Pandora: LArTPC Event Reconstruction

John Marshall for the Pandora Team

16th September 2020

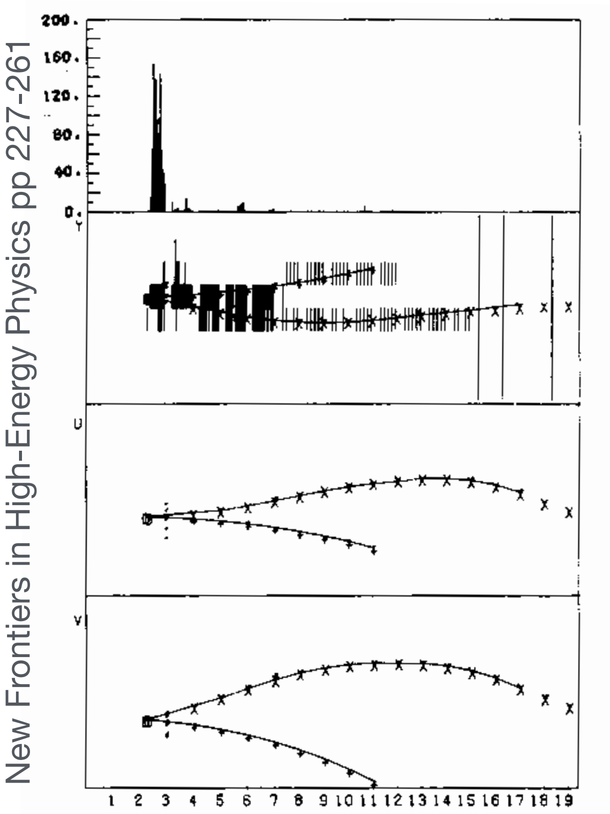
Overview

- **LArTPC detectors and event reconstruction**
- **Details of key Pandora LArTPC pattern-recognition algorithms**
- **Handling LArTPCs w/ multiple volumes and cosmic-ray backgrounds**
- **Pattern-recognition performance**

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

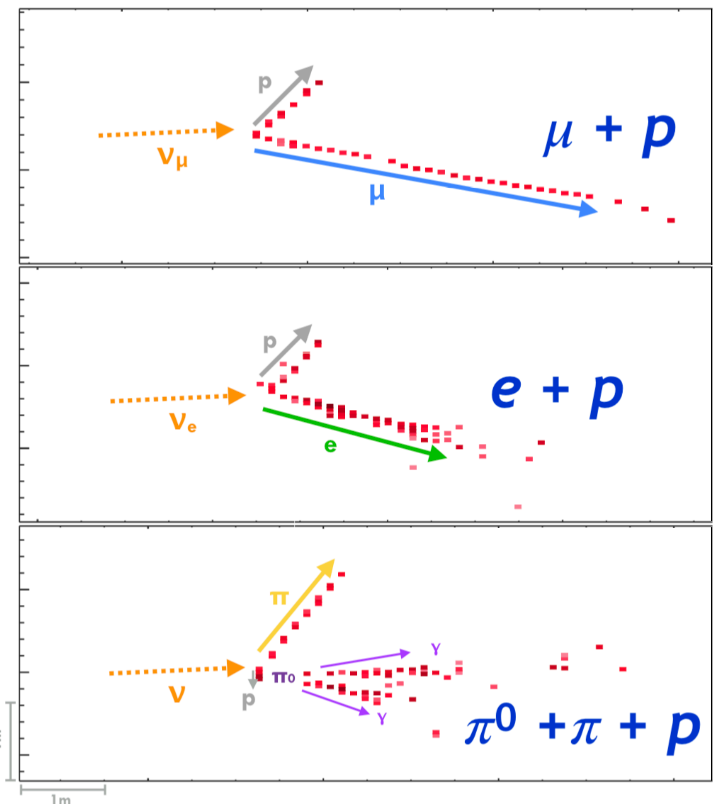
Neutrino Detectors

New Frontiers in High-Energy Physics pp 227-261



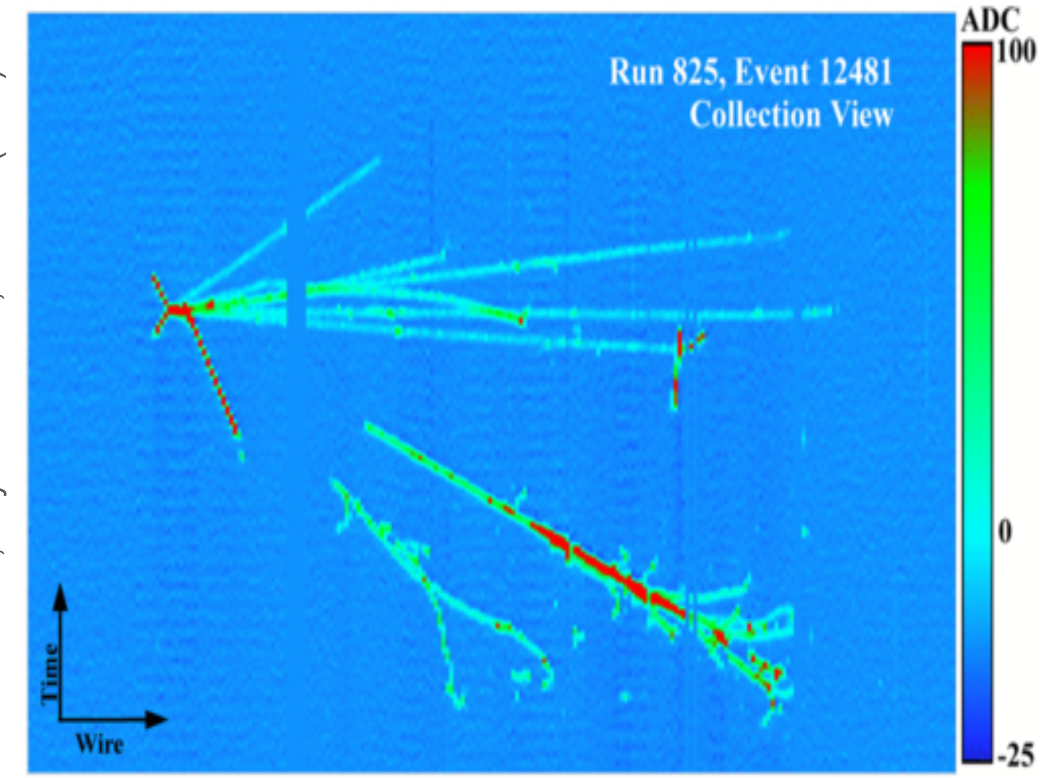
CDHS

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
- Physics sensitivity now depends critically on both hardware and software
  - Need a sophisticated event reconstruction to harness information in the images
- Aim to reconstruct hierarchy of particles of identified types, with measured four-momenta
  - "Particle flow" reconstruction

Liquid-Argon Time-Projection Chamber (LArTPC)

- Charged particles, e.g. produced in neutrino interactions, deposit ionisation trails in liquid argon.
- Ionisation electrons drift in an applied electric field.
- In a single-phase LArTPC, the electrons are detected by a series of wire planes.
- Single-phase LArTPC detectors:
  - Past: ICARUS, ArgoNeuT
  - Current: MicroBooNE, ProtoDUNE, LArIAT
  - Coming soon: SBND, ICARUS@SBN

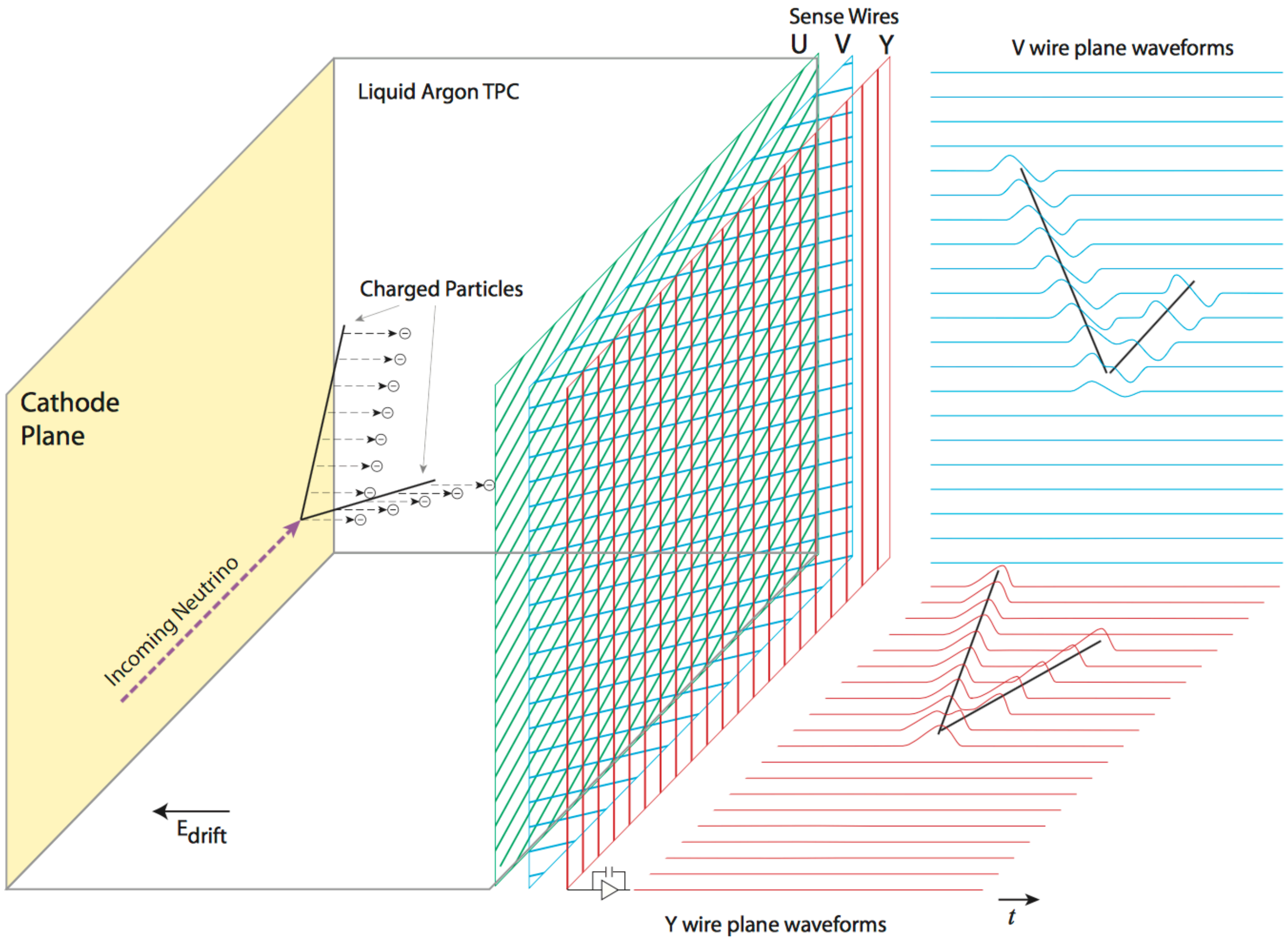
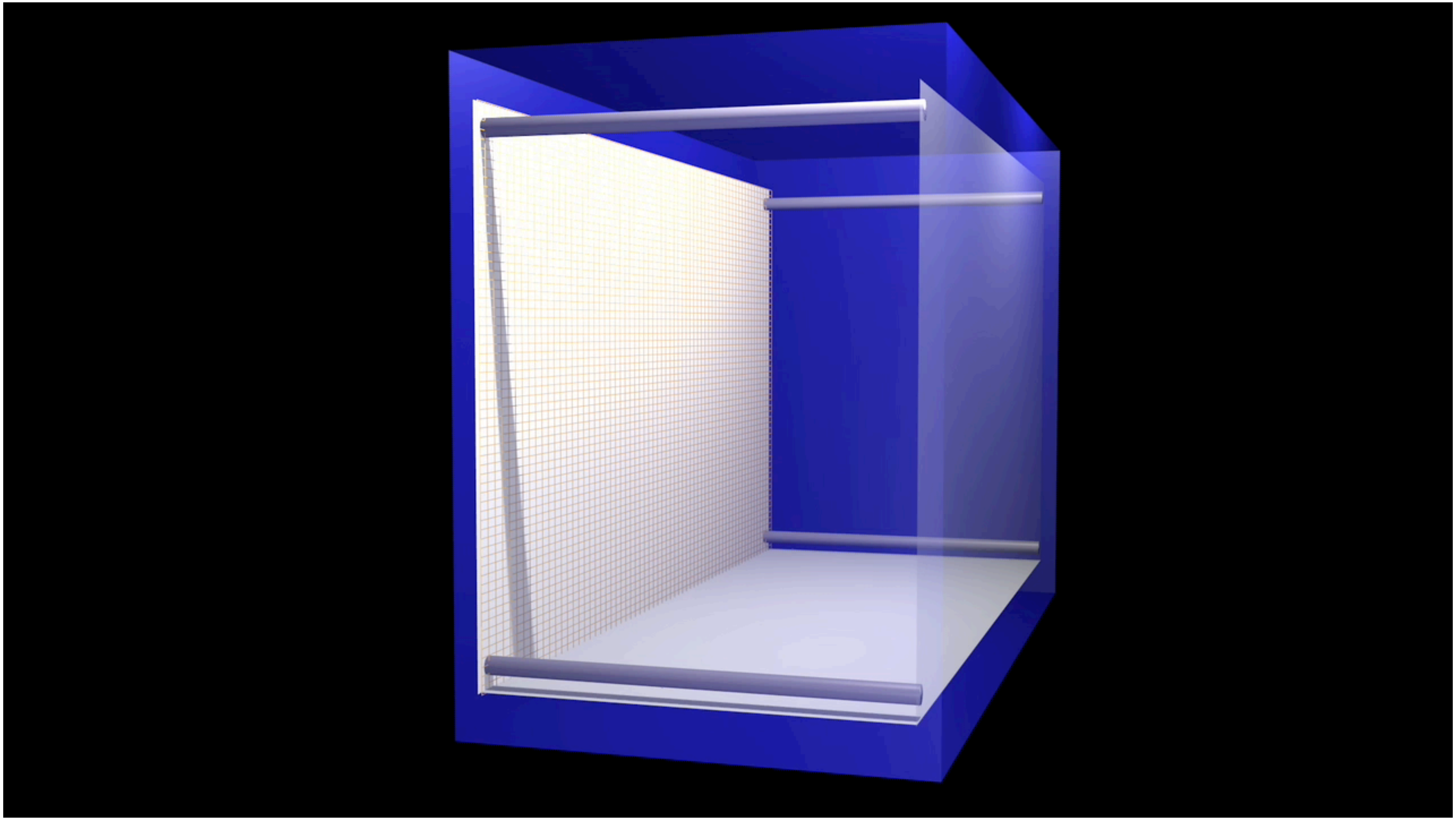


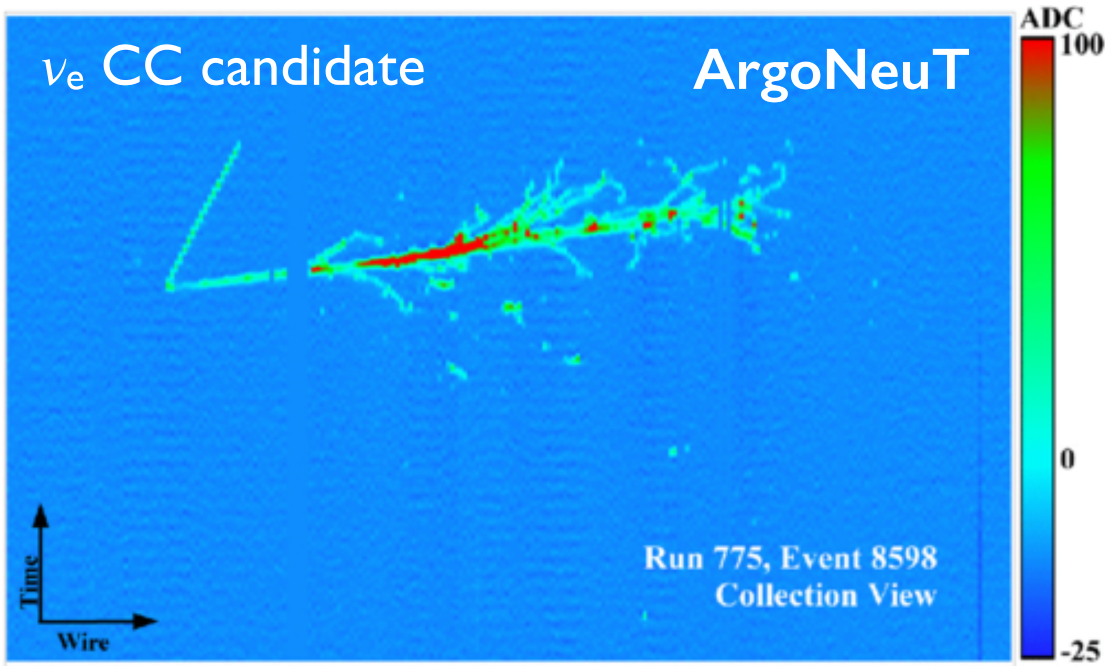
Figure: arXiv:1612.05824

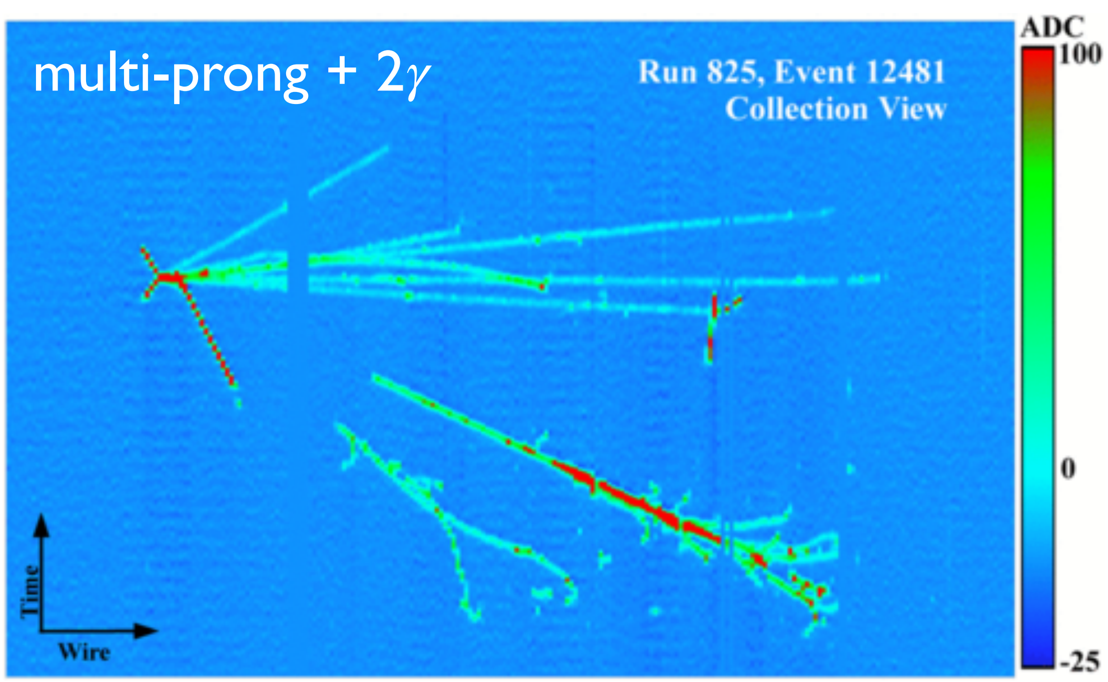
LArTPCs: a somewhat slicker introduction (assuming embedded video works…)



Video: Fermilab

Why Liquid Argon?





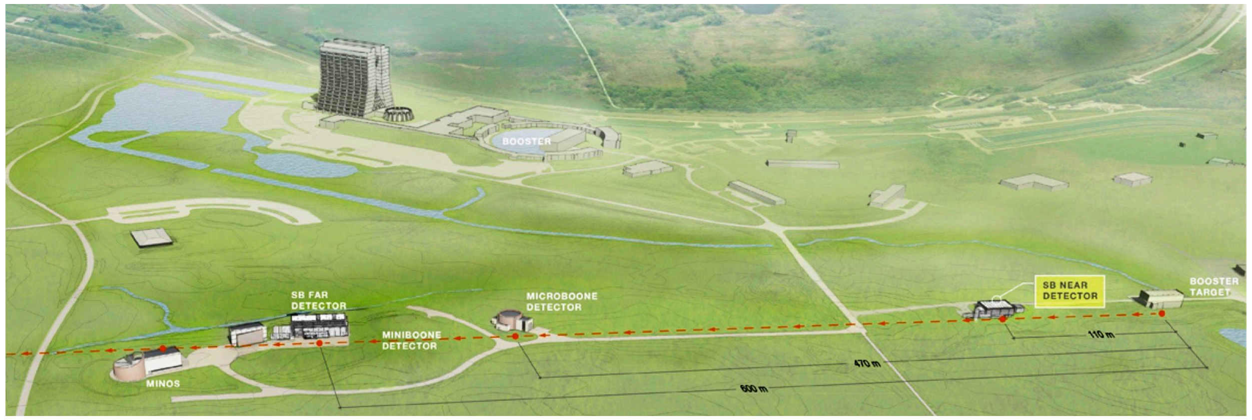
R. Acciari *et al*, Phys. Rev. D 95, 072005 (2017)

- LArTPC detectors are fully active and fine grain, offering superb spatial and calorimetric resolution:
  - Reconstruction of multi-prong final states.
  - Particle identification:
    - $\mu/p/K$ in particle tracks
    - e/γ in electromagnetic showers
- Potential for high efficiency and low backgrounds in most channels
- Scalable to multi-kiloton masses.

LArTPC Programme

One of the key technologies in the current and future neutrino physics programmes is the Liquid-Argon Time-Projection Chamber (LArTPC)

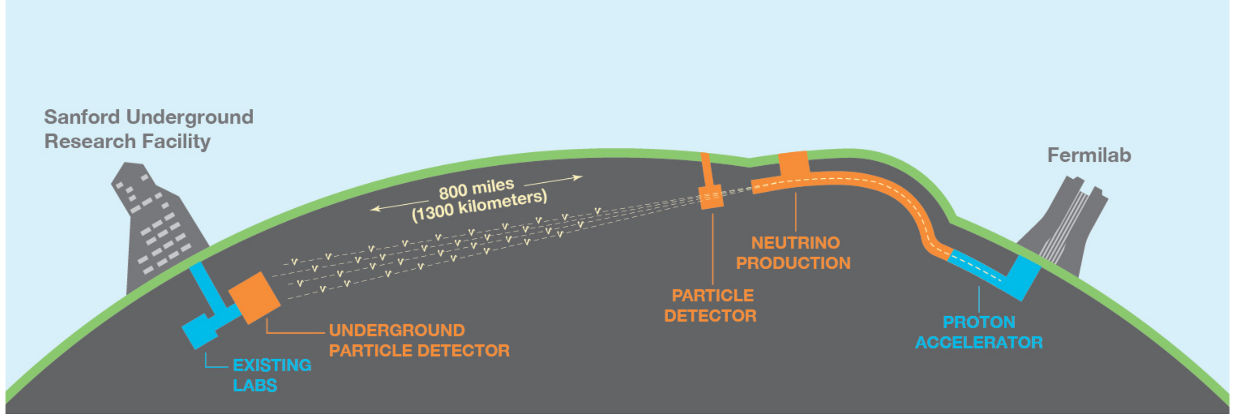
Short-baseline programme



ICARUS **MicroBooNE** **SBND**

- Three LArTPC detectors located along the Booster Neutrino Beam (BNB) at Fermilab
- Main goal is to investigate the potential sterile neutrino signals from LSND and MiniBooNE
- Precision cross-section measurements for neutrino interactions on argon

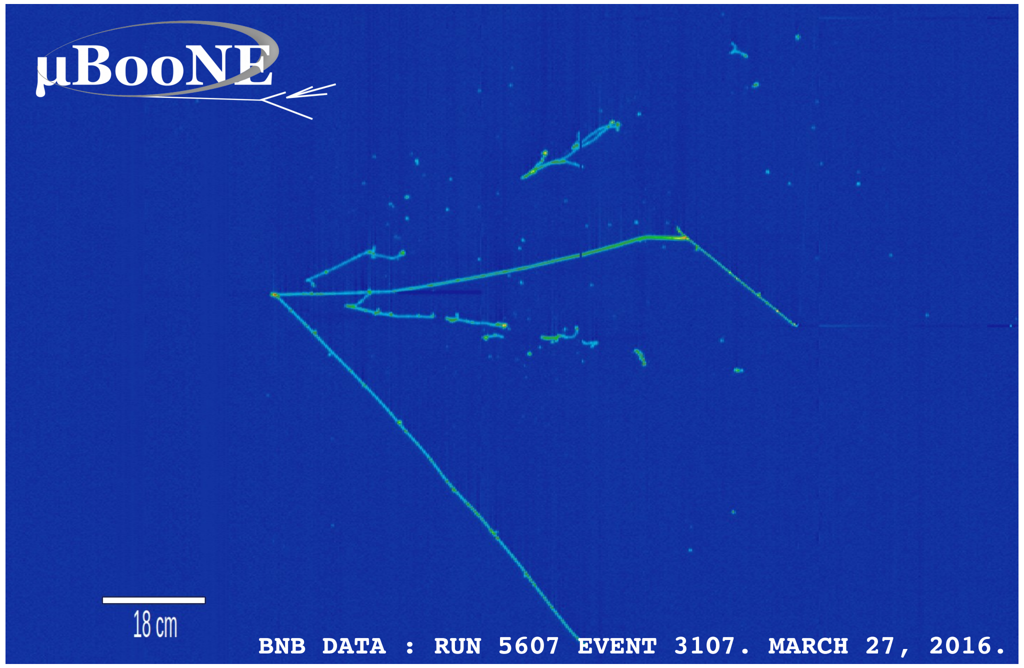
Long-baseline programme

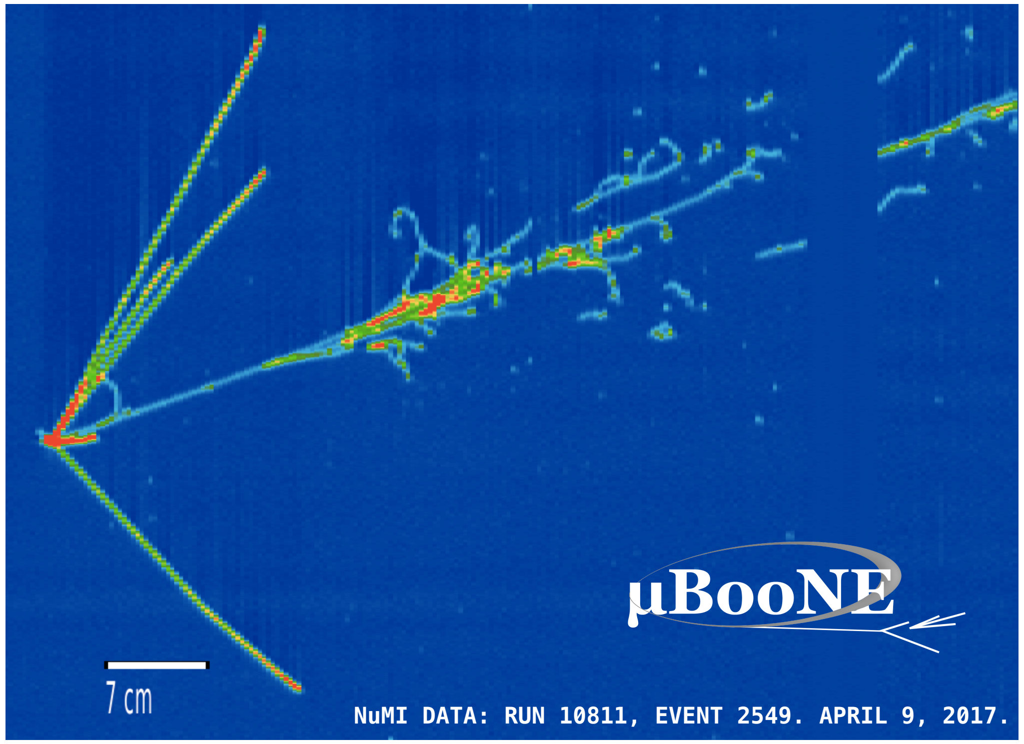


ProtoDUNEs **DUNE**

- Neutrino oscillation physics:
  - Discover CP violation in the leptonic sector
  - Resolve mass hierarchy
  - Test three-flavour paradigm
  - Precision parameter measurement
- Proton decay
- Supernova neutrinos

LArTPC Event Reconstruction

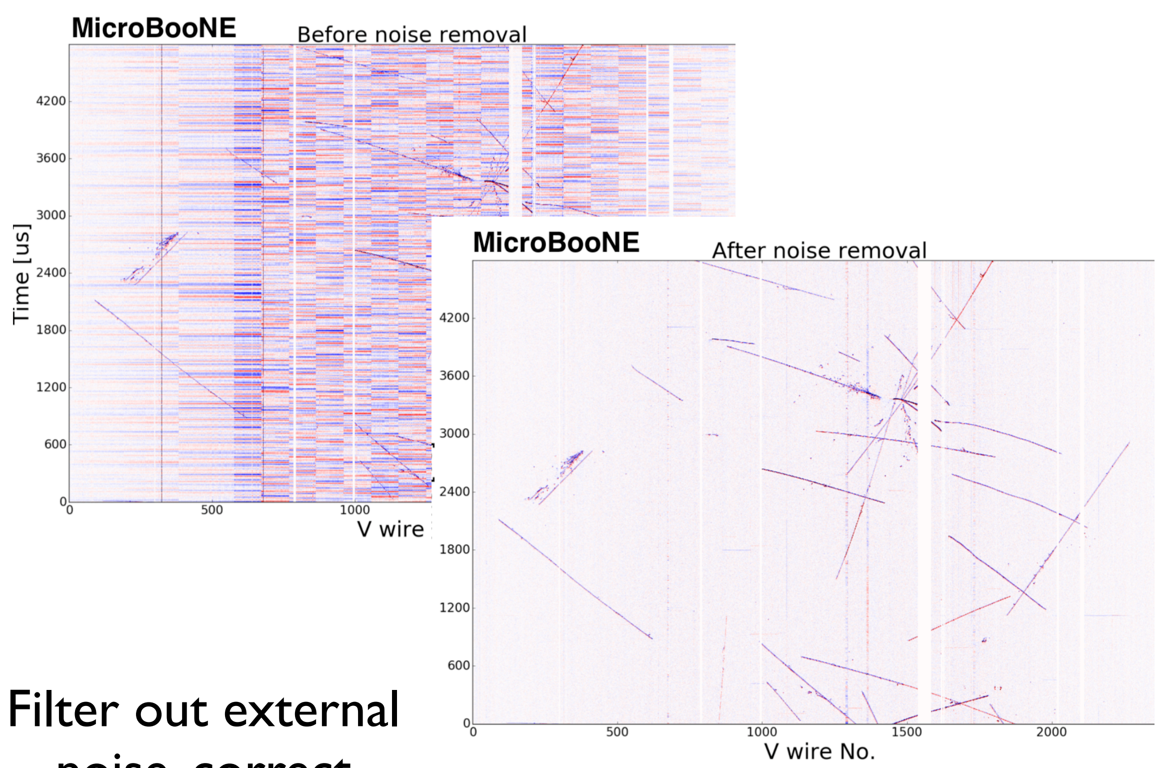




The conversion of raw LArTPC images into analysis-level physics quantities:

- **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…

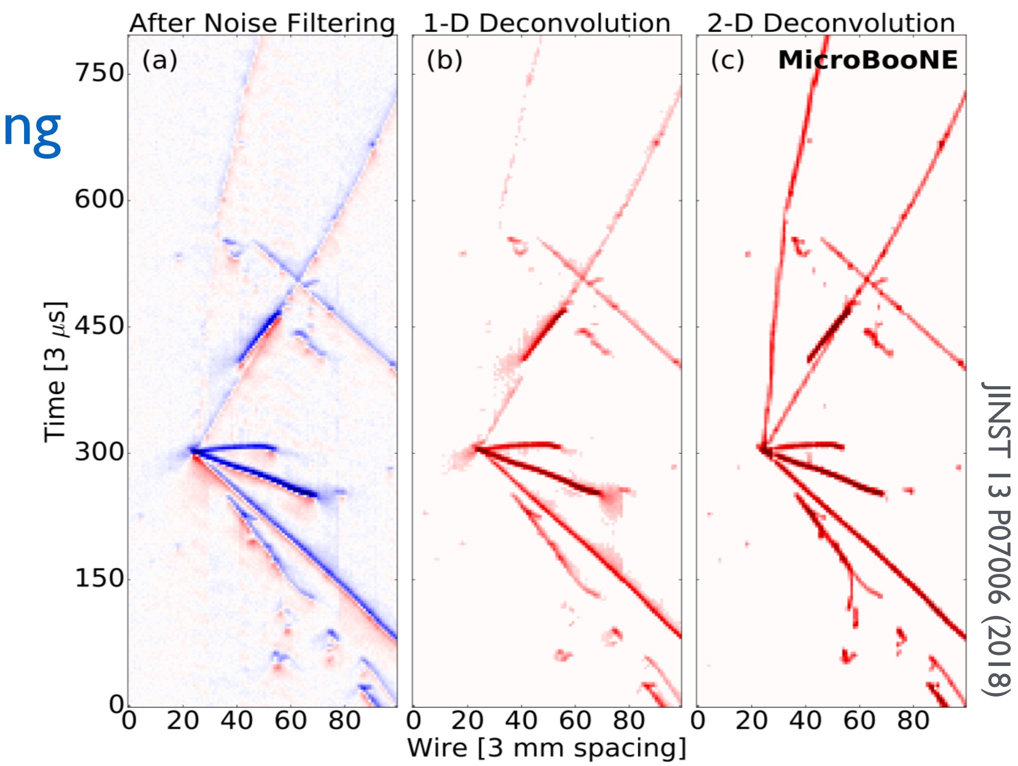
Noise Filtering



Filter out external noise, correct some distortions

JINST 12 P08003 (2017)

