WARWICK
THE UNIVERSITY OF WARWICK

DEEP UNDERGROUND NEUTRINO EXPERIMENT

LArTPC Pattern Recognition with Pandora

John Marshall for the Pandora Team

27th January 2019

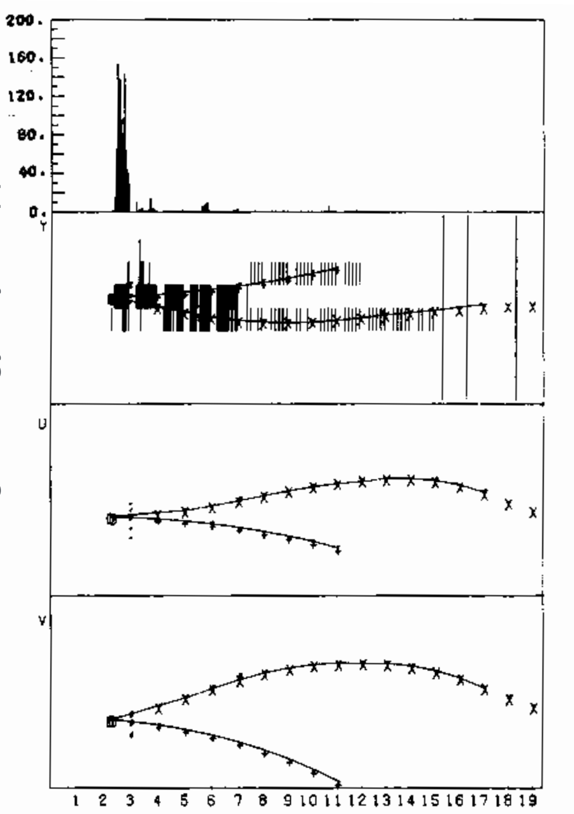
Overview

1. LArTPC event reconstruction
2. Pandora multi-algorithm approach
3. Overview of key Pandora algorithms
4. Pandora highlights at ProtoDUNE-SP

Key references: **Eur. Phys. J. C (2018) 78: 82**
and **Eur. Phys. J. C (2015) 75: 439**

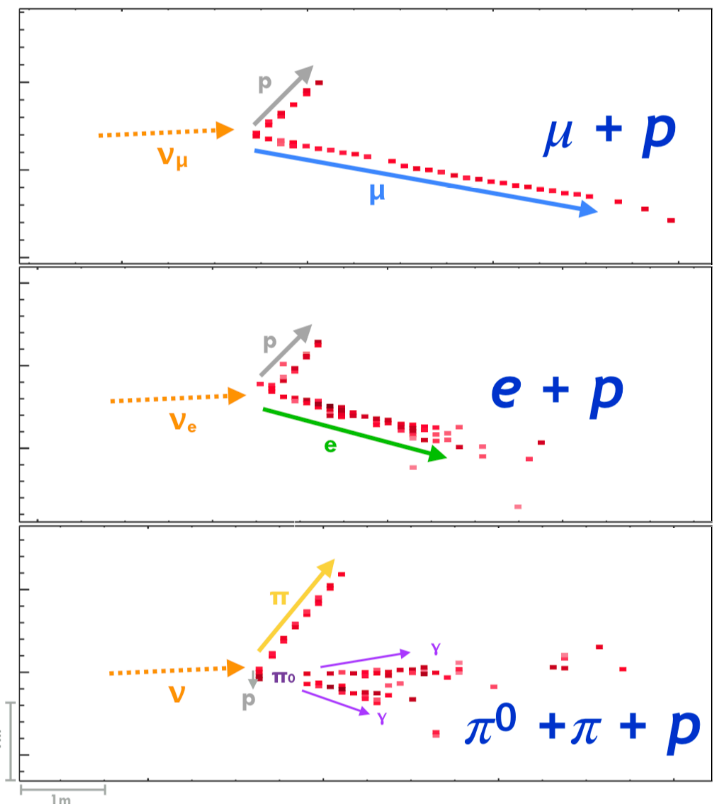
Neutrino Detectors

New Frontiers in High-Energy Physics pp 227-261



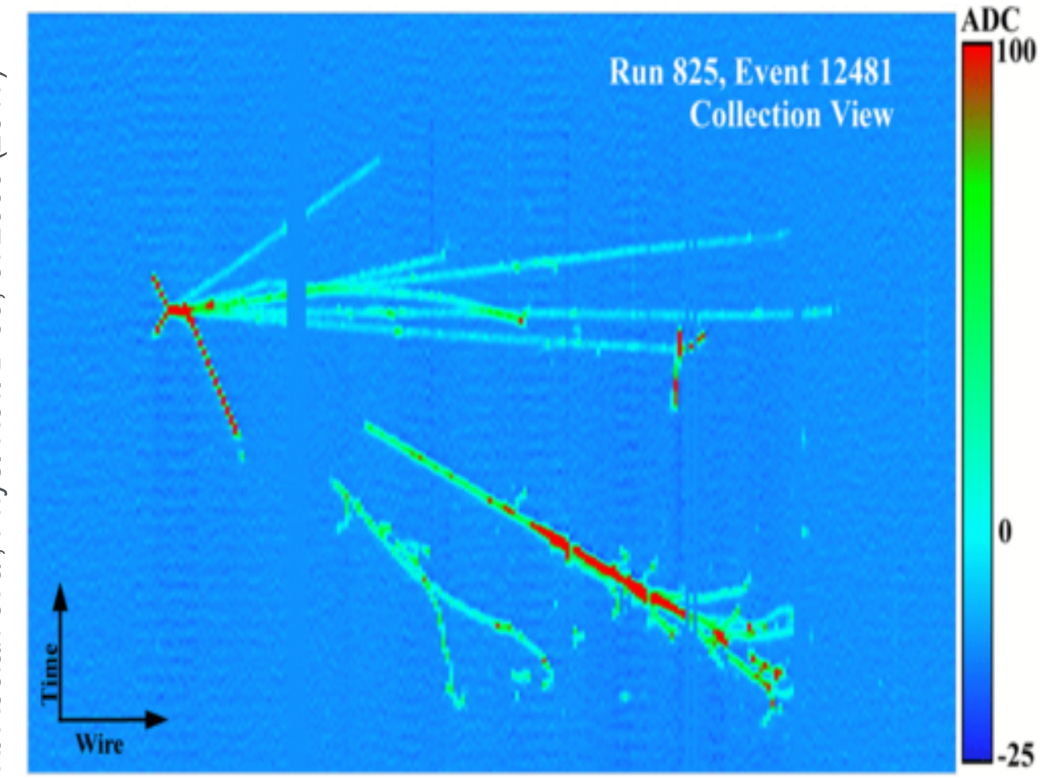
CDHS

DOI: 10.5281/zenodo.1286758



NoVA

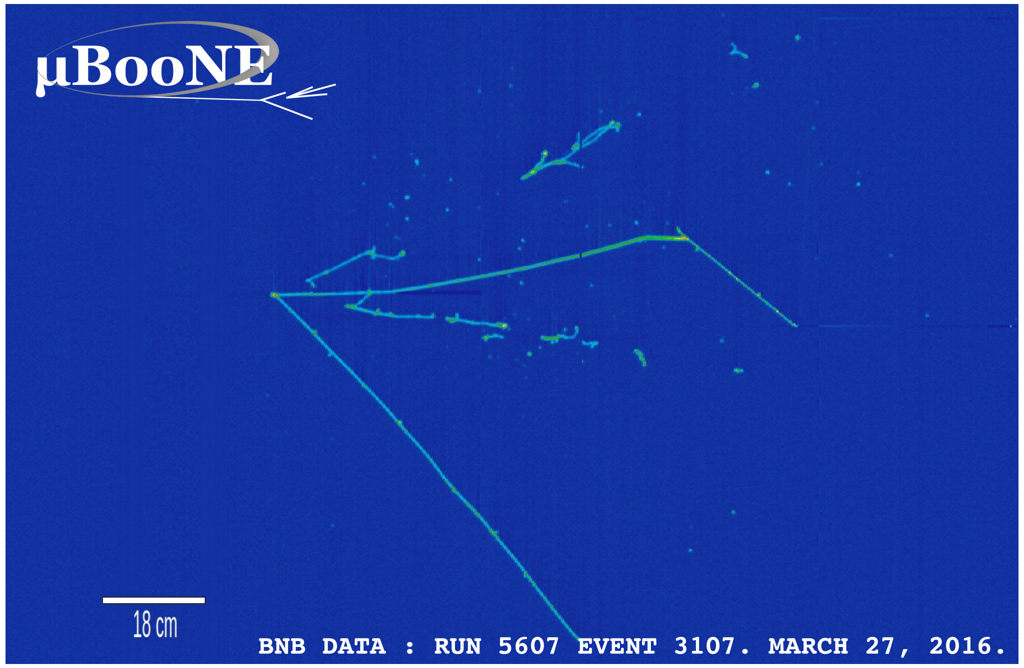
R. Acciari et al, Phys. Rev. D 95, 072005 (2017)

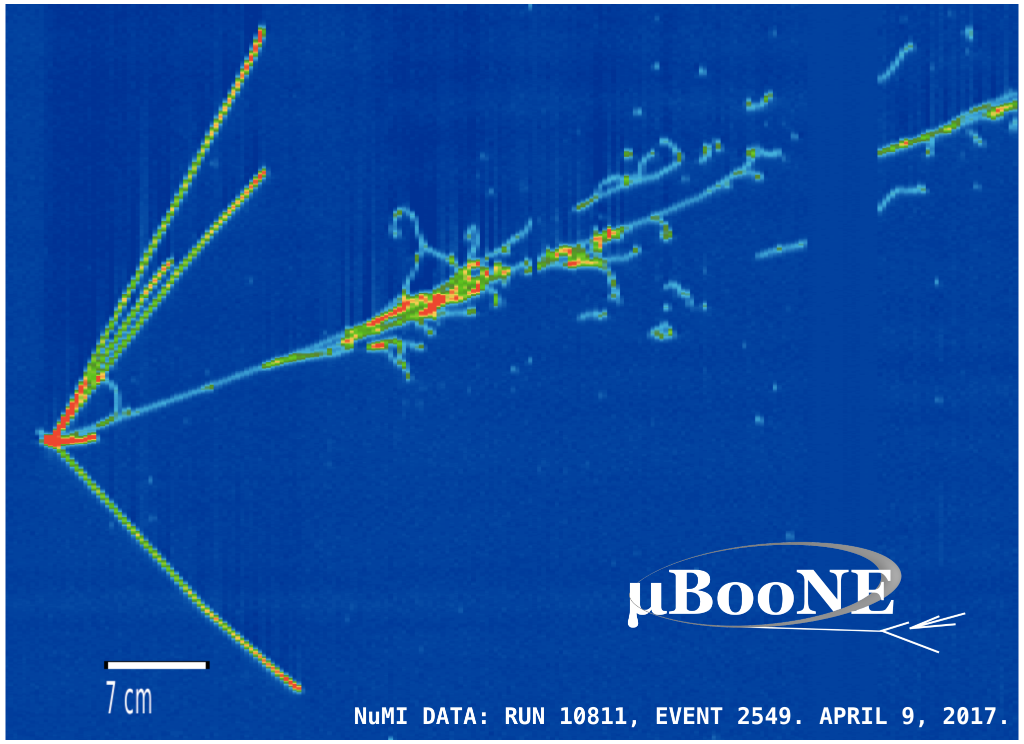


ArgoNeuT

- Evolving detector technologies, with general trend towards imaging neutrino interactions:
  - Emphasis on identifying and characterising individual visible particles
- LArTPCs are fully active and fine grain, offering superb spatial and calorimetric resolution:
  - Need a sophisticated event reconstruction to harness information in LArTPC images
- Physics sensitivity now depends critically on both hardware and software

LArTPC Event Reconstruction





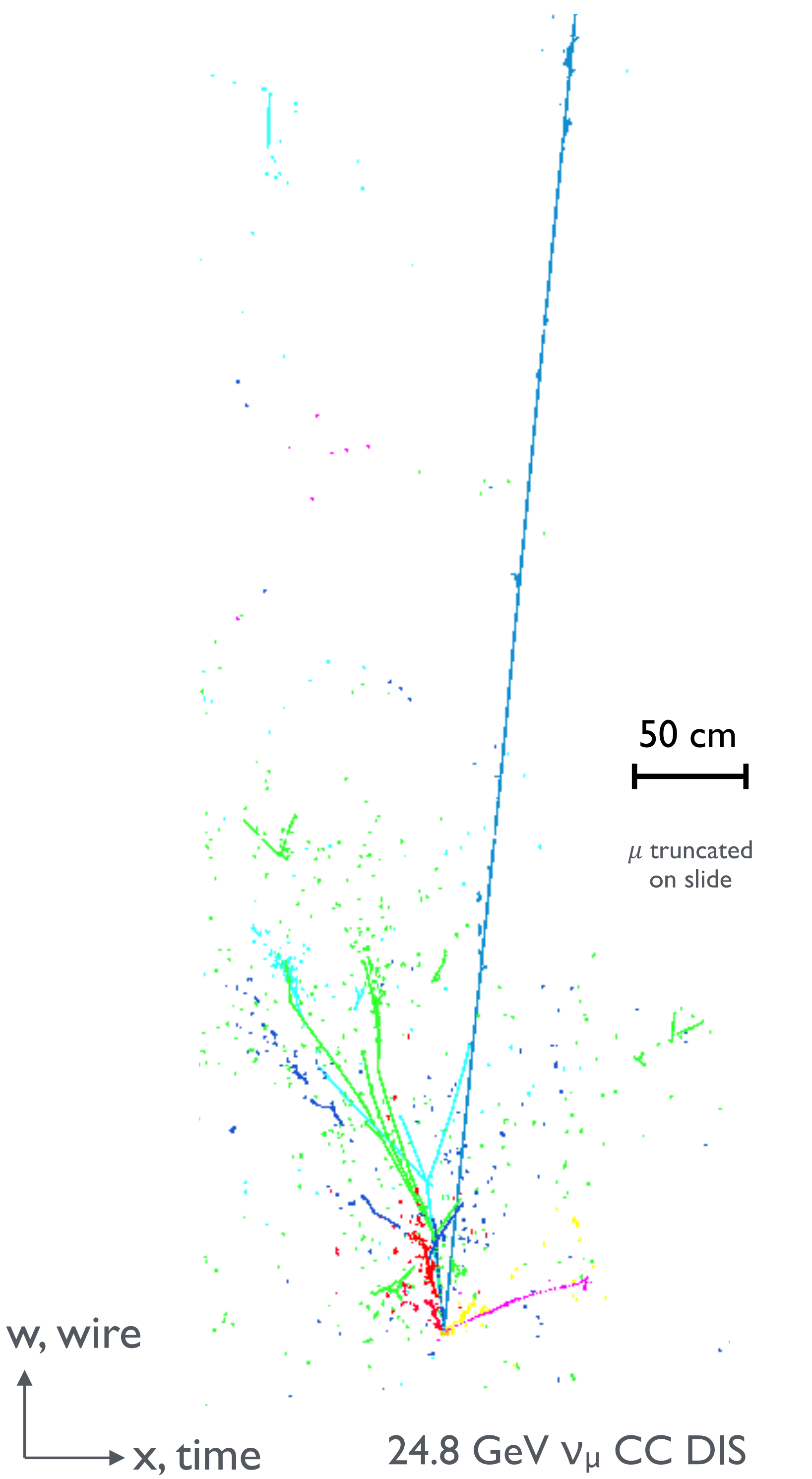
The conversion of raw LArTPC images into analysis-level physics quantities:

- **Low-level steps:**
  - Noise filtering
  - Signal processing
- **Pattern recognition:**
  - The bit you do by eye!
  - Turn images into sparse 2D hits
  - Assign 2D hits to clusters
  - Match features between planes
  - Output a hierarchy of 3D particles
- **High-level characterisation:**
  - Particle identification
  - Neutrino flavour and interaction type
  - Neutrino energy, etc…

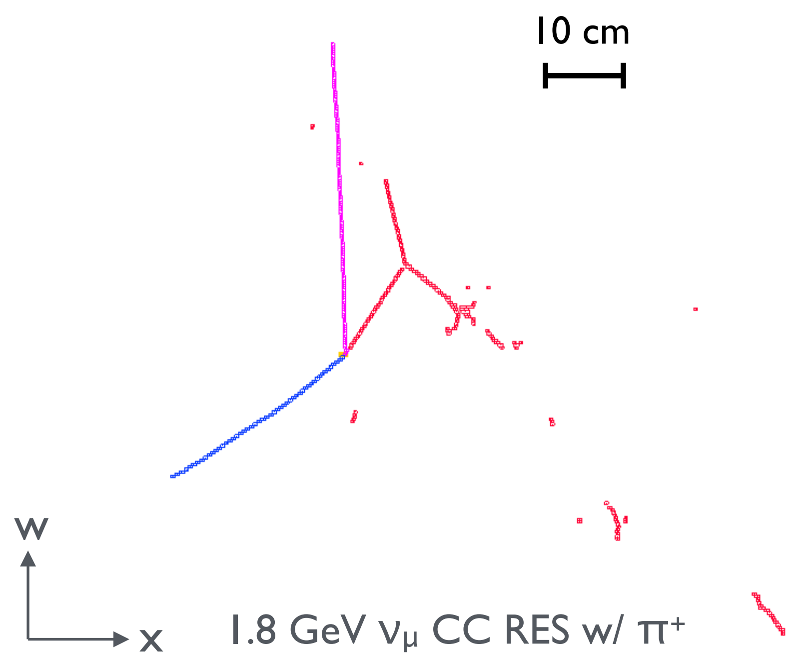
LArTPC Pattern Recognition

It is a significant challenge to develop automated, algorithmic LArTPC pattern recognition

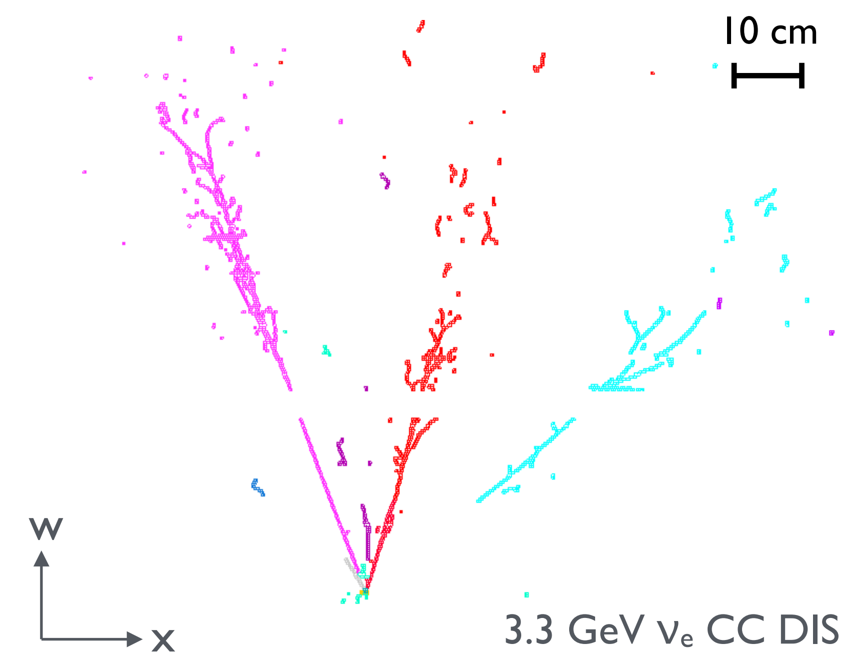
- Complex, diverse topologies:



24.8 GeV ν_μ CC DIS

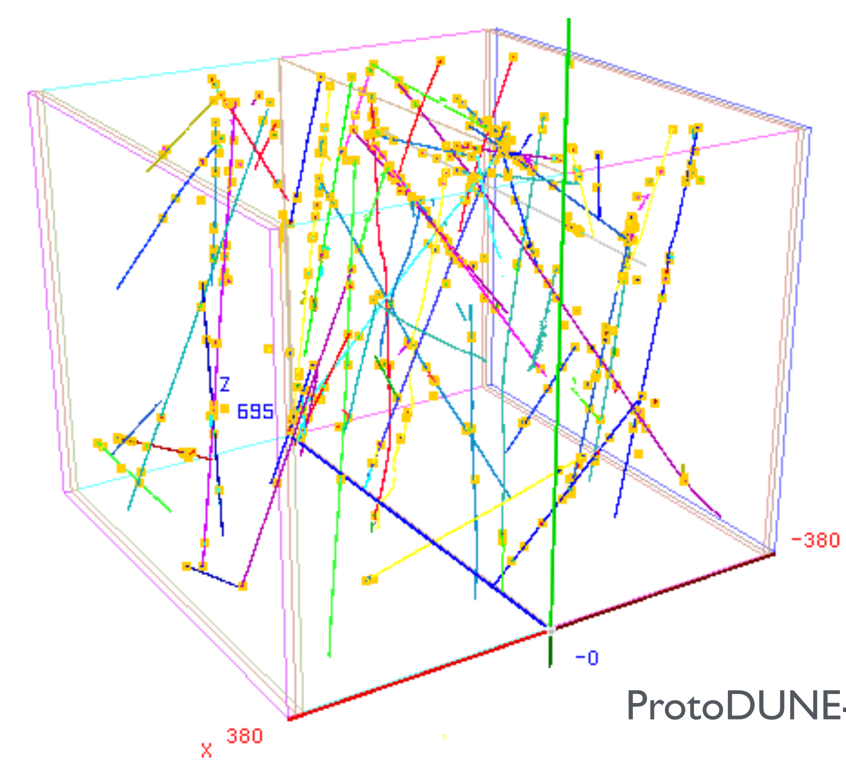


1.8 GeV ν_μ CC RES w/ π^+



3.3 GeV ν_e CC DIS

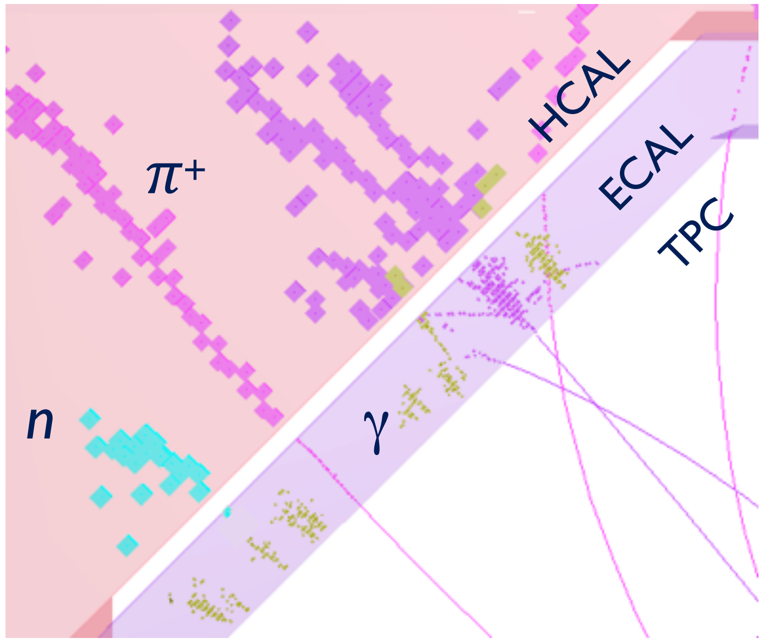
- Also, LArTPCs have long exposures, due to lengthy drift times (up to few ms).
- Significant cosmic-ray muon background in surface-based detectors.



ProtoDUNE-SP

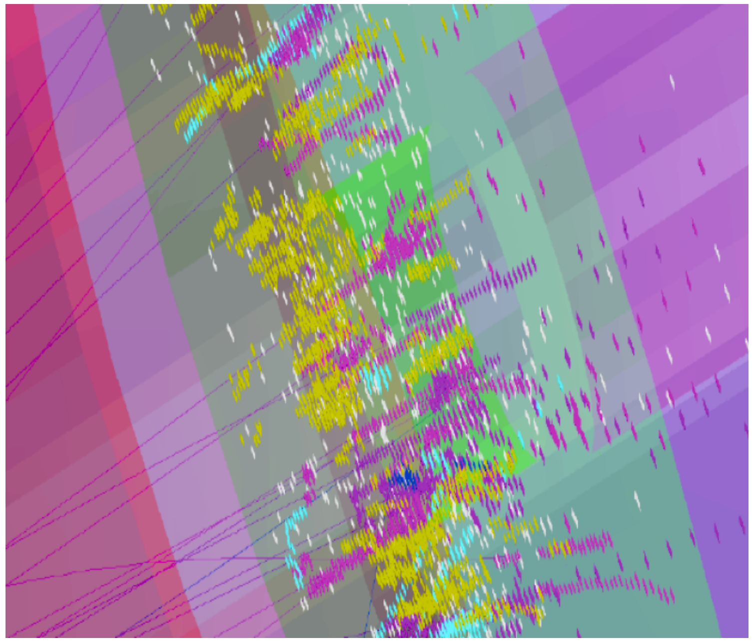
Multi-Algorithm Pattern Recognition

- **Single clustering approach is unlikely to work for such complex topologies:**
  - Mix of track-like and shower-like clusters
- **Pandora project has tackled similar problems before, using a multi-algorithm approach:**
  - Build up events gradually
  - Each step is incremental - aim not to make mistakes (undoing mistakes is hard…)
  - Deploy more sophisticated algorithms as picture of event develops
  - Build physics and detector knowledge into algorithms



Typical ILC event topologies - 3D

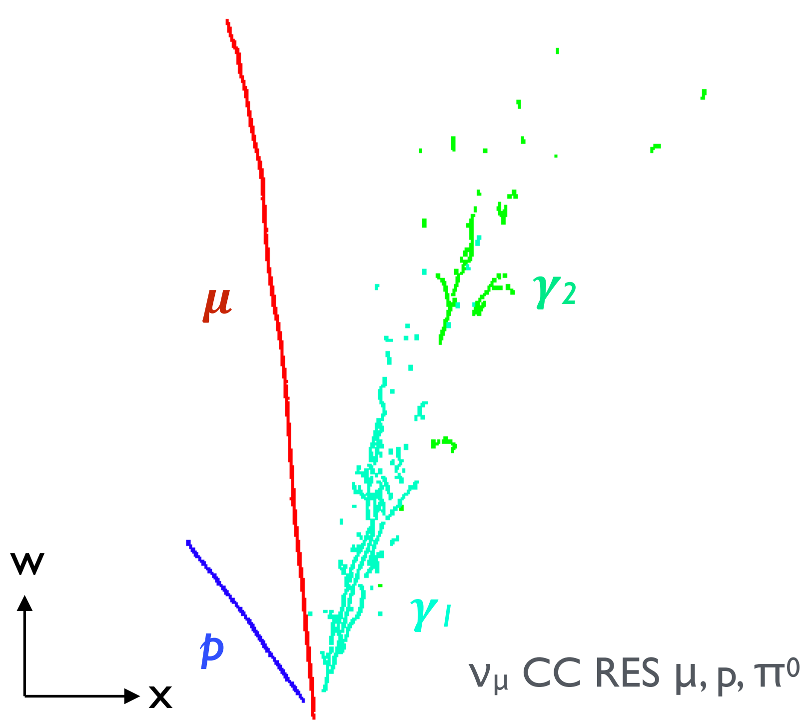
NIMA.2009.09.009 NIMA.2012.10.038



Typical showers in CMS HGCAL - 3D

LHCC-P-008

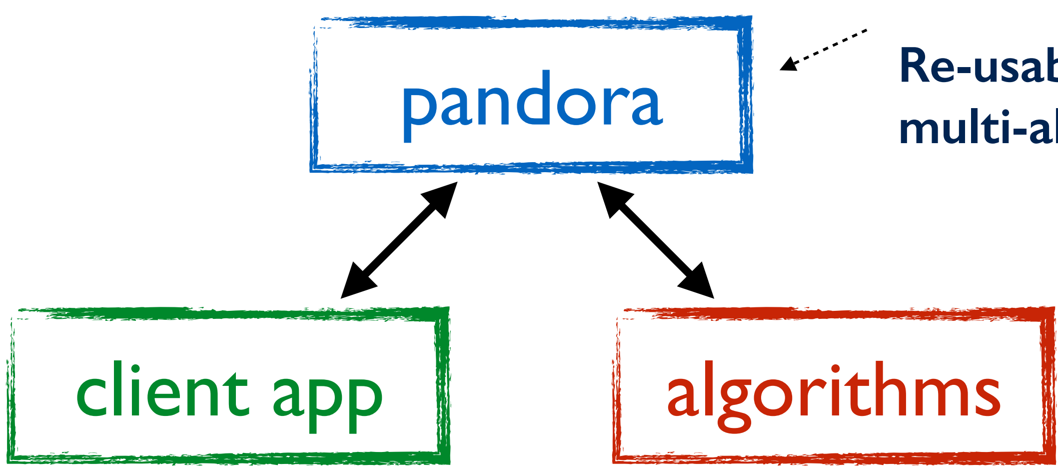
BNB interaction at MicroBooNE - 3 x 2D



Implementation

Pandora Software Development Kit engineered specifically for multi-algorithm approach:

1. Users provide the "building blocks" that define a pattern-recognition problem.
2. Logic to solve pattern-recognition problems cleanly implemented in algorithms.
3. Operations to access/modify building blocks, or create new structures, requested via algs.



pandora sdk and visualisation

Re-usable libraries support multi-algorithm approach

larpandora

Handles input/output LArSoft⟷Pandora

larpandoracontent

137 algorithms and tools, specifically for LArTPC usage

EPJC (2015) 75: 439

github.com/PandoraPFA

Algorithm 1 Cluster creation pseudocode. The logic determining when to create new Clusters and when to extend existing Clusters will vary between algorithms.

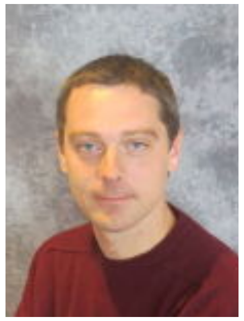
```
procedure CLUSTER CREATION
    Create temporary Cluster list
    Get current CaloHit list
    for all CaloHits do
        if CaloHit available then
            for all newly-created Clusters do
                Find best host Cluster
            if Suitable host Cluster found then
                Add CaloHit to host Cluster
            else
                Add CaloHit to a new Cluster
    Save new Clusters in a named list
```

Simplified **algorithm** implementation

Pandora, A History

c.2006

Mark Thomson creates Pandora project to provide fine-granularity "particle flow" reconstruction for ILC. Starts at back of lecture theatre, at ILC workshop.



2009

Brought **John Marshall** on board for implementation: created Pandora SDK and linear collider particle flow algs used by most/all studies at ILC and CLIC.

2013

JM and **Andy Blake** reunite (previously both on MINOS) to develop a multi-algorithm approach to LArTPC pattern recognition. Join MicroBooNE.

2016

Lorena Escudero (background T2K) joins to help us deliver the pattern recognition for MicroBooNE and plan/develop for DUNE.

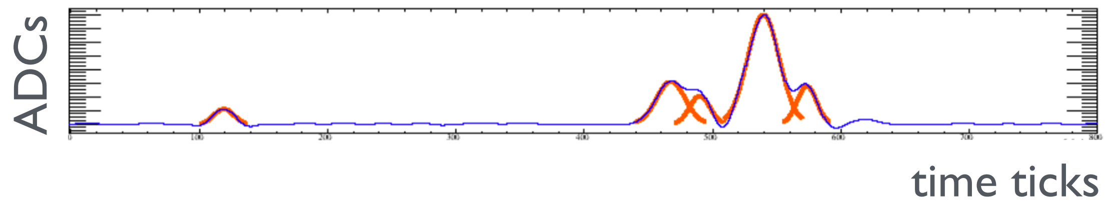
2017

Steve Green (background Pandora for ILC/CLIC) joins to help us deliver the pattern recognition for ProtoDUNE (and plan/develop for DUNE FD).

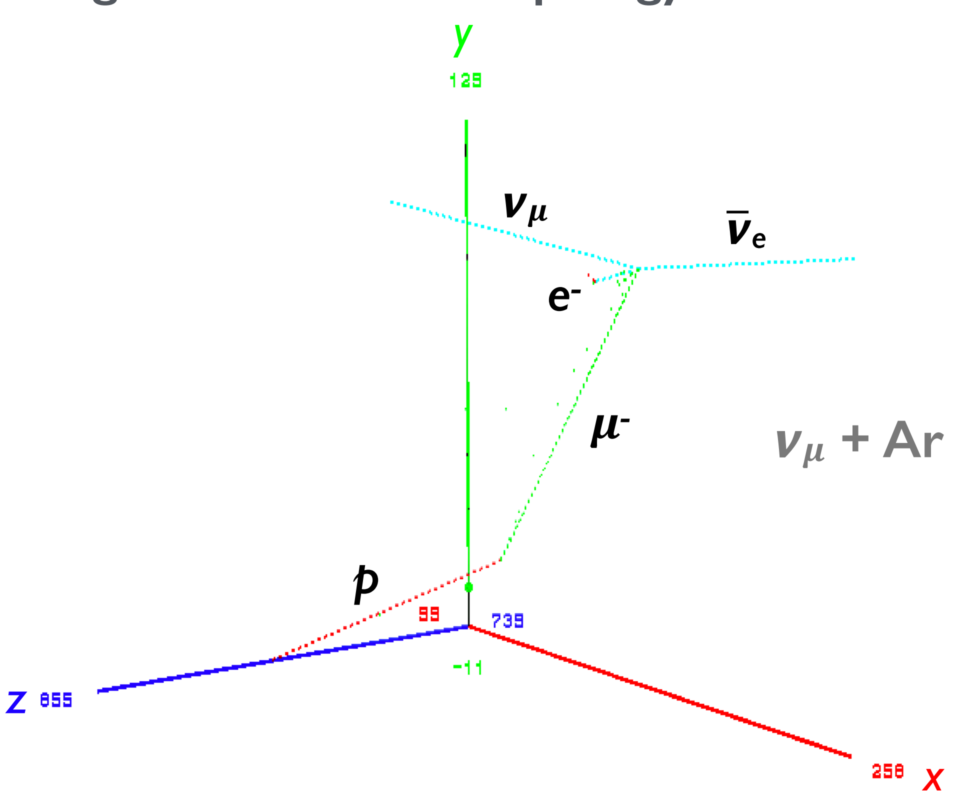
+ Now 7 grad students involved, who deserve a bigger mention than this text box!

