# The Pandora multi-algorithm approach to pattern recognition in LArTPCs

16th September 2019

Reconstruction & Machine Learning in Neutrino Experiments workshop - DESY

**Andrew Smith - For the MicroBooNE & DUNE collaborations**

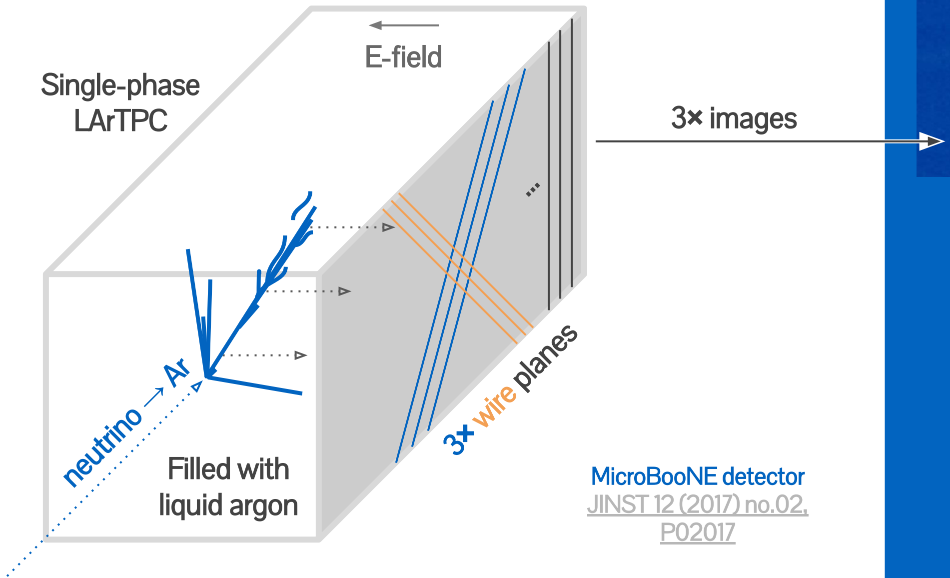
# Overview

- Hope to answer the following questions:
  - Where does pattern-recognition fit into the field of reconstruction, and why is it important?
  - What is Pandora's approach to pattern-recognition for Liquid Argon Time Projection Chamber experiments? - MicroBooNE, ProtoDUNE & DUNE
  - How can I get involved / learn more?

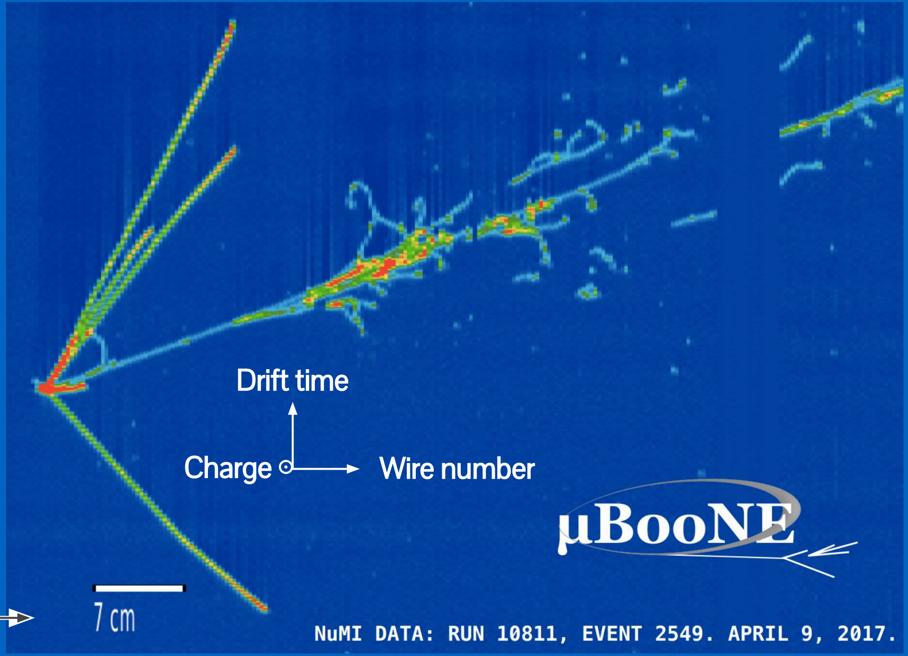
# LArTPC detectors

**LArTPC = Liquid Argon Time Projection Chamber**

- Neutrino interacts with Ar atom
- Resultant particles ionize Ar along their trajectory and also produce scintillation light
- Ionization charge drifts towards wires and produces a signal on all three planes



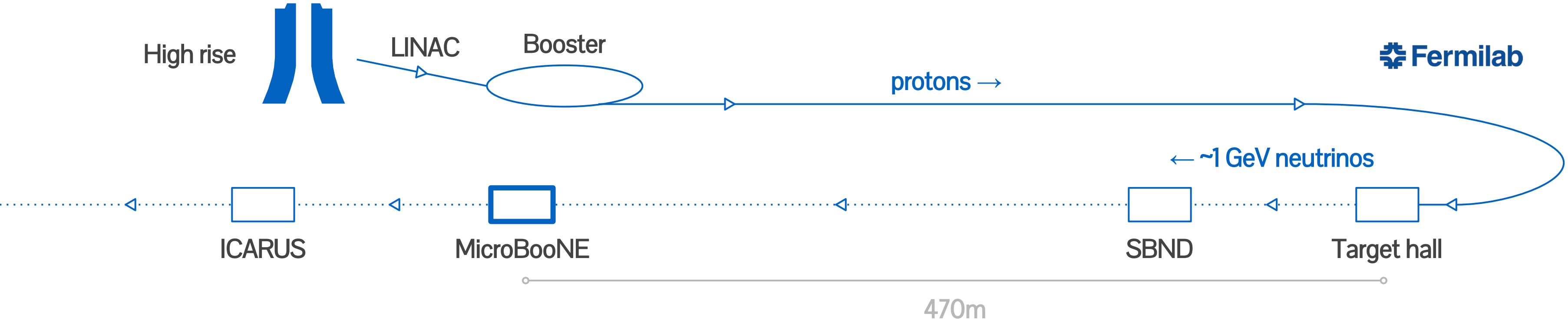

MicroBooNE detector
JINST 12 (2017) no.02, P02017

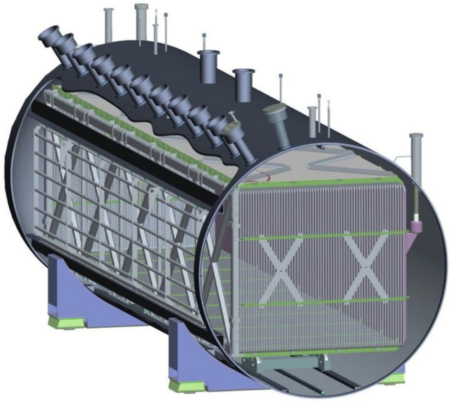



*A neutrino interaction image from one wire plane in MicroBooNE*

- Very high resolution calorimeter - millimeter-scale
- Can resolve individual particles down to low energies
- 3×2D views ⇒ 3D imaging

# The MicroBooNE experiment



### Experiment

- Collaboration of 179 scientists from 31 worldwide institutions
- Taking data since October 2015
- ~$10^5$ neutrino interactions in this time
- Additional off-axis neutrinos from second beam: NuMI

