Task 12.5: Particle Flow Reconstruction

John Back

WARWICK

on behalf of the Task 12.5 institutes

24th April 2023

Overview

Particle Flow Algorithms (PFAs)

State-of-the-art reconstruction for HEP calorimeters and neutrino detectors

Research Groups (main contacts)

- Dual Readout Calorimeters:
  - I. Vivarelli (Sussex), B. Di Micco (INFN Roma-3), S. Vallecorsa (CERN)

- **APRIL**, **A**lgorithm for **P**article **R**econstruction @ **IL**C:
  - G. Grenier (CNRS-IP2I), V. Boudry (CNRS-LLR)

- DUNE Near Detector reconstruction:
  - J. Marshall* & J. Back* (Warwick), M. Uchida & S. Dennis (Cambridge)

    * WP12.5 co-conveners

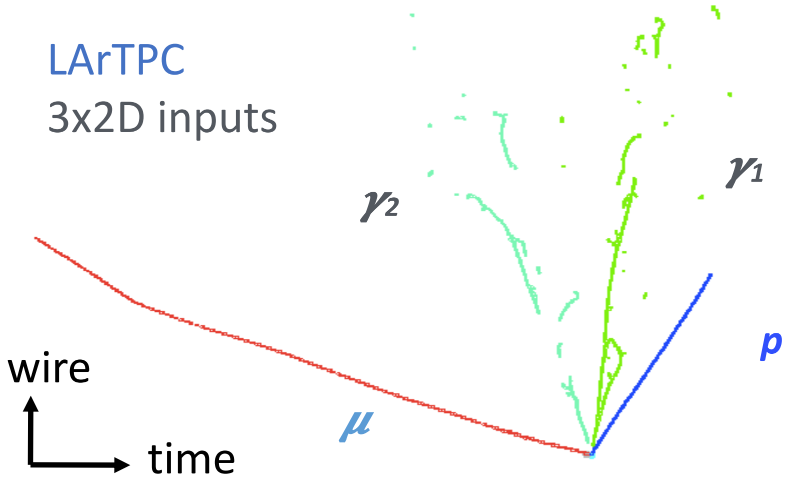
Pandora Software Development Kit

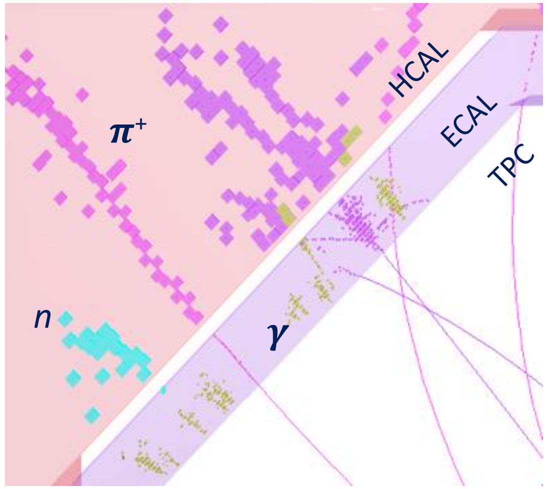
https://github.com/PandoraPFA

A single clustering approach is unlikely to work for complex event topologies:

- Mix of track-like & shower-like clusters
- Use **multi-algorithm** approach using the **Pandora SDK** to **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**
  - Algorithms: can use machine-learning methods & detector physics knowledge

LArTPC
3x2D inputs





ILC/CLIC
3D inputs

Particle Flow for Dual Read-Out Calorimeter

Adelina D'Onofrio[1], Michela Biglietti[1], Biagio Di Micco[2], Iacopo Vivarelli[3] , Sofia Vallecorsa[4]

[1]INFN - Roma Tre

[2]Roma Tre University

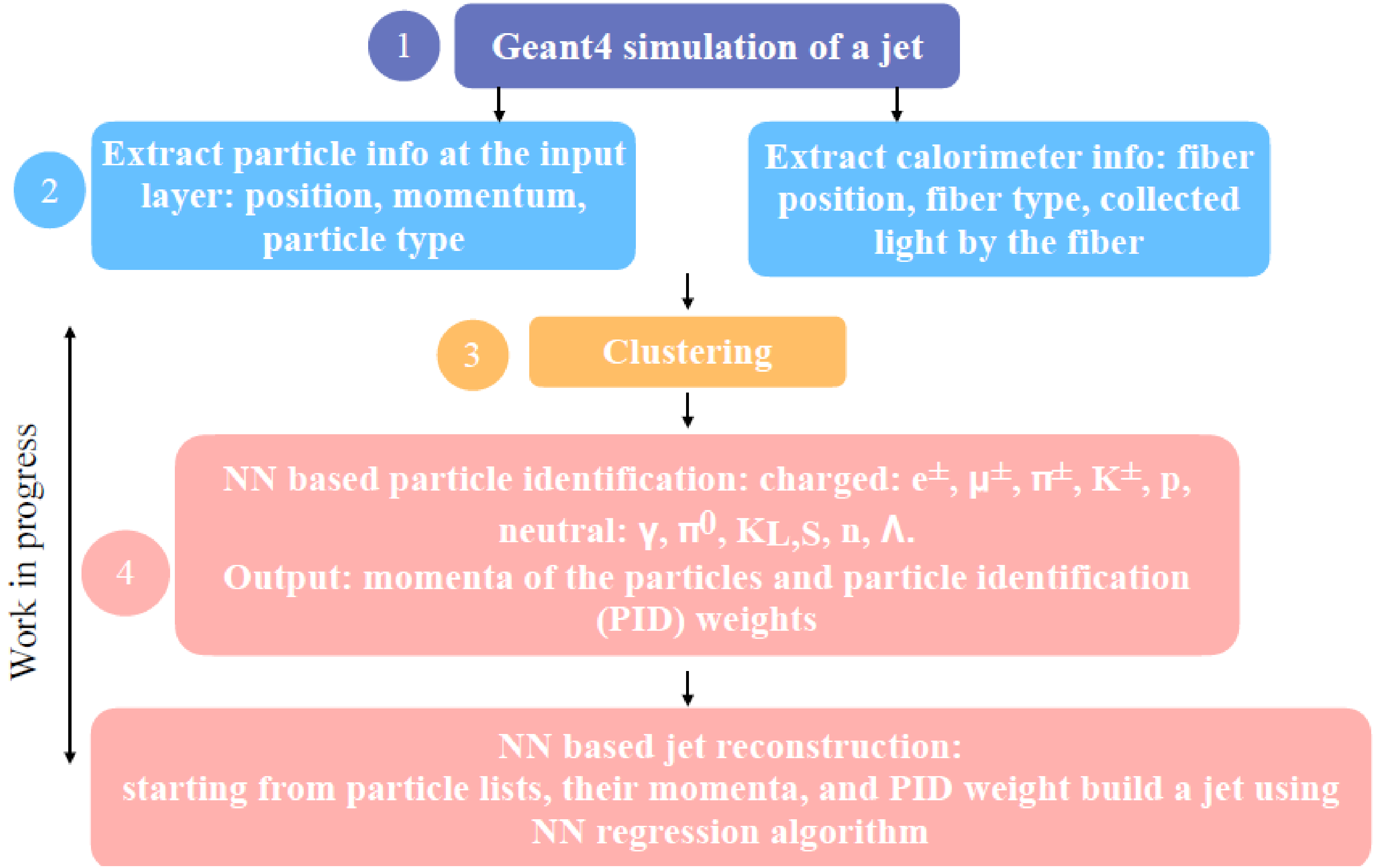
[3]University of Sussex

[4]CERN

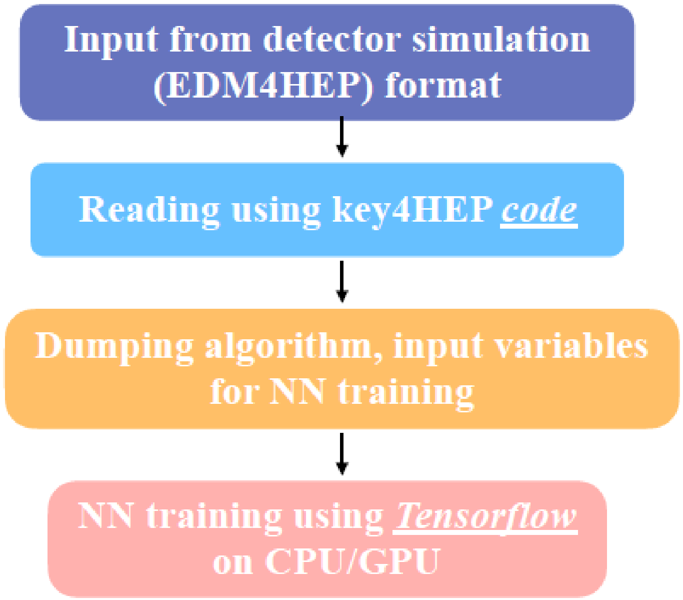
Overview of the Particle Flow Project

- The aim of the project is to build a Neural Network based algorithm that, from a given collection of energy deposits in the calorimeter, is able to completely reconstruct a jet in the detector and maximise the energy resolution of the dual read-out calorimeter

