Recent Pandora Updates

Maria Brigida Brunetti

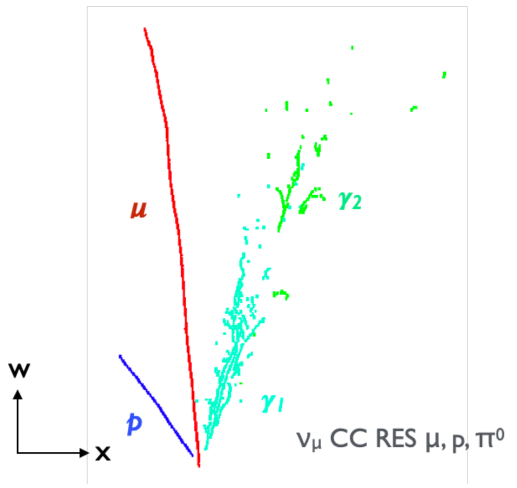
DUNE UK meeting – Lancaster University - 12 January 2023

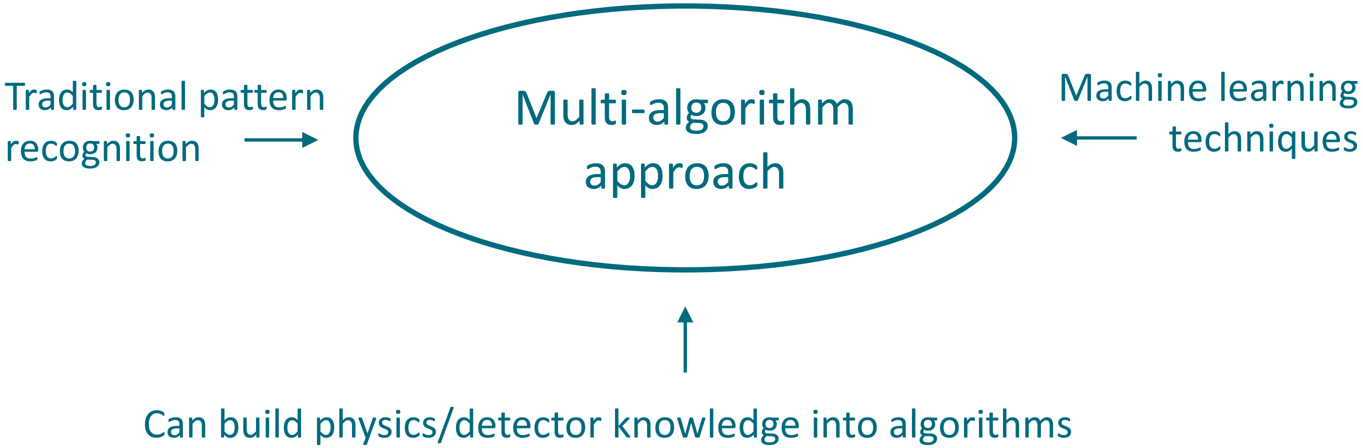
WARWICK
THE UNIVERSITY OF WARWICK

Pandora's approach to reconstruction

- Goal: go from **collections of hits** to **full 3D hierarchies** of particle interactions
- Multi-algorithm approach: cautiously move forwards, processing the images in many logical steps, using the best approach available for each step
- Many algorithms can be combined to form different reconstruction chains, each suited to specific detector/analysis needs

BNB interaction at MicroBooNE - 3 x 2D





Currently used in multiple LArTPC experiments and across different detectors at DUNE

Outline

Current work involves developing the reconstruction with a multi-pronged approach:

Core Pandora development

- Vertexing with Deep Learning *(Andy Chappell)*
- Reclustering mechanics in Pandora

Analysis-driven Pandora development

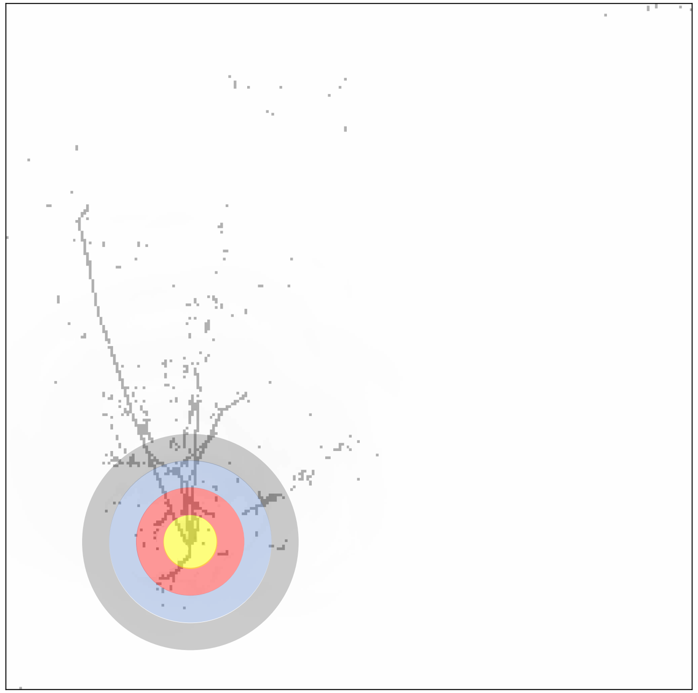
- CP-violation analysis *(Isobel Mawby)*

This talk

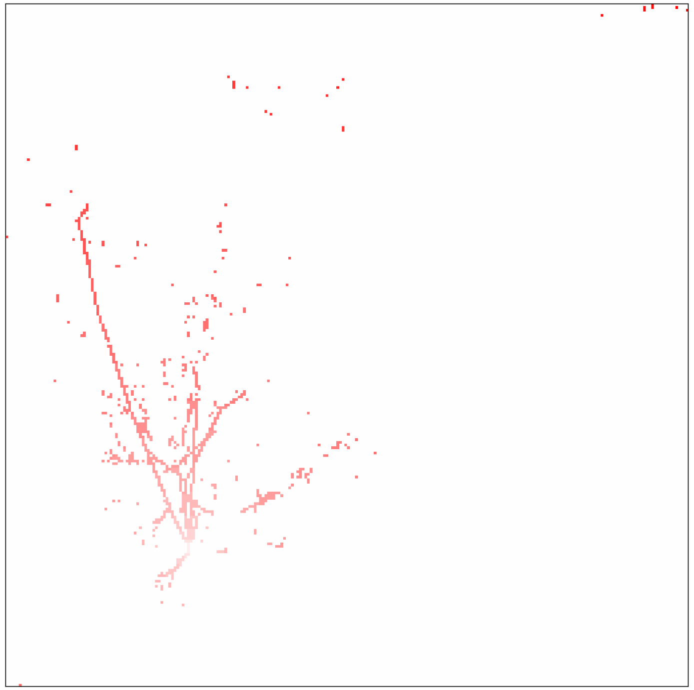
- Supernova neutrinos *(Matthew Osbiston – see his talk in this session!)*

Vertexing network - a quick reminder

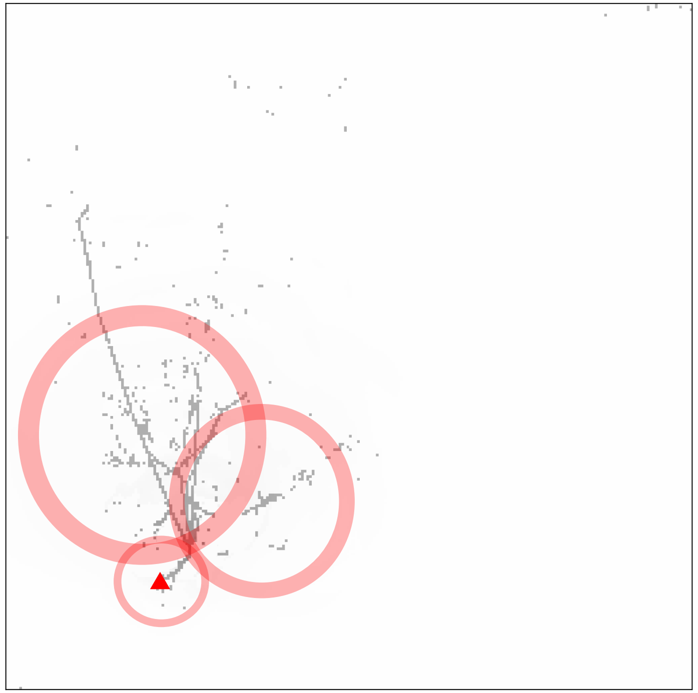
In training hits are assigned a class according to distance from true vertex



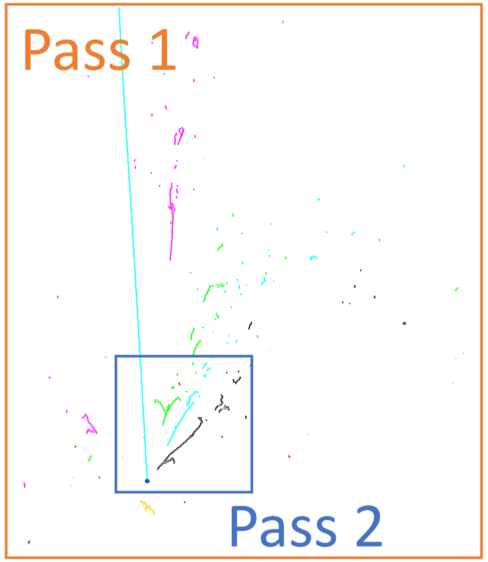
Network trained to learn those distances from input images