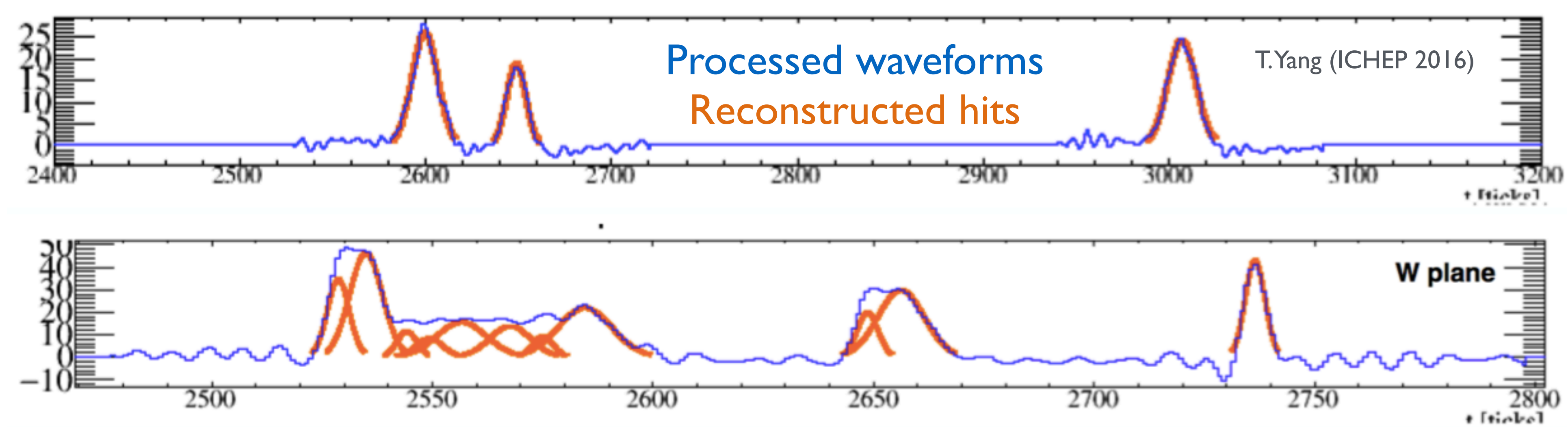
Signal Processing



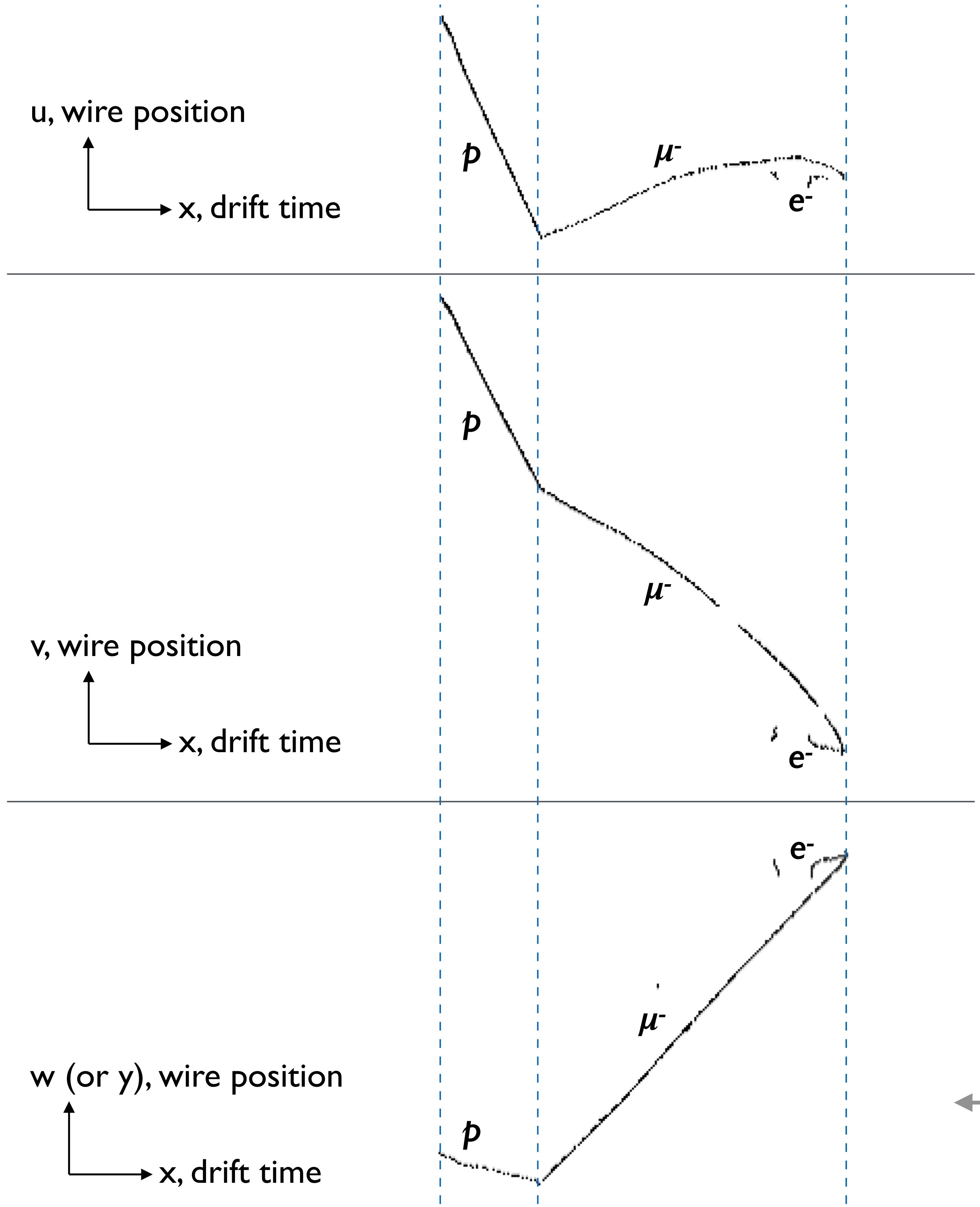
Convert digitised waveform to number of ioniation electrons passing through a wire plane at a given time

Hit Finding

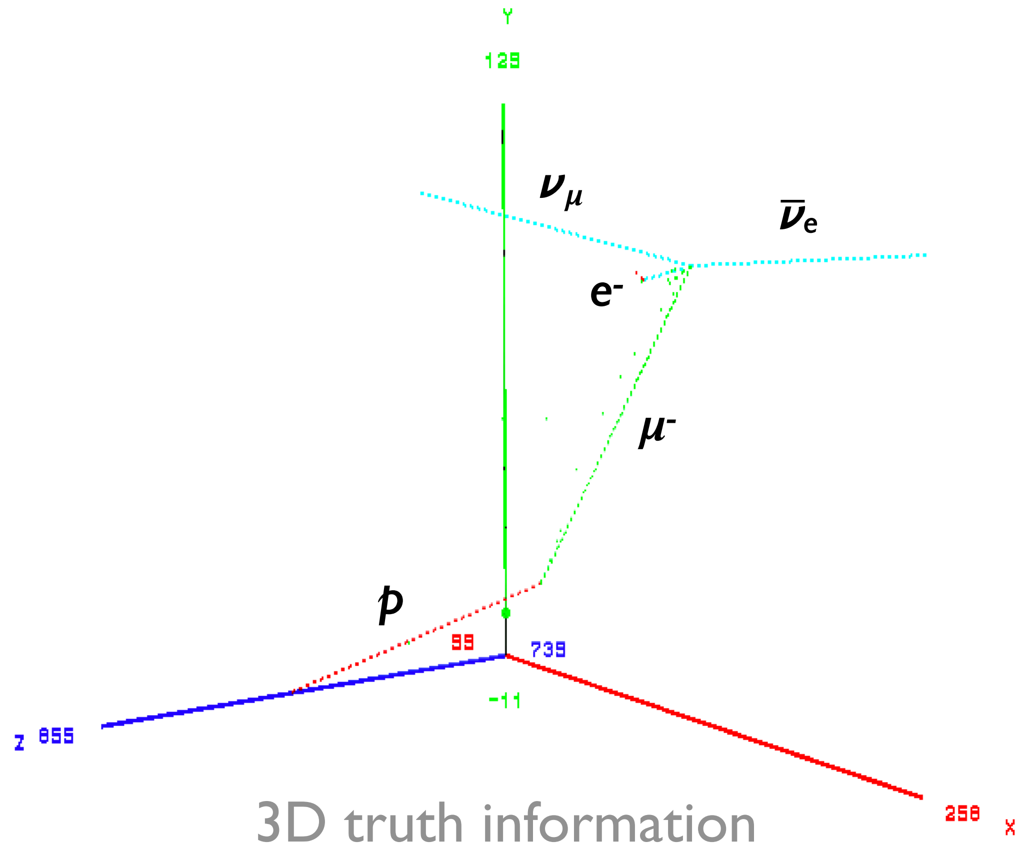
Fit clean waveform with N Gaussians, where N is number of peaks in pulse. Each Gaussian represents a hit.



Inputs to Pattern Recognition



E.g. CC QE: ν_μ + Ar → p + μ^-



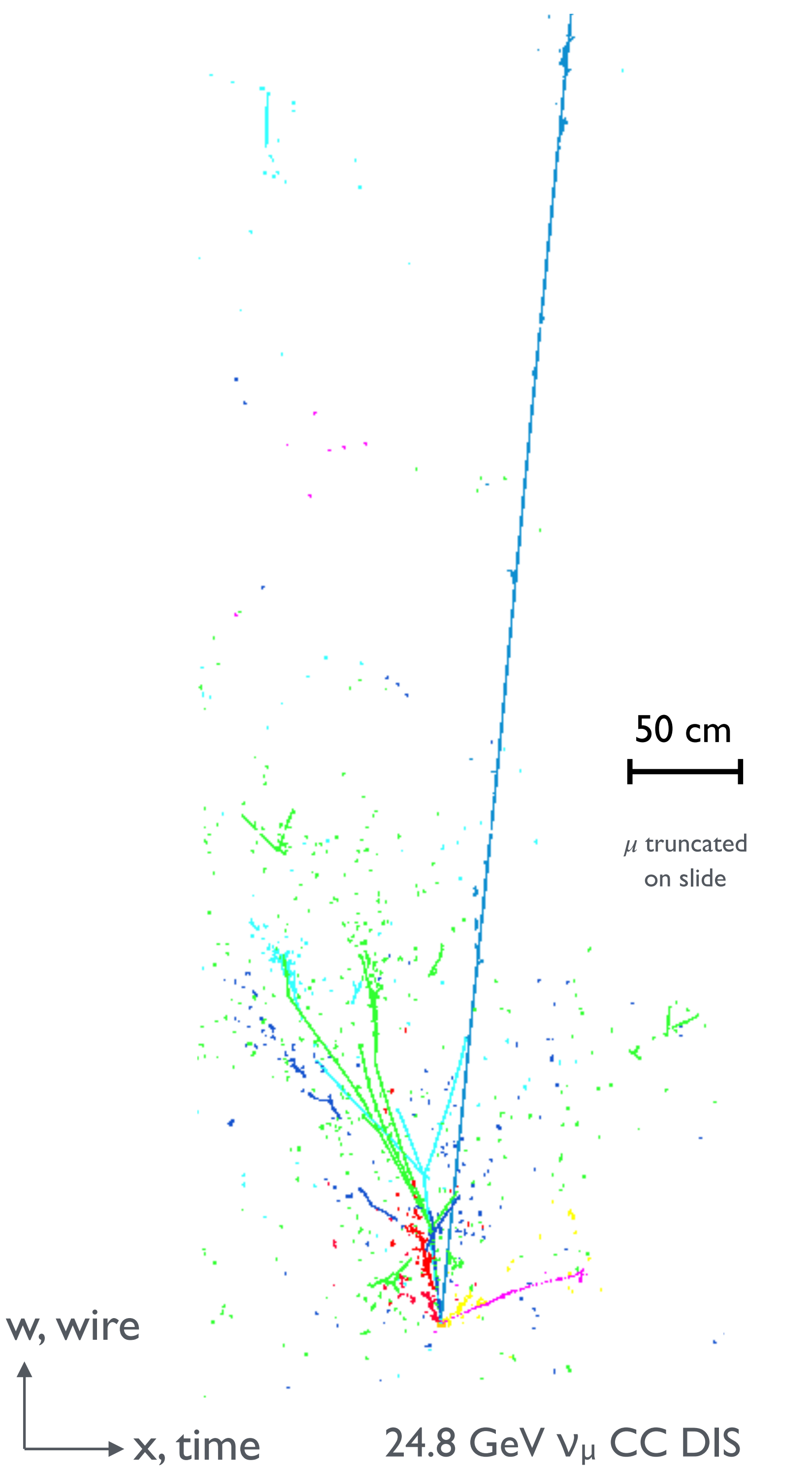
3D truth information

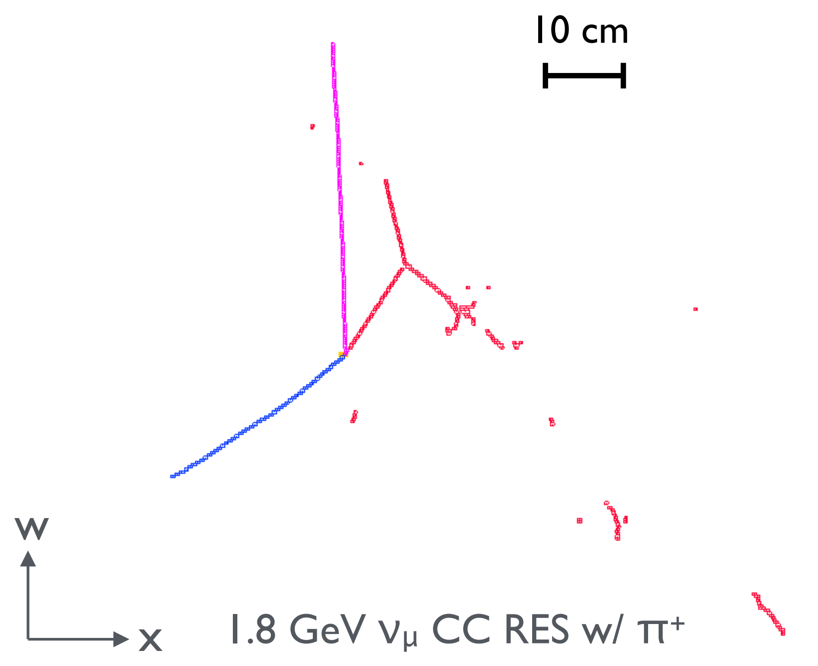
← Three 2D representations with common x coordinate, derived from drift time

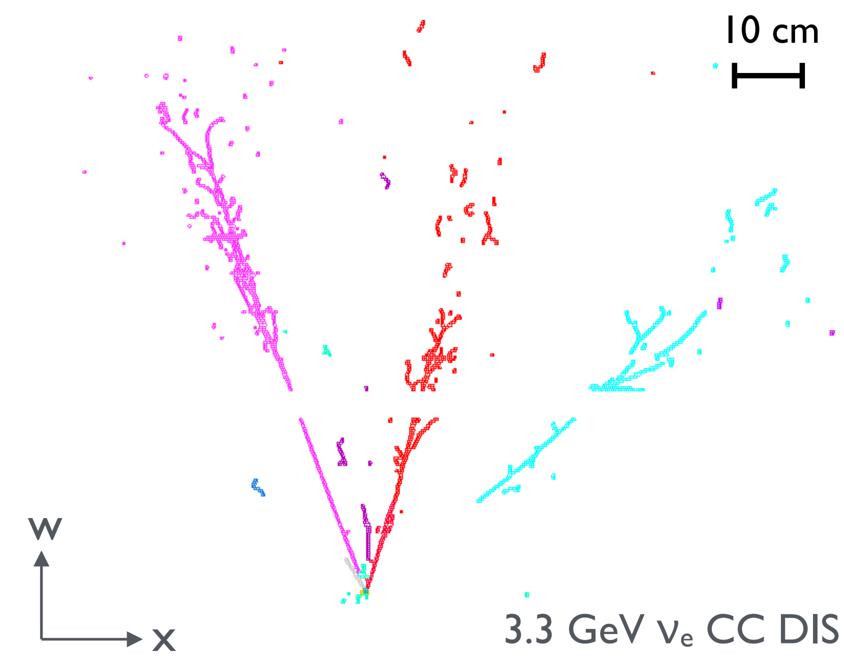
Pattern-Recognition Challenges

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

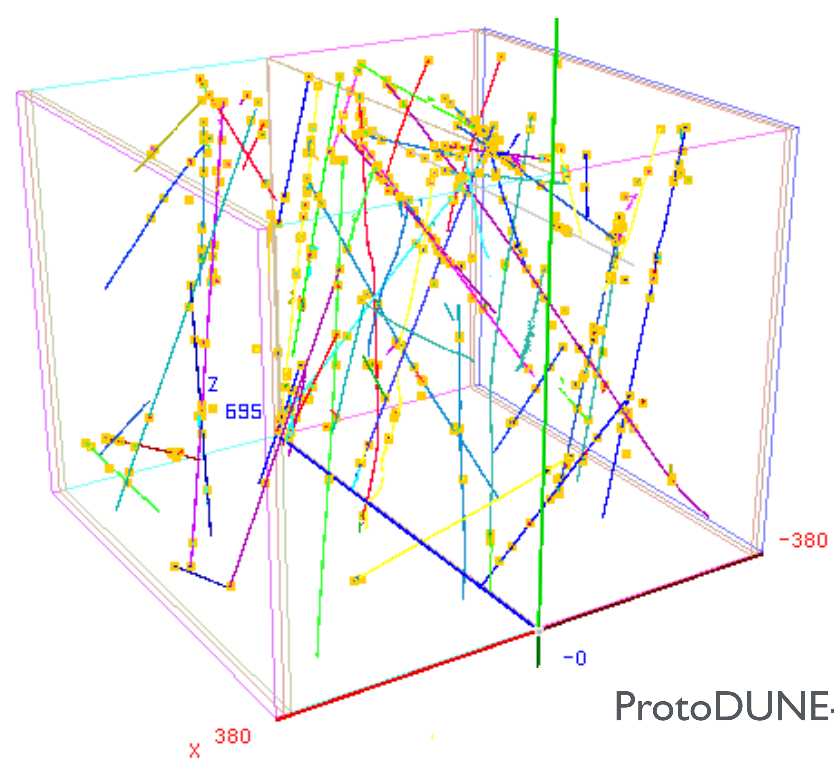
- Complex, diverse topologies:





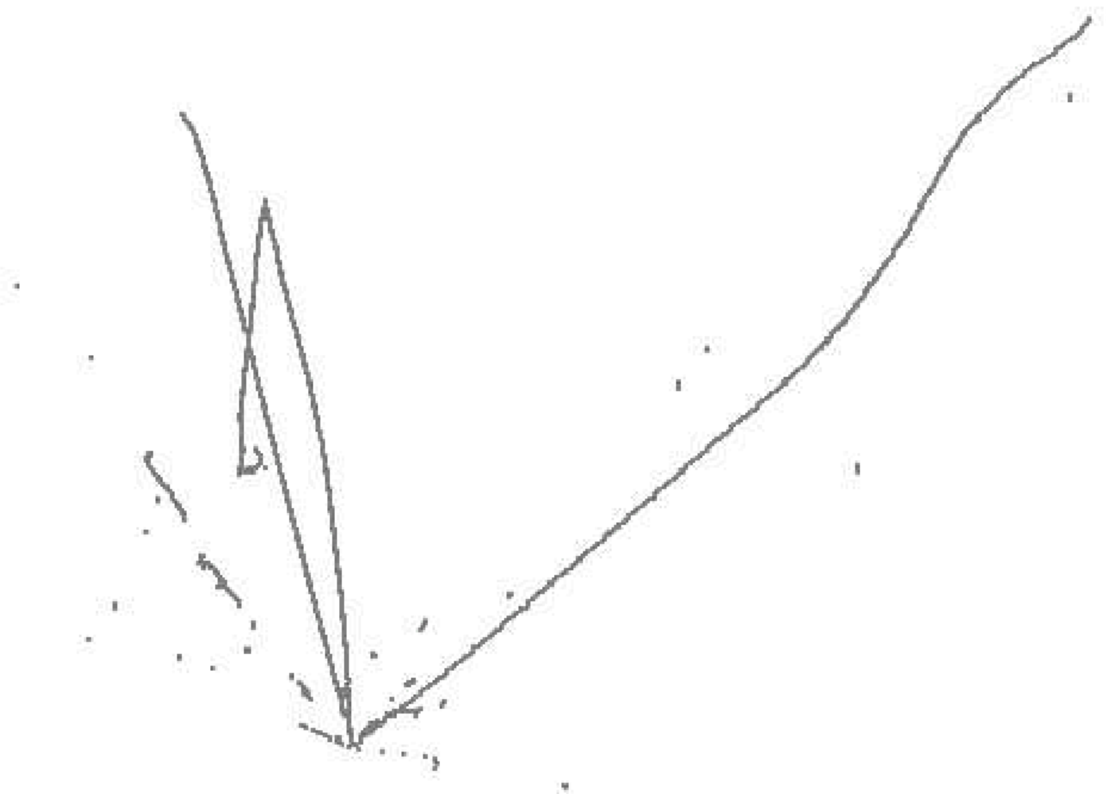


- Also, LArTPCs have long exposures, due to lengthy drift times (up to few ms).
- Significant cosmic-ray muon background in surface-based detectors.



ProtoDUNE-SP

Examples: Collection Plane Images



w, wire

x, time

2.8 GeV ν_μ CC DIS

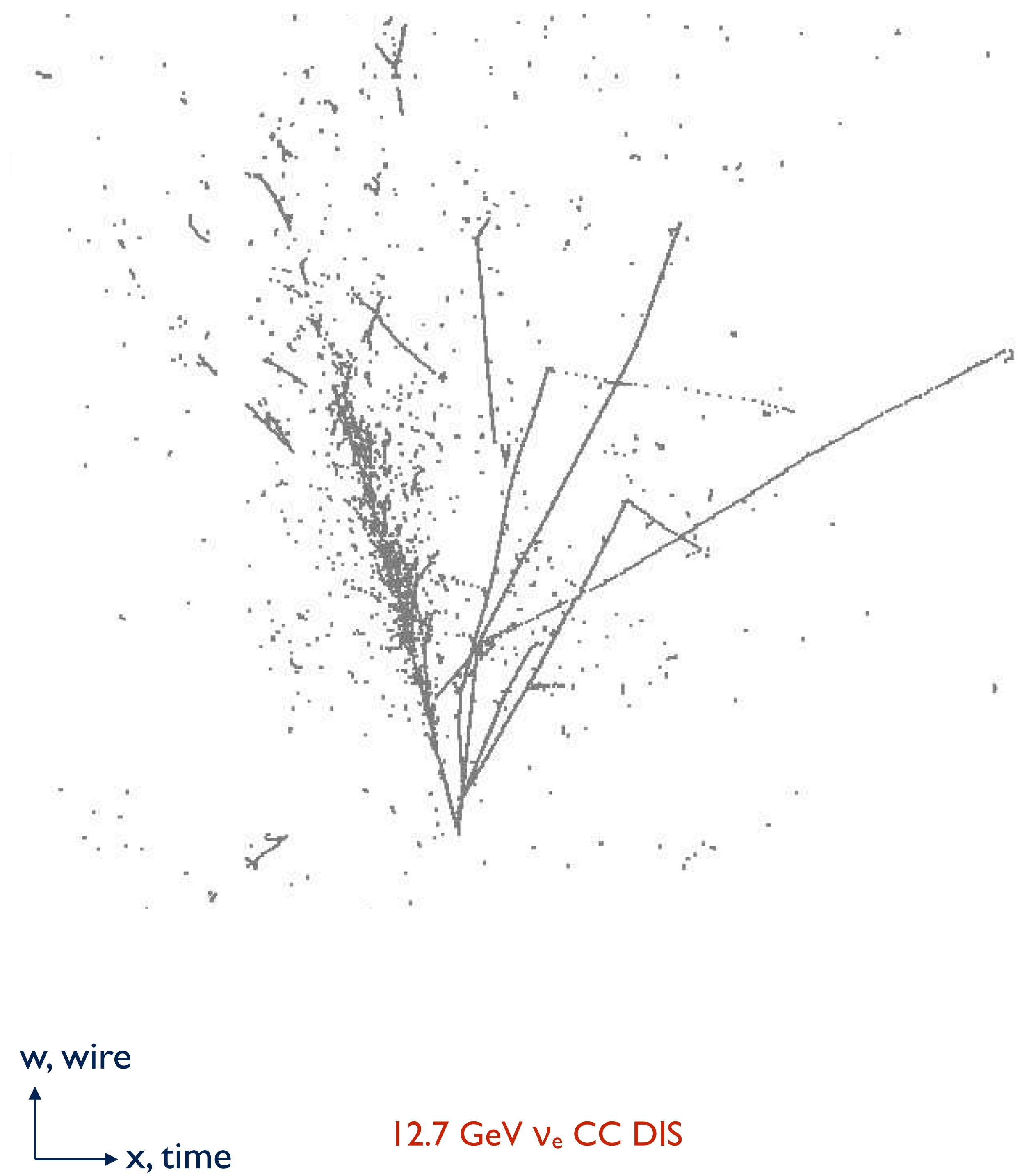
w, wire

x, time

3.2 GeV ν_e CC DIS

Examples: Collection Plane Images

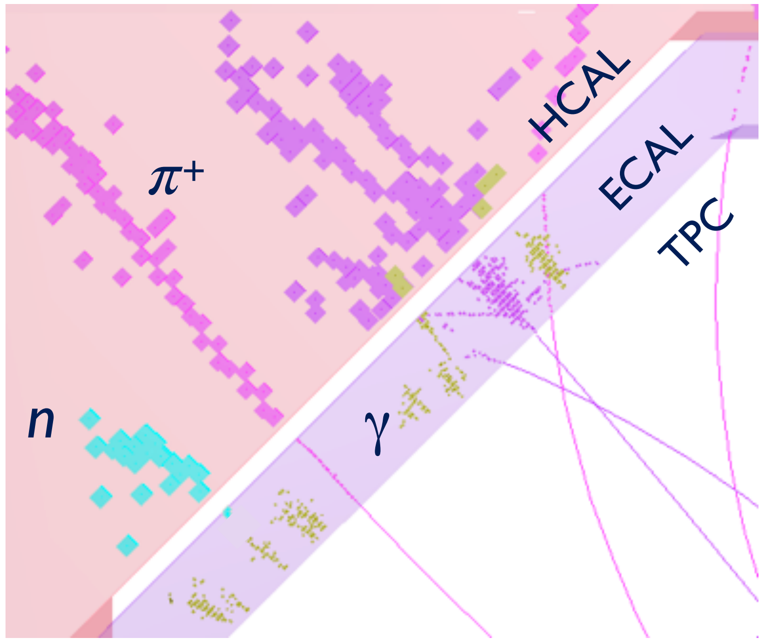
26.8 GeV ν_μ CC DIS



12.7 GeV ν_e CC DIS

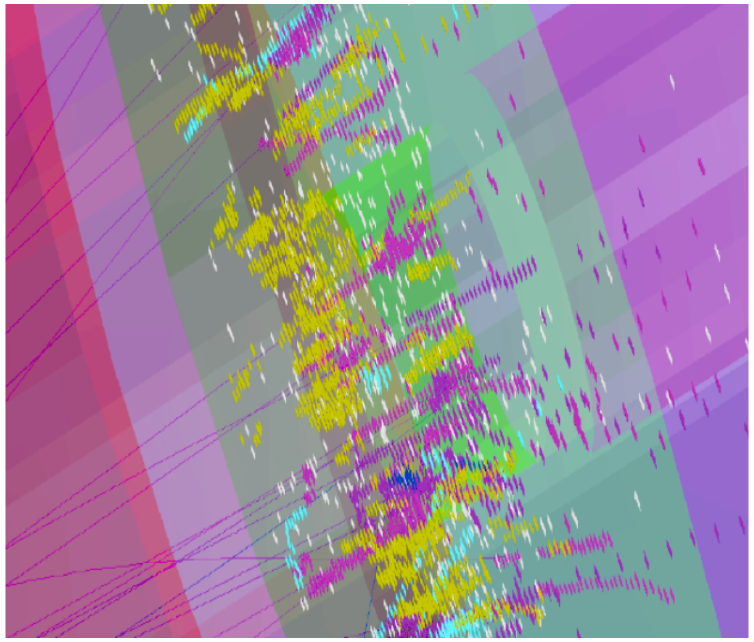
Pandora Multi-Algorithm Approach

- **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

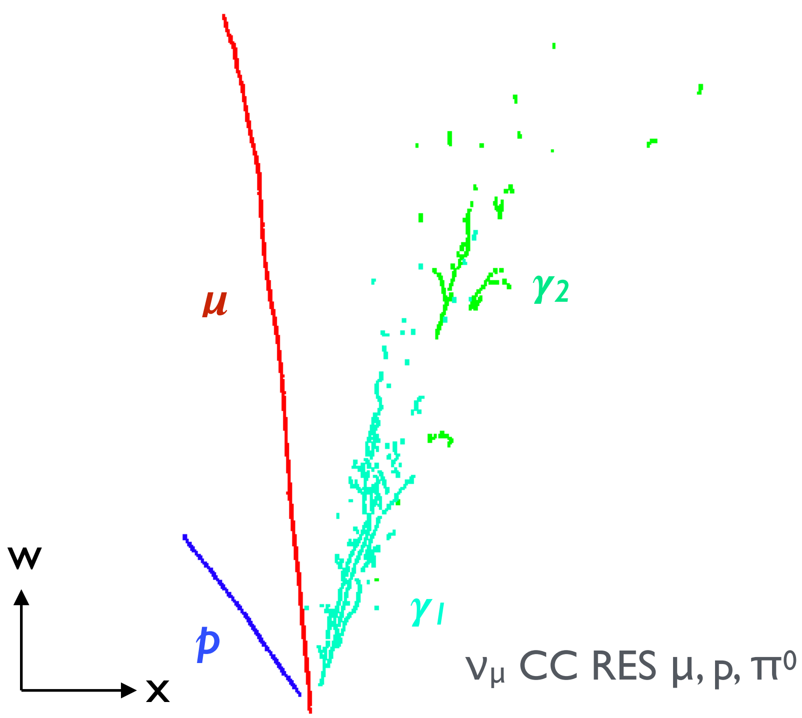
NIMA.2009.09.009 NIMA.2012.10.038



Typical showers in CMS HGCAL - 3D

LHCC-P-008

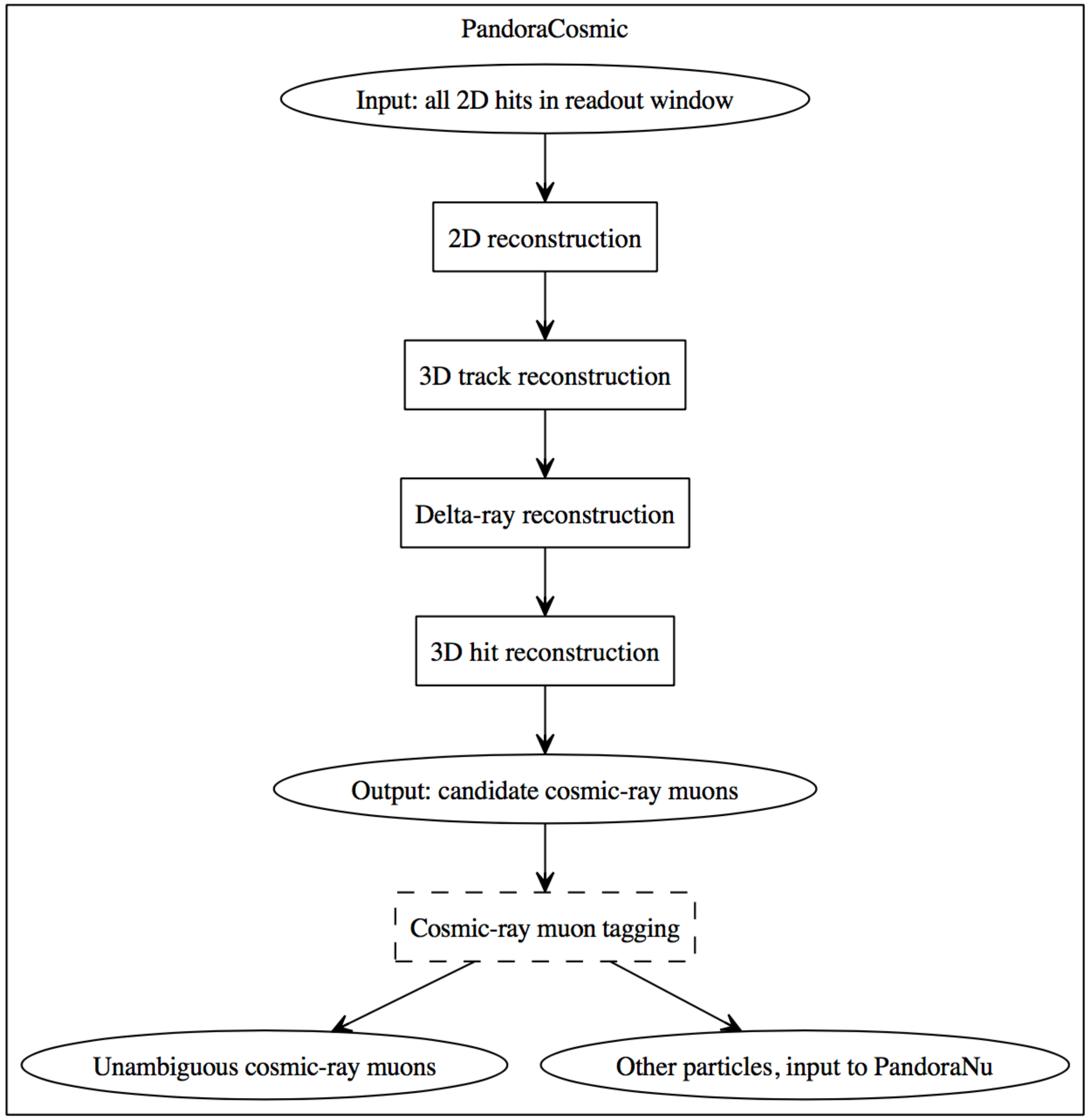
BNB interaction at MicroBooNE - 3 x 2D

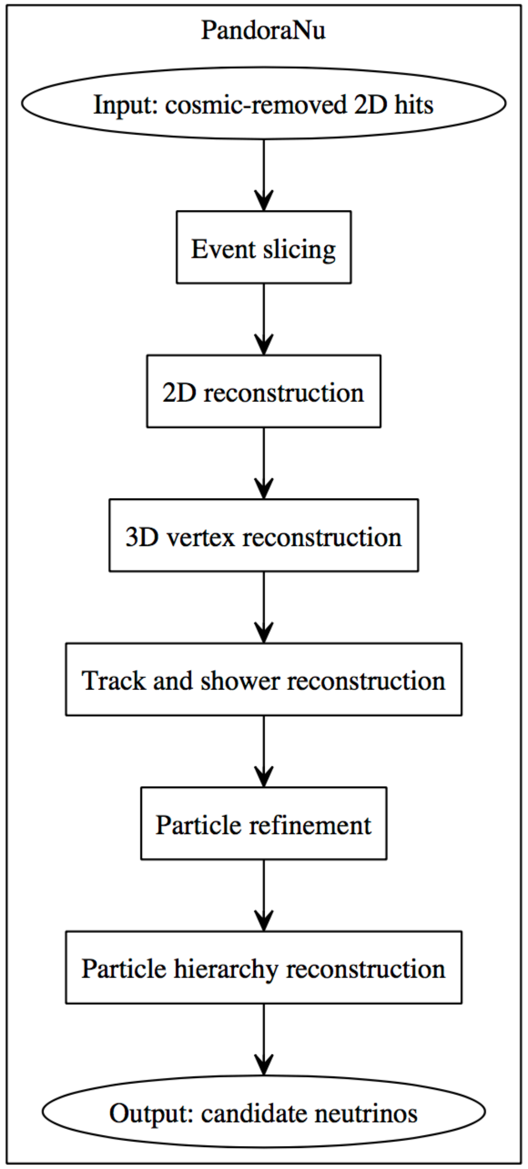


Pandora Algorithm Chains

- **Two algorithm chains (>140 algs) created for LArTPC use, with many algs in common:**
  - **PandoraCosmic:** strongly track-oriented; showers assumed to be delta rays, added as daughters of primary muons; muon vertices at track high-y coordinate.
  - **PandoraNu:** finds neutrino interaction vertex and protects all particles emerging from vertex position. Careful treatment to address track/shower tensions.