Workflow



Software Implementation



NN training using Tensorflow on GPUs

- Tensorflow, interfaced with Keras, is used to build and train a NN on GPUs
- Inputs: energy and position of each hit in the shower generated by the impinging electron and recorded in both S&C fibres—> NN input: 6 kinematic variables (*E, x, y, z, t, flag*) times hit multiplicity

- Two NN approaches tested

DNN approach
1. **10 hidden layers architecture**
2. **20 hidden layers architecture**

- Pro: fully exploits the fibres granularity in the calorimeter
- Cons:
  - Memory issues to process events in the full energy spectrum (0-125 GeV) for input electrons
  - Angular resolution not available

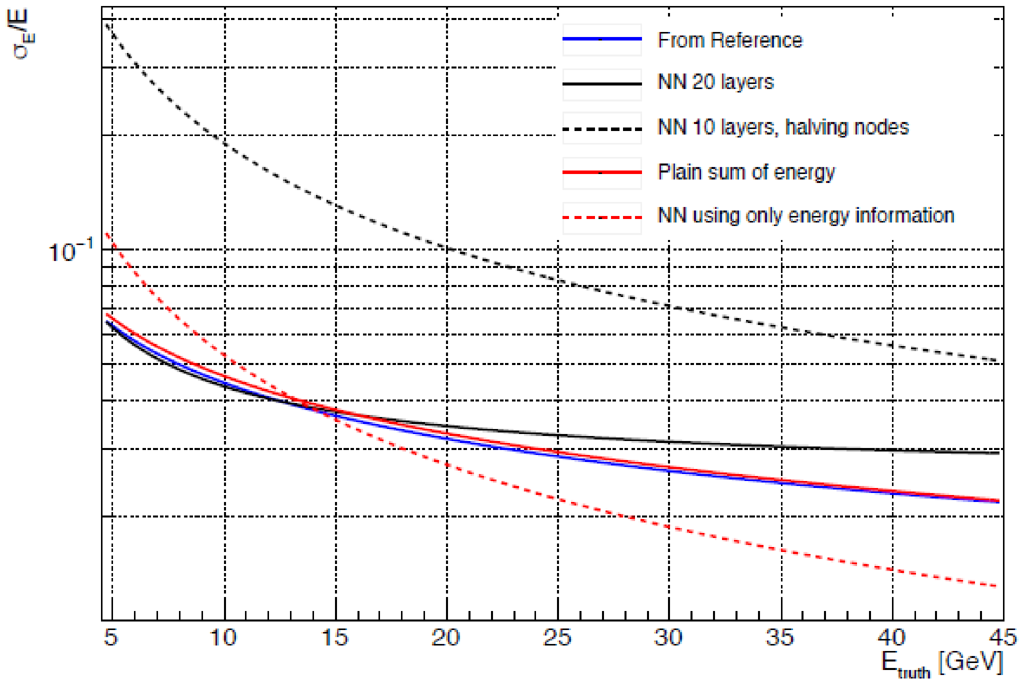
CNN approach
1. **VGG-like architecture**
2. **VGG-like architecture & proto-clustering**

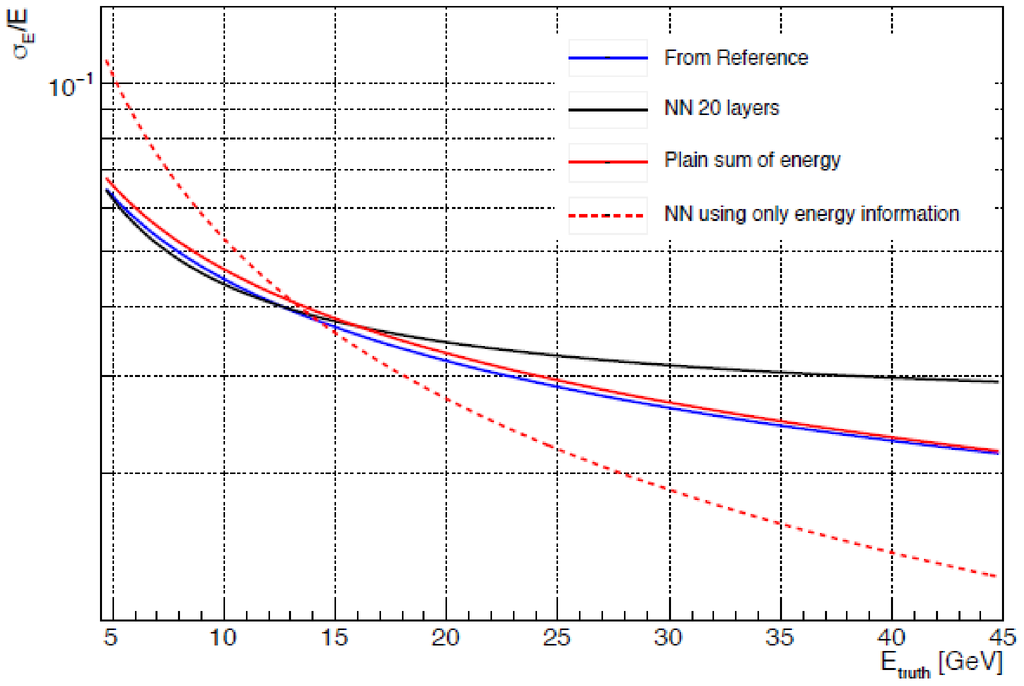
- Pro:
  - Solves the memory issues—>able to exploit electrons info in the full energy range
  - Able to obtain also the angular resolution
- Cons:
  - Further studies needed to improve the energy and angular resolution results

DNN results & next steps

DNN approach

- As a sanity check, we compared our energy resolution results with:
  - A reference https://inspirehep.net/literature/1861660
  - The energy resolution obtained simply summing up the energy deposits in the fibres (S&C)
- The energy resolution improves if we double the NN layers and we keep constant the number of nodes
- **Issue**: the NN performance is still worse than the standard reconstruction —> work in progress
- Next steps: increase the statistics of the simulation &improve the NN performance testing other (CNN) architectures





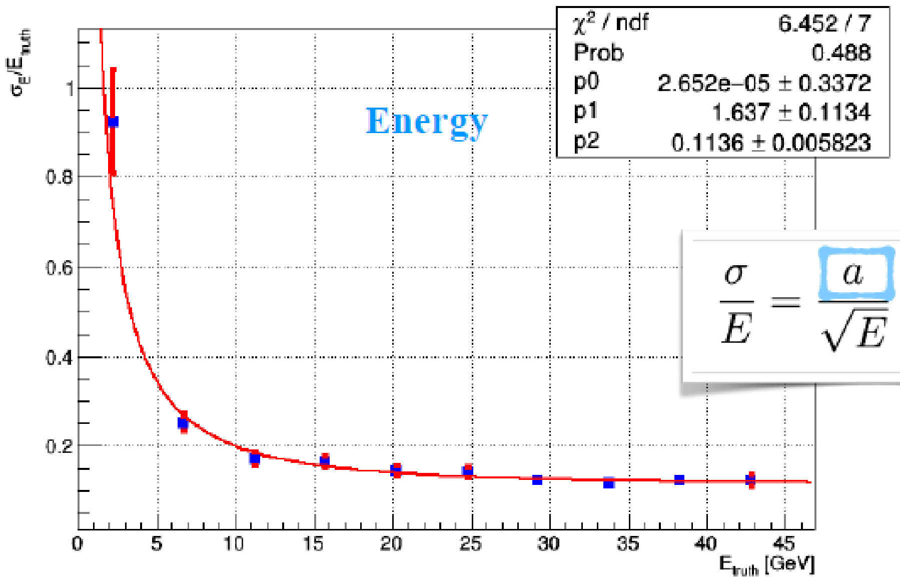
Preliminary results on electron energy resolution