Network infers hit distances and resultant heat map isolates candidate vertex

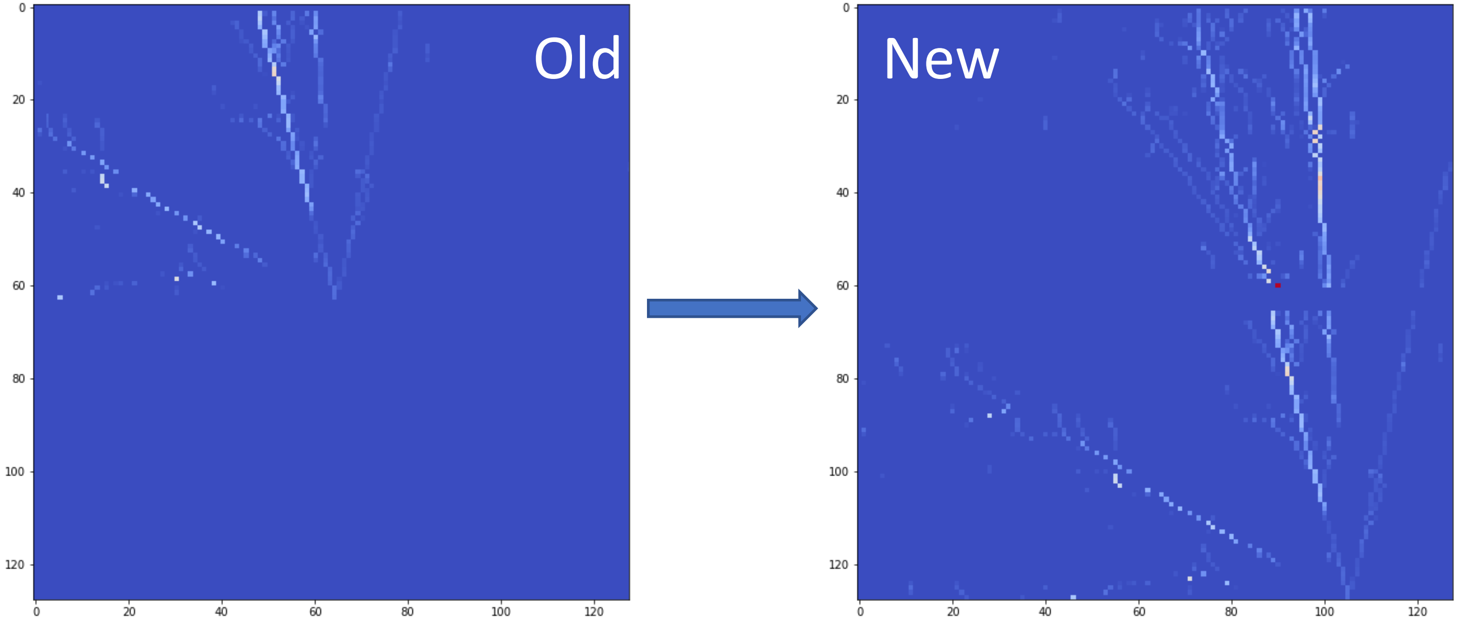


Vertex finding network updates



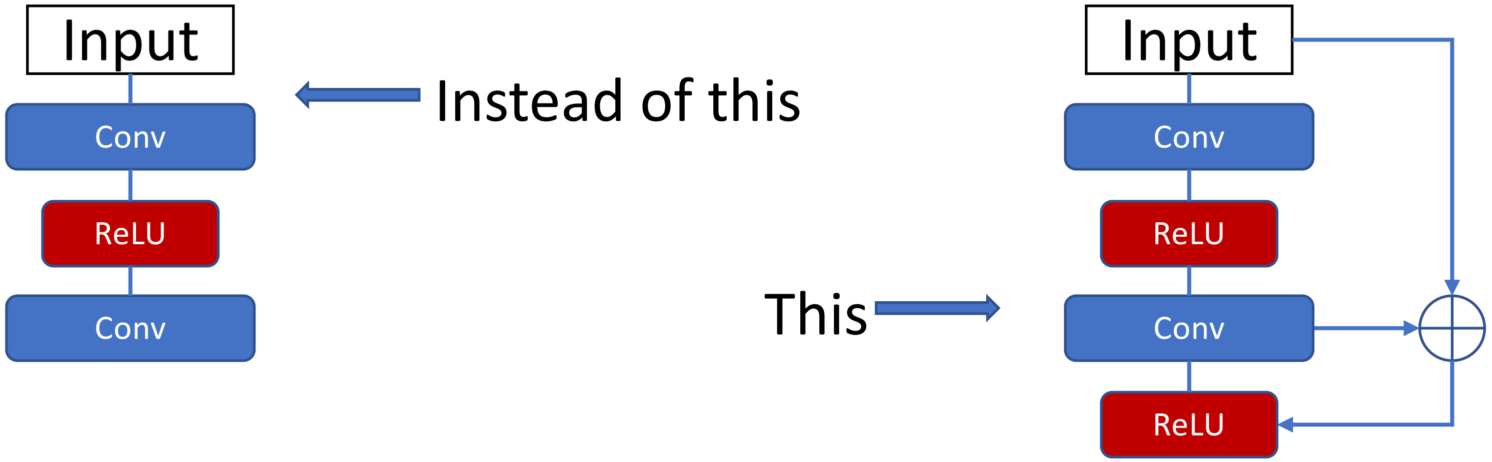
- Added several refinements to the existing network
- World space to pixel space mapping aligned in x in all views
  - Intended to help inter-view correlations for extracting 3D vertex from 2D vertices
  - May allow simultaneous determination of x coordinate via ‘profiling’ views for heat map
- Pass 2 pixel size now same as wire pitch in each view
  - Makes receptive field slightly smaller (previously 0.5 cm pixels) but 1 pixel now 1 wire

- Reframed pass 2 images
  - Previously centred pass 2 inputs on pass 1 estimated vertex – wastes pixels due to boosting
  - Get hit asymmetry left/right, upstream/downstream of pass 1 estimate and frame to better represent hit distribution



Vertex finding network updates (2)

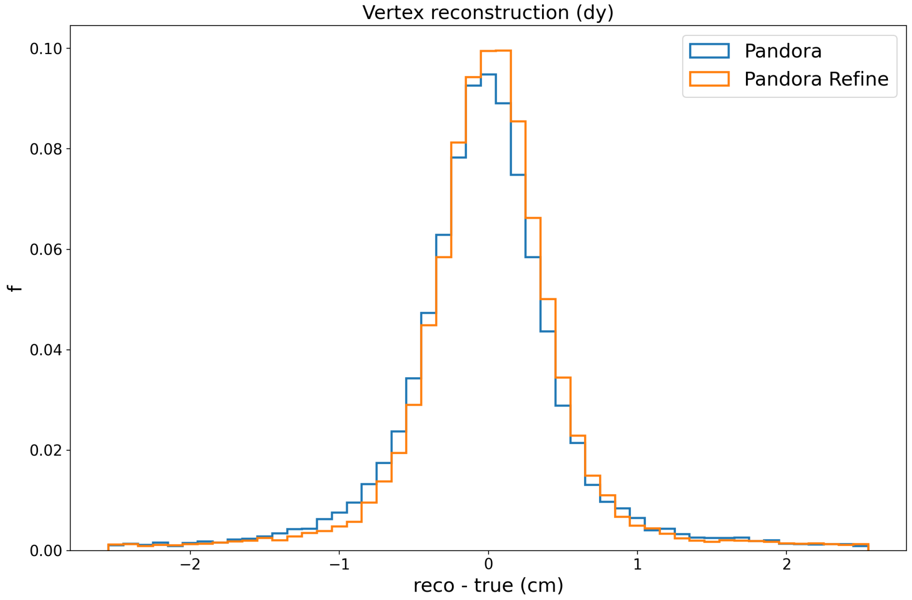
- Minor change to the network architecture
- Replaced the convolutional blocks with ResBlocks (slightly simplified schematic)

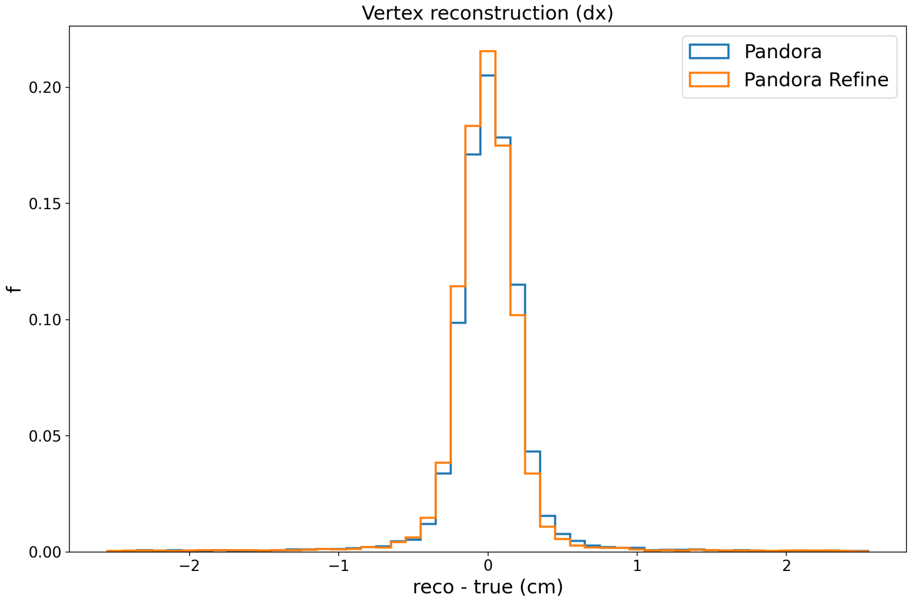


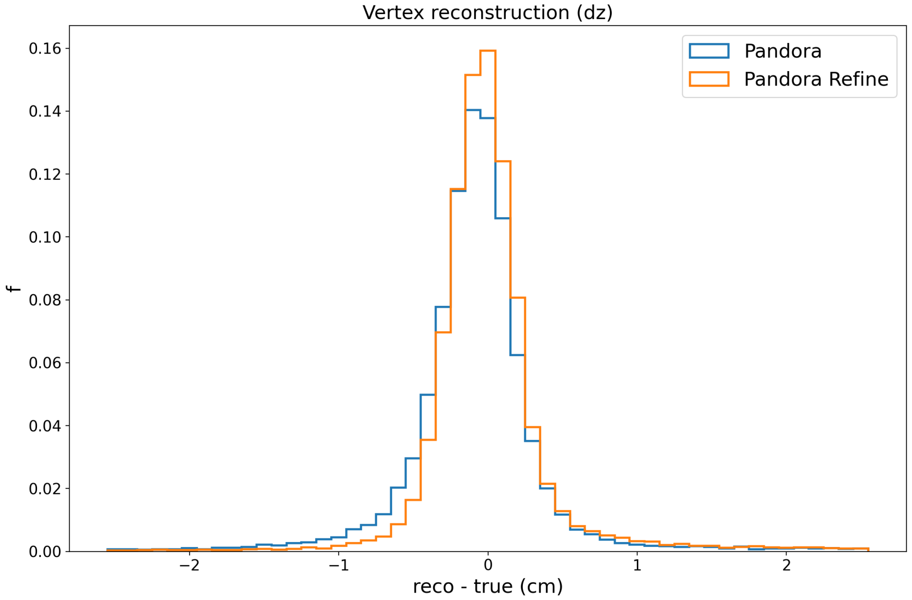
- The shortcut connections introduced by ResNet learn the residuals from input to output rather than the full input to output mapping

DUNE HD FD 1x2x6 accelerator

- Incremental gains in all axes





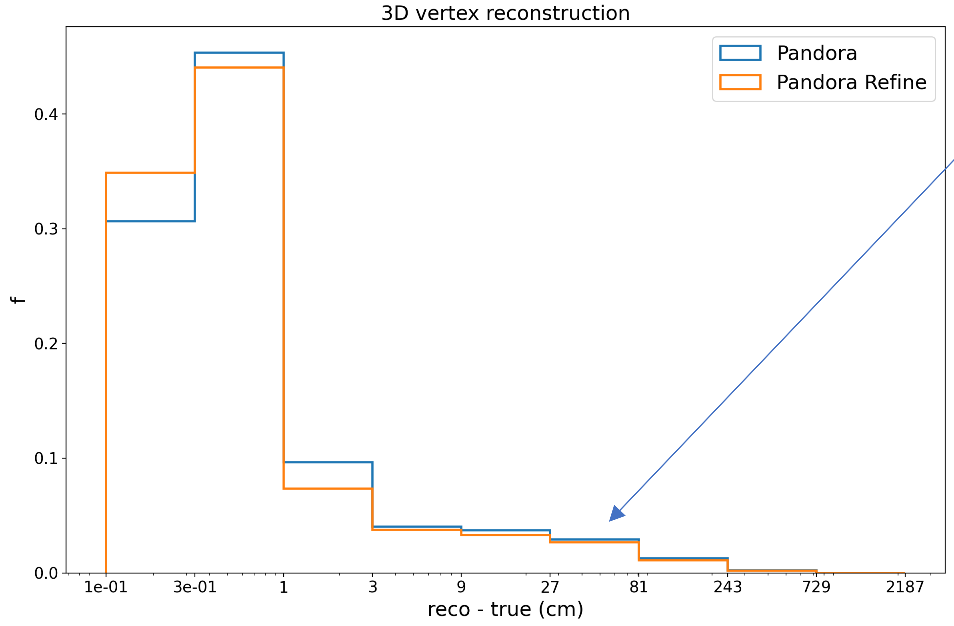


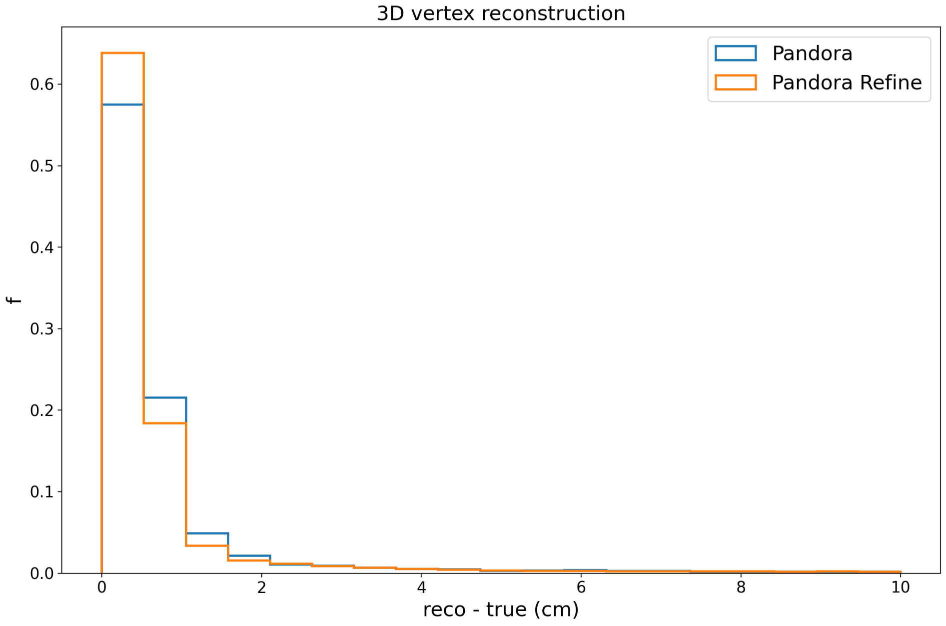
DUNE HD FD 1x2x6 accelerator (2)

- Clear improvement in vertexing precision
- $v_{reco} - v_{true}$ event fractions:

Pandora	Pandora refine
< 1 cm: 78.2%	< 1 cm: 81.6%
< 2 cm: 85.8%	< 2 cm: 87.0%
< 3 cm: 87.9%	< 3 cm: 89.0%
< 5 cm: 89.9%	< 5 cm: 91.0%

Very small improvement in the tail (wasn't expecting much)





