# ProtoDUNE-SP Reconstruction Software Review and Performance

Leigh Whitehead

On behalf of the protoDUNE-SP DRA Group

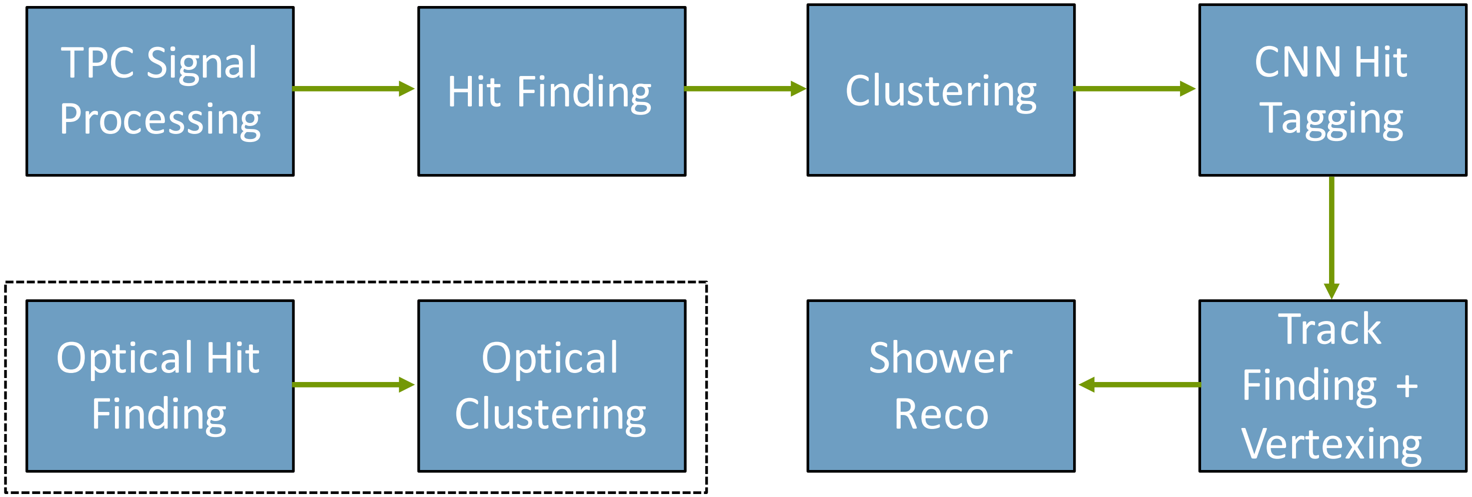
10/05/18

# Introduction

- These slides provide an overview of the material presented in the ProtoDUNE-SP reconstruction software review document
- The reconstruction must provide tools for calibrations, TPC analyses and PD analysis:
  - Efficient cosmic muon reconstruction
  - T0 measurement for as many cosmic muons as possible
  - CNN hit tagging for Michel electron events
- The talk focuses on two main parts:
  - Overview of the algorithms in the reconstruction chain
  - Performance of those algorithms critical to the pion-argon cross section analysis

# Reconstruction Chain Overview

- There are six main steps in the TPC reconstruction chain
  - Some of these steps have different complimentary approaches




- Two steps in the optical information processing

- NB: this figure is demonstrative, other approaches such as WireCell go straight from TPC signals to 3D reconstruction

# TPC Signal Processing

- The goal of the signal processing is to reconstruct the distribution of ionisation electrons arriving on each wire over time
  - Provide clean waveforms from which to begin hit finding
- Current technique based on a 1D convolution
  - Apply Fast Fourier Transform to isolate noise and signal frequencies
- Effectively roles up all sources of signal shaping (amplifiers, electronics response etc) into a Gaussian smearing function

# TPC Signal Processing

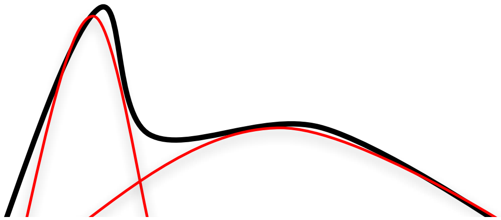
- MicroBooNE recently made an important step forward using a 2D convolution
  - This method will be ported over to protoDUNE soon

- Possible issues for ProtoDUNE:
- Sticky codes: these are incorrect adc values that appear as spikes in the waveform
  - Represent a loss of information but a new ADC code can be formed via interpolation from neighbouring good codes

- Non-linearity of the ADCs
  - Must be dealt with using a calibration scheme.

