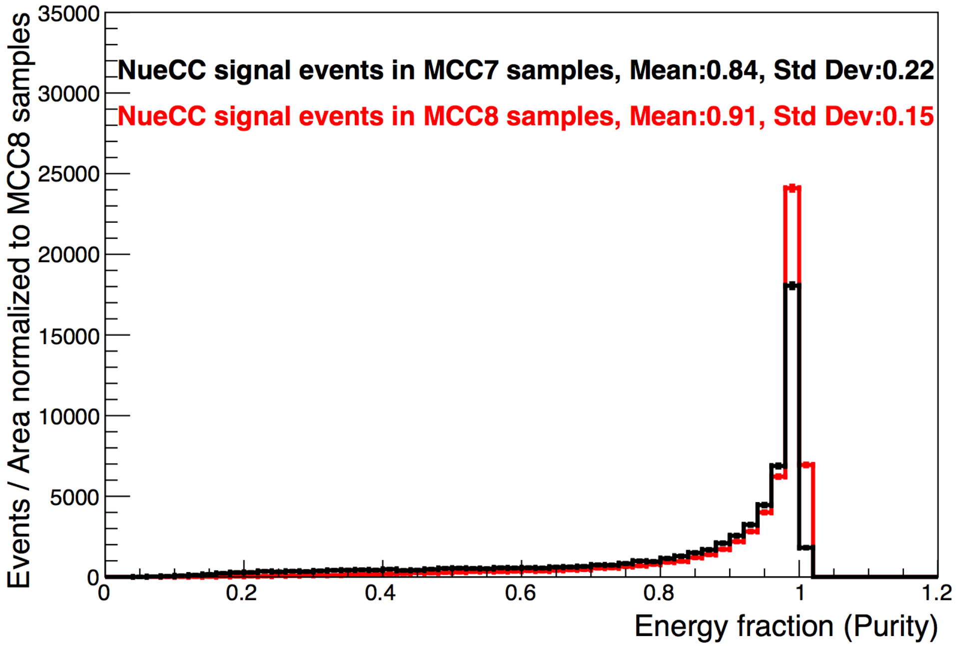
Event Type	ΔX Upstream (cm)	ΔY Upstream (cm)	ΔZ Upstream (cm)
ν_μ CC Pandora	1.18	0.78	0.69
NC Pandora	1.40	0.87	0.81
ν_e CC Pandora	1.12	0.83	0.79

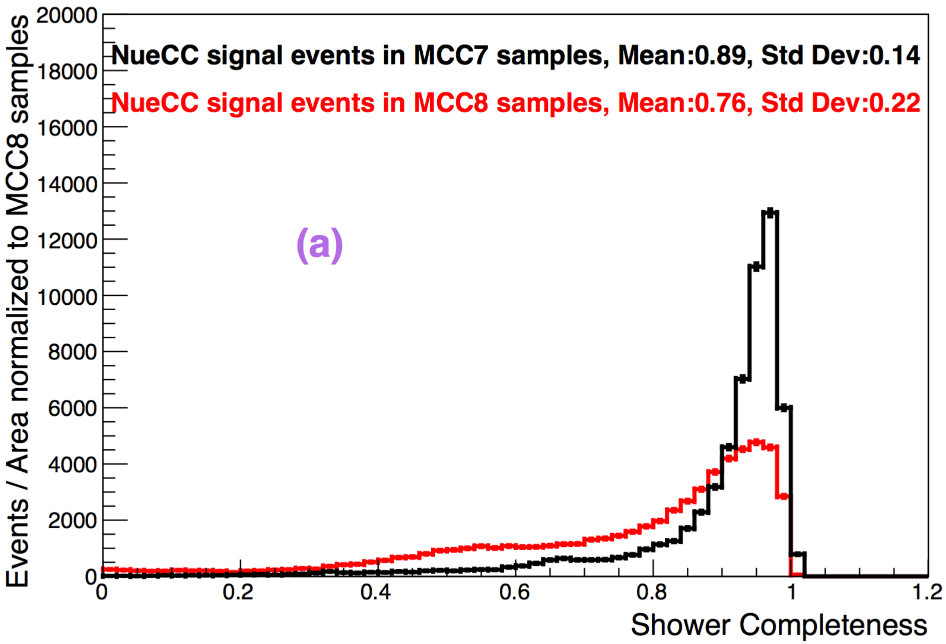
Yujing Sun

Performance Evaluation: Showers

- Evaluate the *completeness* and *purity* of showers electron neutrino signal and background events.

- Latest algorithms (Pandora+CNN in MCC8) have higher purity but lower completeness than previous algorithms (Pandora in MCC7).
  - The underlying cause is still being investigated.
  - An example of why it's critical to have and regularly calculate performance metrics!