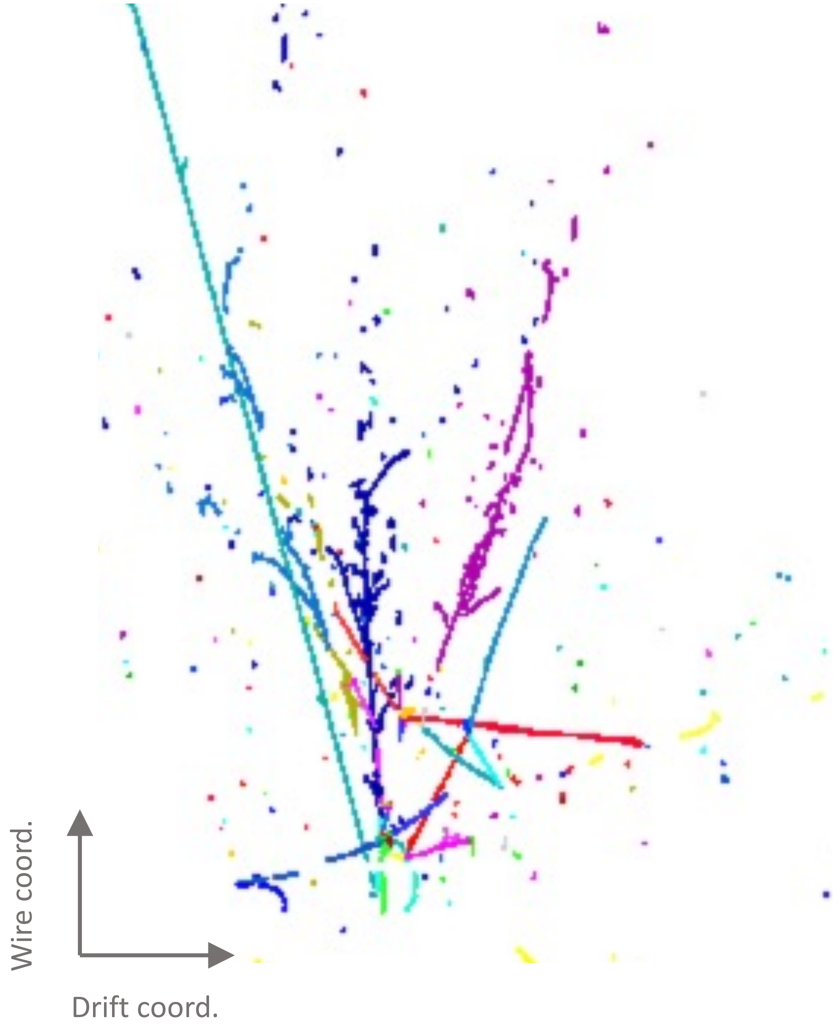
Next steps

WARWICK
THE UNIVERSITY OF WARWICK

- Intend to apply high stats training to atmospheric neutrino samples
  - Will use the accelerator trained network for transfer learning
  - Will look to compare the two networks in both samples to assess the role the beam bias plays in training
- Secondary vertices
  - In principle, this method could detect secondary vertices as well
  - Will use the atmospheric sample (initially) to check feasibility

Shower reconstruction challenges: a reminder

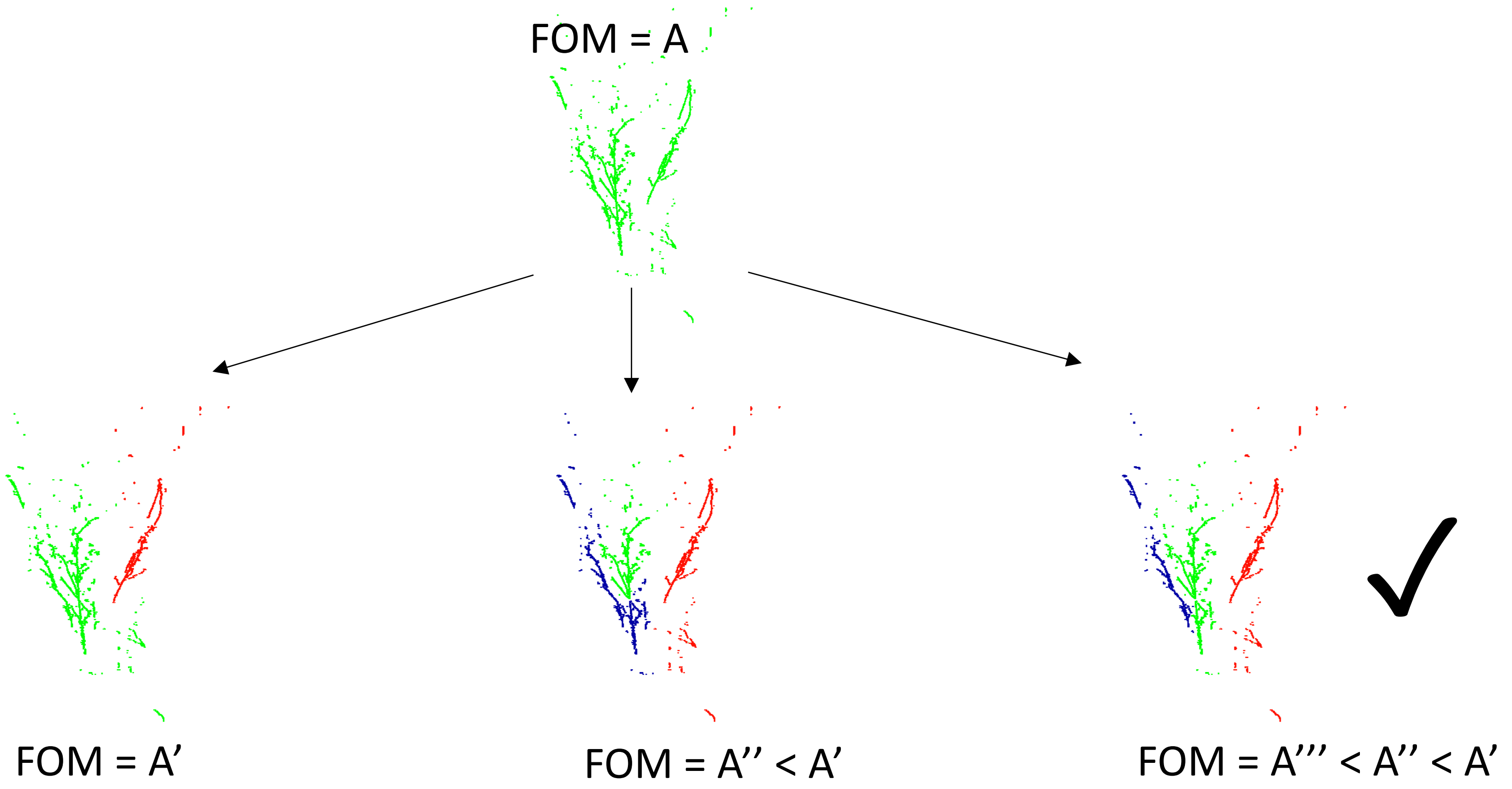
- High-energy interactions in LArTPCs produce complex images with multiple overlapping tracks and showers
- This creates a reconstruction challenge: disentangle merged tracks and showers
- Merging of two or more showers: a relevant problem affecting many different environments and analyses
- In υ_e appearance analyses, merging of photon showers from π^0 decays can lead to mis-tagging the shower as electron-like and mis-labelling the event



6.7 GeV DUNE FD neutrino interaction
cheated pattern recognition

Reclustering: general idea

- Aim: tackle shower splitting with a reclustering paradigm (Pandora was designed for this)



Reclustering example

- The main algorithm runs many encapsulated clustering algorithms
- Each produces a list of clusters and has an associated Figure of Merit (FOM)
- The main algorithm picks cluster outcome with best FOM

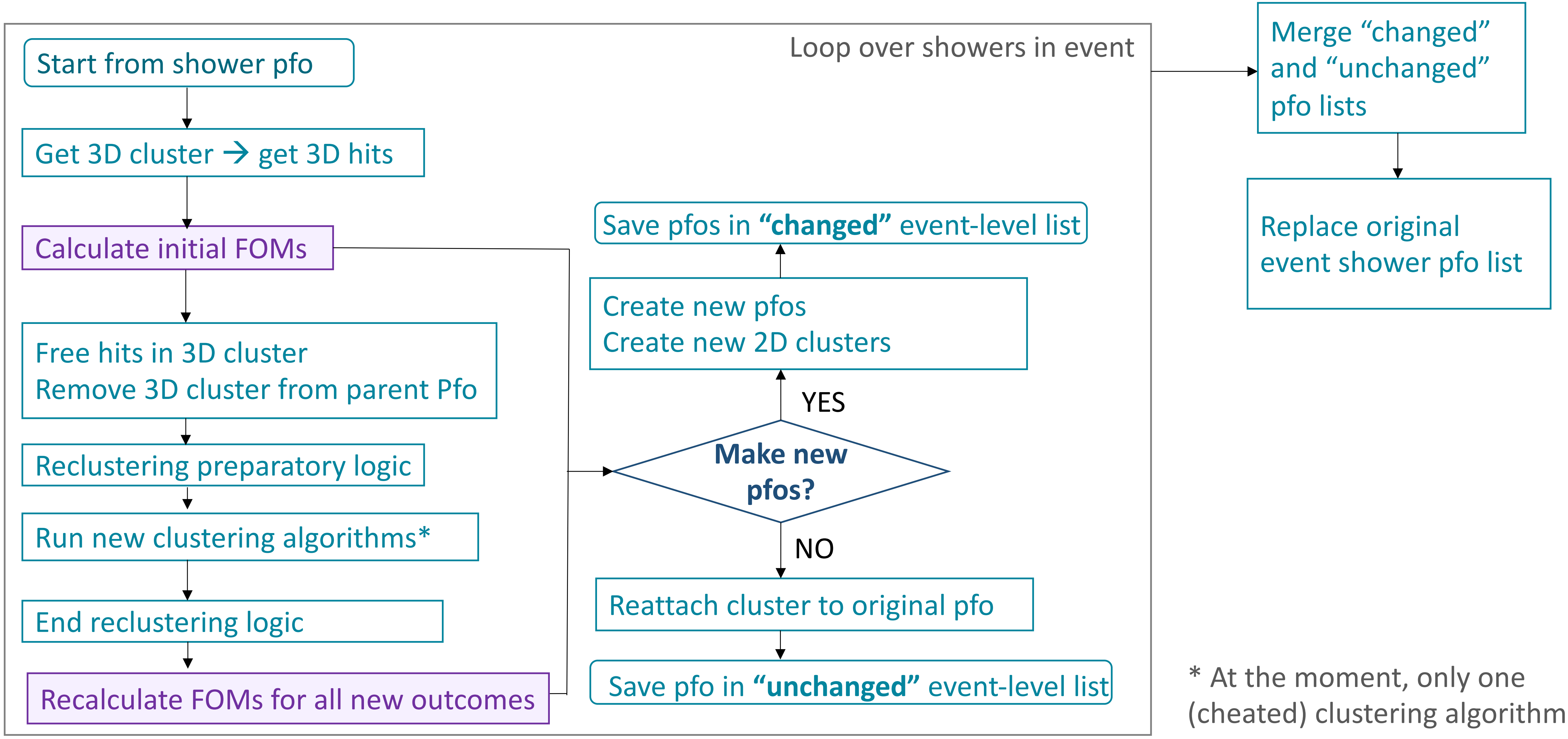
- Pick best outcome rather than fine-tuning algorithms → robust approach

Reclustering: a roadmap - part 1

1. Proof of principle of calorimetry-driven reclustering **See previous talks**
2. Implement reclustering mechanics in Pandora **Implemented, under refinement**
3. Implement cheated FOM and cheated clustering algorithm, with aims: **Implemented**
   - Test mechanics
   - Get performance ceiling
4. Patrec-driven performance improvements with cheated reclustering **Ongoing**

Will now discuss 2. 3. and 4.

