A novel approach in the Pandora multi-algorithm reconstruction to tackle challenging topologies at DUNE

Maria Brigida Brunetti on behalf of the DUNE Collaboration

Neutrino Physics and Machine Learning 2024, 27 June 2024, ETH Zurich

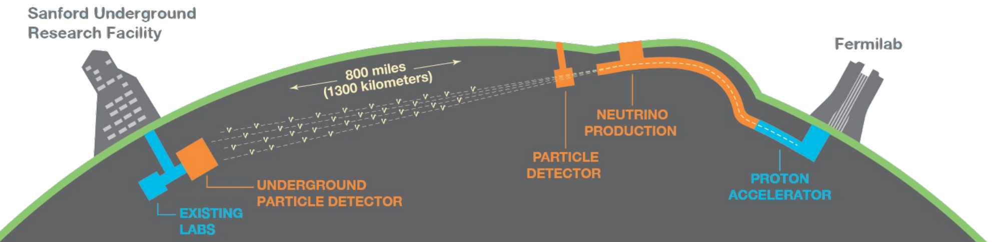
The Deep Underground Neutrino Experiment

- Long baseline (1285 km) oscillation experiment
- 2 MW → 2.4 MW beam at Fermilab (**most intense ν beam in the world**)
  - Wide-band energy
- Liquid argon time projection chamber technology
- Near detector at Fermilab
- Four 17 kton Far Detector modules at SURF
- 1.5 km underground location

High-precision measurements
of neutrino mixing and oscillation fundamental parameters (including CPV)

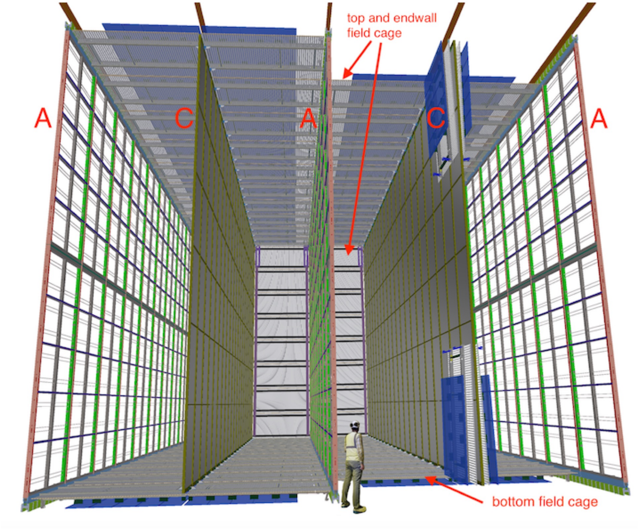
Astrophysical neutrinos
Supernova and solar neutrinos

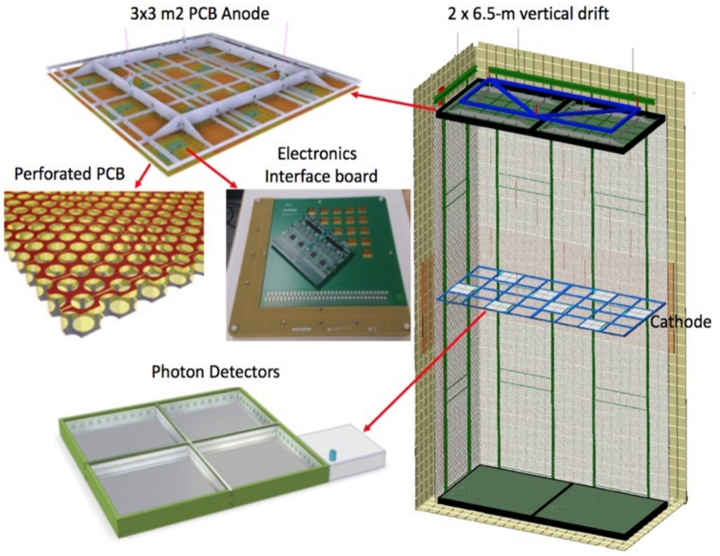
Probe new physics
including nucleon decay