Input: 3x2D images, known wire positions [cm] vs. recorded positions from drift times [cm]

E.g. Hits found for an individual wire:

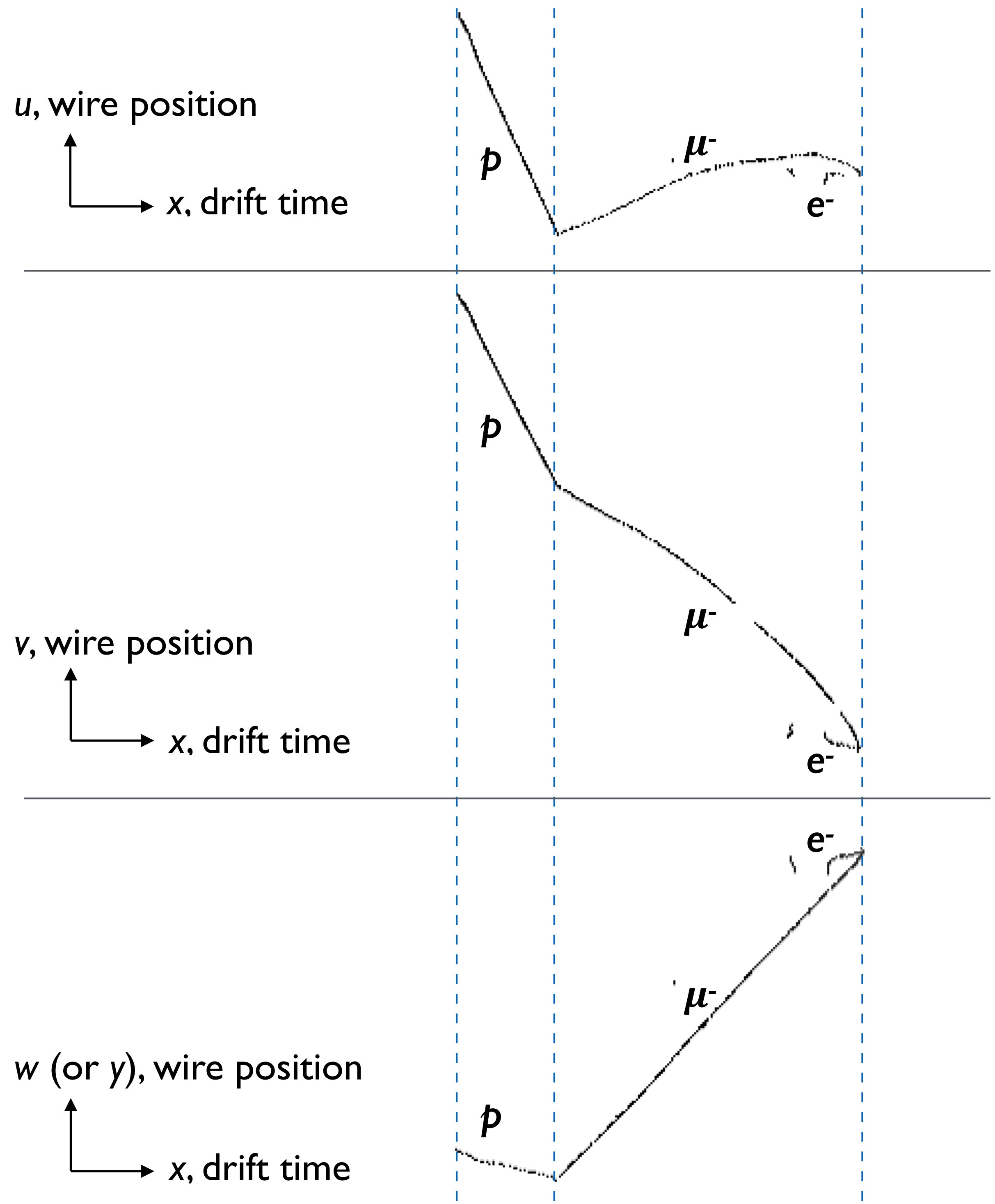


E.g True 3D event topology:



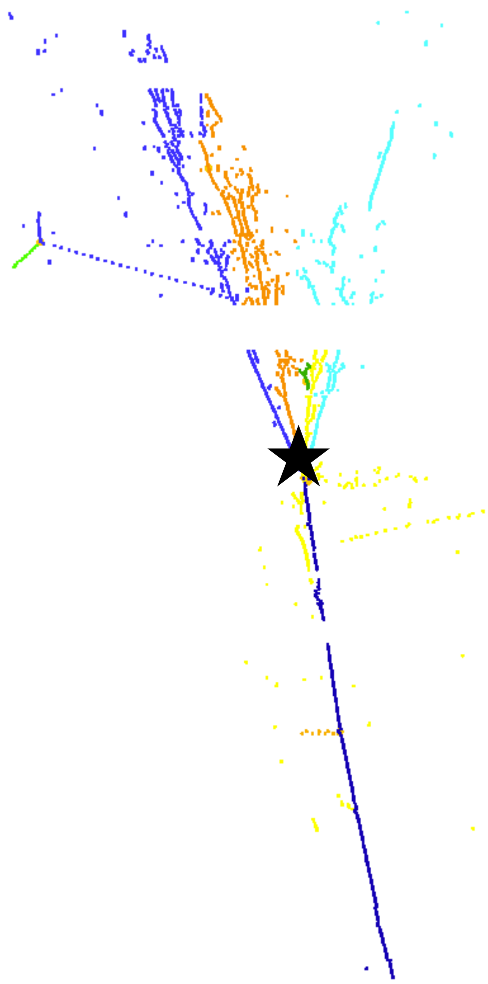
$\nu_\mu + \mathrm{Ar} \rightarrow p + \mu^-$

3x2D representations with common coordinate derived from drift time, "x"



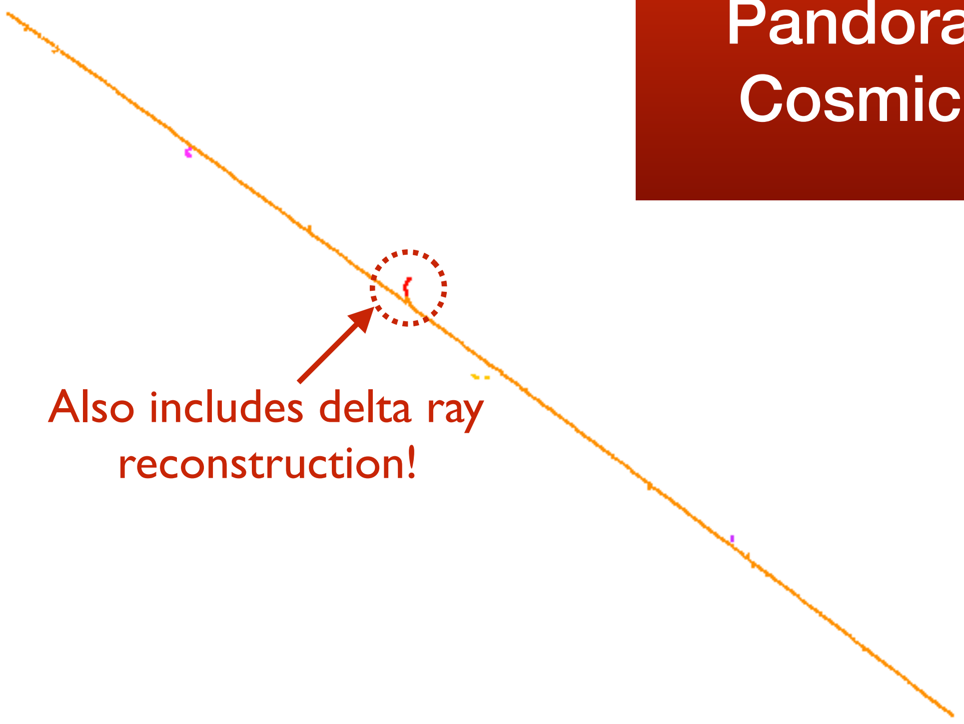
Pandora Algorithm Chains

- **Use multi-algorithm approach to create two algorithm chains for LArTPC usage.**
  - Consolidated reconstruction uses these chains to guide reconstruction for all use cases: Cosmic rays ✔, Multiple drift volumes ✔, Arbitrary wire angles ✔, 2 or 3 wire planes ✔



Pandora Test Beam

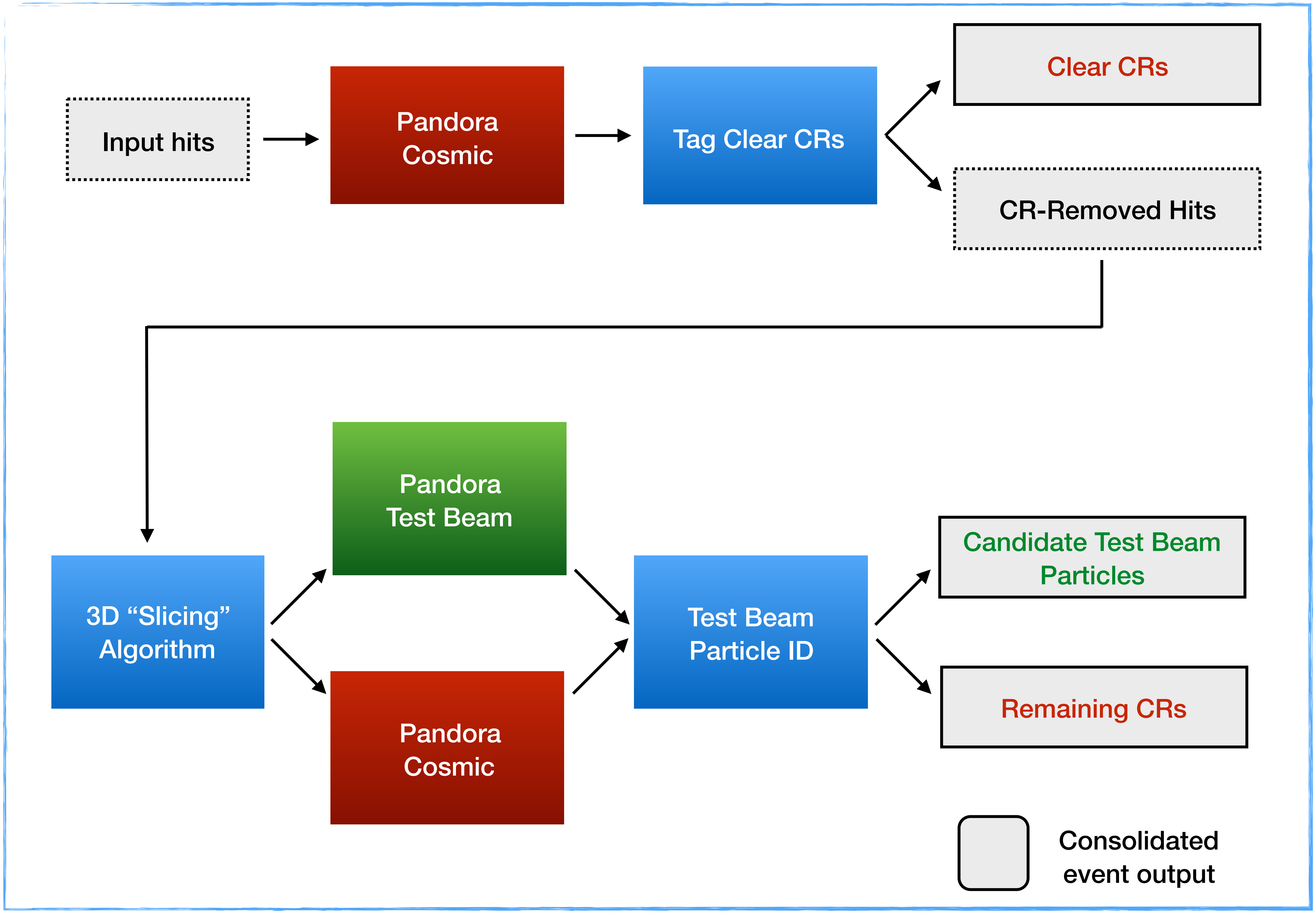
Targets reconstruction of particles emerging from an identified vertex ★



Pandora Cosmic

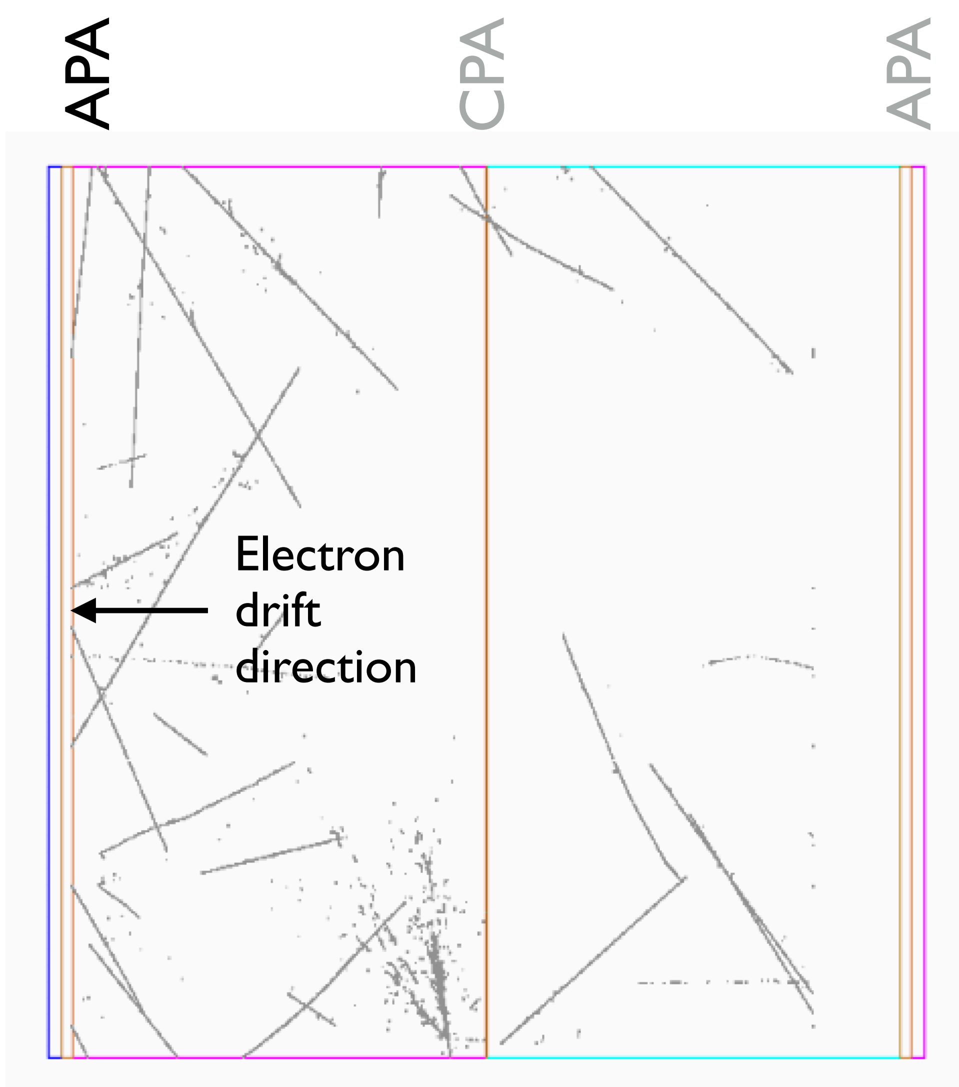
Targets the reconstruction of straight-line particles in the detector (e.g. cosmic rays)

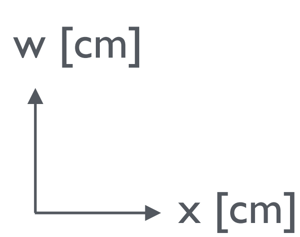
Consolidated Reconstruction

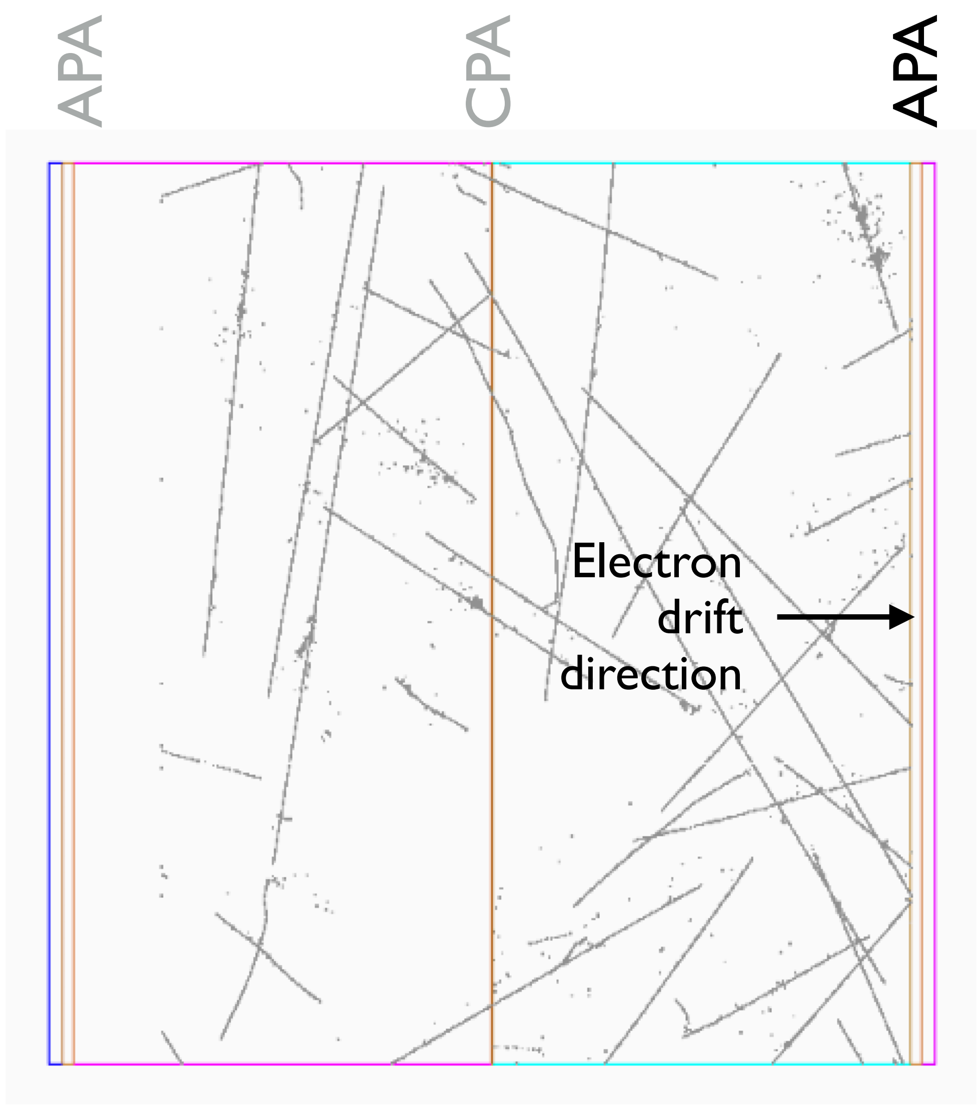


Event Reconstruction at ProtoDUNE-SP

- **Multiple "drift volumes", complex topologies and significant cosmic-ray activity:**
  - A fantastic workout for LArTPC pattern recognition!





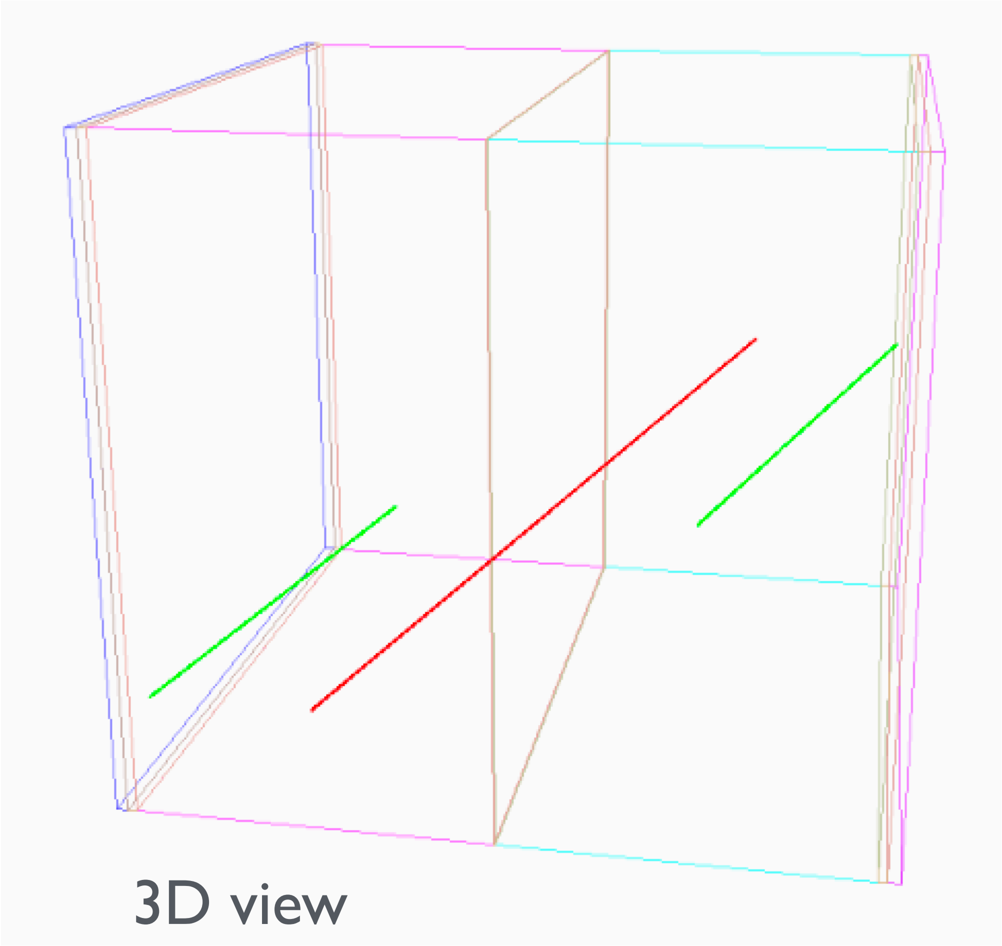


APA: Anode Plane Assembly
CPA: Cathode Plane Assembly

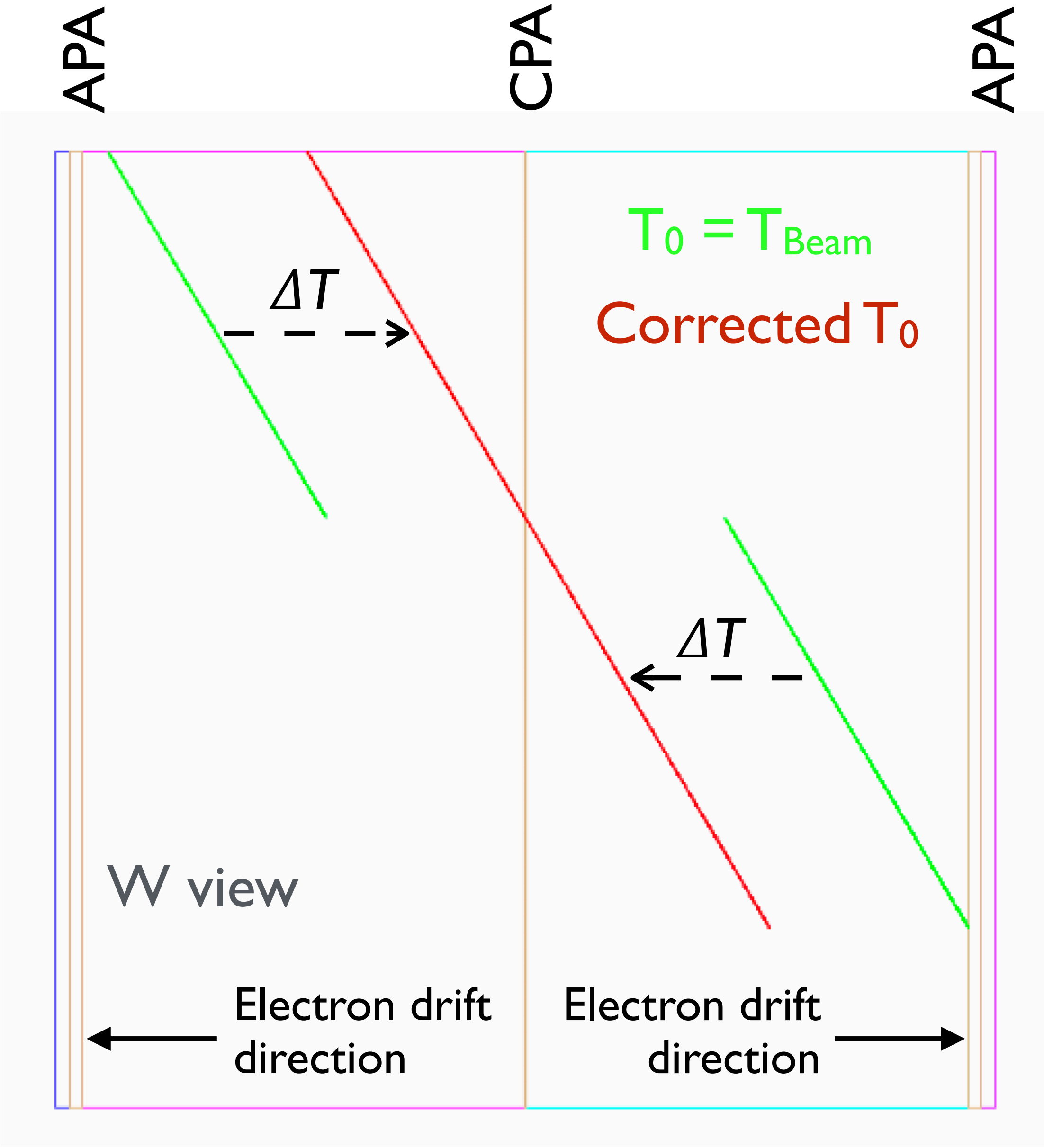
1. Reconstruct cosmic-ray muons independently for each volume of detector

Stitching and T_0 Identification

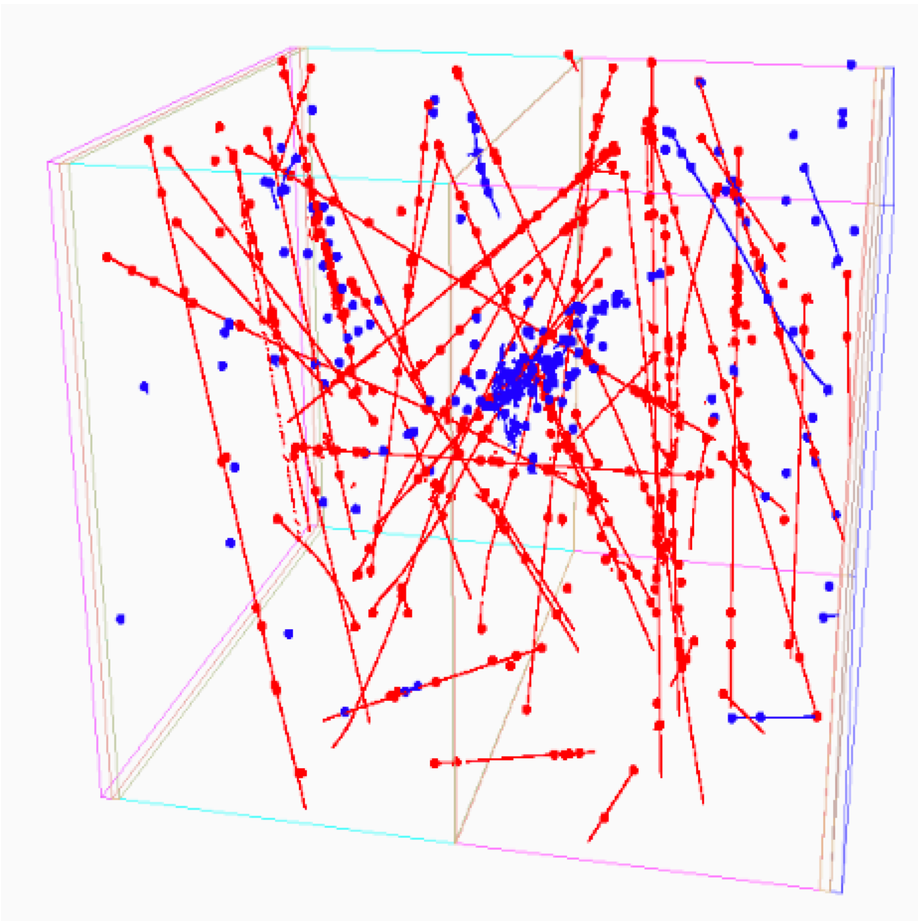
- **For detectors with multiple drift volumes, can determine the true particle time if it crosses an enclosed cathode (or anode) plane. This process is called "stitching".**
  - By shifting pairs of reconstructed particles in different drift volumes by an equal amount in drift time, cosmic rays (with a different T_0 to the target TB/ν) can be identified.



2. Stitch together any cosmic rays crossing between volumes, identifying T_0



Cosmic Ray Tagging and Slicing

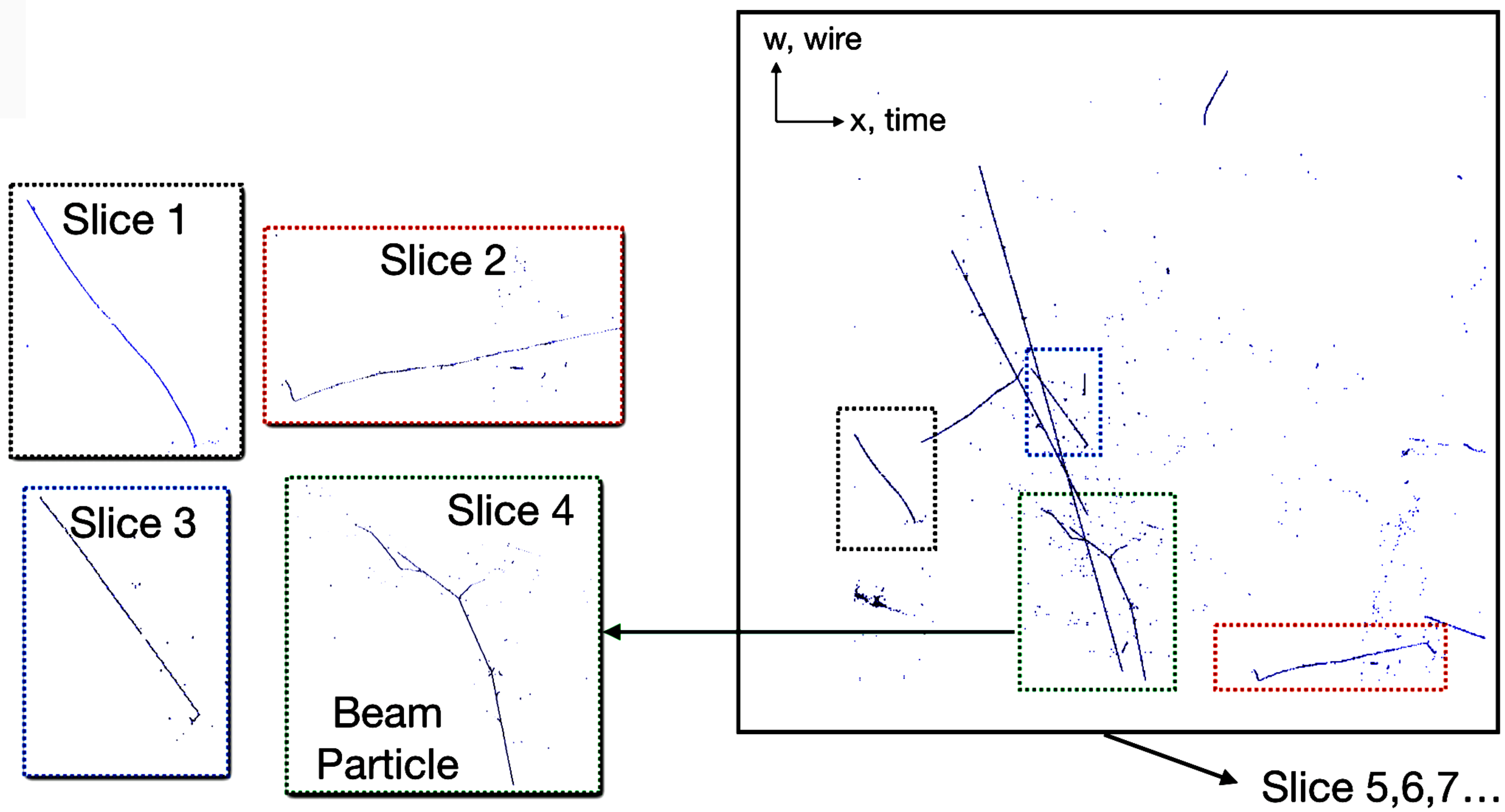


3. Identify clear cosmic rays (red) and hits to reexamine under test beam hypothesis (blue)

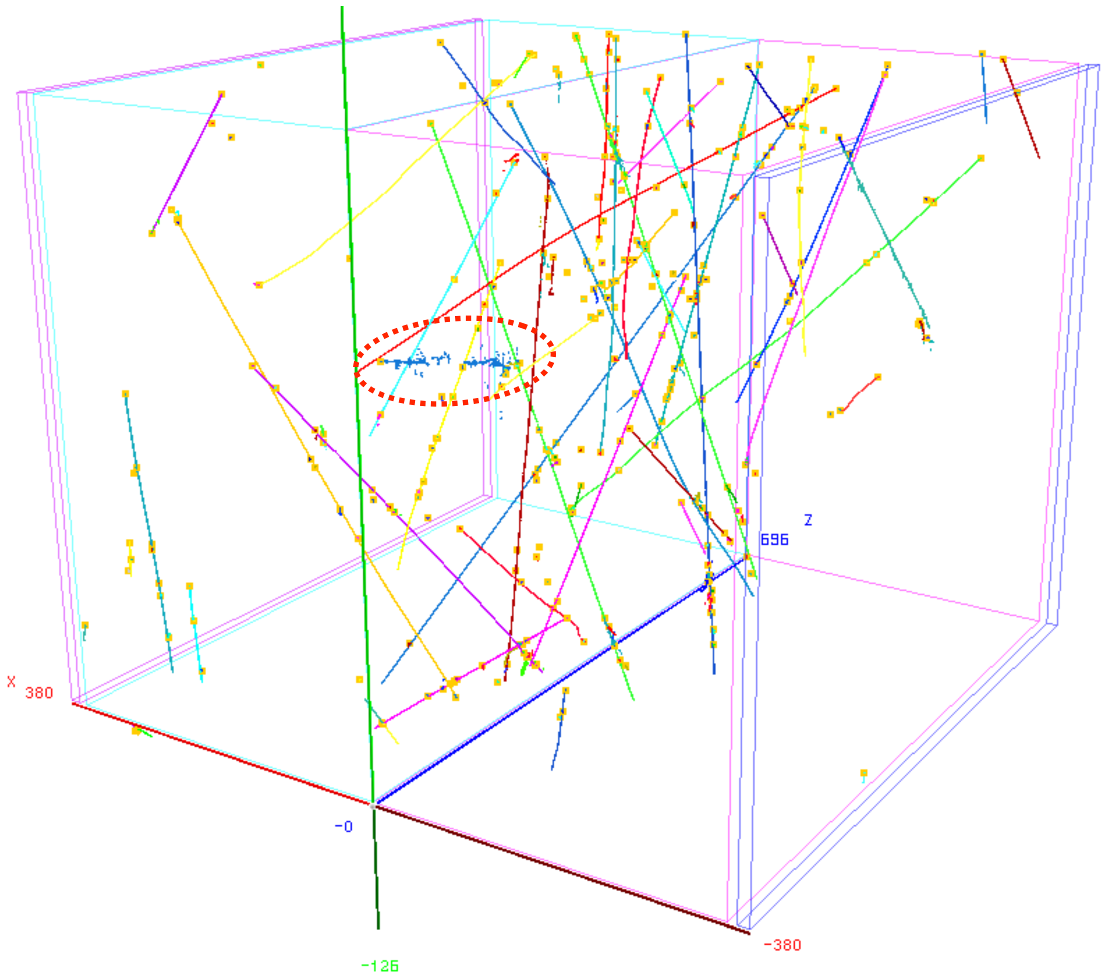
Clear cosmic rays:

- Particles appear to be "outside" of detector if $T_0 = T_{Beam}$
- Particles stitched between volumes using a $T_0 \neq T_{Beam}$
- Particles pass through the detector: "through going"

- Slice/divide blue hits from separate interactions
- Reconstruct each slice as test beam particle
- **Then choose between cosmic ray or test beam outcome for each slice**

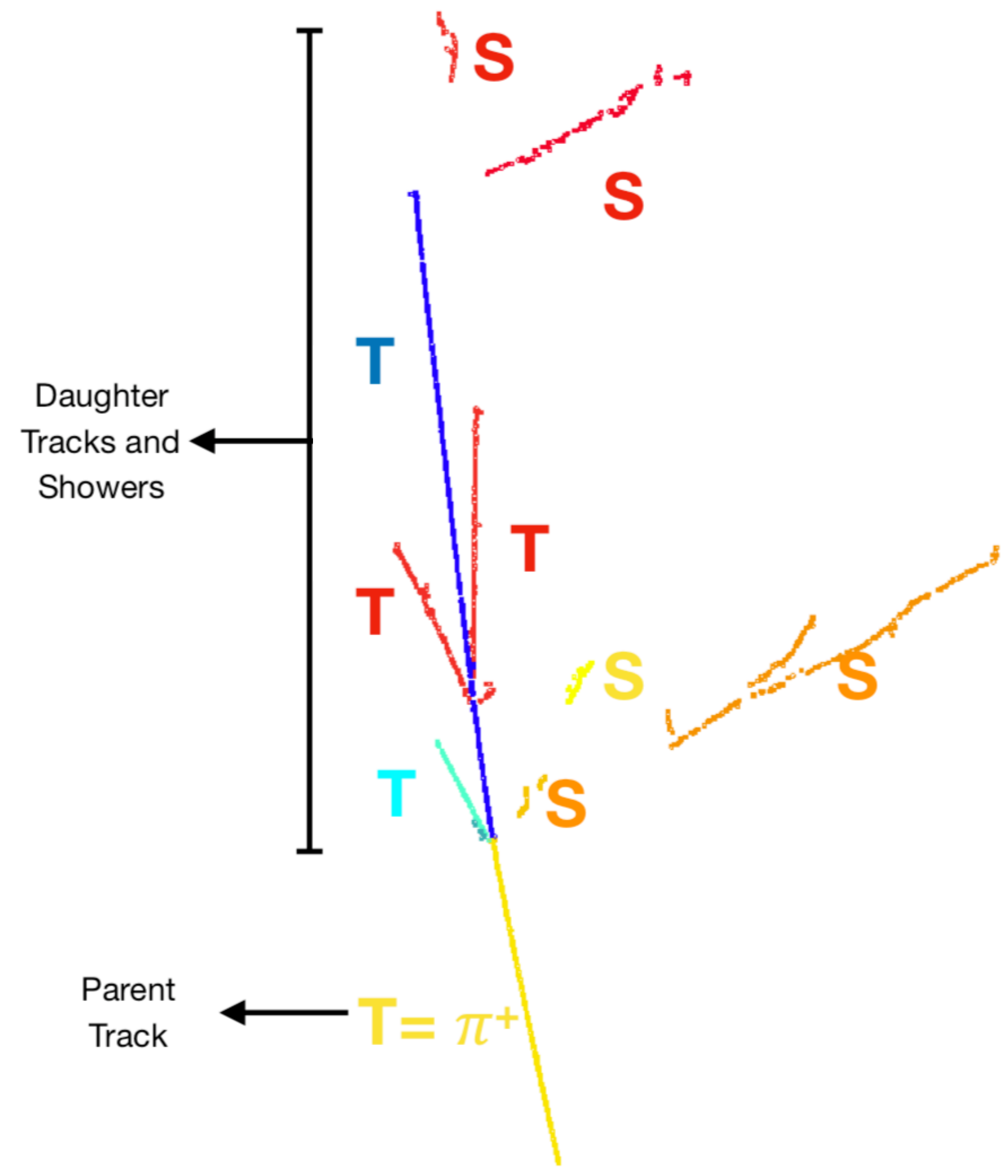


Consolidated Output



E.g. Reconstruction output: test beam particle (electron)

and: *N* reconstructed cosmic-ray muon hierarchies

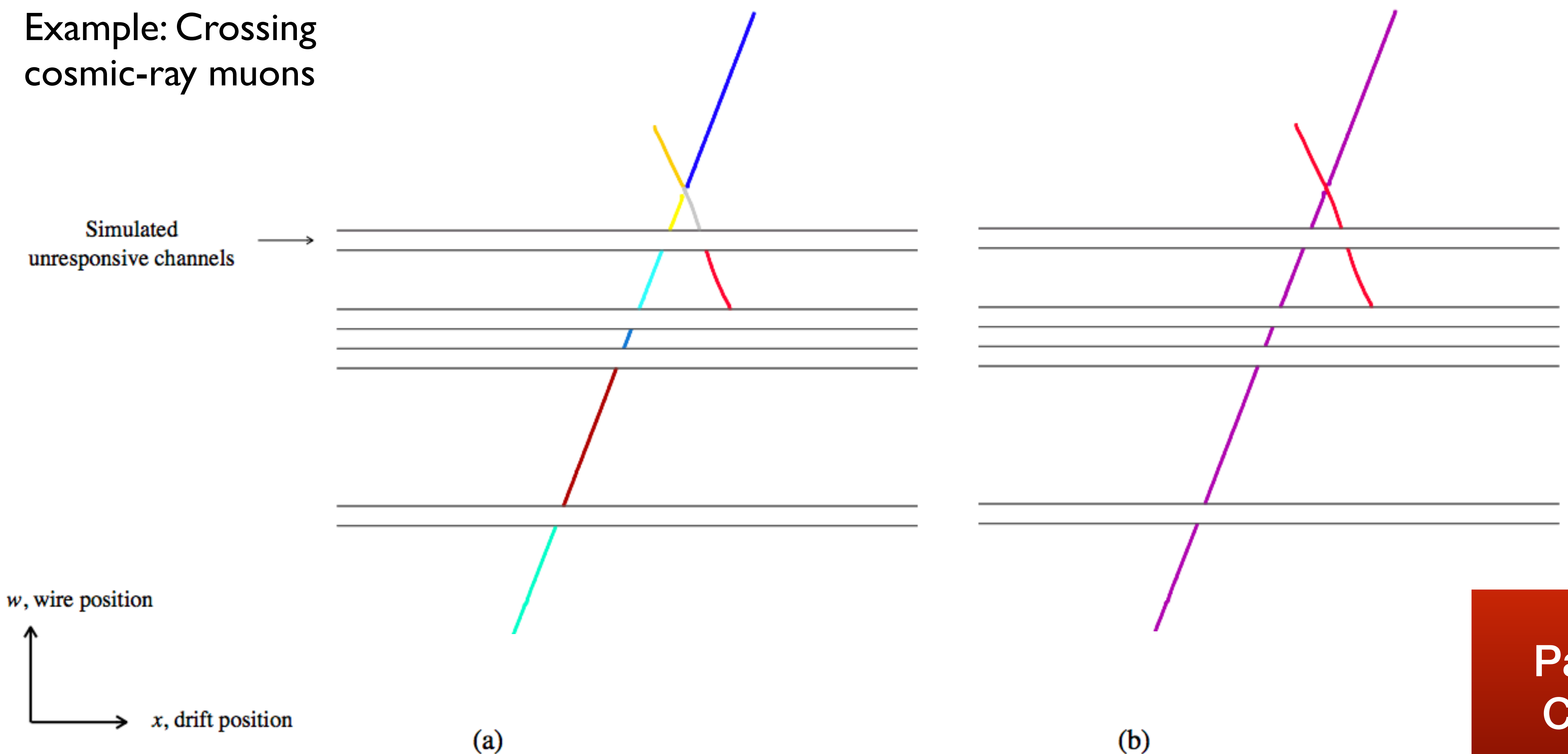


E.g. Test beam particle: charged pion

Cosmic-Ray Muon Reconstruction - 2D

- For each plane, produce list of 2D clusters that represent continuous, unambiguous lines of hits:
  - Separate clusters for each structure, with clusters starting/stopping at each branch or ambiguity.
- Clusters refined by series of 15 cluster-merging and cluster-splitting algs that use topological info.

Example: Crossing cosmic-ray muons

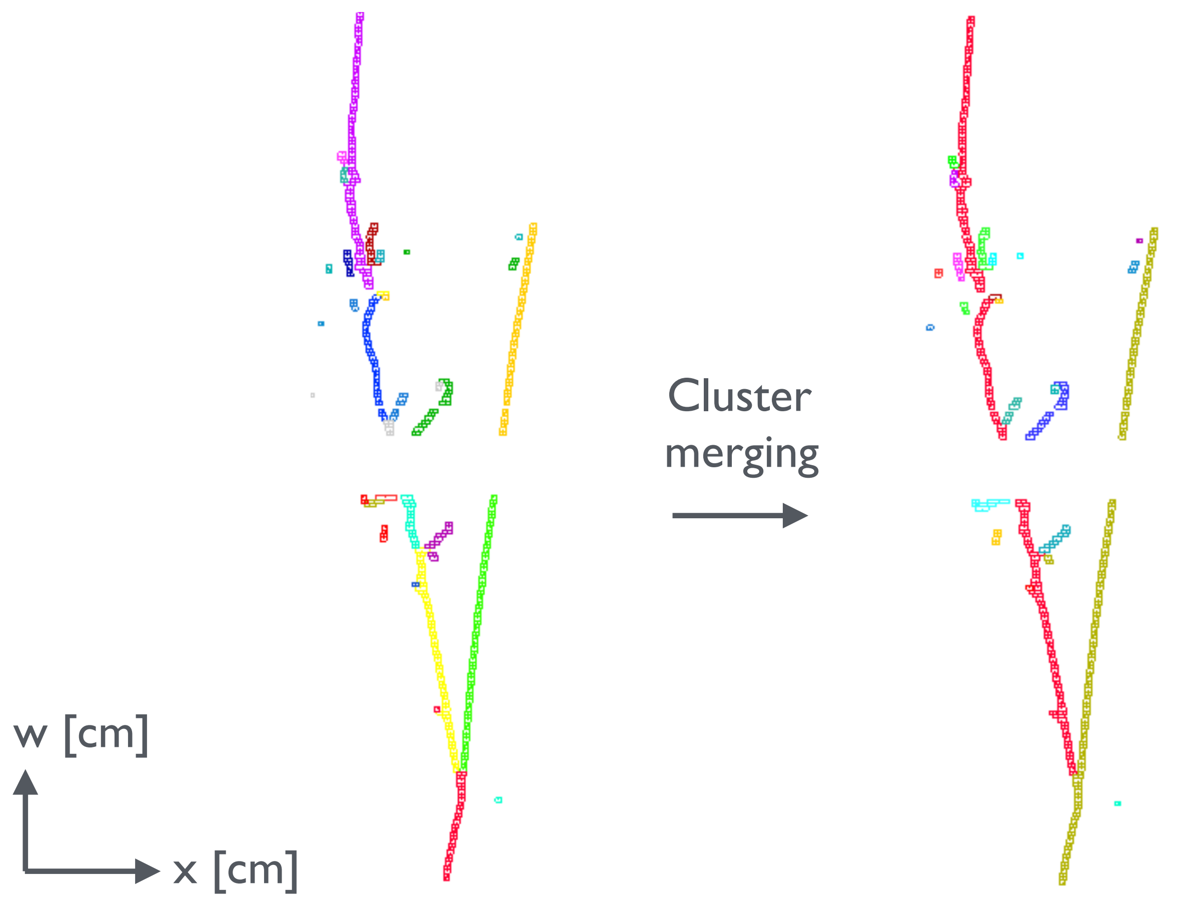


Pandora Cosmic

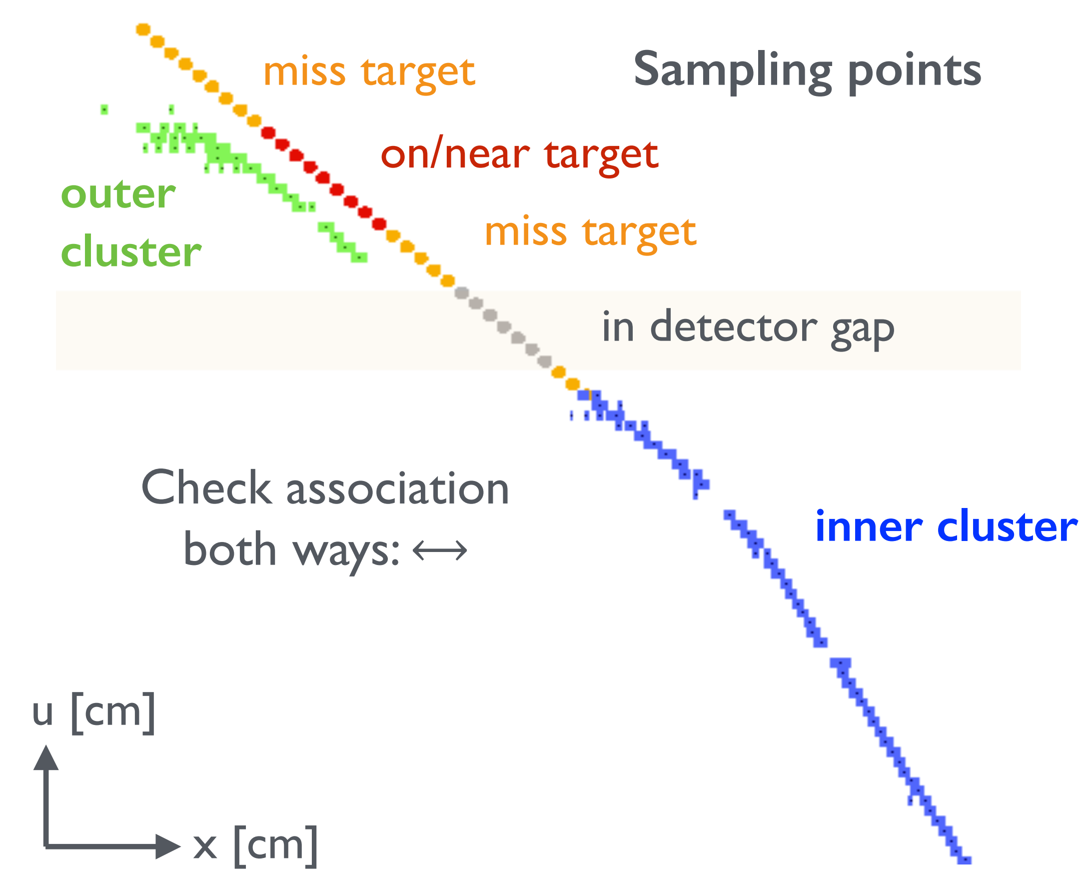
Topological Association - 2D

- **Cluster-merging algorithms identify associations between multiple 2D clusters and look to grow the clusters to improve completeness, without compromising purity.**
  - The challenge for the algorithms is to make cluster-merging decisions in the context of the entire event, rather than just considering individual pairs of clusters in isolation.

E.g. LongitudinalAssociation

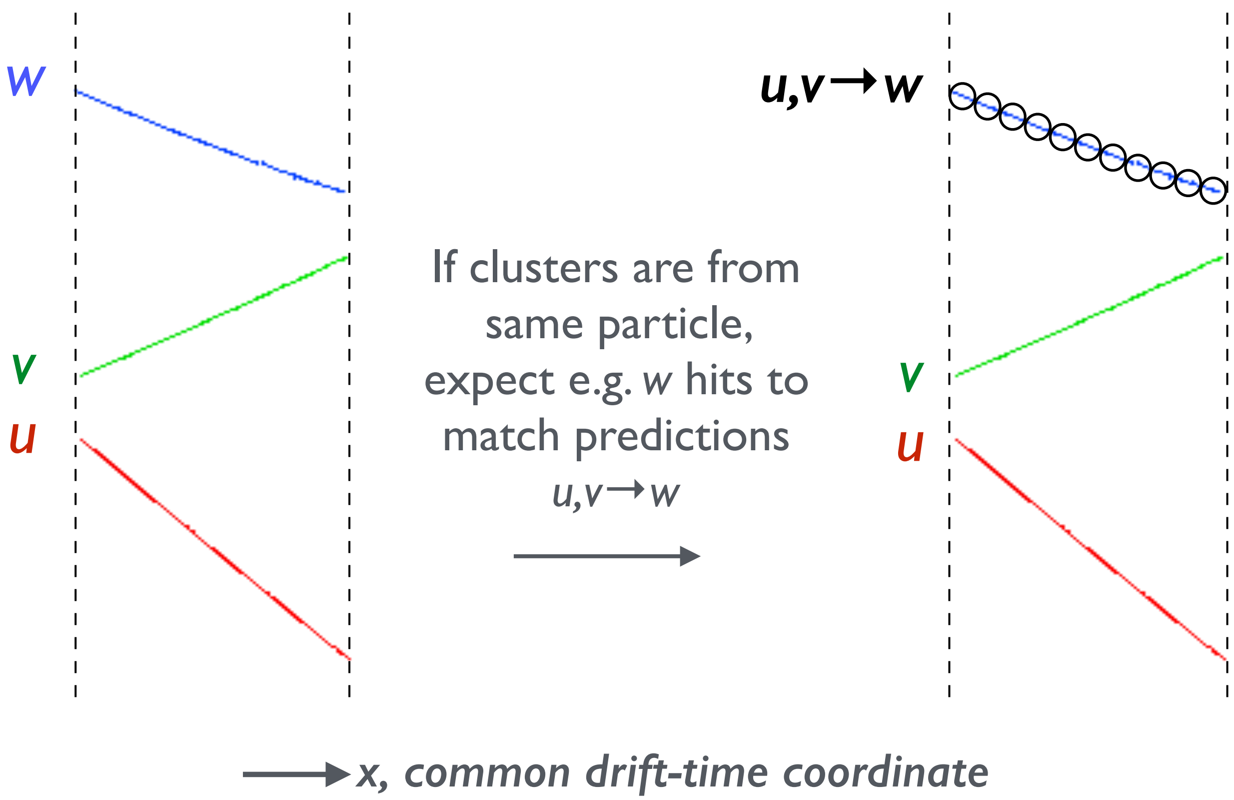


E.g. CrossGapsAssociation

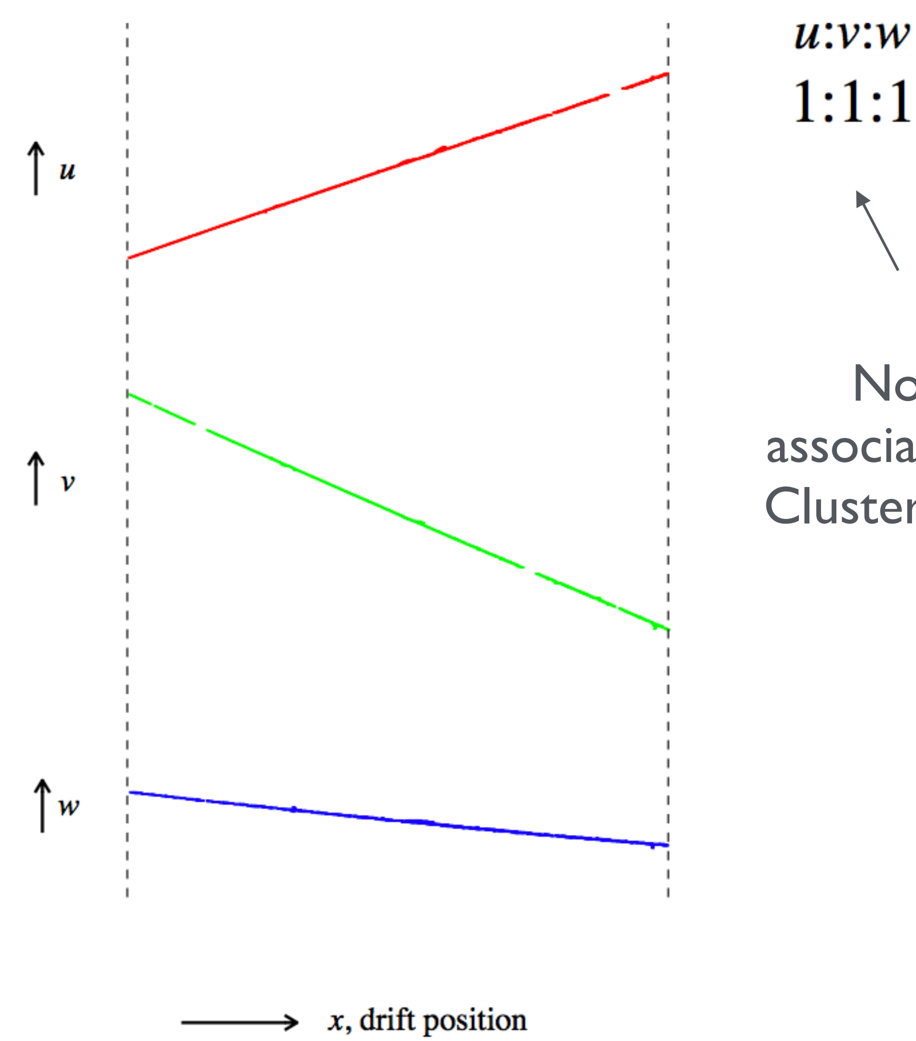


Track Pattern Recognition - 3D

- **Our original input was 3x2D images of charged particles in the detector.**
- **Should now have reconstructed three separate 2D clusters for each particle:**
  - Compare 2D clusters from *u*, *v*, *w* planes to find the clusters representing same particle.
  - Exploit common drift-time coordinate and our understanding of wire plane geometry.



- **Store matching information for all cluster combinations, then carefully examine results:**

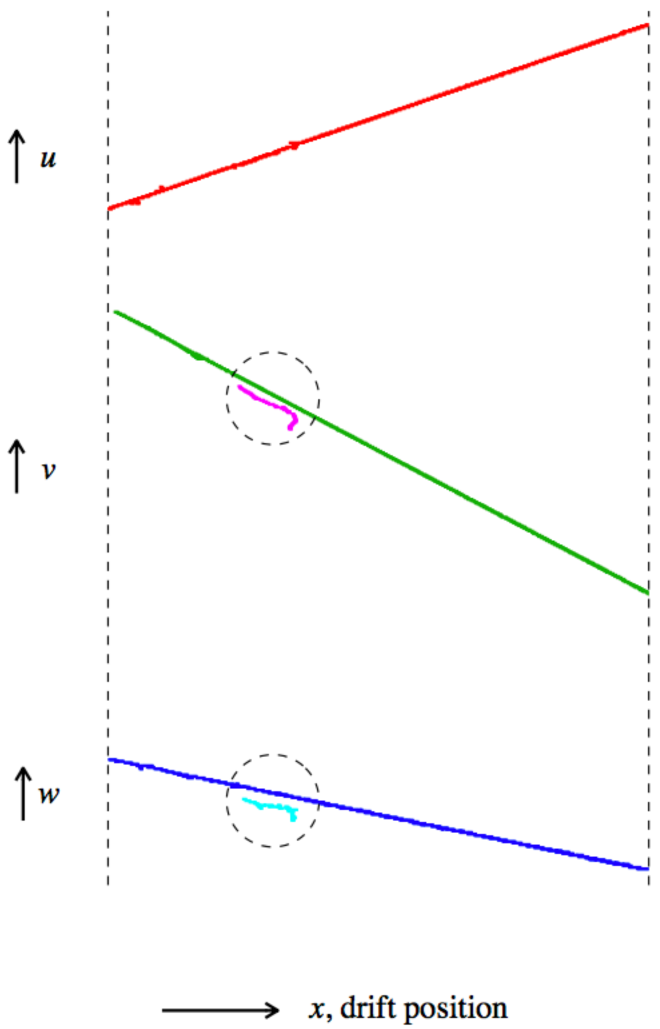


Easiest cases first: unambiguous matches, demanding that the common x-overlap is 90% of the x-span for all three clusters.

Clear Tracks

Track Pattern Recognition - 3D

- **Begin to use cluster-matching information to resolve ambiguities and improve 2D PatRec:**
  - **Simple ambiguities first: clusters matched in multiple ways, but one combination "best"**

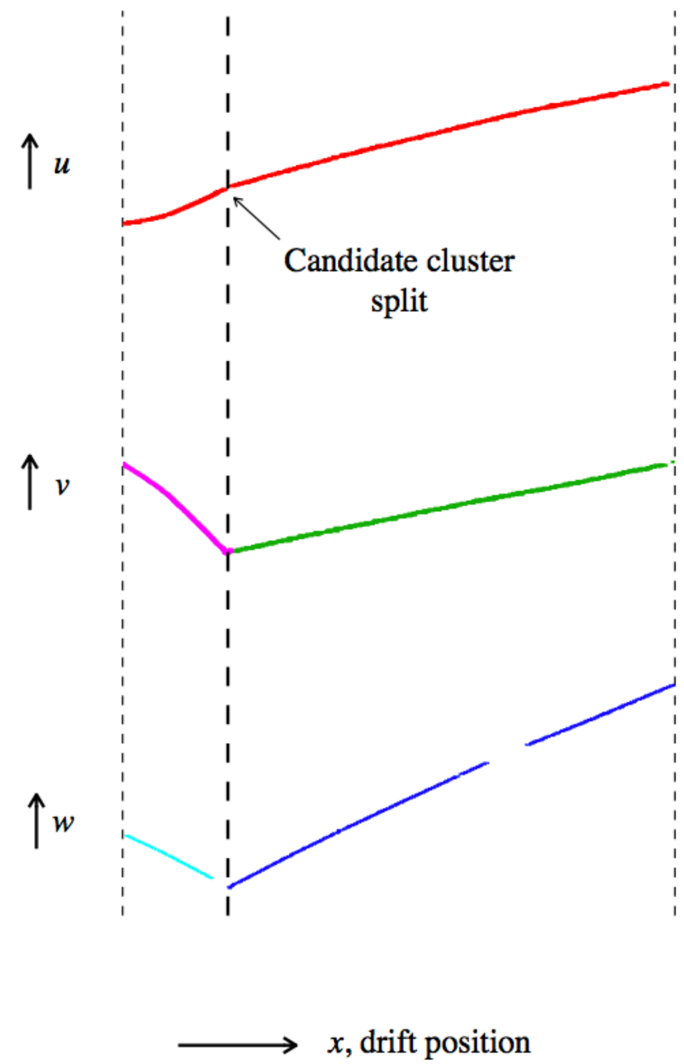


u:v:w
1:2:2

Ringed clusters in *v* and *w* views also match *u* cluster, so matching ambiguous

Long Tracks

- **Approach really comes to life when the 2D clustering "disagrees" between wire planes:**
  - **Automated detection of 2D PatRec issues, with treatment for specific cases, e.g.:**



u:v:w
1:2:2

Two clusters in *w* and *v*, matched to common *u* cluster.

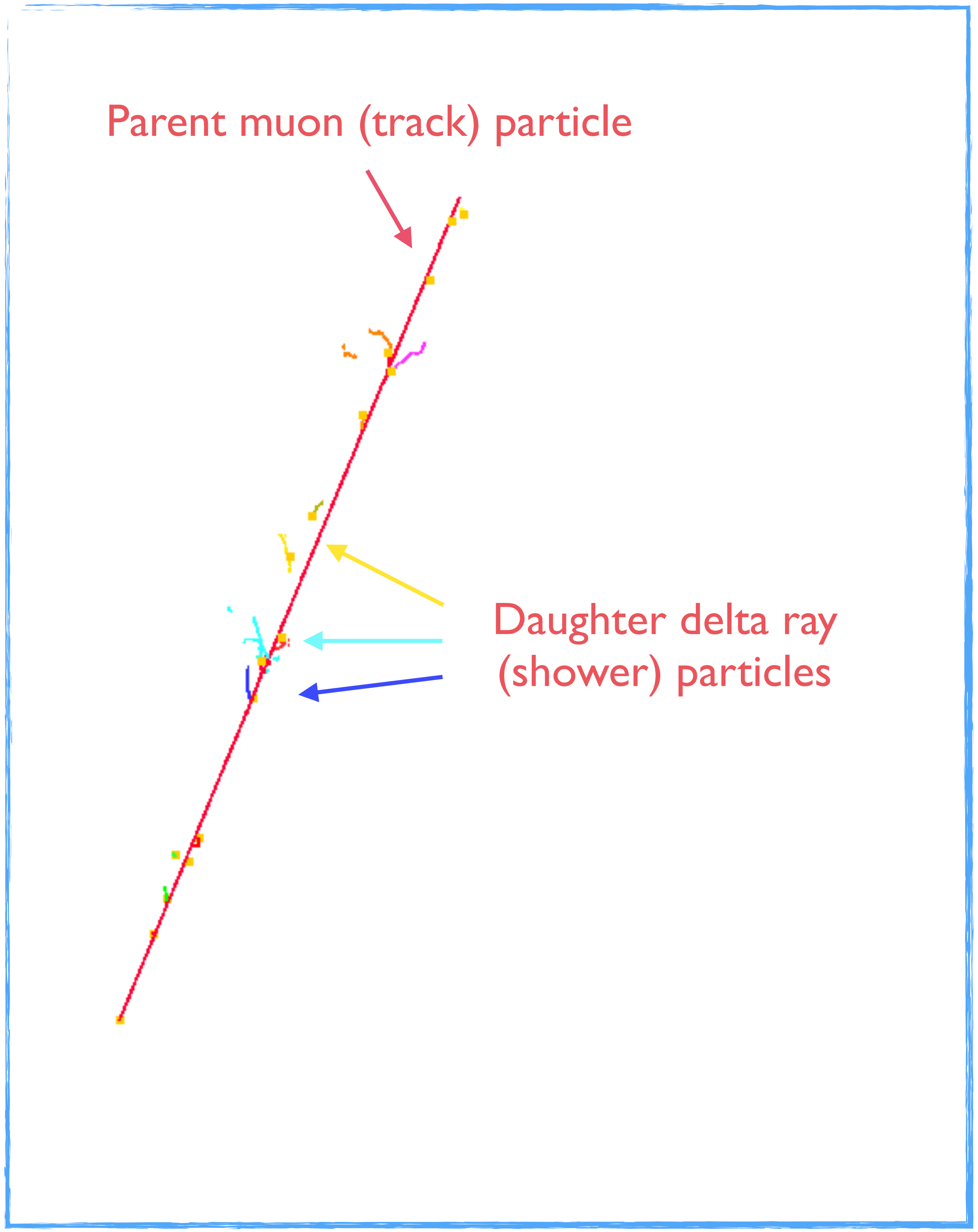
Split *u* cluster; truly two particles.

Check to see if is a kink topology in 3D

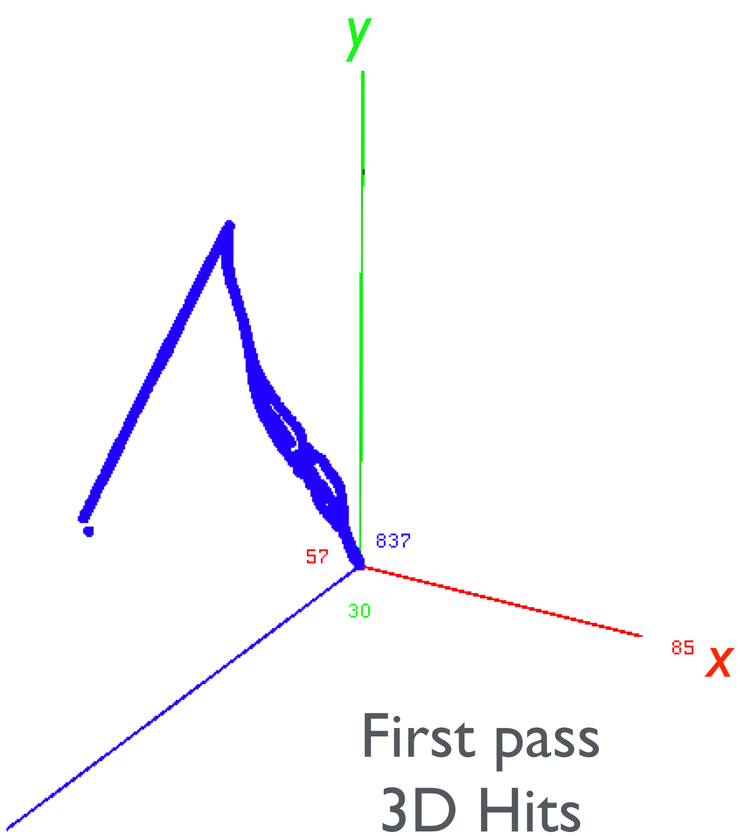
Overshoot Tracks

Delta-Ray Reconstruction - 2D, 3D

- **Assume any 2D clusters not in a track particle are from delta-ray showers:**
  - Simple proximity-based reclustering of hits, then topological association algs.
  - Delta-ray clusters matched between views, creating delta-ray shower particles.
  - Parent muon particles identified and delta-ray particles added as daughters.