Reclustering mechanics

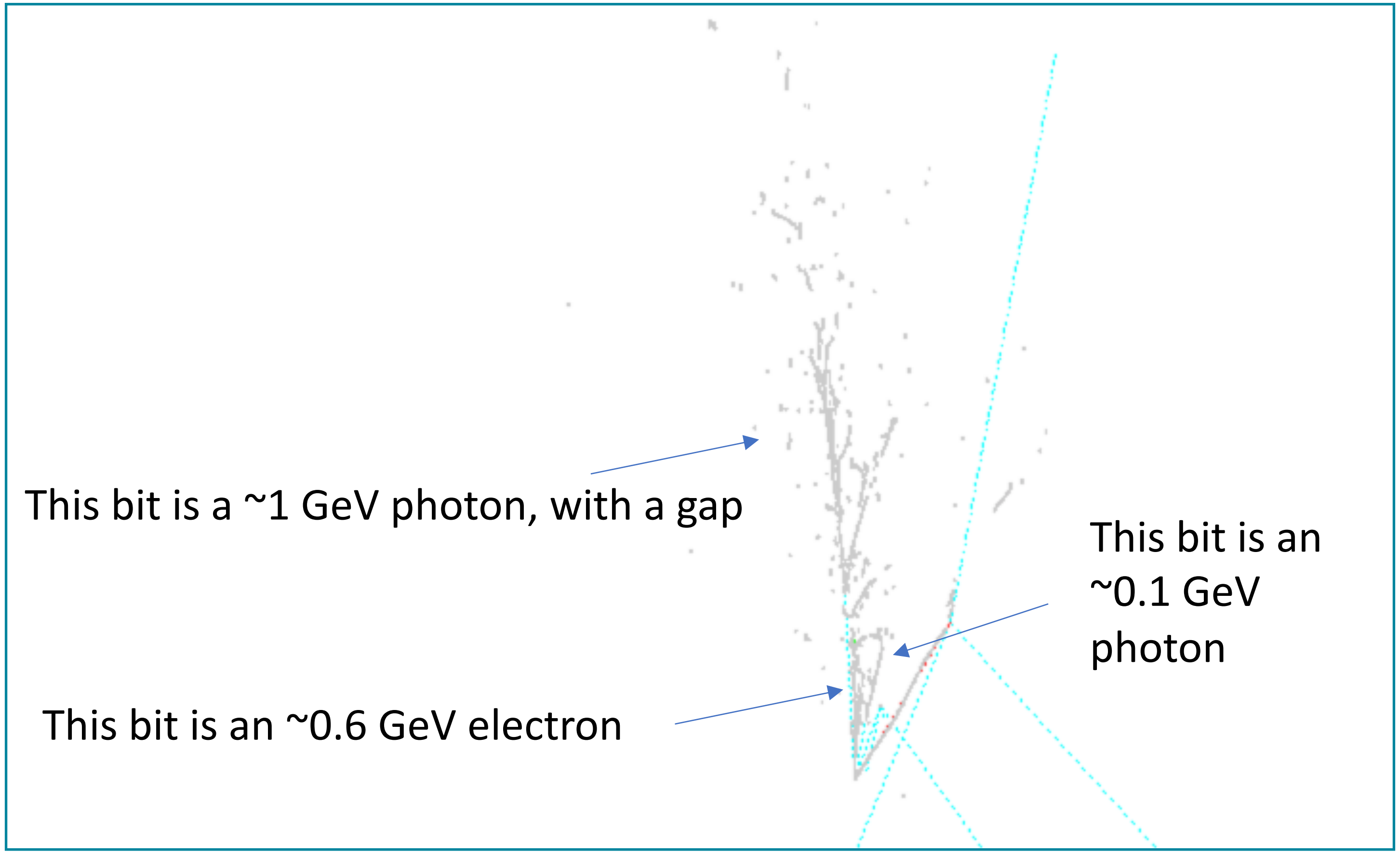


* At the moment, only one (cheated) clustering algorithm

A cheated reclustering example (cheated FOM and cheated clustering)

Start from a merged shower pfo.

Collinear showers → This case may be hard to tackle with real clustering algos, but longitudinal shower profiles may help

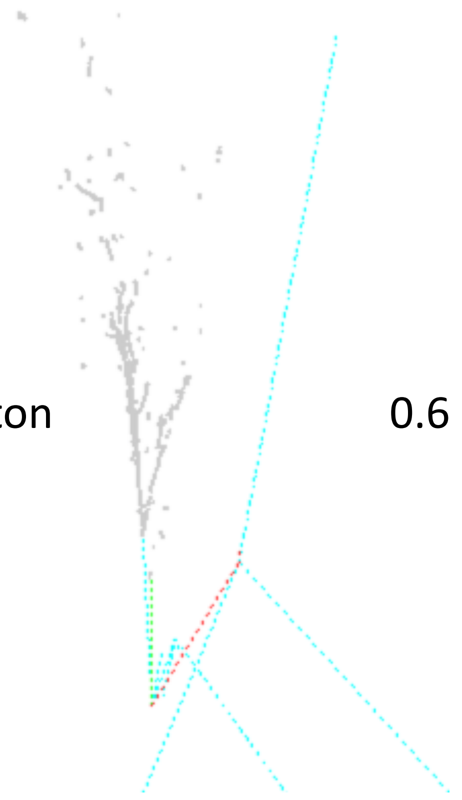


These showers have been merged together!

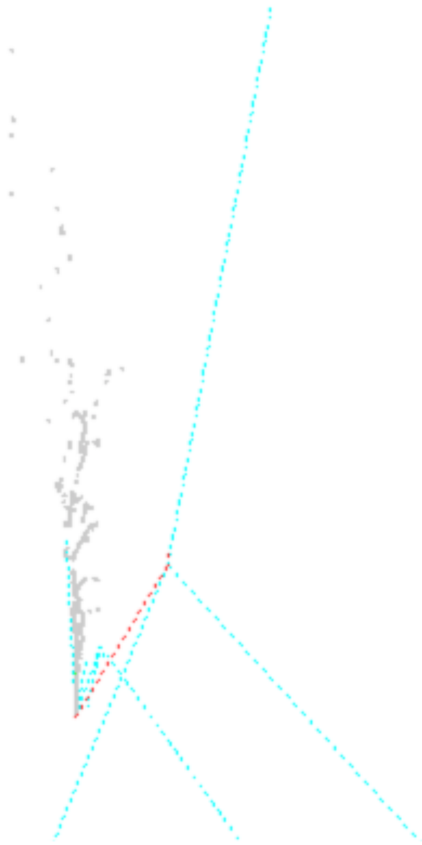
A cheated reclustering example (2)

New 3D clusters returned by cheated clustering algorithm

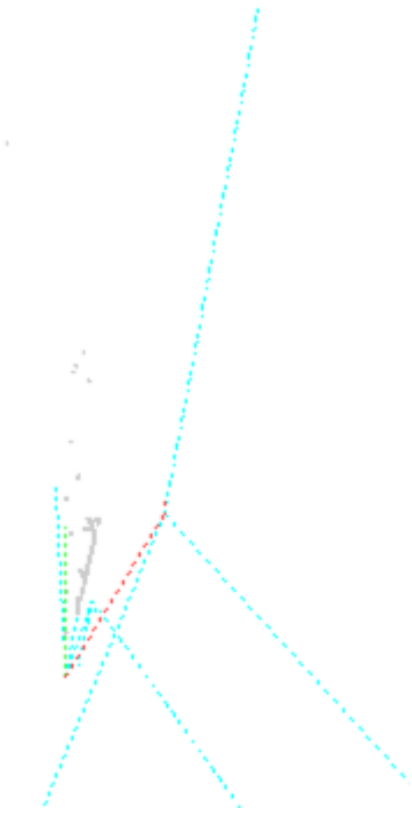
1 GeV photon



0.6 GeV electron



0.1 GeV photon



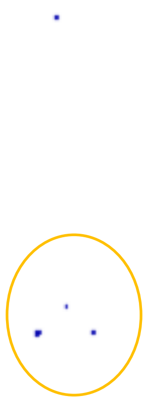
In future, different clustering algorithms can be called, each giving a different set of clusters

A cheated reclustering example (3)

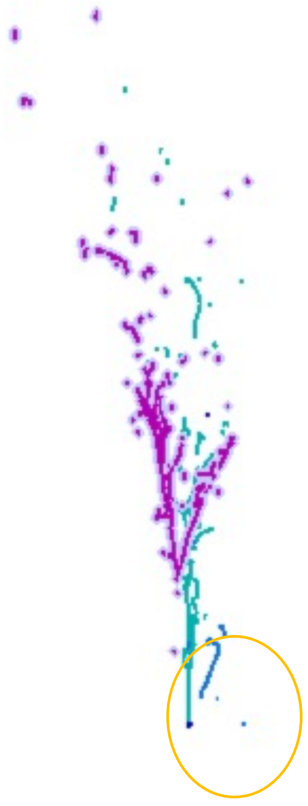
Once the new 3D clusters have been made, the algorithm needs to: take care of 2D clusters, make new pfos, and do some bookkeeping

Initial 2D cluster

New 2D clusters



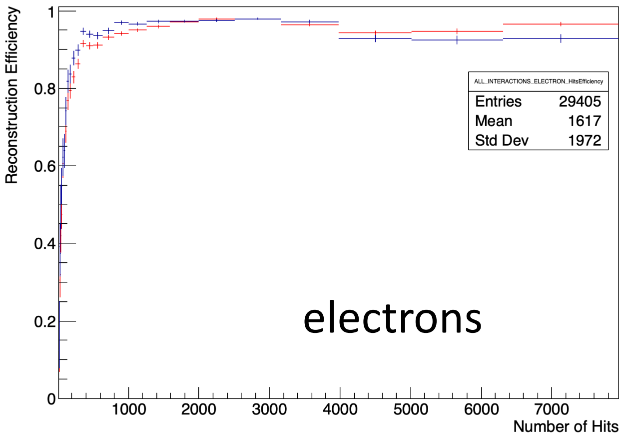
…We have a few leftover hits →

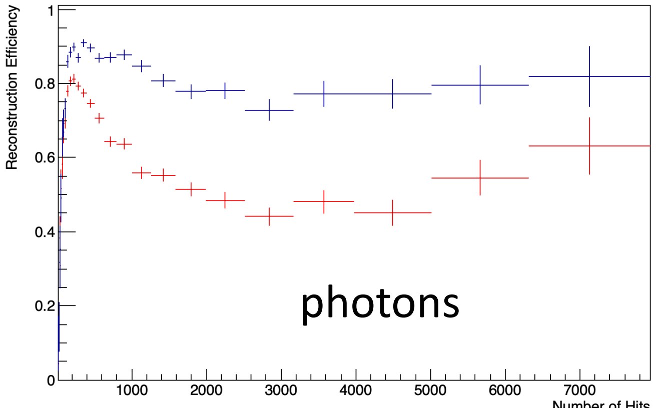


Assign them to nearest cluster

Highlights: cheated reclustering performance

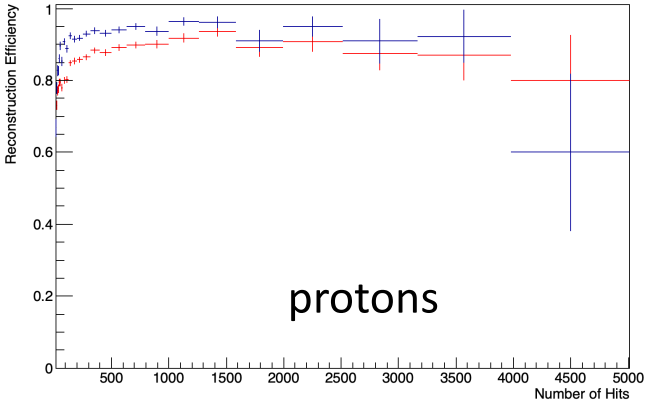
- A first look at cheated reclustering performance with Pandora standard patrec metrics

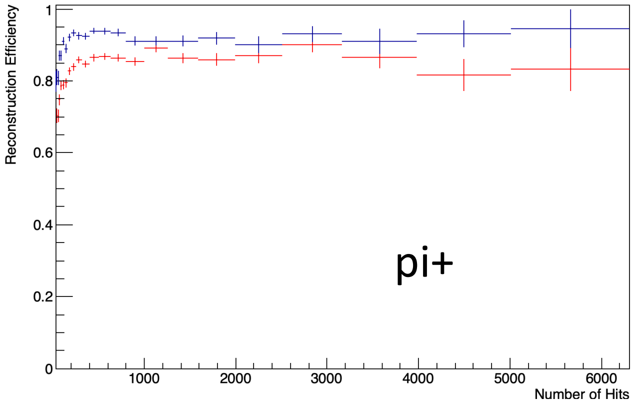




Investigate electron efficiency when using reclustering

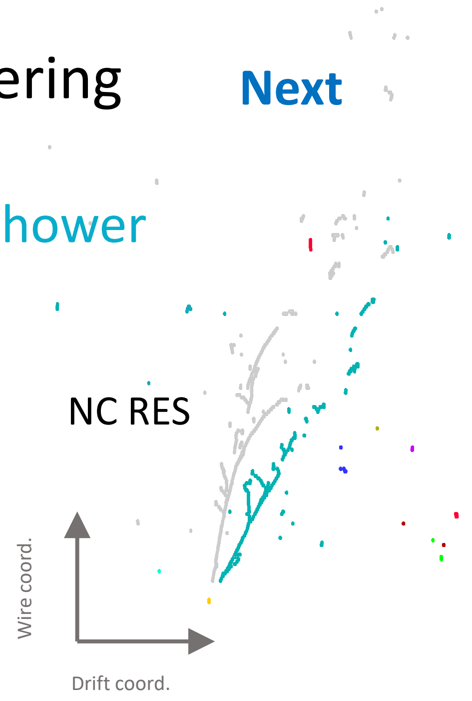
Red = standard
Blue = reclustering





Reclustering: a roadmap - part 2

5. Analysis-driven performance improvements with cheated reclustering **Next**
   - π^0 mass reconstruction
     (merging of photon showers from π^0 decays can lead to mis-tagging the shower as electron-like and mis-labelling the event)
   - δ_{CP} sensitivity studies



6. Replace cheated FOM **Ongoing**
   - Transverse and longitudinal calorimetric FOMS
   - Topological FOMs

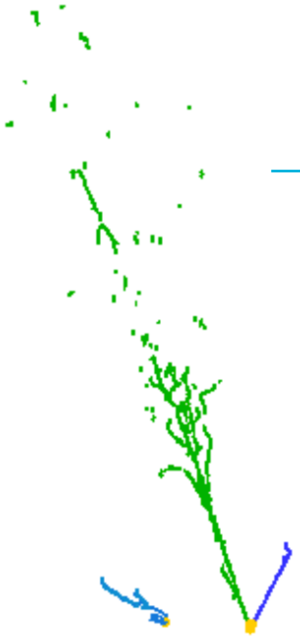
7. Replace cheated clustering with novel algorithms, e.g. using calorimetry **To-do**

8. Explore the use of ML/DL **To-do**

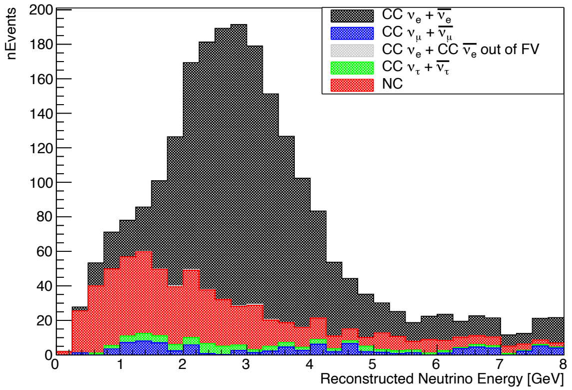
CP-Violation analysis

Credit to **Dom Brailsford** (Lancaster University) for initial development and continued support and discussion – thank you!