CNN approach

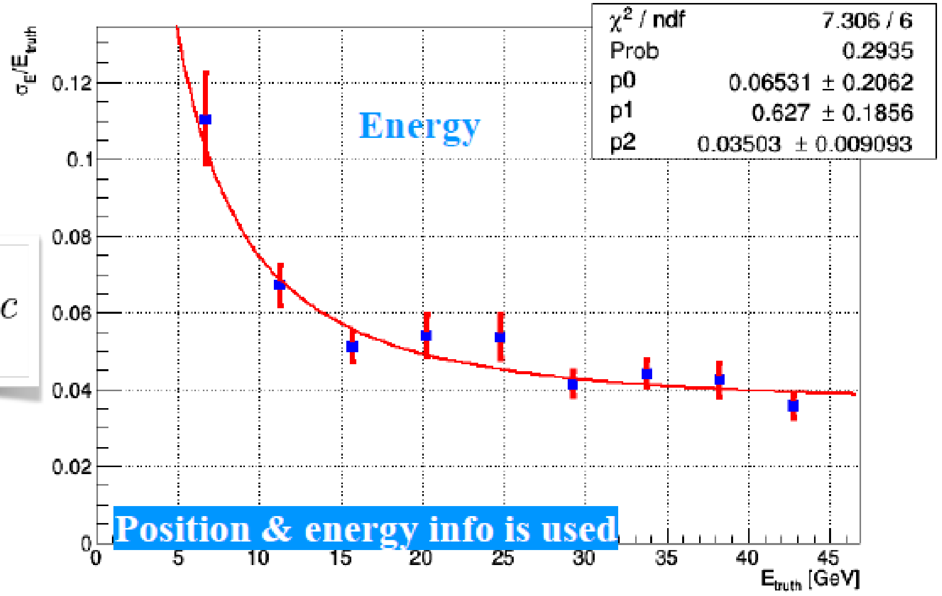
1. **VGG-like architecture w/o proto-clustering**

CNN approach

2. **VGG-like architecture with proto-clustering**



$$\frac{\sigma}{E} = \frac{\boxed{a}}{\sqrt{E}} \oplus \frac{b}{E} \oplus c$$



- 10k simulated electrons

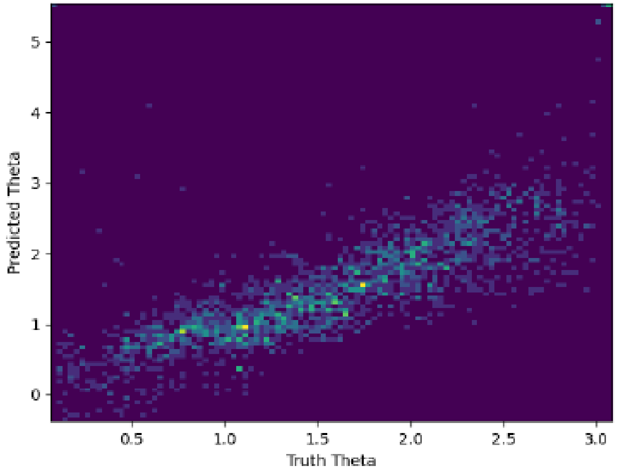
Hyperparameter optimisation:
applied on batch size & start learning rate
Batch Size = 64, Start Learning Rate = 10^{-4}

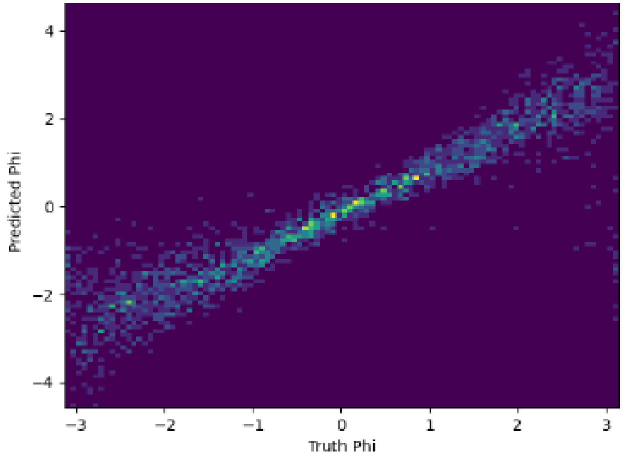
**Batch size: it is a number of samples processed before the model is updated*

**Learning rate: it is a hyper-parameter used to govern the pace at which an algorithm updates or learns the values of a parameter estimate*

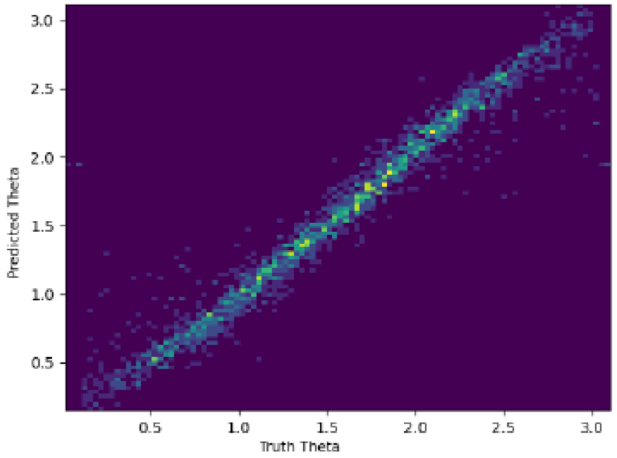
Preliminary results on electron angular resolution

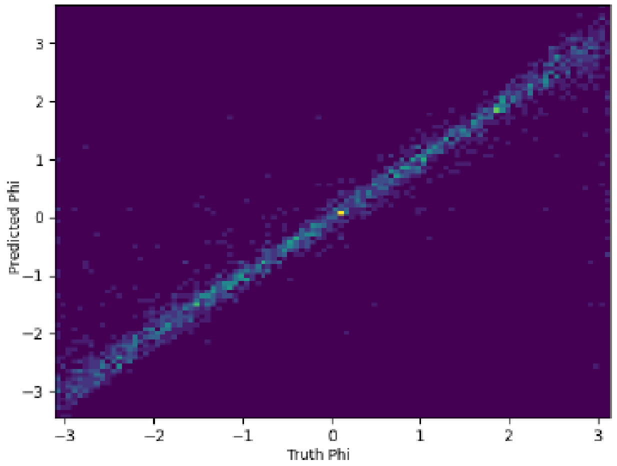
CNN approach
1. VGG-like architecture w/o proto-clustering





CNN approach
2. VGG-like architecture with proto-clustering





- Improvements observed if a pro-clustering is applied

Next Steps

- Further explore the possibility to perform a clustering before of feeding the NN
  - Using Pandora algorithms
- Test alternative NN approaches. like GNN
- Increase the number of electrons in the simulations
- Perform analogous studies in the case of other input particles like pions and kaons

Optimised APRIL PFA for hadronic jets

G. Grenier (CNRS-IP2I), V. Boudry (CNRS-LLR)

- Released software tools to generate samples for calibrating APRIL for the ILD: https://github.com/SDHCAL/SDHCAL_ILD_prod

- Masters students' 2022 summer internships on calibration:
  - Dijet MC generator level jet energy resolution
    - Finding energy ranges for photons & neutral hadrons to achieve accurate calibration
  - Optimisation of SiW-ECAL energy resolution
    - Comparing hit counting and energy sums, especially for low energy photons (< 10 GeV)

- CALICE test beam participation over 2022 summer
  - SDHCAL beam test completed 28th Sept