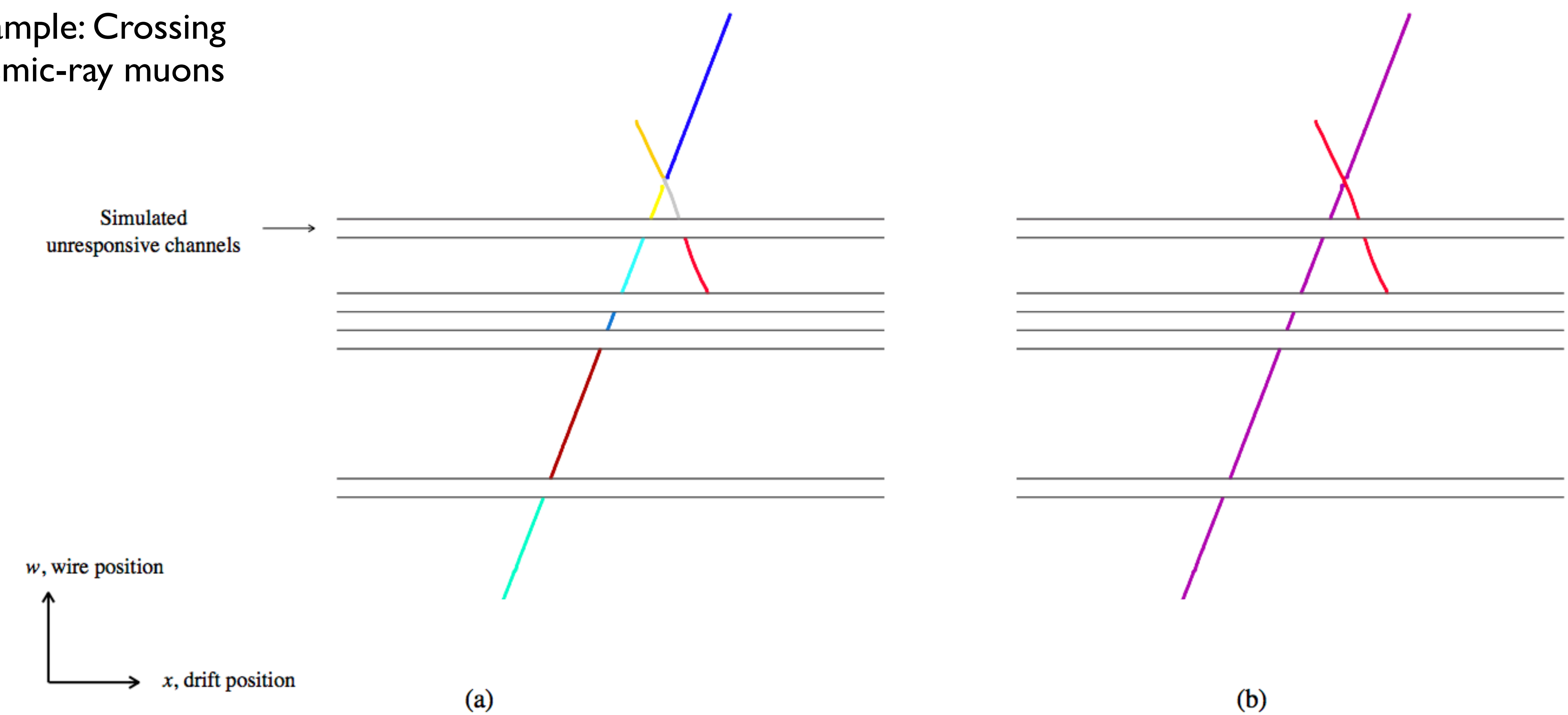
Initially use a two-pass approach:
Input to PandoraNu excludes hits from unambiguous cosmic rays.





- 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 16 cluster-merging and cluster-splitting algs that use topological info.

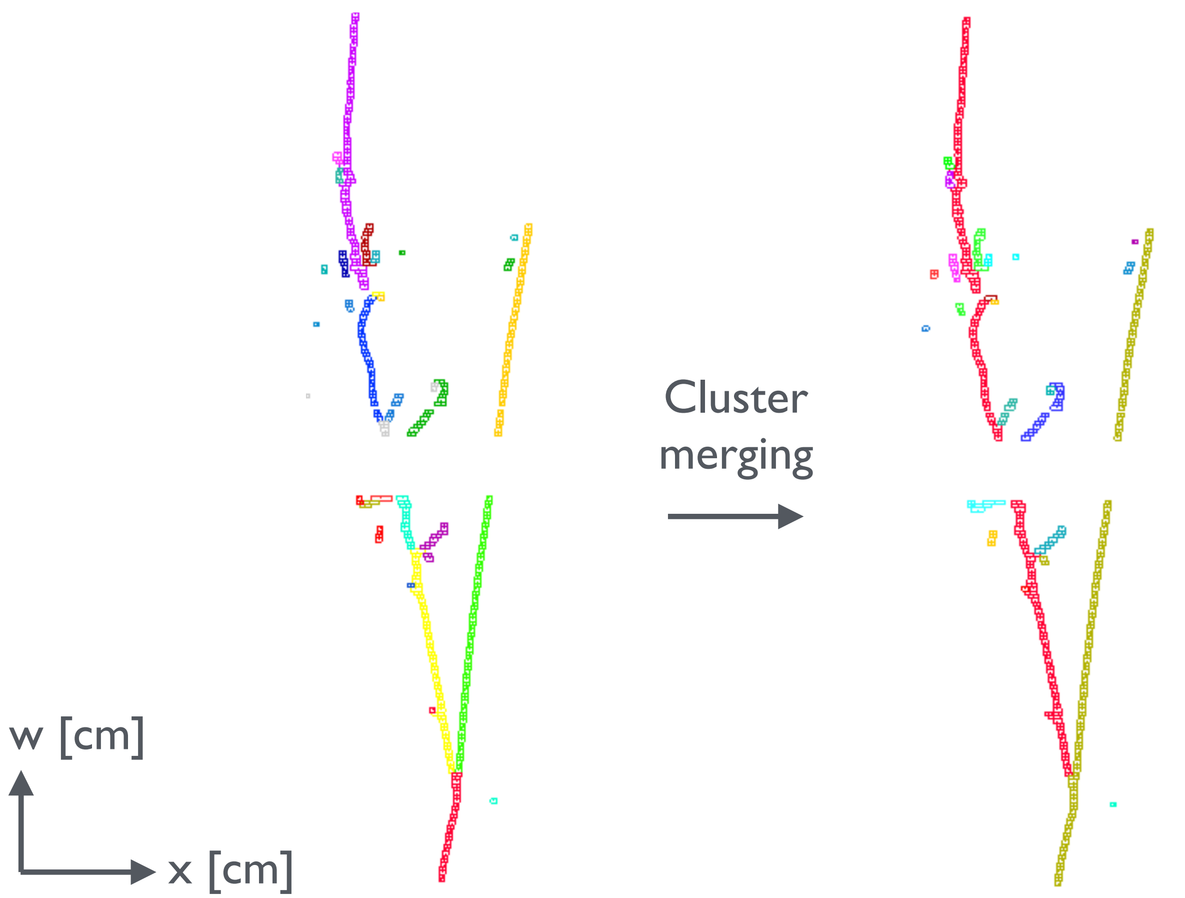
Example: Crossing cosmic-ray muons



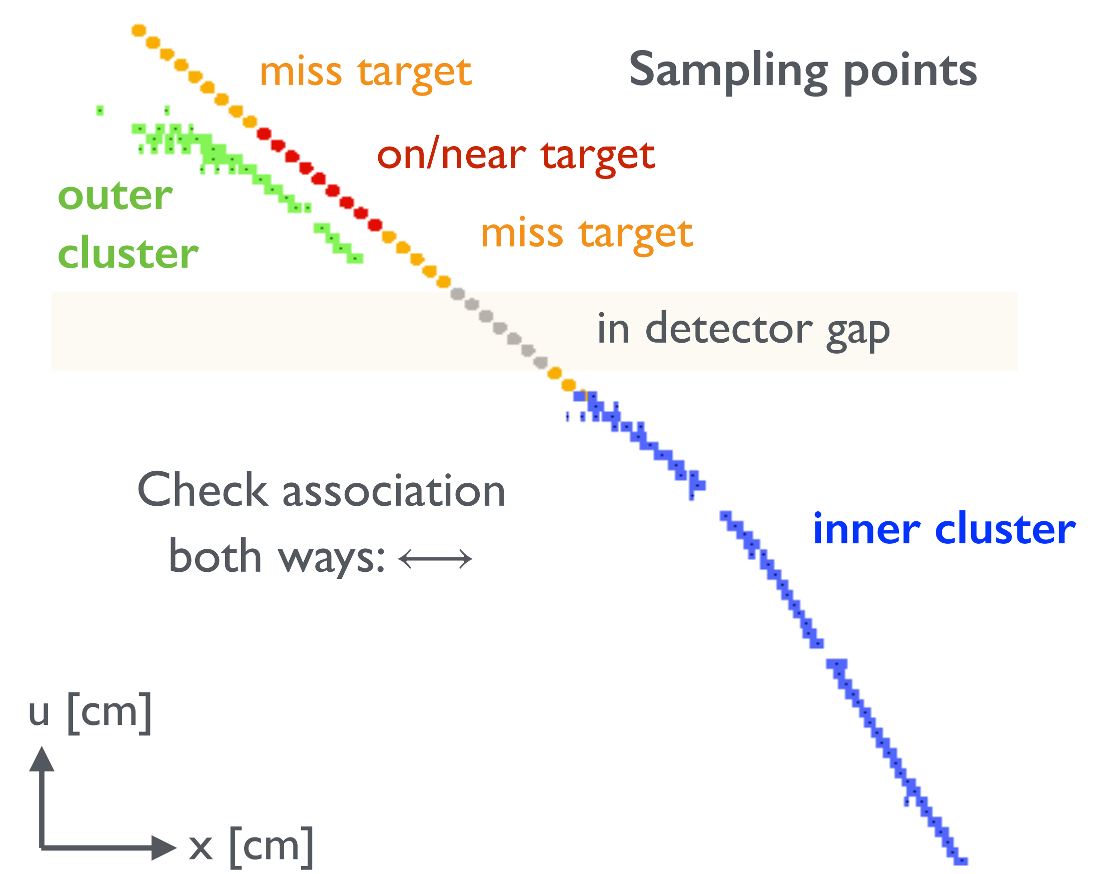
Topological Association - 2D

- **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.
  - Typically need to provide a definition of association (for a given pair of clusters), then navigate forwards and backwards to identify chains of associated clusters that can be safely merged.

E.g. LongitudinalAssociation

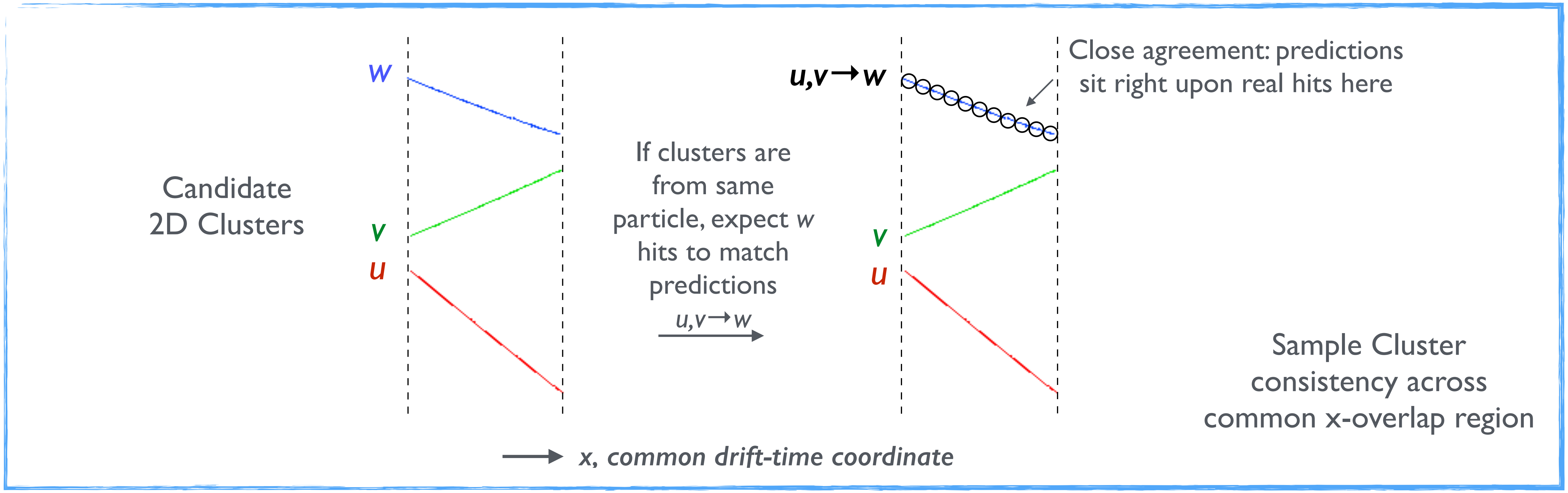


E.g. CrossGapsAssociation



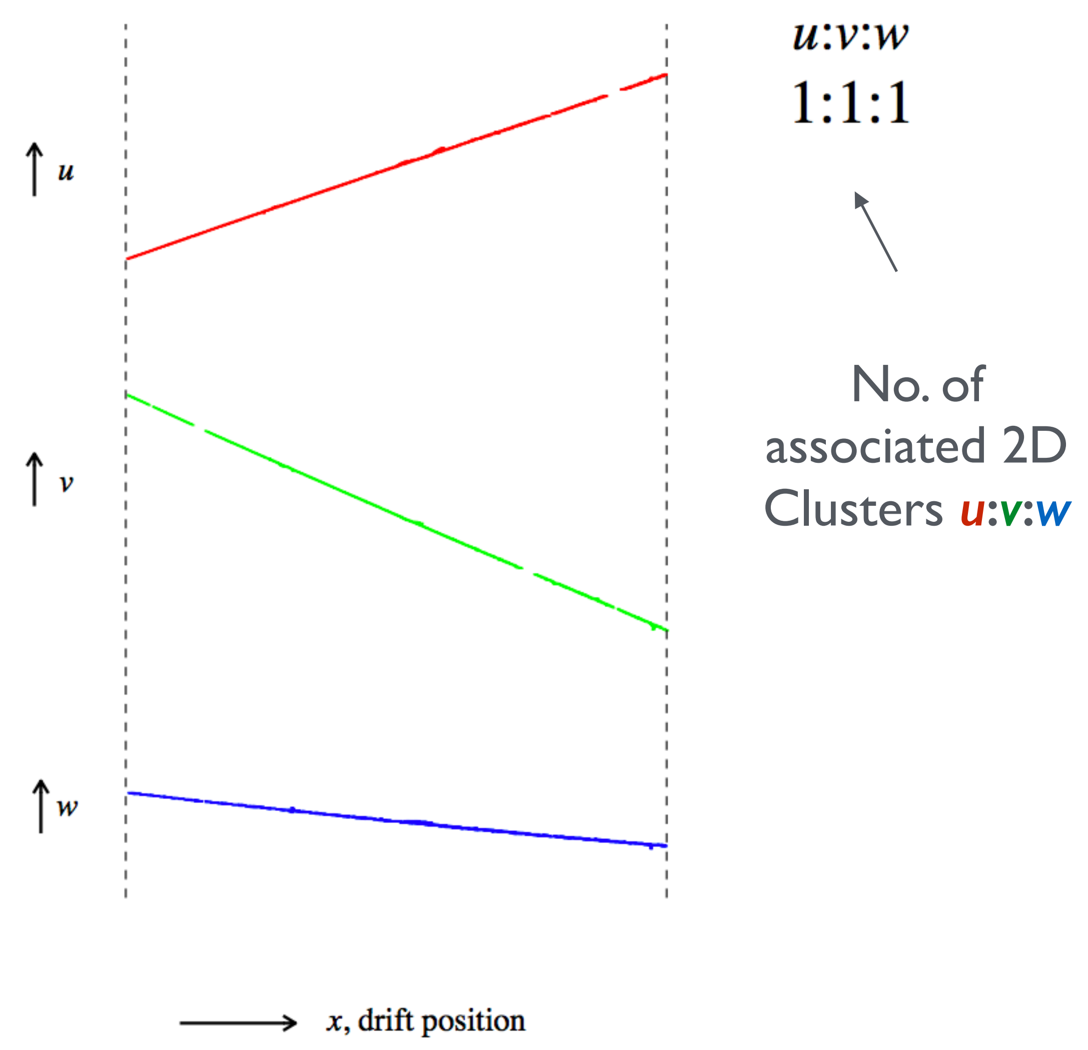
Track Pattern Recognition - 3D

- **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.
  - At given *x*, compare predictions {*u*,*v*→*w*; *v*,*w*→*u*; *w*,*u*→*v*} with cluster positions, calculating χ^2



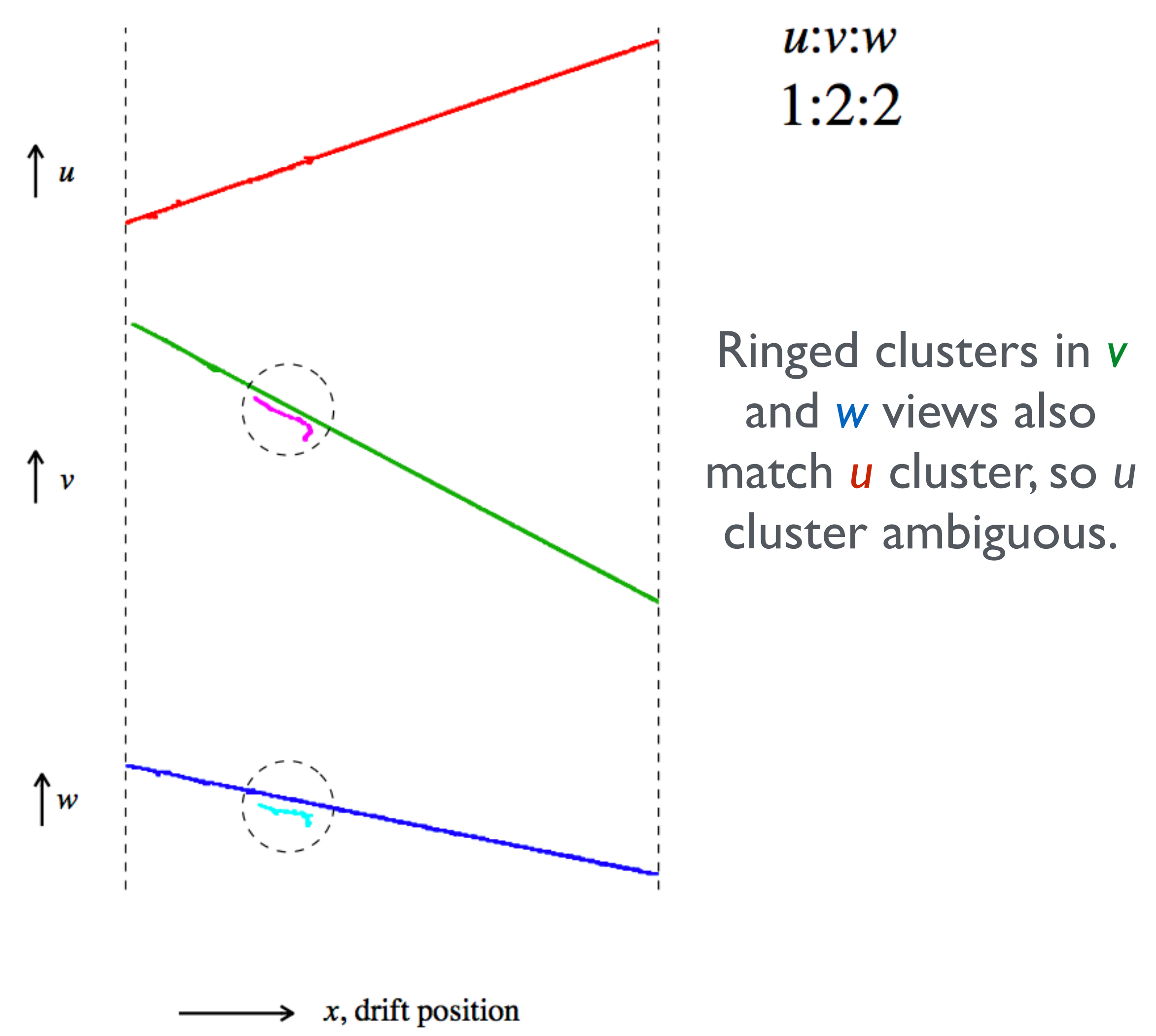
Store all results in a 3D array, recording *x*-overlap span, no. of sampling points, no. of "matched" sampling points and a χ^2 - documents all 2D cluster-matching ambiguities.

3D array stores overlap details for all trios of 2D clusters. Tools make 2D reco changes to resolve any ambiguities. If a tool makes a change (e.g. splits a cluster), all tools run again.



Find unambiguous elements, demanding that the common x-overlap is 90% of the x-span for all three clusters.

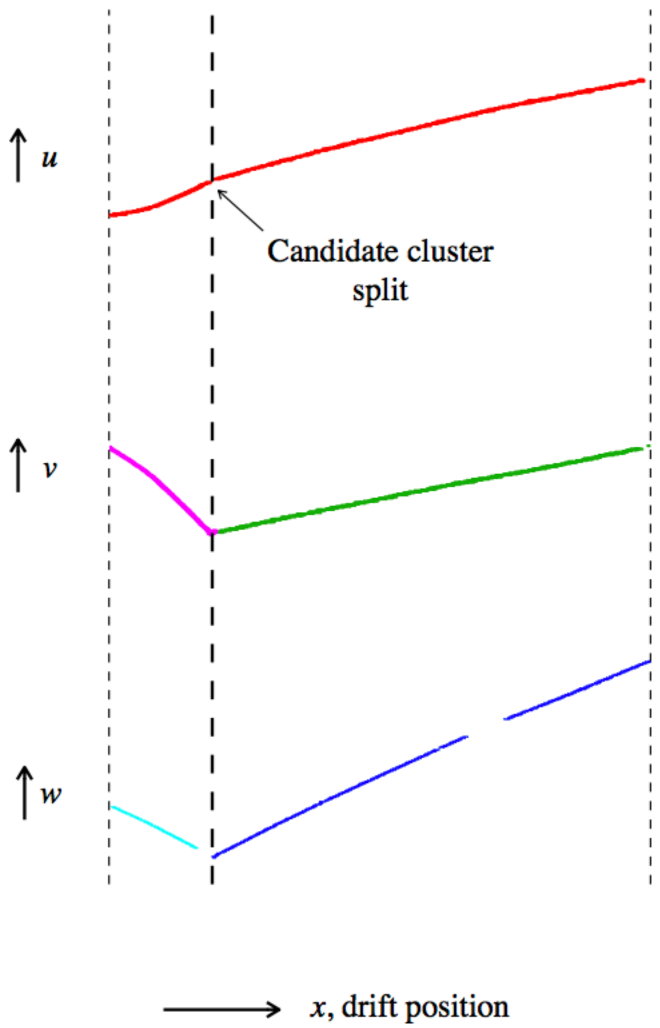
Clear Tracks Tool



Resolve **obvious** ambiguities: clusters are matched in multiple configurations, but one element is much "better" than others.

Long Tracks Tool

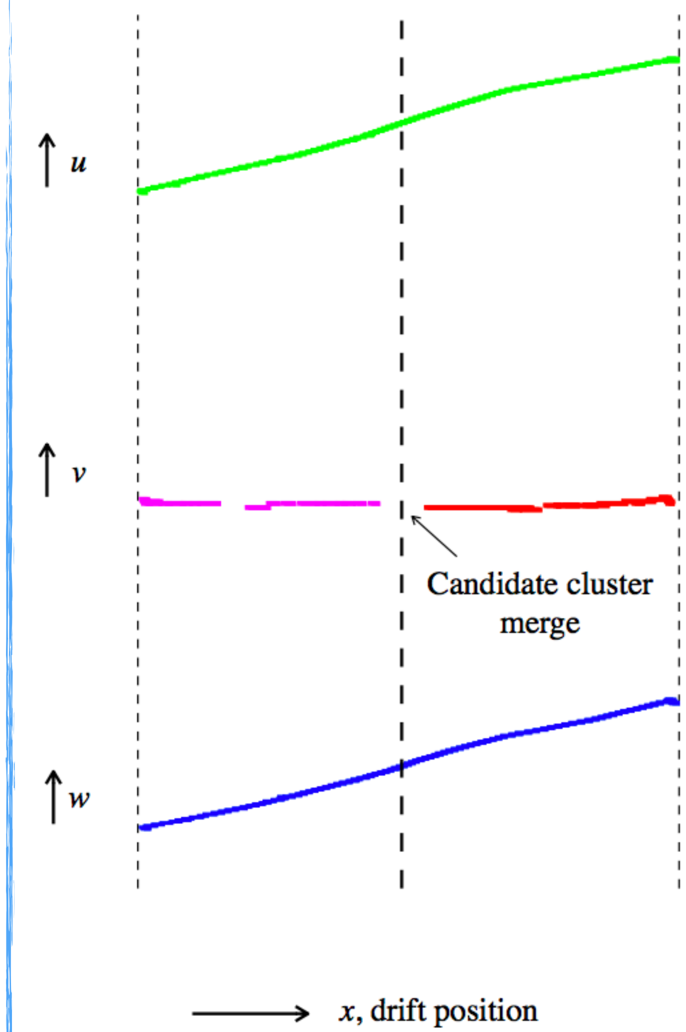
Track Pattern Recognition - 3D



u:*v*:*w*
1:2:2

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

Overshoot Tracks Tool



u:*v*:*w*
1:2:1

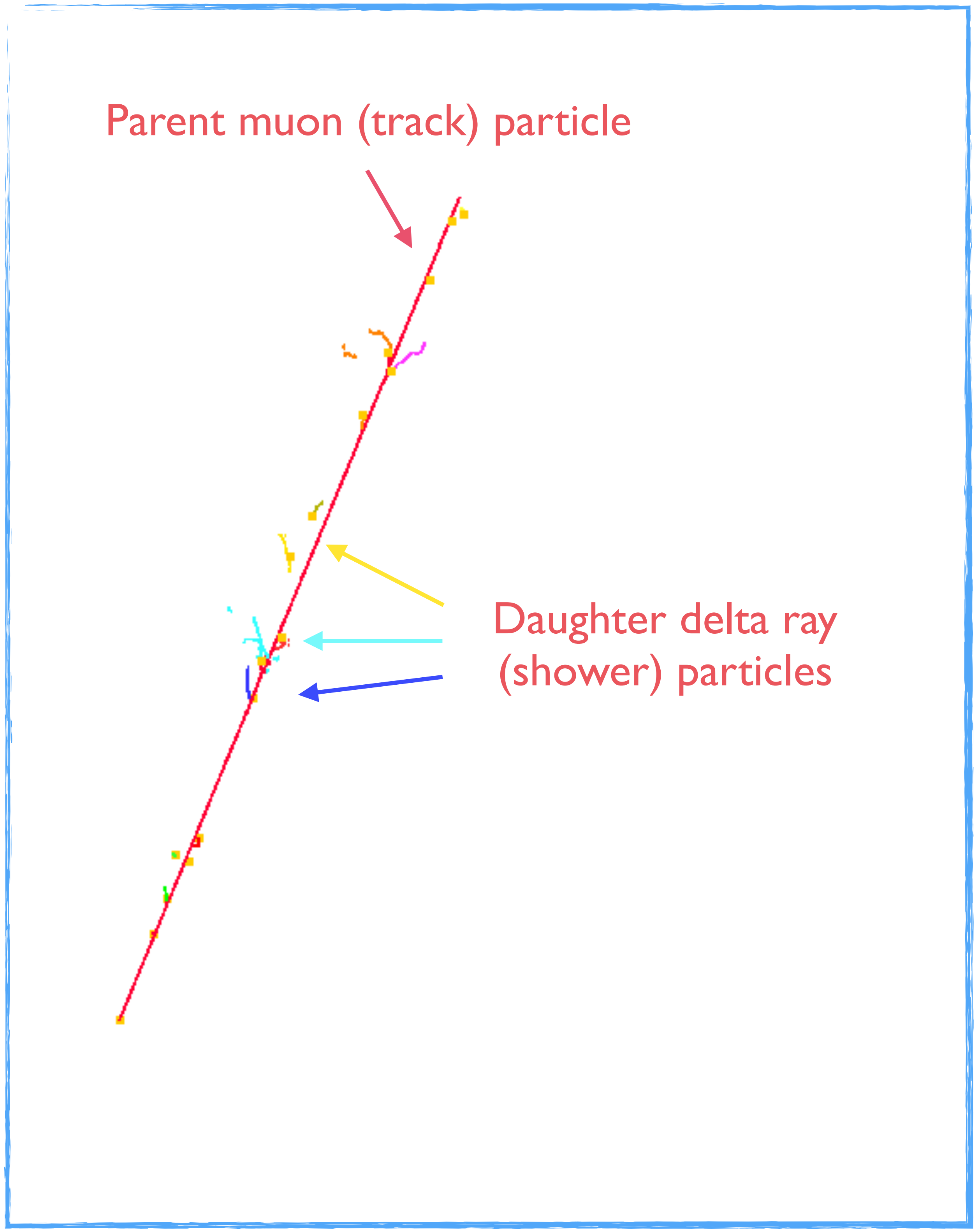
Two clusters in *v* view, matched to common clusters in *u* and *w* views.
Merge *v* clusters.