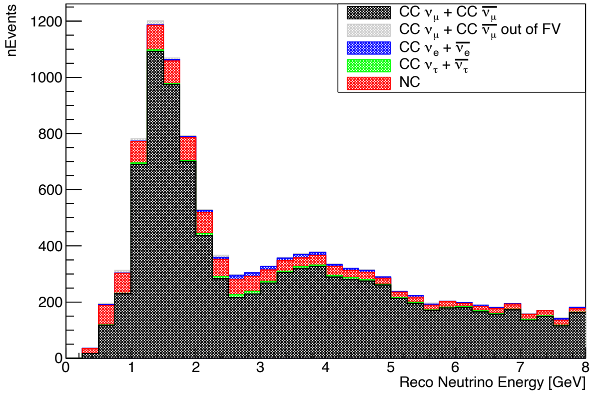
Pandora pattern recognition → Particle characterisation → nue/numu selection → Neutrino energy estimation → CP-Violation metrics

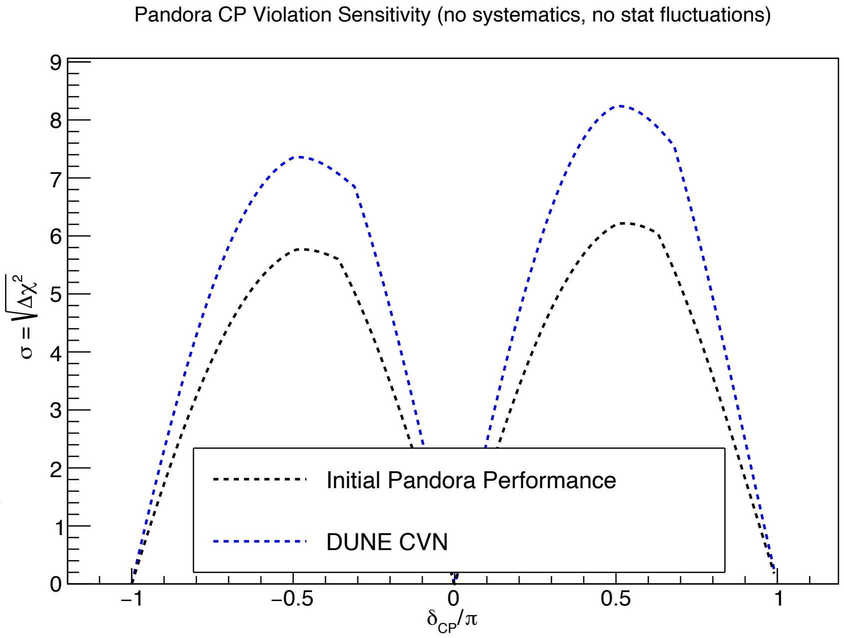


→ Is nue selected? →



→ Is numu selected? →

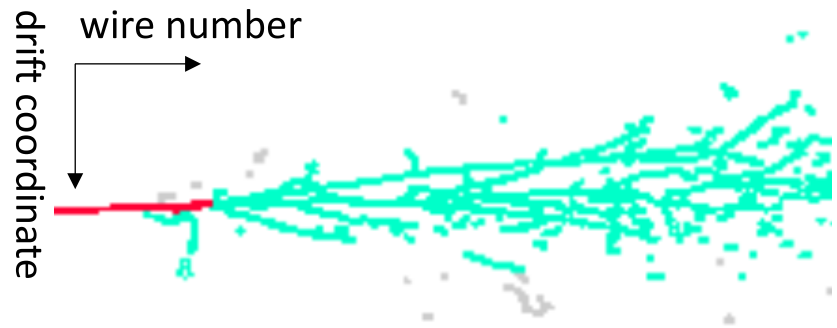




Limited Electron/Photon Separation

- CP-violation analysis performance is limited by our ability to separate electrons and photons

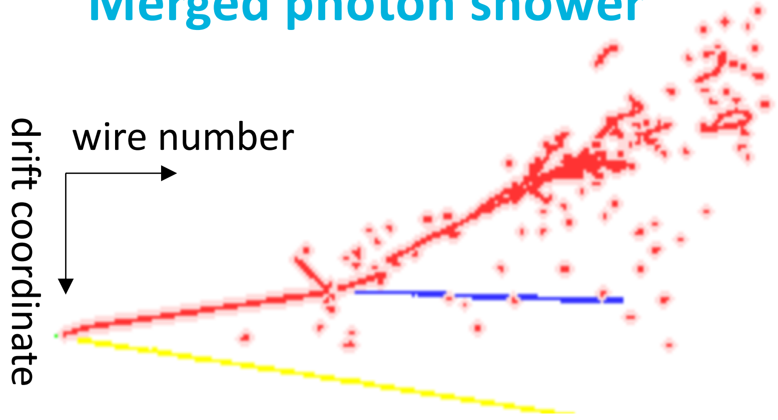
Broken electron shower



$\Longrightarrow$ photon-like gap

$\Longrightarrow$ high *dEdx*

Merged photon shower



$\Longrightarrow$ electron-like gap

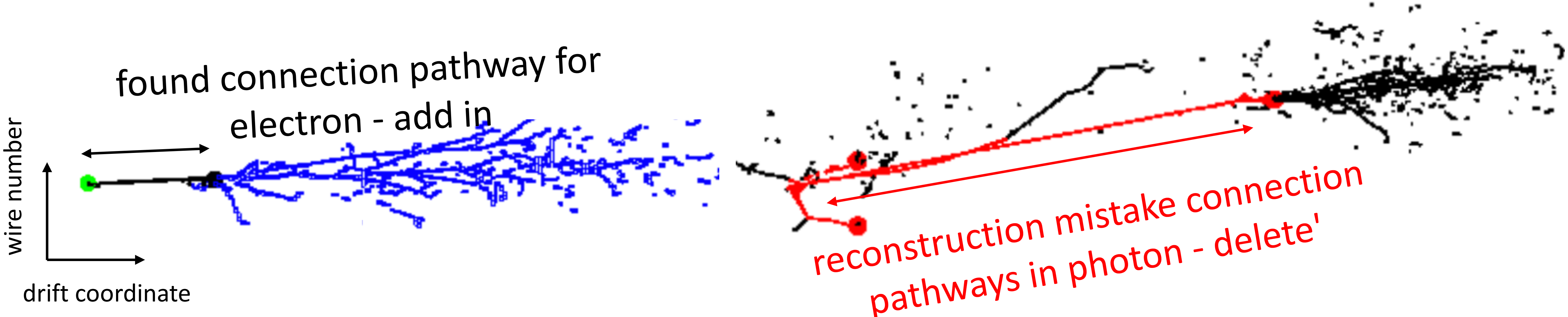
$\Longrightarrow$ incorrect *dEdx*

To improve:

1. Correct the reconstruction errors that smear the electron/photon separation variables in the electron-like BDT
2. Add more information to the BDT to aid the electron/photon separation

The Multi-Algorithm Approach

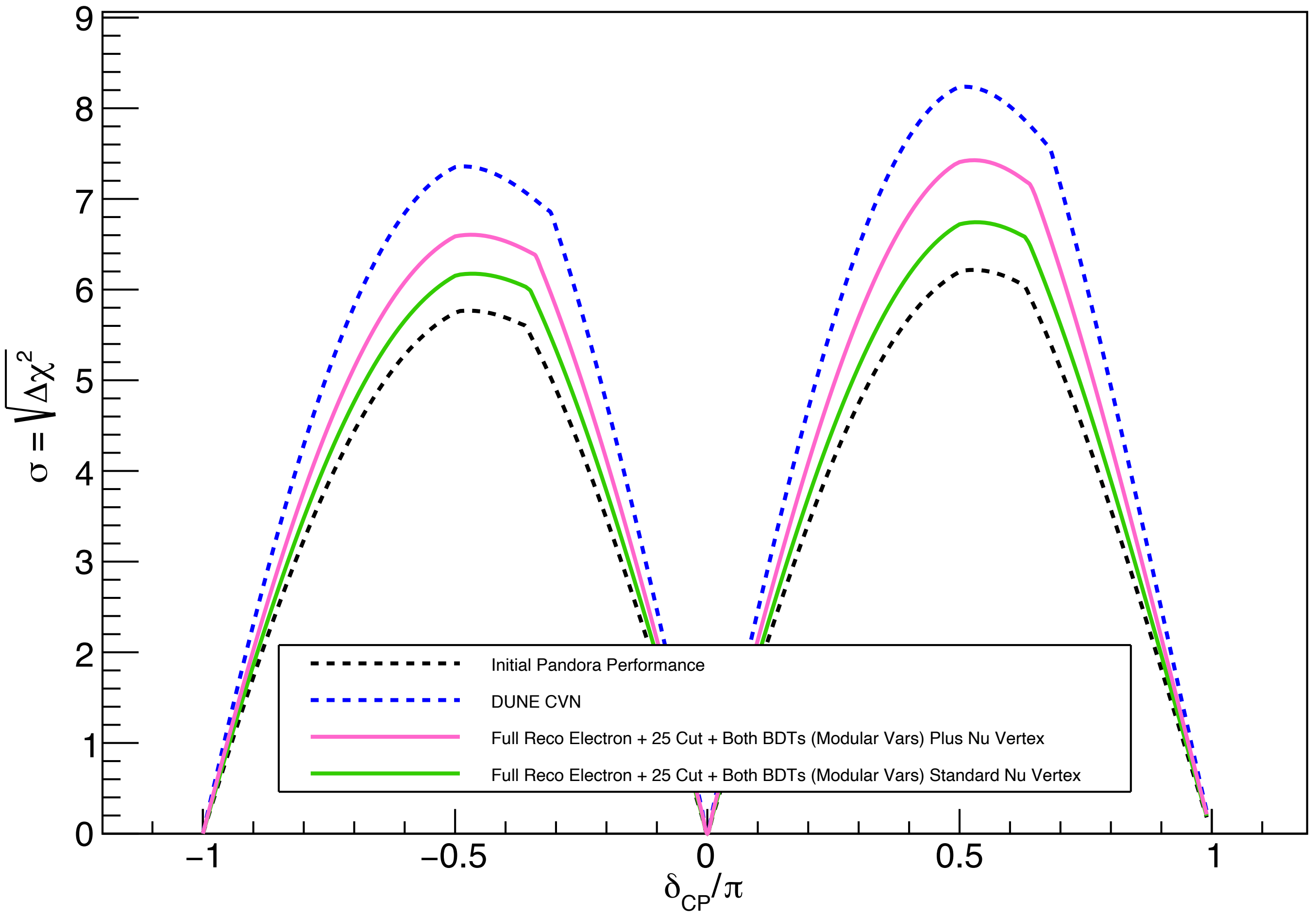
- The Pandora multi-algorithm approach allows us to create a **specifically designed algorithm** to improve the reconstruction of the initial shower region
- In this algorithm we
  - Find the **connection pathways** that the electron (photon) should have (has) followed to get back to the neutrino vertex



  - **Decide** whether the connection should be there or not (using a specially designed BDT)
  - **Add or remove** the connection pathway