# Hit Finding

- The hit finding ”GausHitFinder” algorithm searches for the number of peaks in a waveform



- After finding each of the N peaks the distribution is fitted with N Gaussian functions

- Each one of these N Gaussian fits forms the basis of an individual hit object (recob::Hit)

# Hit Disambiguation

- The wrapped induction wires of the APAs give a non one-to-one mapping of an electronics channel ID to a wire ID
  - Each channel ID maps to a number of wire IDs (on both sides of the APA)

- Whilst protoDUNE has TPCs only on one side of the APAs the wires are wrapped and an algorithm must be used to identify the correct wire ID for a signal on a given channel ID

- ProtoDUNE-SP uses SpacePointSolver as the default algorithm...
  - 10x faster than the previous method developed for the 35t
  - More accurate in ProtoDUNE (not the case for the FD)
  - Improved and faster tracking efficiency with linecluster + pma
  - Details of the process on the next slide

# Space Point Solver

- SpacePointSolver aims to convert three 2D views into a single collection of 3D space points
- Matches triplets of wires across three views matching closely in time – often there can be multiple candidate triplets
  - Resolves ambiguities by minimising the difference between the predicted and observed charges on the induction wires
- Designed as the first step towards a fully 3D reconstruction for FD neutrino interactions
- For ProtoDUNE we will initially use it to perform disambiguation
  - More accurate and faster than the aforementioned disambiguation