Undershoot Tracks Tool

- Use all connected clusters to assess whether this is a true 3D kink topology.
- Modify 2D clusters as appropriate (i.e. merge or split) and update cluster-matching details.
- Initial ClearTracks tool then able to identify unambiguous groupings of clusters and form particles.

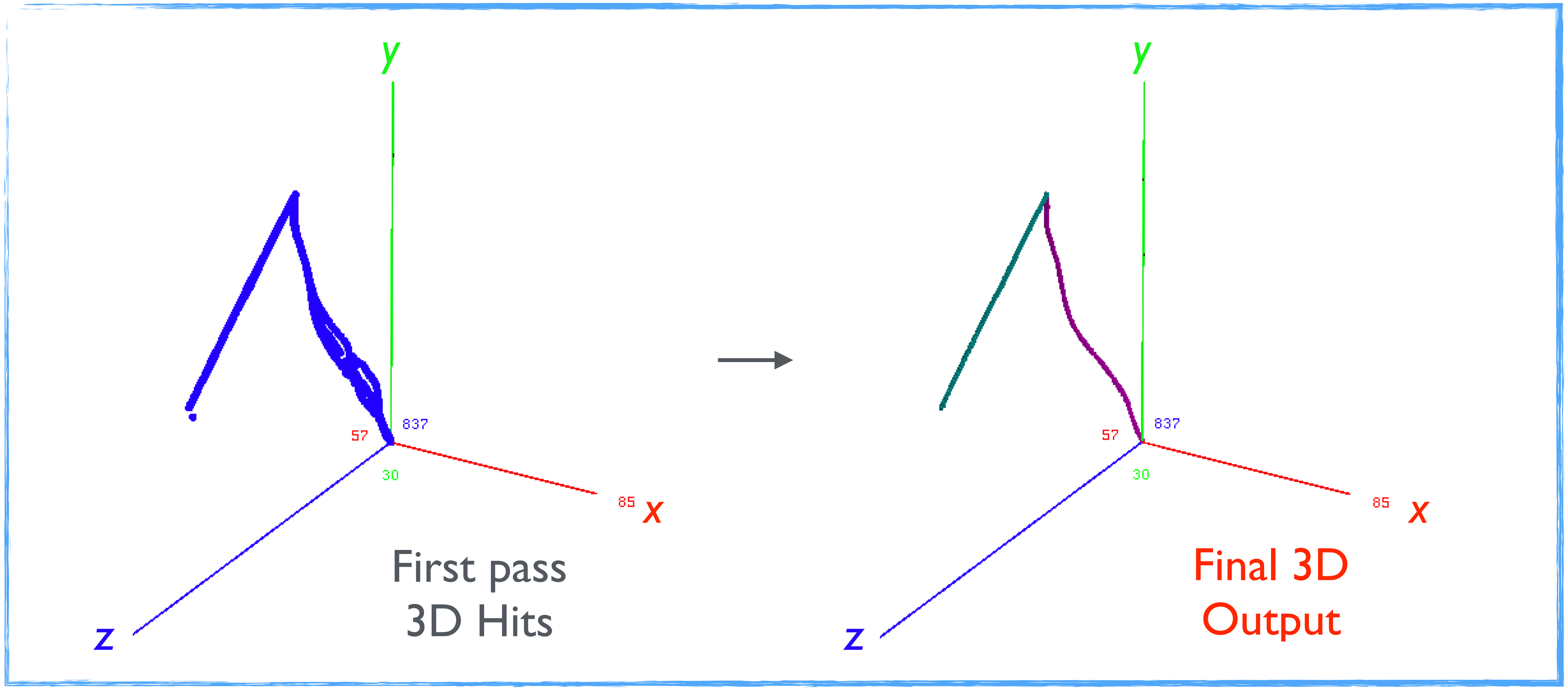
Delta-Ray Reconstruction - 2D, 3D

- **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.

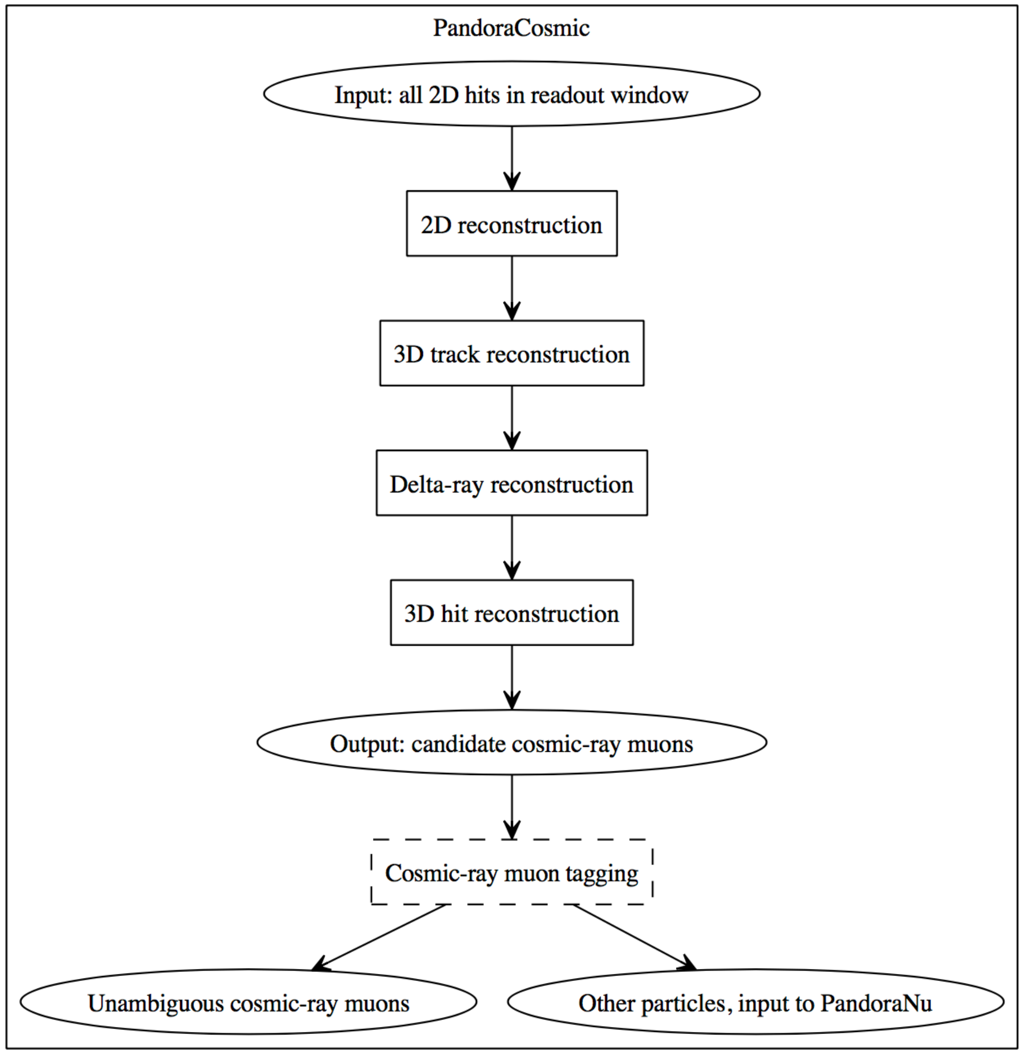


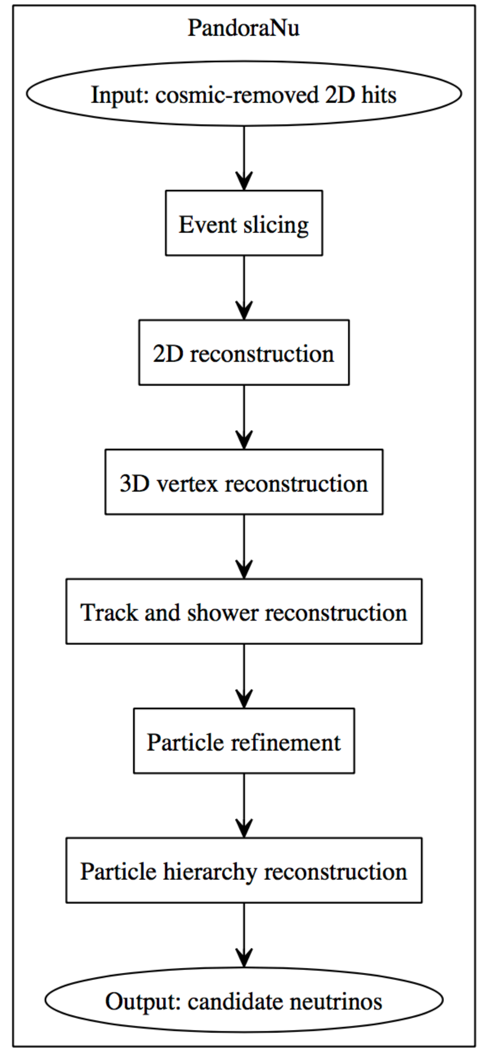
3D Hit/Cluster Reconstruction

- 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



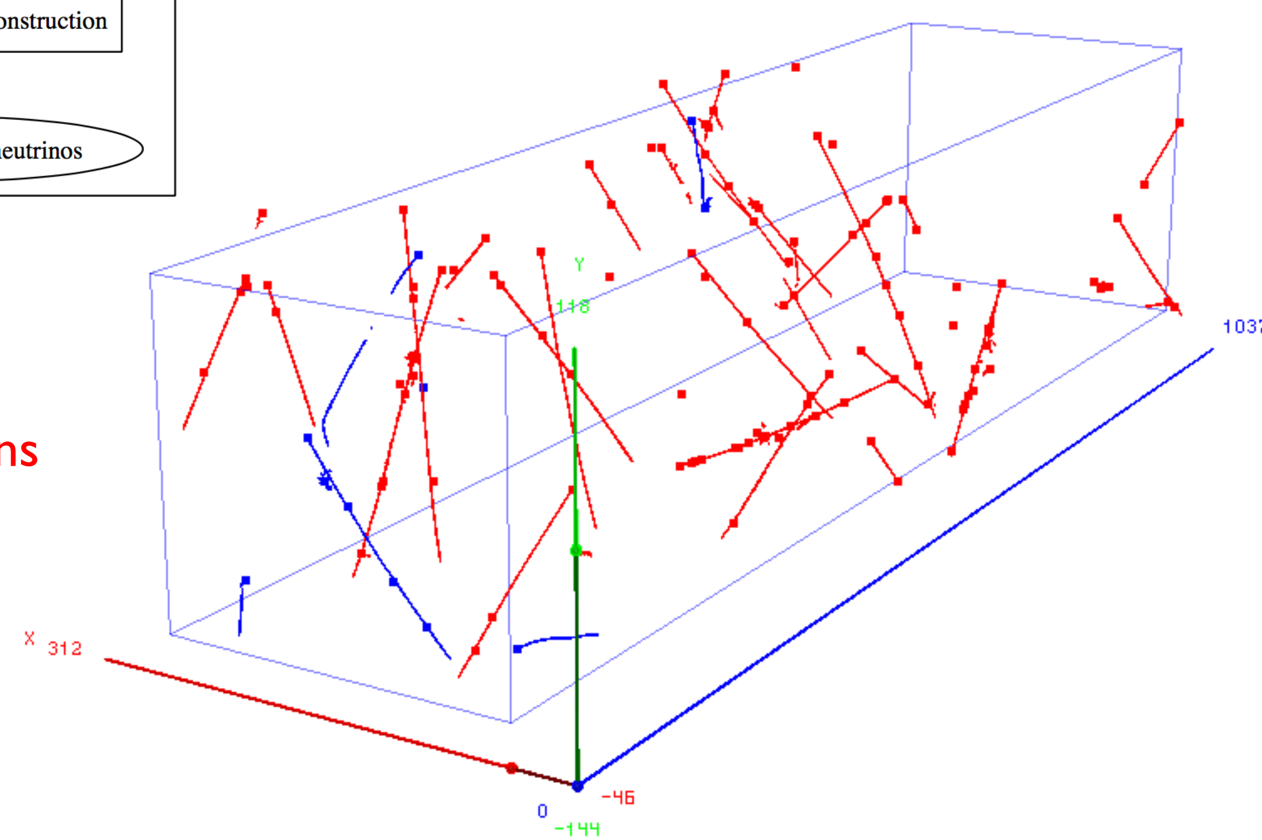
PandoraCosmic → PandoraNu





Other particles, input to PandoraNu

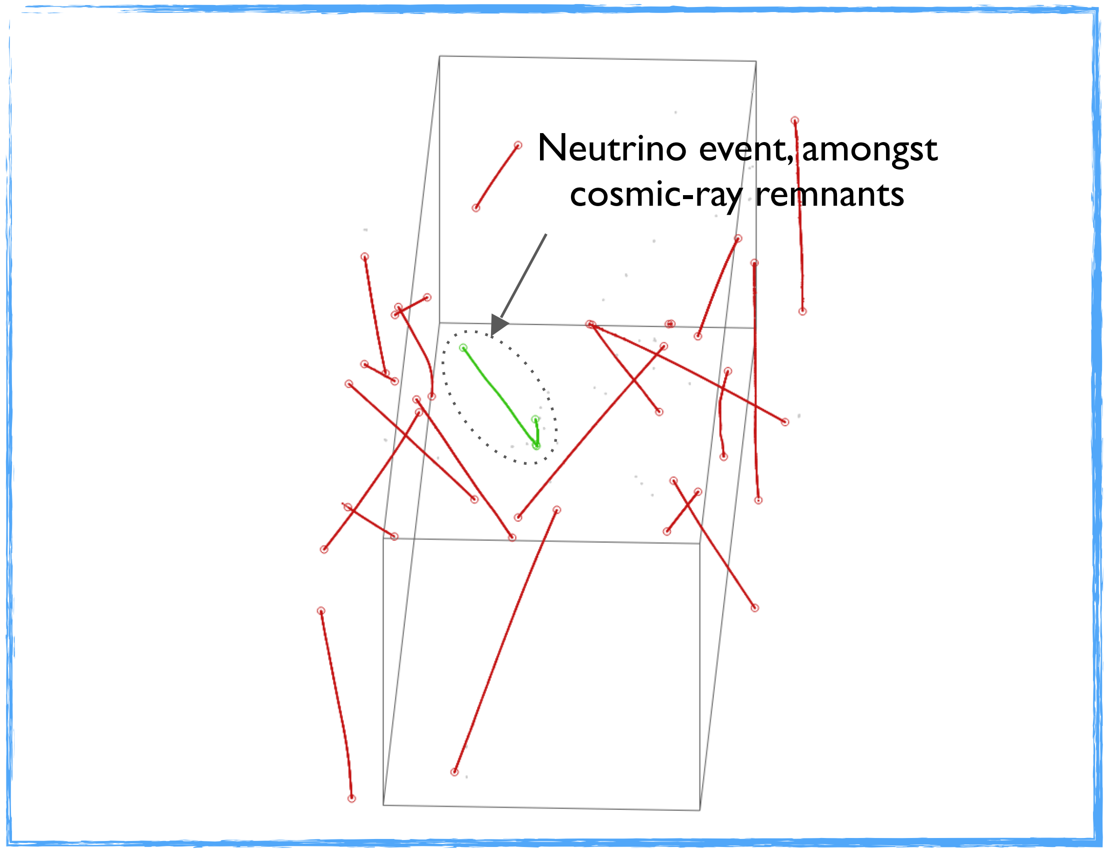
Unambiguous cosmic-ray muons

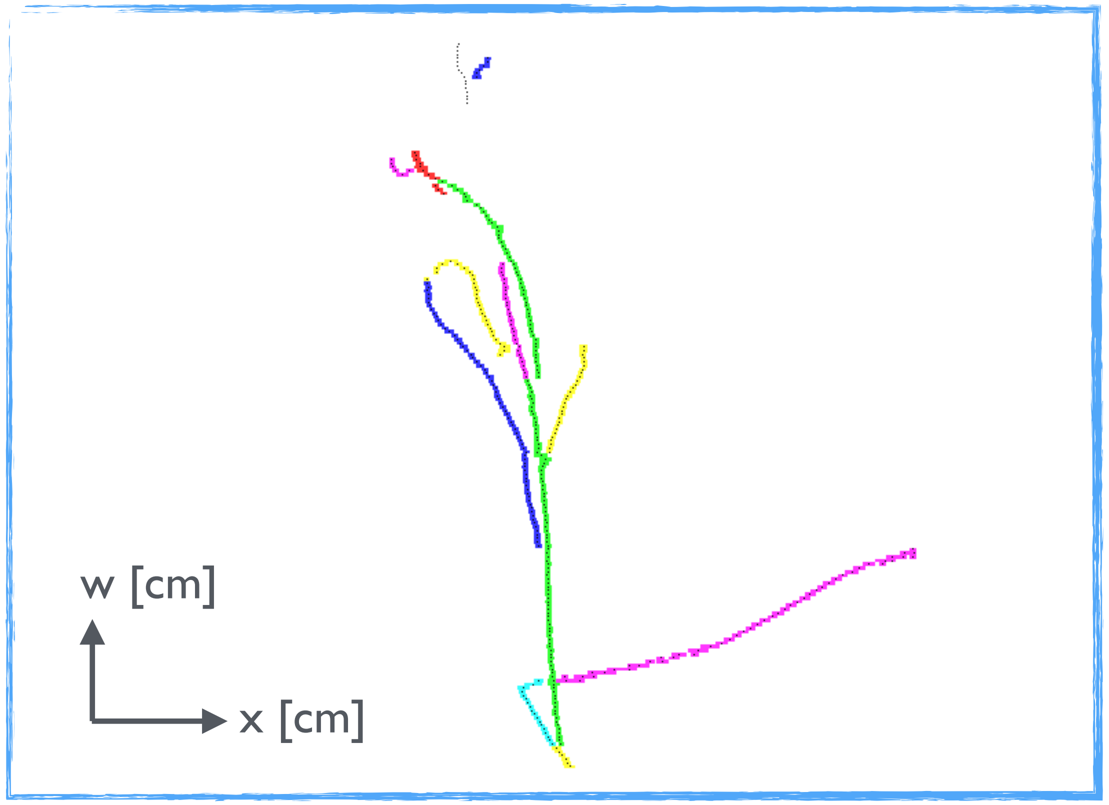


Neutrino Reconstruction

- **Must be able to deal with presence of any cosmic-ray muon remnants.**
  - Run fast version of reconstruction, up to 3D hit creation
  - "Slice" 3D hits into separate interactions, processing each slice in isolation.
  - Each slice ⇒ candidate neutrino particle.

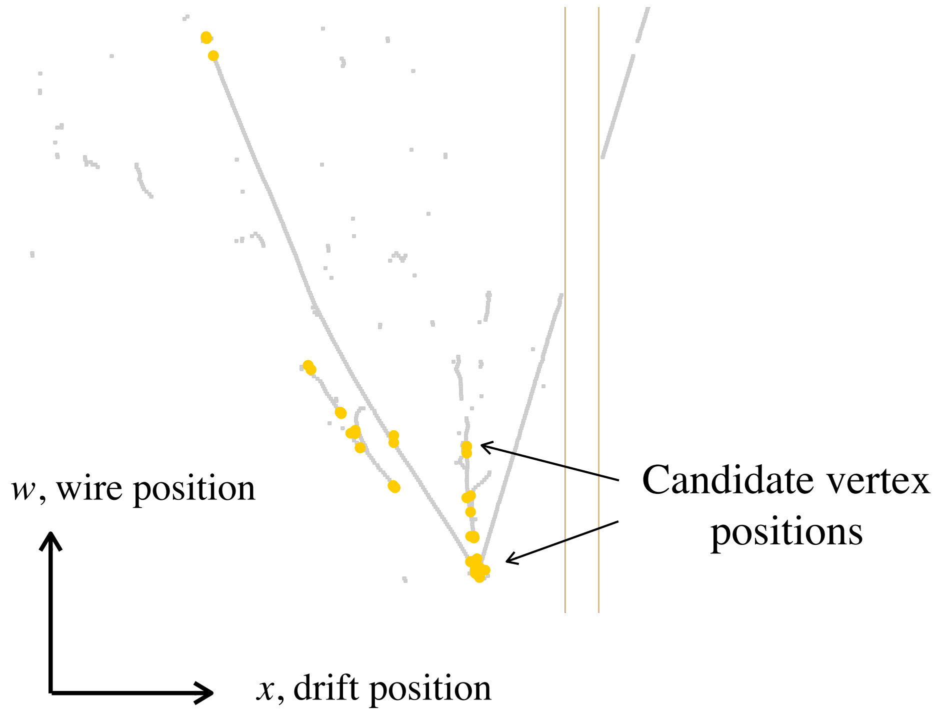
- **Neutrino pass reuses track-oriented clustering and topological association.**
  - Topological association algs must handle rather more complex topologies.
  - Specific effort to reconstruct neutrino interaction vertex.
  - More sophisticated efforts to reconstruct showers.

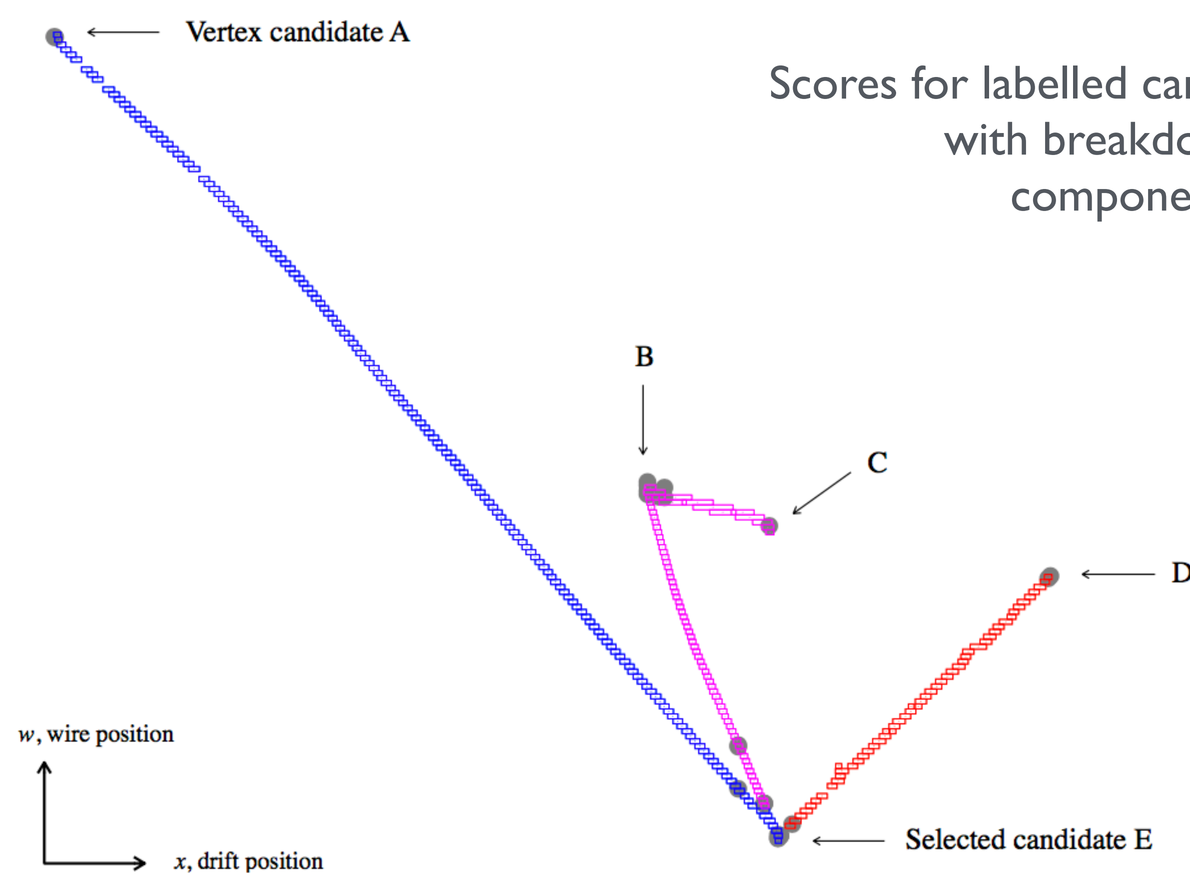




Search for neutrino interaction vertex:

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.
   - Selection uses Boosted Decision Trees and a Convolutional Neural Network.





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

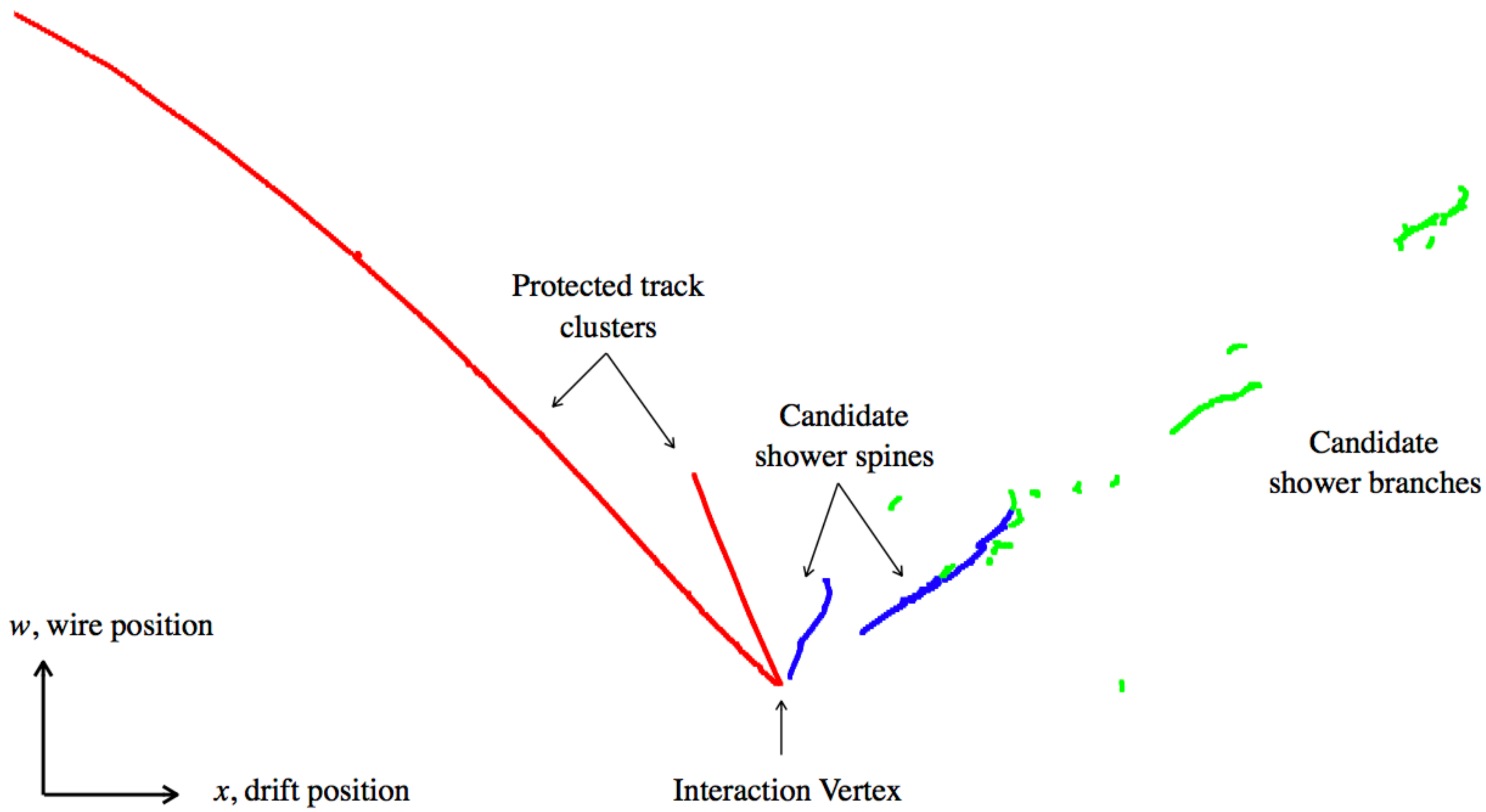
Downstream usage:

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

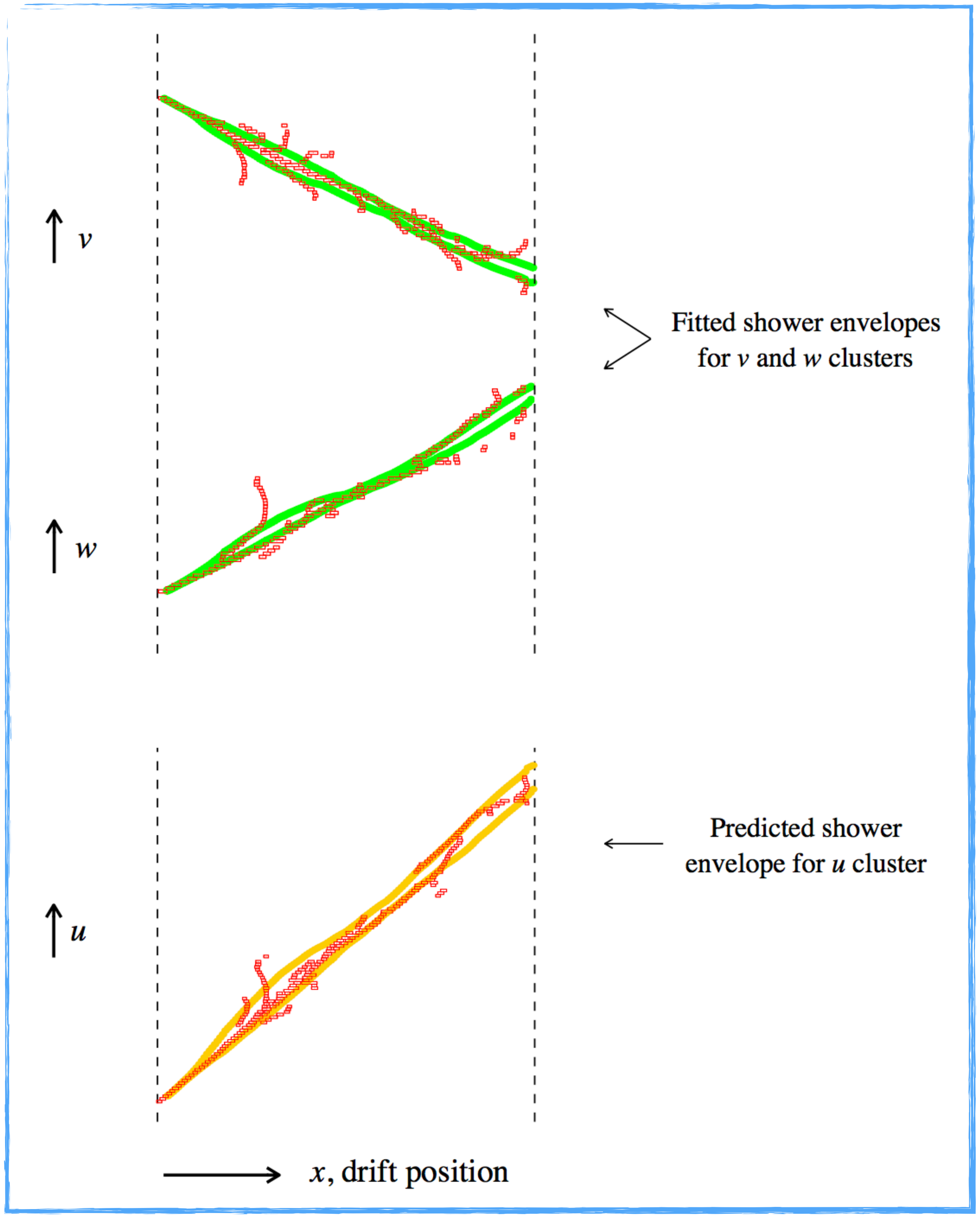
Shower Reconstruction - 2D

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.