Improved Performance

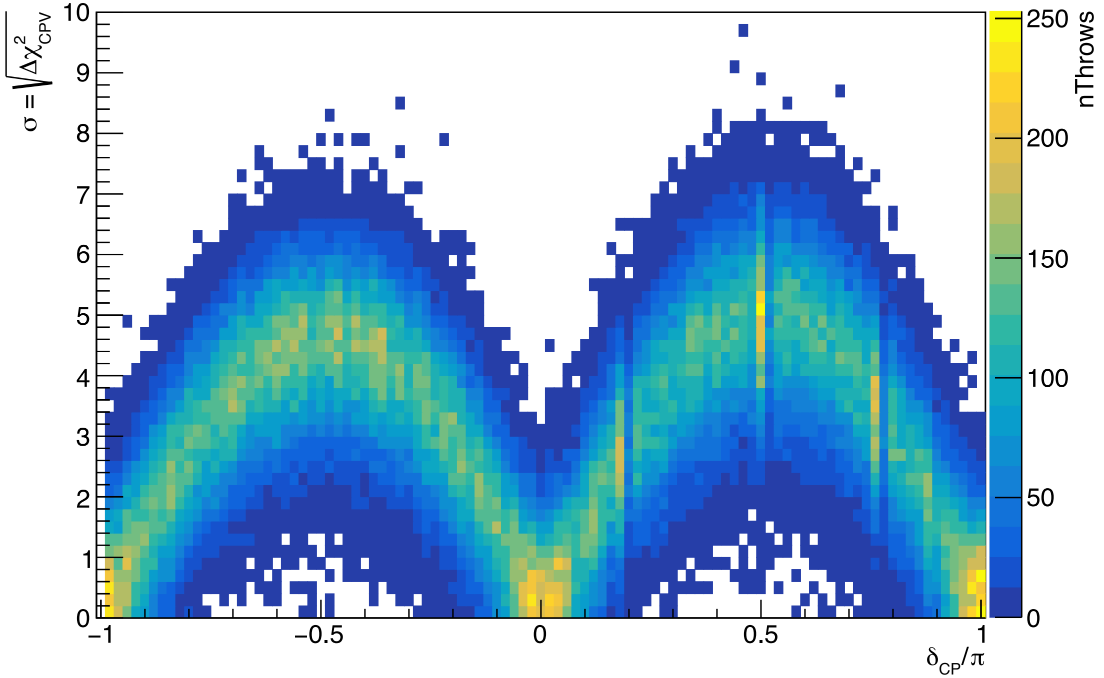
- Extra electron-like BDT variables were created that examine the found **connection pathway**
- With these improvements, **substantial** sensitivity gains were achieved!
- These are **furthered** with an improved neutrino vertexing procedure

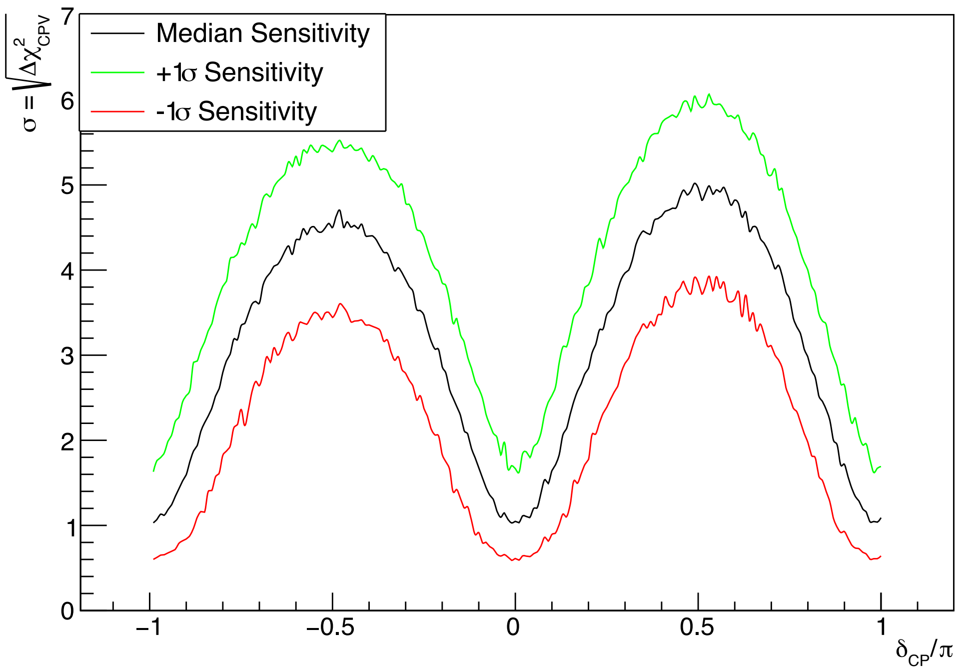


Validating Results: Systematics

- The results incorrectly assume the MC model is a correct reflection of reality
- In response to this, incorporate our uncertainties into the sensitivity estimate

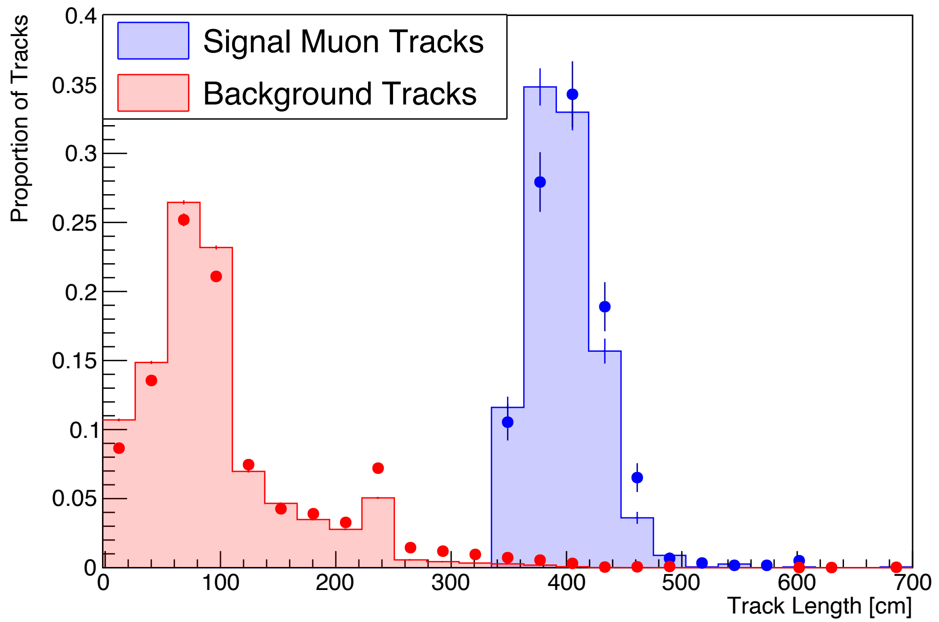
Osc. Param. Uncertainties | **Cross Section Systematics** | **Detector/Energy Systematics** | **Flux Systematics**

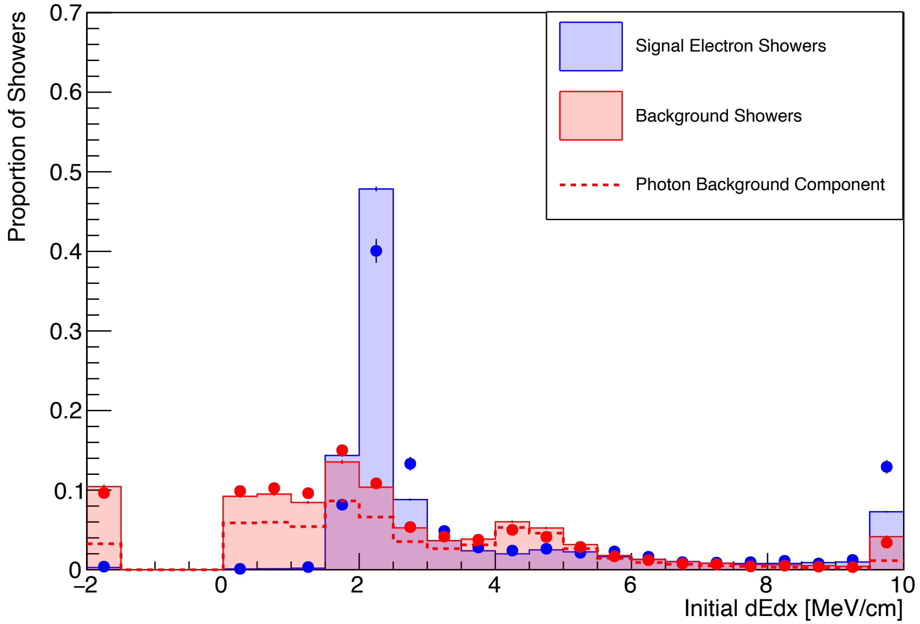




Validating Results: Data

- Can also compare CP-violation analysis performance in MC and data
- Well kind of...
  - ProtoDUNE is serviced by the CERN test beam i.e. not neutrino data
  - Compare performance of electron-like and muon-like BDTs in ProtoDUNE MC and data





- Agreement is very good! Expect that found CP-violation performance can be extended to data

Conclusions

- Pandora's multi-algorithm paradigm allows several approaches to development and optimization to be pursued in parallel, e.g. pure patrec- and analysis-driven

- **Core development to benefit many analyses/environments**
  - Sizeable improvements with new vertexing network
  - Reclustering to improve shower reconstruction

- **Analysis-driven Pandora development**
  - First very successful example: CP-violation analysis
  - New areas of analysis-driven development started: SN neutrinos (see Matt's talk)

Cheated clustering

- Start from merged shower
- Loop over 3D hits
- For each, get parent 2D hit address, and find main contributing MC particle
- Separate 3D hits in different 3D clusters based on parent main contributing MC particle
- Return new 3D clusters to main reclustering algorithm, that will take care of creating the corresponding new pfos and 2D clusters

Cheated FOM

- Loop over 3D hits and find parent (2D hit) address for each
- Via the parent address, find the MC particle that contributes most charge
- **Purity** = fraction of hits where main energy contribution is from the main contributing MC particle
- FOM for initial shower $= 1 - purity$
- After a clustering algorithm has run, N new 3D clusters have been made
- With the cheated clustering algorithm, the cheated FOM at the end is 0 for all new clusters
- I define the new FOM = maximum of the FOMs for the new clusters

π^0 mass reconstruction

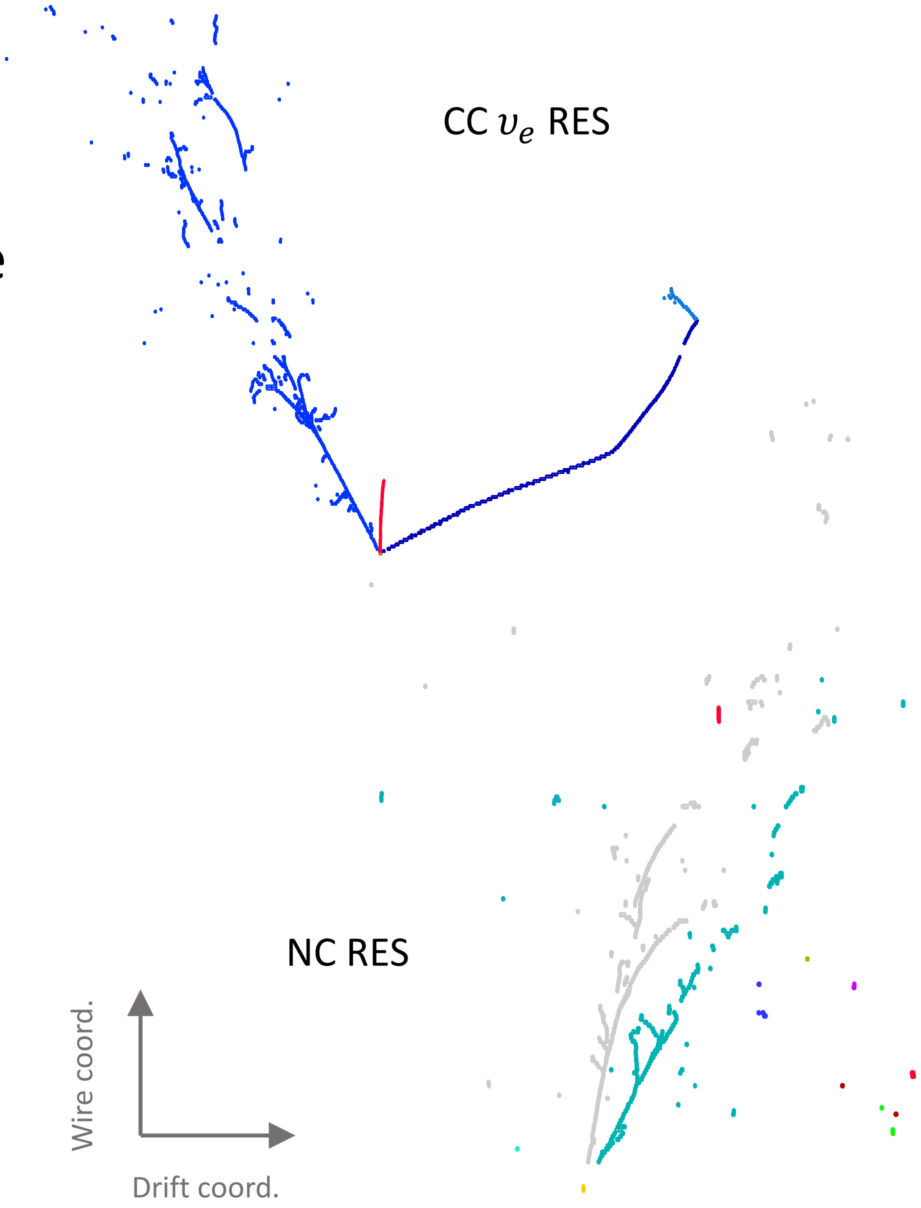
- In υ_e appearance analyses, merging of photon showers from π^0 decays can lead to mis-tagging the shower as electron-like and mis-labelling the event

- Leading causes of performance loss

  https://indico.fnal.gov/event/53402/

- I am working on optimising Pandora shower reconstruction, and will benchmark it with π^0-mass reconstruction-related metrics first

(Neutrino interactions in the 1x2x6 HD FD)



Analysis-driven Pandora optimisation

1) Choose analysis metrics

2) Identify main issues and develop new targeted algorithms

3) Assess metrics improvements

- This approach complements pattern recognition-driven metrics

  MicroBooNE Pandora paper: arXiv:1708.03135v1

  Andy C.'s new patrec metrics: https://indico.fnal.gov/event/50215/contributions/232770/

- The use of analysis-driven metrics as a guide for Pandora optimisation has been successfully demonstrated by I. Mawby in deltaCP sensitivity studies

Splitting merged showers: reclustering

- We can implement **reclustering** in Pandora (main novel design feature!)