Optimised APRIL PFA for hadronic jets

APRIL PFA Have looked into calibration for ILD option 2

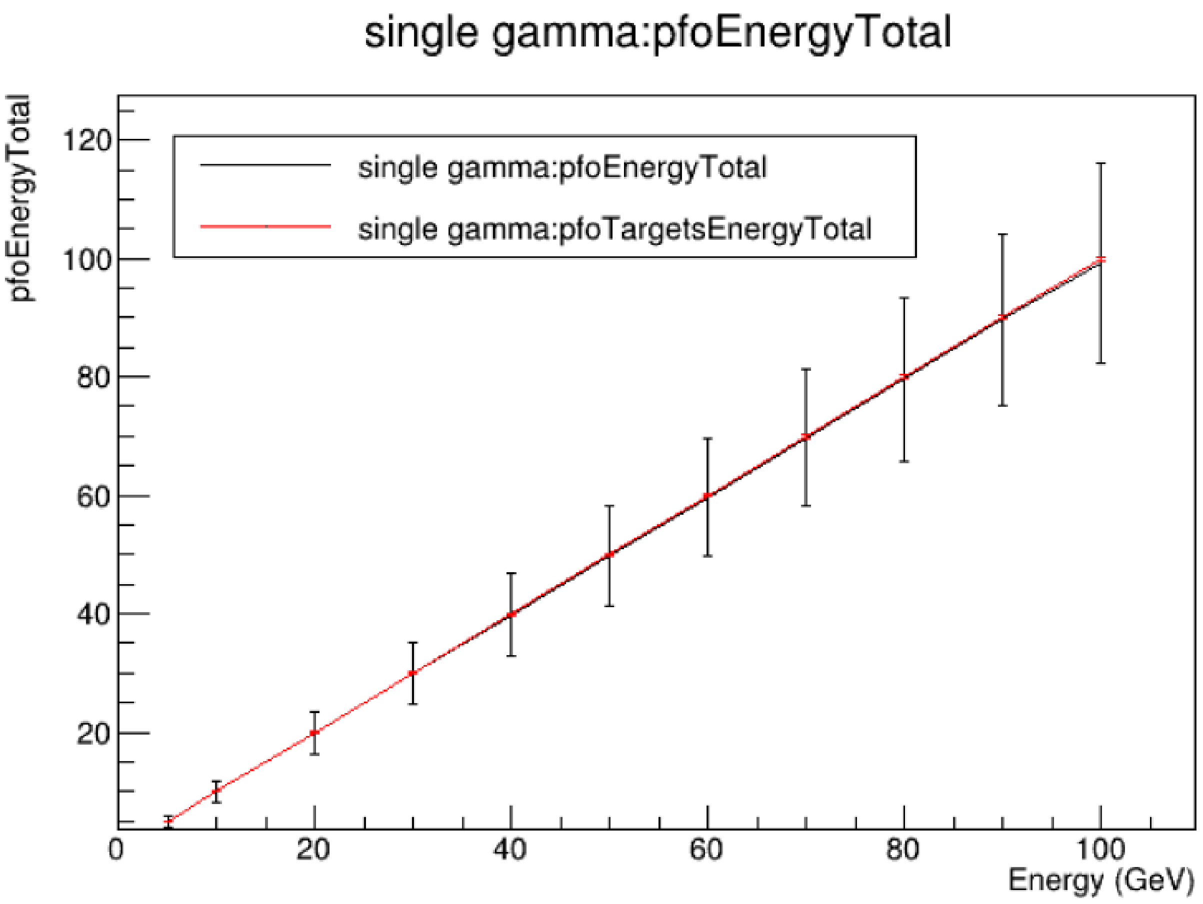
1) So far only ILD option 1 have been fully calibrated. Main change between the two are a change in Hadronic Calorimeter (from analogue AHCAL to semi-digital SDHCAL)
2) Have produced samples of $q\bar{q}$, single muons, single gammas and single klongs
   1) All samples have been reconstructed using ilcsoft and performing Pandora reconstruction with PerfectPFA.
   2) **Issue : no charged PFO reconstructed in $q\bar{q}$ sample.**
3) Using calibration procedure described in CalibrationPandoraAnalysisExplained.tex from PandoraPFA/LCPandoraAnalysis/doc/ on github.com
   1) Two steps method : first step calibrate the digitiser and second Pandora itself. **First step out of date.**
   2) Standard ilcsoft pandora assumes linear energy reconstruction for the hits : each hits has an energy (attributed by the digitisation process) and cluster energy is the sum of its hits energy.
      1) OK for ECAL and AHCAL.
      2) **Far from optimal for SDHCAL**.

Optimised APRIL PFA for hadronic jets

Pandora calibration for ILD option 2

1) **Calibration for ECAL : OK**

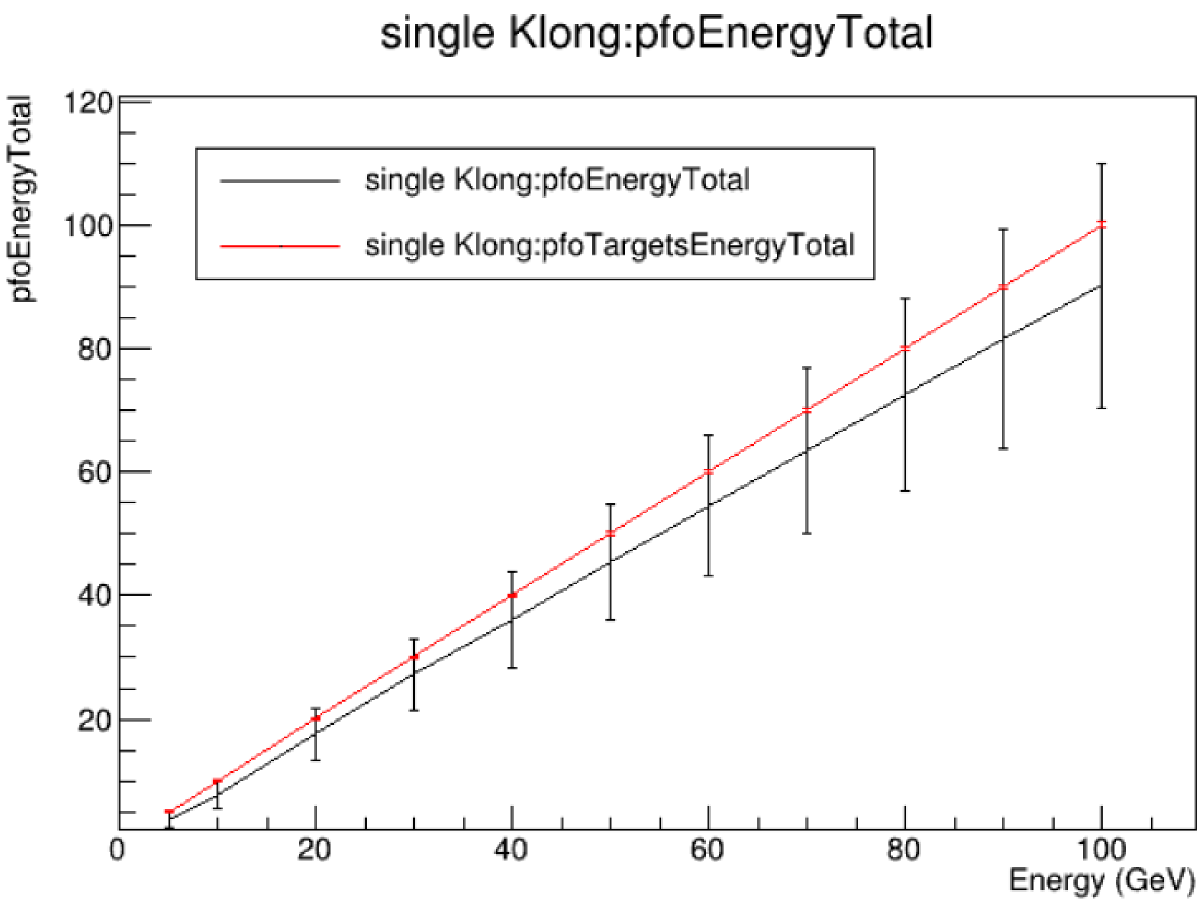
2) In the plot, error bars represents the width of the energy distribution for single gammas.



Optimised APRIL PFA for hadronic jets

Pandora calibration for ILD option 2

1) Calibration for SDHCAL : doesn't look that good but closest inspection shows
    1) Endcap is OK (see next slide)
    2) Barrel is too low (see next slide)
2) **SDHCAL is correctly calibrated but it lacks a correction** to correct cluster energy depending on the incidence angle of the cluster particles.
3) In the left plot, error bars represents the width of the energy distribution for single klongs.
4) Next step : implement angle correction.