The Horizontal and Vertical Drift Far Detector

TPC size: 12.0m × 14.0m × 58.2 m. Drift length: 3.5 m



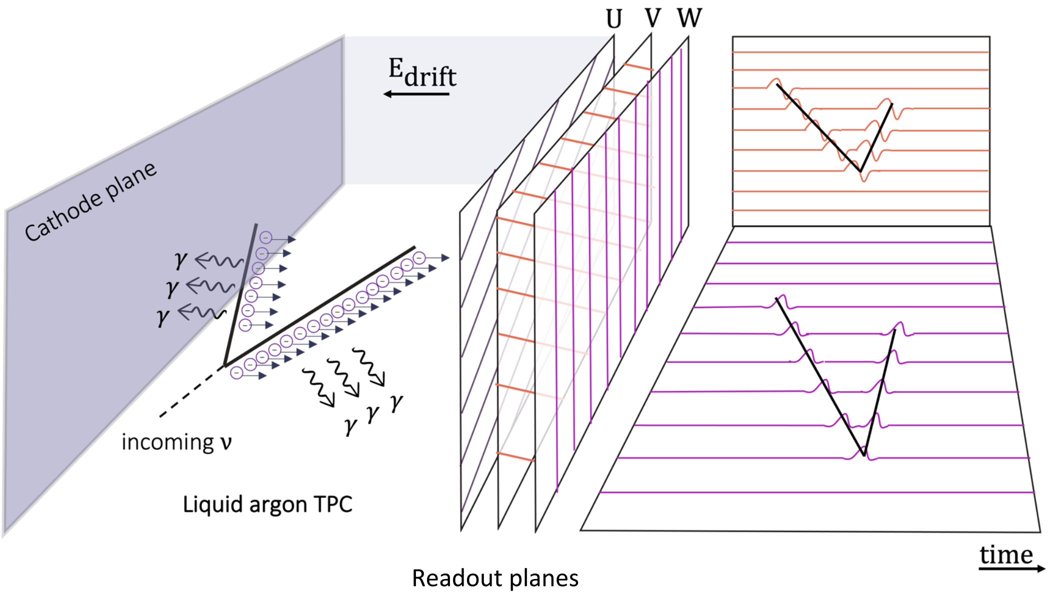


- Modular wire-based charge readout
- 4 drift volumes defined by 5 arrays of anode and cathode planes

- PCB-based charge readout
- 2 drift volumes defined by a cathode plane, and 2 PCB-based anode planes

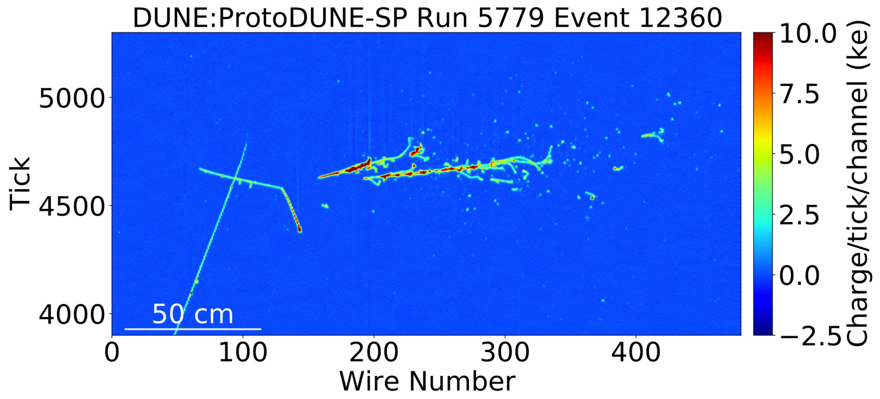
Liquid Argon Time Projection Chamber (LArTPC)