# Space Point Solver

- Example of the algorithm performance at the FD

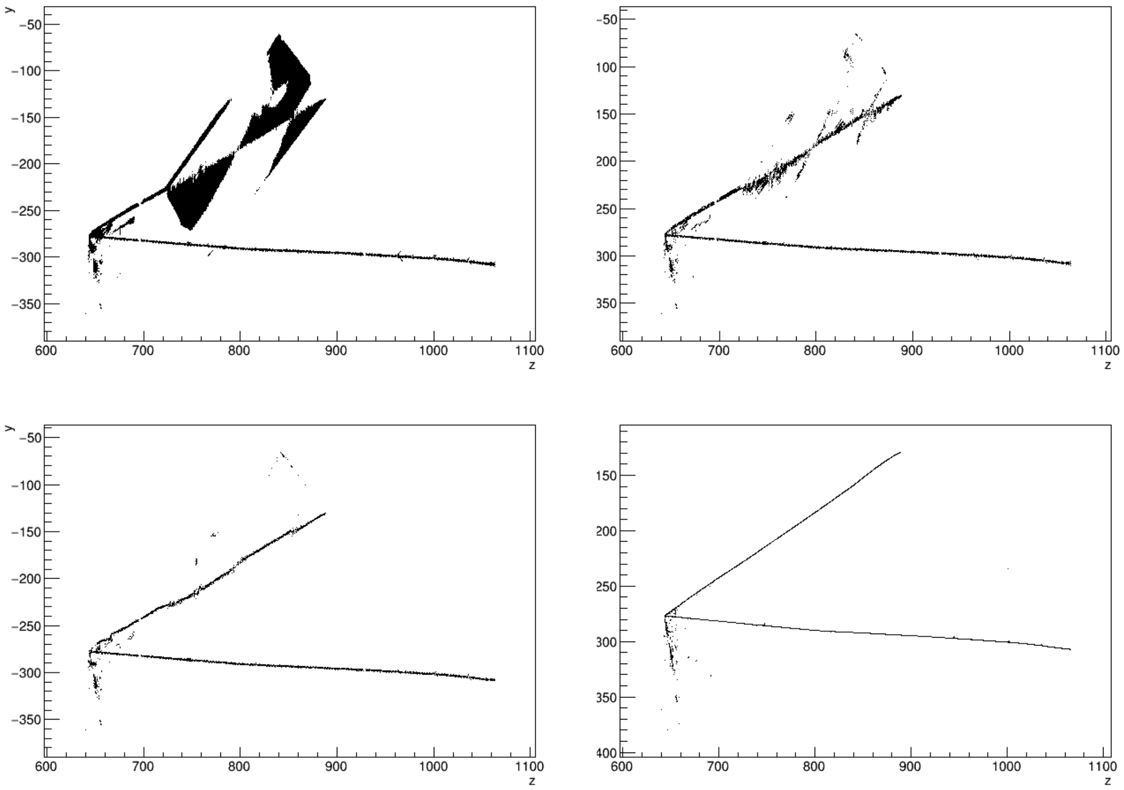


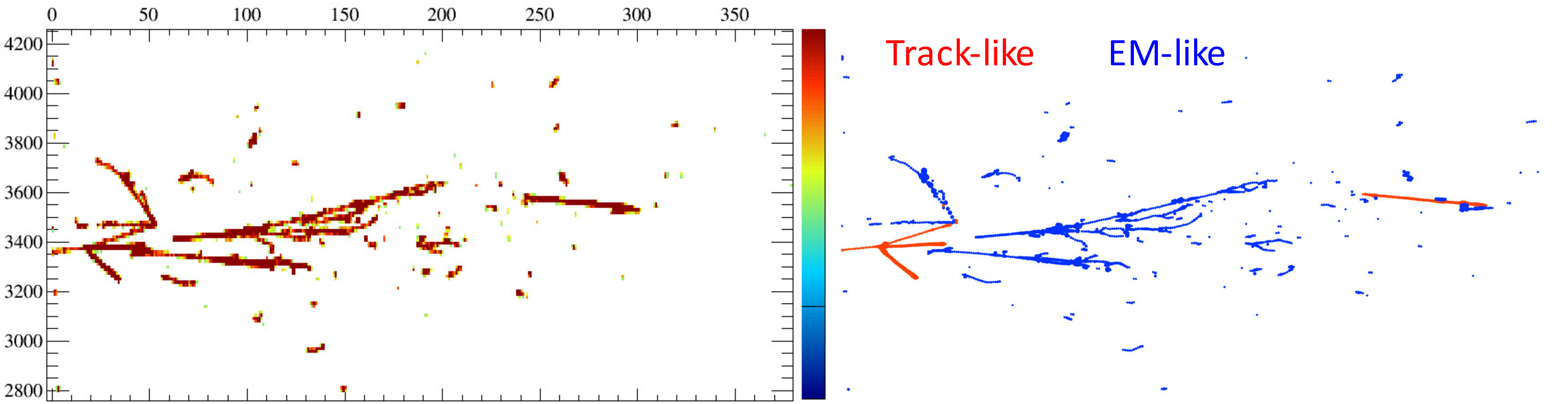

Figure 1: Performance of SpacePointSolver on a simulated FD neutrino interaction. The first panel shows the position of all triplet coincidences in the $zy$ view (looking from the side of the detector), displaying multiple ambiguous regions. The second and third panels show the solution with and without regularization, the regularization disfavouring various erroneous scattered hits. The final panel shows the true charge distribution, demonstrating that the fidelity of the regularized reconstruction.

# Clustering

- We have two clustering approaches as of MCC10:
  - LineCluster and TrajCluster
- Both methods aim to form clusters using a short line-like seed cluster and searching for similar hits to extend the cluster to produce 2D clusters of associated hits
- TrajCluster is more complex than LineCluster
  - Can match together the clusters from the 2D views into 3D
  - Tags shower-like clusters
- Pandora (see later) has its own set of clustering algorithms

# CNN Hit Tagging

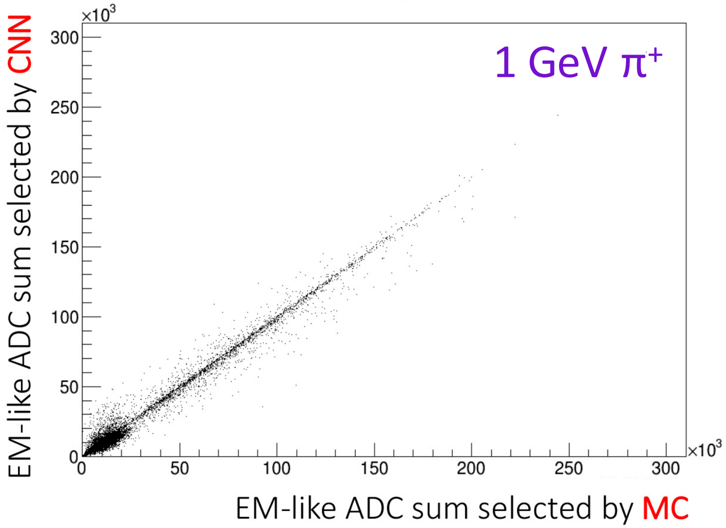
- The hit-tagging CNN takes the hits from the clustering step as input
  - It classifies each hit as track-like or EM-like, and then also as how Michel-like it is

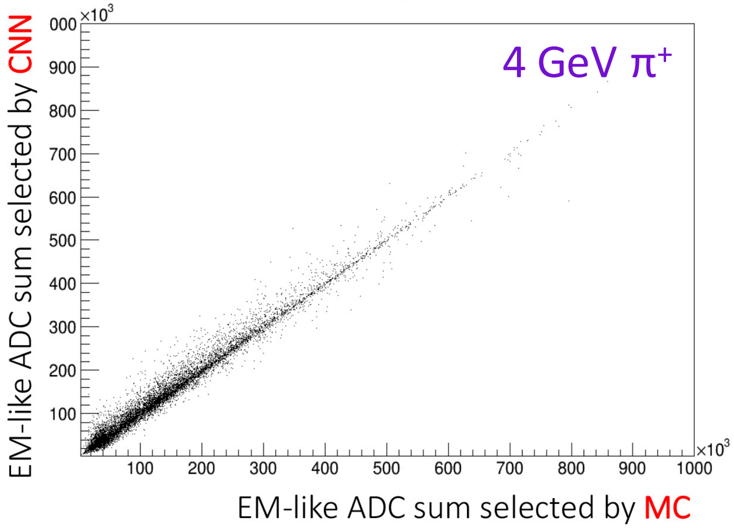



- It considers each view separately and classifies hits in each view in the same way

# CNN Hit Tagging

- Example performance for beam $\pi^+$ events
  - Show the total EM-like tagged ADC total from the CNN compared to the true total EM-like ADC




- Output from the CNN used in numerous places
  - Allow tracking algorithms to purely focus on track-like hits
  - Michel-like hits used for the Michel electron analysis
  - EM-like hits used for electron and $\pi^0$ reconstruction and analyses

# Tracking - PMA

- Projection Matching Algorithm (PMA) was developed as a 3d reconstruction tool for particle trajectories in ICARUS
- It natively creates 3D track objects by minimising the distance to hits in all three views simultaneously
- It also performs track vertexing allowing for the creation of extended and complex structures of interactions
- There have been some updates for the specific challenges of ProtoDUNE...