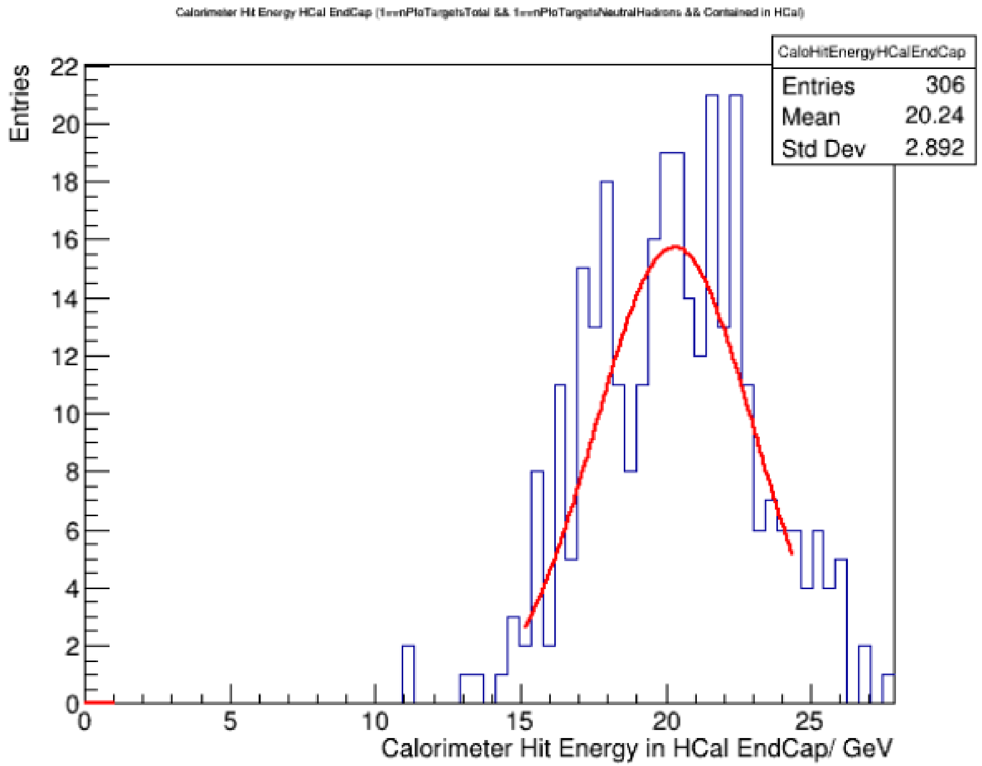


Optimised APRIL PFA for hadronic jets

20 Gev single klongs.

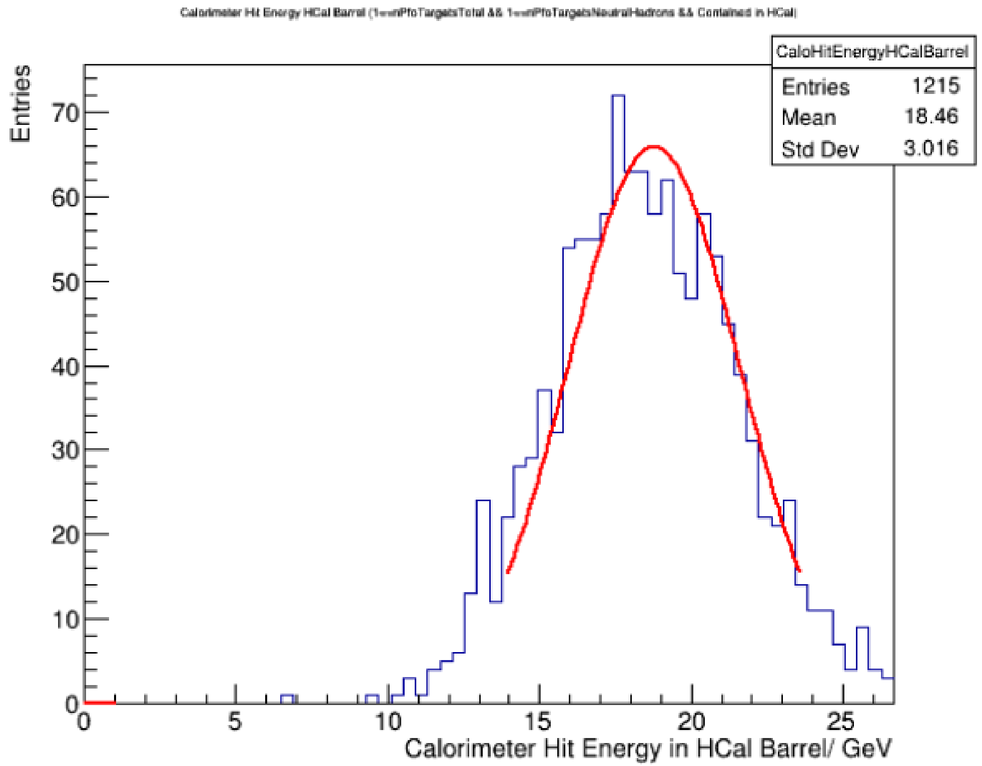
SDHCAL Endcap

1) Particles fly perpendicularly to layer surfaces.



SDHCAL Barrel

1) Particles do not fly perpendicularly to layer surfaces.



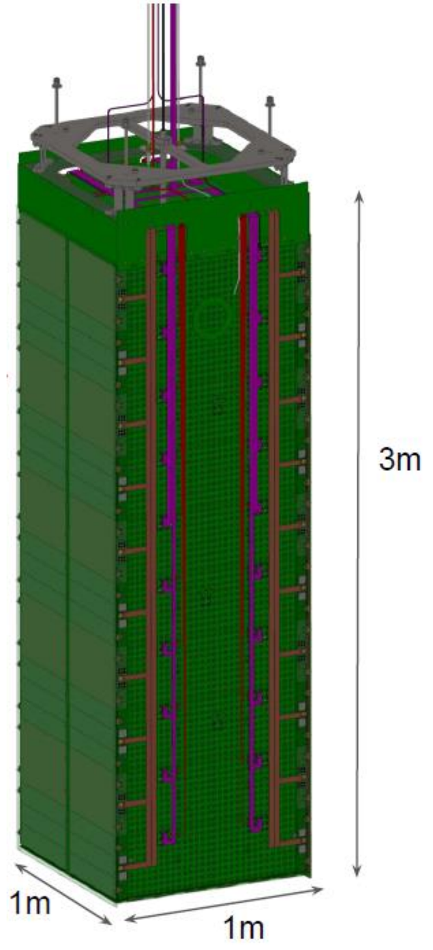
Optimised APRIL PFA for hadronic jets

- Assess PFA-only workflow from Marlin xml/processors for ILD Higgs MC samples
- Continue setting up APRIL for ILD option 2 (SDHCAL)
  - Revive current "how-to-run" APRIL PFA used for ILD option 1 (AHCAL)
    - Rémi Été & Bo Li, old version of DDMarlinPandora
  - Unresolved issue with using Pandora in ilcsoft, needs bug fix
  - Include energy correction in Pandora-like PFA & check calibration
  - Include APRIL in Pandora: DDMarlinPandora or other Key4Hep-Pandora interface?
- Explore adding time information in PFA & shower reco
  - Standalone study looks promising (not yet with PandoraSDK)
- Develop tools to compare PFA results between algorithms
  - PFO energy resolution, reco efficiency & purity
- Further develop AMSTER (reclustering for APRIL) and use it with PandoraSDK

Reconstruction for the DUNE Near Detector (ND)

John Back & John Marshall (Warwick),
Steve Dennis, Jingyuan Shi, Melissa Uchida, Leigh Whitehead (Cambridge), Tingjun Yang (Fermilab),
Munera Alrashed (Kansas State), Richie Diurba & Anja Gauch (Bern), Aleena Rafique (Argonne)

- ND LAr = 7x5 array of 1x1x3 m^3 modules,
  optically segmented LAr TPCs, 3D pixel readout
- 2x2 prototype: data taking during 2023



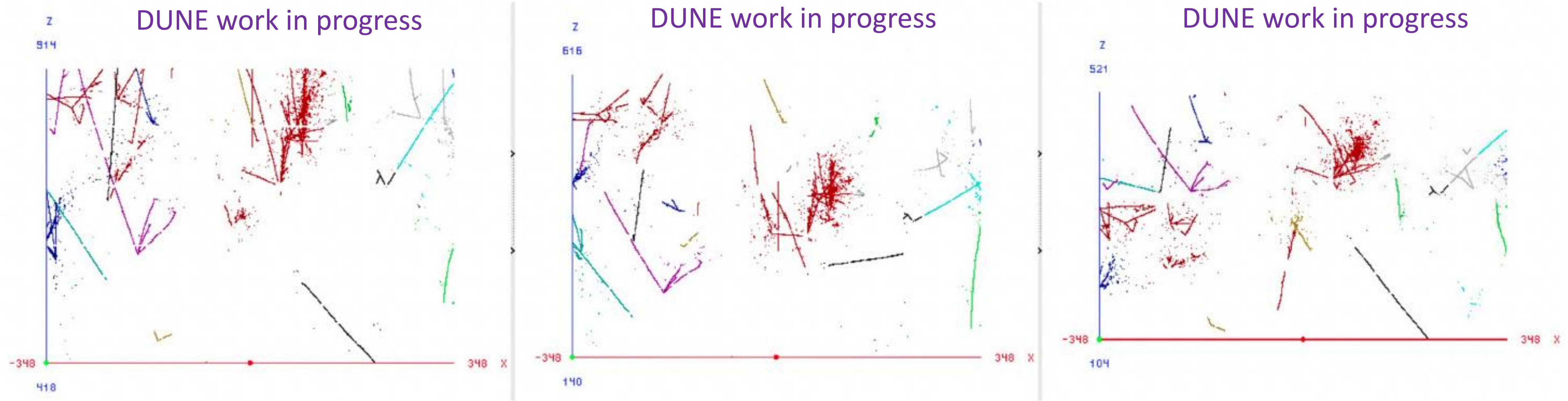
- Using **Pandora** for reconstructing 2x2 data
- "2x2 simulation challenge" underway
  - Centrally produced multi-neutrino events
  - larnd-sim digitisation applied to Geant4 (edep-sim) hits
  - HDF5 format; decoded for Pandora input

- Expect ~50 ν interactions per sec for 7x5 ND LAr
  - LBNF 120 GeV, 1.2 MW proton beam on graphite target
  - Secondary $\pi \rightarrow \mu\nu$: $7.5\text{x}10^{13}$ protons per beam "spill" (1.2 sec)

Multi-neutrino interactions reco: Slicing

Break up complex spills into independent ν interactions: **1 slice $\cong$ 1 ν**

1. Group together hits into Particle Flow Objects (PFOs) without using vertices
2. Find main ν vertex for each slice to improve PFOs