  (Following ILC's strategy https://arxiv.org/abs/1506.05348)

1. **Choose figures of merit (FOMs)** to assess whether a shower is merged
2. Start from a list of shower particles, and calculate initial figures of merit
3. Remove the 3D clusters from the parent particle, and **make 3D hits available**
4. Iterate over many encapsulated clustering algorithms. Each will produce a **new list of candidate 3D clusters**
5. Calculate FOMs for all outcomes, and pick best one (or keep original)

→ Integrate with ML/DL approach

Splitting merged showers: reclustering (3)

- First reclustering implementation in Pandora currently under development/testing
- Testing strategy:

1. Fully cheating FOM (make correct decision on whether to split shower)
2. Fully cheating clustering using hit-level MC truth information (returns N new 3D clusters based on the parent MC that contributes most)

How do pattern recognition metrics change?

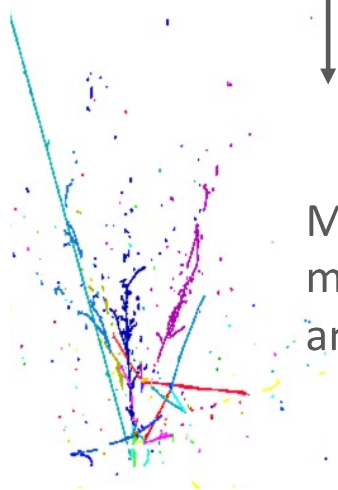
How do π^0 invariant mass reconstruction metrics change?

→ Indication on maximum patrec- and analysis-driven performance improvements achievable with a reclustering strategy

Shower energy profiles

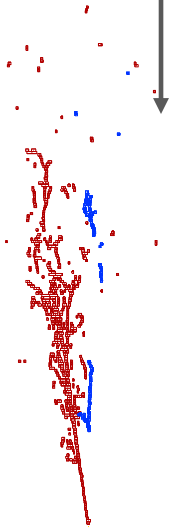
- Use hit calorimetric information to drive reclustering decision via a FOM
- Explore use of shower energy profiles: see previous CM presentation https://indico.fnal.gov/event/50217/contributions/241544/
- Start from **transverse profiles**, look at **longitudinal profiles** at a later stage

(*)



Most useful when merged showers have angular separation

(*)

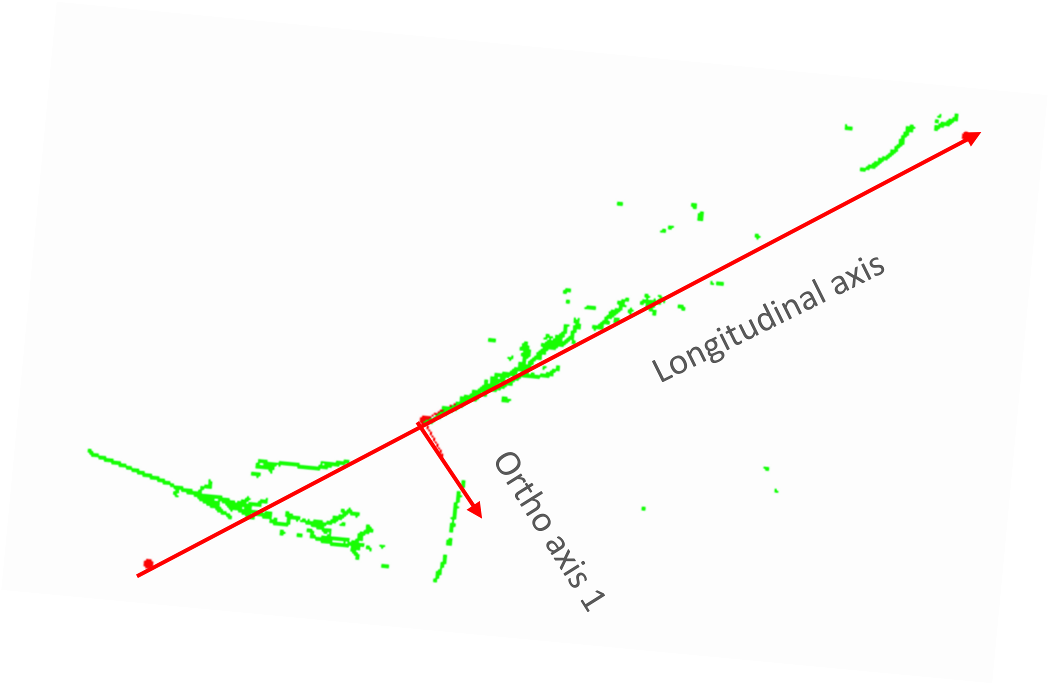


Most useful when two merged showers start at different depths along the principal axis

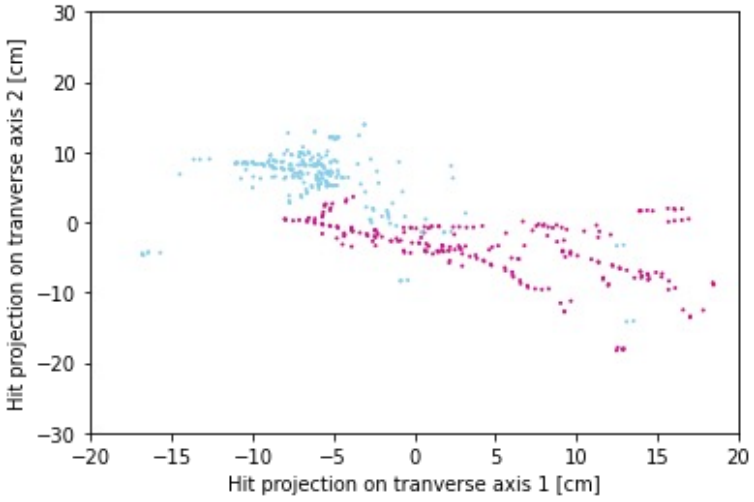
*cheated patrec

Transverse shower energy profiles

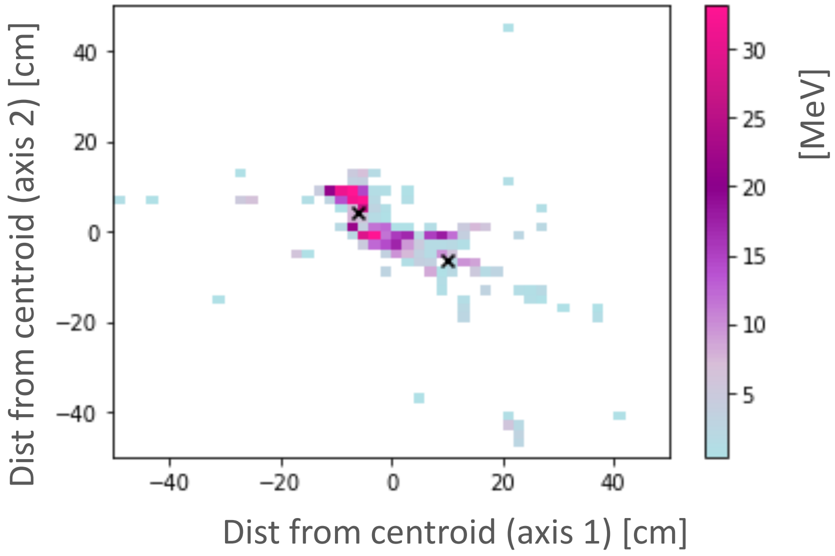
- Find shower principal axis
- Project all hits onto transverse plane
- Produce 2D binned profile summing energy depositions for hits that fall in same bin



Hit projection on transverse plane
Displaying most contributing MC particle

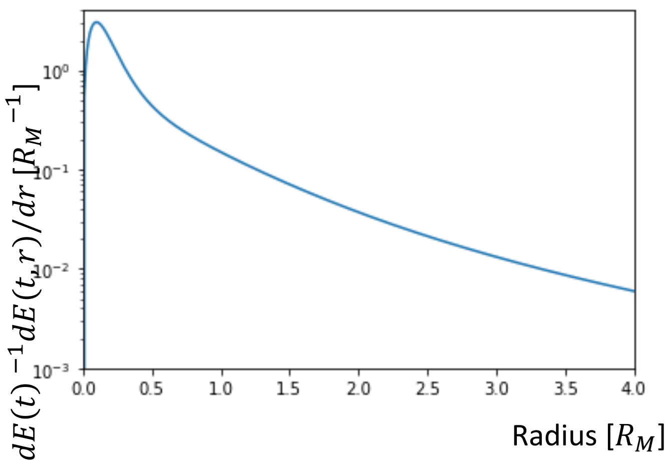


Normalised binned observed energy profiile

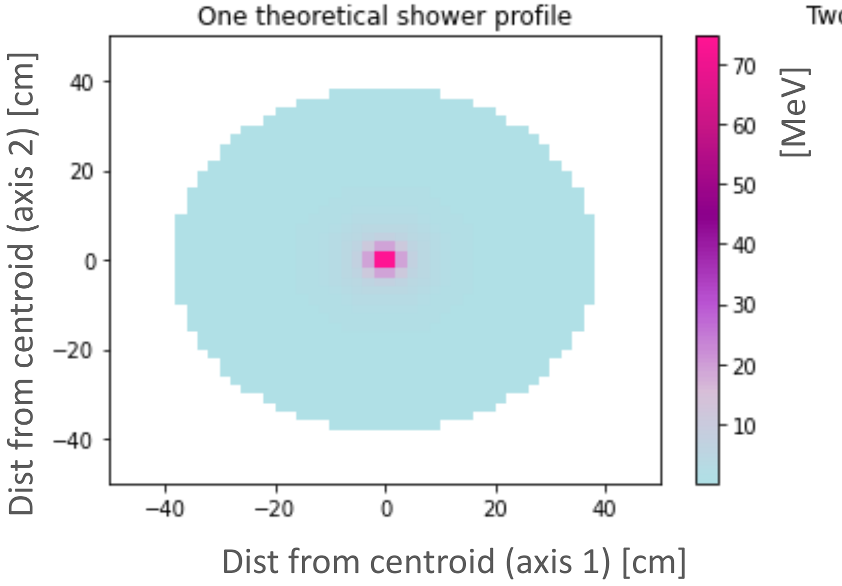


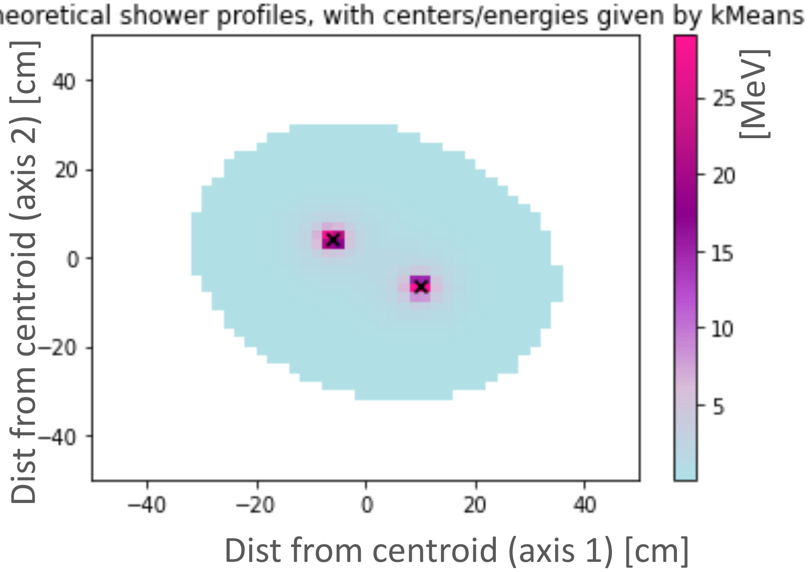
Transverse shower energy profiles (2)

- Transverse shower profile parametrization from [arXiv:hep-ex/0001020v1](https://arxiv.org/abs/hep-ex/0001020v1) (Grindhammer)
- Derive 2D binned expected profile. Can combine many profiles together
- Need external input for center positions and relative energy (e.g. for kMeans)
- At this stage, not focusing on optimal parametrisation, but exploring use of calorimetry
  - May get improvements by simply counting shower cores. Can optimise later



Example expected radial profile for a single 500 MeV shower (normalized)