### Physics

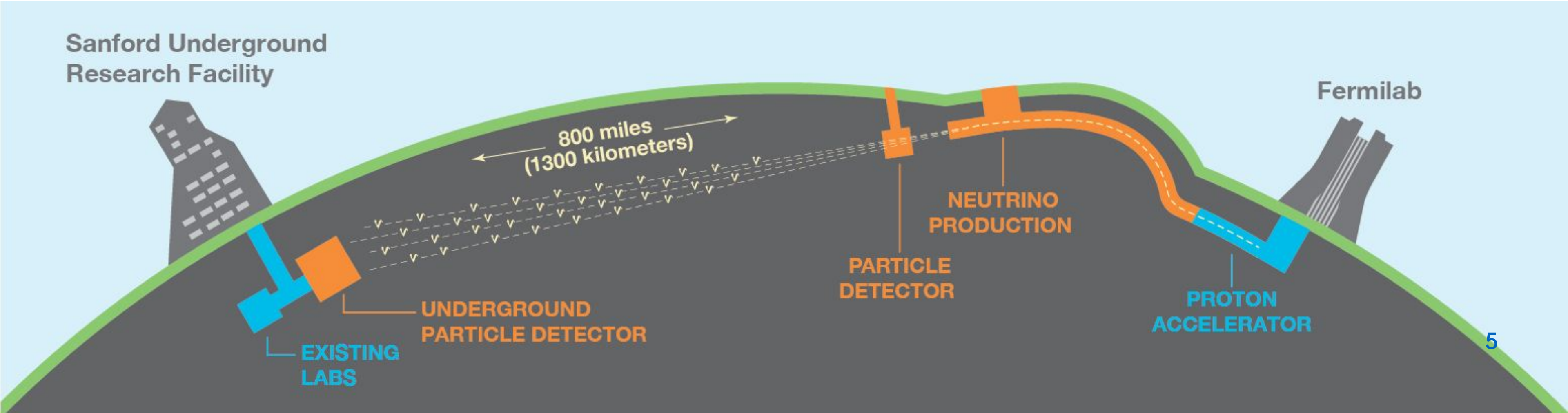
- Study previous anomalous results: excess of $\nu_e$ at low energies
- Measure suite of precision $\nu$ on Ar cross sections
- Hardware and software R&D for future experiments such as DUNE

# DUNE and ProtoDUNE

- **DUNE** = Deep Underground Neutrino Experiment
  - Groundbreaking for DUNE long baseline neutrino facility in 2017
- **ProtoDUNE single phase** = a prototype for DUNE
  - Based at CERN, first data 2018
  - Important test-beam data for DUNE
- Collaboration of over 1000 scientists from more than 30 countries

**Physics**

- Neutrino oscillation - Do neutrinos violate CP?
- Astroparticle physics, supernovae
- Proton-decay?



# From images to Physics

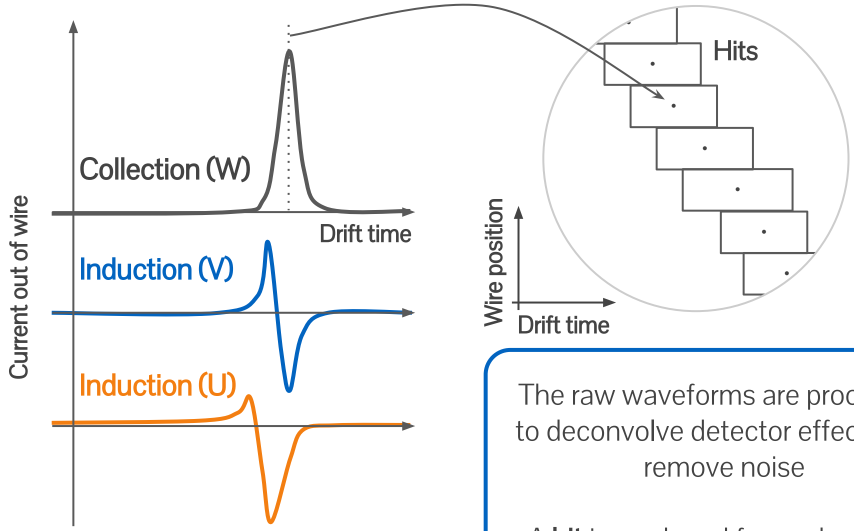
The reconstruction pipeline

→ Low-level image processing → Hit finding → **Pattern recognition** → Particle fits: Tracks, Showers → Calorimetric reconstruction → Particle Identification →

# From images to Physics

**The reconstruction pipeline**

Low-level image processing → Hit finding → Pattern recognition → Particle fits: Tracks, Showers → Calorimetric reconstruction → Particle Identification




*Cartoon wire responses to ionization charge drifting past a wire on induction and collection planes*

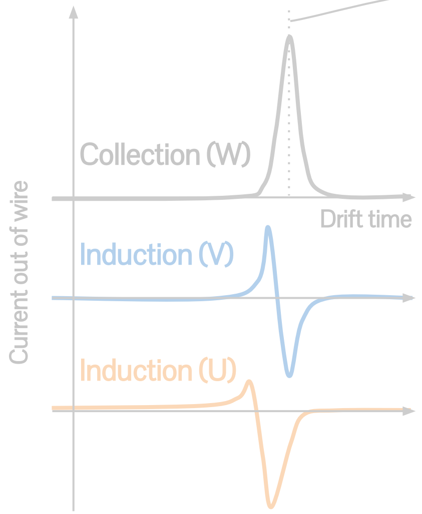
The raw waveforms are processed to deconvolve detector effects and remove noise

A **hit** is produced for peaks in the processed waveforms. These form the input to the patrec stage

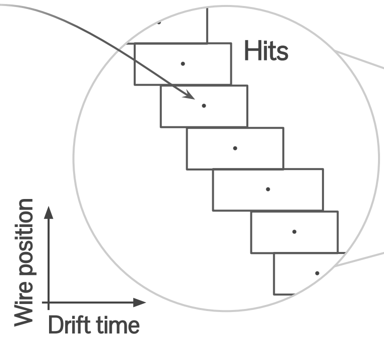
# From images to Physics

**The reconstruction pipeline**

Low-level image processing → Hit finding → **Pattern recognition** → Particle fits: Tracks, Showers → Calorimetric reconstruction → Particle Identification



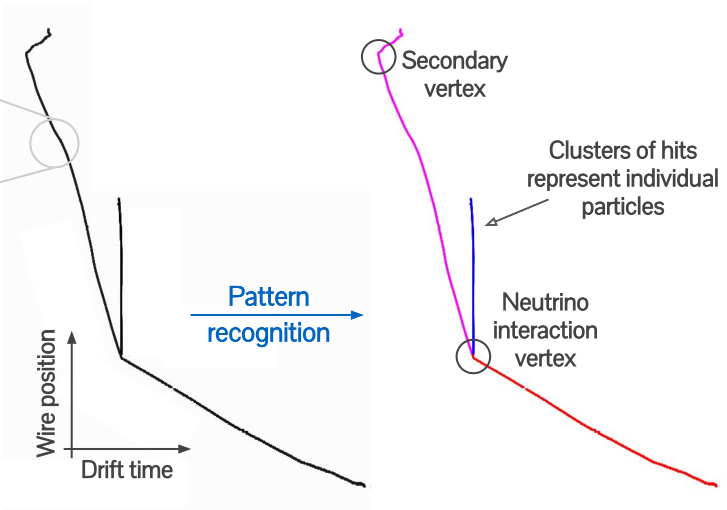

*Cartoon wire responses to ionization charge drifting past a wire on induction and collection planes*




The main job of the patrec is to:

Cluster the hits together to represent individual particles

Identify the hierarchical relationship between particles




*Hits from the collection plane for a simulated neutrino interaction in MicroBooNE, before and after patrec*

Eur.Phys.J. C78 (2018) no.1, 82

# The Pandora project

SDK

Application

Algorithms

Pandora

- General purpose open-source framework for pattern recognition
- Initially used for future linear collider experiments, but now well established on many LArTPC experiments too!
- Can be used standalone or as part of the LArSoft framework

| GitHub Repository | Software development kit | µBooNE Algorithms |
|---|---|---|
| github.com/PandoraPFA | Eur.Phys.J. C75 (2015) no.9, 439 | Eur.Phys.J. C78 (2018) no.1, 82 |