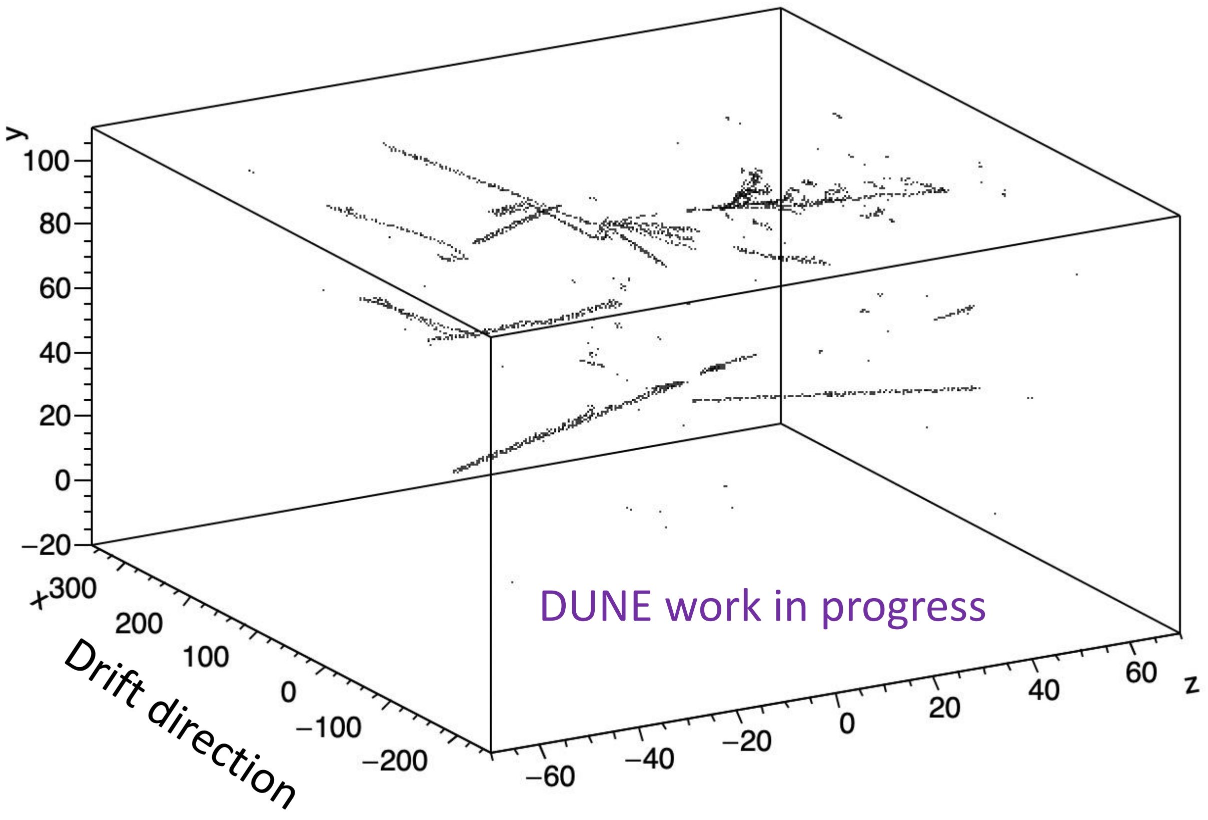


2D projections of reconstructed PFOs, **1 colour = 1 slice**

Looks reasonable, but needs improvement

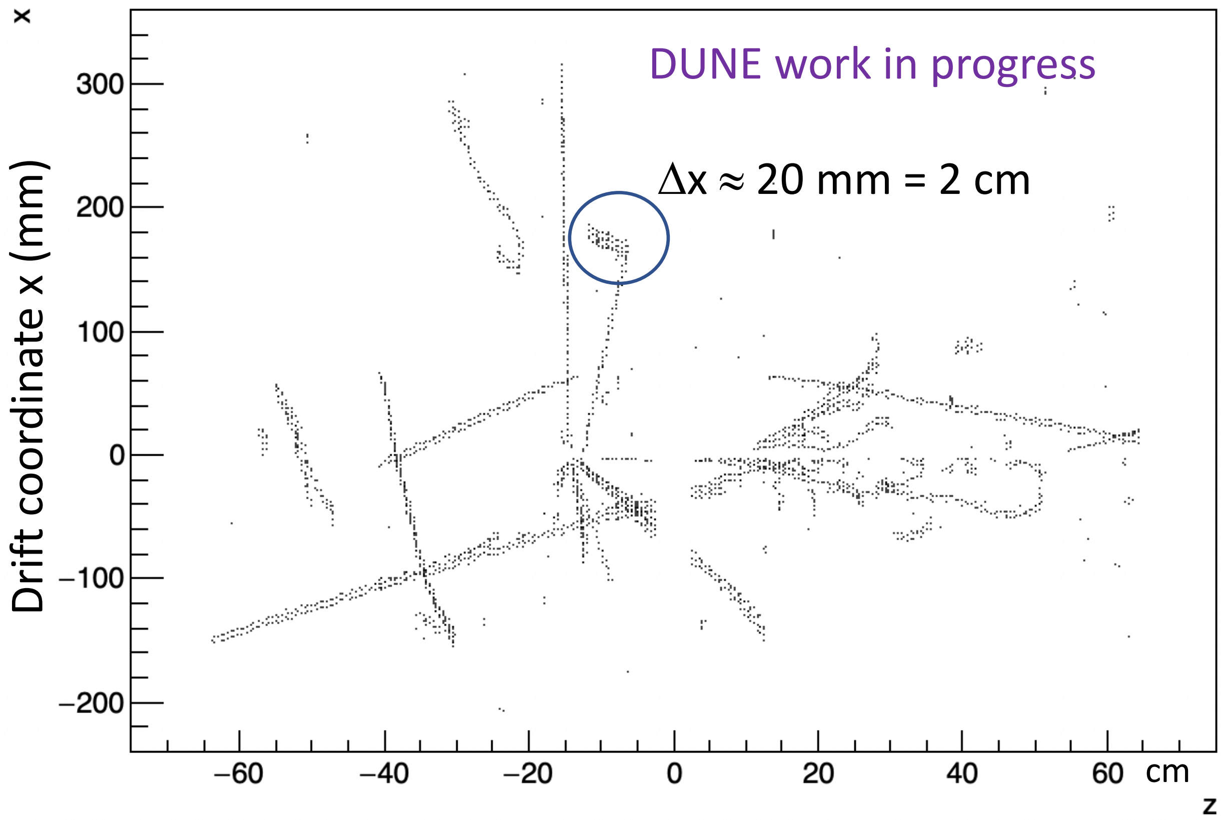
Need to quantify **performance** using Pandora **Hierarchy** tools (developed by Andy Chappell, Warwick)

2x2 simulation challenge example event



Several neutrino interactions

2D projection



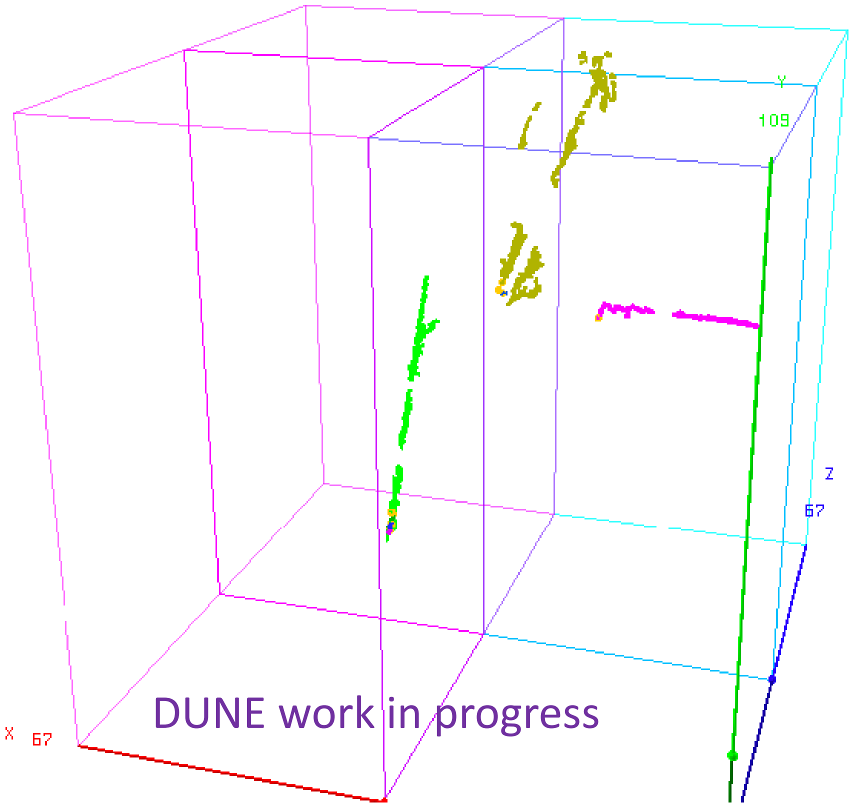
Tracks appear "fuzzy": few cm spread of hits along drift x
Caused by charge deposition affecting neighbouring pixels
DUNE ND calibration: smoothing algorithm in progress
(charge reweighting)