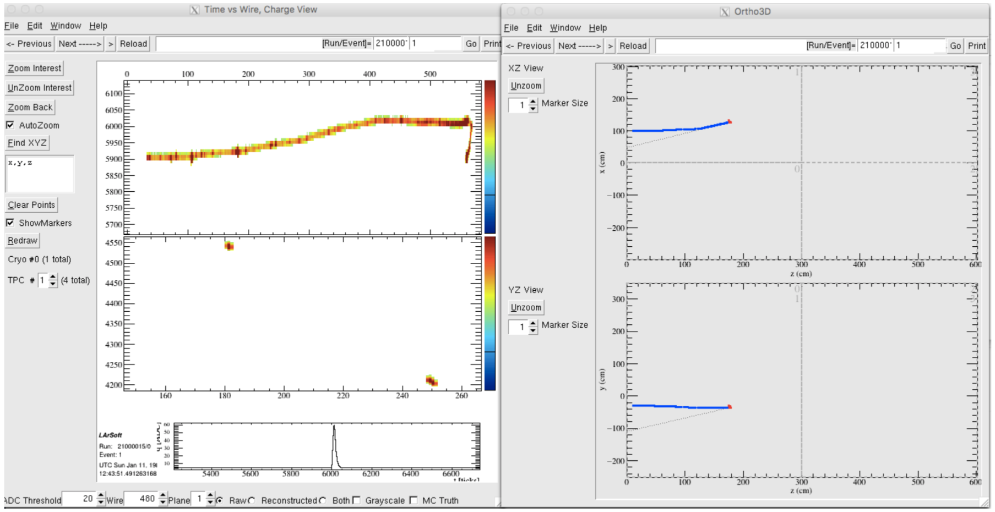


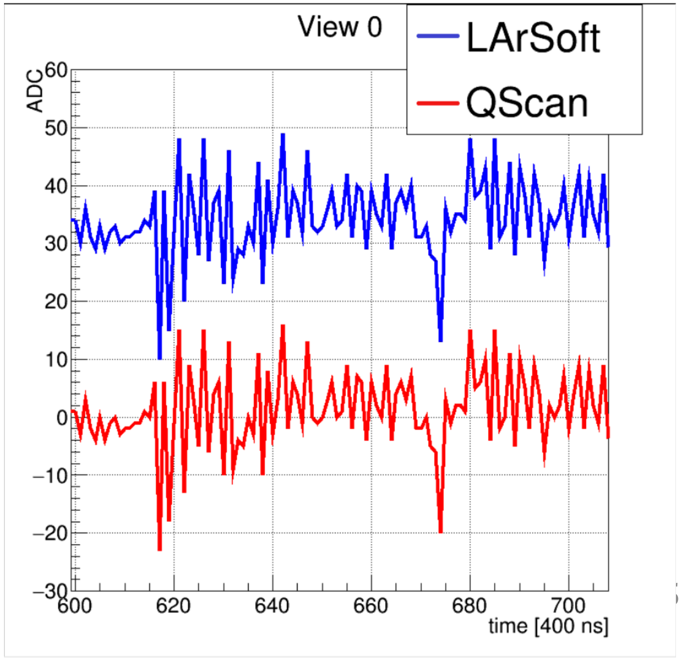
Reddy Pratap Gandrajula

Dual Phase Integration

- Rotate the drift direction
  - Dual Phase drift is vertical, not horizontal
  - Turned out to be a deep LArSoft assumption
  - On-going work

Gianluca & Balint





- Read in raw data from 3x1x1 dual phase prototype.