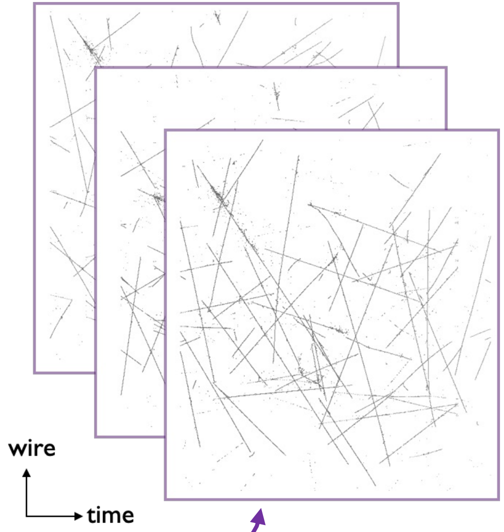
- Use scintillation and ionization to find 3D position of particles and interactions
- Drift charge recorded by several readout planes, with different orientations, forming images
- Light collected by photon detection system



LArTPC images

Exquisite **tracking** and **calorimetry** **capability**





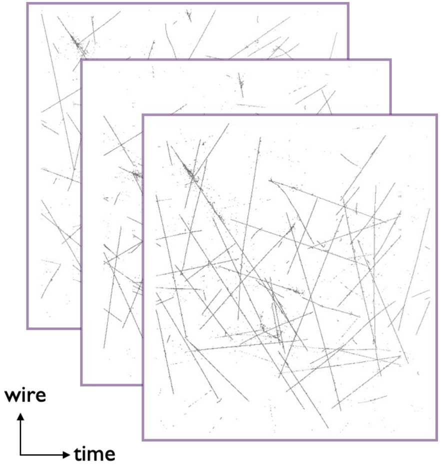
Signal processing and hit finding steps to go from raw images to **pattern recognition inputs**

3 x 2D collections of hits

hit = energy measurement on a given wire, at a given time

Pandora reconstruction multi-algorithm approach

- Reconstruction framework in use at all neutrino LArTPC experiments, and applied at linear colliders
- Many logical steps (> 100 algorithms) to go from input hits to 3D hierarchies
- Build different techniques, including deep learning, and physics and detector knowledge in the pattern recognition algorithms