UNIVERSITY OF CAMBRIDGE

Lancaster University

WARWICK
THE UNIVERSITY OF WARWICK

µBooNE

Imaging Cosmic And Rare Underground Signals

SHORT-BASELINE NEAR DETECTOR SBND

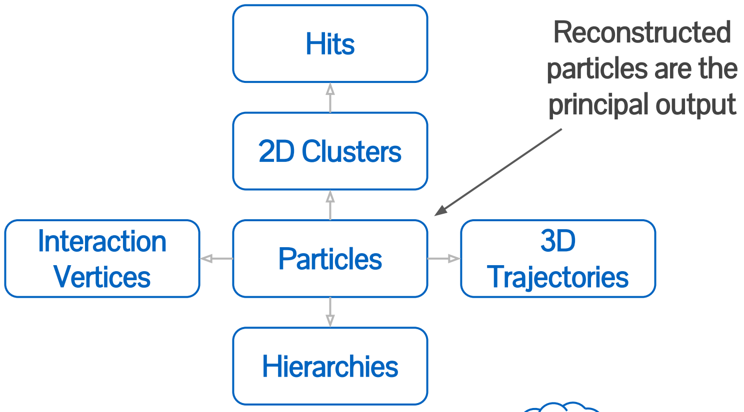
DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

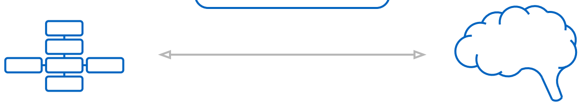
clic

ilc

# The event data model

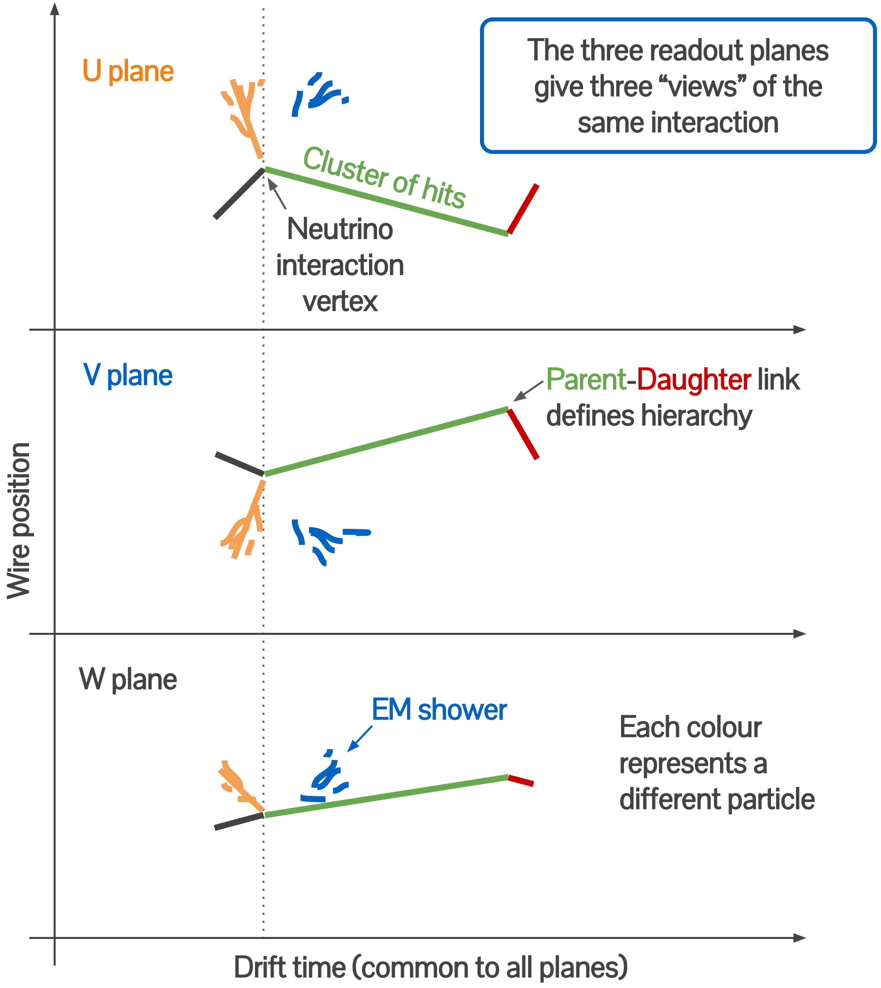
- We encapsulate our current understanding of the event via the **event data model**
- After the pattern recognition is finished, these are the objects which are available in LArSoft for downstream analysis



Event data model

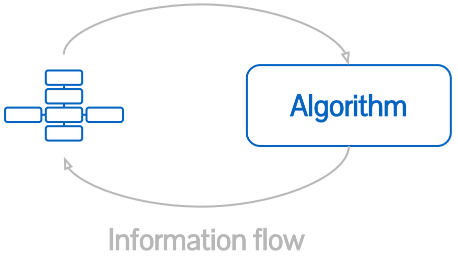
Understanding of the event




*A cartoon LArTPC pattern recognition problem*

# Pandora’s multi-algorithm approach

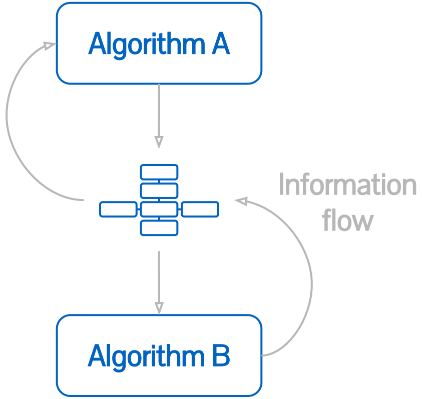
- Break the problem up into smaller well defined tasks and develop targeted algorithms for each task
    - E.g. Cluster together two hits if …




**Algorithms update our current understanding of the event by modifying the event data**

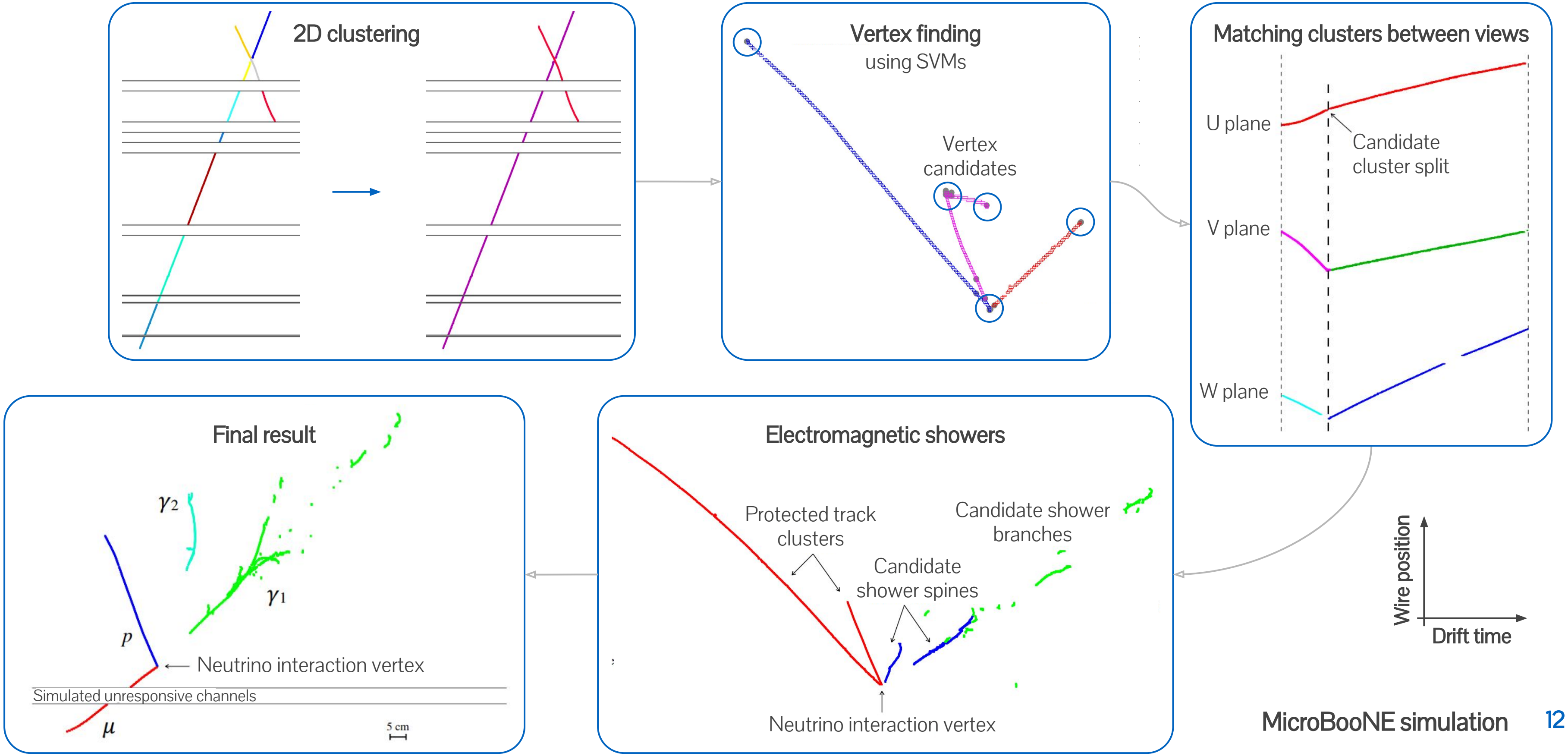
- Algorithm complexity varies from simple cuts up to more advanced machine learning techniques
- The application runs ~100 algorithms to gradually build our understanding until a complete picture of the event develops