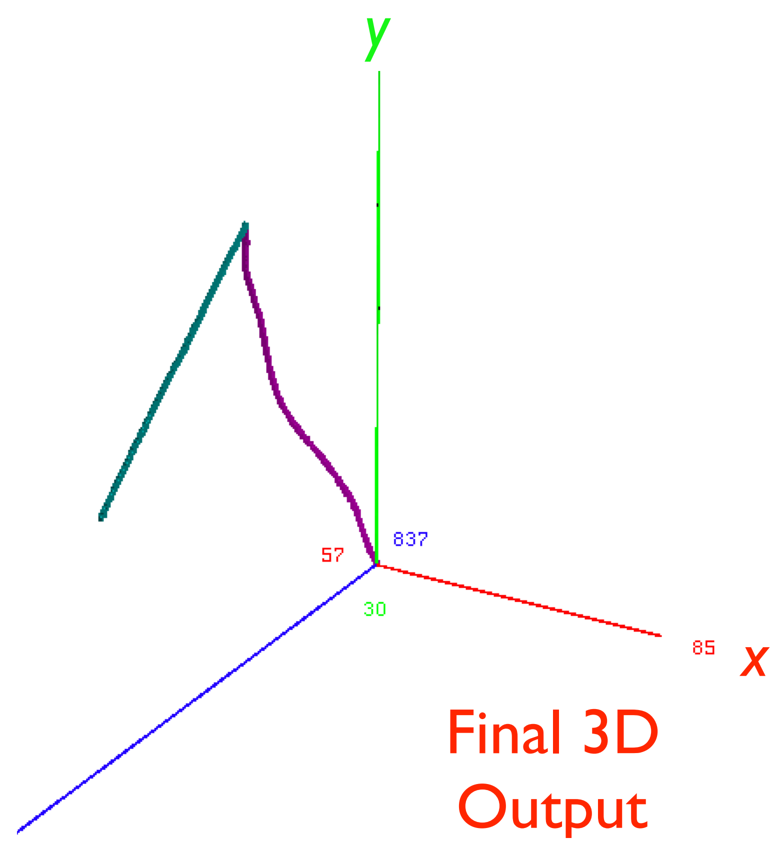


- For each 2D Hit, sample clusters in other views at same x, to provide u_{in}, v_{in} and w_{in} values
- Provided u_{in}, v_{in} and w_{in} values don't necessarily correspond to a specific point in 3D space
- Analytic expression to find 3D space point that is *most consistent* with given u_{in}, v_{in} and w_{in}
  - $\chi^2 = (u_{out} - u_{in})^2 / \sigma_u^2 + (v_{out} - v_{in})^2 / \sigma_v^2 + (w_{out} - w_{in})^2 / \sigma_w^2$
  - Write in terms of unknown y and z, differentiate wrt y, z and solve
  - Can iterate, using fit to current 3D hits (extra terms in χ^2) to produce smooth trajectory



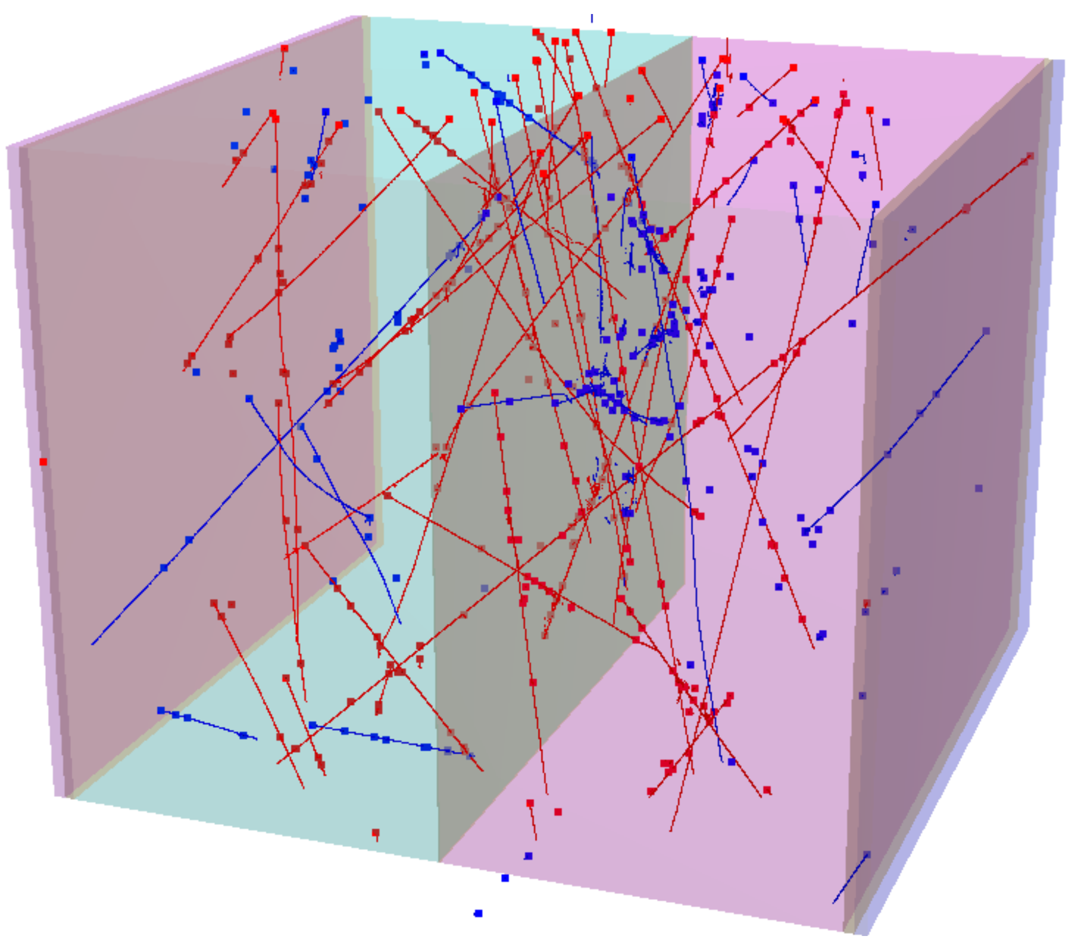
First pass 3D Hits

→

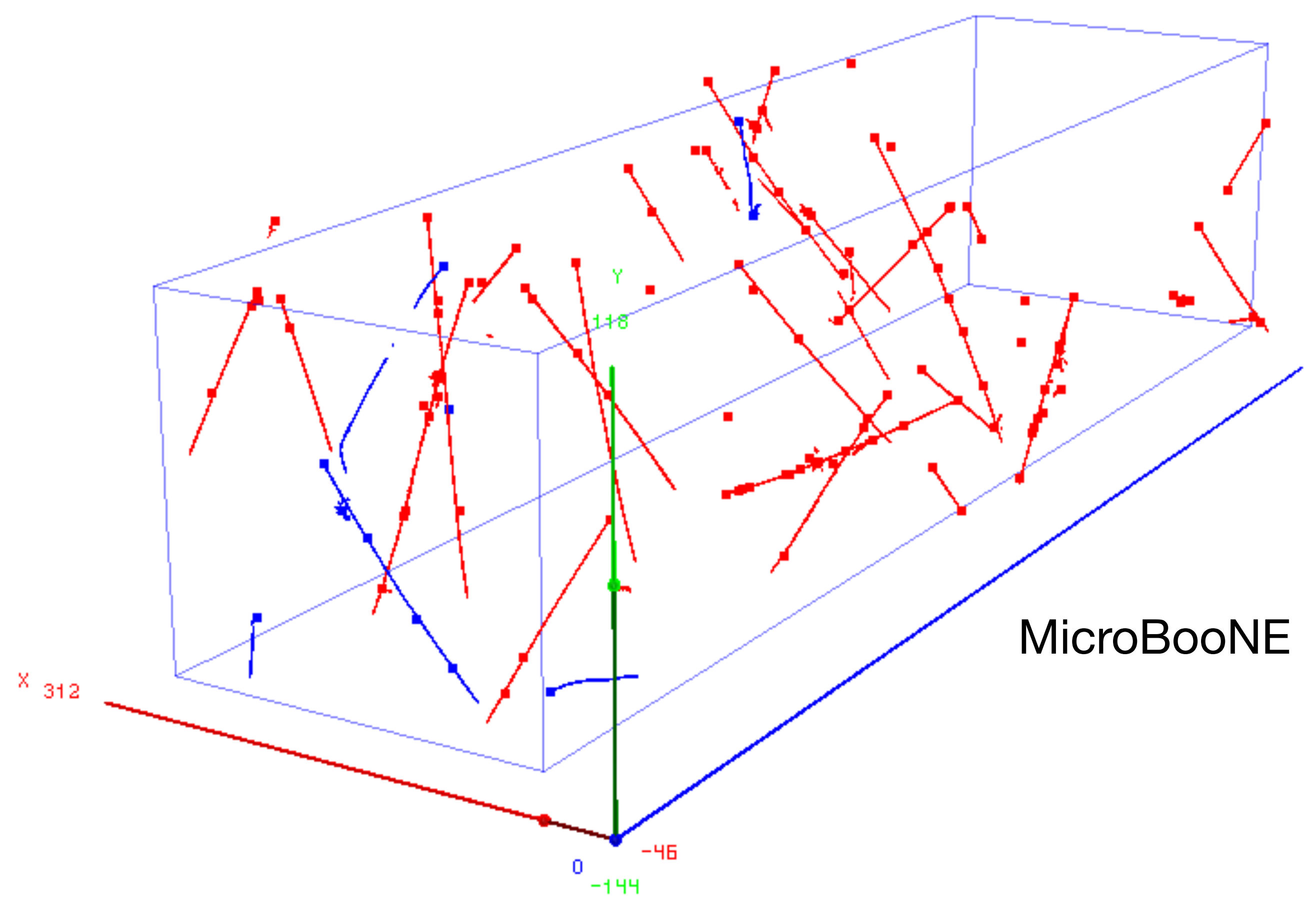


Final 3D Output

PandoraCosmic → PandoraTestBeam



ProtoDUNE-SP



MicroBooNE

Clear cosmic-ray muons

Ambiguous particles

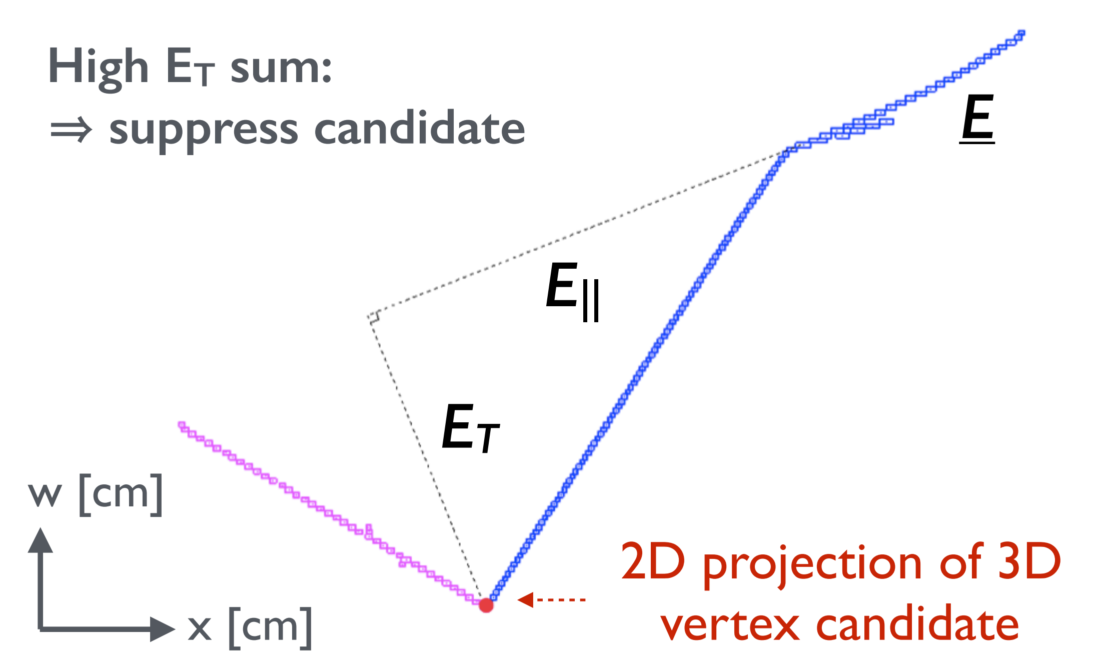
- Hits in ambiguous particles are divided into *slices*.
- Each slice is passed to PandoraTestBeam.

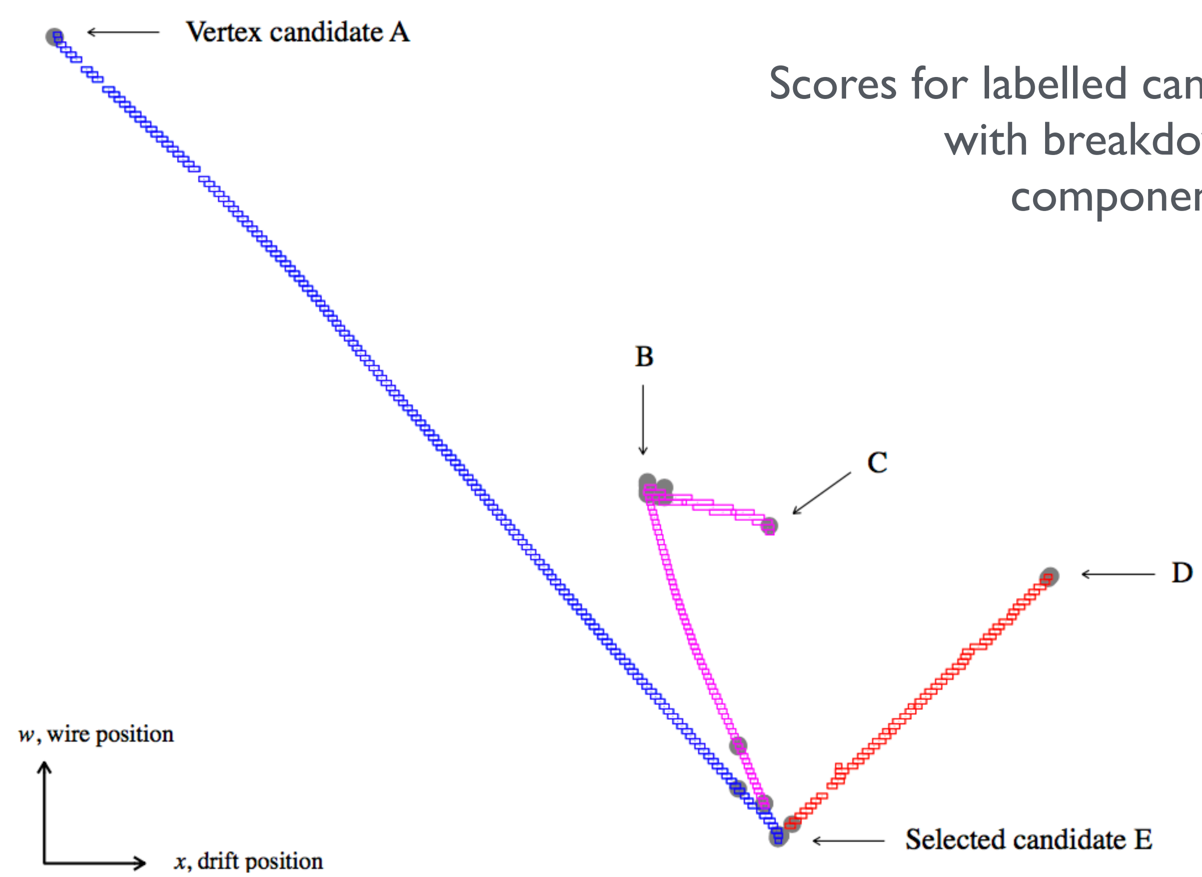
Pandora Test Beam

Vertex Reconstruction - 3D

Search for beam particle interaction vertex:

1. Use pairs of 2D clusters to produce list of possible 3D vertex candidates.
2. Examine candidates, calculate a score for each and select the best.





Scores for labelled candidates, with breakdown into component parts:

Candidate	S	$S_{\text{energy kick}}$	$S_{\text{asymmetry}}$	$S_{\text{beam deweight}}$
A	4.9E-07	3.5E-06	1.00	0.14
B	1.3E-02	3.1E-02	0.99	0.42
C	1.1E-03	2.4E-03	0.95	0.46
D	5.7E-10	1.1E-09	1.00	0.52
E	9.0E-01	9.0E-01	1.00	0.99

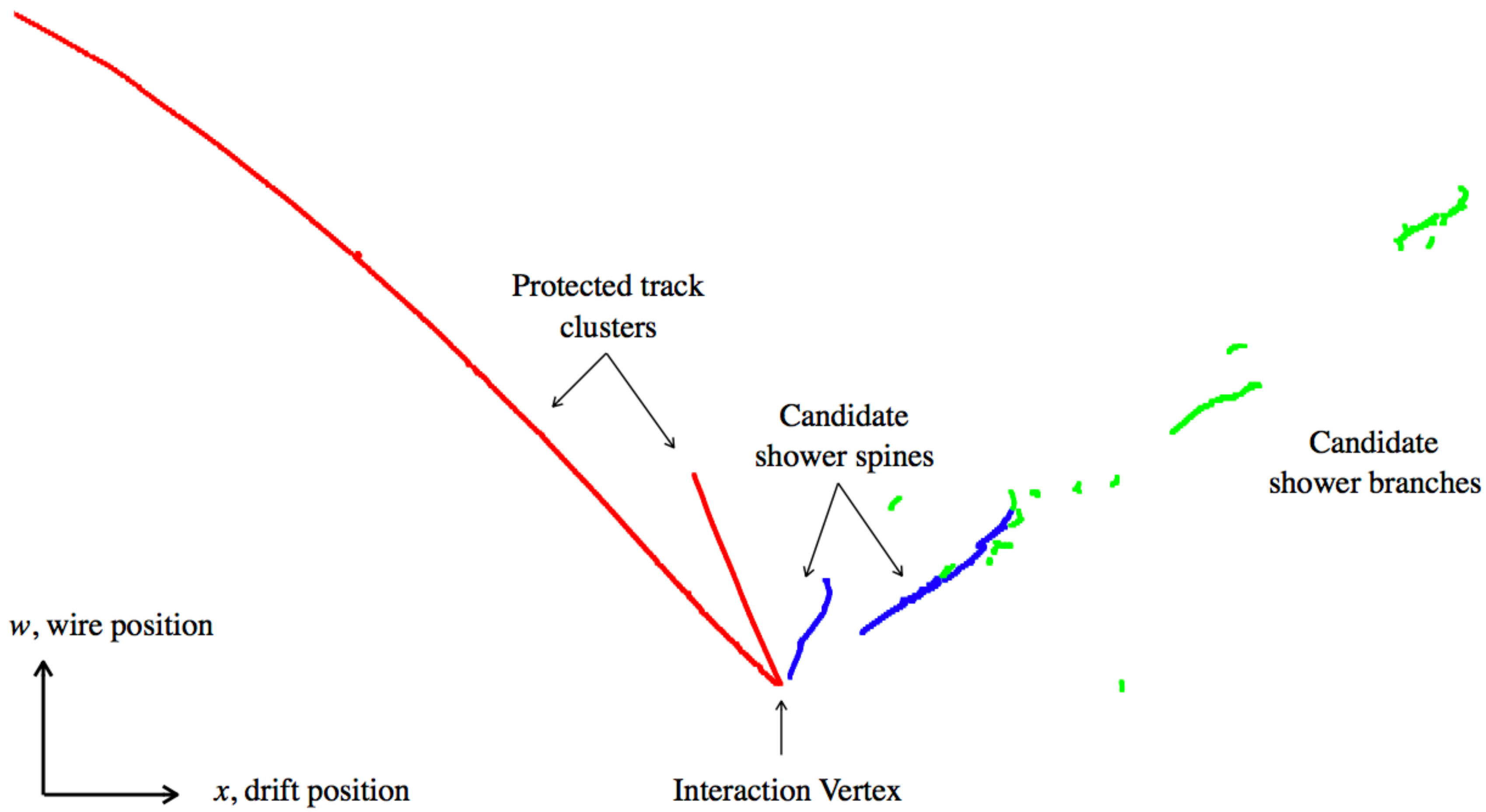
Downstream usage:

- Split 2D clusters at projected vertex position.
- Use vertex to protect primary particles when growing showers.

Shower Reconstruction - 2D

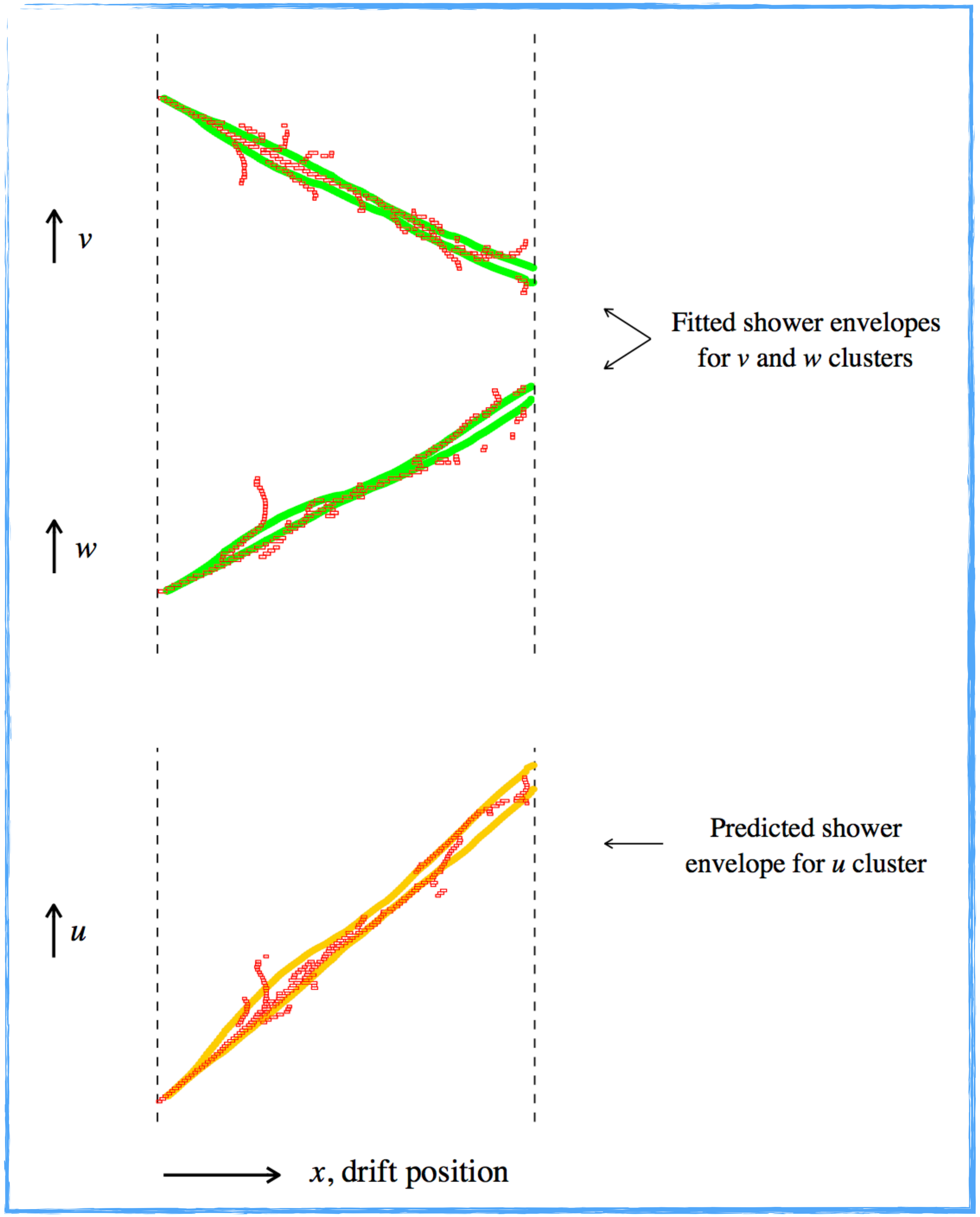
Track reconstruction exactly as in PandoraCosmic, but now also attempt to reconstruct primary electromagnetic showers, from electrons and photons:

- Characterise 2D clusters as track-like or shower-like, and use topological properties to identify clusters that might represent shower spines.
- Add shower-like branch clusters to shower-like spine clusters. Recursively identify branches on the top-level spine candidate, then branches on branches, etc.



Shower Reconstruction - 3D

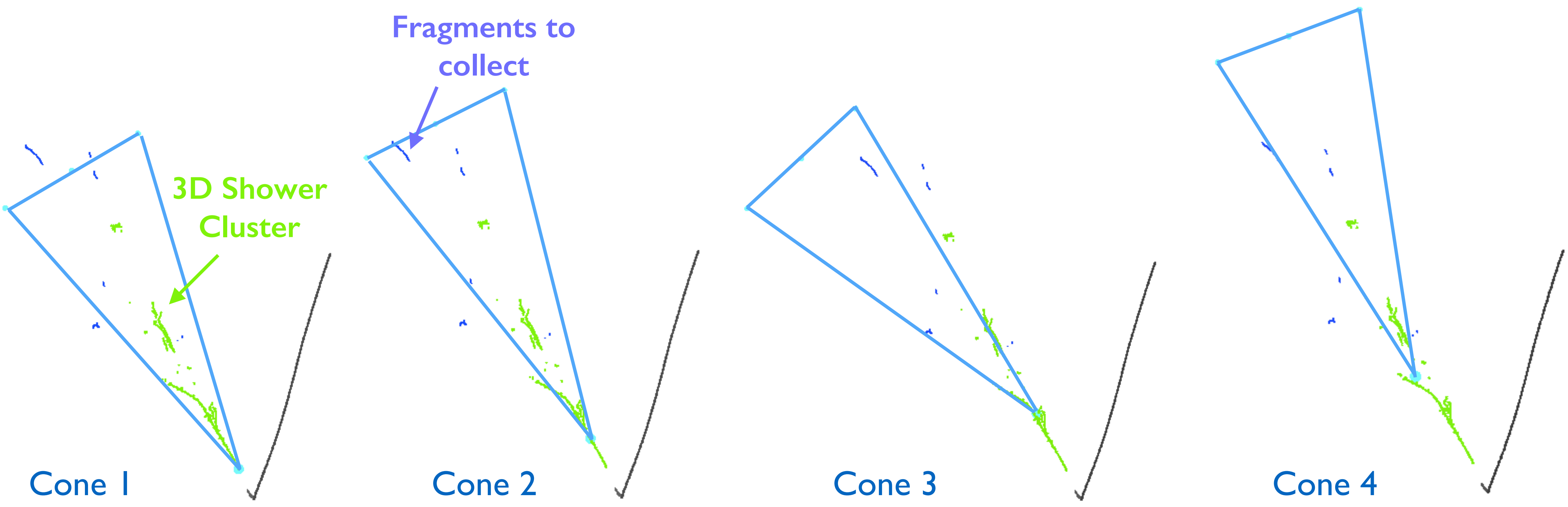
- **Reuse ideas from track reco to match 2D shower clusters between views:**
  - Build a tensor to store cluster overlap and relationship information.
  - Overlap information collected by fitting shower envelope to each 2D cluster.
  - Shower edges from two clusters used to predict envelope for third cluster.



Particle Refinement - 2D, 3D

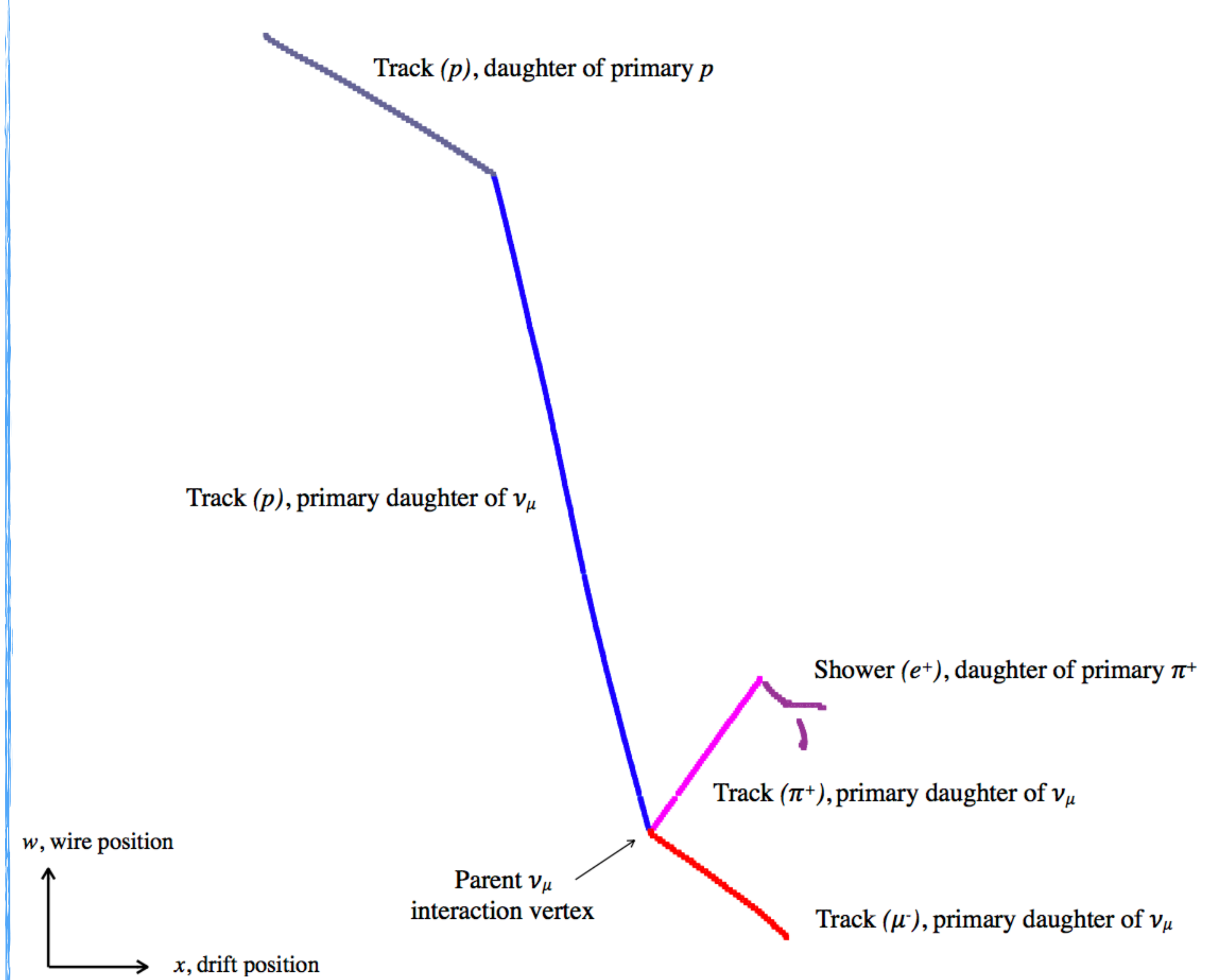
Series of algs deal with remnants to improve particle completeness (esp. sparse showers):

- Pick up small, unassociated clusters bounded by the 2D envelopes of shower-like particles.
- Use sliding linear fits to 3D shower clusters to define cones for merging small downstream shower particles, or picking up additional unassociated clusters.
- If anything left at end, dissolve clusters and assign hits to nearest shower particles in range.