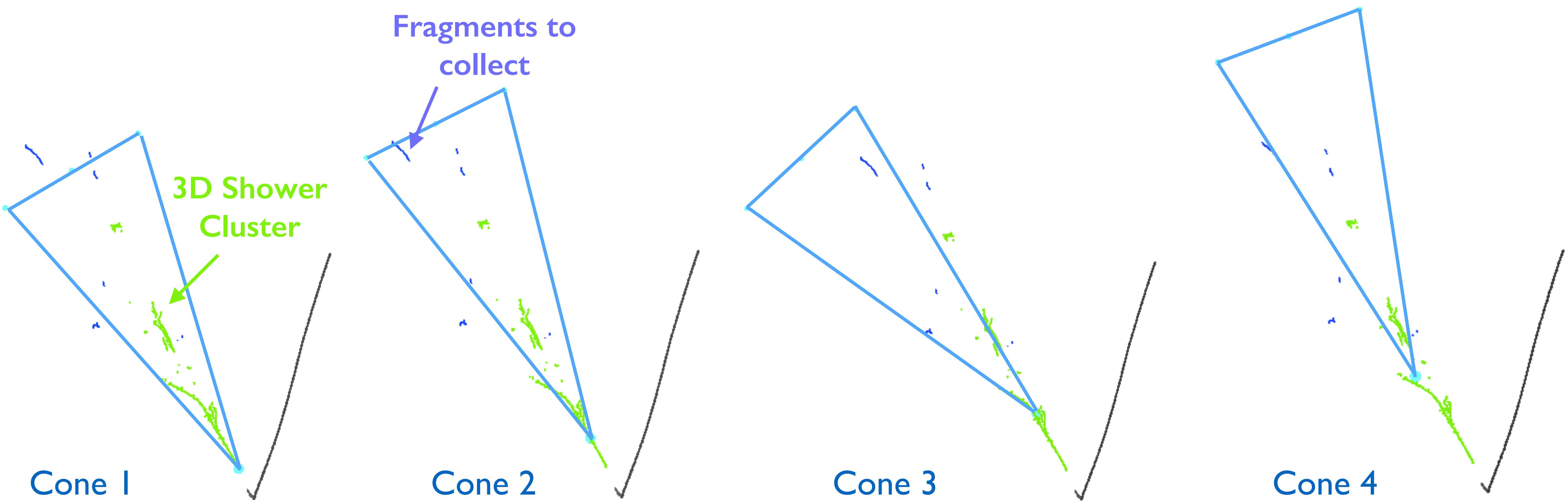
- **Reuse ideas from track reco to match 2D shower clusters between views:**
  - Build a 3D array to store cluster overlap and relationship information.
  - Overlap information collected by fitting shower envelope to each 2D cluster.
  - Shower edges from two clusters used to predict envelope for third cluster.



Particle Refinement - 2D, 3D

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

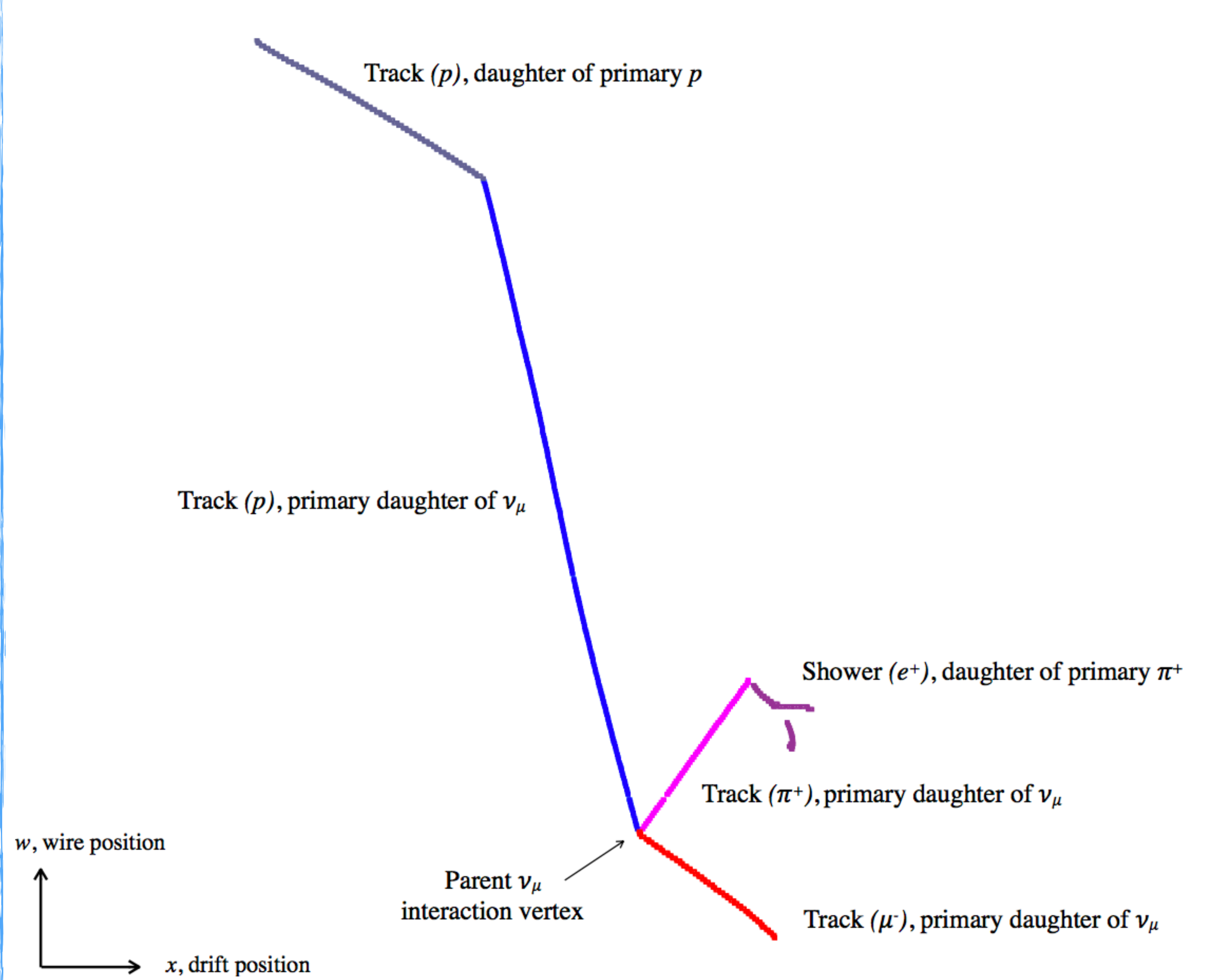
- Pick up small, unassociated clusters bounded by the 2D envelopes of shower-like particles.
- Use sliding linear fits to 3D shower clusters to define cones for merging small downstream shower particles, or picking up additional unassociated clusters.
- If anything left at end, dissolve clusters and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



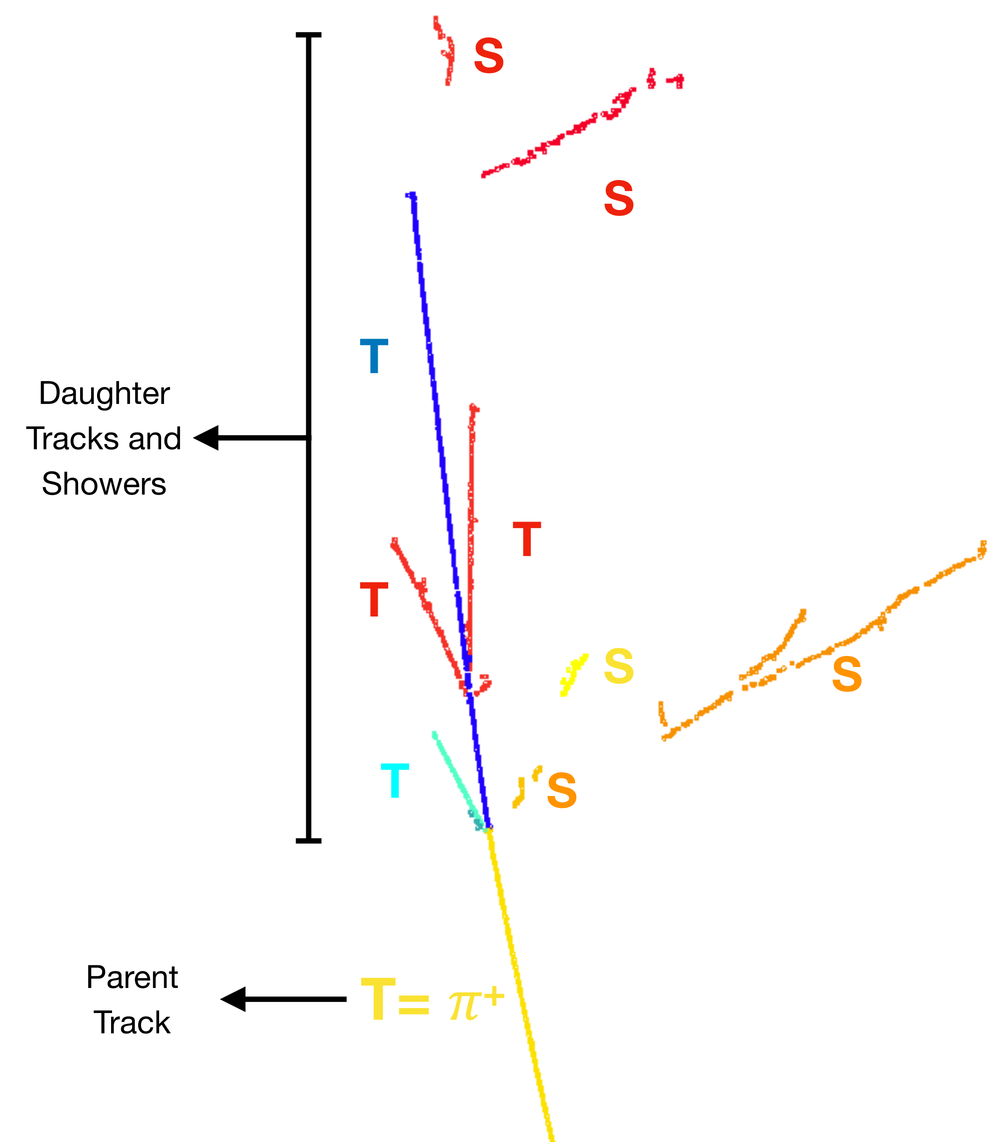
Particle Hierarchy Reconstruction - 3D

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

MicroBooNE simulation



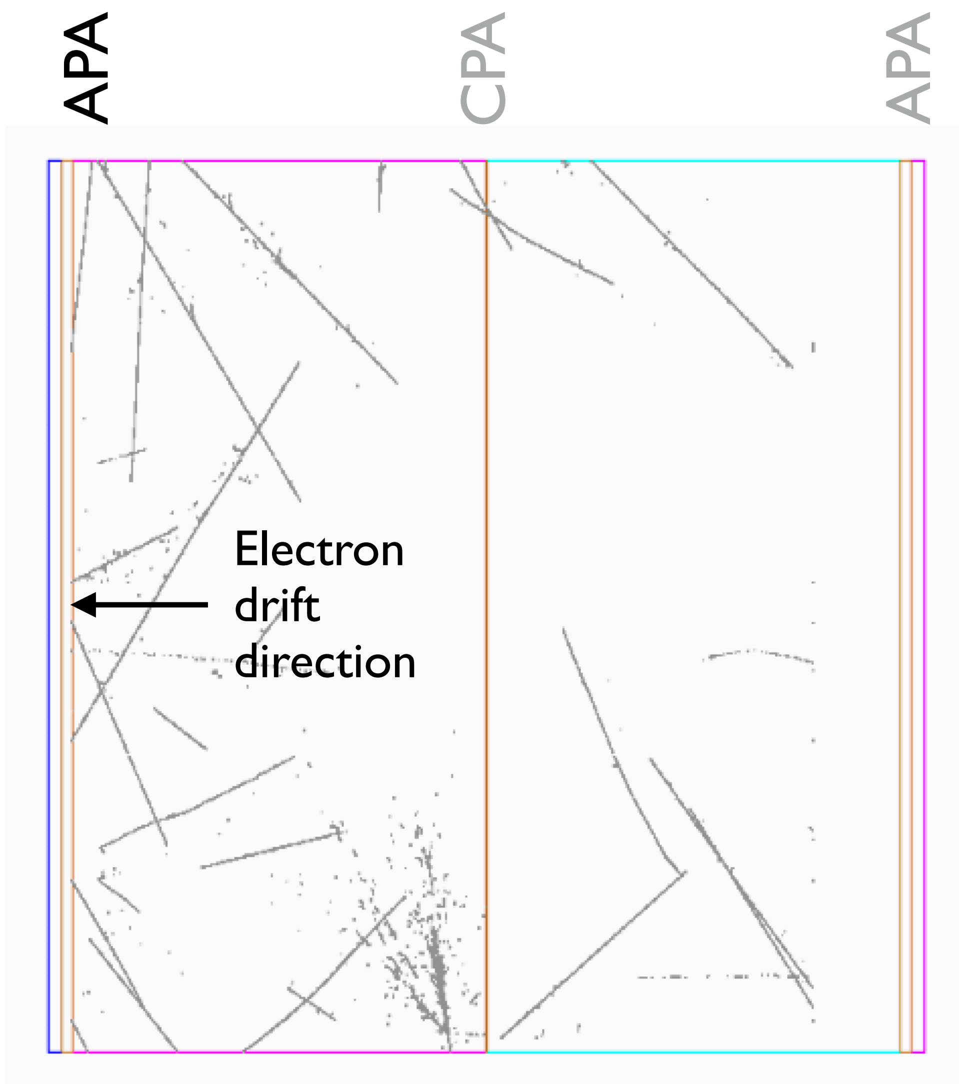
ProtoDUNE-SP simulation

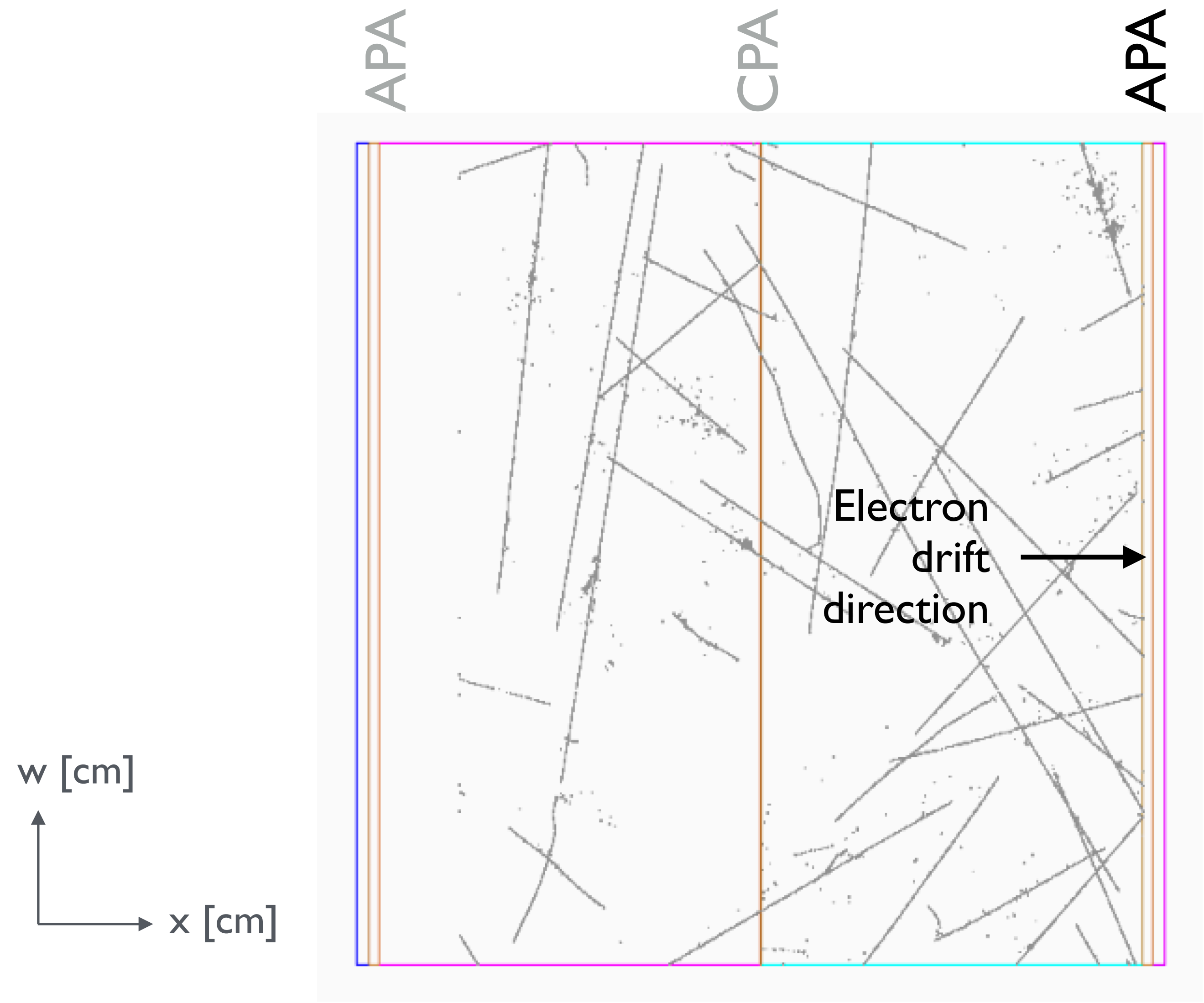


EPJC (2018) 78:82

Example: Reconstruction at ProtoDUNE-SP

- **Single Phase DUNE Far Detector prototype, exposed to test beam at CERN**
- **Multiple "drift volumes", complex topologies and significant cosmic-ray backgrounds:**
  - An ideal testing ground for LArTPC pattern recognition



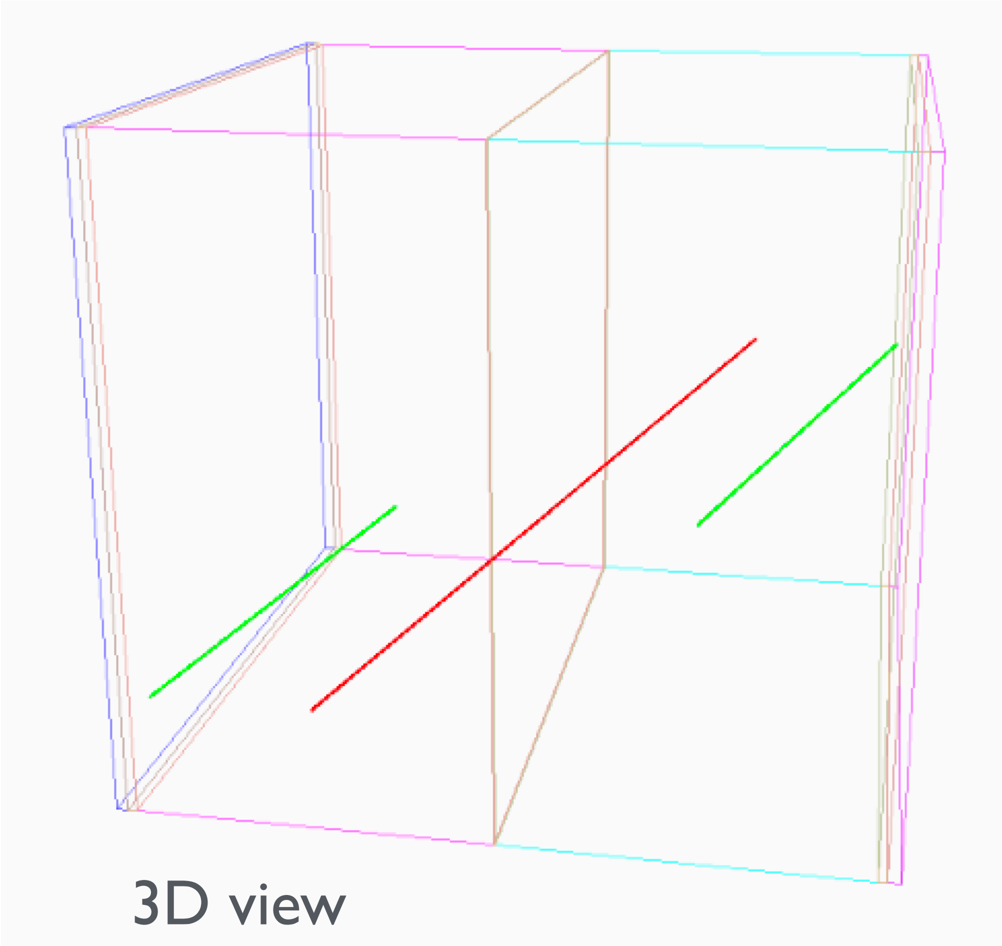


APA: Anode Plane Assembly
CPA: Cathode Plane Assembly

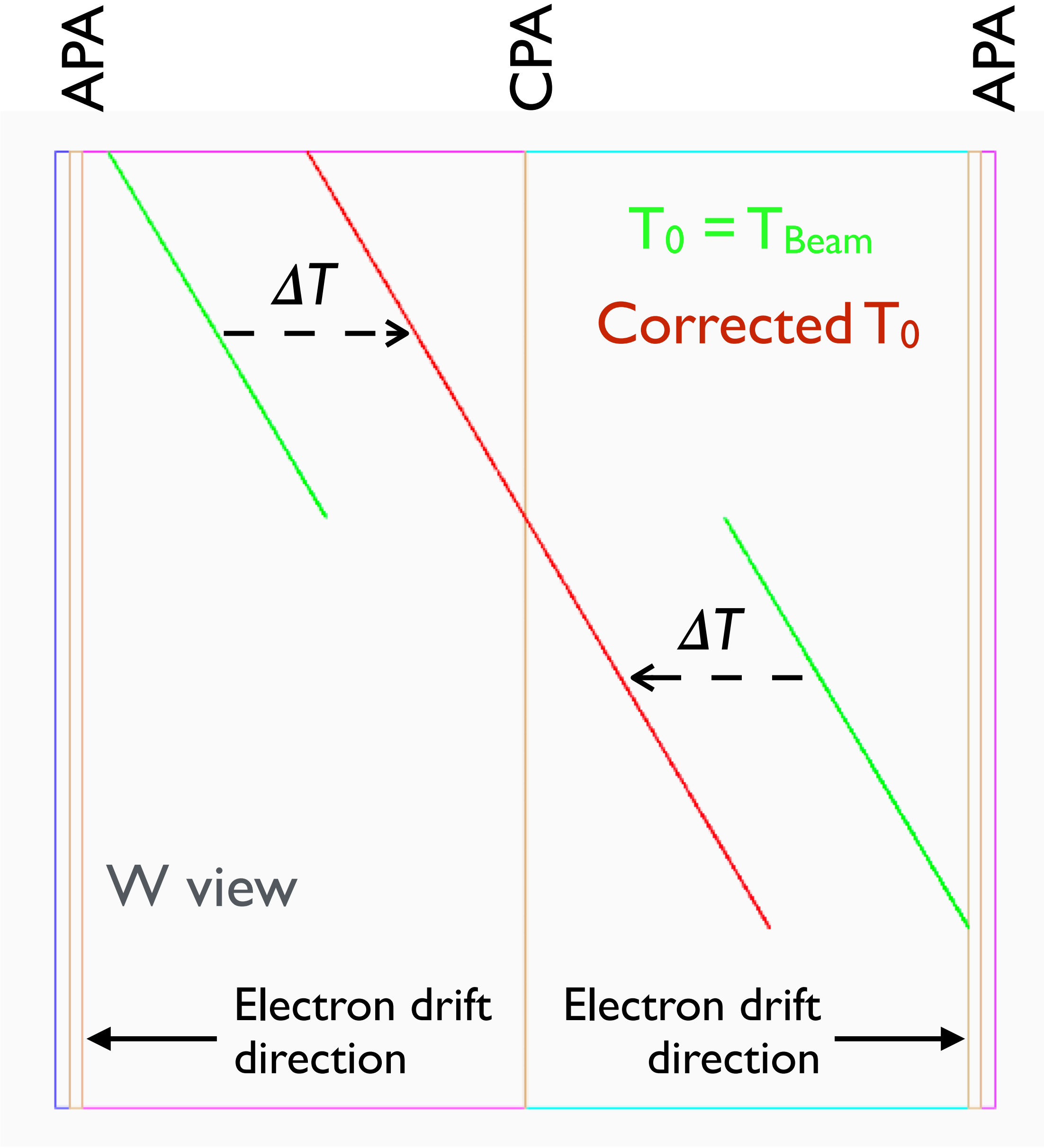
1. Reconstruct cosmic-ray muons independently for each volume of detector

Stitching and T_0 Identification

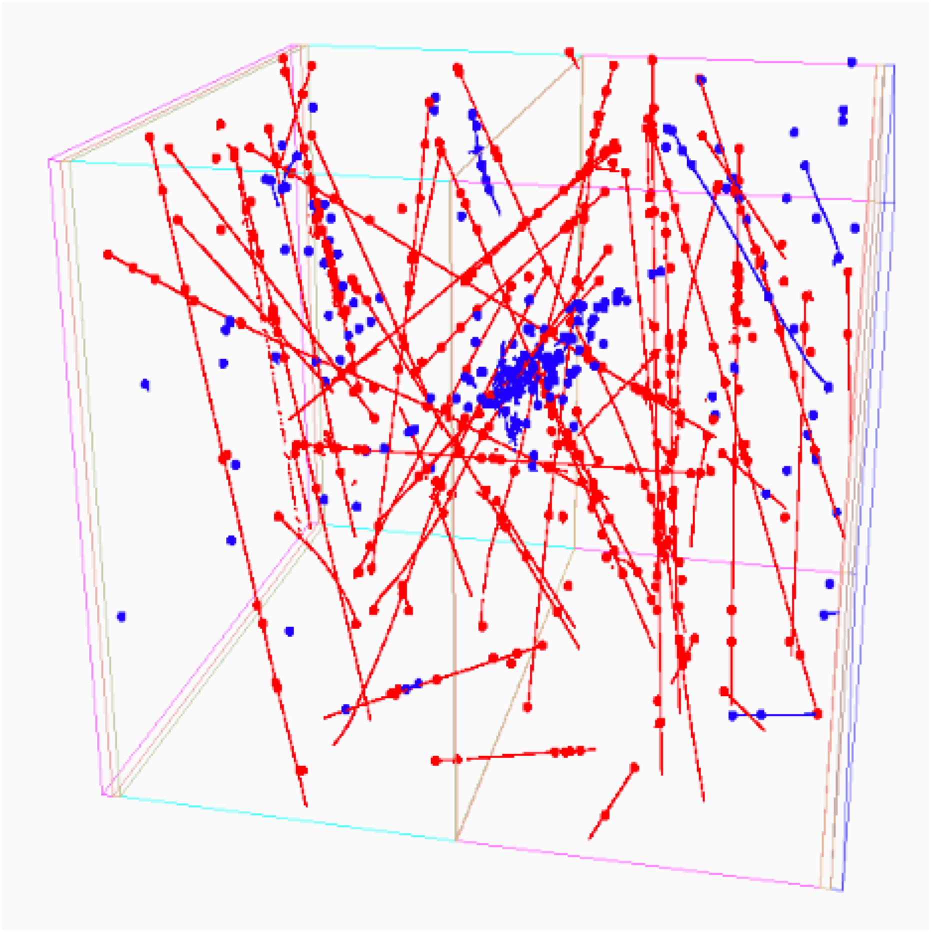
- **In a LArTPC image, one coordinate derived from drift times of ionisation electrons:**
  - But, only know electron arrival times, not actual drift times: need to know start time, T_0
  - For beam particles, can use time of beam spill to set T_0, but unknown for cosmic rays
  - Place all hits assuming T_0=T_{Beam}, but can identify T_0 for any cosmic rays crossing volumes



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



Cosmic Ray Tagging and Slicing

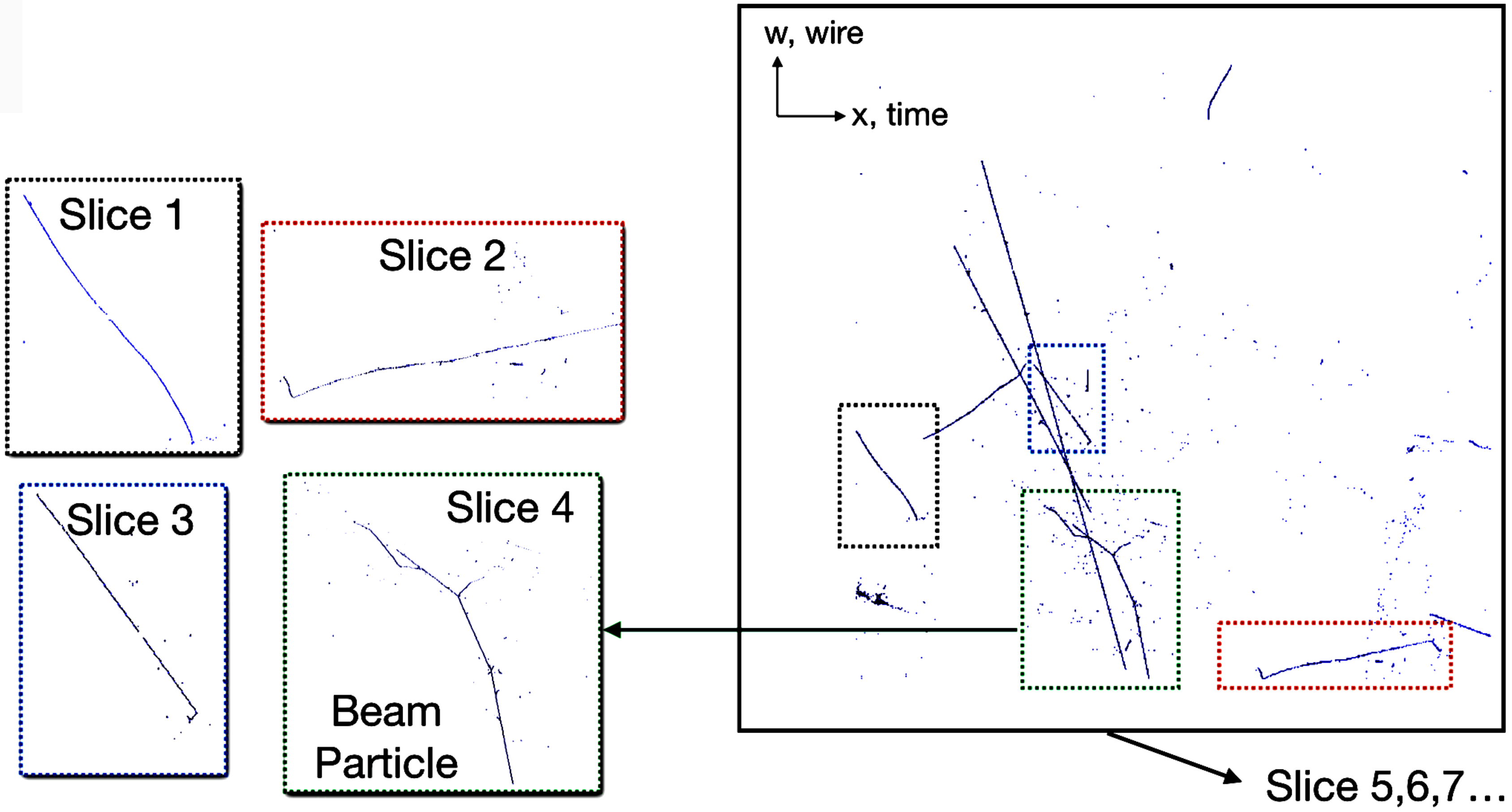


3. Identify clear cosmic rays (red) and hits to reexamine under test beam hypothesis (blue)

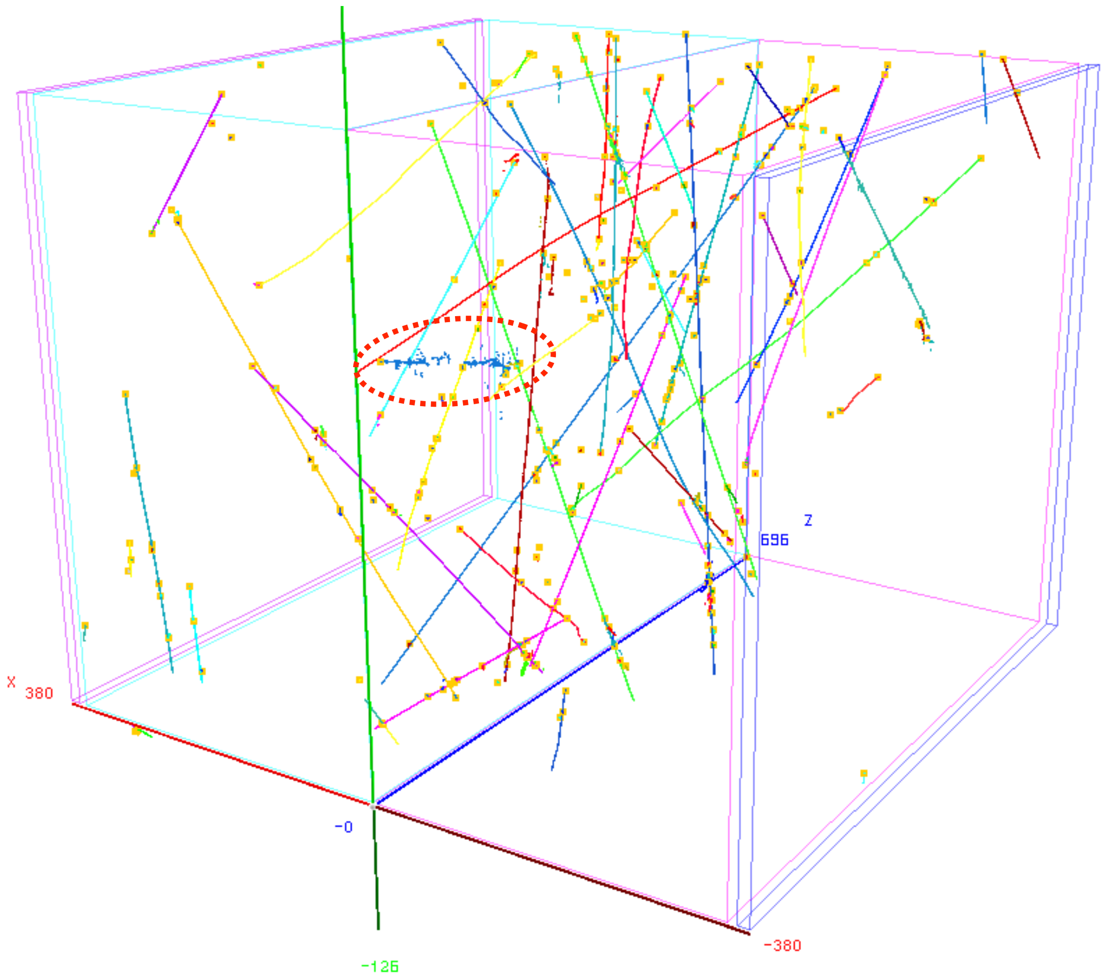
Clear cosmic rays:

- 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**

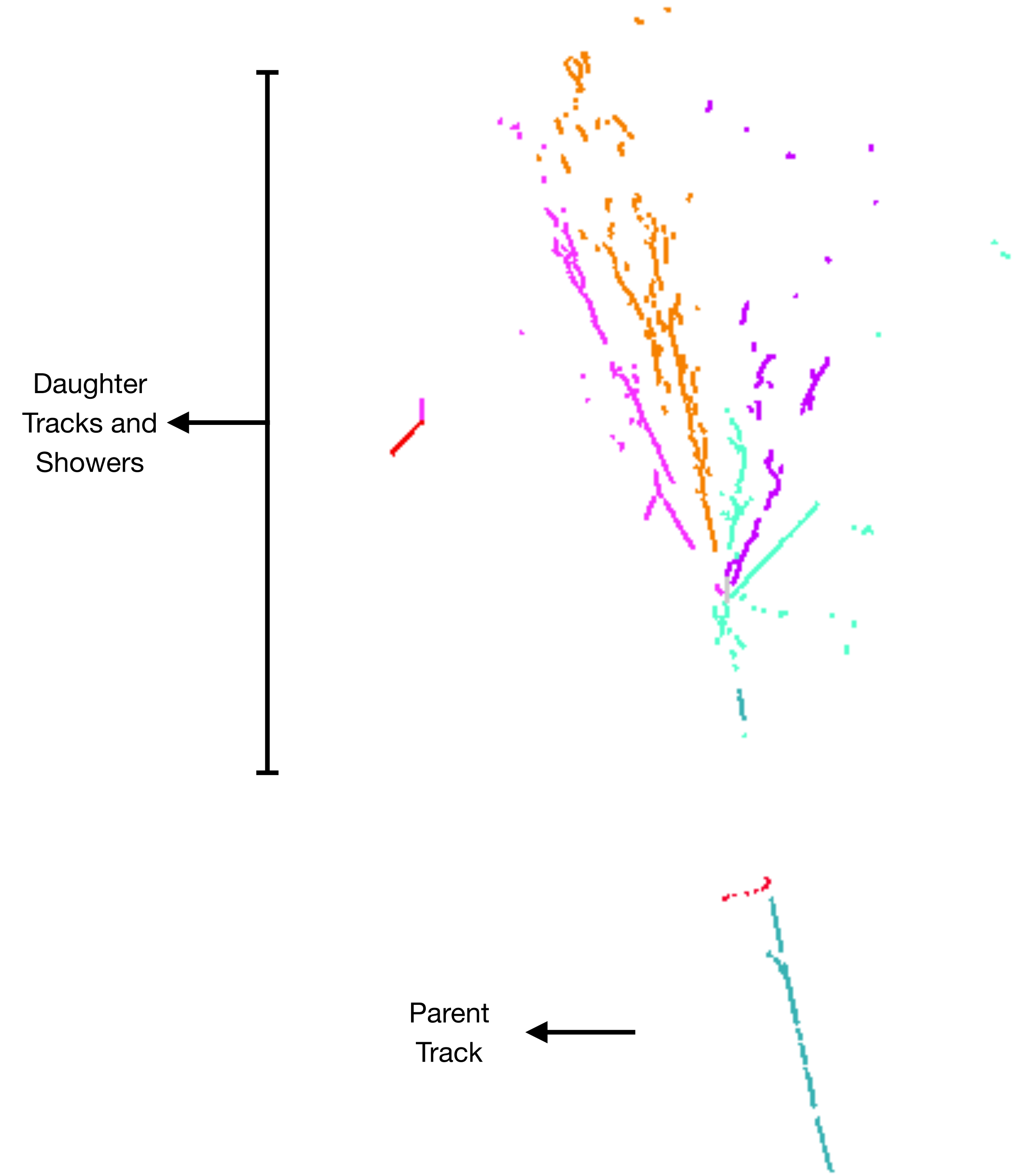


Consolidated Output



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

and: N reconstructed cosmic-ray muon hierarchies

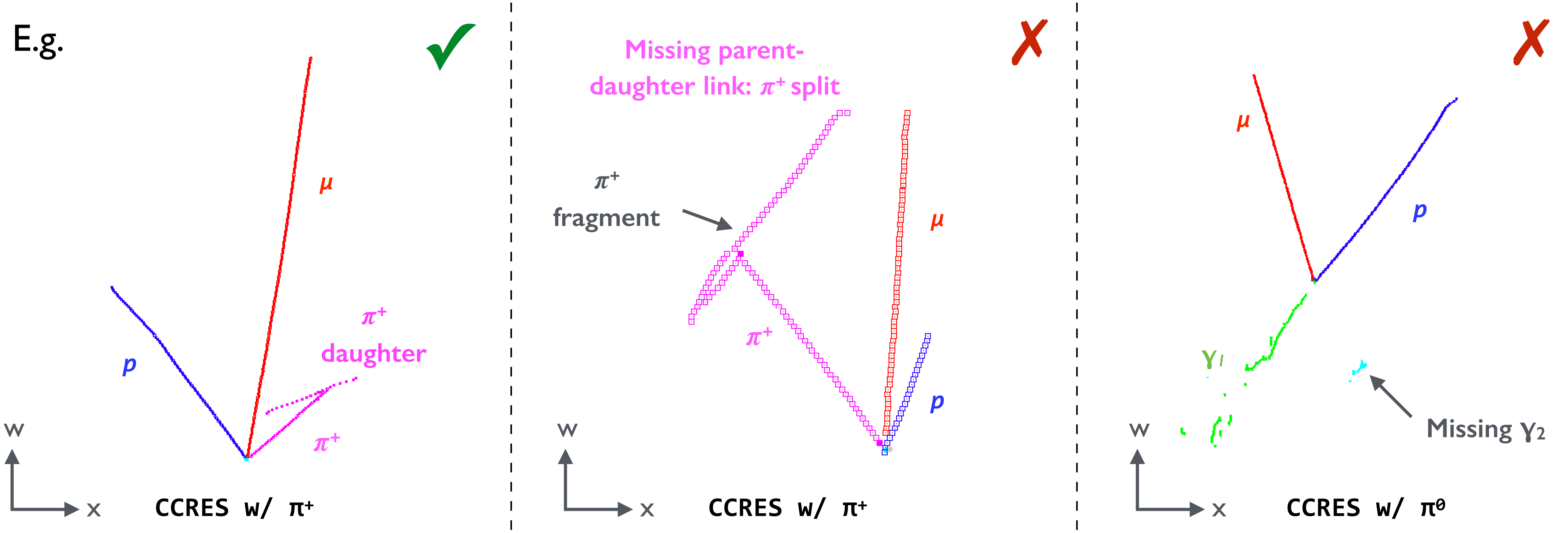


E.g. Test beam particle: charged pion

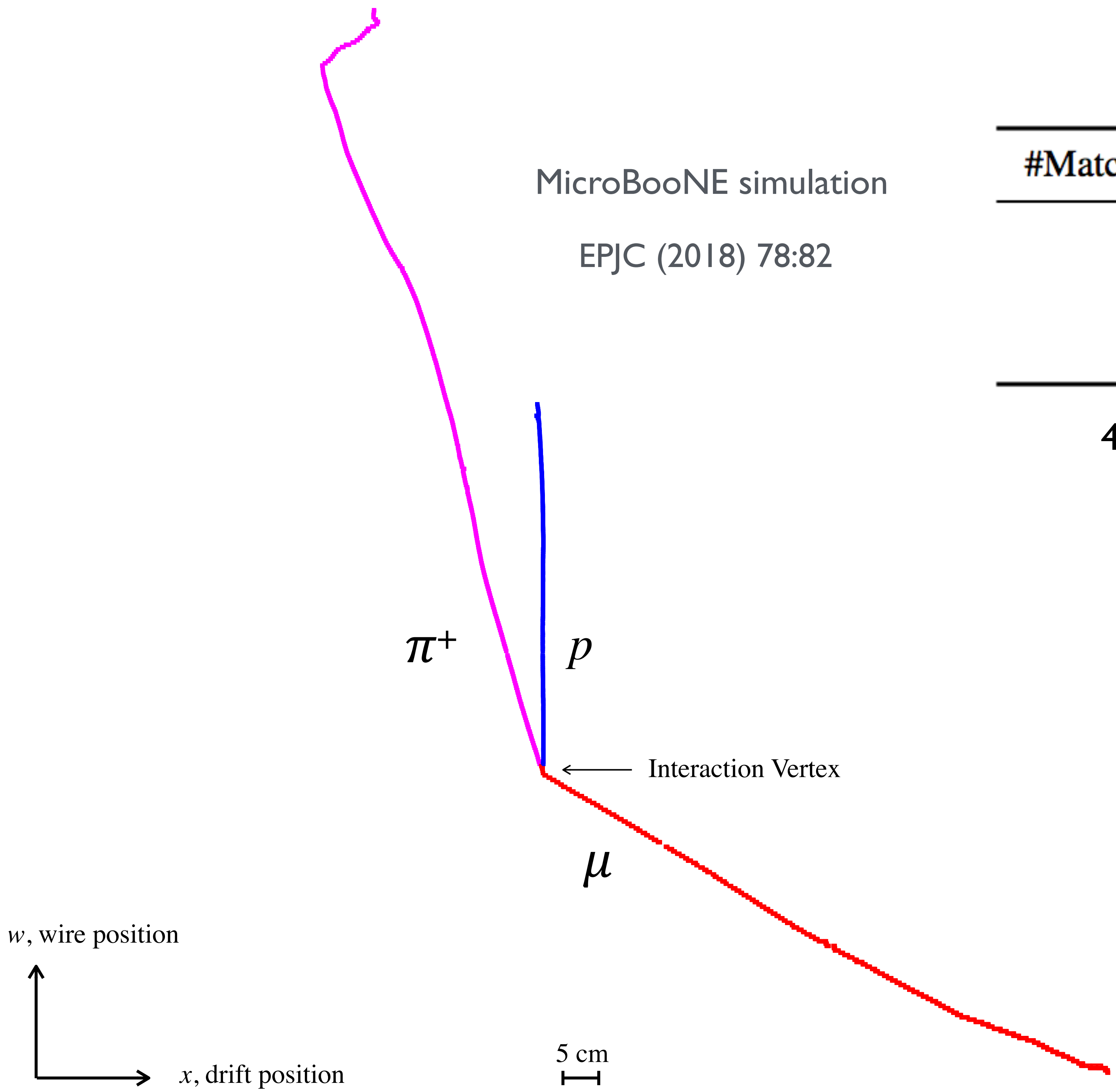
Assessing Pattern-Recognition Performance

- Assess performance for simulated events, using a 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.



CC RES: ν_μ + Ar → μ^- + p + π^+



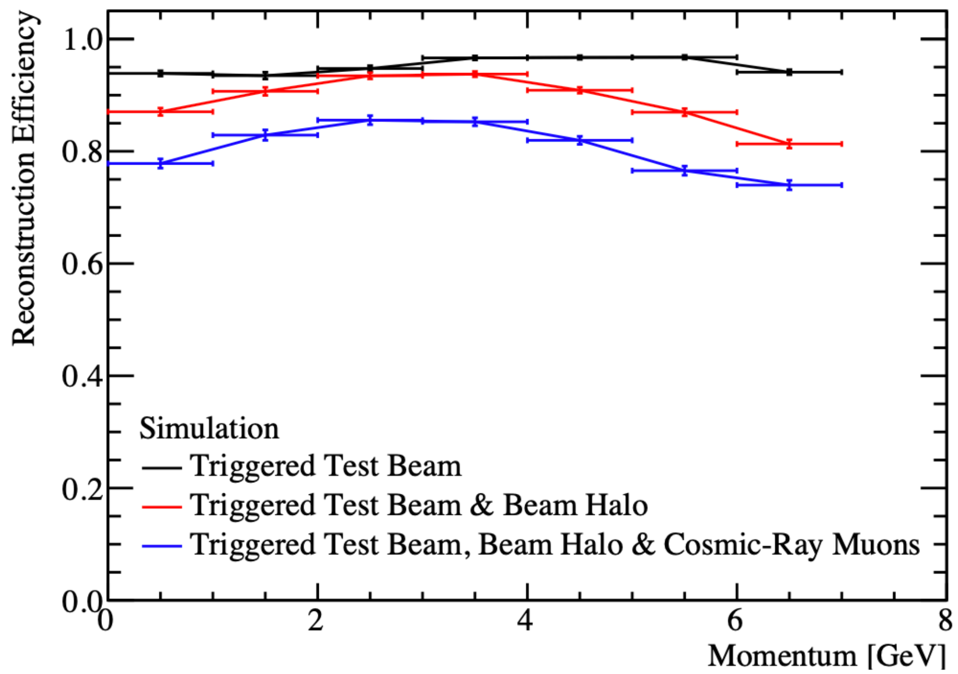
Three-track topology: CC ν_μ interactions with resonant charged pion production:

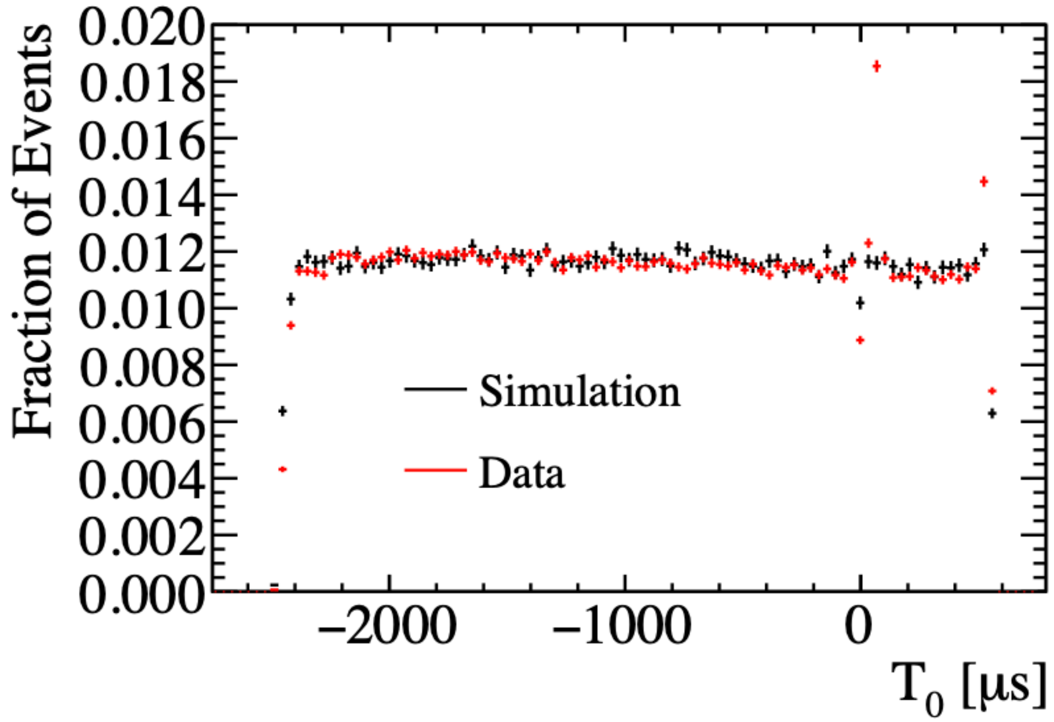
#Matched Particles	0	1	2	3+
μ	3.5%	95.1%	1.4%	0.0%
p	9.0%	86.8%	4.0%	0.3%
π^+	6.9%	80.9%	11.4%	0.8%

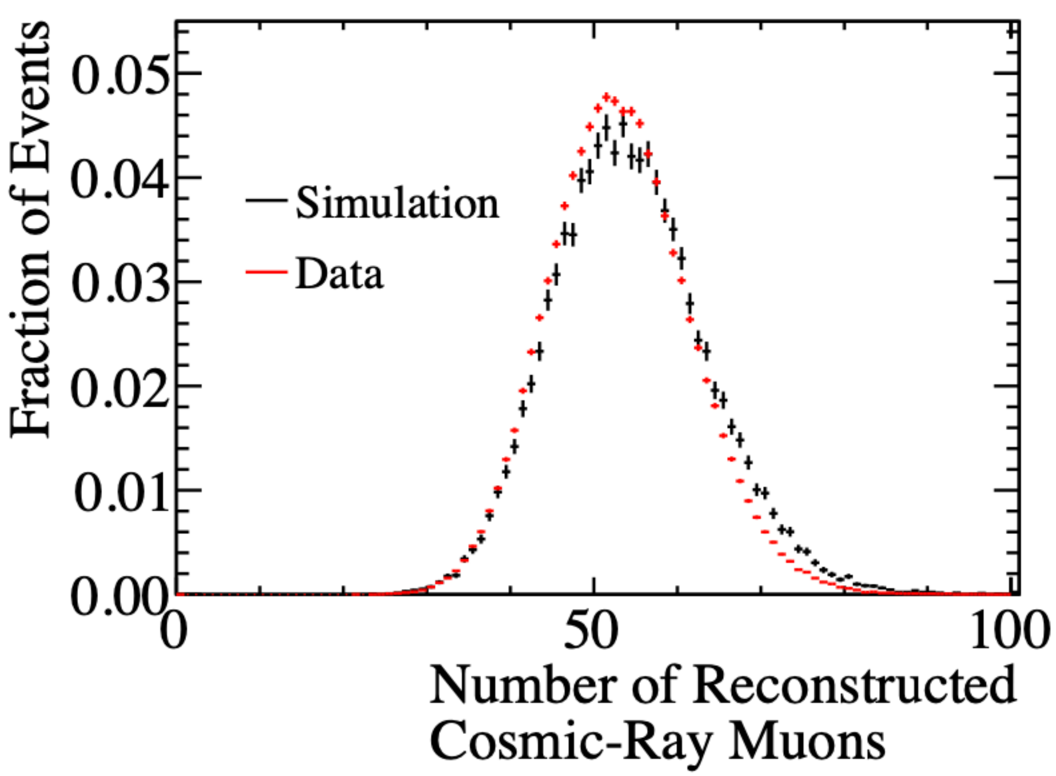
47,754 events, 70.5% have exactly one reco particle matched to each target.

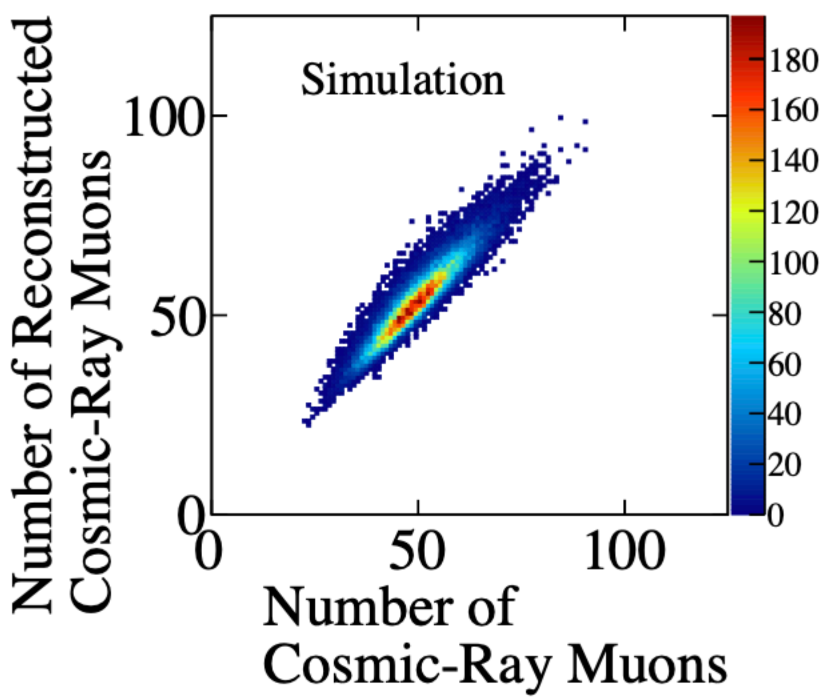
- Performance for μ and p similar to that reported for quasi-elastic events.
- π^+ interactions can lead to hierarchy of visible particles. If reconstructed separately (without parent-daughter links), π^+ is reportedly split.

1.1 GeV ν_μ

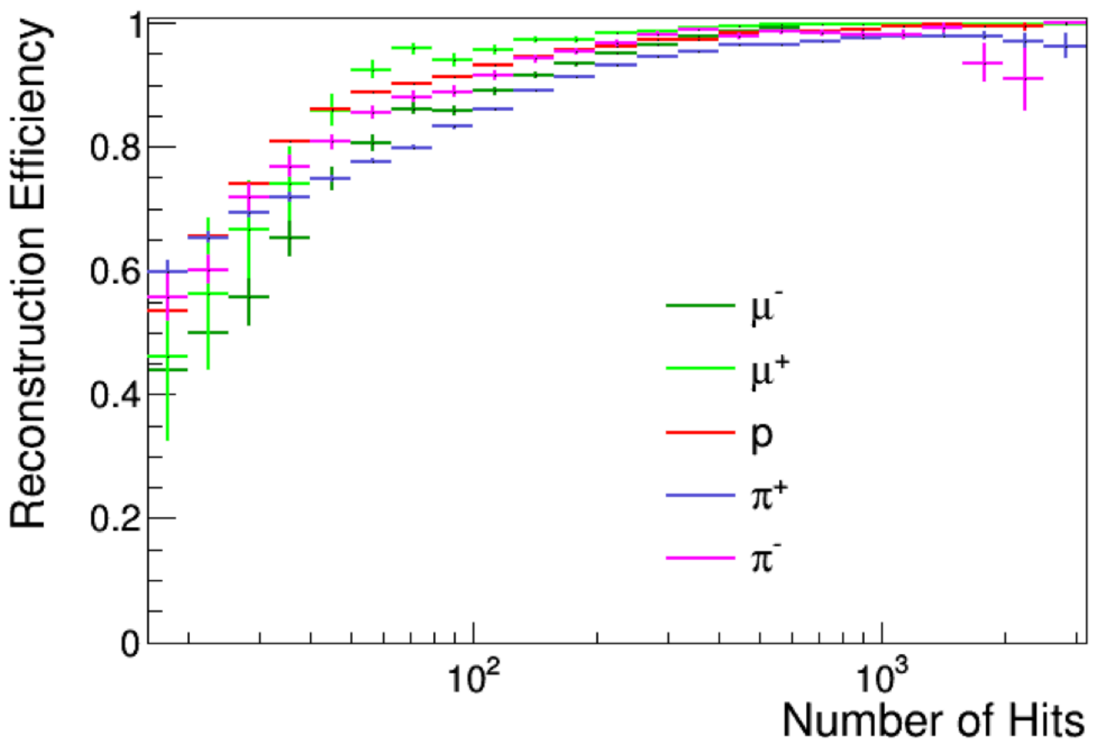


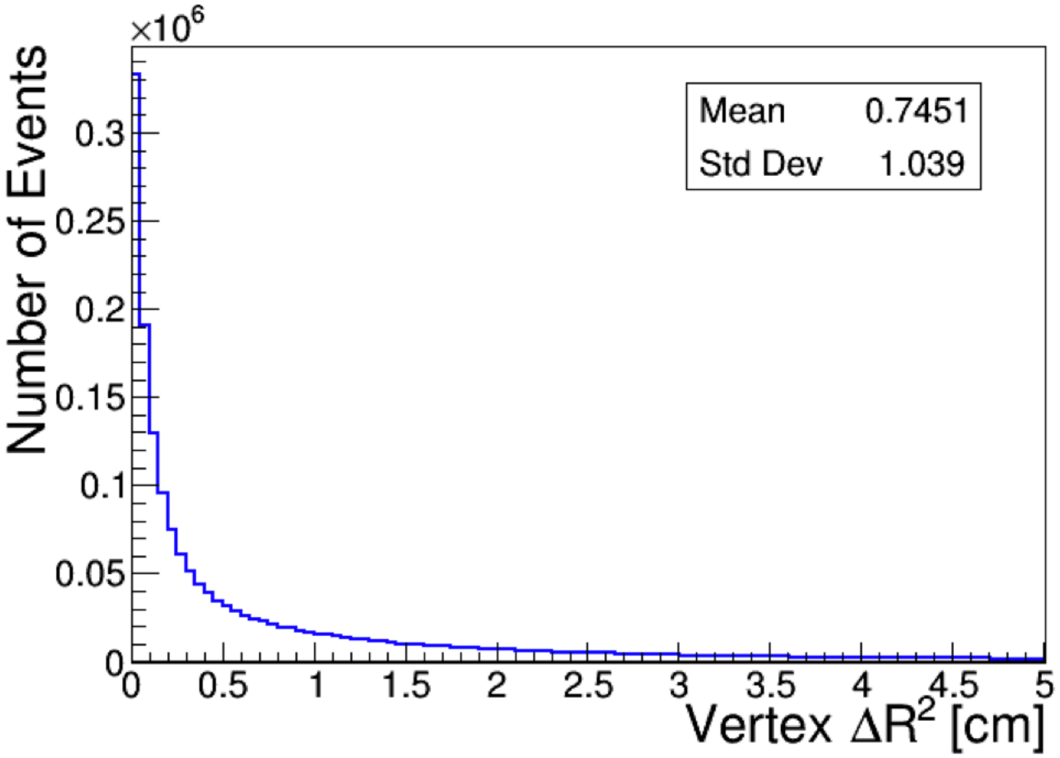


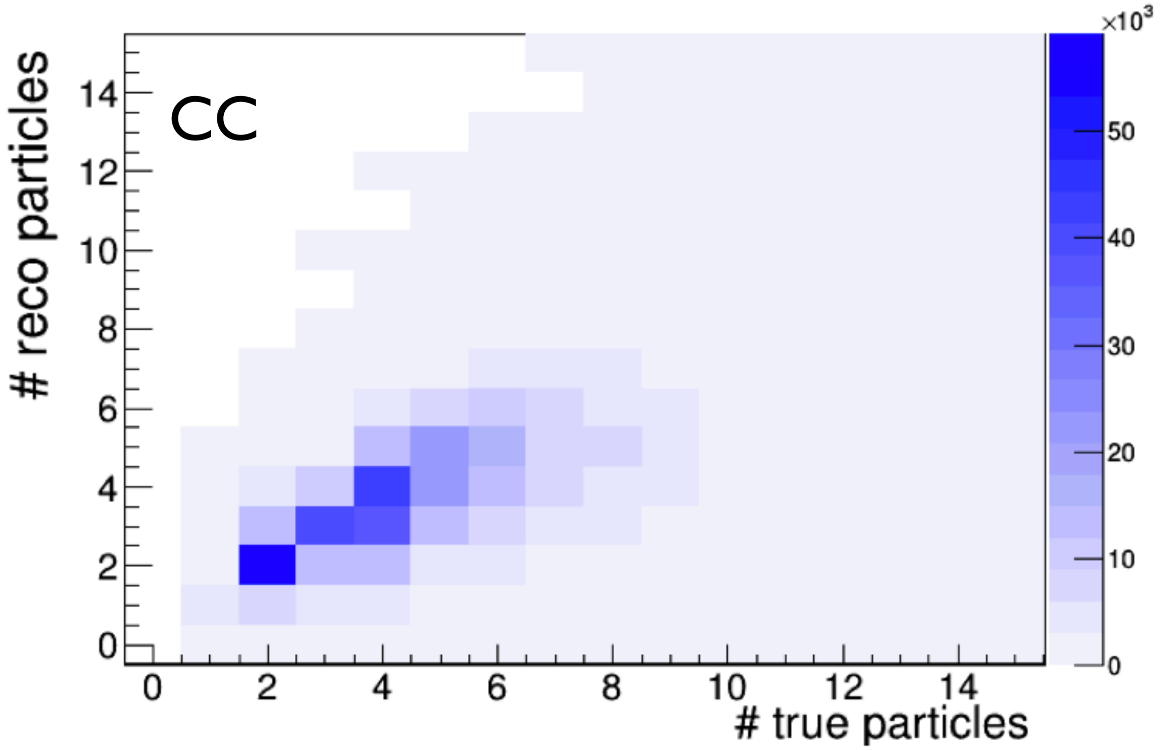


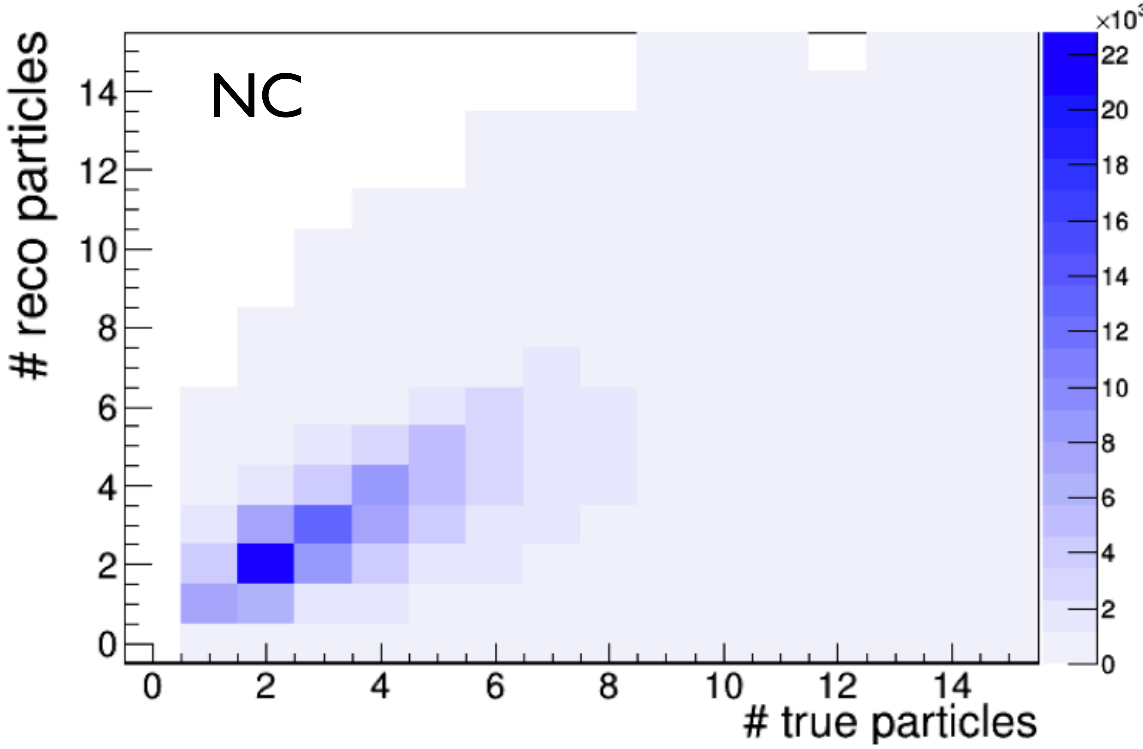


https://arxiv.org/pdf/2002.03005.pdf







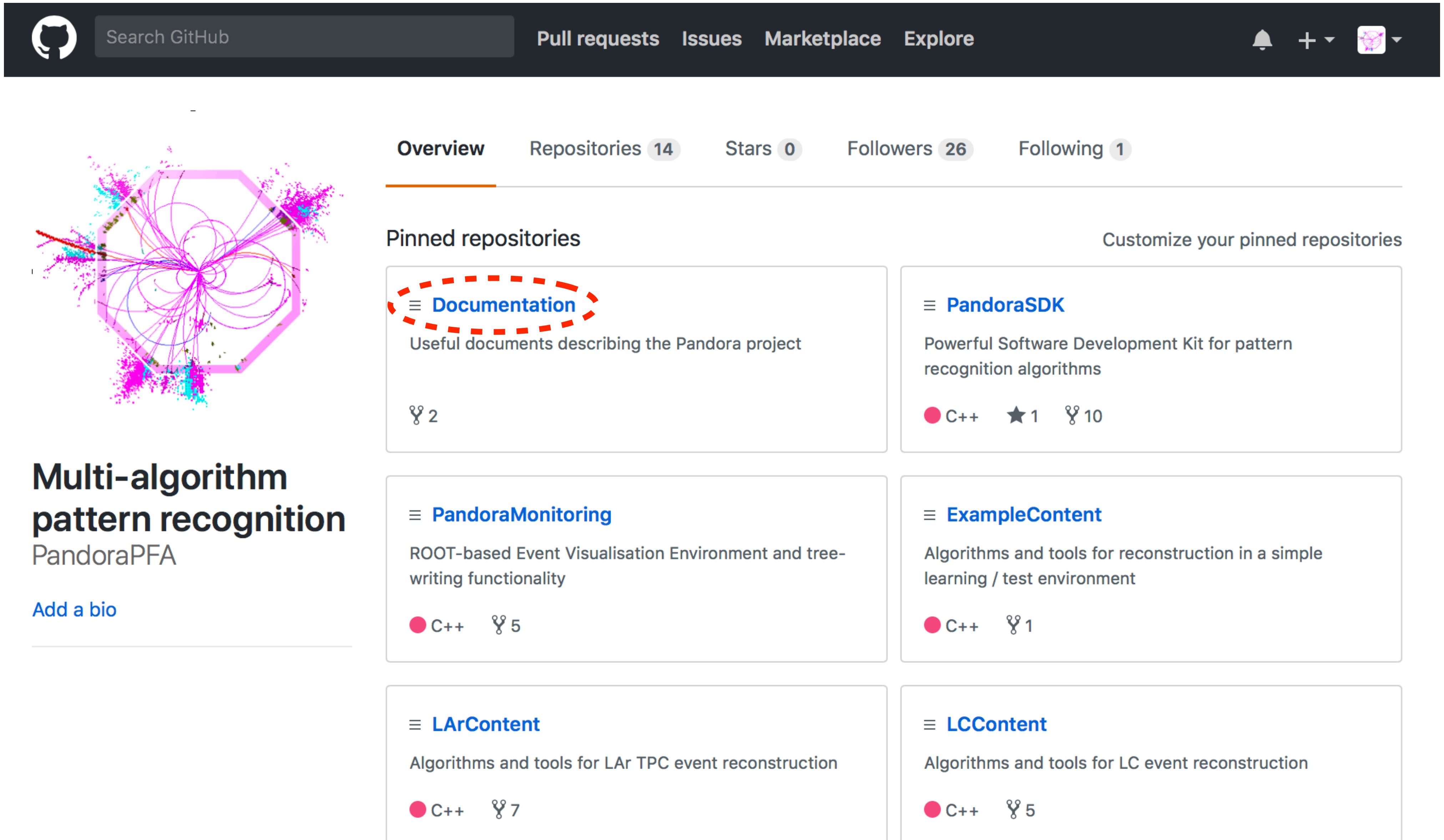


https://arxiv.org/pdf/2002.03005.pdf

Summary

- The use of Liquid Argon technology is one of the cornerstones of the current and future neutrino programmes.
- 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.
  - Output is a carefully-arranged hierarchy of reconstructed particles, each corresponding to a distinct track or shower.