# Tracking - PMA

- Cathode stitching:
  - Associate tracks either side of the cathode and form a single track
  - The shift required in the drift direction to do this gives the track T0
  - NB: this also works for anode stitching in those geometries that require it

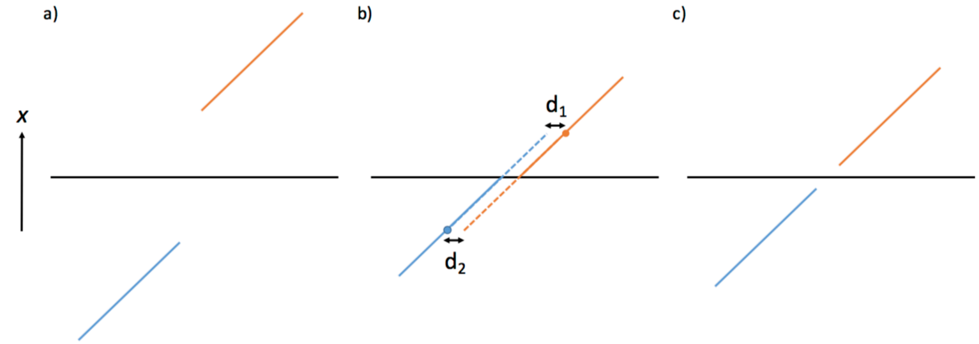



Figure 2: Illustration of the track stitching process. a) Tracks that truly cross the cathode are seen with an equal and opposite shift in the drift direction ($x$). b) The tracks are shifted to the cathode, and an extrapolation is performed from the $n^{th}$ point closest to the cathode on each track to the x coordinate of the $n^{th}$ point on the other track. The two distances between the tracks and extrapolated positions $d_1$ and $d_2$ are summed. c) The sum of the distances is minimised by varying the shift within a few centimetres and the tracks are merged together if $d_1 + d_2 < s_t$.

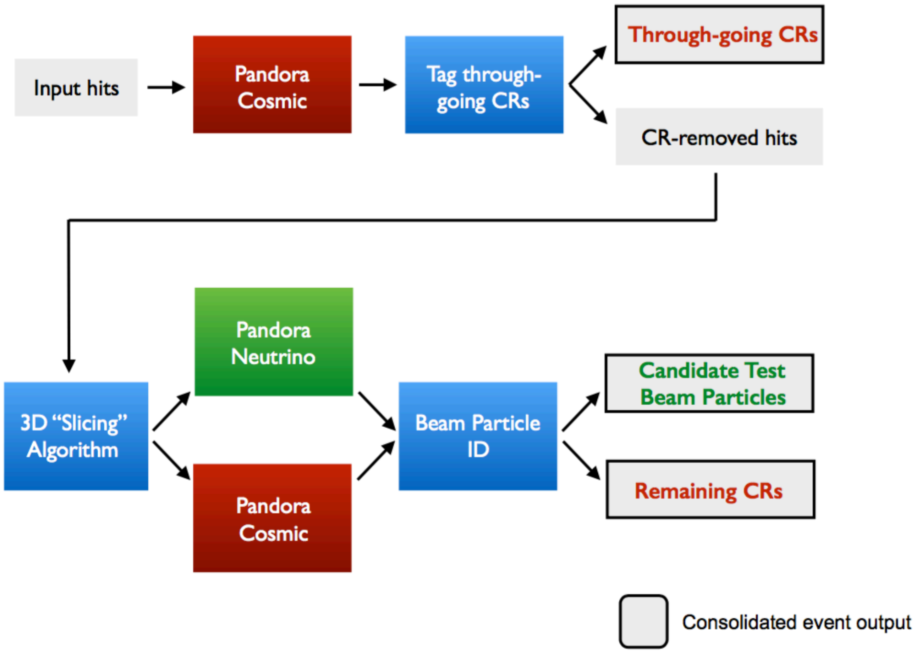
- Cosmic-ray tagging

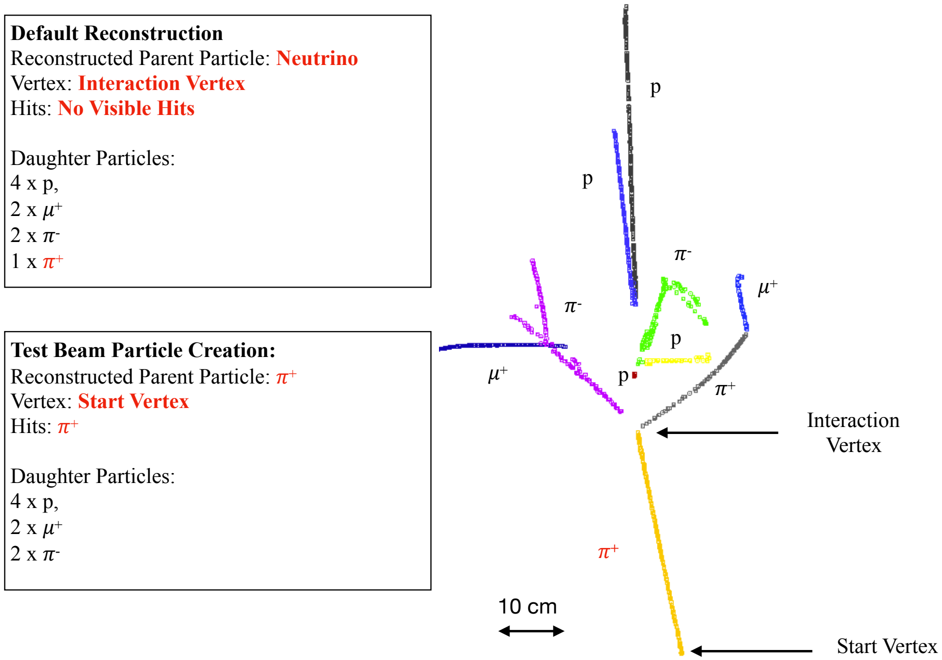
- Use the hit-tagging CNN to reconstruct only track-like objects

