Task 12.5: Particle Flow Reconstruction

John Back

WARWICK

on behalf of the Task 12.5 institutes

18th March 2024

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 101004761.

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

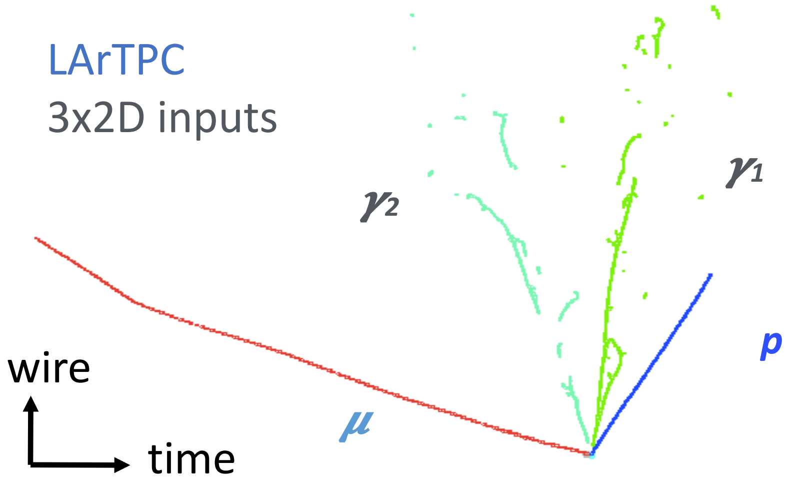
https://github.com/PandoraPFA

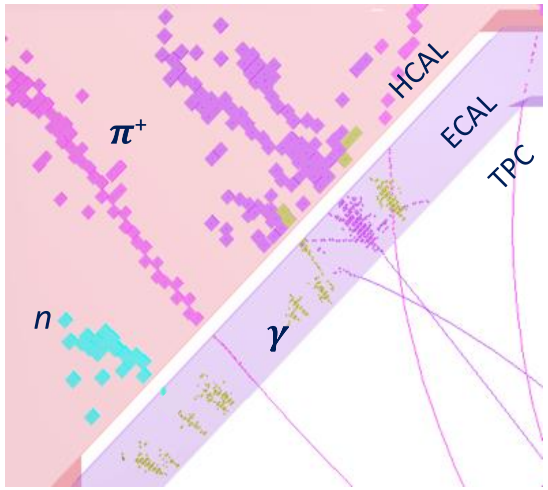
A single clustering approach is unlikely to work for complex event topologies:

- 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

Jet Particle Flow reconstruction in Dual Readout Calorimeters

B. Di Micco, S. Vallecorsa, I. Vivarelli

M. Biglietti, A. Farilla, R. Di Nardo, A. D'Onofrio

Dual Readout Calorimeter

Improve energy resolution of hadronic calorimeters: measuring the EM fraction using scintillating & Cerenkov fibre readout

Calorimeter layout simulated with GEANT-4 for the IDEA detector (**I**nnovative **D**etector for future **e**+e- **A**ccelerators)