Pandora 3D ND reco developments

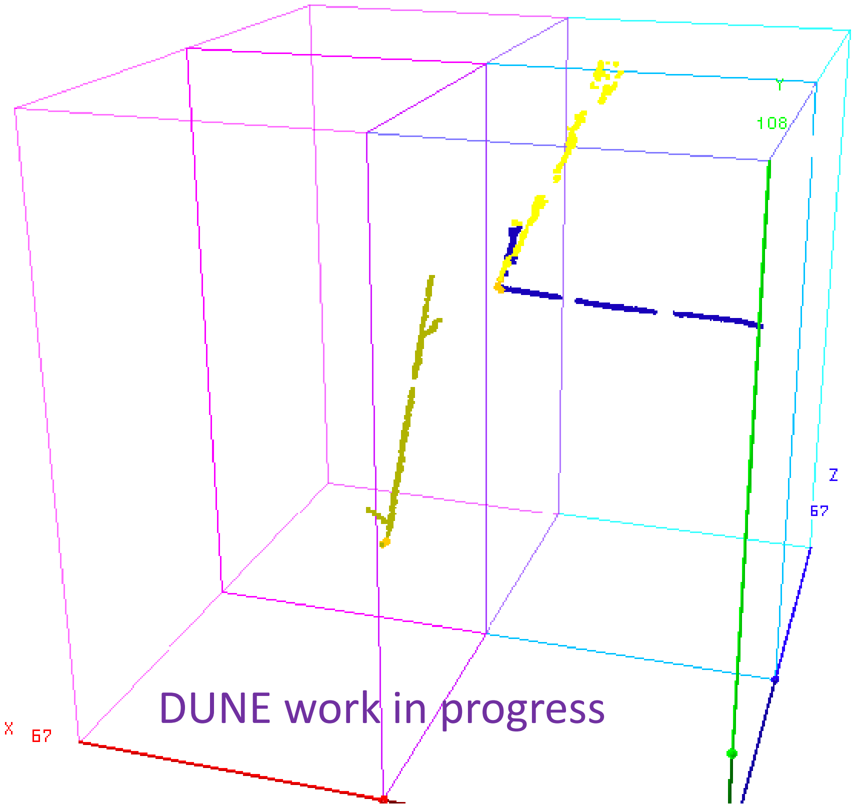
- Initial energy hits clustering performed using 3D coords
- Cluster merging and refinement
  - Apply existing 2D algorithms to the 3D hit clusters
  - Currently does not use y information (to be addressed)
- Clusters projected into 2D
  - Use neutrino reco algorithms based on 2D projections (e.g. MicroBooNE)
- Create Particle Flow Objects
  - Match 2D hits back to 3D
  - Build final tracks and showers