# Pandora

- Pandora employs a multi-algorithm approach to gradually build up a complete interaction
  - Used successfully on MicroBooNE
- Events are sliced into regions of interest ideally containing hits from a single primary
  - The hits in these regions are passed through two reconstruction chains: one optimised for cosmics, the other for neutrinos
- In the case of protoDUNE, the neutrino reconstruction chain becomes the beam particle reconstruction
  - Along with the addition of a specific module that re-organises the final interaction given that there is an incoming beam particle and not a neutrino interaction vertex

# Pandora




- Pandora will then decide whether a given slice contains a beam or cosmic particle using a BDT
  - Gives candidate beam particles and cosmic-rays as output

# Shower Reconstruction

- Pandora produces shower objects as part of the full primary particle interaction description
- The EMShower algorithm takes the Pandora outputs and reconstructs full 3D showers
  - It also takes the output from the CNN to reject non EM-like hits
  - Position and momentum four-vectors
  - dE/dx in the initial region of the shower – provides electron / photon ID
- Did not run as part of Monte Carlo Challenge (MCC) 10
  - Testing currently underway and will be re-introduced in MCC11

# Calorimetry and PID

- The calorimetry algorithms are required to convert the ADC to a final dE/dx for reconstructed tracks

- Firstly a conversion from ADC to charge is performed
  - Account for charge loss due to impurities
  - Provides dQ/dx

- In order to convert from dQ/dx to dE/dx need to account for charge quenching:
  - Apply Birk's or the modified Box model

# Calorimetry and PID

- Examples from the FD:
- Bottom right plot shows alternative PID method called PIDA
- PIDA uses dE/dx and residual range to separate species

$$PIDA = \frac{1}{N} \sum_{R_i=0}^{R_i<30\,\mathrm{cm}} \left(\frac{dE}{dx}\right)_i R_i^{0.42}$$