Kevin Fusshoeller

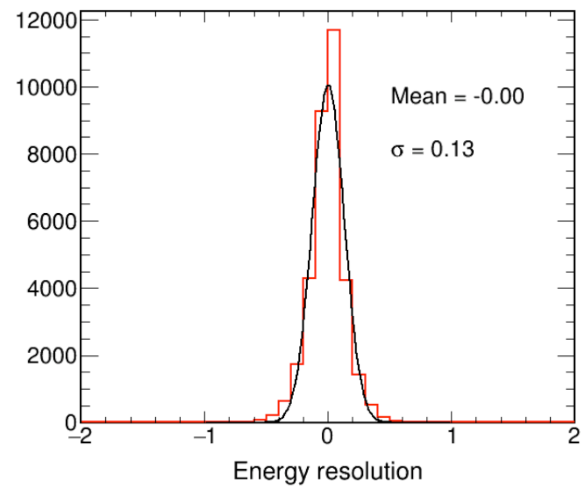
High-Level Reconstruction

Support work in the physics groups

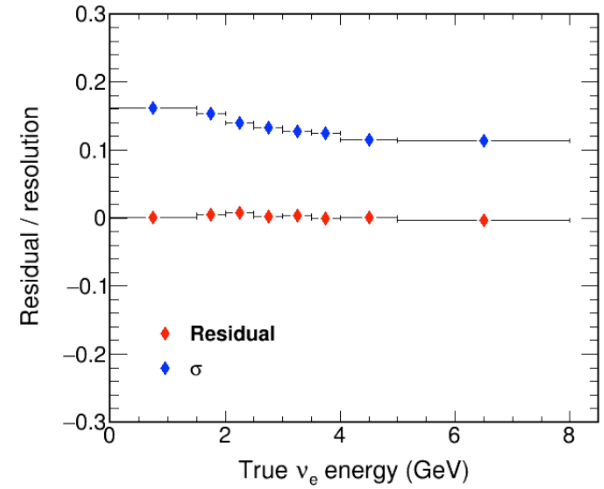
Energy Reconstruction

- Reasonable performance achieved.
  - 10-15% for ν_e
  - 20% for ν_μ
- However, ad hoc corrections to hadronic energy are needed.
  - Must account for neutral particles.

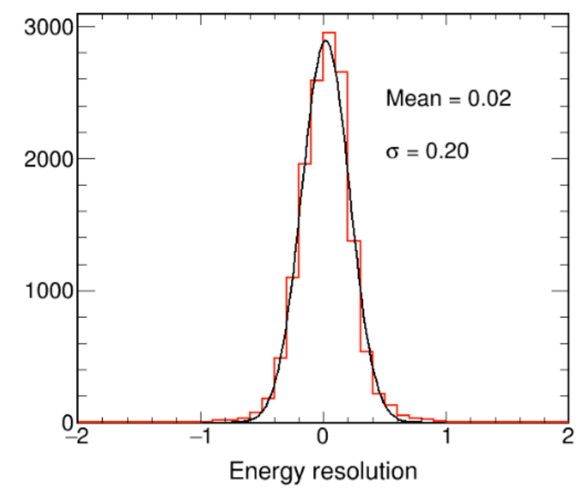
Overall ν_e energy resolution



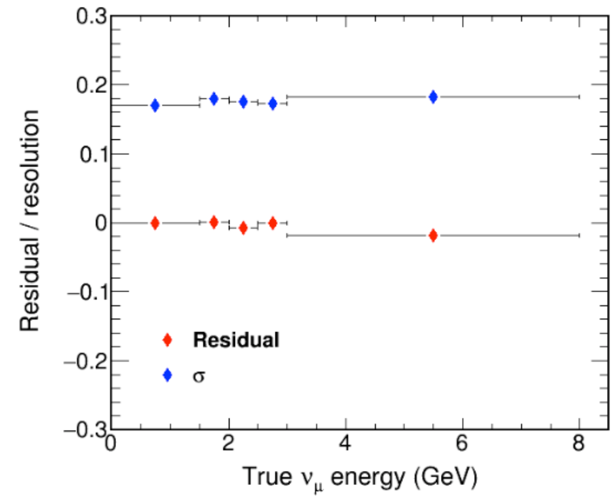
ν_e energy resolution as function of true energy