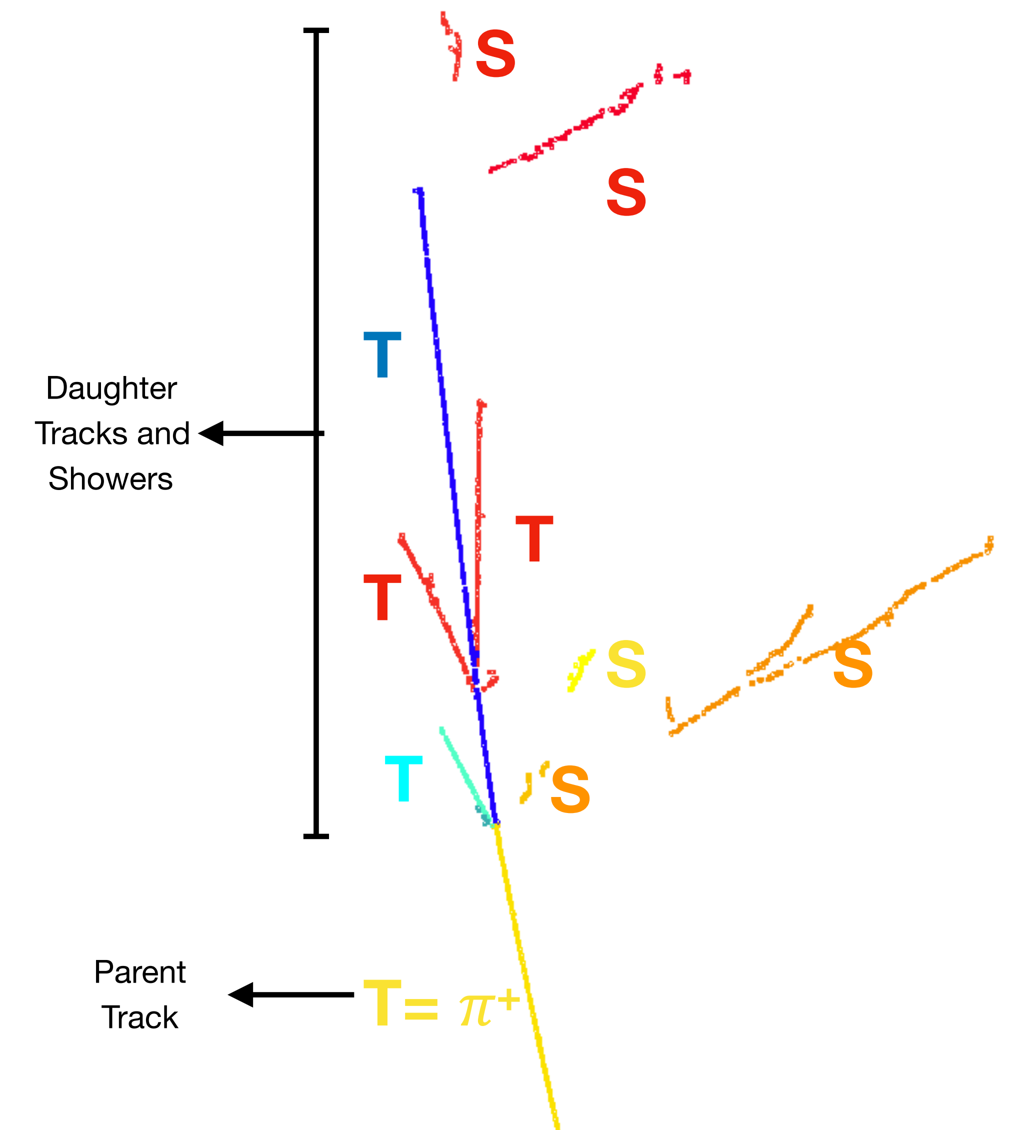


Use 3D clusters to organise particles into a hierarchy, working outwards from interaction vtx:

Simulated ν_μ Pandora Reconstruction at MicroBooNE



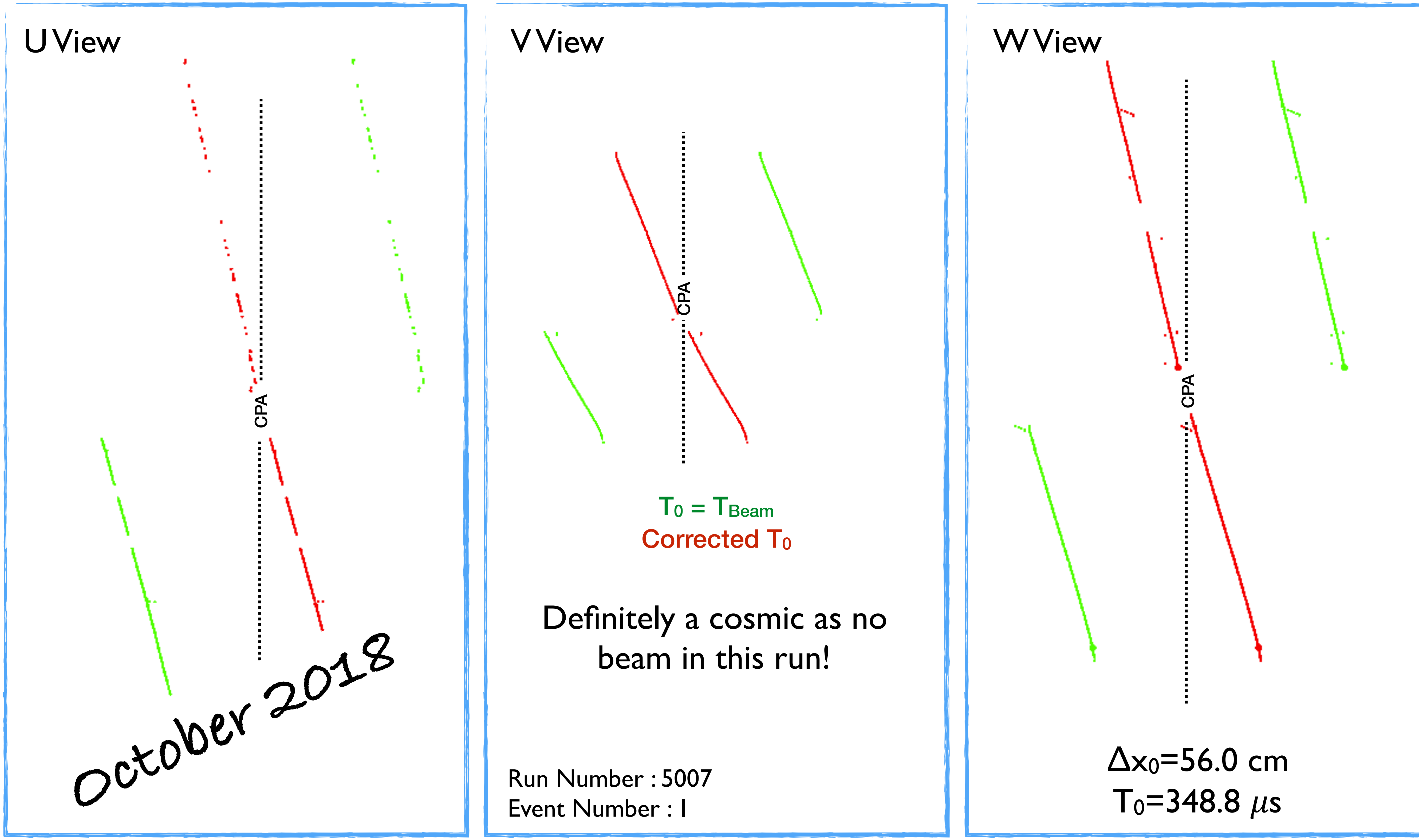
Simulated π^+ Pandora Reconstruction at ProtoDUNE-SP



EPJC (2018) 78:82

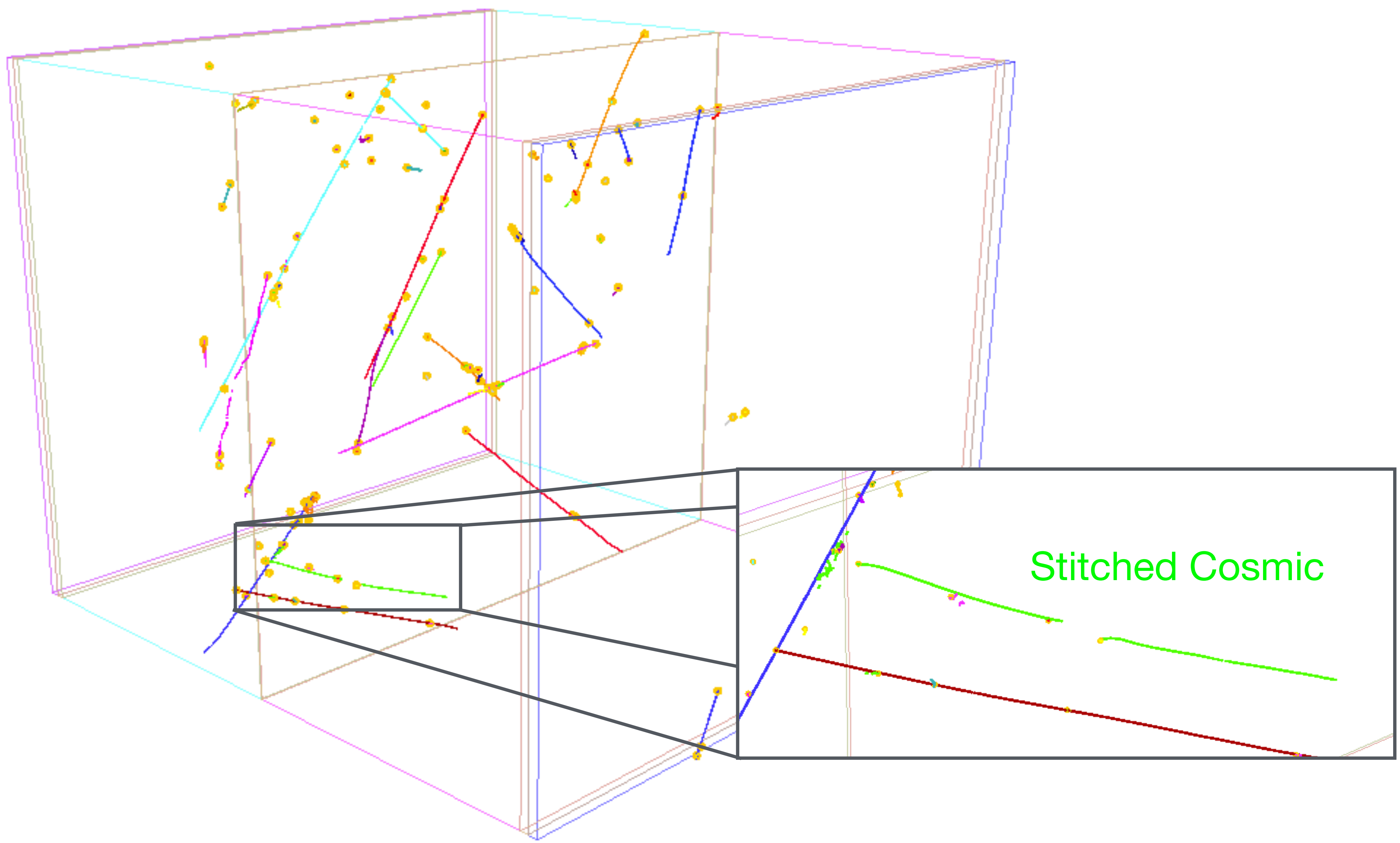
Pandora: ProtoDUNE-SP Cosmic Data

Finish with a few highlights from the application of Pandora to ProtoDUNE-SP data:



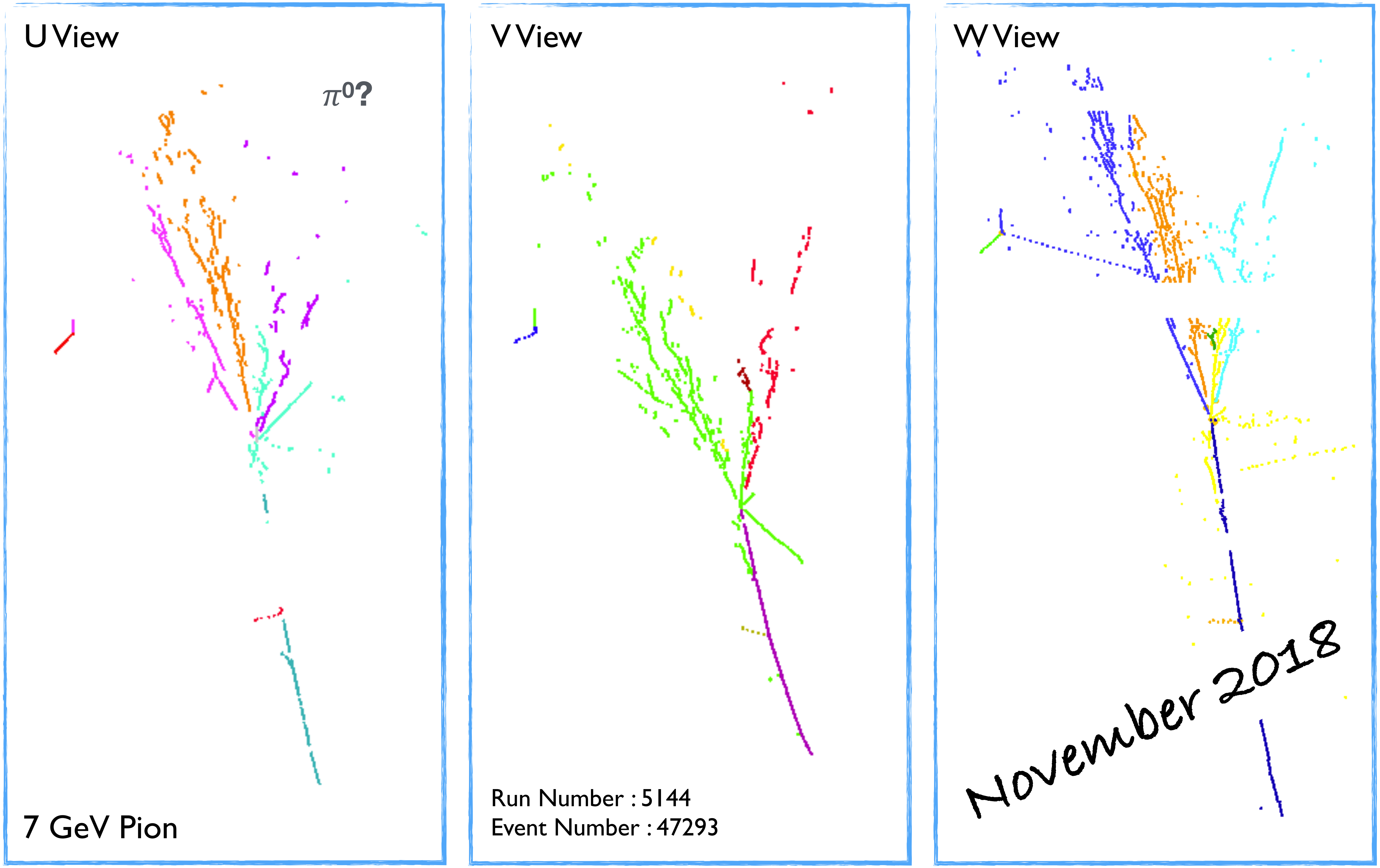
Pandora: ProtoDUNE-SP Cosmic Data

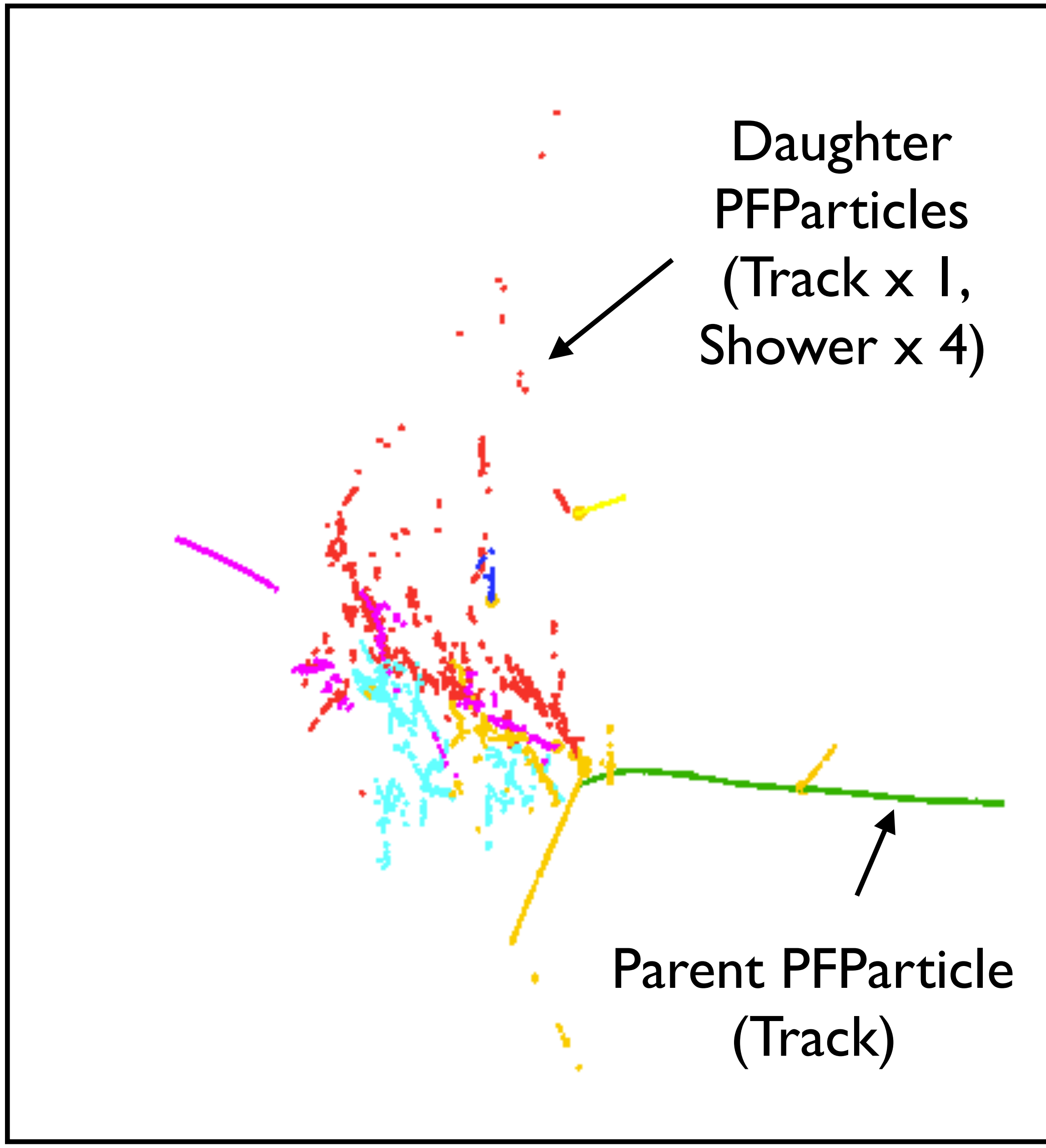
Full 3D Reconstruction



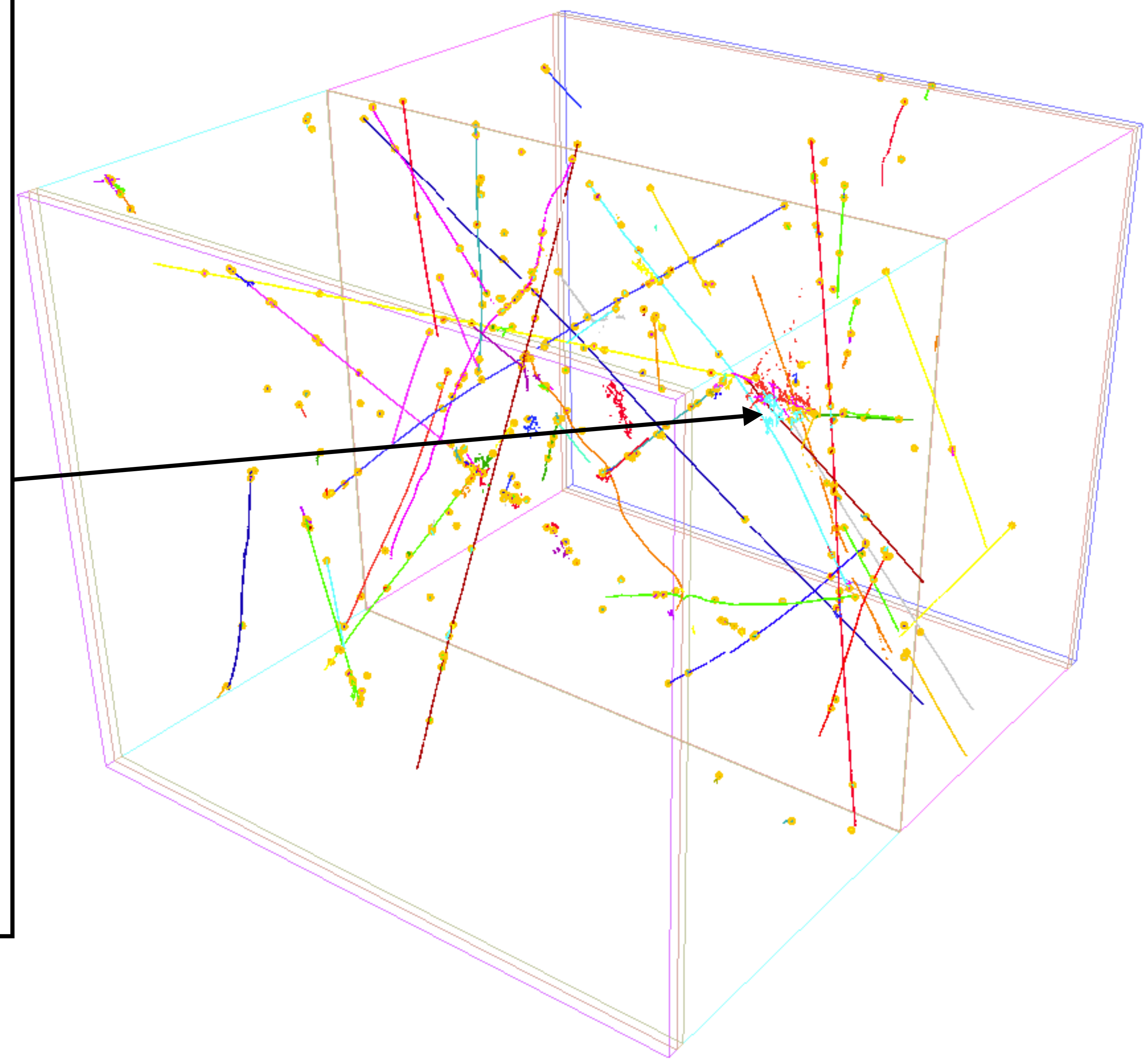
First example of stitching for real LArTPC data!

Pandora: ProtoDUNE-SP Test Beam Data





Full 3D Reconstruction



Pandora ID identifies this as a test beam particle.

Future Development

- **We're increasingly using machine-learning approaches (all with manual feature extraction, so far) to drive pattern-recognition decisions in key algorithms:**
  - Identification of interaction vertices,
  - Track-like vs. shower-like classification during and after pattern recognition,
  - Decision to use PandoraTestBeam or PandoraCosmic outcomes for slices.

- **Promising outlook for combining multi-algorithm and machine-learning approaches, with both aspects increasing in sophistication:**
  - Solve lots of smaller problems using machine learning,
  - Algs write info, for external training, and read training outputs to drive decisions.

- **Pandora is important part of UK DUNE construction proposal: optimistic we will have four full-time PDRAs, each for a six-year period, with two brand-new posts.**

Concluding Comments

- **High-performance reconstruction techniques are required in order to fully exploit the imaging capabilities offered by LArTPCs:**
  - **Pandora multi-algorithm approach uses large numbers of decoupled algorithms to gradually build up a picture of events.**
  - **Algorithm developers and analysers need to work together to ensure fidelity of reconstruction at point of usage, and ensure optimal use of our LArTPC images.**

- **Lorena:** Pandora tutorial today
- **Steve:** Pandora ProtoDUNE developments, Wednesday

Thanks for your attention!

Pandora LAr TPC Reconstruction

Pandora is an open project and new contributors would be extremely welcome.
We'd love to hear from you and we will always try to answer your questions.

Pandora SDK development	John Marshall (John.Marshall@warwick.ac.uk) Mark Thomson (thomson@hep.phy.cam.ac.uk)
LAr TPC algorithm development	John Marshall (John.Marshall@warwick.ac.uk) Andy Blake (a.blake@lancaster.ac.uk)
MicroBooNE integration	Andy Smith (asmith@hep.phy.cam.ac.uk)
ProtoDUNE integration	Steven Green (sg568@hep.phy.cam.ac.uk)
DUNE FD integration	Lorena Escudero (escudero@hep.phy.cam.ac.uk)
Graduate students	MicroBooNE: Joris Jan de Vries, Jack Anthony, Andy Smith ProtoDUNE: Stefano Vergani DUNE: Jhanzeb Ahmed, Mousam Rai, Ryan Cross

WARWICK THE UNIVERSITY OF WARWICK

UNIVERSITY OF CAMBRIDGE

Lancaster University

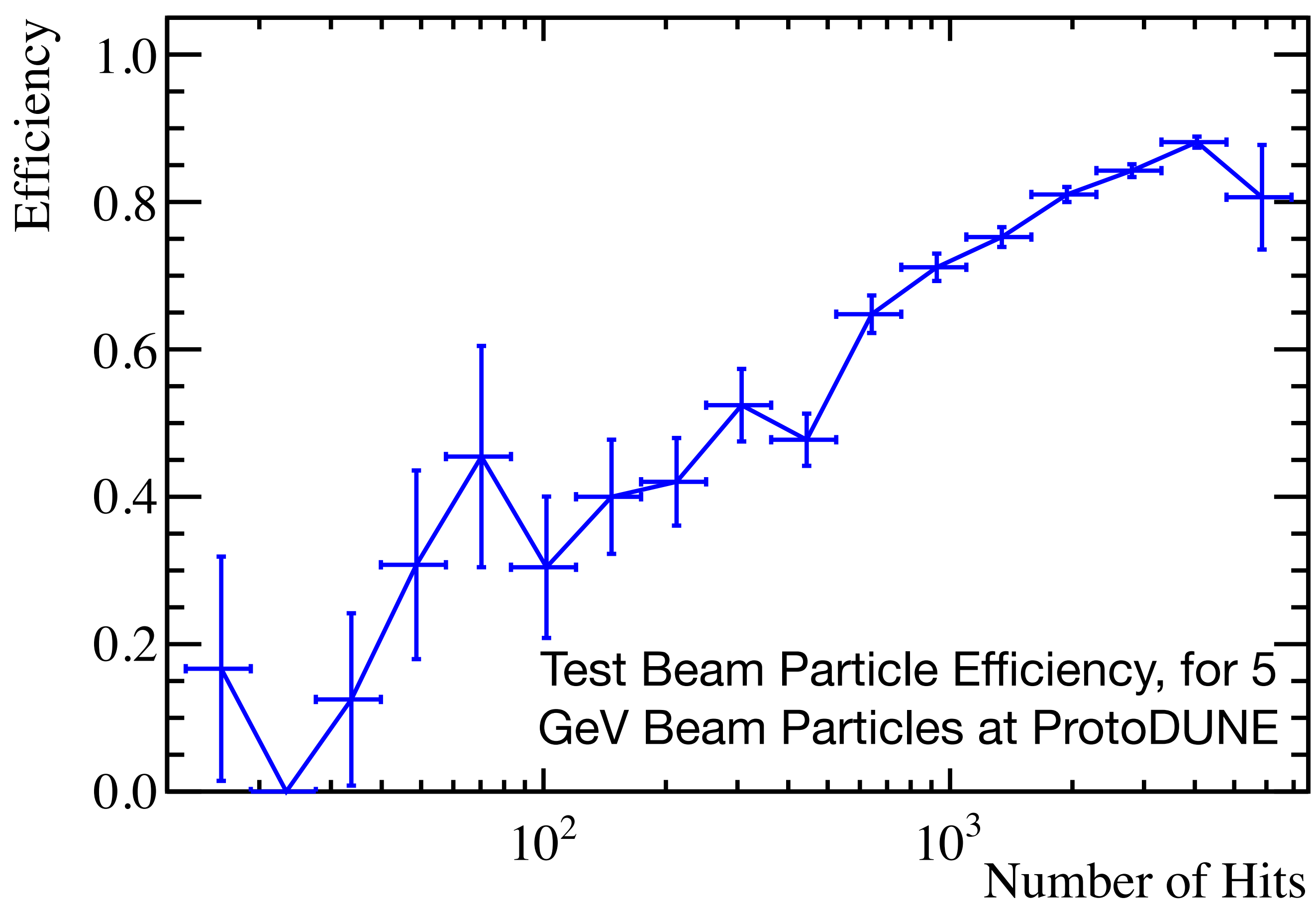
https://github.com/PandoraPFA

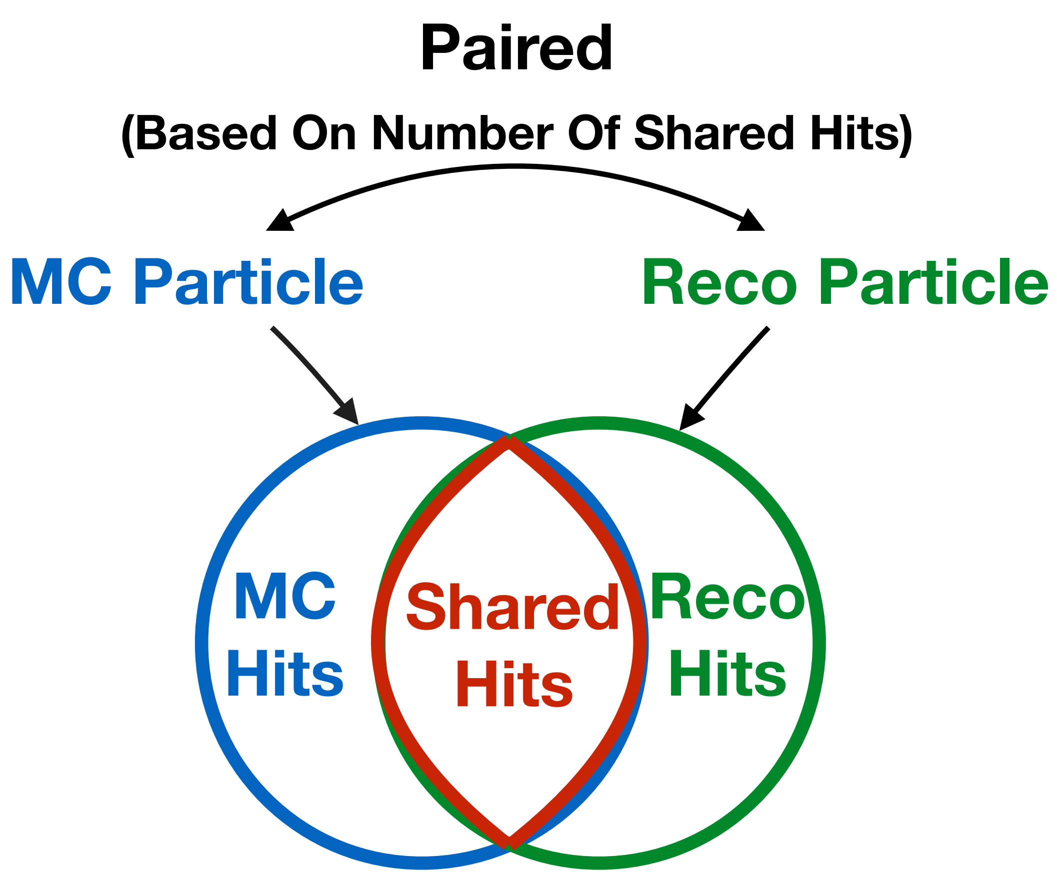
https://pandorapfa.slack.com

ProtoDUNE Performance Metrics

- There are many ways of assessing pattern recognition performance. In Pandora, for ProtoDUNE, we primarily use the Efficiency:

"Fraction of target MCParticles with at least one matched reconstructed particle."





*Purity = nSharedHits / nRecoHits > 50%

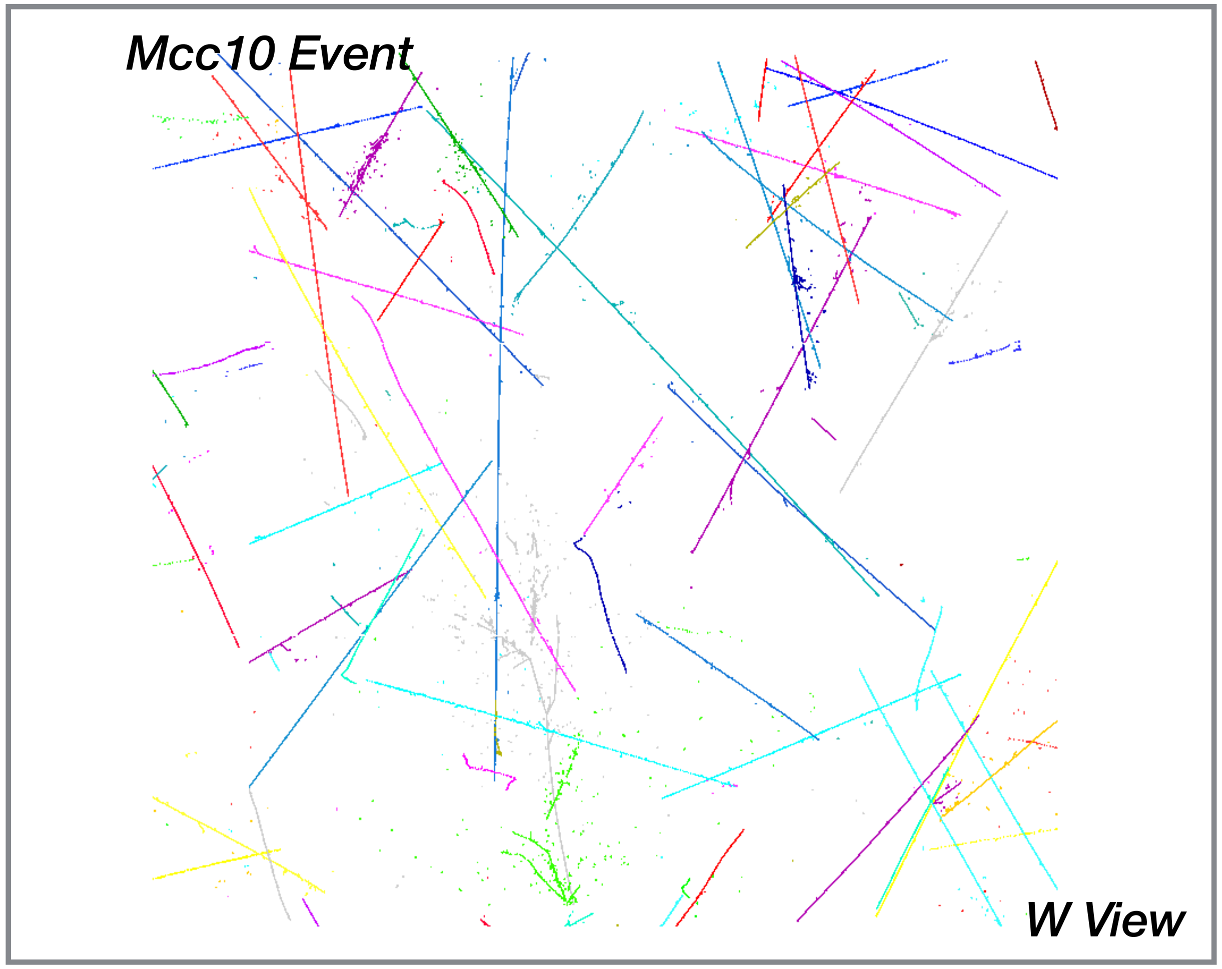
*Completeness = nSharedHits / nMCHits > 10%

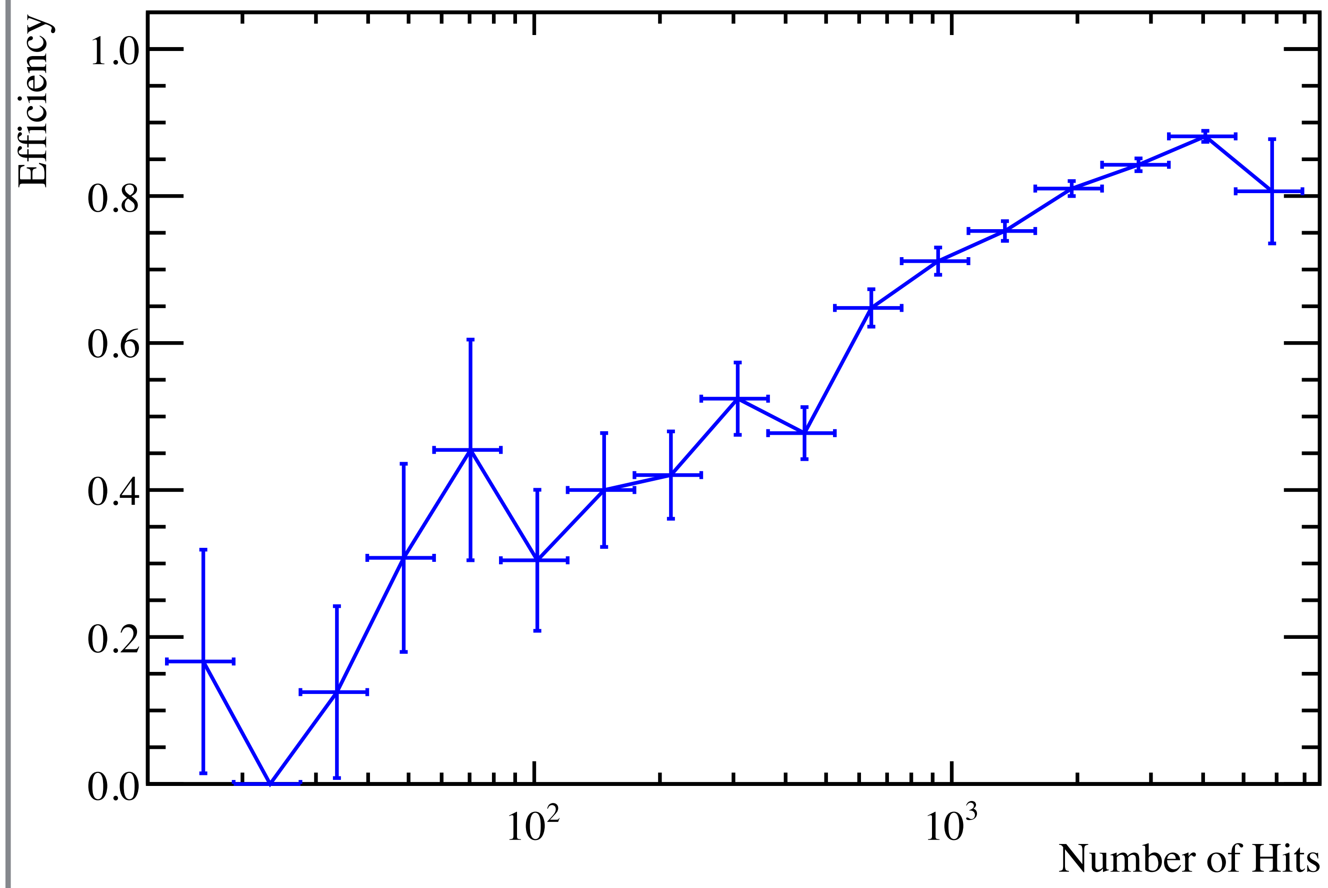
- This metric folds in effects** from cosmic-ray pattern recognition, cosmic-ray tagging, slice creation, both the cosmic-ray and neutrino slice reconstructions and test beam particle identification.