- Iteration is used to allow 2-way information flow between algorithms




- Iteration provides powerful feedback loops - a technique that Pandora frequently utilizes

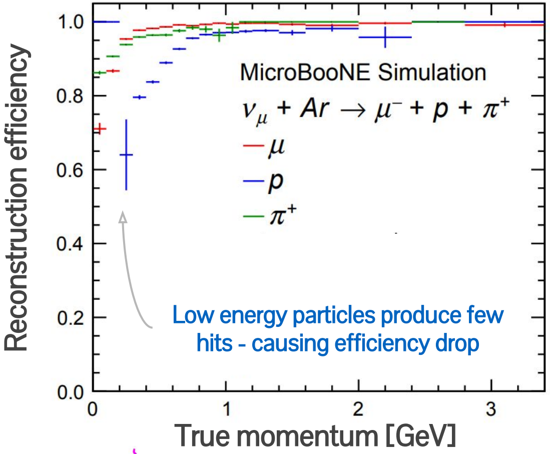
# Pandora’s algorithms for neutrino interactions

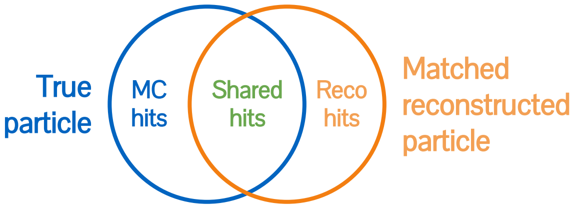



MicroBooNE simulation

# Performance - case study

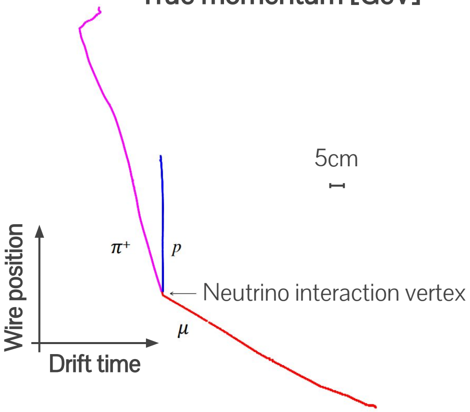
For performance in other channels, see:
[Eur.Phys.J. C78 (2018) no.1, 82]

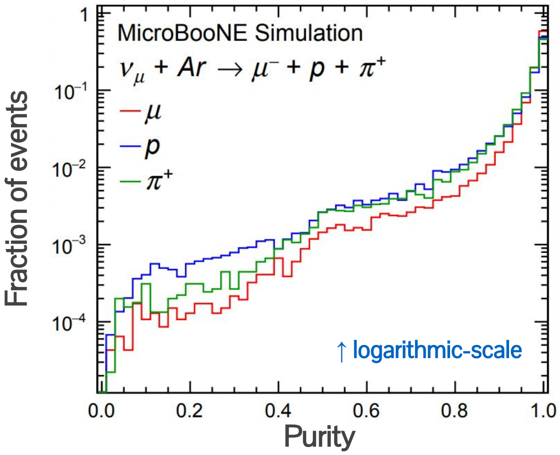


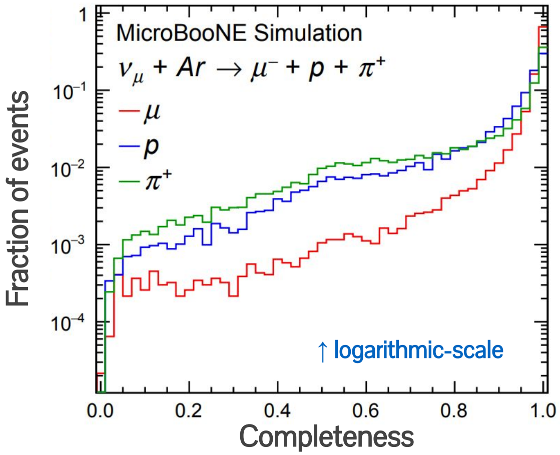




Purity = # Shared hits / # Reco hits

Completeness = # Shared hits / # MC hits



# Pandora's algorithm chains

## Neutrino / test-beam chain

- As described on previous slides, algorithms are designed for neutrino or test-beam interactions
- Identify the primary interaction vertex early in the pattern recognition to inform later algorithms
- Includes special chains of algorithms for electromagnetic showers

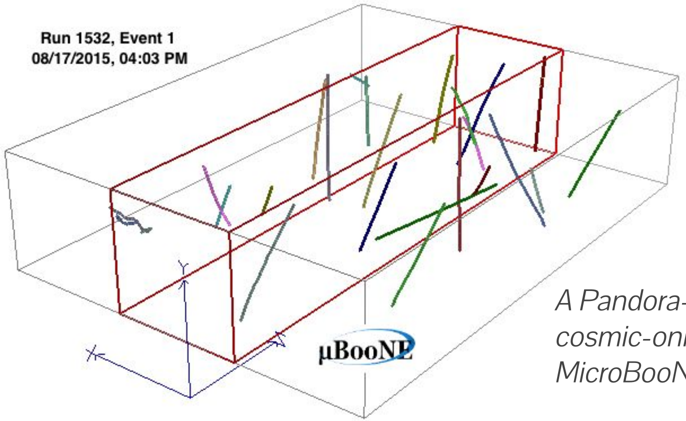
Each algorithm chain works well on the types of interactions it's designed for

For surface detectors, we need a way of dealing with events that contain **both** neutrino/test-beam interactions and cosmic-rays