Overall

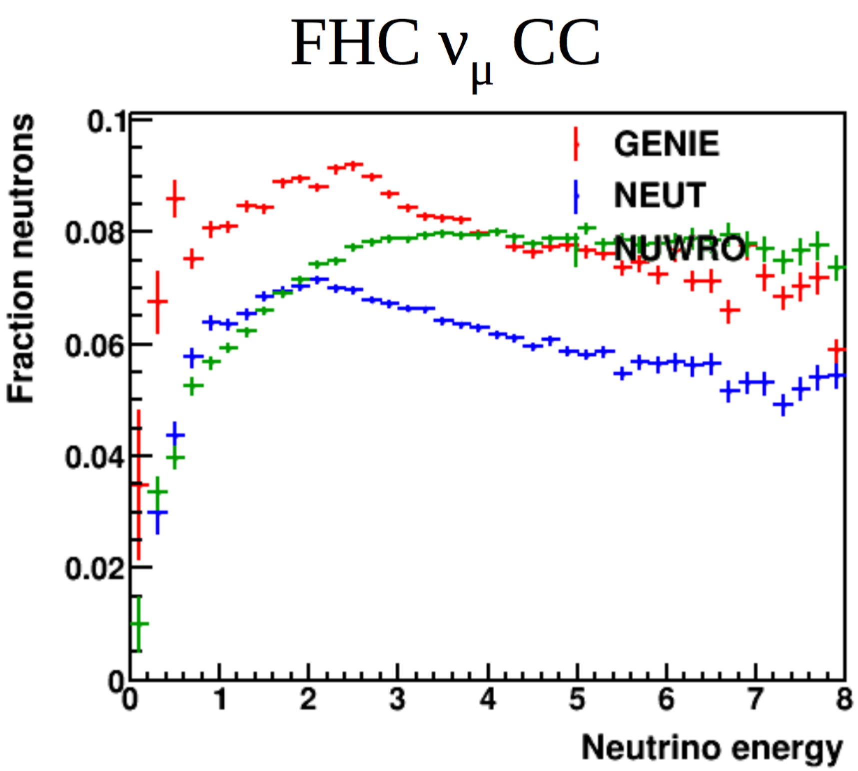


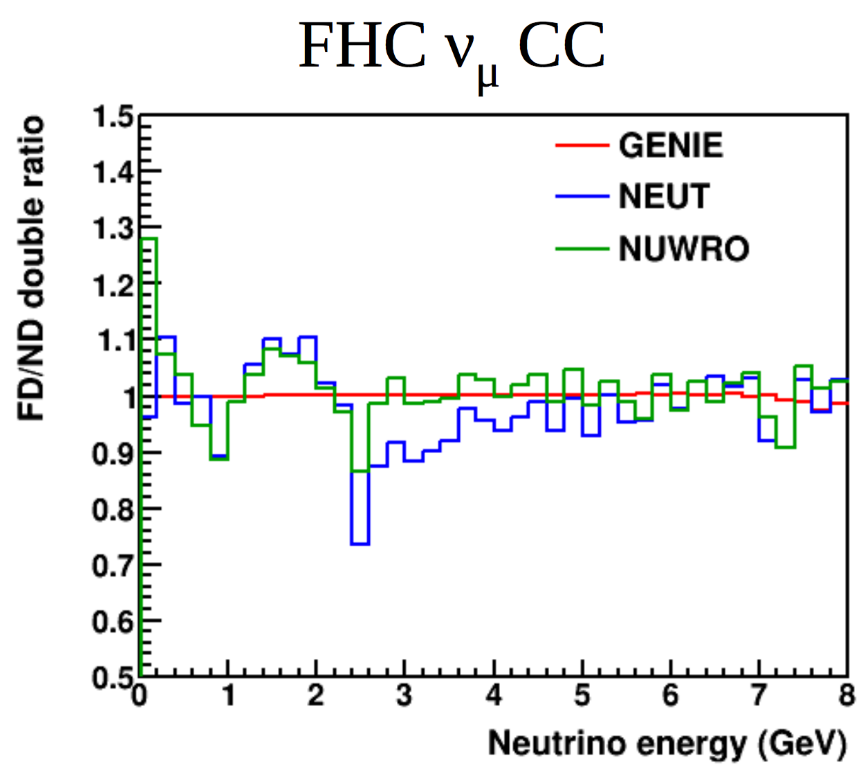
As function of true ν_μ energy



Nick Grant

Correcting for Lost Energy to Neutrons





- The amount of energy going into neutrons varies with different generators.
- Simply collecting up the visible energy available from the neutrons in ND and FD still leaves a 10% affect.

Chris Marshall

6-month Look Ahead

- Simulation
  - TPC
    - Improved modeling of electronics and noise based on experience in running detectors.
    - Full integration of WireCell simulation improvements
  - Photon detectors
    - Incorporation of more detailed timing

6-month Look Ahead

- PMA/CNN
  - Improve reconstruction of vertex, showers
  - Continued development of track/shower separation.
- Pandora
  - Continued improvements in multi-TPC support, optimization for DUNE energies.
- Wirecell
  - TPC signal processing
    - Integrate the TPC signal processing based on 2D deconvolution and ROI selection with LarSoft (developed in the context of MicroBooNE data analysis) for usage of protoDUNEand DUNE
    - Develop code for the data reduction for protoDUNE after TPC signal processing
  - 3D Imaging
    - Achieve low-memory tiling for protoDUNE/DUNE wire geometry (i.e. taking into account the wrapped wires)
    - Implement L1 regularization (i.e. compressed sensing) with 3D imaging
  - 3D pattern recognition
    - Revisit the development of 3D pattern recognition

6-month Look Ahead

- Strengthen connections between single and dual phase efforts
  - Dual phase in LArSoft
  - Track/Shower CNN applied to both
  - Photon simulation
- Support high-level reconstruction efforts in the physics groups.
  - Energy reconstruction
  - Also working to improve short track efficiency – a critical need identified by the NDK group.