Reconstructed particles have to be correctly tagged to count towards the efficiency!

ProtoDUNE Performance Metrics

- The Pandora test beam reconstruction is good and ready for real data.
- Inefficiency are primarily due to:



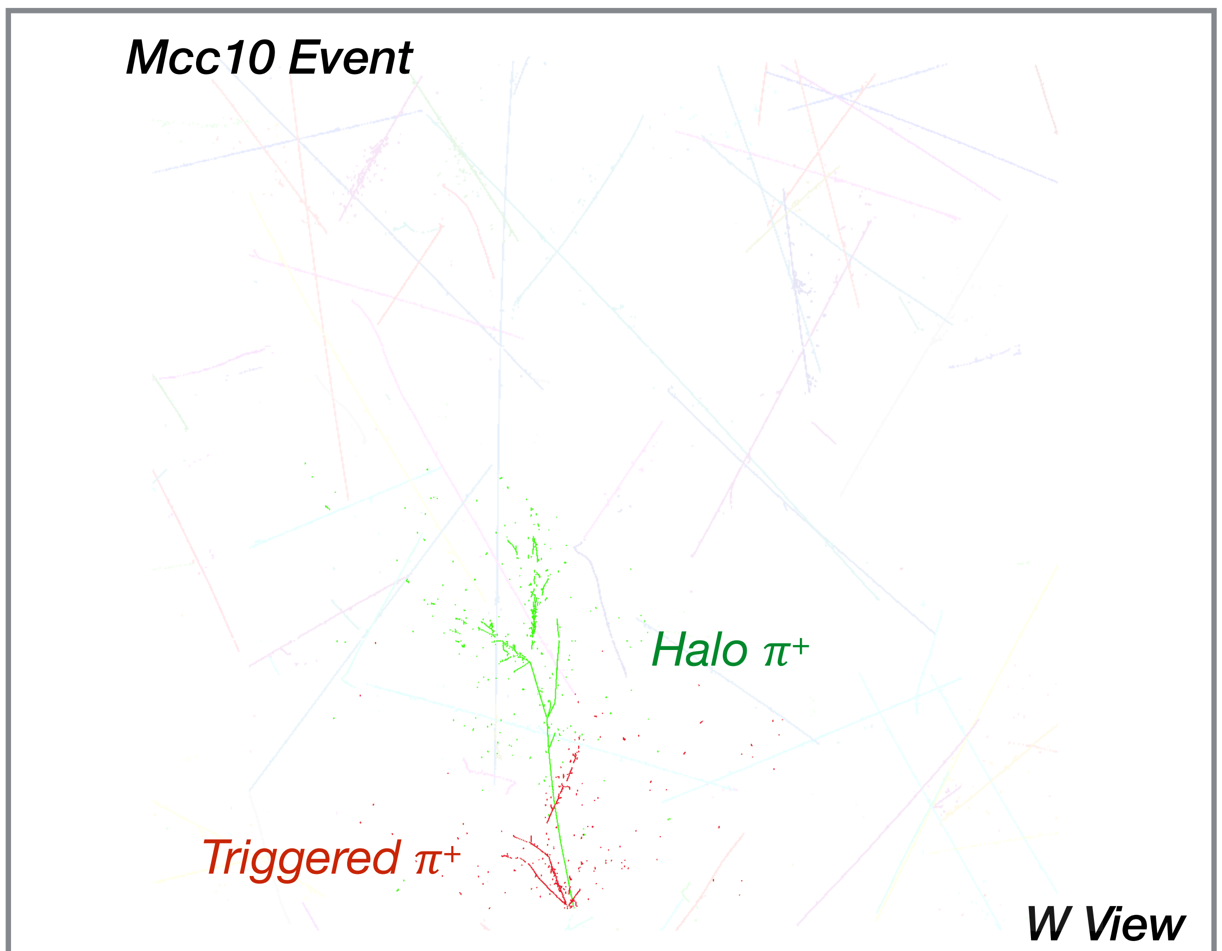


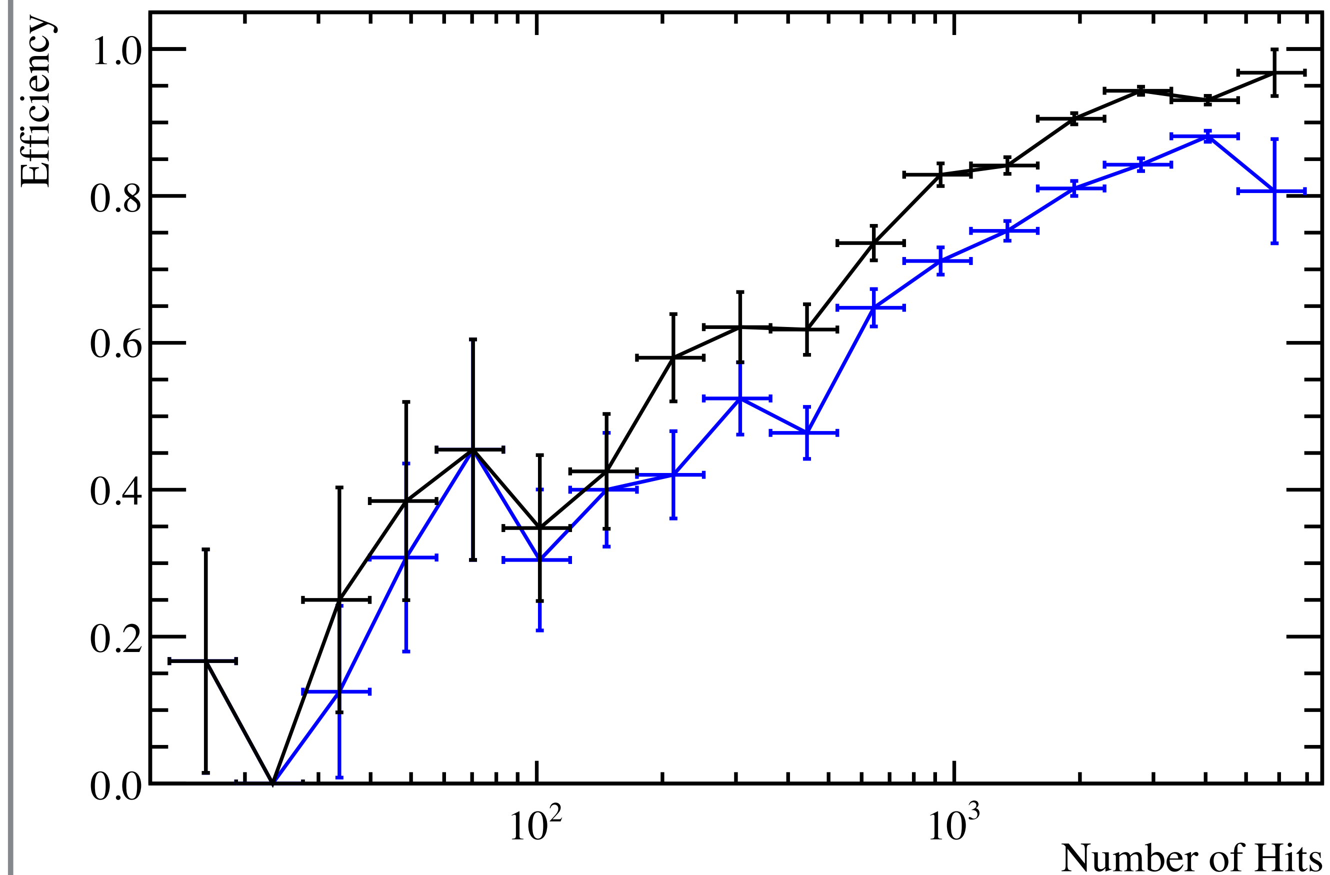
	Integrated Efficiency [%]
Beam & Cosmics	78.88±0.47

This metric folds in effects from cosmic-ray pattern recognition, cosmic-ray tagging, slice creation, both the cosmic-ray and neutrino slice reconstructions and test beam particle identification.

ProtoDUNE Performance Metrics

- The Pandora test beam reconstruction is good and ready for real data.
- Inefficiency are primarily due to:
    - Cosmic overlay

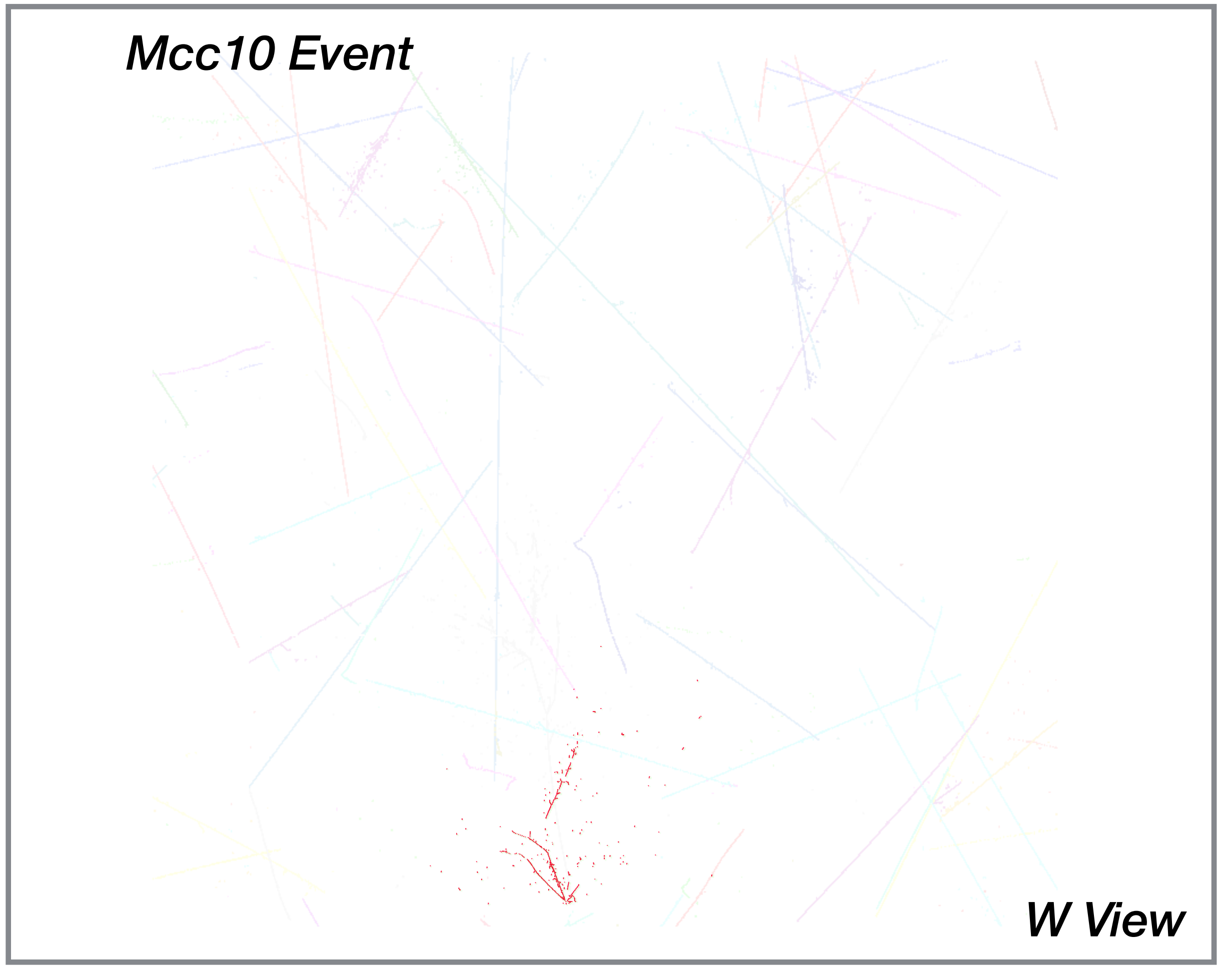


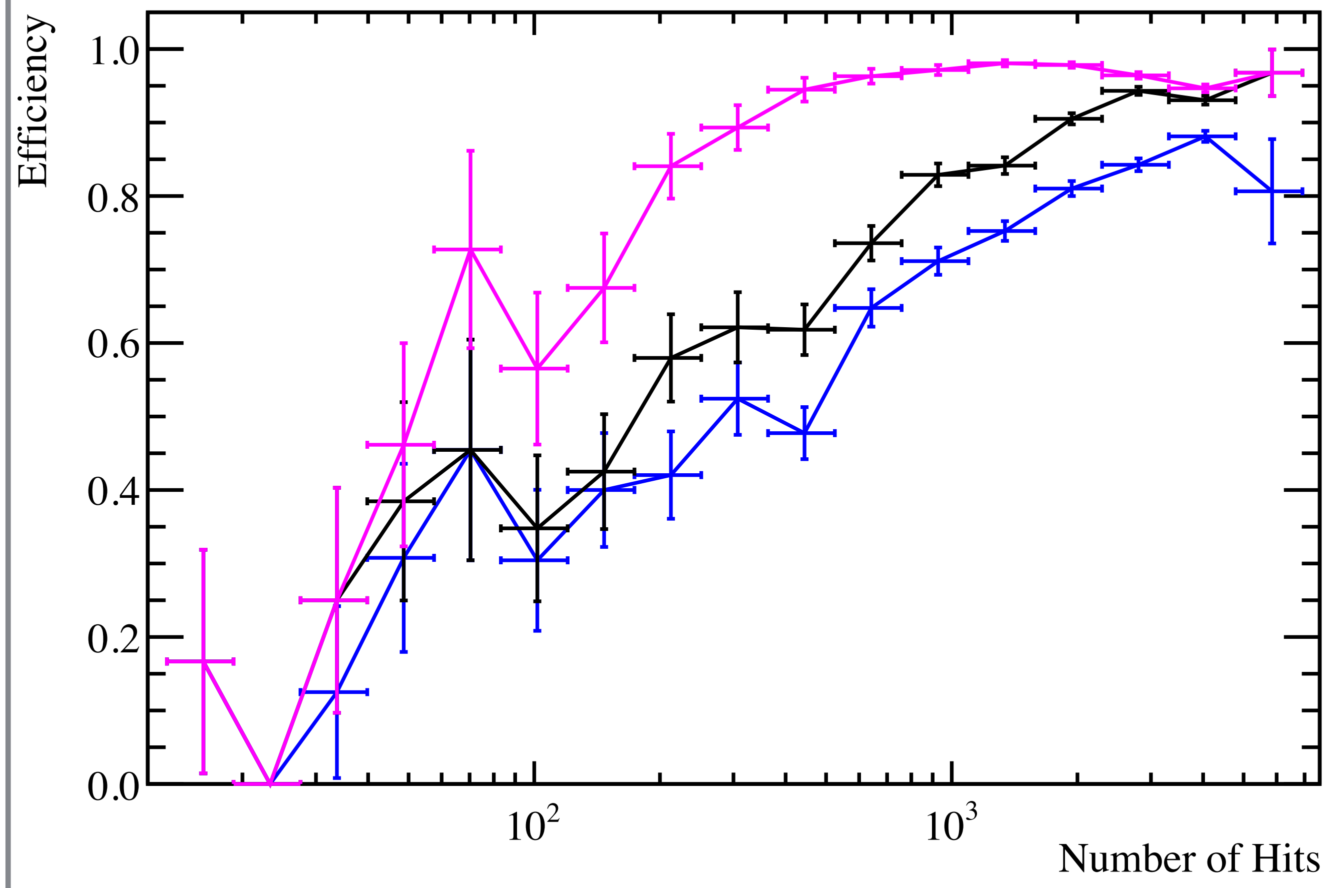


	Integrated Efficiency [%]
Beam & Cosmics	78.88±0.47
Beam (Triggered + Halo)	87.58±0.38

ProtoDUNE Performance Metrics

- The Pandora test beam reconstruction is good and ready for real data.
- Inefficiency are primarily due to:
  - Cosmic overlay
  - The Beam Halo.

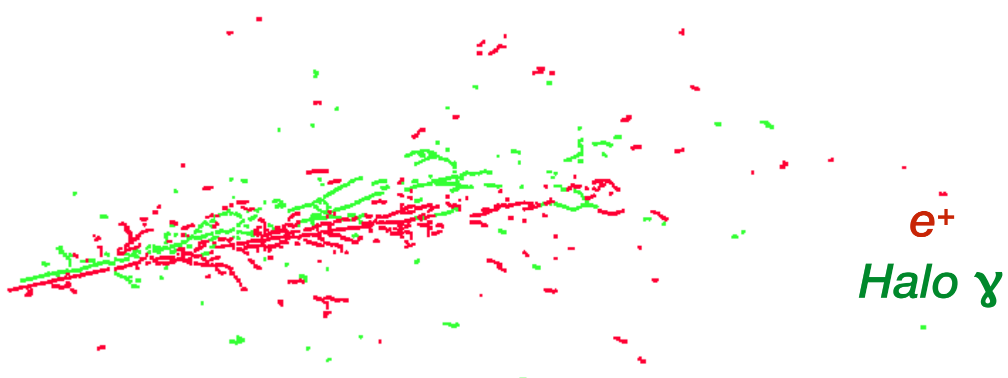




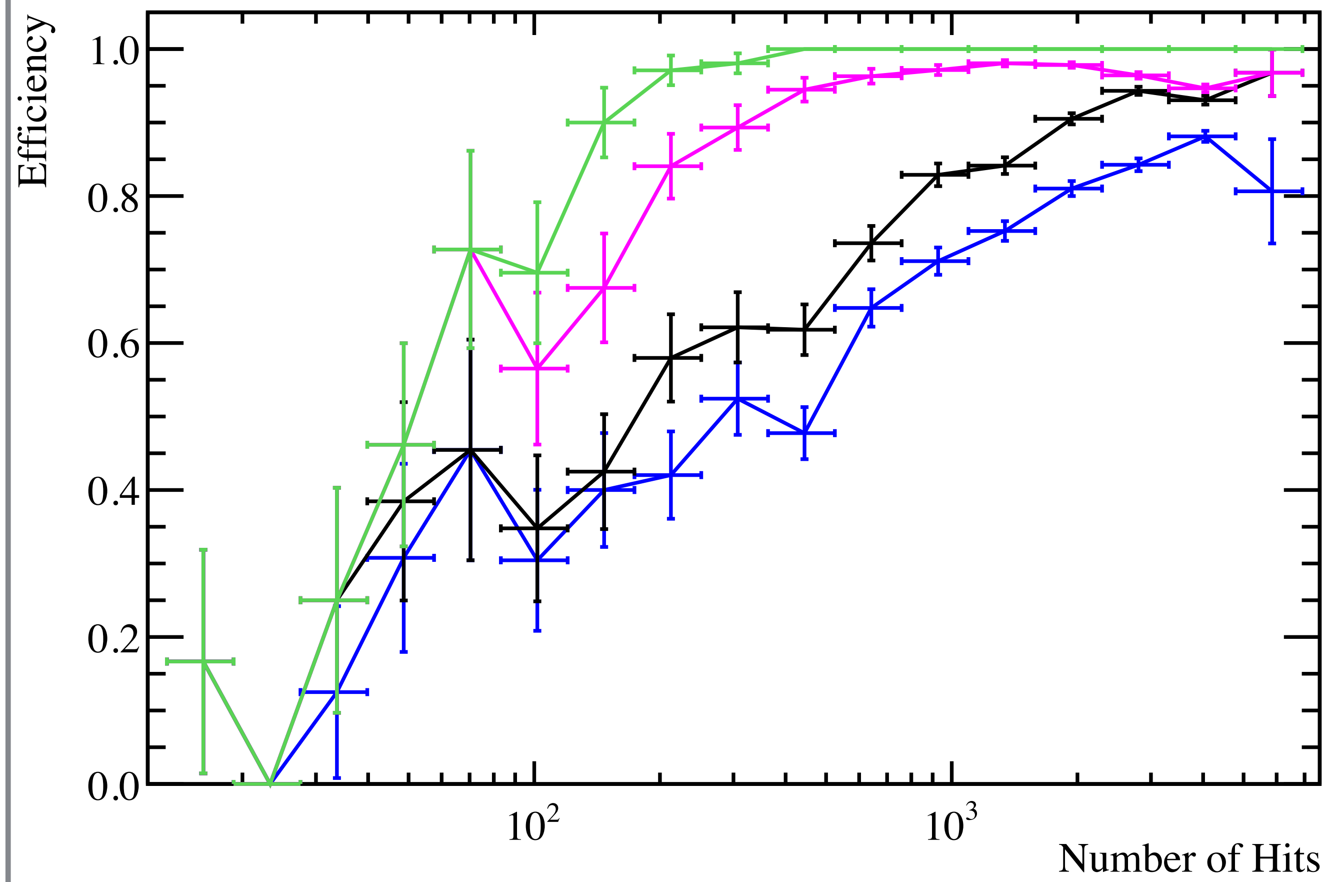
	Integrated Efficiency [%]
Beam & Cosmics	78.88±0.47
Beam (Triggered + Halo)	87.58±0.38
Beam (Triggered)	95.68±0.23

ProtoDUNE Performance Metrics

- There is a small inefficiency from Pandora slicing and test beam particle ID too.
- The e± reconstruction efficiencies suffer from the effect of the halo due to missing MC links to bremsstrahlung photons.

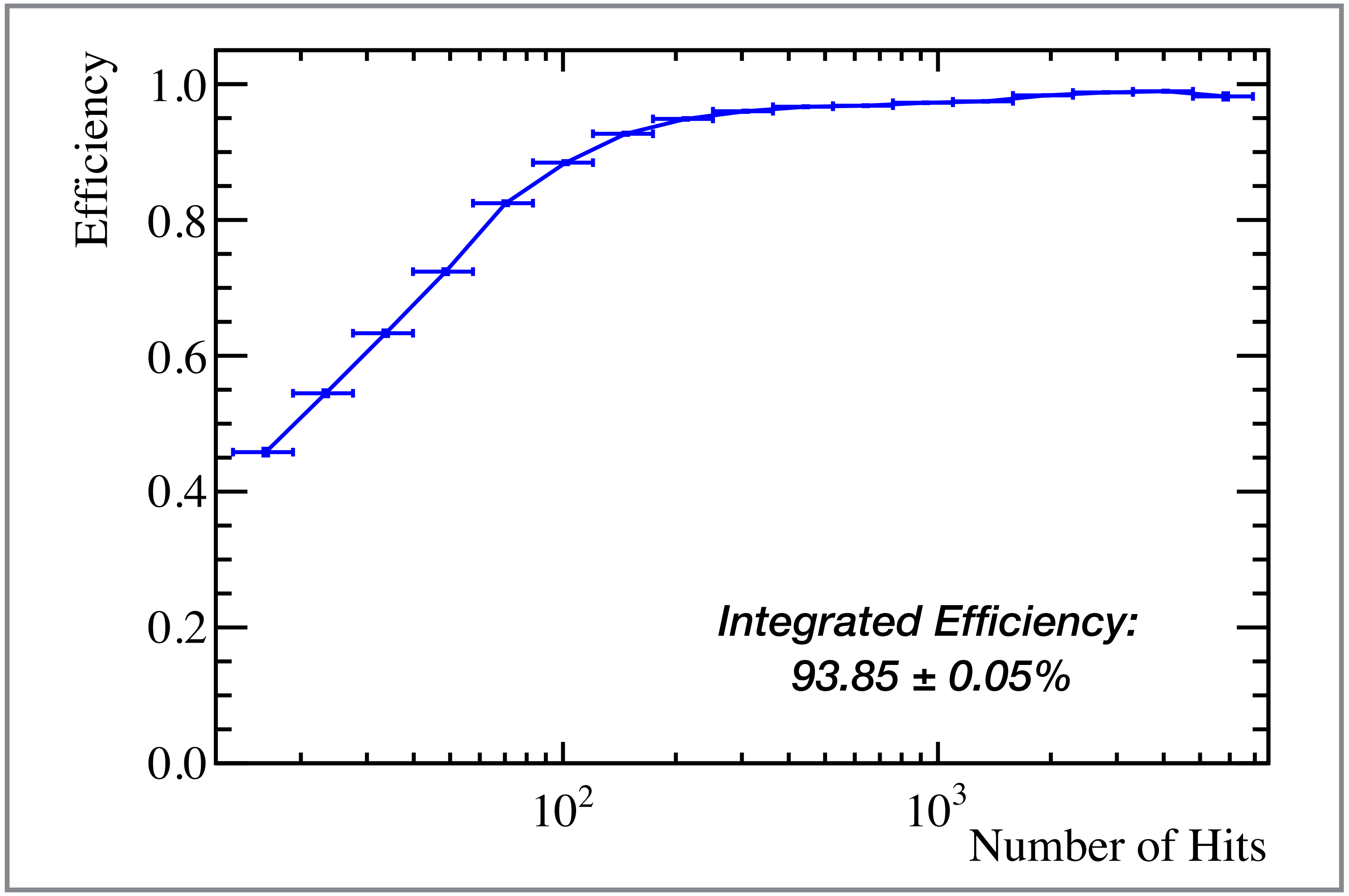


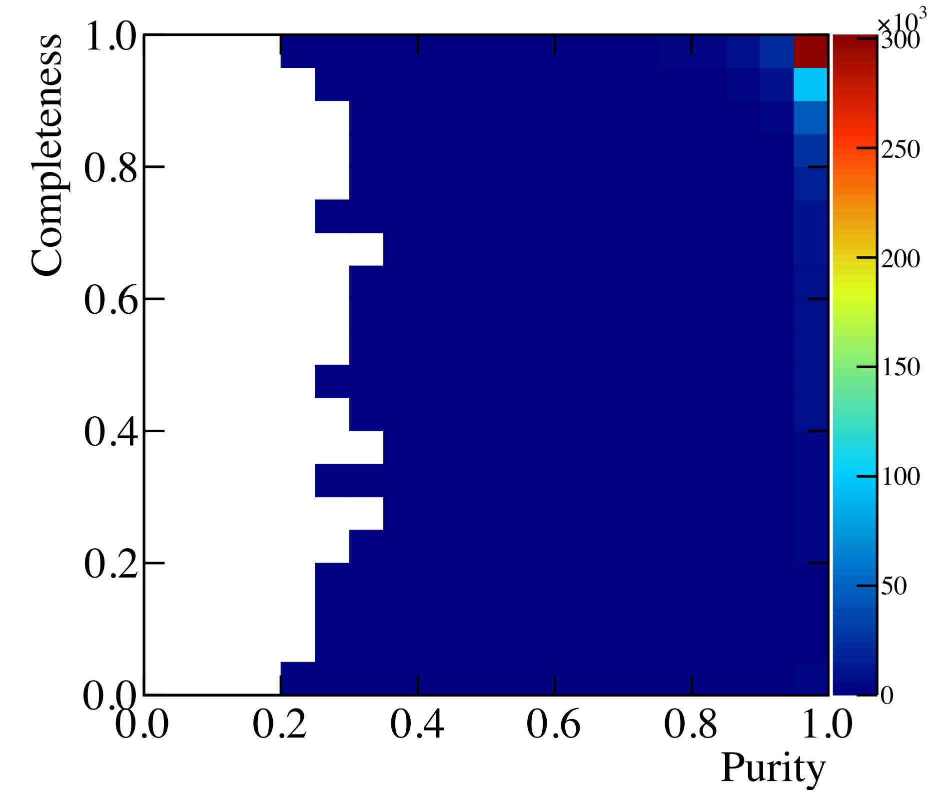
e+ Beam Only	Integrated Efficiency [%]
Beam (Triggered + Halo)	72.7±0.8293
Beam (Triggered)	90.8±0.5905

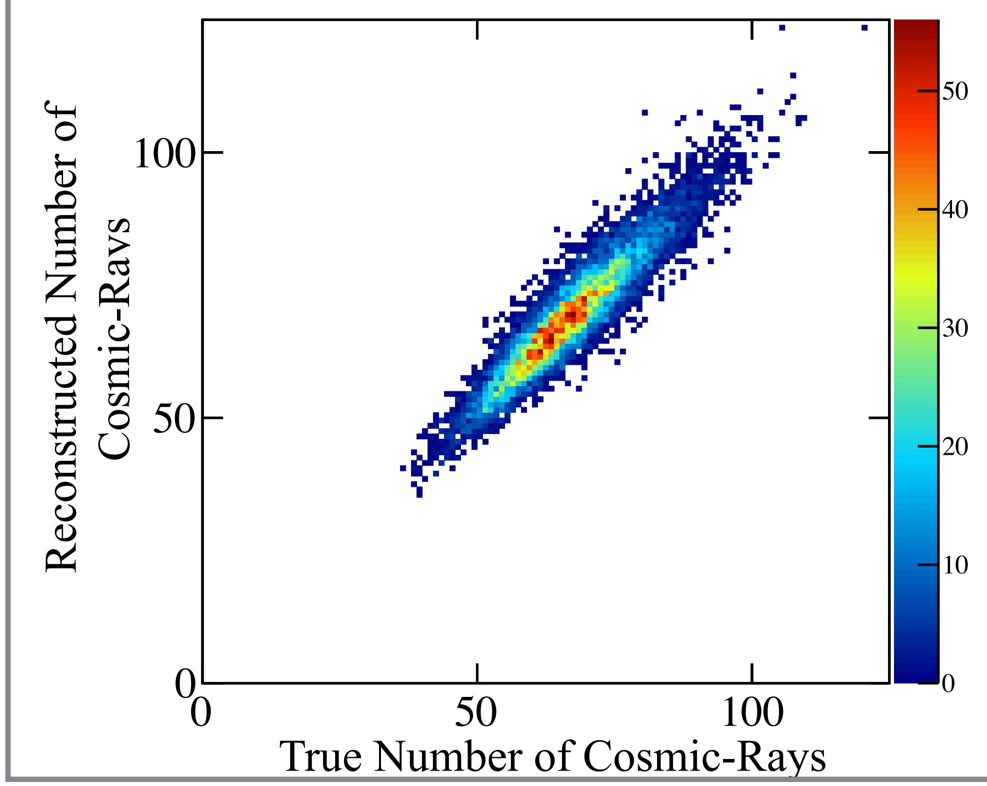


	Integrated Efficiency [%]
Beam & Cosmics	78.9±0.5
Beam (Triggered + Halo)	87.6±0.4
Beam (Triggered)	95.7±0.2
Beam (Test Beam Reco, No Slicing)	99.5±0.1

ProtoDUNE Performance Metrics

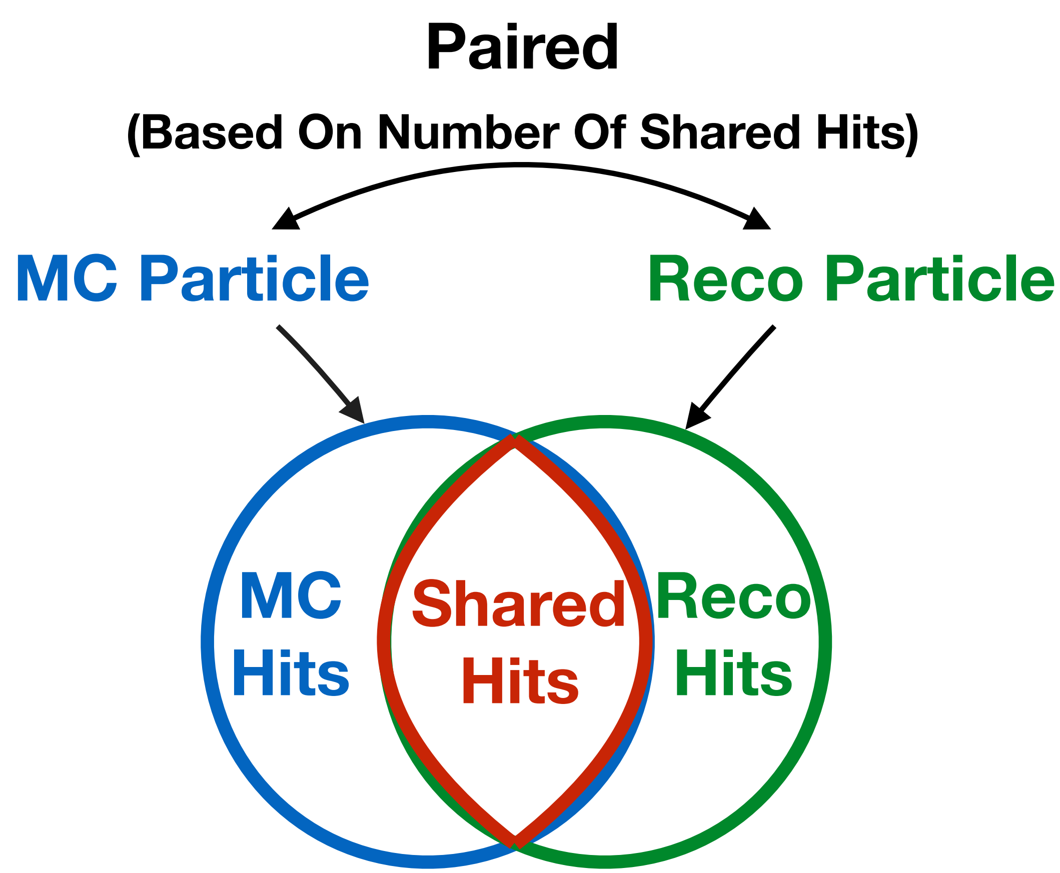






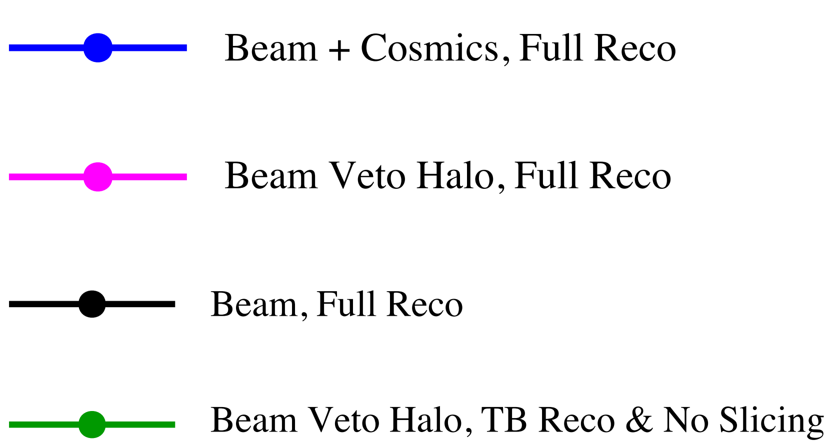
- Pandora reconstruction for cosmic rays is highly efficient.
- The purity and completenesses for cosmic rays, which including all secondaries e.g. delta rays, is good despite large number of cosmic rays per event.