**Solution**: "Consolidated approach"

## Cosmic-ray chain

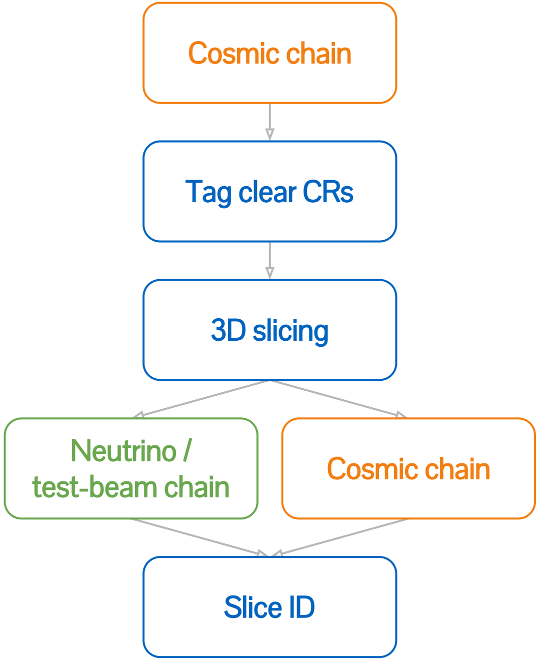
- Optimised to reconstruct cosmic-ray muons and daughter delta-rays
- More strongly track-oriented than the neutrino / test-beam chains
- Includes algorithms to identify and reconstruct delta-rays of energetic cosmic-rays



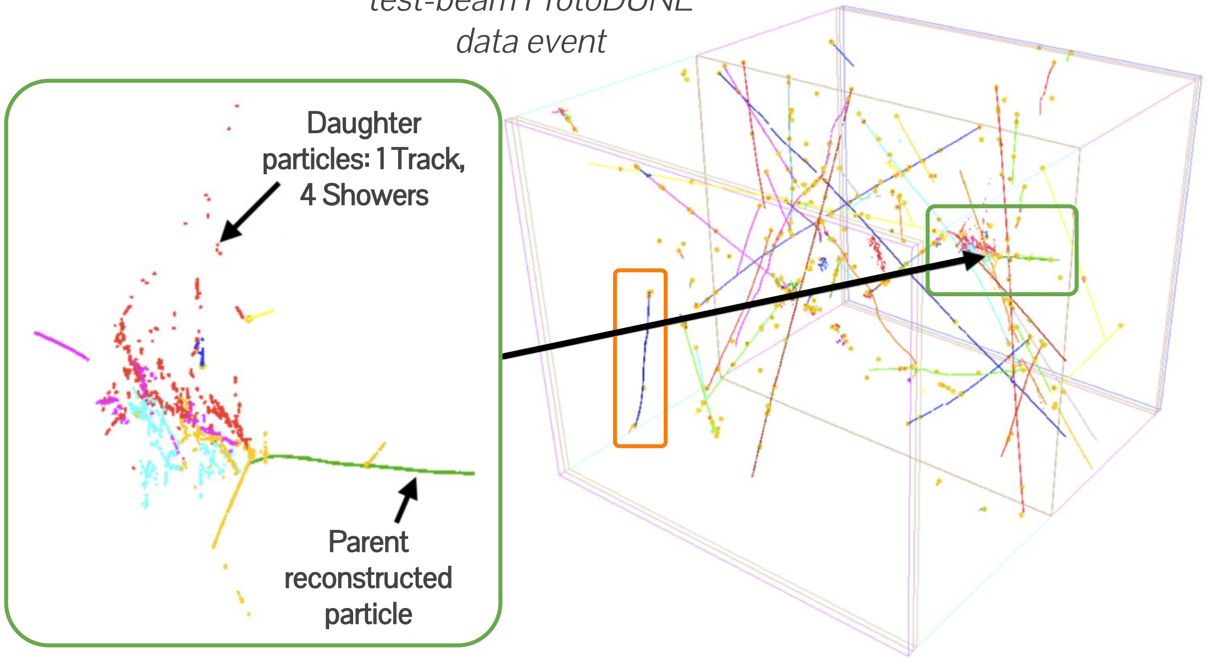

*A Pandora-reconstructed cosmic-only data event in MicroBooNE*

# Pandora's consolidated approach

- Use two different algorithm chains to handle **neutrinos / test-beam** interactions and **cosmic-rays**



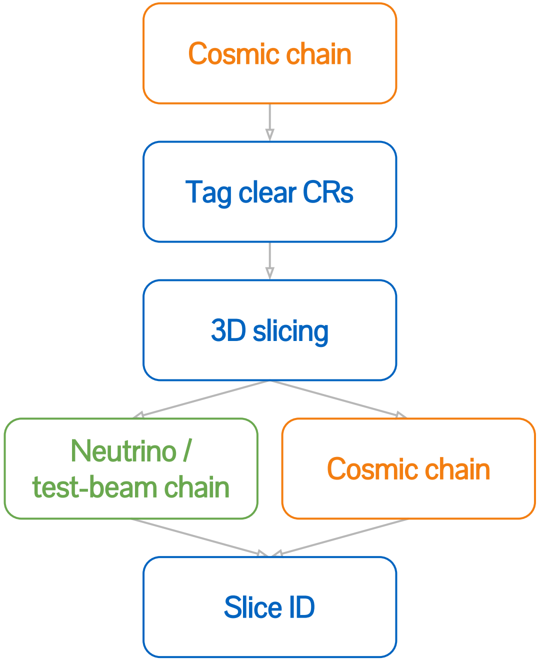

*A Pandora-reconstructed test-beam ProtoDUNE data event*



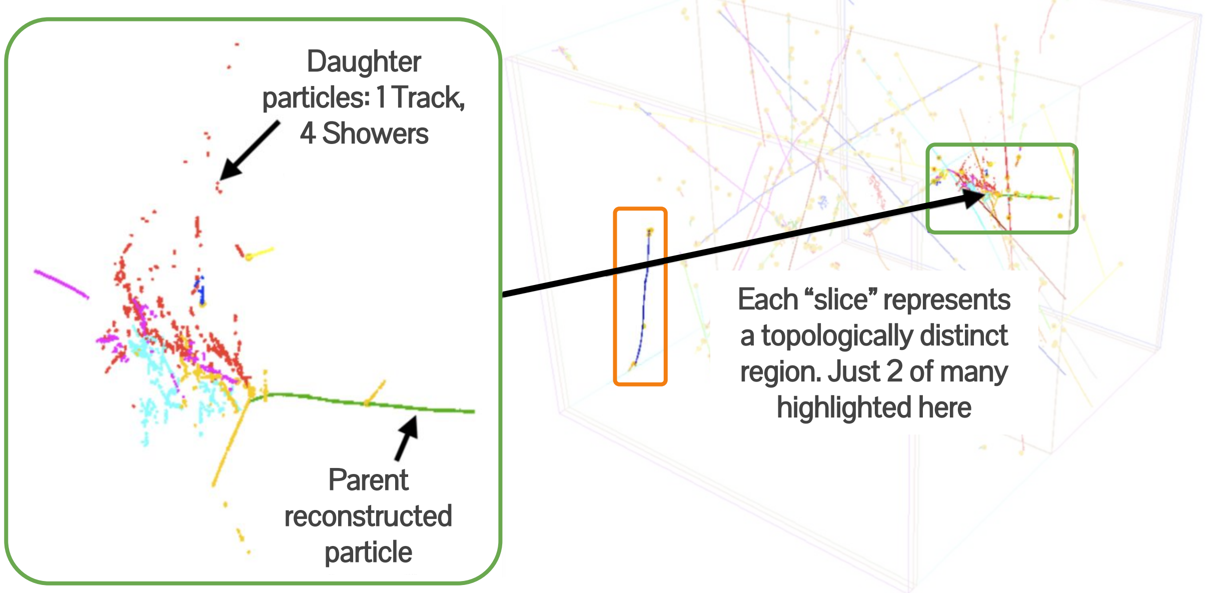

*"Slice" is reconstructed using the test-beam algorithm chain*

# Pandora's consolidated approach

- Use two different algorithm chains to handle **neutrinos / test-beam** interactions and **cosmic-rays**




*A Pandora-reconstructed test-beam ProtoDUNE data event*




*"Slice" is reconstructed using the test-beam algorithm chain*

# Slice identification

- Each experiment uses its own slice ID logic. When integrated into the LArSoft framework, we have access to all detector subsystems (not just the TPC) to make this decision. Depending on the experiment this includes:

**Time projection chamber (TPC)**
Use topological features to identify the correct slice

**Light detection systems**
Much faster timing resolution than TPC.
Match TPC to light signals to exploit this information

**Cosmic-ray tagging (CRT) system**
Match signals between external CRTs and TPC and use their timing resolution to identify cosmic-rays

# Summary

- LArTPCs are capable of imaging neutrino interactions with very high resolution

- Pandora provides a robust, general purpose solution to pattern-recognition to exploit these rich images
    - Our approach is to use many targeted algorithms to build an understanding of the event
    - Some problems are best solved using more traditional hand-designed algorithms, others are great candidates for machine learning techniques
    - We aim to carefully exploit the most appropriate solution for each problem
    - Capable of handling complex neutrino interaction topologies in dense cosmic-ray environment

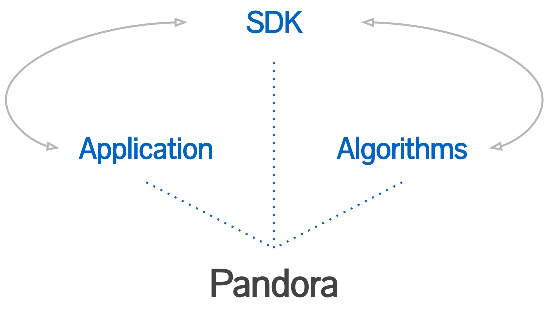
- This approach has been used successfully in MicroBooNE and ProtoDUNE, with continued active developments that will see Pandora used for DUNE

# Further information and tutorials

# Papers and documentation

The Pandora SDK paper details the design of the software development kit and how algorithms interface with the application that is running Pandora

See the next slide for information about using Pandora's outputs within the LArSoft framework (application)




The Pandora MicroBooNE paper gives details of Pandora's algorithms in MicroBooNE at the time of publication, but generally applicable to other LArTPC experiments too

All Pandora code is self-documented using doxygen and is available on github

https://github.com/PandoraPFA

**Software development kit**
Eur.Phys.J. C75 (2015) no.9, 439

**μBooNE Algorithms**
Eur.Phys.J. C78 (2018) no.1, 82

# Previous workshops & hands-on exercises

**Links on this page go to the slides for the past workshops**

- Multi-day Pandora workshop in Cambridge, UK - 2016
  - Talks about how the algorithms work and step-by-step exercises about how you might develop a new algorithm using Pandora!

- LArSoft workshop in Fermilab - 2019
- LArSoft workshop in Manchester, UK - 2018
- Workshop on advanced computing & machine learning, Paraguay - 2018
  - Talks and exercises about running and using Pandora within LArSoft, including tutorials on using Pandora's custom event display

- Experiment specific resources:
  - ProtoDUNE analysis workshop, CERN - 2019
  - MicroBooNE Pandora workshop, Fermilab - 2018

# Pandora team for LArTPC 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<br>Mark Thomson | john.marshall@warwick.ac.uk<br>thomson@hep.phy.cam.ac.uk |
| **LArTPC algorithm development** | John Marshall<br>Andy Blake | <br>a.blake@lancaster.ac.uk |
| **MicroBooNE integration**<br>**ProtoDUNE integration**<br>**DUNE FD integration** | Andy Smith<br>Steven Green<br>Lorena Escudero | asmith@hep.phy.cam.ac.uk<br>sg568@hep.phy.cam.ac.uk<br>escudero@hep.phy.cam.ac.uk |
| **Graduate students** | MicroBooNE<br>ProtoDUNE<br>DUNE | Joris Jan de Vries, Jack Anthony, Andy Smith<br>Stefano Vergani<br>Jhanzeb Ahmed, Mousam Rai, Ryan Cross |