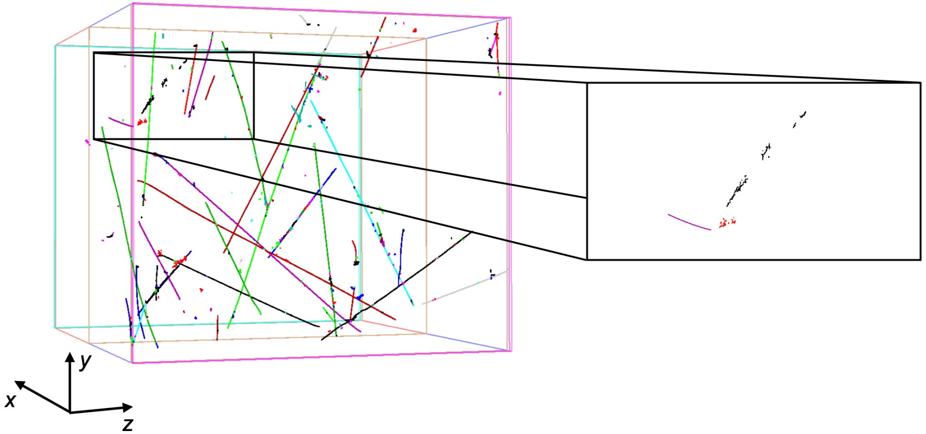
2D pattern recognition

2D → 3D matching

Vertex finding

Hierarchy building

Track/shower ID

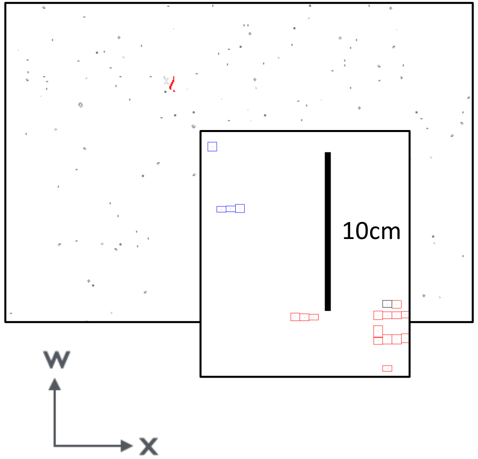


Machine learning in Pandora

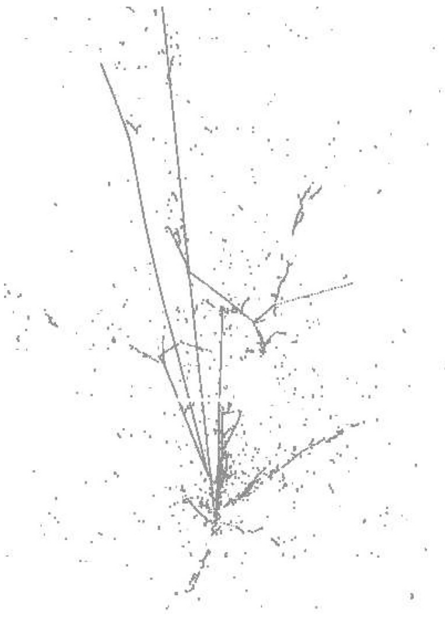
- Very diverse event topologies motivate exploiting different techniques
- Broad use of machine learning, and especially Deep Learning (DL)

Highlights:

- Neutrino interaction vertex finding in DUNE FD (see Andy Chappell's talk)
- Neutrino interaction vertex finding in the DUNE ND (AIDAinnova)
- Neutrino interaction vertex finding and background rejection in low-energy neutrino signatures
- Hierarchy building (see Isobel Mawby's talk)
- **New reclustering approaches (this talk)**



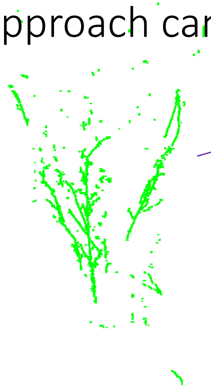
30 MeV
ν_e CC CEvNS



26.8 GeV
ν_μ CC DIS

Reclustering approach

- Overlapping showers are challenging to reconstruct, and impact DUNE physics goals
  - Showers from π^0 mistaken as single electron are background in appearance analyses
- Tackle via new reclustering approach (STFC-funded project)
- A similar approach was successfully used at [linear colliders]
- Approach can be applied to other topologies, e.g. including tracks



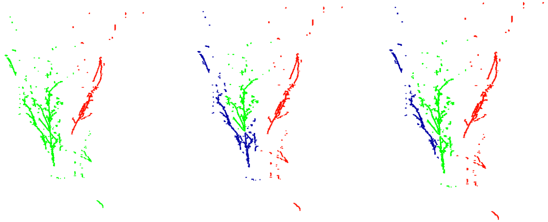
Consider particles **at the end of the reconstruction chain**

Did we make a mistake?

Does something look wrong? **Is a better reconstruction outcome based on the 3D hits possible?**

Redistribute 3D hits

Run **multiple algorithms** proposing alternative clustering of the hits



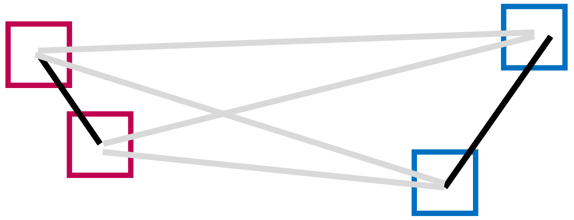
Pick best outcome

Clustering 3D hits with Graph Neural Networks

- We are investigating DL techniques, such as transformer networks, to identify target topologies
- In this presentation, will assume that target topologies have already been identified
- Start from 3D hits, aim at proposing new possible clustering outcomes