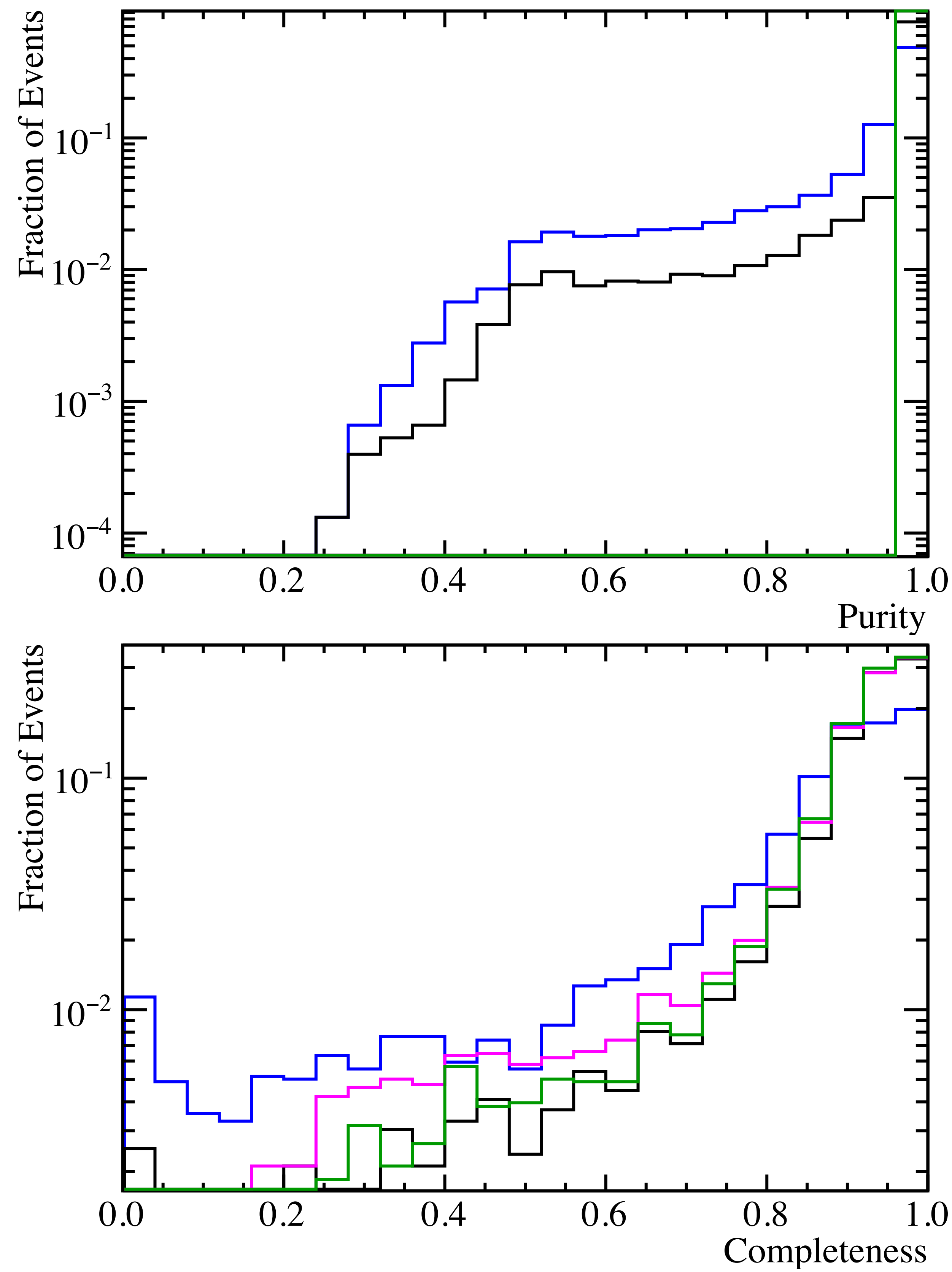
ProtoDUNE Performance Metrics



Purity = nSharedHits / nRecoHits

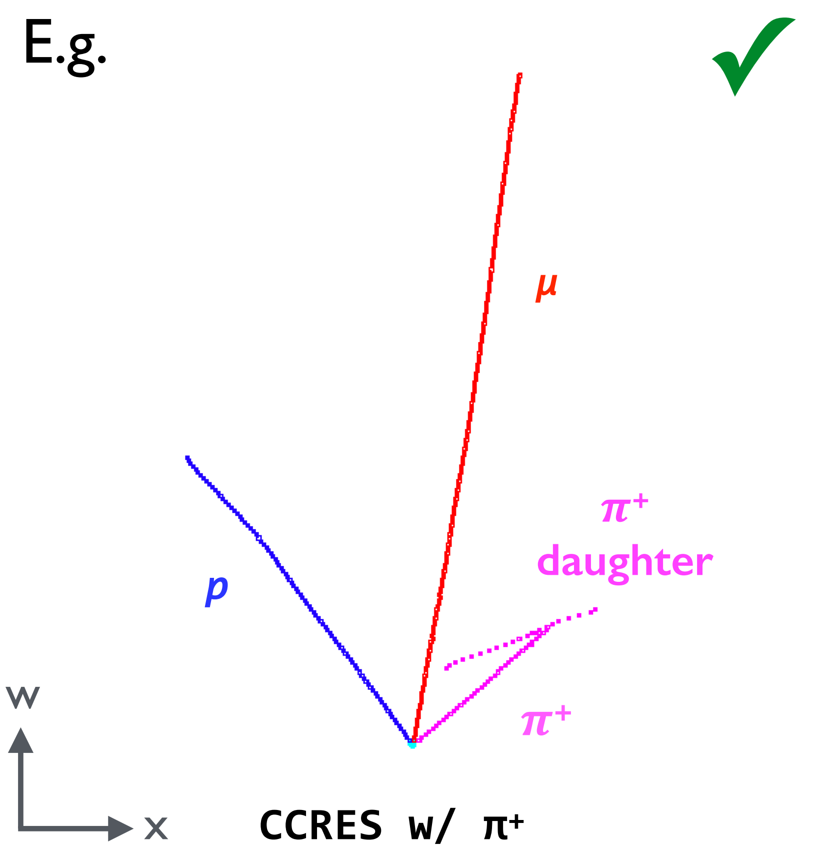
Completeness = nSharedHits / nMCHits

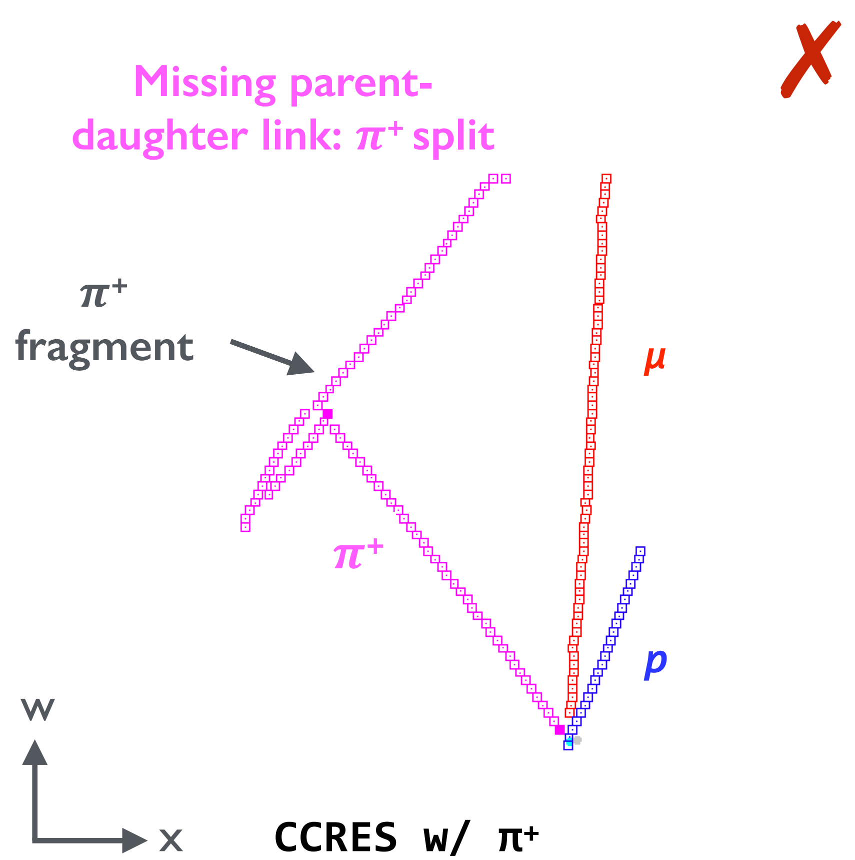


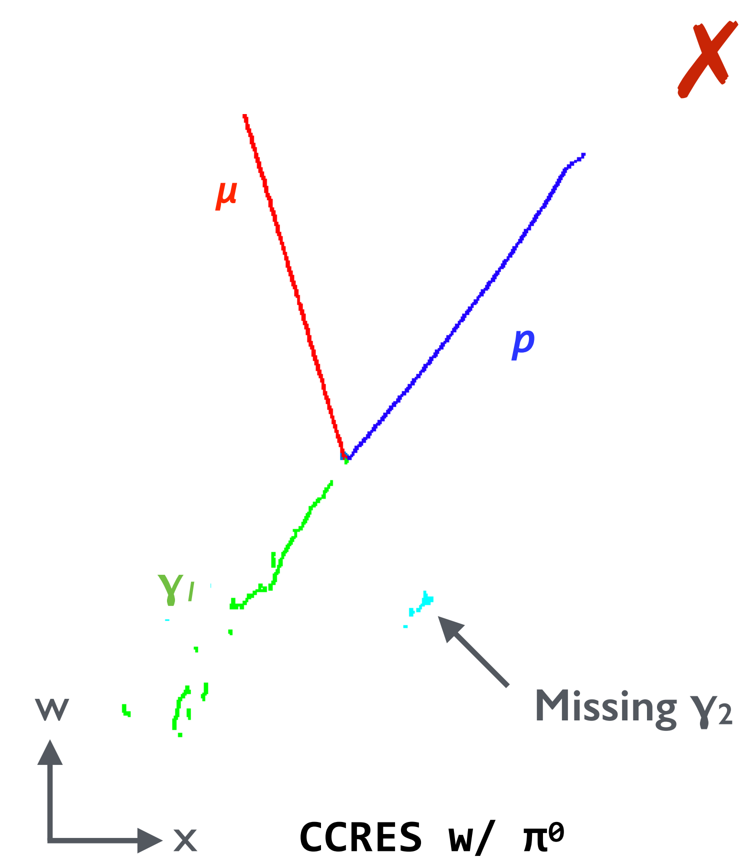


- To assess performance for simulated **MicroBooNE** events, used selection of event topologies.
- Examine fraction of events deemed "correct" by ***very strict*** pattern-recognition metrics:
  - Consider exclusive final-states where all true particles pass simple quality cuts (e.g. nHits)
  - Correct means exactly one reco primary particle is matched to each true primary particle

E.g.







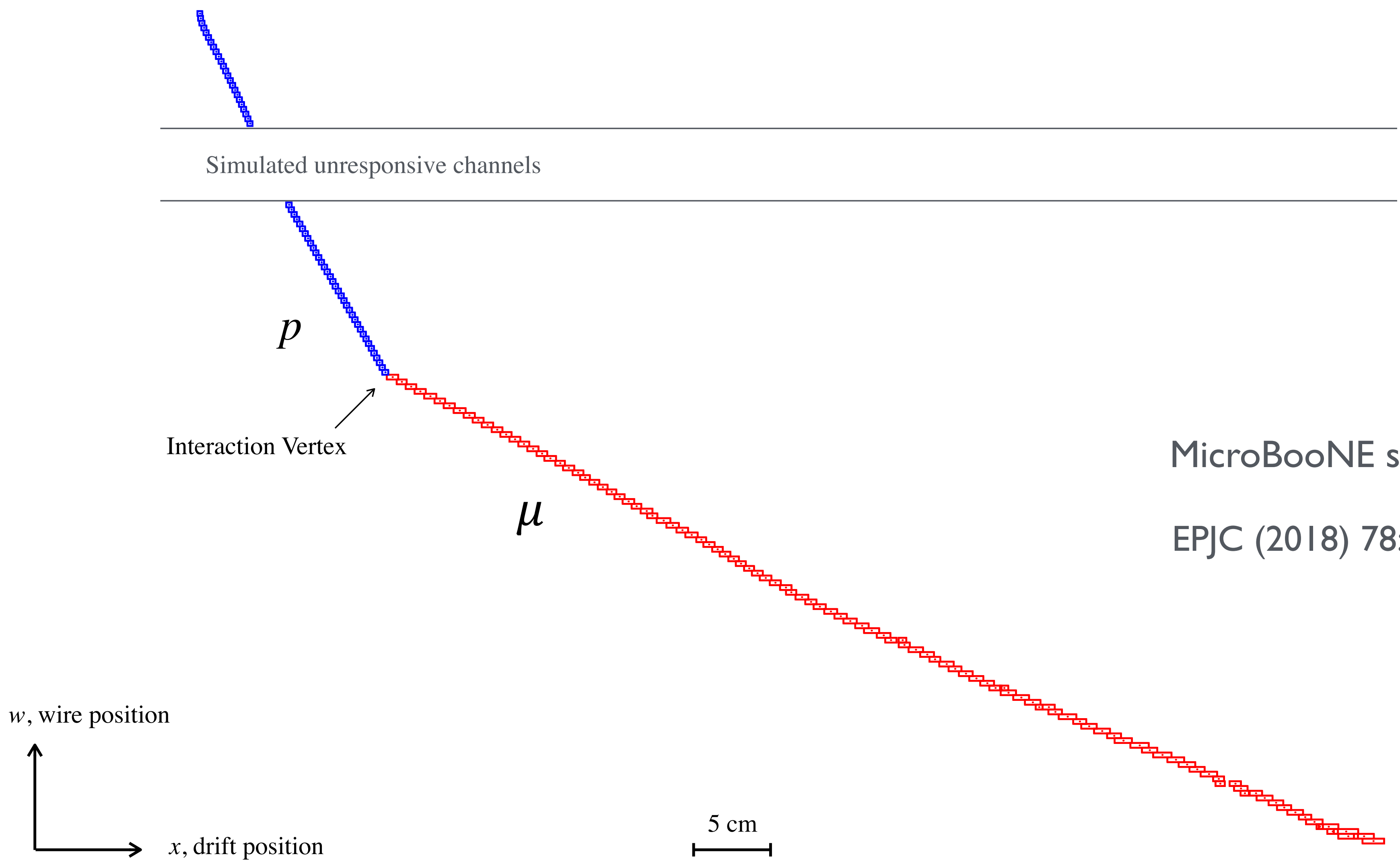
BNB CC QE: $\nu_\mu + \text{Ar} \rightarrow \mu^- + p$

Clean topology: ν_μ CC QE interactions with exactly one reconstructable muon and one reconstructable proton in visible final state:

No cosmic rays here

#Matched Particles	0	1	2	3+
μ	1.3%	95.8%	2.9%	0.1%
p	8.9%	87.3%	3.6%	0.2%

53,168 events, 86.0% have exactly one reco particle matched to each target.

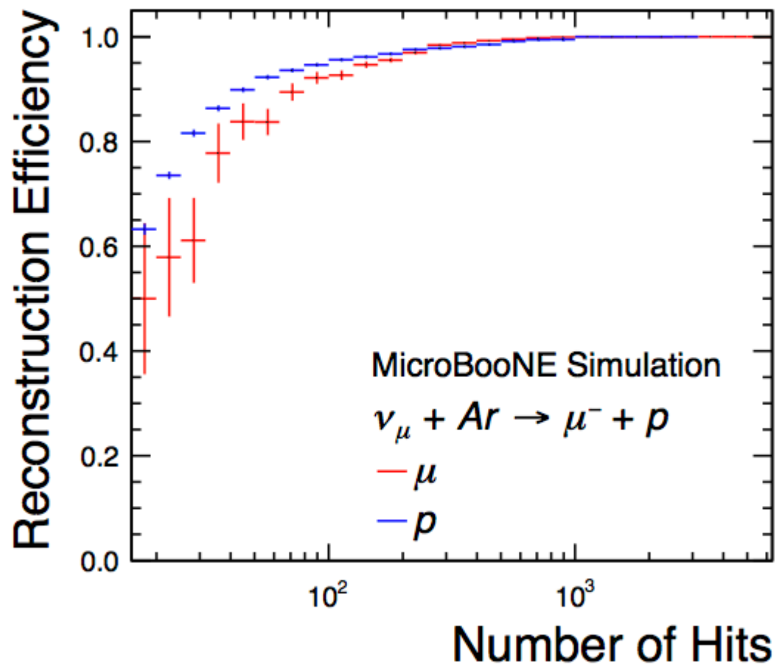


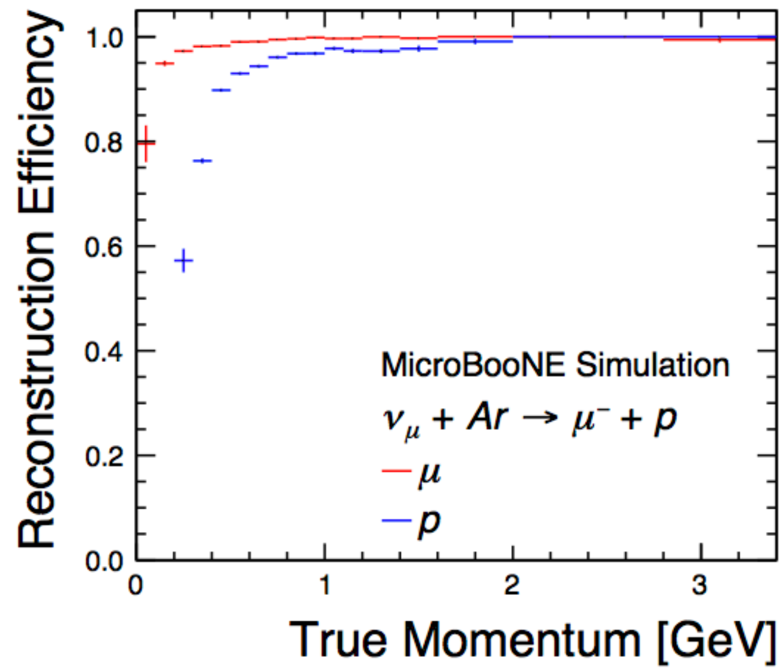
MicroBooNE simulation

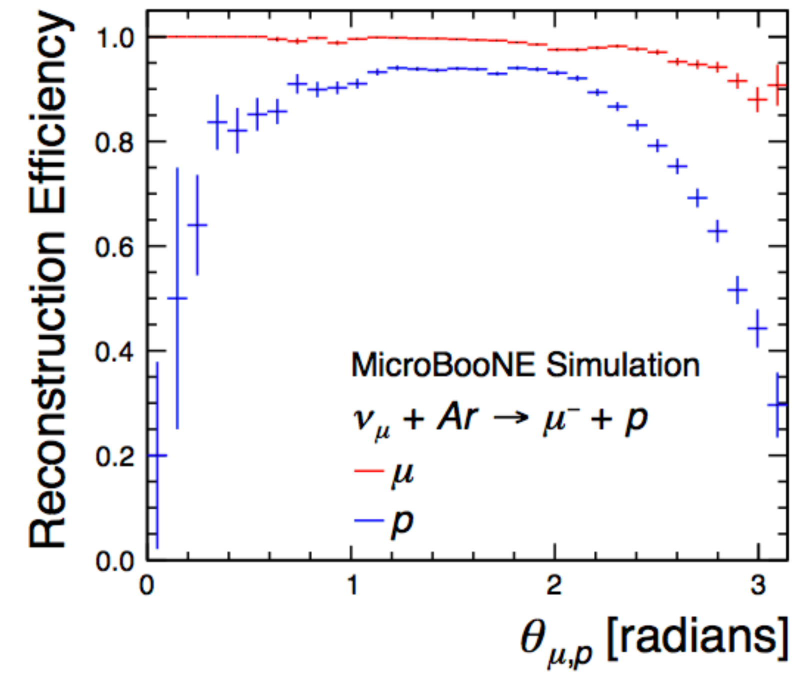
EPJC (2018) 78:82

520 MeV ν_μ

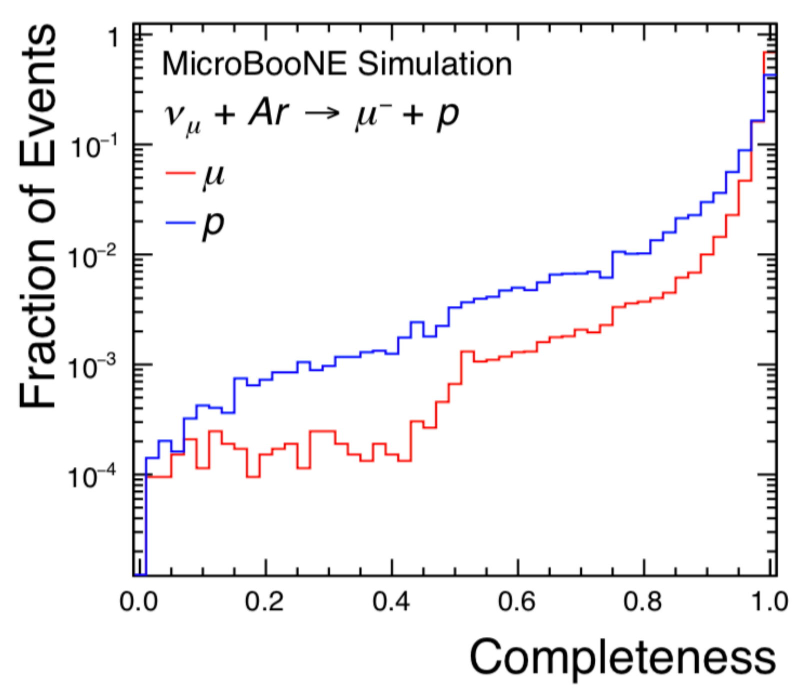
BNB CC QE: $\nu_\mu + Ar \rightarrow \mu^- + p$

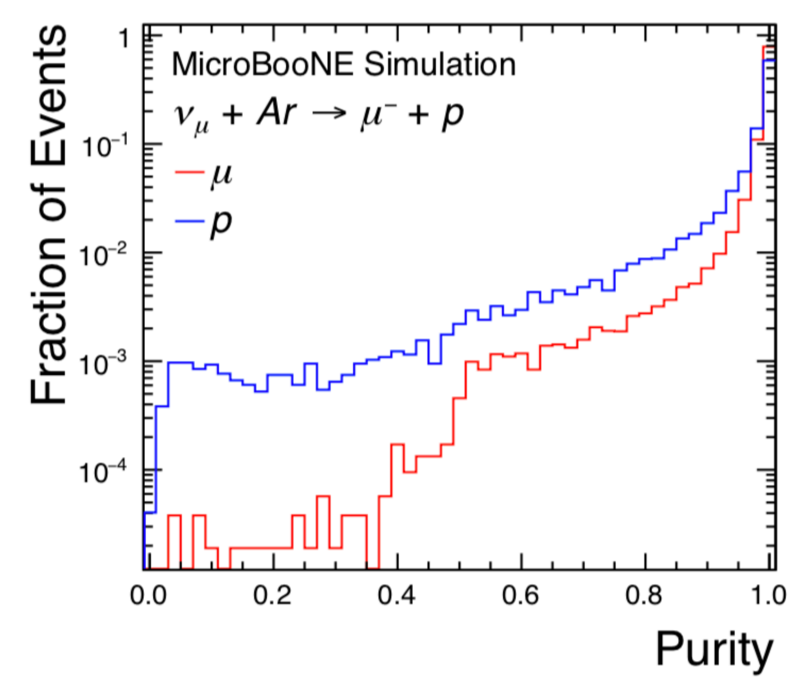






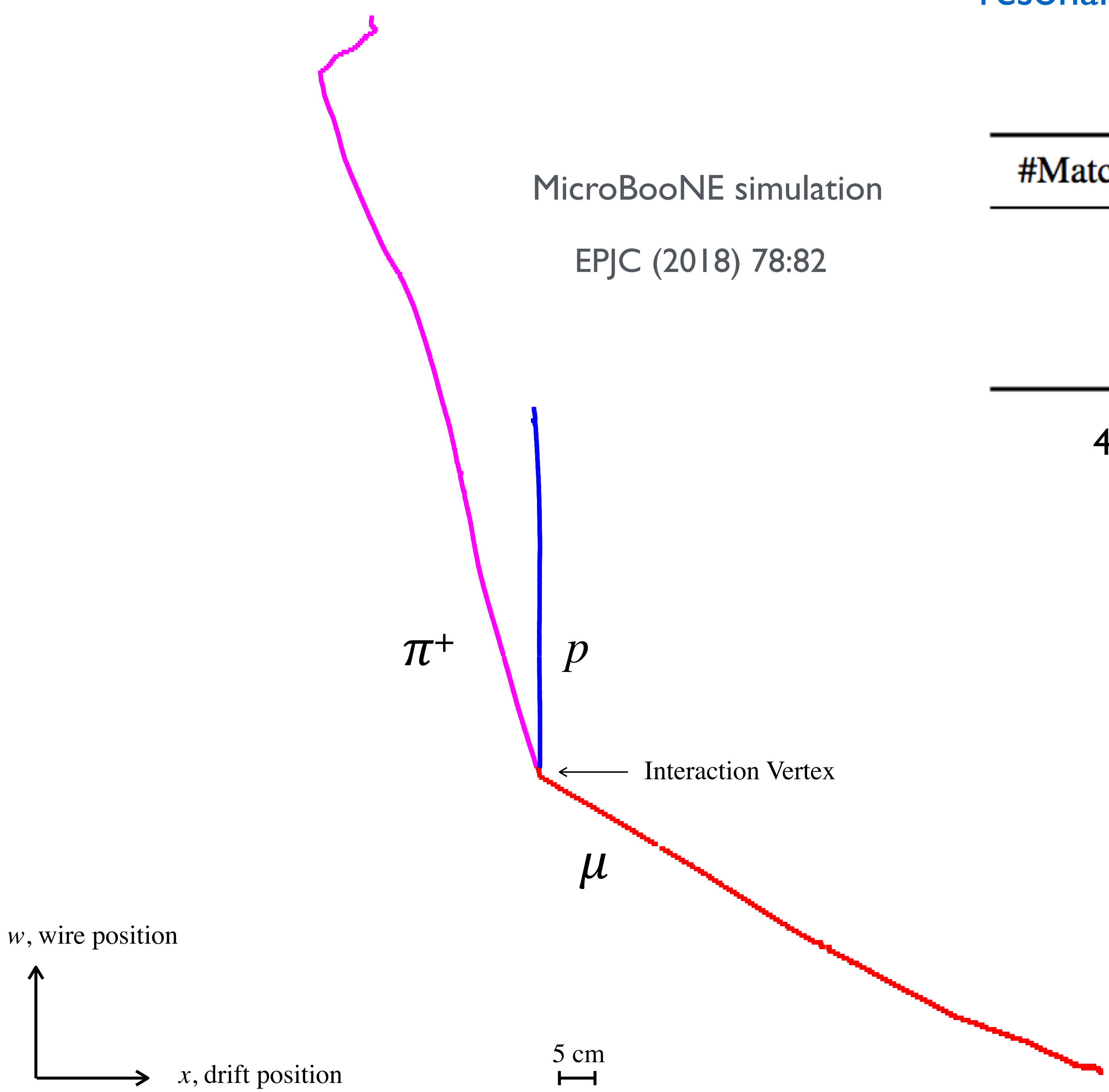
- The most common failure mechanism is merging muon and proton into a single reconstructed particle.
- Single particle is matched to target with which it shares most hits, which will preferentially be the muon.





EPJC (2018) 78:82

BNB CC RES: $\nu_\mu + \text{Ar} \rightarrow \mu^- + p + \pi^+$



Three-track topology: CC ν_μ interactions with resonant charged pion production:

#Matched Particles	0	1	2	3+
μ	3.5%	95.1%	1.4%	0.0%
p	9.0%	86.8%	4.0%	0.3%
π^+	6.9%	80.9%	11.4%	0.8%

47,754 events, 70.5% have exactly one reco particle matched to each target.

- Performance for μ and p similar to that reported for quasi-elastic events.
- π^+ interactions can lead to hierarchy of visible particles. If reconstructed separately (without parent-daughter links), π^+ is reportedly split.

1.1 GeV ν_μ

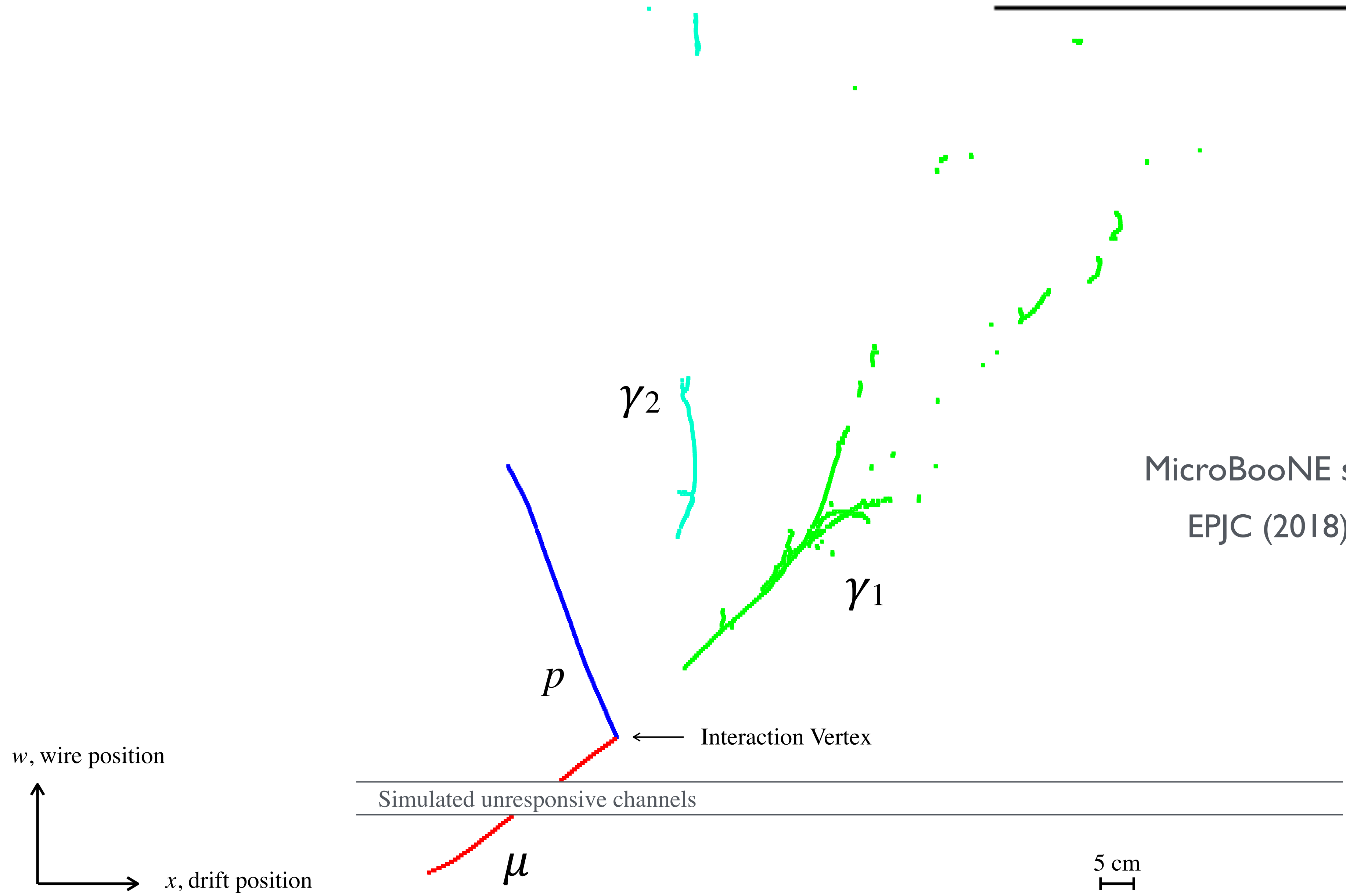
BNB CC RES: $\nu_\mu + \text{Ar} \rightarrow \mu^- + p + \pi^0$

Two-photon topology: CC ν_μ interactions with resonant neutral pion production:

#Matched Particles	0	1	2	3+
μ	3.7%	94.8%	1.5%	0.0%
p	9.9%	85.5%	4.3%	0.3%
γ_1	6.8%	88.0%	4.8%	0.4%
γ_2	29.9%	66.4%	3.6%	0.2%

17,939 events, 49.9% have exactly one reco particle matched to each target.