Pandora reco of 2x2 simulation events

MicroBooNE Reconstruction

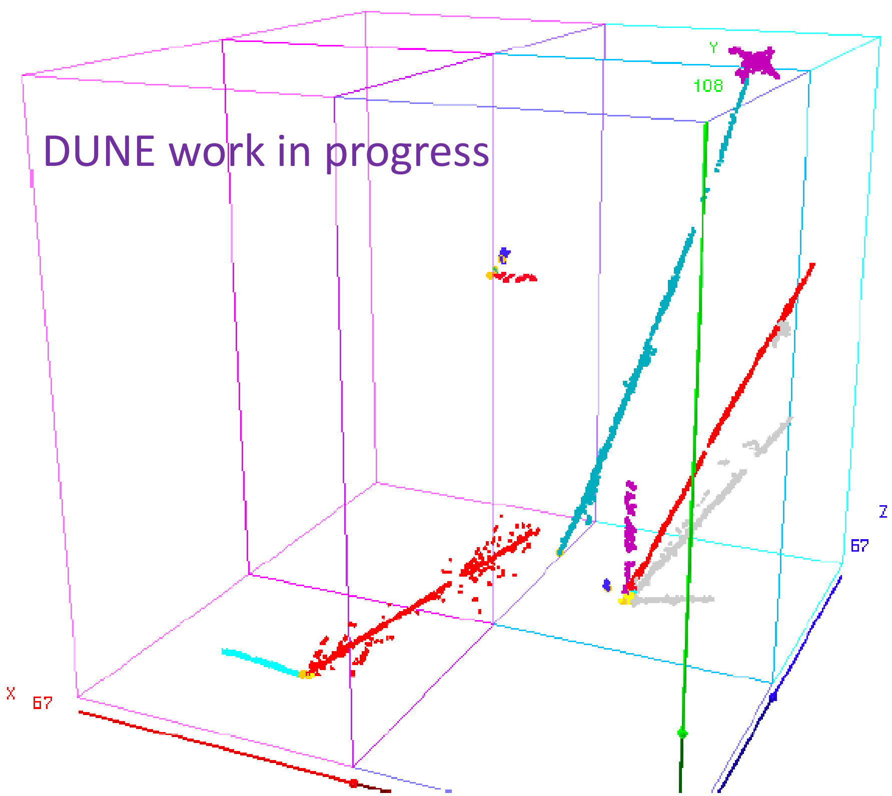


“3D” ND Reconstruction

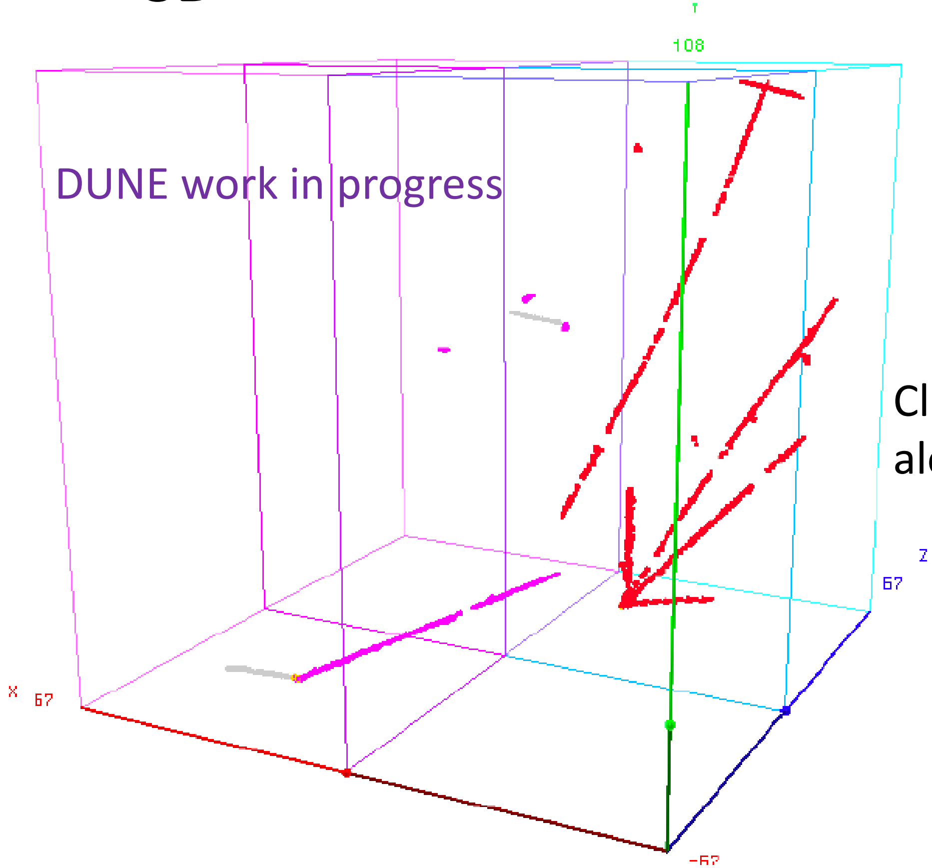


Pandora reco of 2x2 simulation events

MicroBooNE Reconstruction



"3D" ND Reconstruction



Clusters overlap along y projection

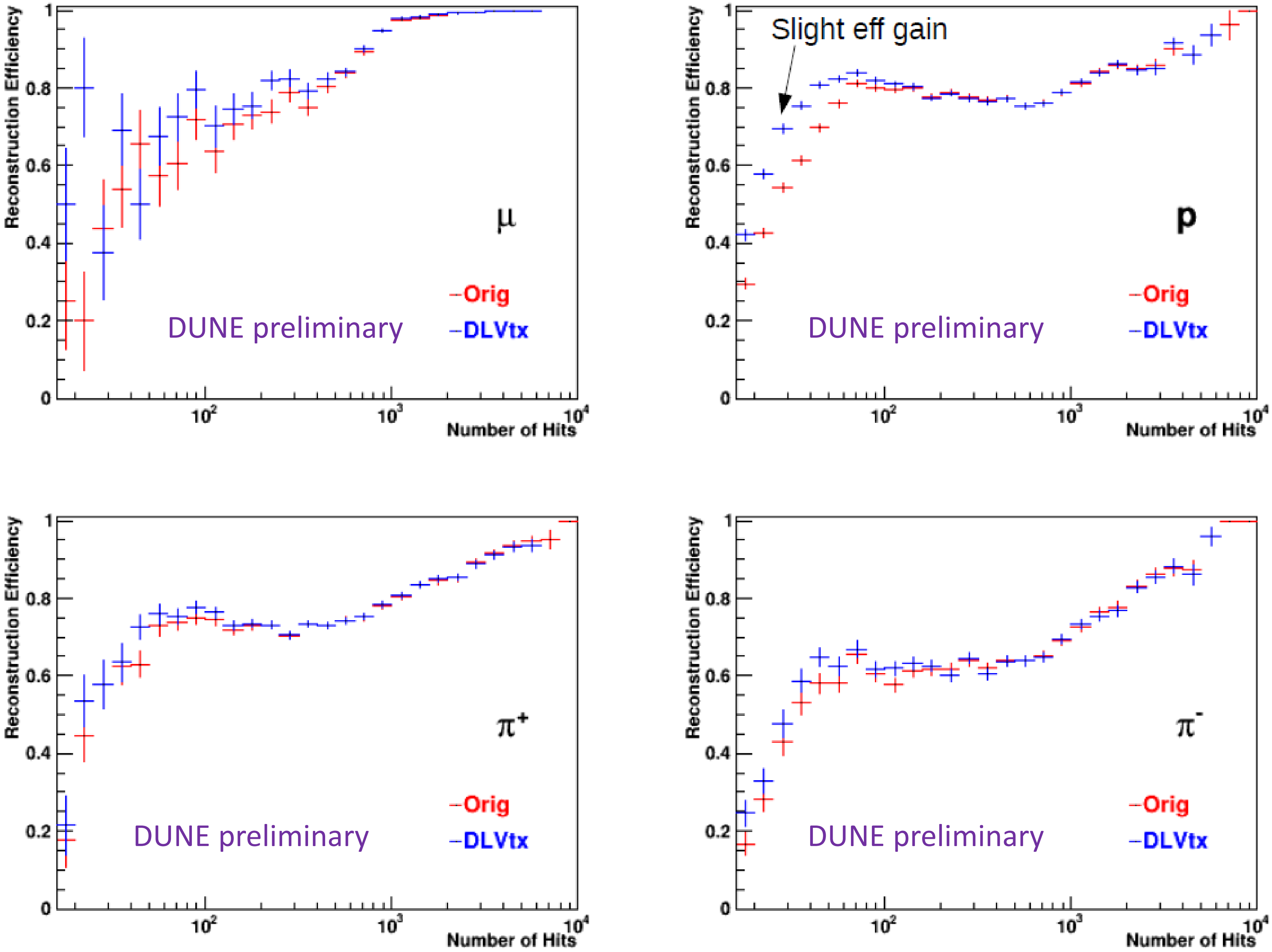
Deep Learning Vertexing

- Trying out Deep Learning (DL) vertexing for DUNE ND
  - Pandora **MicroBooNE neutrino** algorithms
  - **LArDLVertexing** algorithm: trained for **DUNE far detector** (A Chappell)
  - Reusing algorithm & parameters for DUNE ND: **no retraining done yet**

- Using 100k single ν_μ interaction simulated events (no spills)

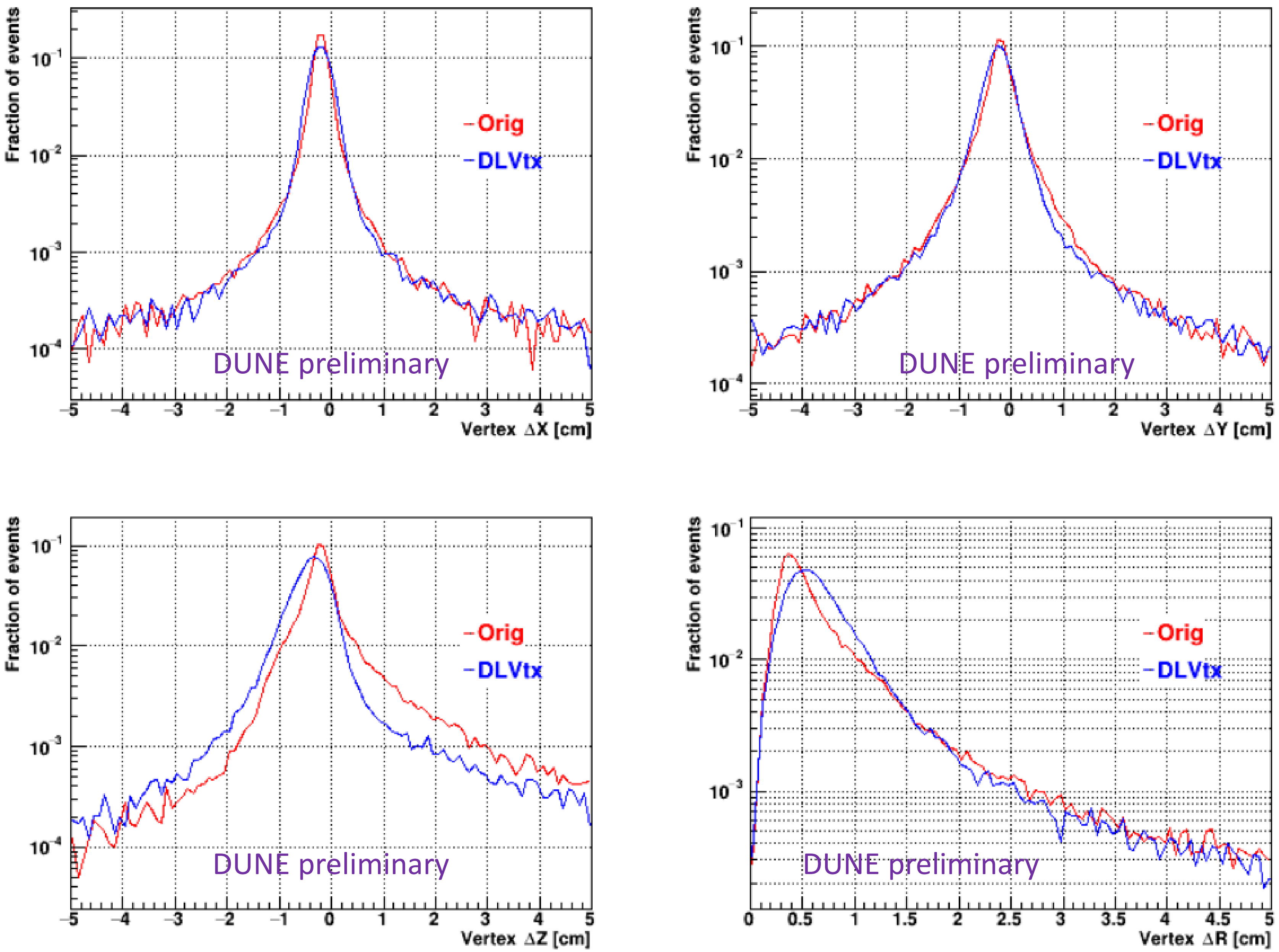
- Comparing ND reco performance with/without DL vertexing
  - Original: MicroBooNE neutrino algorithms only
  - DLVtx: MicroBooNE neutrino algorithms with DL vertexing

ν_μ events: particle reco eff vs number of hits



Original
DLVtx

ν_μ events: reco – MC vertex residuals (log scale)



Original
DLVtx

Peak ΔR ≈ 4 mm
Peak ΔR ≈ 5 mm

Using network trained for LArTPC far detector

Pandora DUNE ND summary

- LArRecoND package created for reconstruction developments

- Reconstruction for 2x2 LAr ND prototype
  - Using data converted from HDF5 format files
  - 2x2 simulation challenge
  - Developing 3D methods

- 3D algorithms
  - Initial 3D hit clustering
  - Cluster merging: 2D projections, need changes to also use **y** coord
  - Final tracks & showers: match 2D hits back to 3D

- Deep learning vertexing
  - Needs to be retrained for ND

Summary

- Dual readout calorimeters
  - NN jet reconstruction & PID
  - TensorFlow, looking at other options (PyTorch?)
- APRIL
  - Energy calibration
  - Developing reco for SDHCAL (ILD option 2)
  - Include APRIL in Pandora
- DUNE
  - Near Detector reconstruction
  - 2x2 LArTPC prototype, 3D info, slicing, deep learning vertexing