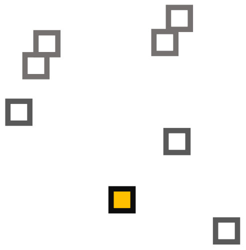
Graph Neural Networks (GNNs) motivation

- Straightforward representation of a 3D cluster of hits as a graph:
  - One 3D cluster = 1 graph
  - Nodes = hits
  - Node features = hit *x*, *y*, *z*, charge
- Can imagine pairs of hits to be connected by a positive edge if most energy was deposited by the same true particle



Grey line = false edge
Black line = true edge

Message-passing layers

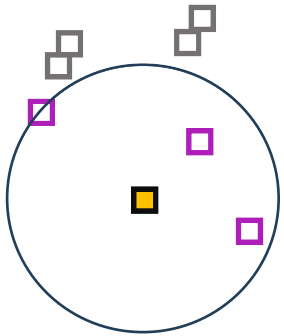


Leverage GNN message-passing so that local features know about broader cluster structure

Using GraphSAGE architecture

- For each layer, aggregate nearby nodes features to node under consideration (e.g. via averaging)
- Only consider nearby hits within radius, e.g. 10 cm
- Stacking together many layers means concatenating features from further away

Message-passing layers

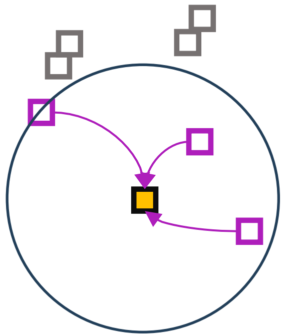


Leverage GNN message-passing so that local features know about broader cluster structure

Using GraphSAGE architecture

- For each layer, aggregate nearby nodes features to node under consideration (e.g. via averaging)
- Only consider nearby hits within radius, e.g. 10 cm
- Stacking together many layers means concatenating features from further away

Message-passing layers

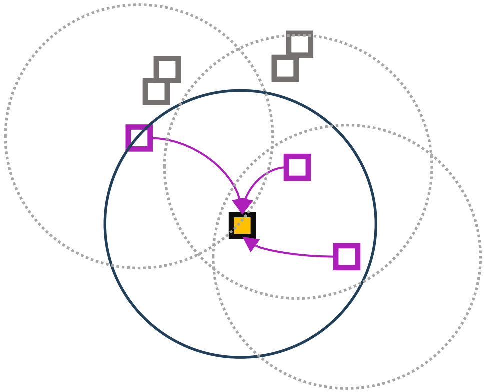


Leverage GNN message-passing so that local features know about broader cluster structure

Using GraphSAGE architecture

- For each layer, aggregate nearby nodes features to node under consideration (e.g. via averaging)
- Only consider nearby hits within radius, e.g. 10 cm
- Stacking together many layers means concatenating features from further away

Message-passing layers



Leverage GNN message-passing so that local features know about broader cluster structure

Using GraphSAGE architecture

- For each layer, aggregate **nearby nodes** features to **node under consideration** (e.g. via averaging)
- Only consider nearby hits within radius, e.g. 10 cm
- Stacking together many layers means concatenating features from further away

Two GNN-based approaches for clustering

A. Supervised edge prediction:
Find all pairs of hits
that are truly connected
(this talk)

B. Unsupervised clustering
Assign a cluster label to each hit
Future development

Message-passing layers

Linear predictor layer

Linear transformation + softmax

	A	B
Output:	Per-hit-pair predicted edge score	Per-hit cluster assignment
Loss function:	Binary cross-entropy	Non-task specific minCUTpool losses
Post-processing necessary?	Yes	No

Edge prediction GNN performance example

Training and testing samples

- About 130 shower particles
- Number of hits > 400 and < 700
- with > 30% contamination
- 90%-10% training-testing split