#hits γ_1 > #hits γ_2



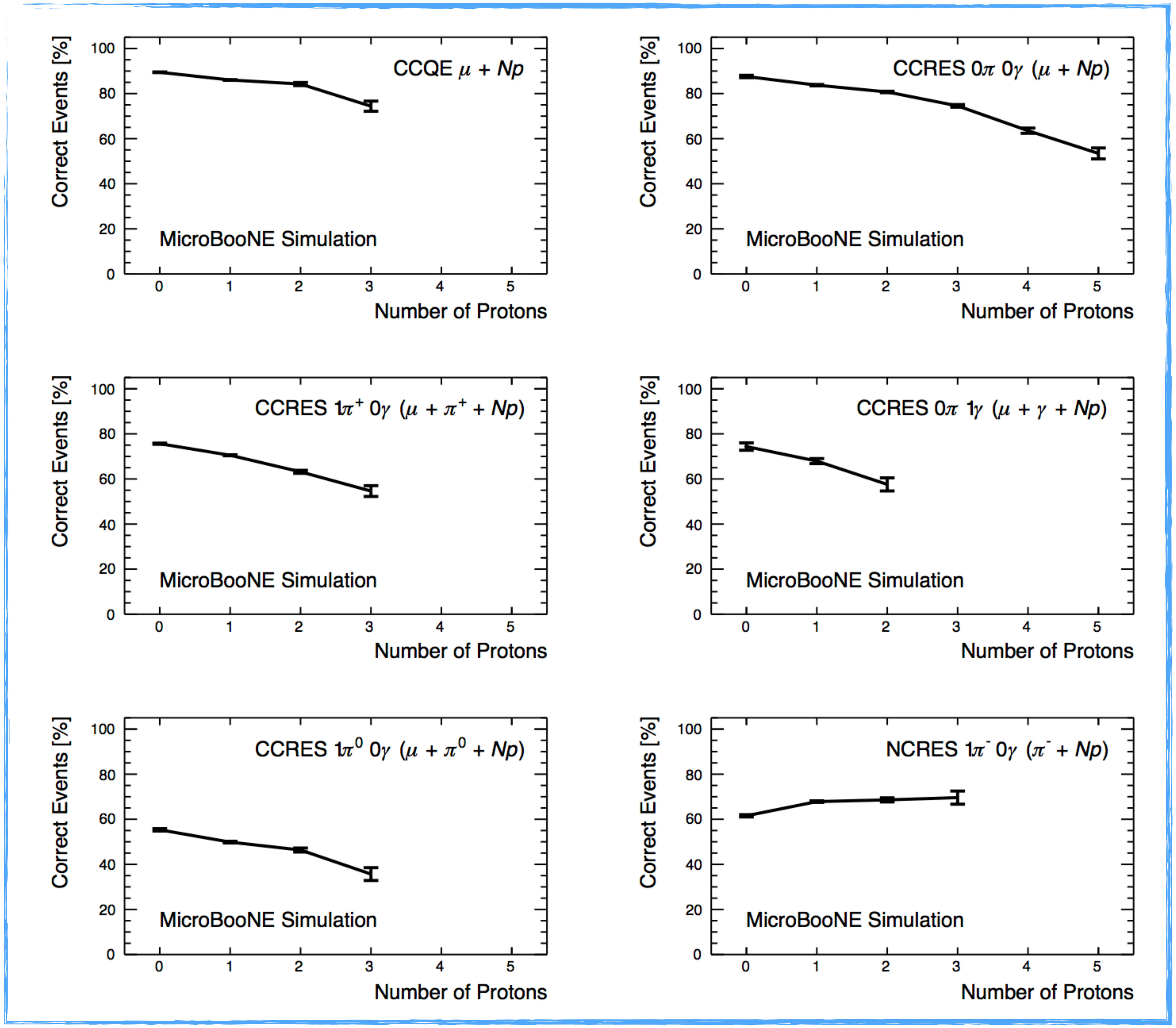
MicroBooNE simulation

EPJC (2018) 78:82

1.4 GeV ν_μ

Selection of Exclusive Final States

- Assess larger selection of exclusive final states using correct event fraction.
- Recall aim: a general purpose reconstruction for diverse event topologies.

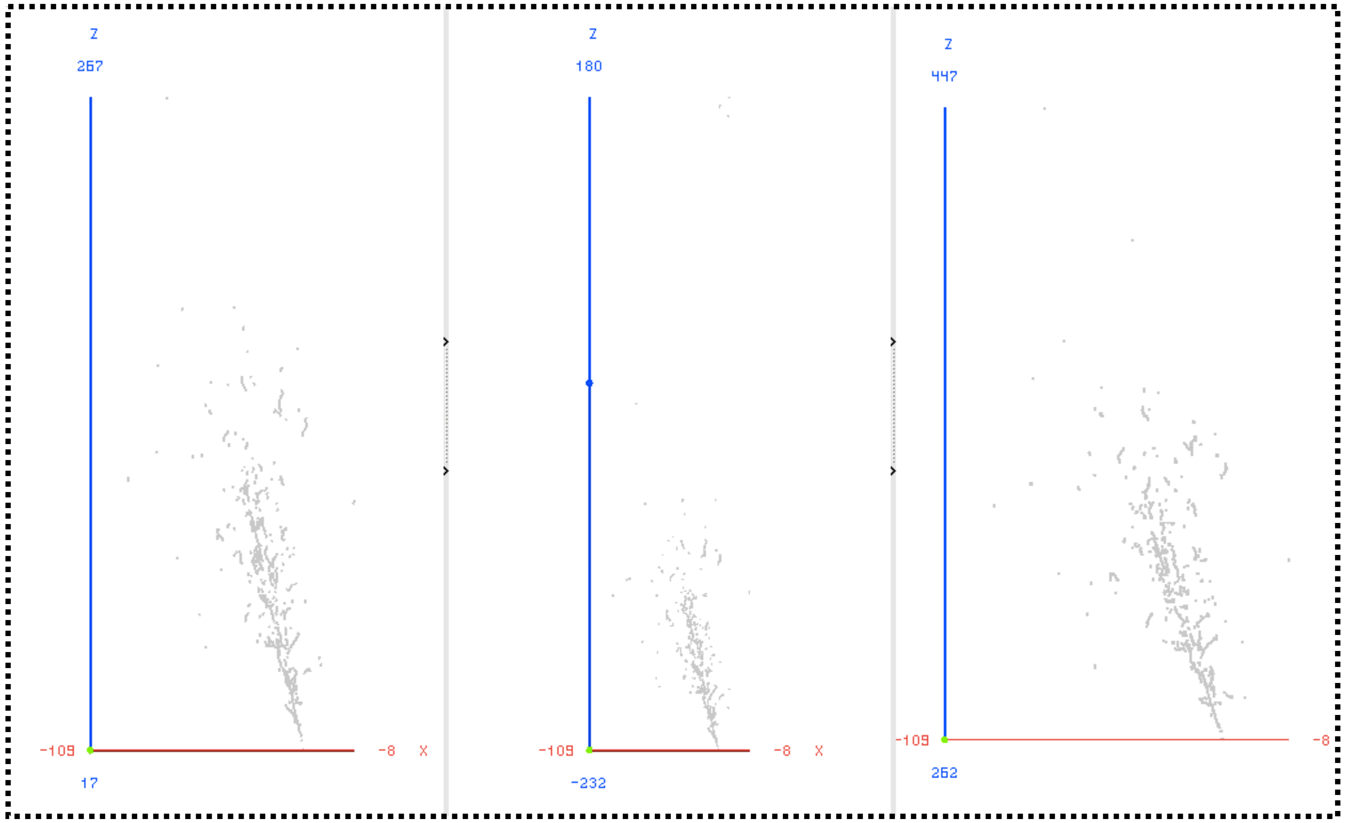


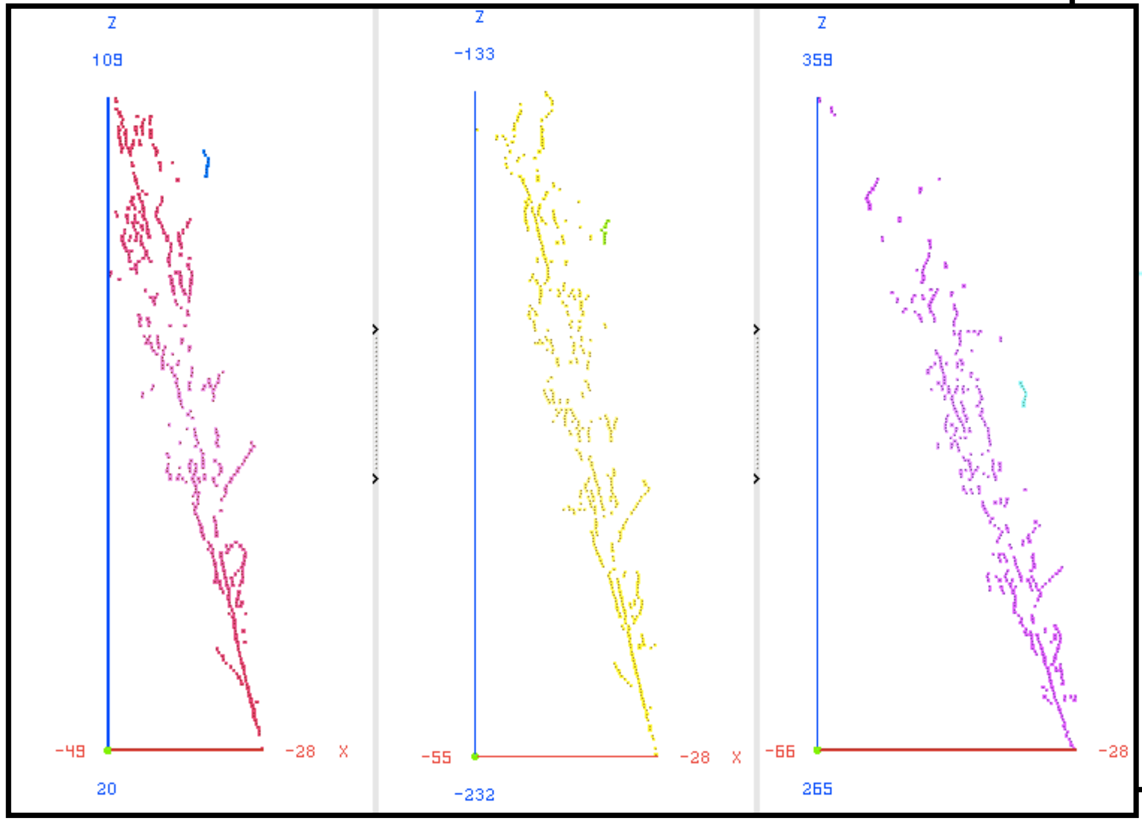
EPJC (2018) 78:82

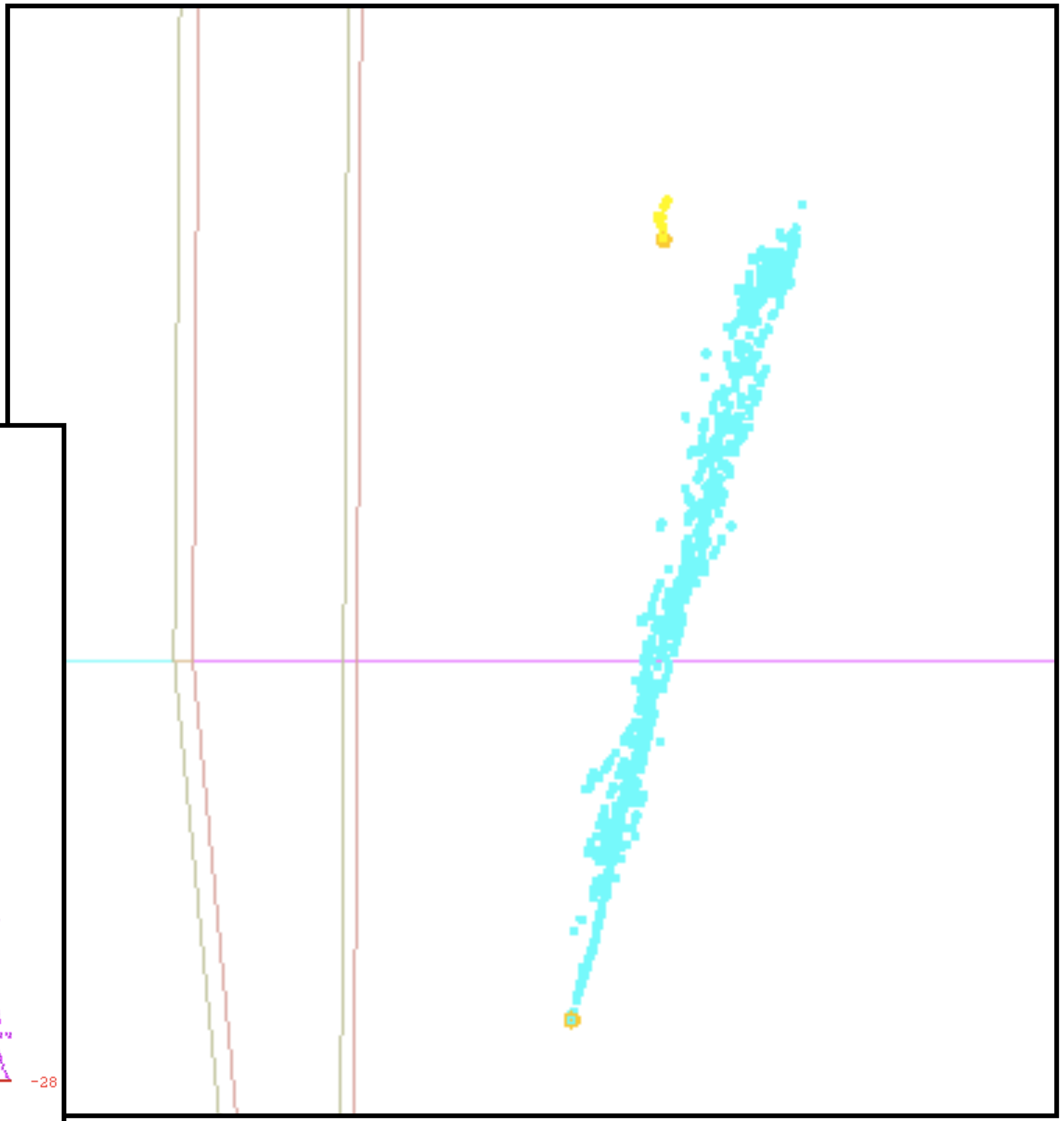
PandoraTestBeam Output

ProtoDUNE-SP

- Positron test beam particle



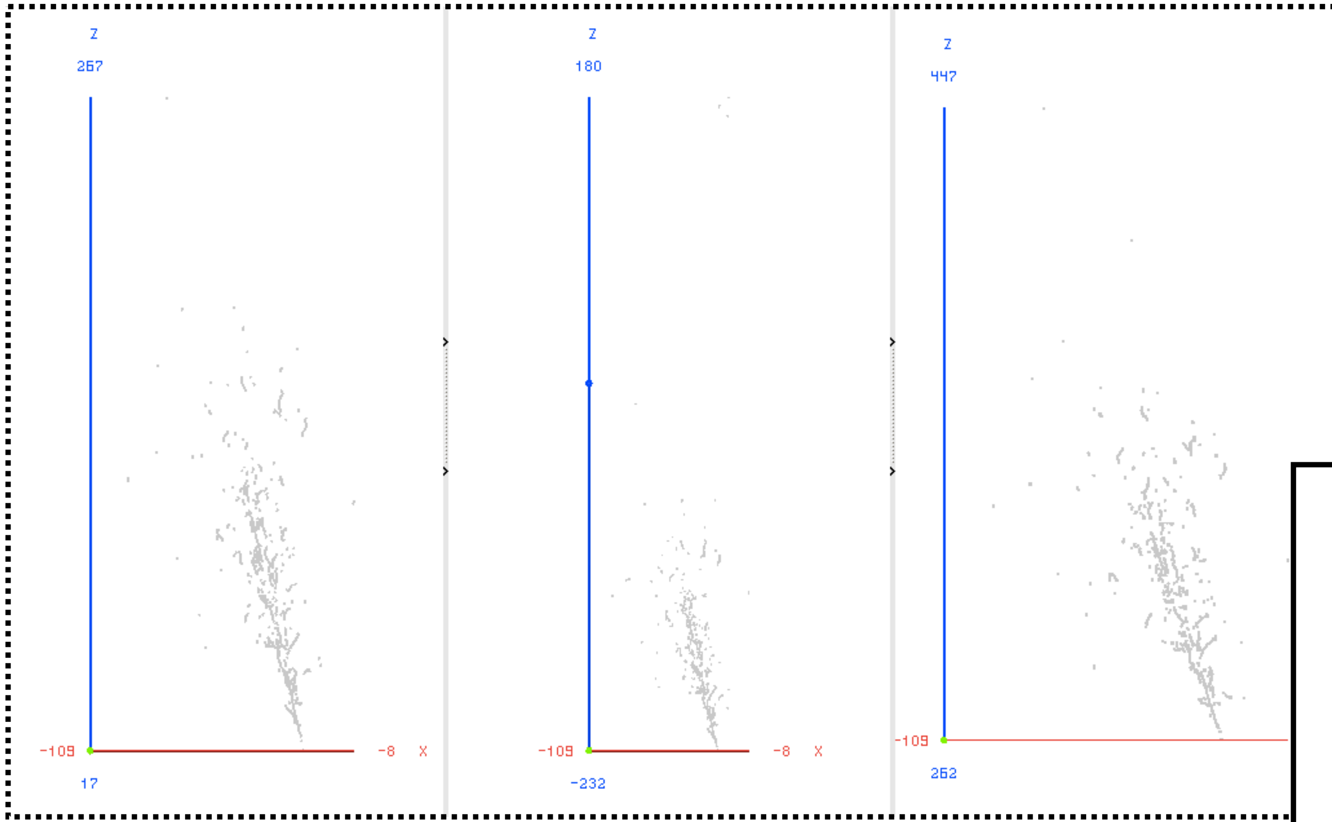


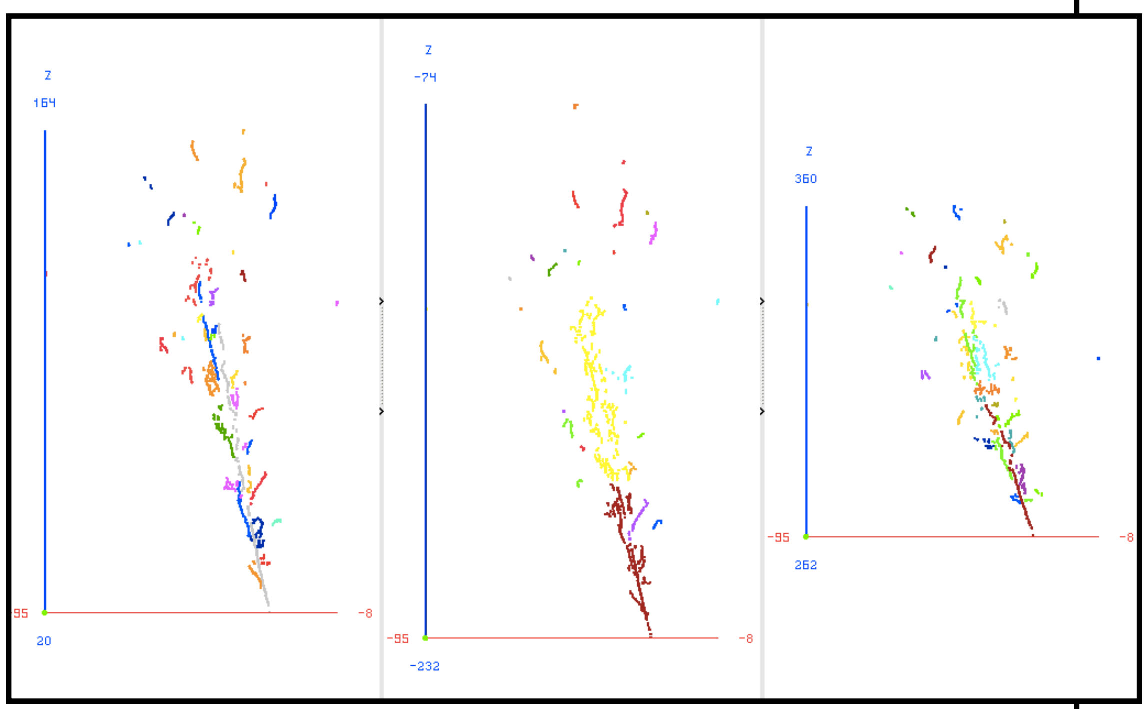


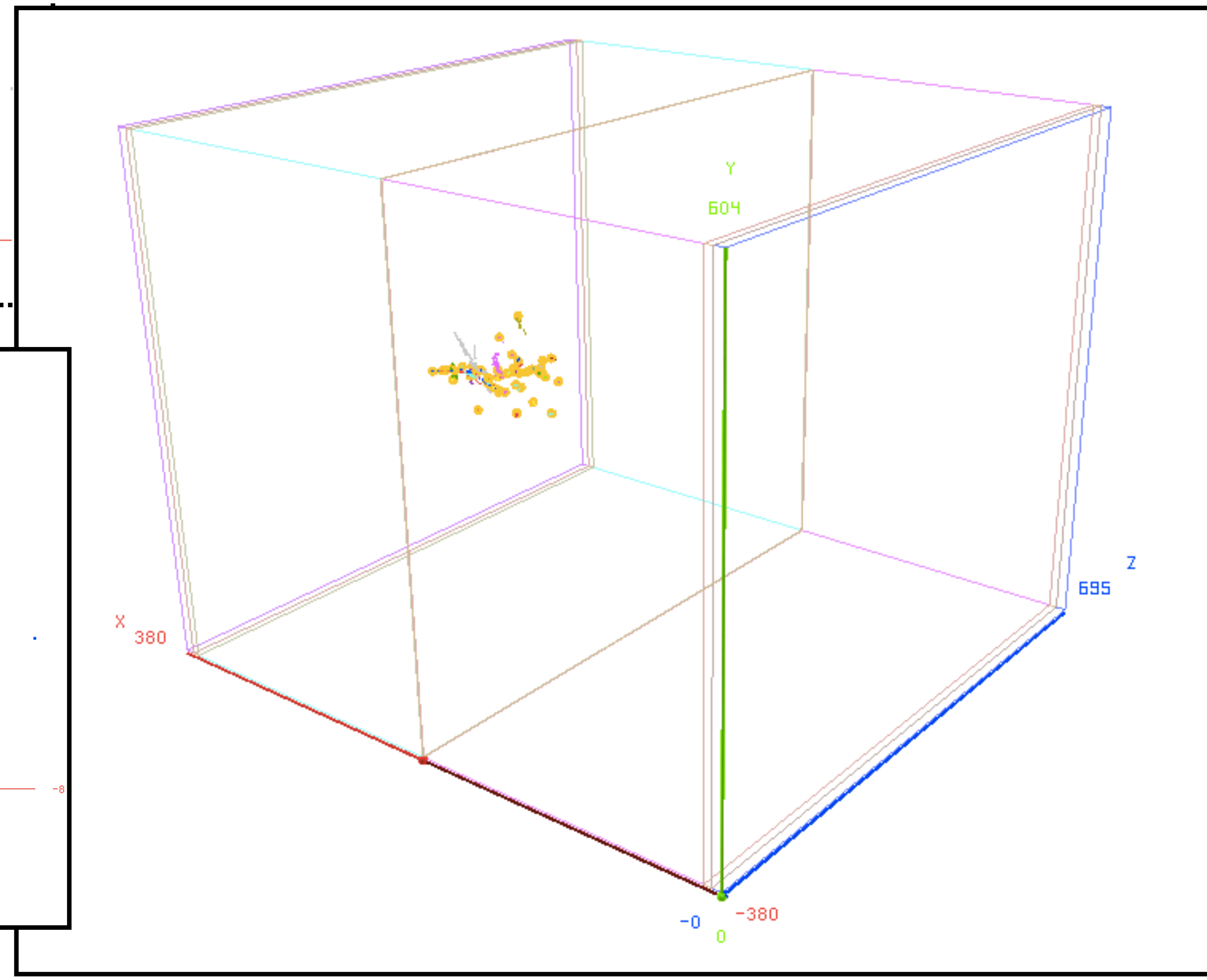
PandoraCosmic Output

ProtoDUNE-SP

- Positron test beam particle



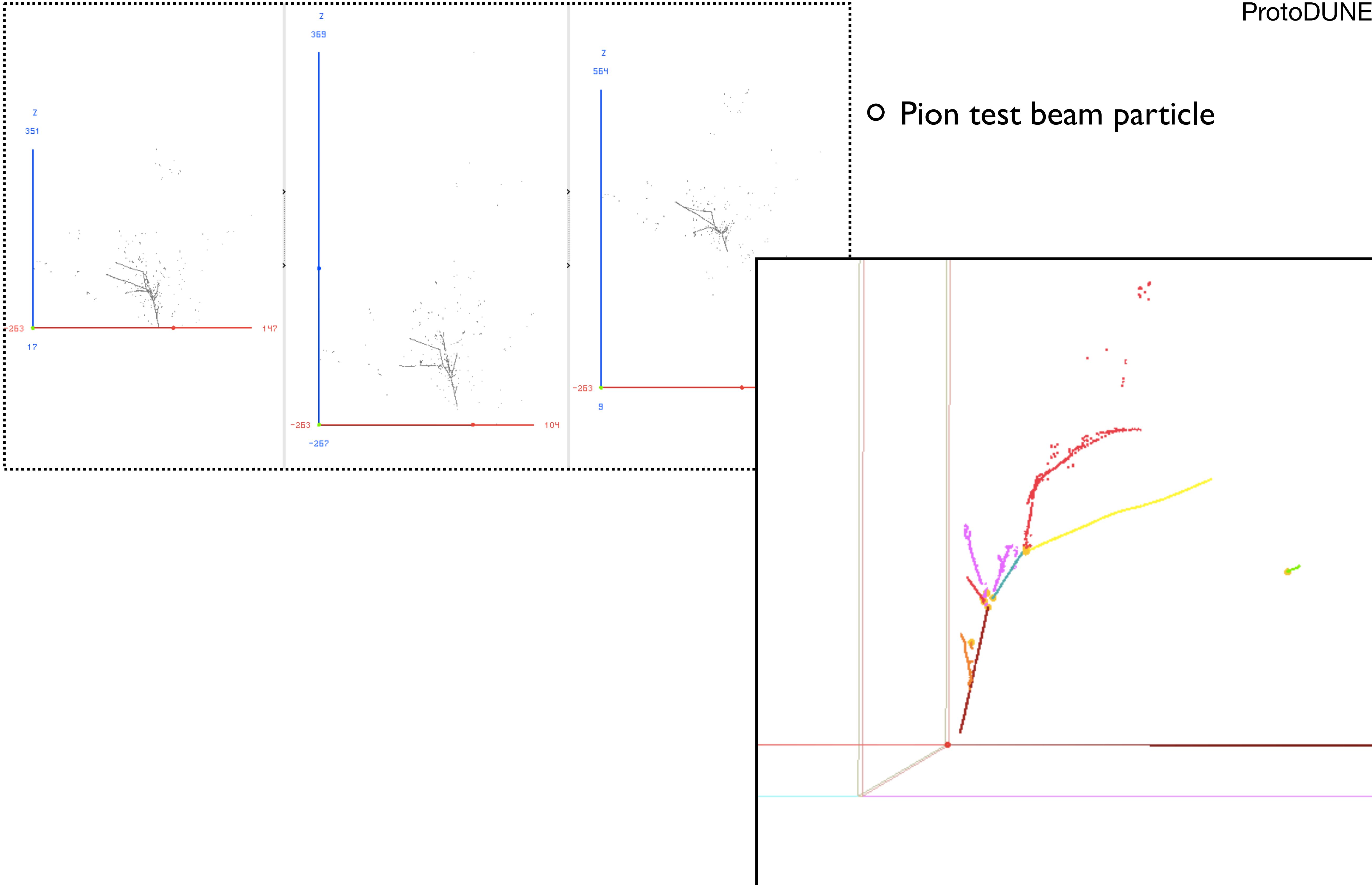




PandoraTestBeam Output

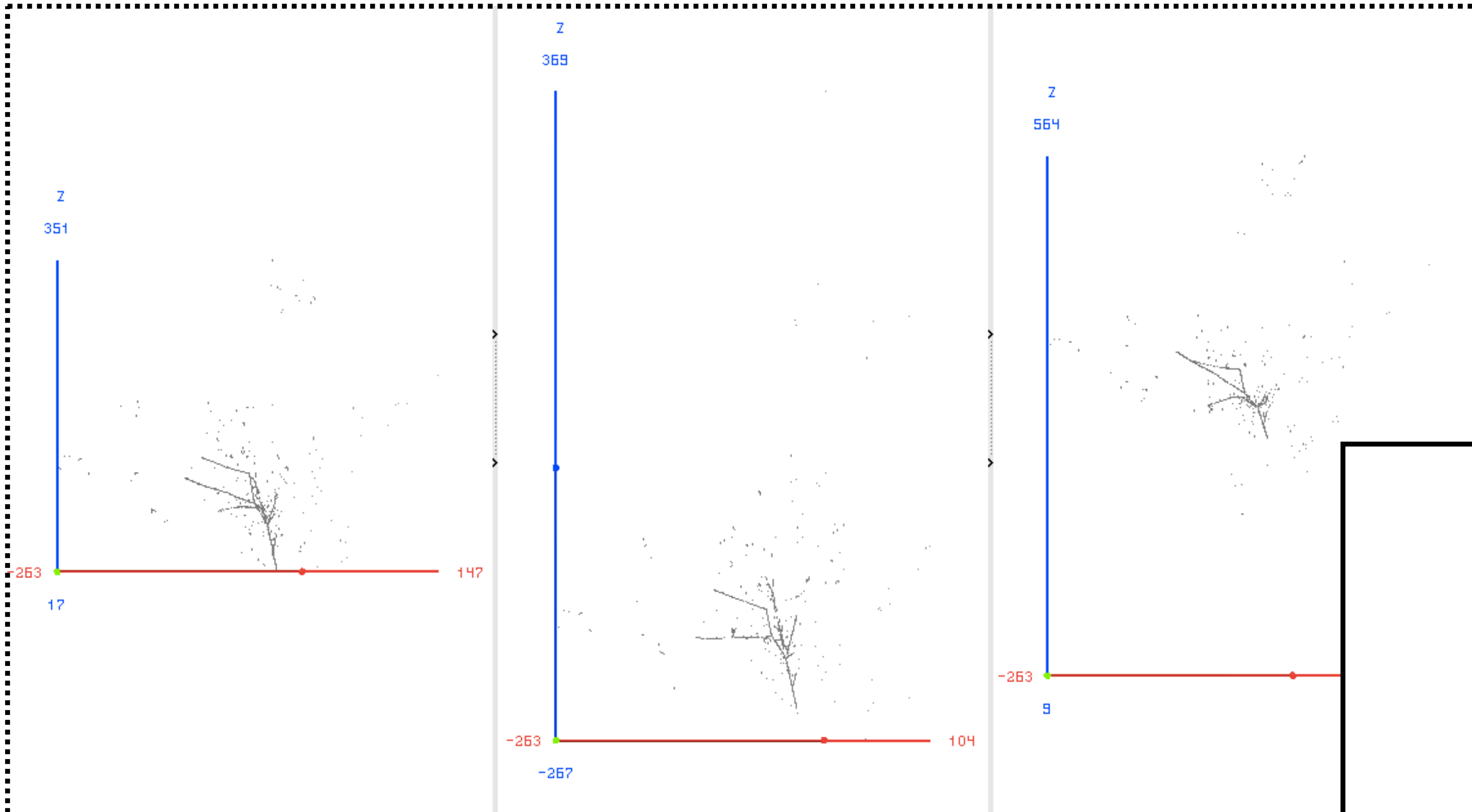
ProtoDUNE-SP

- Pion test beam particle

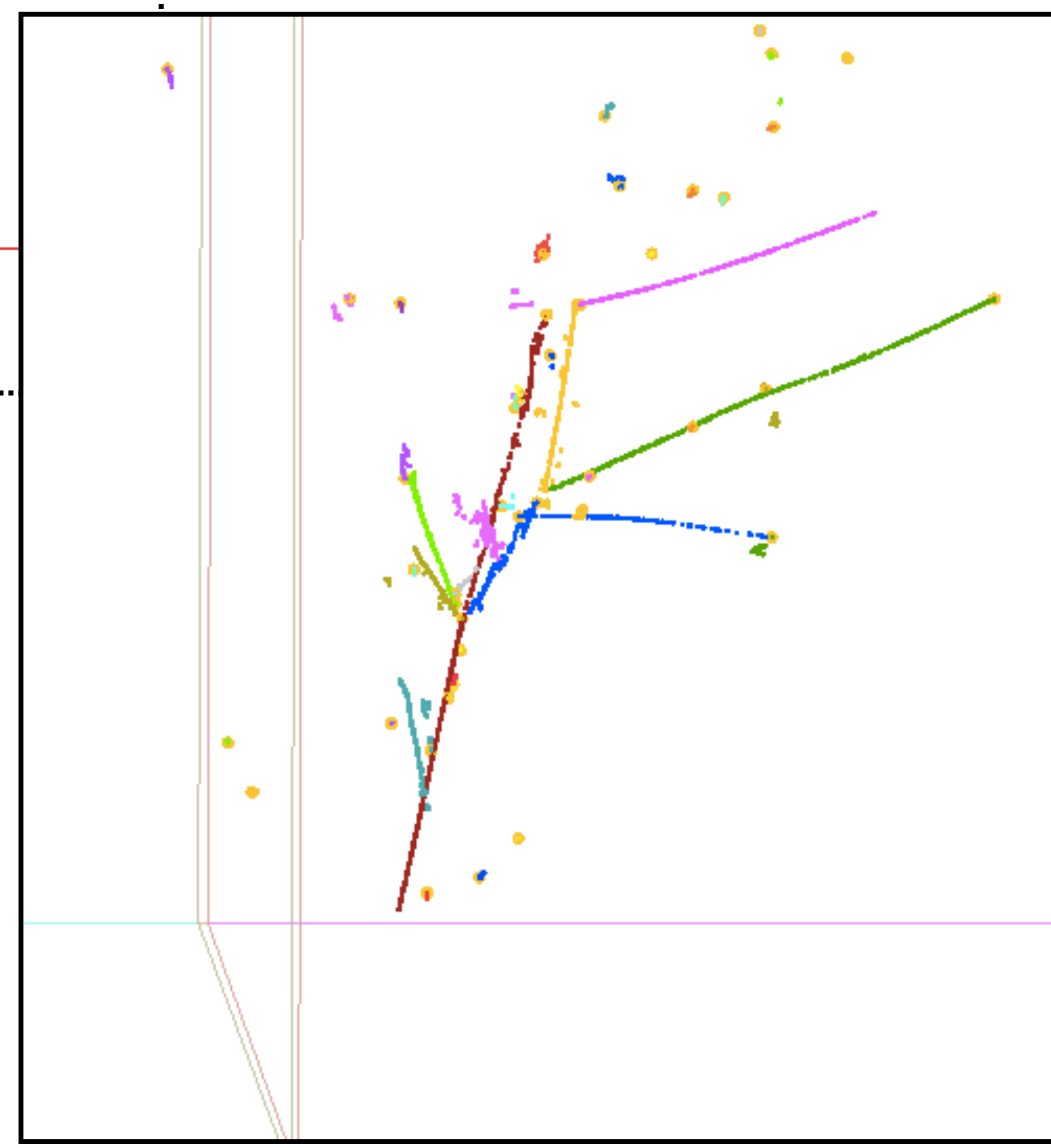


PandoraCosmic Output

ProtoDUNE-SP



- Pion test beam particle



- Track driven through test beam interaction because the Pandora Cosmic algorithm chain is trying to reconstruct cosmic rays.