github.com/PandoraPFA

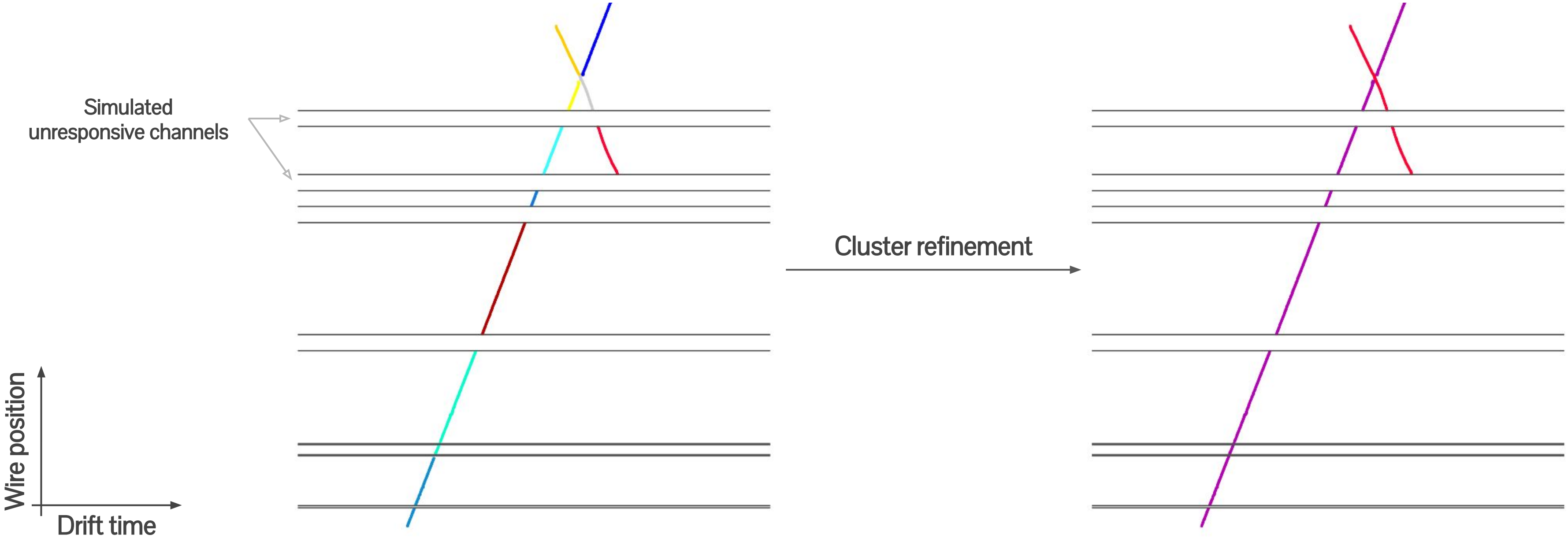
PandoraPFA.slack.com

# Backup

# 2D clustering

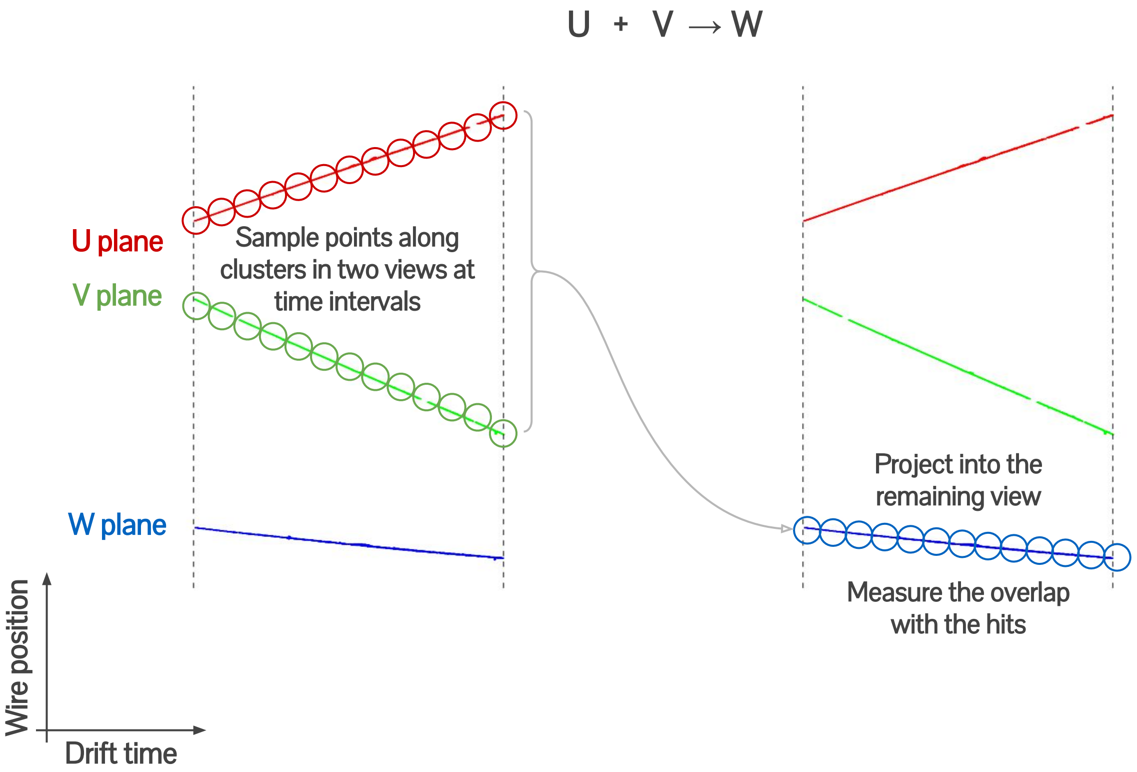
MicroBooNE simulation

- Start by producing clusters with high purity, stopping at any bifurcation / ambiguity
- Run multiple cluster merging & splitting algorithms to improve completeness
    - Consider all cluster pairings simultaneously to make decisions in the context of the whole event



# Matching clusters between views

MicroBooNE simulation




- For every combination of clusters - one from each view - calculate a $\chi^2$ which measures their consistency

Overlaps

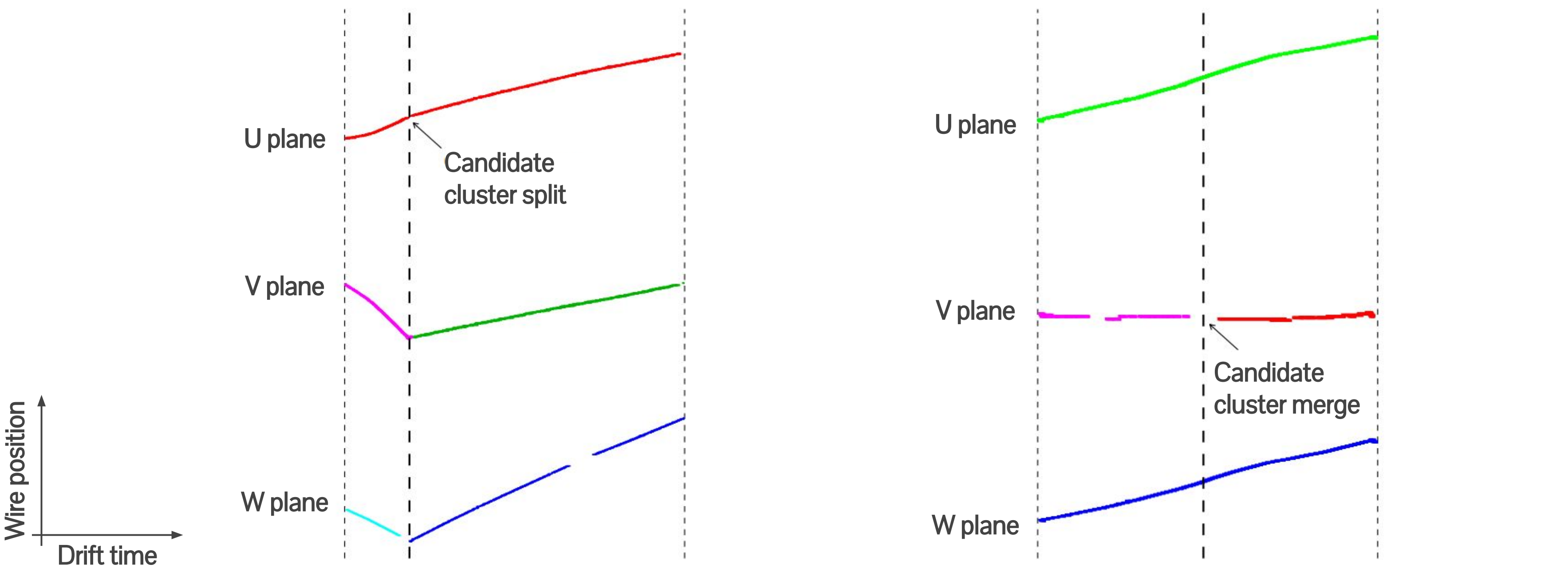
$$\left.\begin{array}{l} U + V \rightarrow W \\ V + W \rightarrow U \\ W + U \rightarrow V \end{array}\right. \Rightarrow \chi^2$$

- Construct a rank-3 “tensor” with elements $\chi^2_{uvw}$, where u, v & w label the clusters from each view
- The more “diagonal” the tensor, the better we have clustered the hits

⇒ Iteratively run algorithms which modify the clusters to diagonalise the tensor!

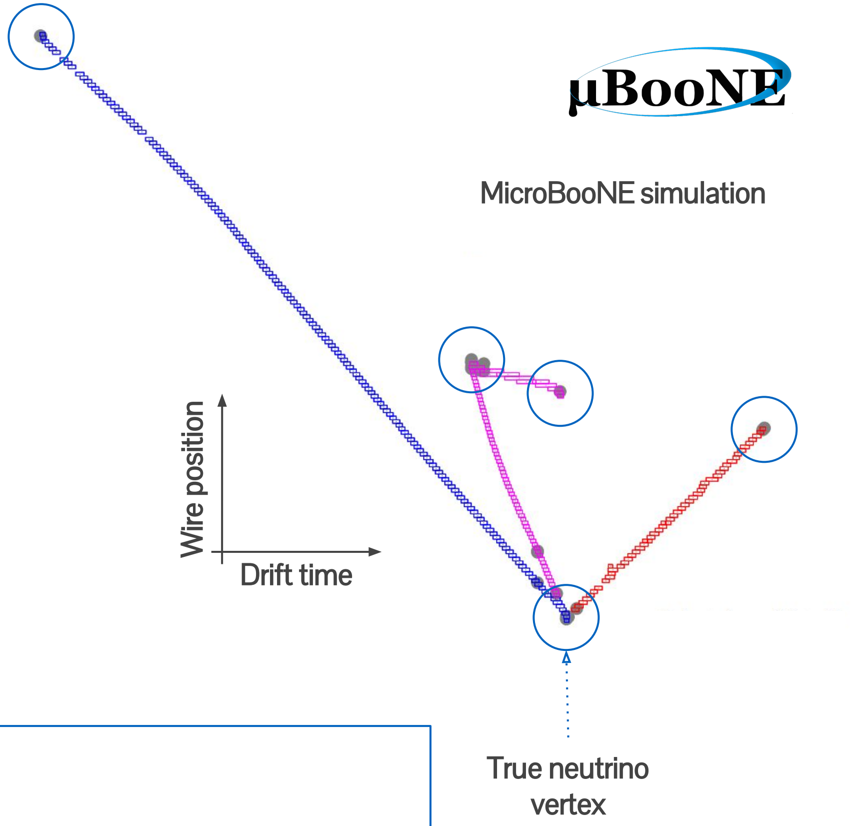
# Diagonalising the tensor

- Run many algorithms iteratively that look to group / split / merge clusters at positions motivated by clusters in the other views
- Keep going until no further refinements can be made and the tensor is as diagonal as possible!



# The neutrino vertex

- Consider all 2D cluster endpoints, from all views - and use the matching clusters in other views to determine the corresponding 3D points
- Use these 3D points as candidate neutrino interaction vertices
- Use a number of topological and calorimetric features, fed into a support vector machine (SVM) to get a score for each candidate
- Choose the candidate with the highest SVM score!