WOULD YOU LIKE TO KNOW MORE?



https://github.com/PandoraPFA

https://pandorapfa.slack.com

WARWICK
THE UNIVERSITY OF WARWICK

UNIVERSITY OF CAMBRIDGE

Lancaster University

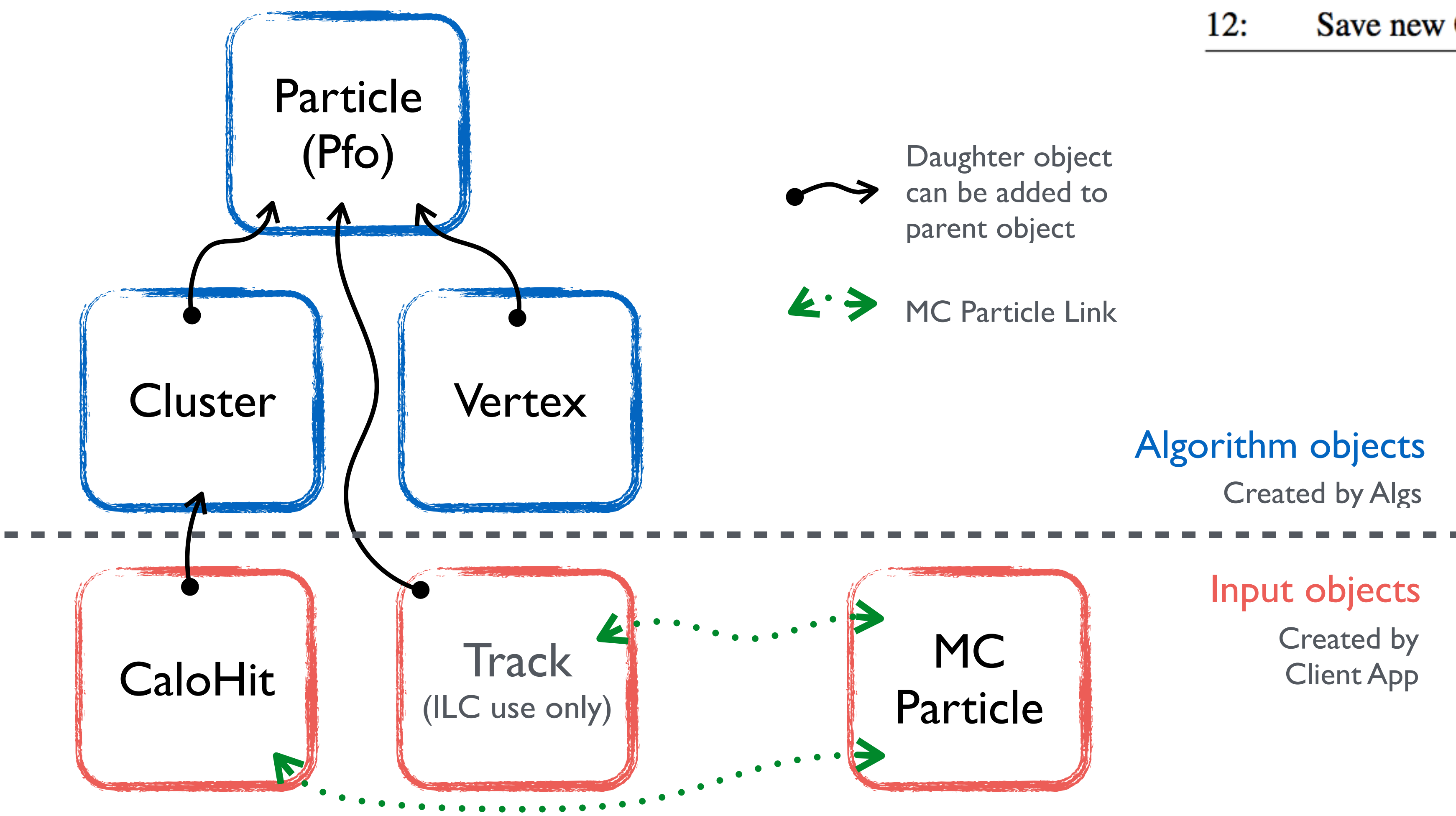
Additional slides

- **Algorithms contain high-level logic and concentrate on the important bits: physics/patrec ideas.**
  - Pandora software framework provides functions to access objects, make new objects, modify existing objects, etc.