→ 1M edges
(equal number positive and negative)

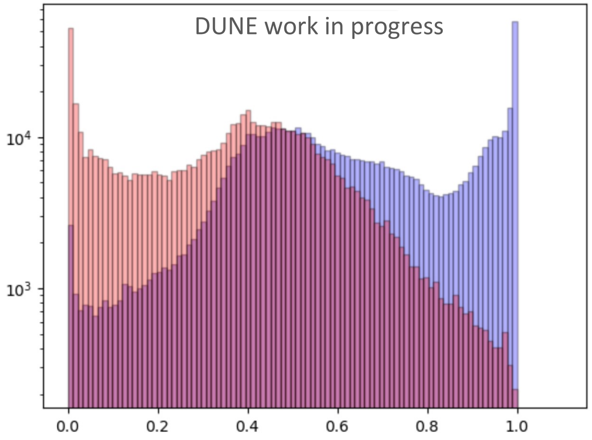
Architecture

- 10 SAGEConv layers, 1 predictor layer, 16 hidden channels

Predicted edge scores

- 0.5 cut → typical prediction accuracy around 75-80%
- Network could be tuned for further separation
- However, a harder cut on scores may already give good performance in clustering

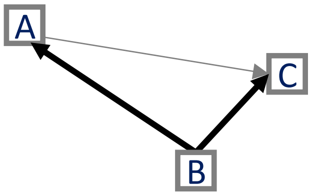
Predicted edge scores



Blue distribution: true positive edges
Red distribution: true negative edges

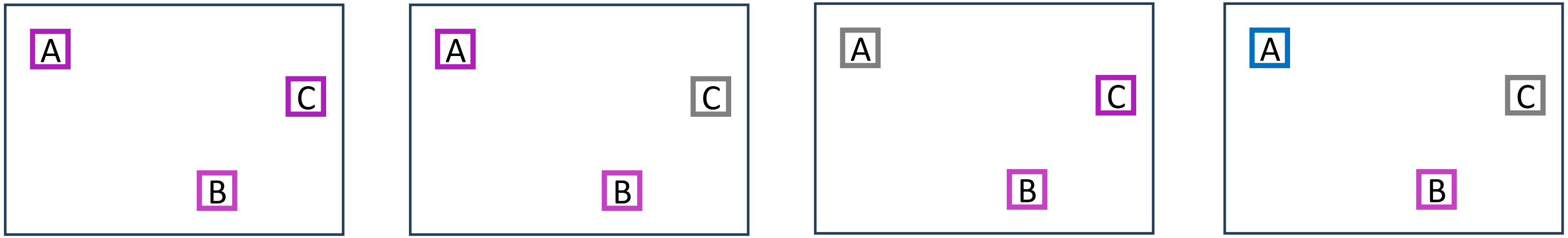
How to use the predicted scores to guide 3D hit clustering?

How to use the GNN output



- There will be ambiguities as the network output is not perfect
- B is predicted to be strongly connected to A and C, but A and C are not predicted to be strongly connected together

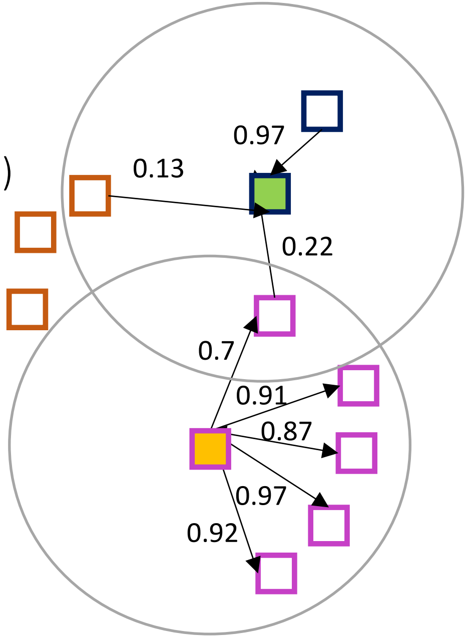
- Four different outcomes



GNN output in clustering algorithms

- Explored the idea of an "average connection score" to quantify how strongly connected each hit is to its local neighbourood.