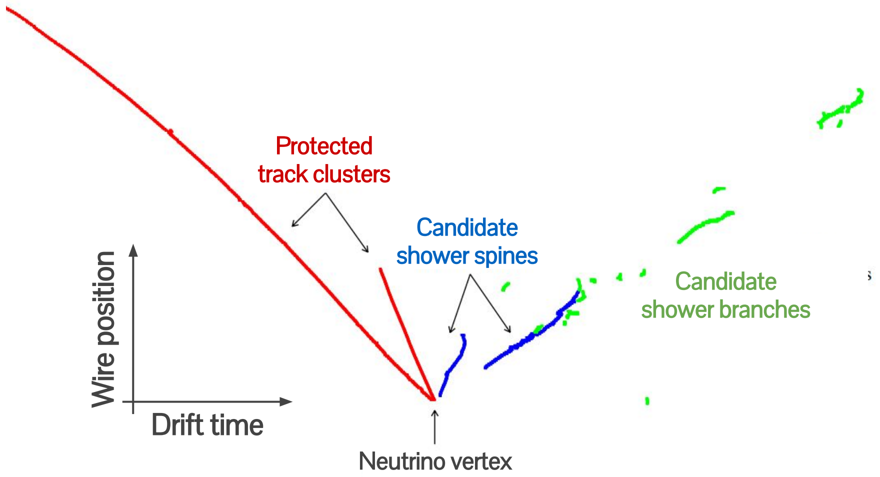


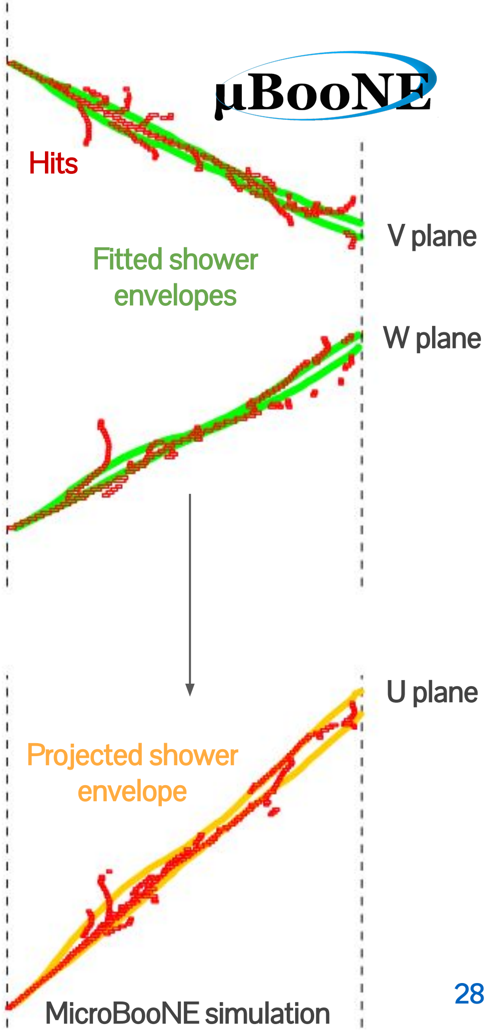

**Downstream usage**

- Split 2D clusters at the projected vertex - helps identify low-energy protons
- Protect primary particles when searching for EM showers in later steps

# Electromagnetic showers

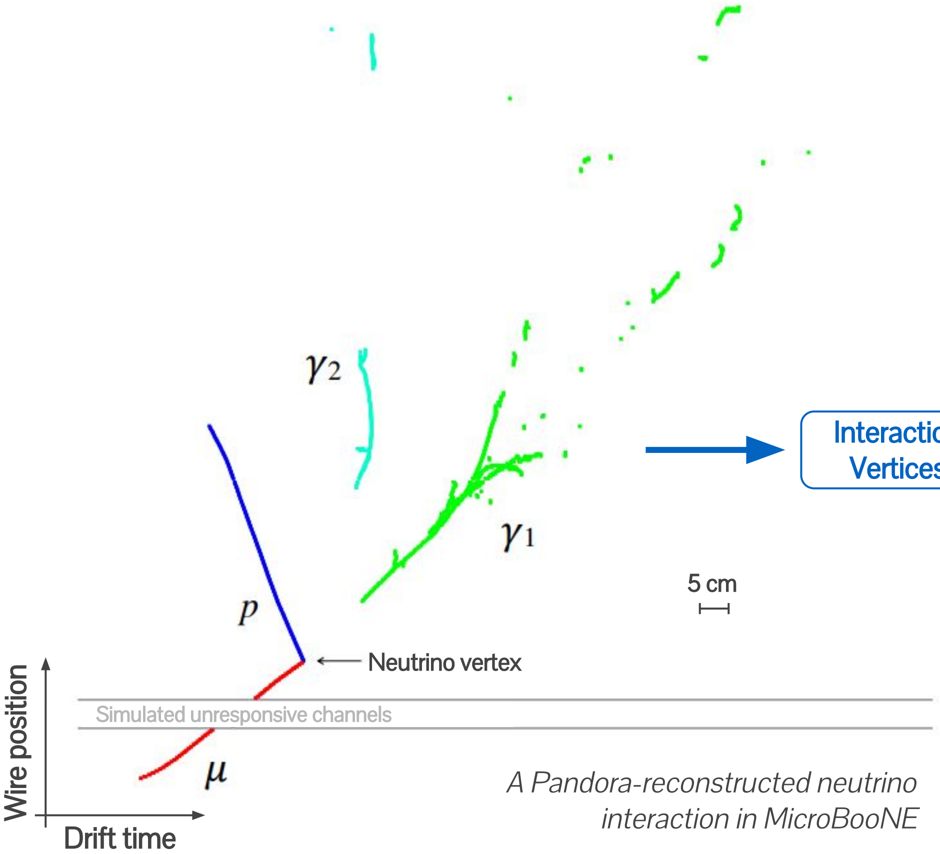
- Characterise 2D clusters as track-like or shower-like using topological features
- Find shower "spines" - protecting track-like clusters from the vertex
- Grow the showers by merging the spines with shower-like branch clusters
- Re-use the "tensor" mechanic to match the showers between views, and iterative through another set of algorithms to diagonalise the tensor

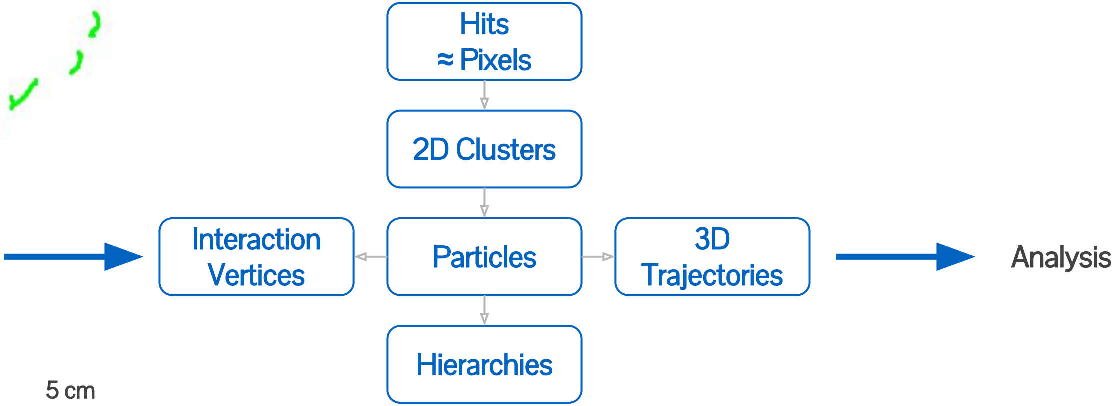


# The final product!

MicroBooNE simulation




*A Pandora-reconstructed neutrino interaction in MicroBooNE*