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
```

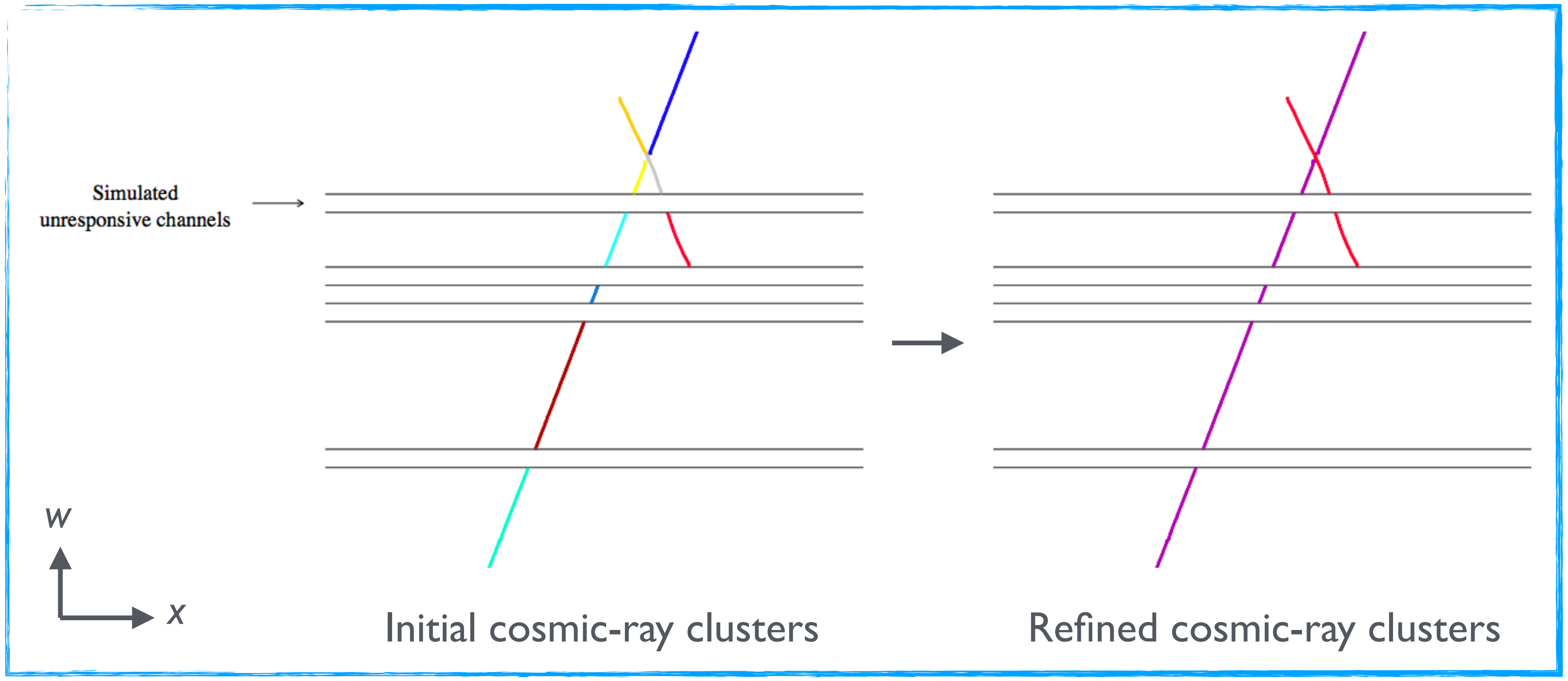
Example alg structure

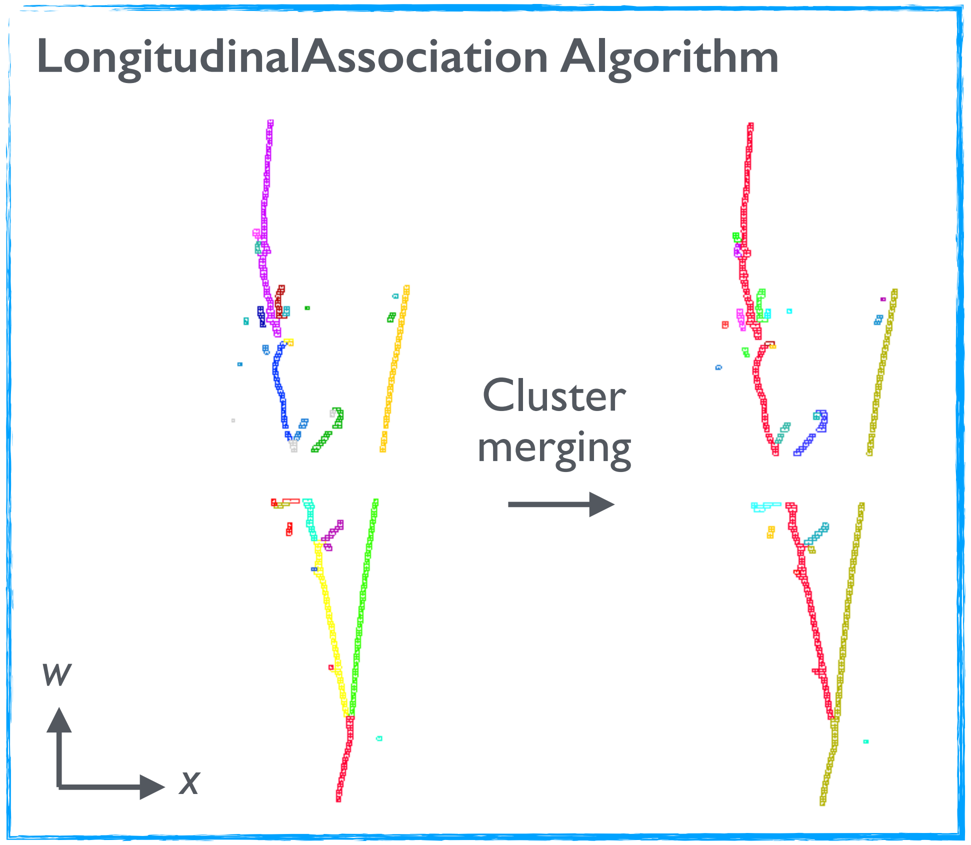


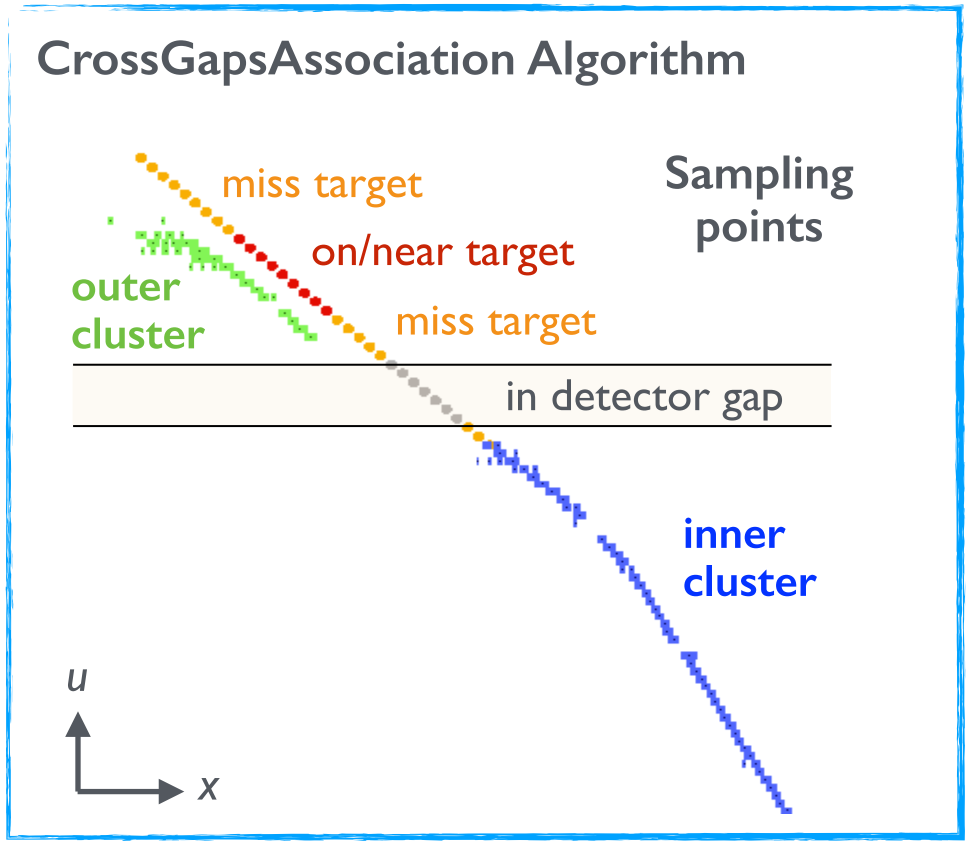
Event Data Model

Pandora - "Traditional" Approaches

- For each wire plane, create a list of 2D clusters that represent continuous, unambiguous lines of hits:
  - Separate clusters for each structure, with clusters starting/stopping at any branch or ambiguity.



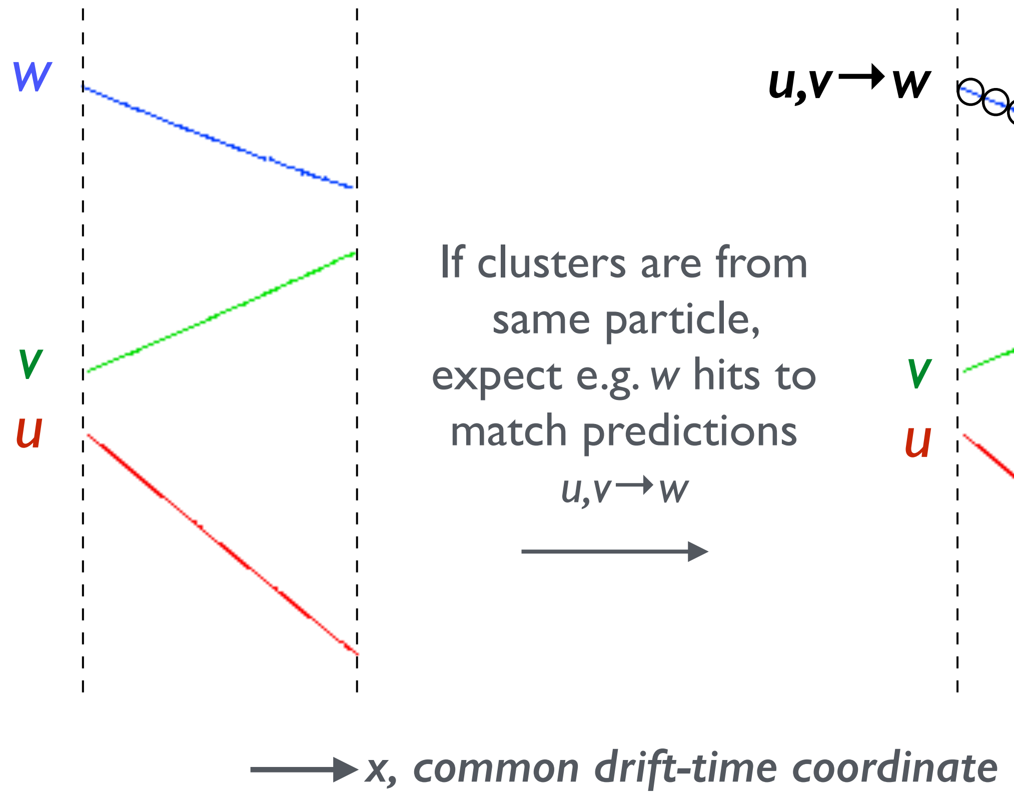




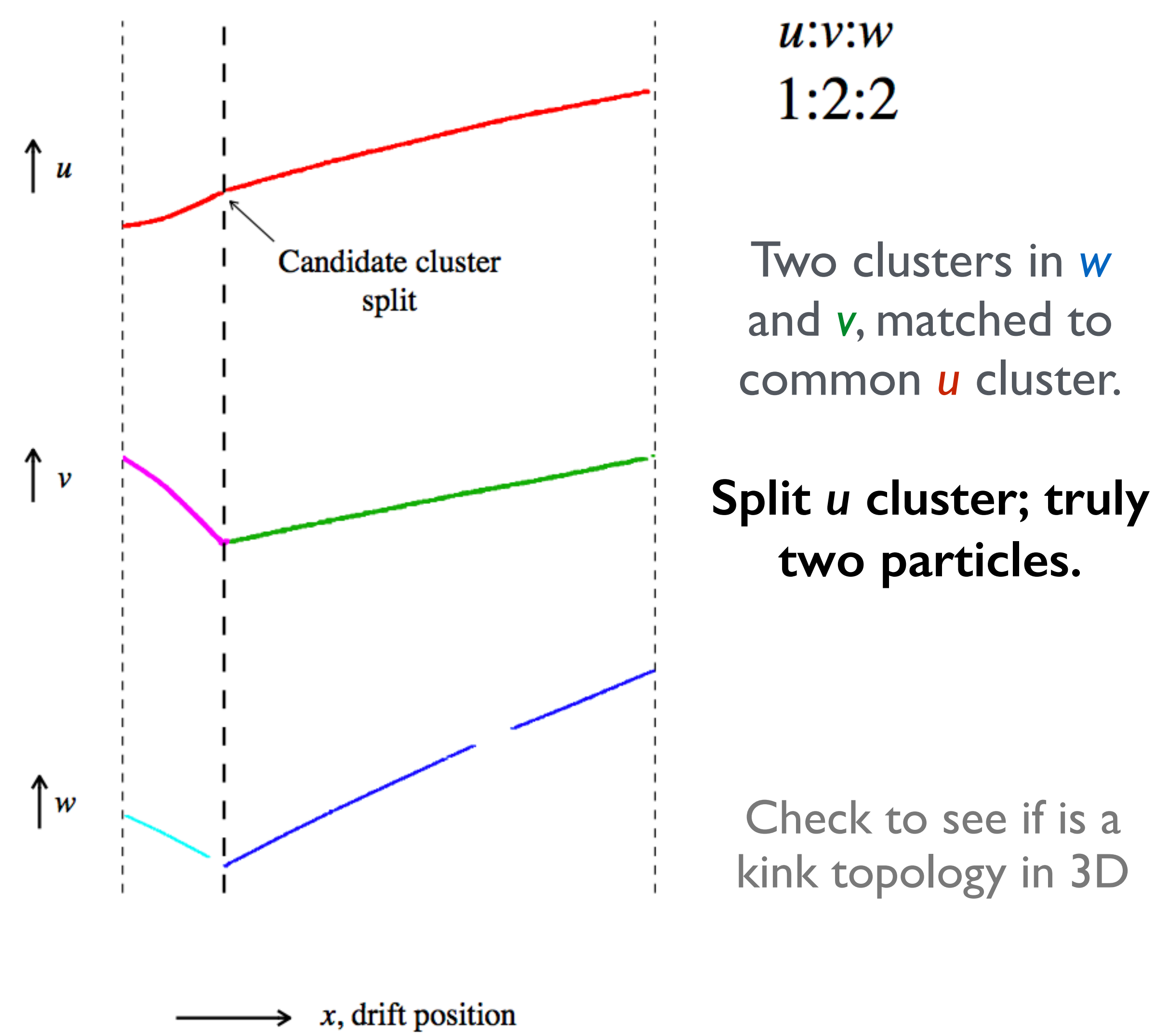
Initial clusters are refined by a series of cluster-merging and cluster-splitting algorithms that use topological info.

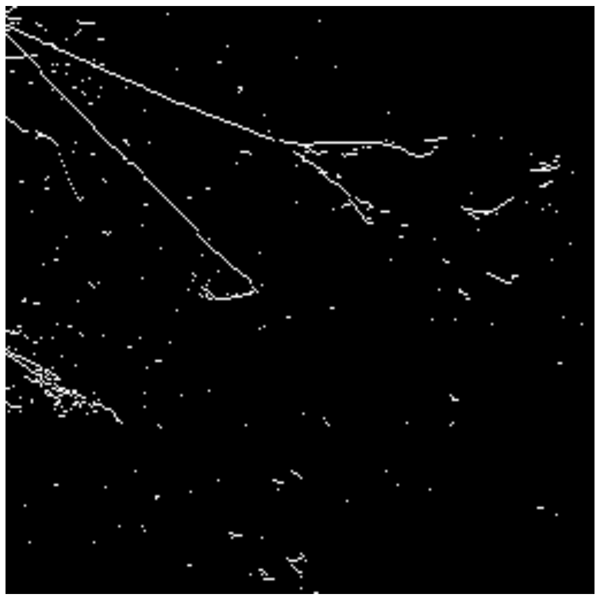
Pandora - "Detector-Physics" Approaches

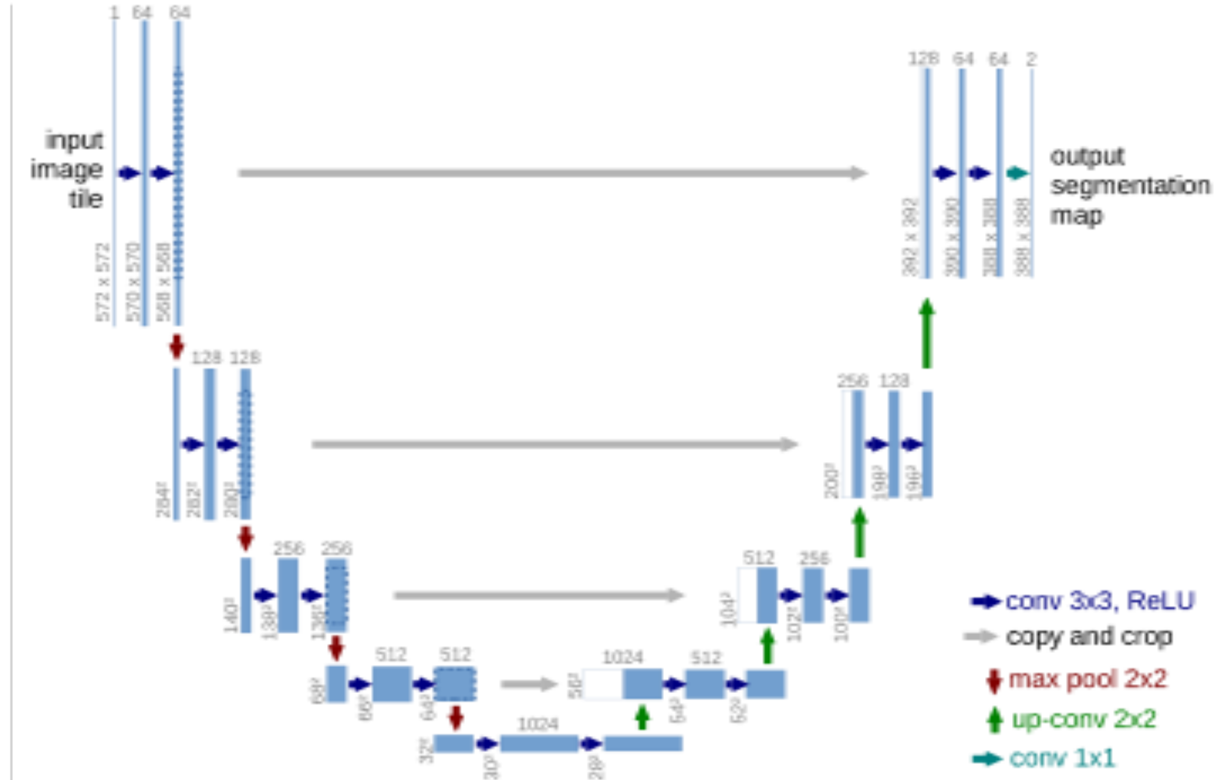
- **Our original input was 3x2D images of charged particles in the detector.**
- **Should now have reconstructed three separate 2D clusters for each particle:**
  - Compare 2D clusters from *u*, *v*, *w* planes to find the clusters representing same particle.
  - Exploit common drift-time coordinate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

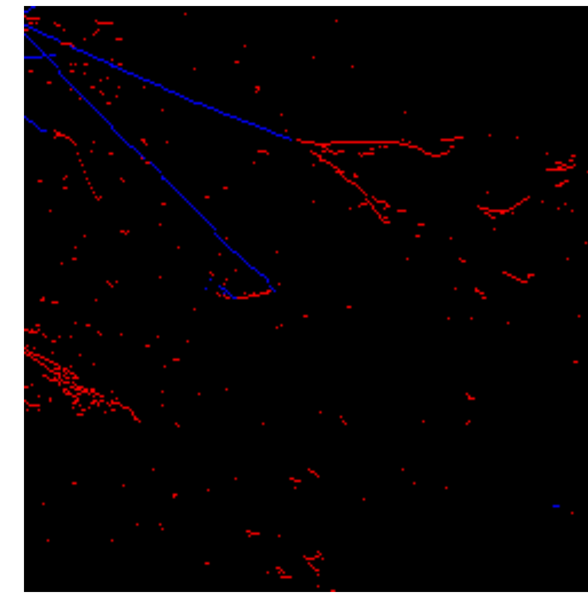


- **Approach really comes to life when the 2D clustering "disagrees" between wire planes:**
  - **Automated detection of 2D PatRec issues, with treatment for specific cases, e.g.:**

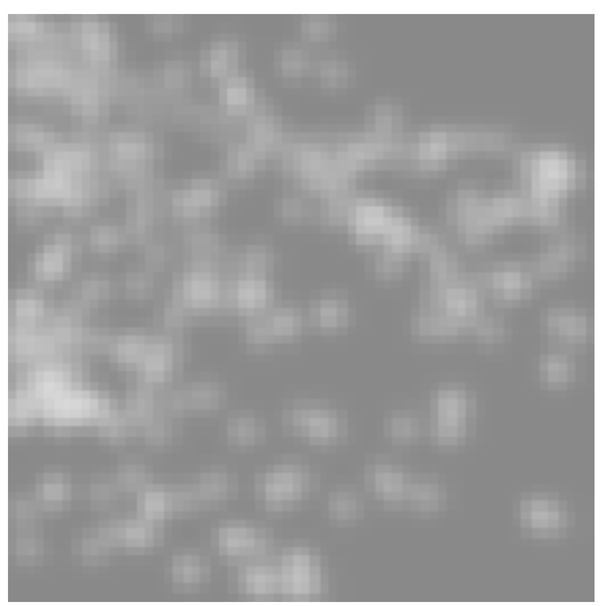








https://arxiv.org/abs/1505.04597



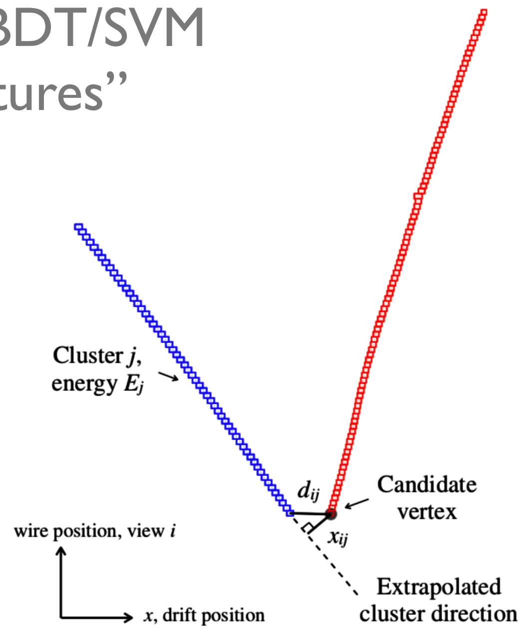
- Convolution filters identify features
- Track/Shower probability map constructed

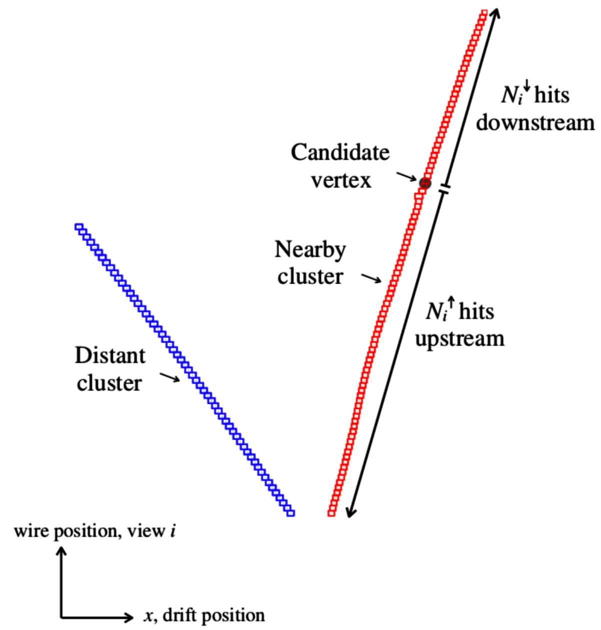
Vertex Reconstruction - 3D

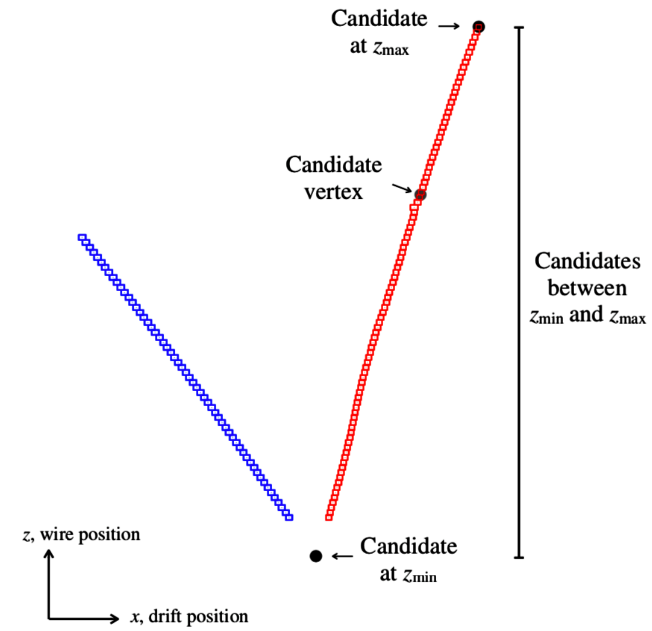
Interaction vertex is an important feature point in our LArTPC images:

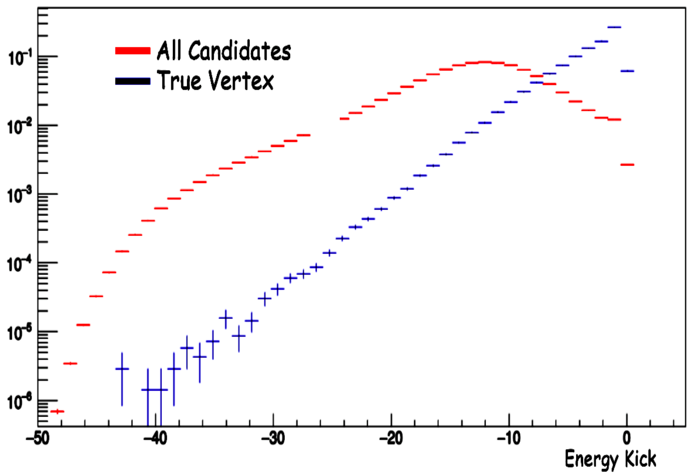
- Continued development, ever-more sophisticated approaches to finding 3D vertex position
- Boosted Decision Trees (BDTs) or Support Vector Machines (SVMs) to select best candidate
- Exploit Convolutional Neural Networks (CNNs)

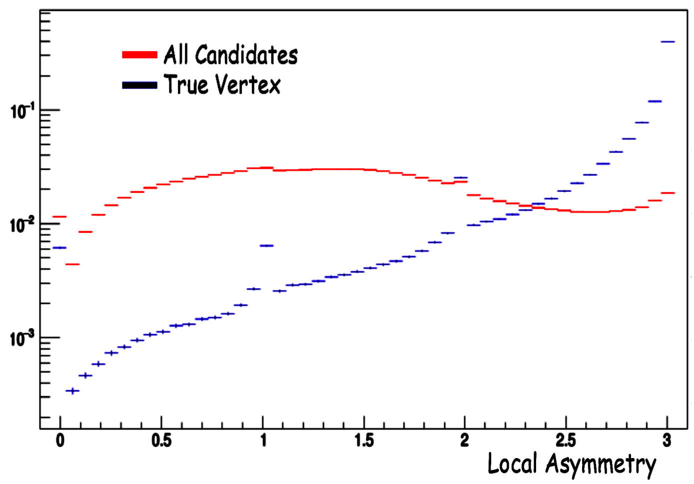
E.g BDT/SVM "features"

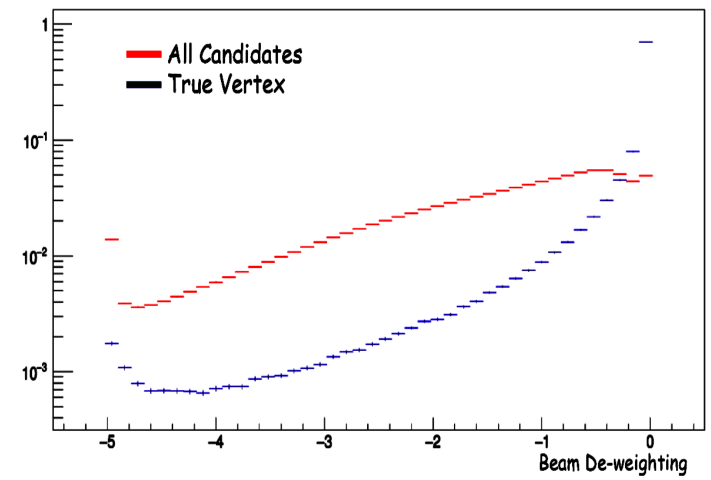




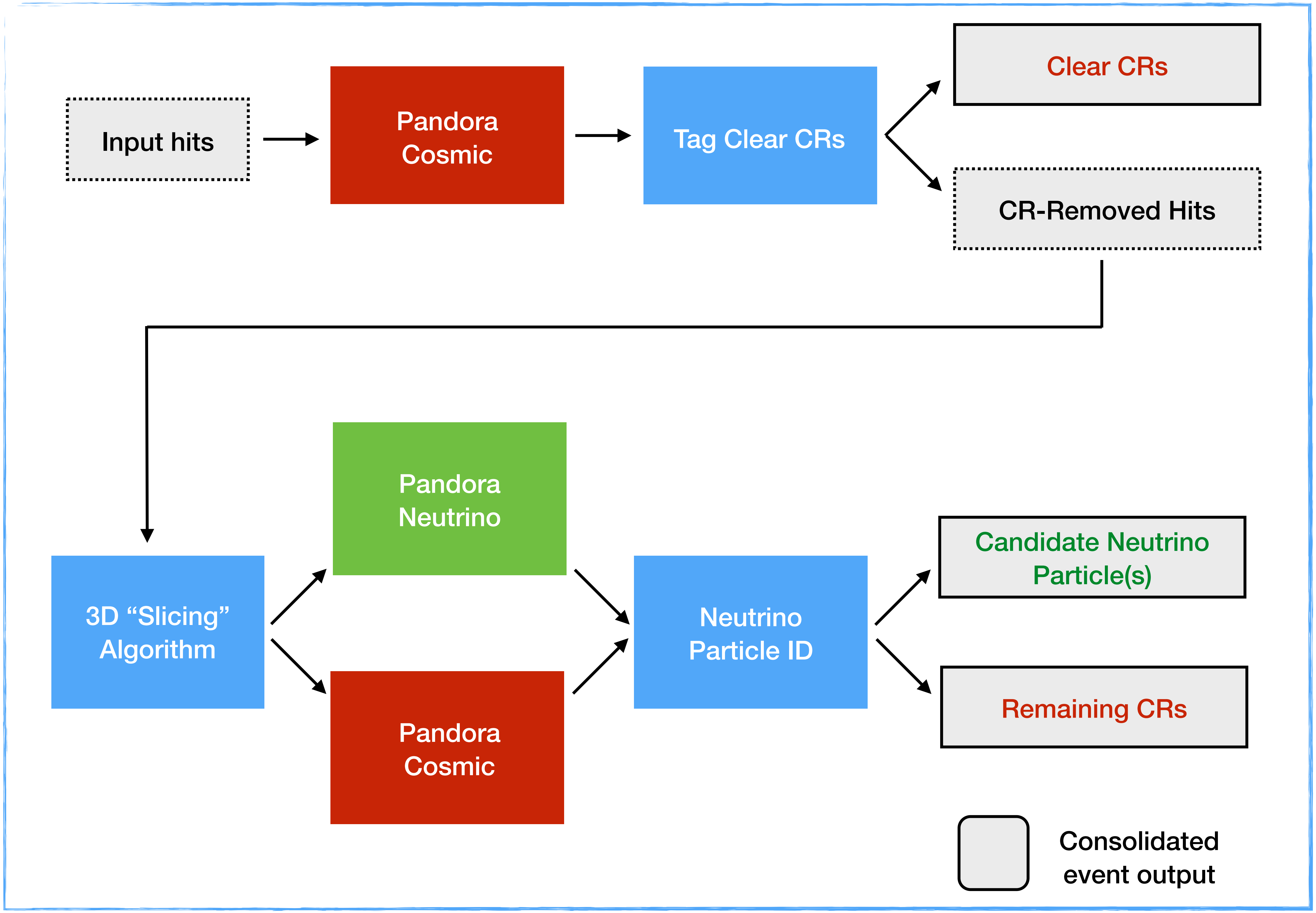








Pandora - Consolidated Reconstruction



Pandora - Performance Metrics

1. **Determine target MCParticle associated to each hit**
   - Use MCParticle hierarchy to determine primary "targets" for reco
   - Associate hits to target MCParticle making largest E contribution

Target MCParticles must satisfy quality cuts

Reco/MCParticles matches must satisfy quality cuts.

2. **Match reco particles to target MCParticles**
   - For each combination of reco particle and target MCParticle, find the number of shared hits; fold all daughter particles, in both reco and MCParticle hierarchies, back into parent primaries
   - **Interpret** raw/comprehensive matching information to clarify pattern recognition performance:
     1. Find strongest (most shared hits) match between any reco particle and target MCParticle
     2. Repeat step i, using reco and MCParticles at most once, until no further matches possible
     3. Assign any remaining reco particles to target MCParticle with which they share most hits

3. **Define performance metrics**
   - Efficiency: Fraction of target MCParticles with at least one matched reco particle
   - Completeness: Fraction of MCParticle true hits shared with the reco particle
   - Purity: Fraction of hits in reco particle shared with the target MCParticle

   - **Match exactly one reco particle to each target MCParticle ⇒ Event is "correct"**

Pandora - Performance Metrics

- In practice, some MCParticles not reconstructable. Targets must satisfy quality cuts:
- ≥15 hits in total, at least five hits in at least two views.
- Target must deposit >90% E in these hits.
- Plus, ignore all hits which are downstream of far-travelling neutron in MC hierarchy.

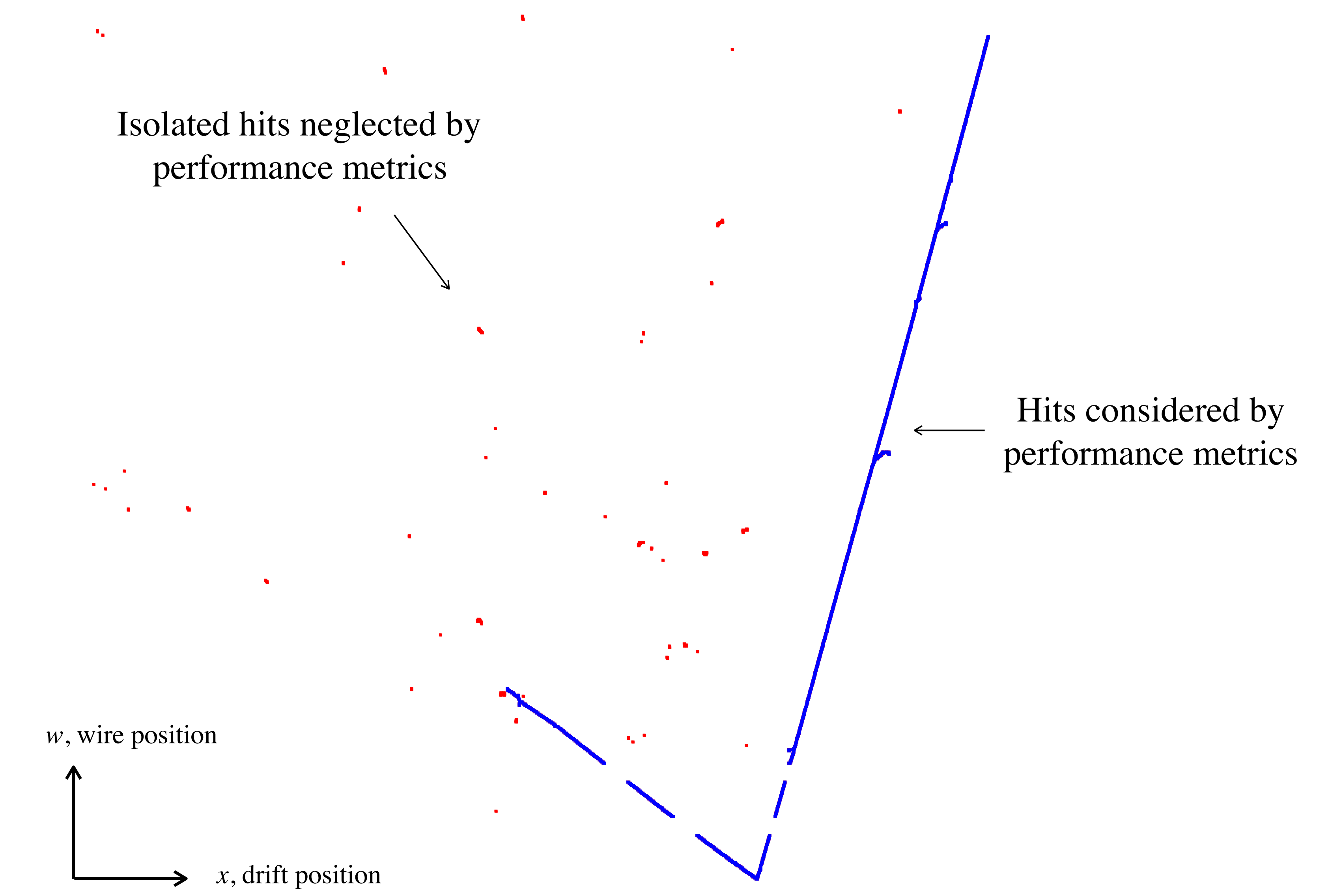


Fig. 8: The hits that are considered (blue) and neglected (red) in the construction of pattern recognition performance metrics for a typical CC ν_μ event at MicroBooNE. By considering the MCParticle hierarchy, hits that will likely form part of an isolated and diffuse topology are not used to identify or characterise the reconstructable target MCParticles in an event.

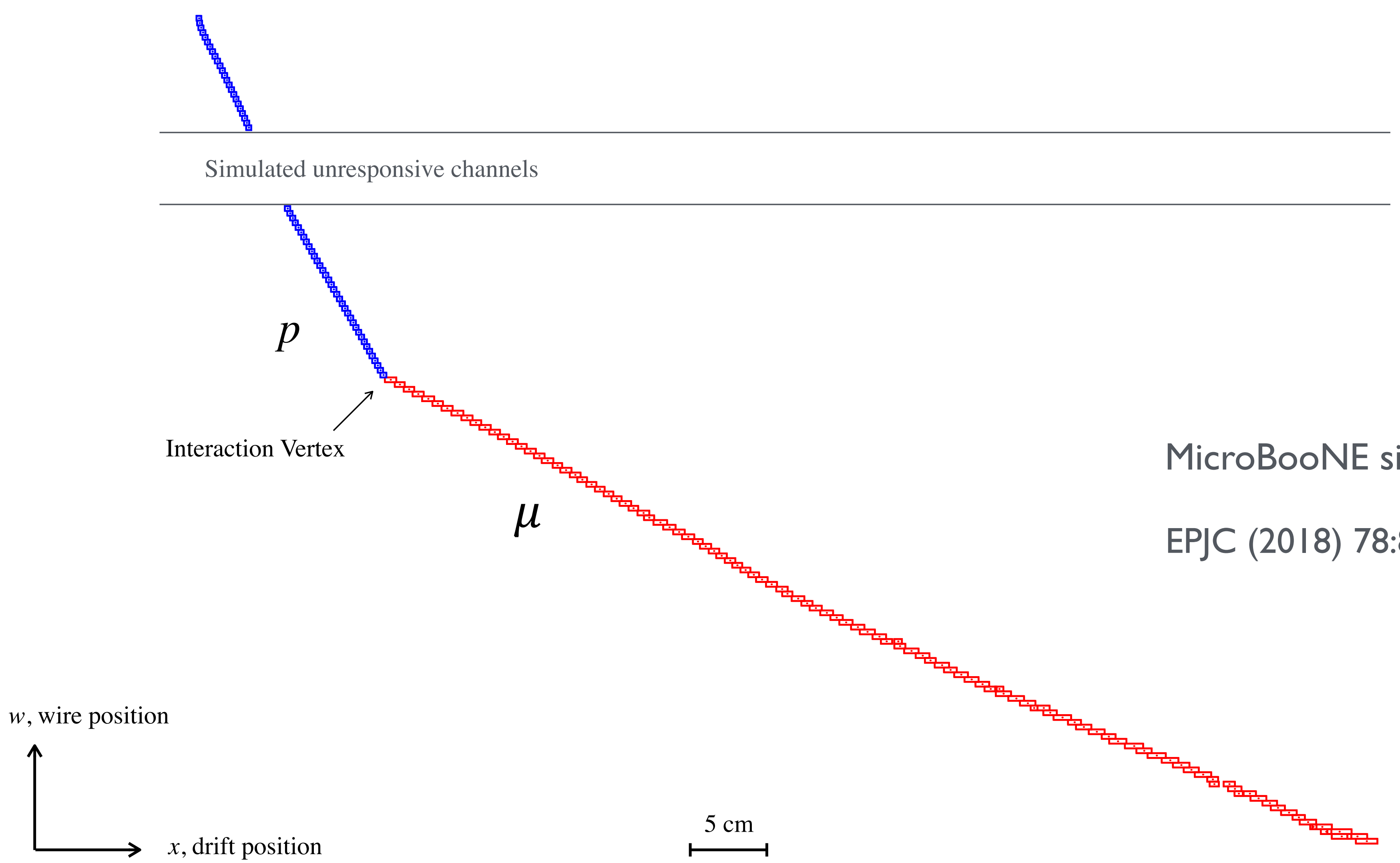
MicroBooNE - CC QE: $\nu_\mu + \text{Ar} \rightarrow \mu^- + p$

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.

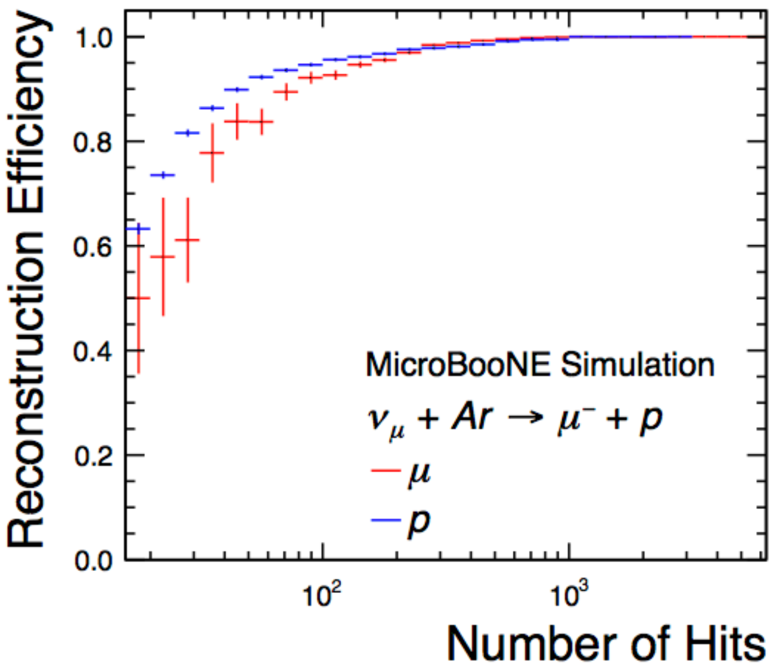


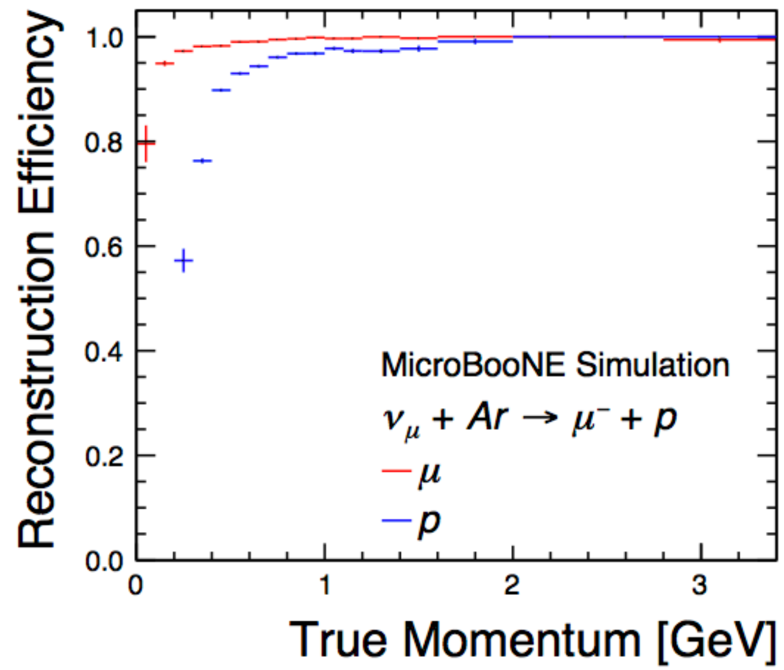
MicroBooNE simulation

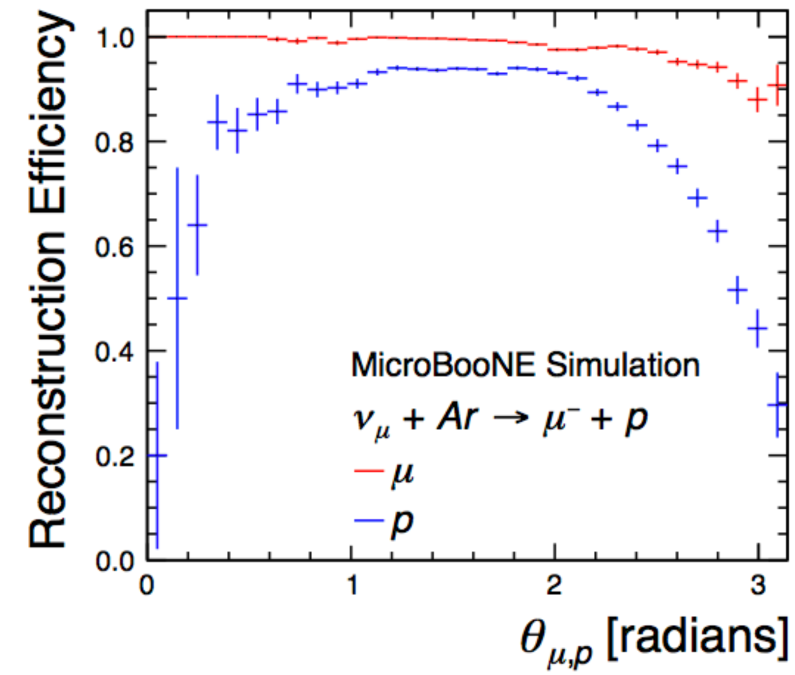
EPJC (2018) 78:82

520 MeV ν_μ

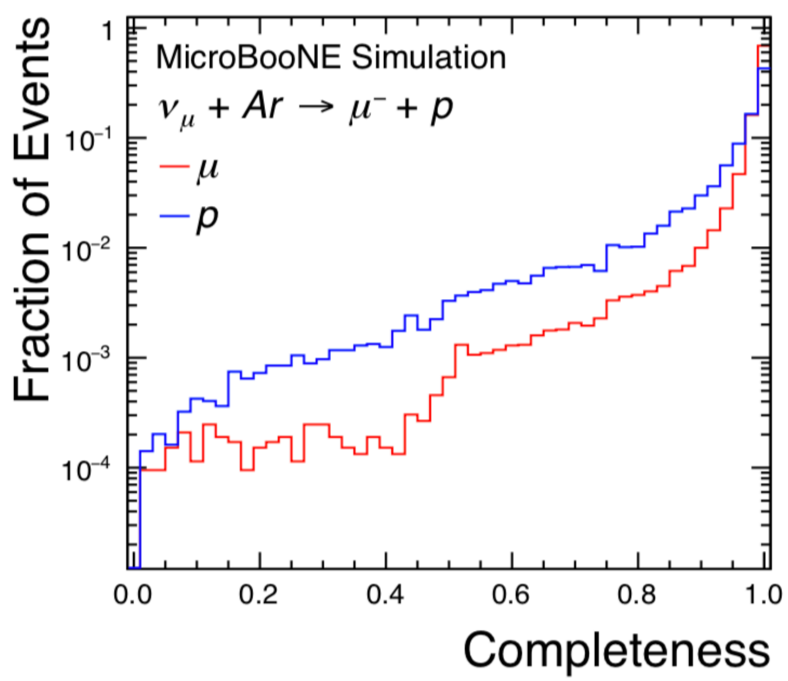
MicroBooNE - CC QE: $\nu_\mu + \mathrm{Ar} \rightarrow \mu^- + p$

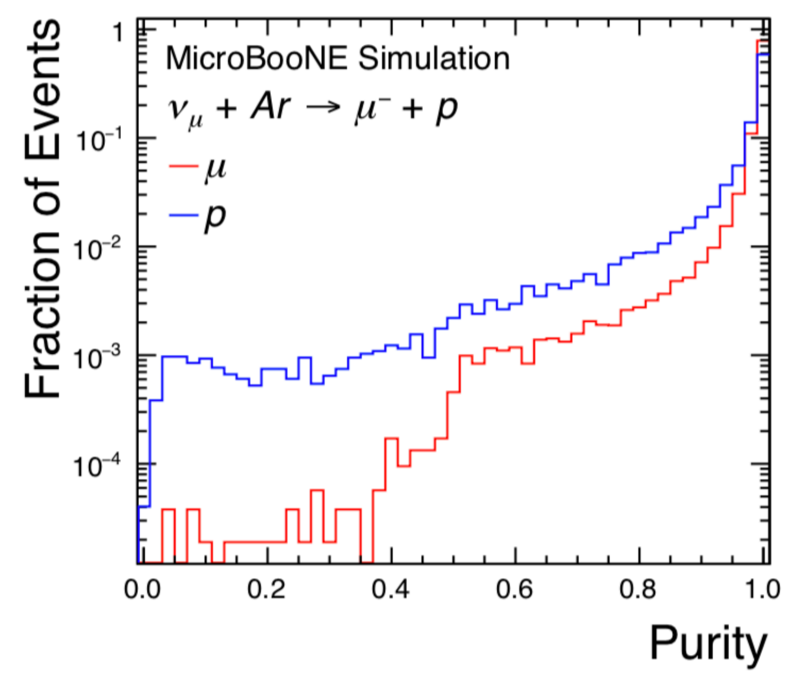






- The most common failure mechanism is merging muon and proton into a single reconstructed particle.
- Single particle is matched to target with which it shares most hits, which will preferentially be the muon.





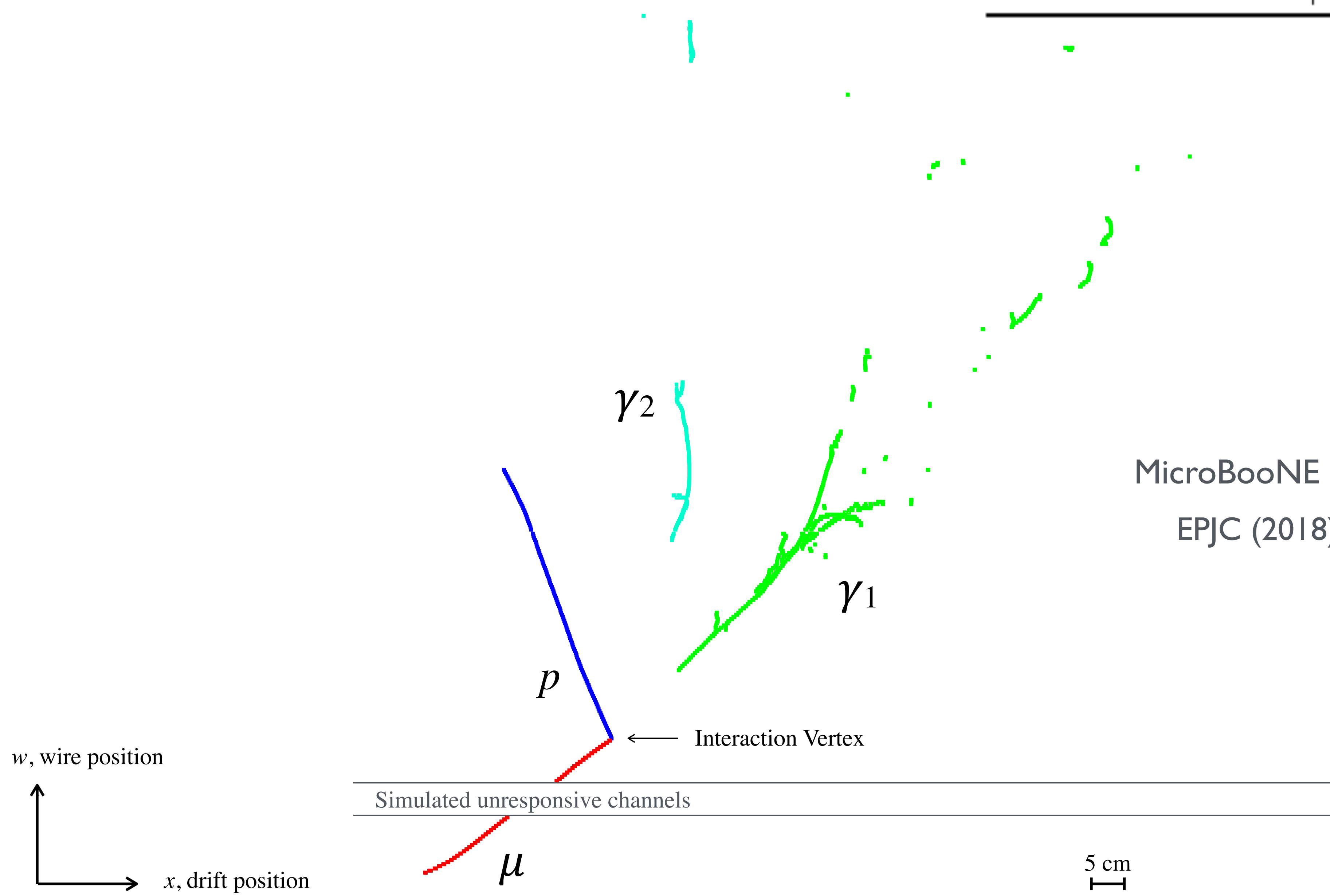
EPJC (2018) 78:82

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

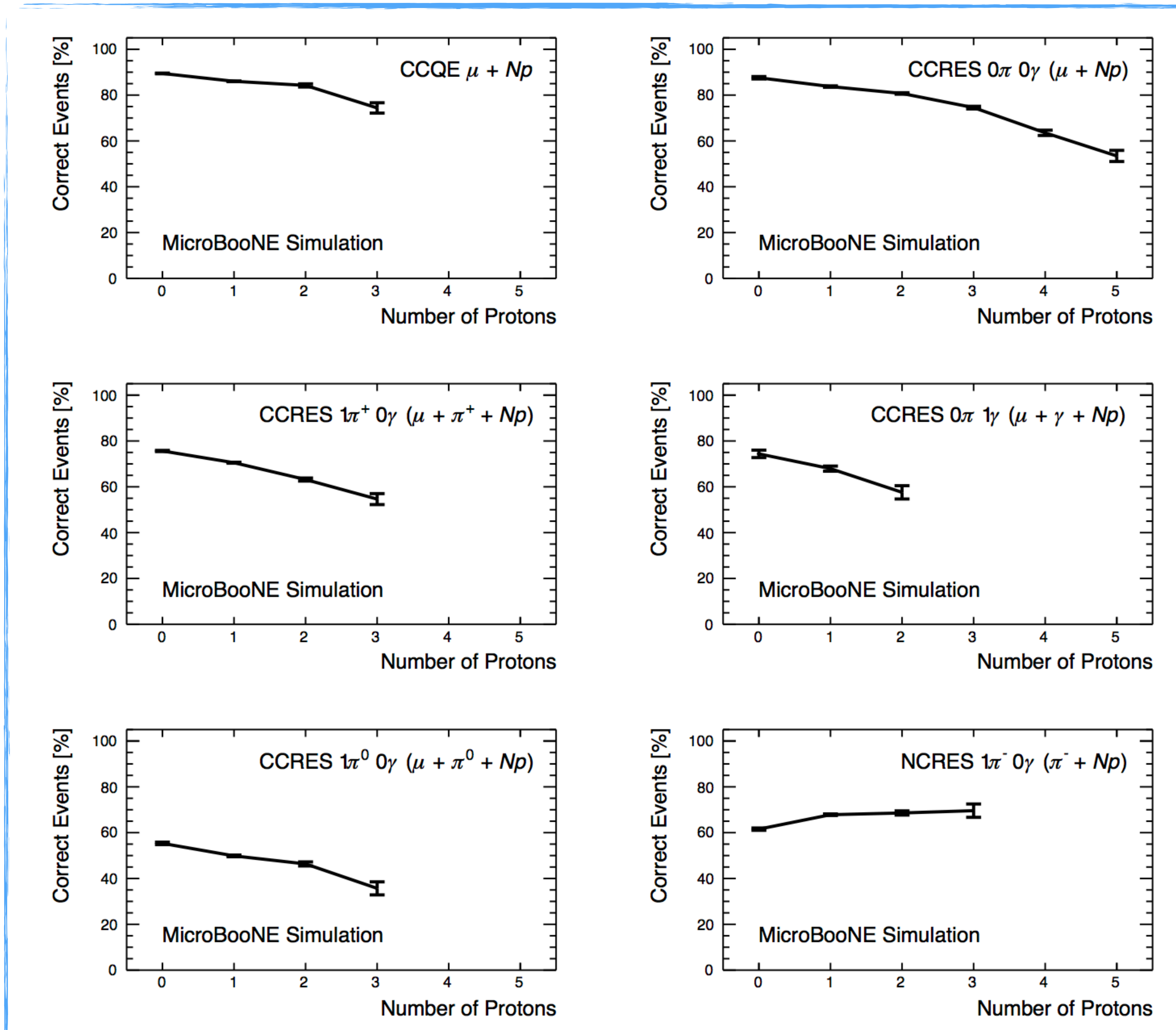


MicroBooNE simulation

EPJC (2018) 78:82

1.4 GeV ν_μ

- 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