$$\text{Average connection score (for each hit)} = \frac{\text{(sum of edge scores above threshold within 10 cm)}}{\text{(\# neighbouring hits within 10 cm)}}$$



Hope to show splitting points/regions, or differences between particles
→ Use to split cluster

Merged particles score examples

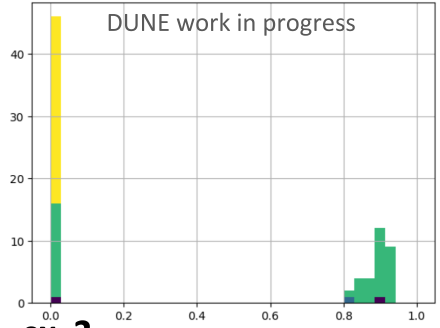
Average connection score

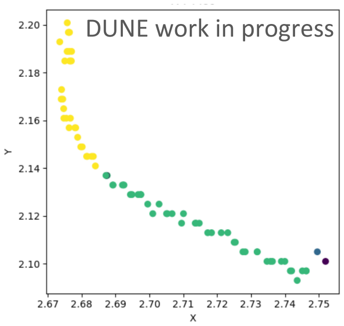
X-Y projection

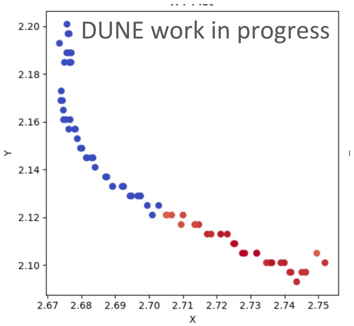
Colour code: True MC ID

Colour code: Average connection score

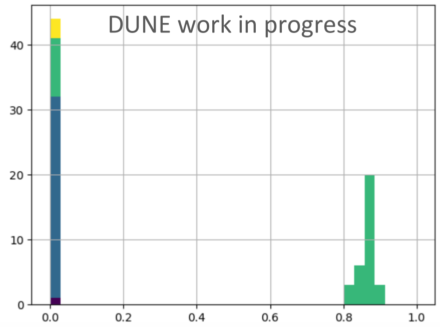
ex. 1

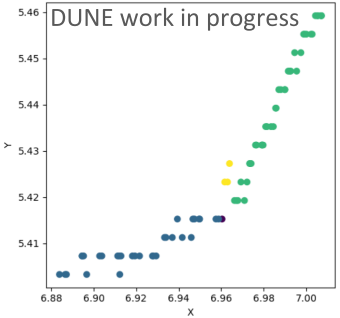


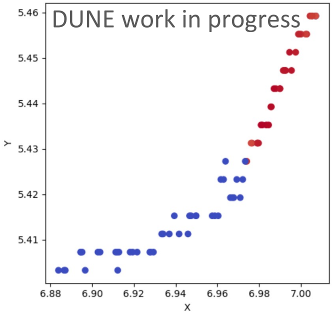




ex. 2







These topologies are "track-like". Scores may be less useful in more shower-like topologies

Summary

- LArTPCs yield very high resolution images of particle interactions
- State-of-the-art reconstruction crucial to achieve DUNE physics goals
- Some topologies, such as overlapping showers, pose special challenges
- Pandora’s multi-algorithm approach is well placed to tackle these challenges
- A reclustering paradigm, under development in Pandora, allows exploiting multiple techniques, including deep learning
- Exploring using GNNs, based on graph interpretation of 3D particles
- Initial training of network to predict true connections between hits
- Exploring use of output scores in new clustering algorithms
- In parallel, exploring an unsupervised clustering approach