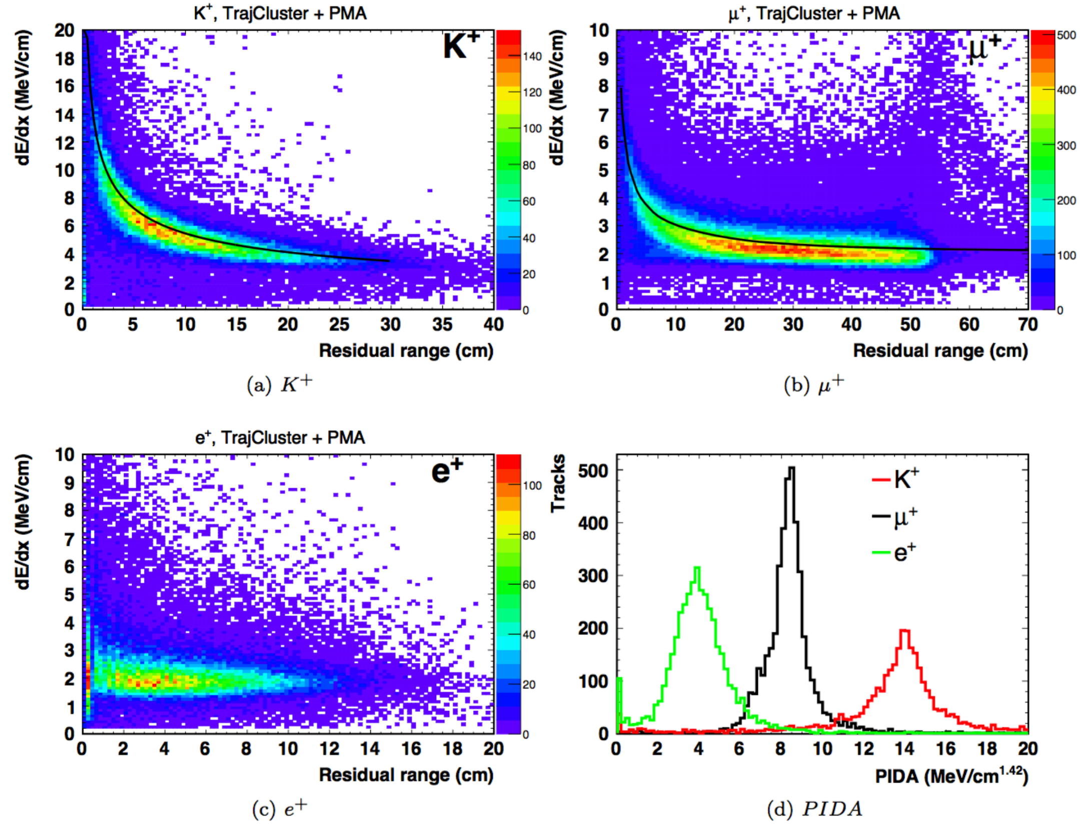



Figure 6: (a)(b)(c): $dE/dx$ as a function of the residual range for reconstructed $K^+$, $\mu^+$ and $e^+$ tracks in proton decay events ($p \to \bar{\nu}K^+$, $K^+ \to \mu^+ \to e^+$). The black curves are theoretical predictions. (d): $PIDA$ distributions of the reconstructed $K^+$, $\mu^+$ and $e^+$ tracks.

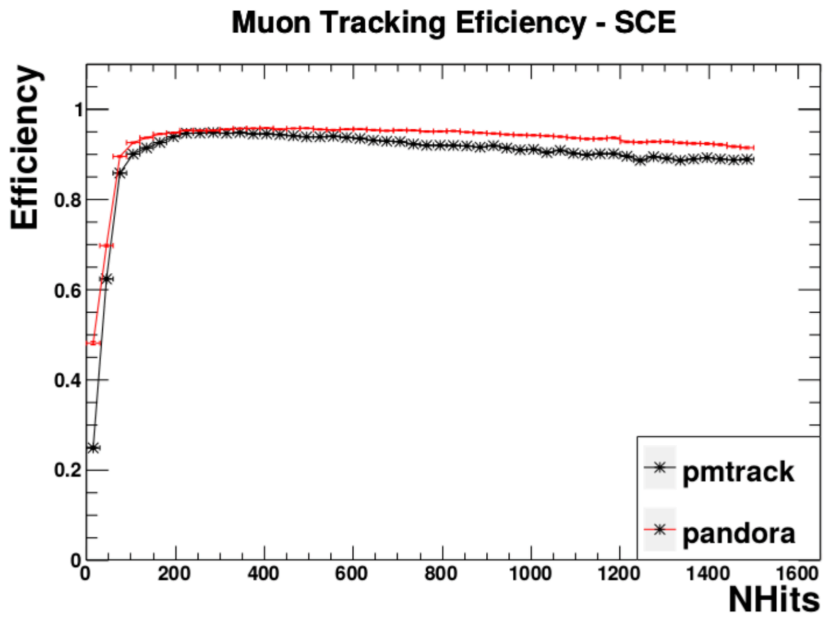
- dE/dx curves one of the first goals from ProtoDUNE beam data

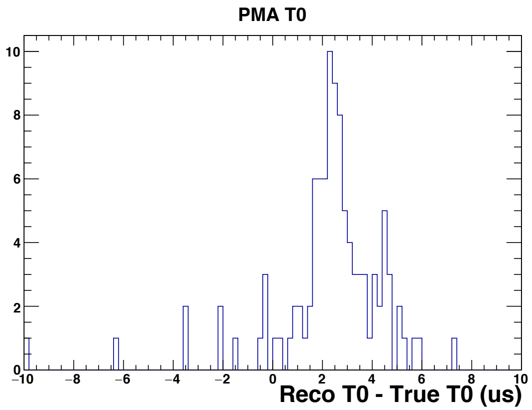
# Critical Path for the Pion Analysis

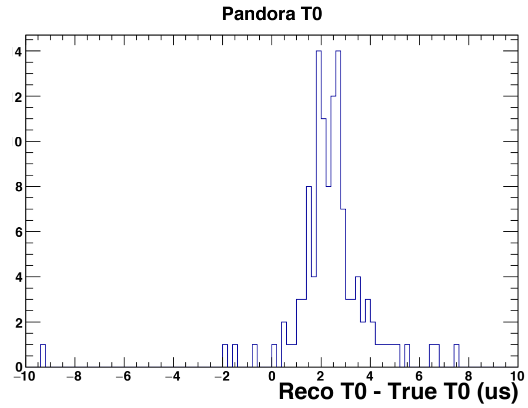
- The primary physics goal for protoDUNE-SP is the measurement of the inclusive pion-argon cross section
  - See Stefania's talk from the morning session for more details
- The algorithms explicitly required for this analysis are a subsample of those previously described
- We need:
  - Reconstructed cosmic muons with $T_0$ for calibrations
  - Rejection of cosmic rays and identification of $\pi^{\pm}$ for the analysis
- Track reconstruction is key here

# Cosmic-ray Track Reconstruction

- We need to efficiency and accurately reconstruct and identify cosmic rays
- T0-tagged cosmics needed for detector calibration
- Need to reject as many cosmics as possible for the beam analyses