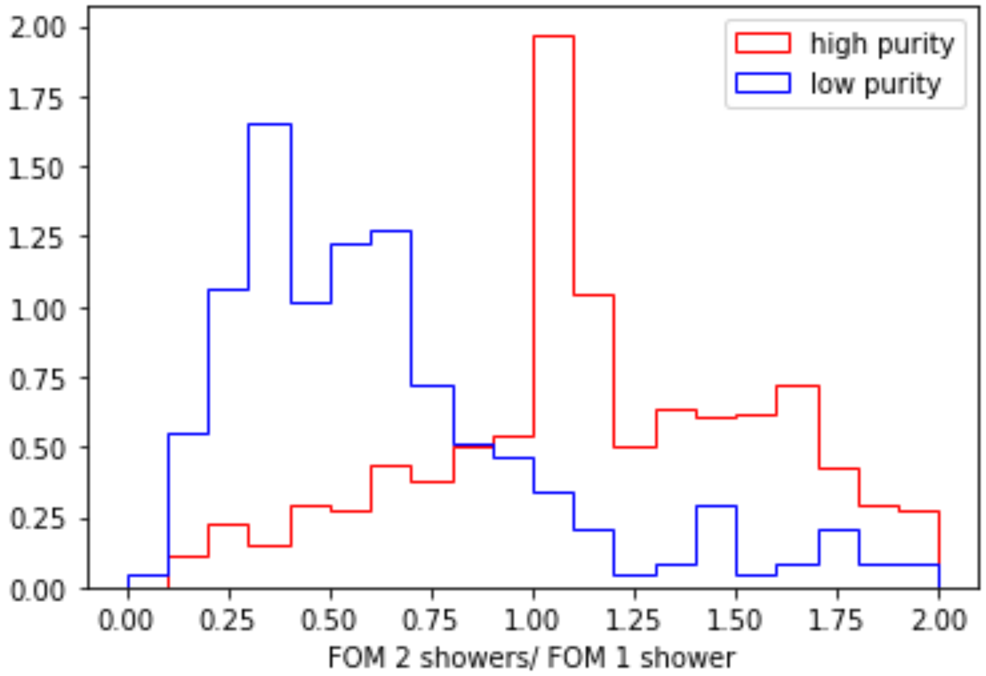
Splitting decision procedure

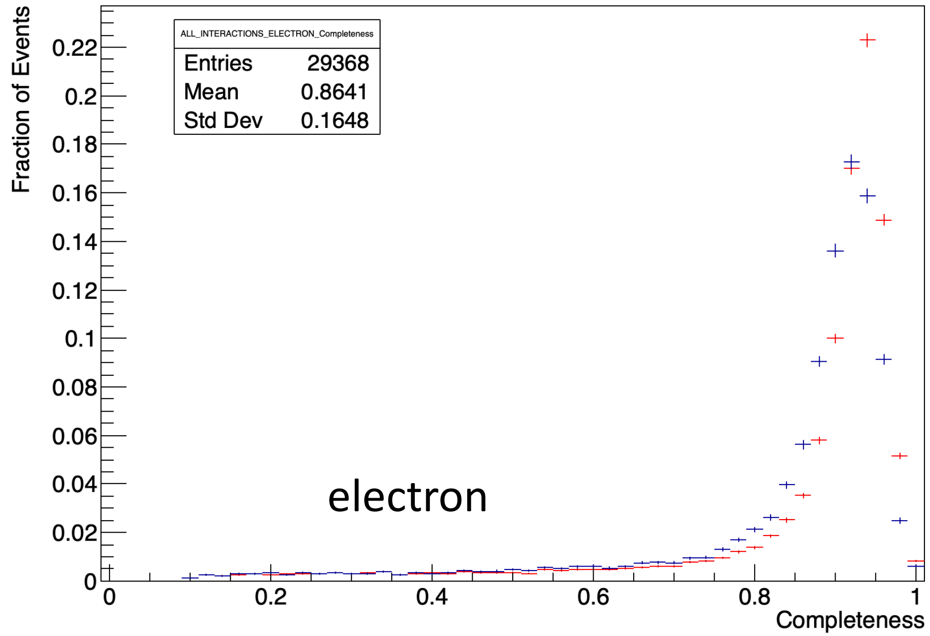
- Does the observed profile agree better with a 1 shower, or a 2 showers prediction?
- Compare bin-by-bin the observed profile to the predictions for different numbers of showers
- Define a FOM:

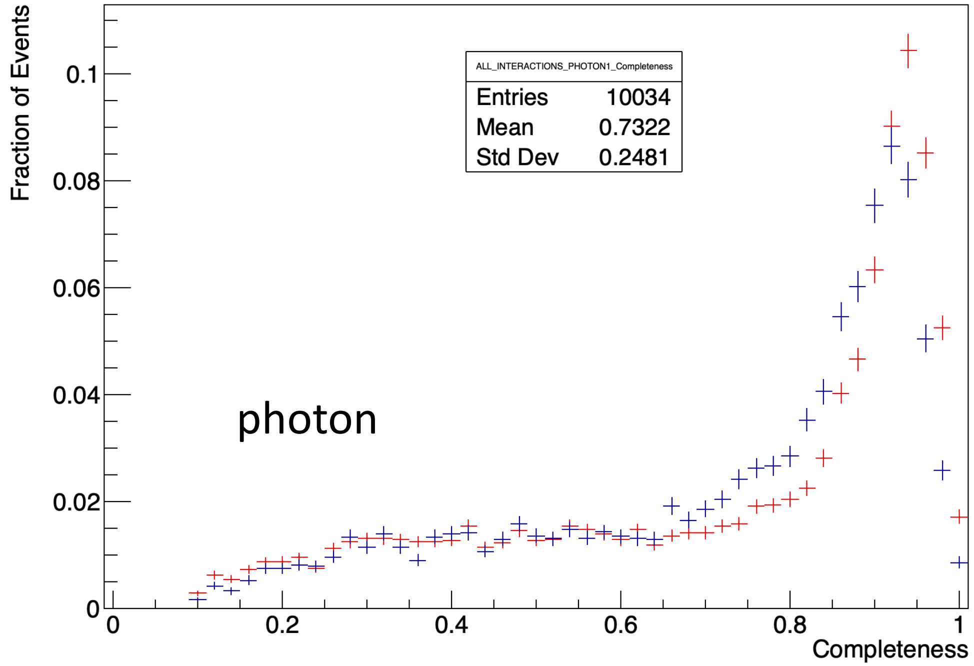
$$FOM = \frac{\sum_{Bins} \frac{(N_{OBS} - N_{EXP})^2}{N_{EXP}}}{total\ Energy}$$

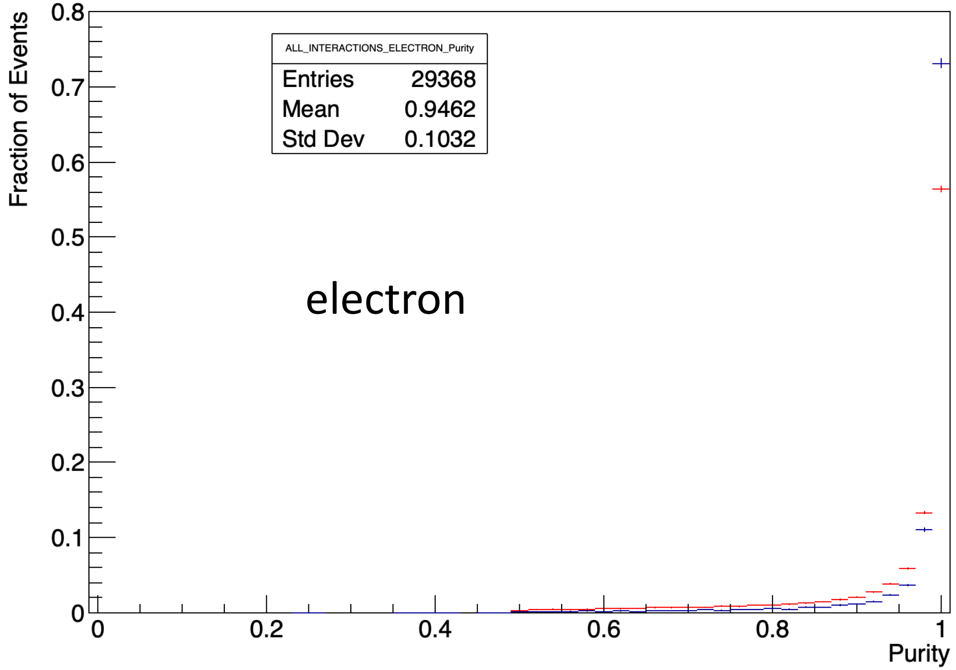
- If the FOM is smaller for a 2-shower prediction than for a 1-shower prediction, can decide to split cluster
- Can combine different FOMs, both calorimetric and topological, in multivariate approach

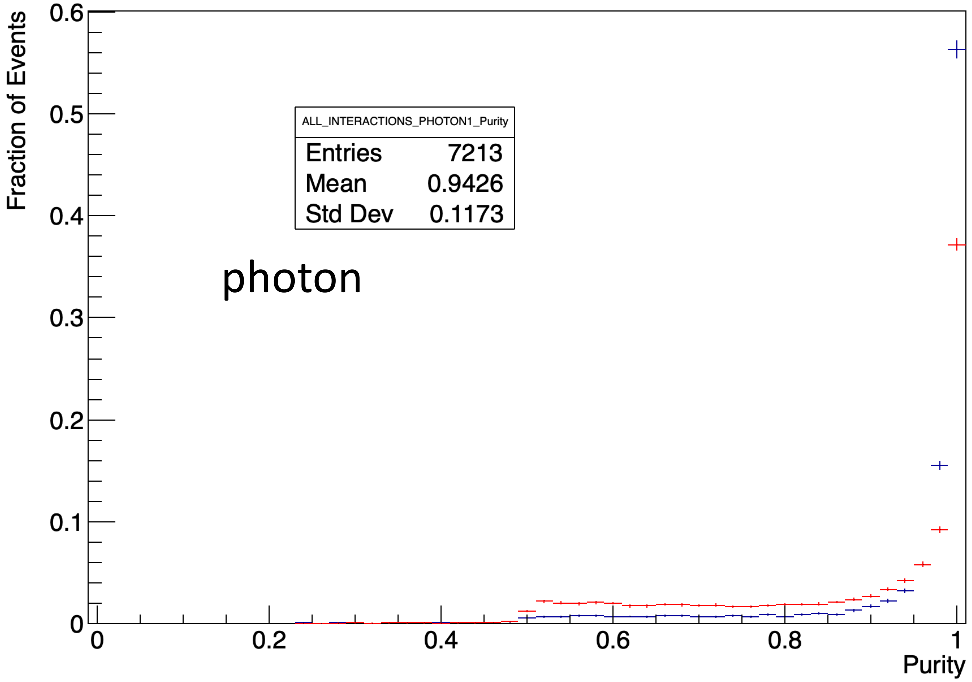
1000 showers sample
Ratio of FOM(2 showers) and FOM(1 shower) shows good separation




ALL_INTERACTIONS_ELECTRON_Completeness
Entries 29368
Mean 0.8641
Std Dev 0.1648
electron
Fraction of Events
Completeness


ALL_INTERACTIONS_PHOTON1_Completeness
Entries 10034
Mean 0.7322
Std Dev 0.2481
photon
Fraction of Events
Completeness


ALL_INTERACTIONS_ELECTRON_Purity
Entries 29368
Mean 0.9462
Std Dev 0.1032
electron
Fraction of Events
Purity


ALL_INTERACTIONS_PHOTON1_Purity
Entries 7213
Mean 0.9426
Std Dev 0.1173
photon
Fraction of Events
Purity

Transverse profiles - parametrization

- from arXiv:hep-ex/0001020v1

$$f(r) = \frac{1}{dE(t)} \frac{dE(t,r)}{dr},$$

$$\begin{aligned} f(r) &= p f_C(r) + (1-p) f_T(r) \\ &= p \frac{2 r R_C^2}{(r^2 + R_C^2)^2} + (1-p) \frac{2 r R_T^2}{(r^2 + R_T^2)^2} \end{aligned}$$

t = longitudinal shower depth in units of radiation length
τ = t/T = shower depth in units of the depth of the shower maximum
r = radial distance from the shower axis in Moliere radius units
E = shower energy in units of critical energy

ADC->MeV conversion factor=0.0075 MeV/ADC
Argon properties
https://pdg.lbl.gov/2014/AtomicNuclearProperties/HTML/liquid_argon.html

$$\begin{aligned} R_{C,hom}(\tau) &= z_1 + z_2 \tau \\ R_{T,hom}(\tau) &= k_1 \{\exp(k_3(\tau - k_2)) + \exp(k_4(\tau - k_2))\} \\ p_{hom}(\tau) &= p_1 \exp\left\{\frac{p_2 - \tau}{p_3} - \exp\left(\frac{p_2 - \tau}{p_3}\right)\right\} \end{aligned}$$

with

$$\begin{aligned} z_1 &= 0.0251 + 0.00319 \ln E \\ z_2 &= 0.1162 + -0.000381 Z \\ k_1 &= 0.659 + -0.00309 Z \\ k_2 &= 0.645 \\ k_3 &= -2.59 \\ k_4 &= 0.3585 + 0.0421 \ln E \\ p_1 &= 2.632 + -0.00094 Z \\ p_2 &= 0.401 + 0.00187 Z \\ p_3 &= 1.313 + -0.0686 \ln E \end{aligned}$$