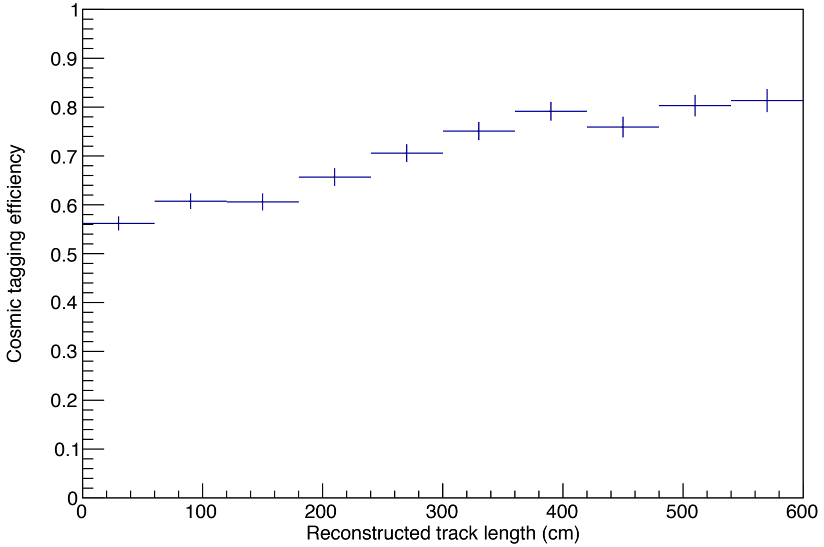


# Cosmic-ray Muon Tagging - PMA

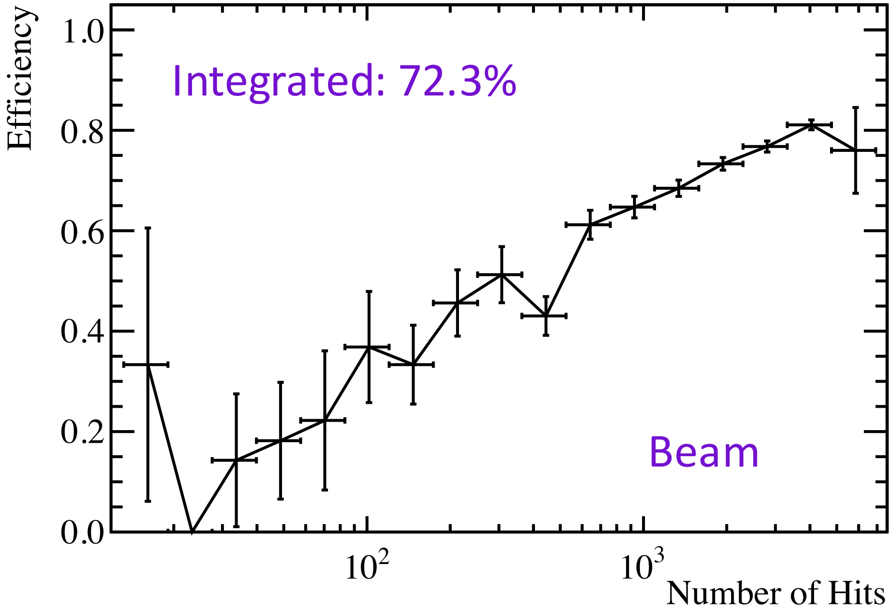
- Cosmic rays tagged in PMA using a number of techniques
  - Measured $T_0$ incompatible with beam
  - Reconstructed outside the detector in drift direction assuming $T_0 = 0$
  - Track either:
    - Crosses TPC top to bottom
    - Crosses TPC from top to front/back
    - Enters TPC from the top and stops
  - Integrated efficiency = 70%
  - Analysis level cuts will remove more of these cosmics

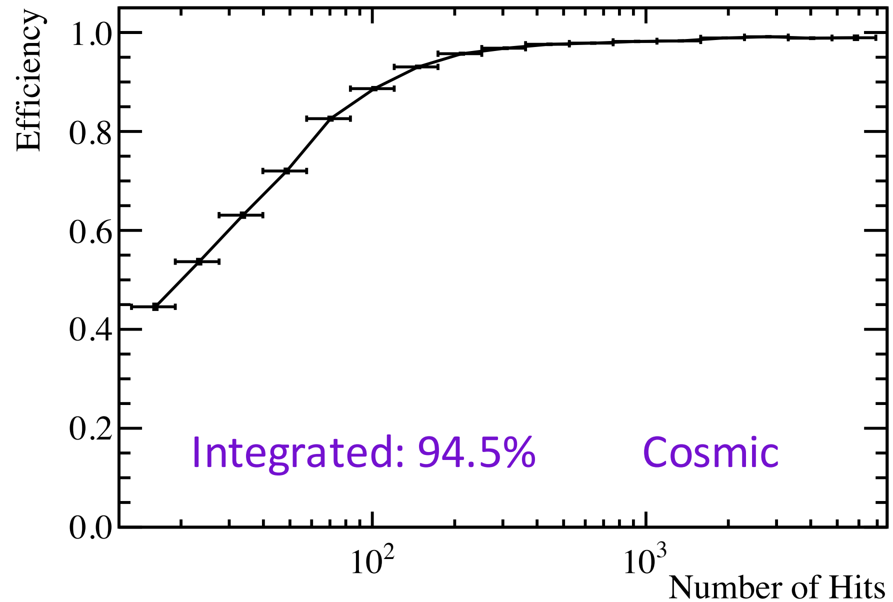



- Purity ~ 93% – the $T_0$ tagging methods can also tag beam backgrounds, but not the signal particles we are interested in
- T0 tagged cosmics are critical for the detector calibration

# Cosmic / Beam ID - Pandora

- The ProtoDUNE-SP version of Pandora aims to label the particles as either of cosmic or beam origin

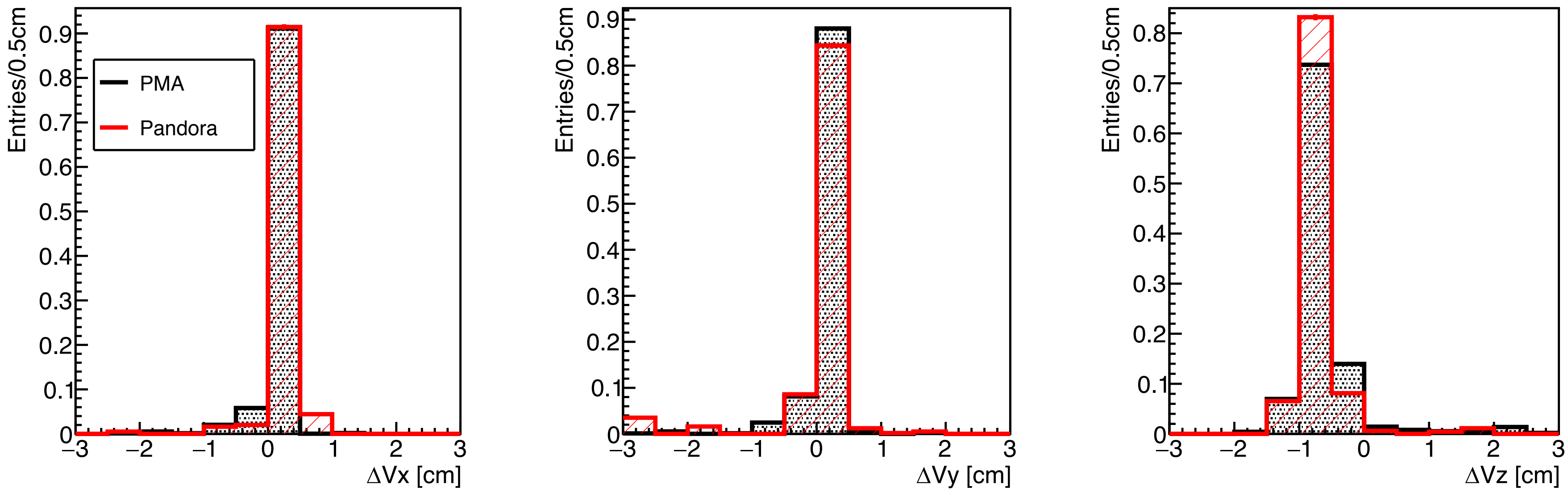



- NB: If a slice contains more true cosmic hits than beam hits it is considered as cosmic thus causing the apparent low efficiency in the beam particle plot
  - This purely means Pandora finds these events ambiguous
  - These are not lost for downstream algorithms and analyses where beam – TPC matching will reclaim many of these.

# Pion Entry Point

- A key element of this analysis is matching the beam particle to the correct track inside the TPC drift volume
  - This has been studied using both PMA and Pandora




- The x and y components look mostly the same with SCE, but we get a 20cm(!) shift in z. Important to correct for this!

# Summary

- The reconstruction software is in good shape
    - It will provide the required samples for the required calibrations and the primary physics goal

- Those tools required for the pion-argon cross section are the algorithms needed to reconstruct pion and muon tracks
    - Demonstrated that these are working well in simulation

- The algorithms not included in the critical path are still very important
    - Needed for other secondary goals ($\pi^0$, beam electrons etc)
    - Important for the developers to test their algorithms on protoDUNE data ahead of implementing them in the DUNE FD

# Cosmic / Beam ID - Pandora

- The ProtoDUNE-SP version of Pandora aims to label the particles as either of cosmic or beam origin

| Test Beam Particle Momentum [GeV] | Beam Particle Reconstruction Efficiency [%] | |
|---|---|---|
| | BDT Beam Particle ID | Cheated Beam Particle ID |
| 1 | $77.4 \pm 1.1$ | $82.8 \pm 1.0$ |
| 3 | $77.6 \pm 1.9$ | $87.6 \pm 1.5$ |
| 5 | $72.3 \pm 0.6$ | $83.3 \pm 0.5$ |
| 7 | $70.8 \pm 0.5$ | $88.1 \pm 0.3$ |

- NB: If a slice contains more true cosmic hits than beam hits it is considered as a loss of efficiency. This is something that can be reclaimed at analysis time