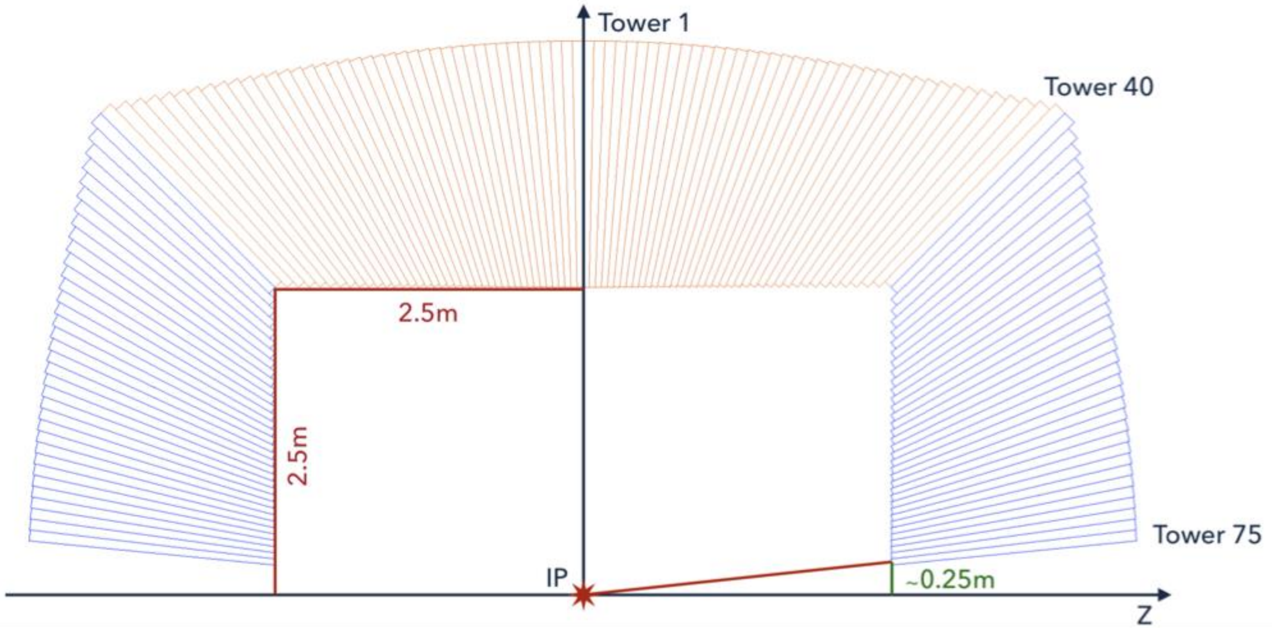
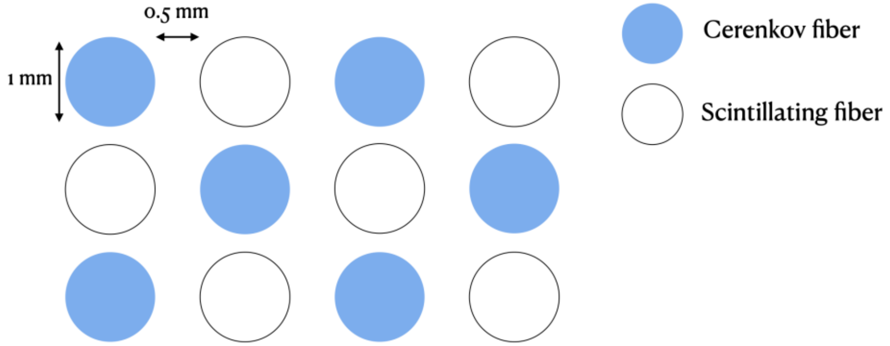
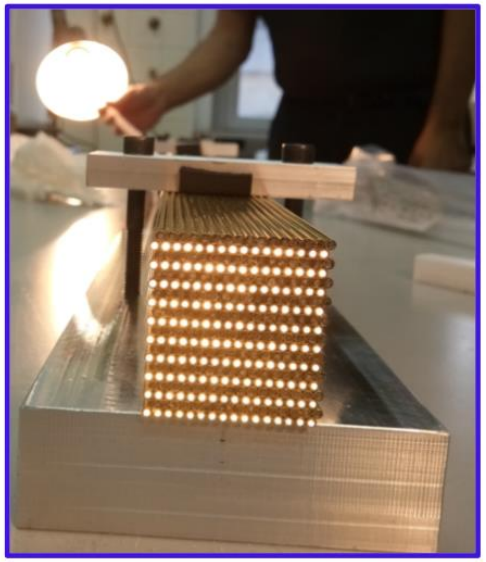


Figure 1: Sketch of a single slice of the IDEA calorimeter.

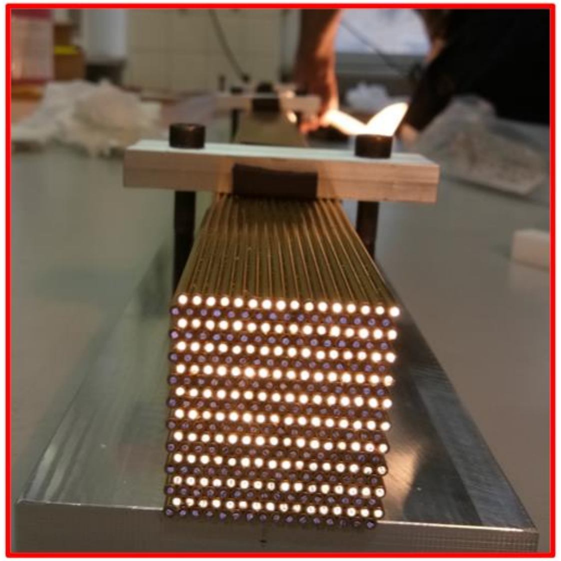


Scintillating fibers: polystyrene
Cherenkov fibers: PMMA (PolyMethyl MethAcrylate)

Passive material: copper



Scintillation fibers



Cherenkov fibers

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

1. Geant4 simulation of a jet ✓
2. Extract particle info at the input layer: position, momentum, particle type — Extract calorimeter info: fiber position, fiber type, collected light by the fiber ✓
3. Clustering
4. NN based particle identification: charged: $e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$, p, neutral: γ, π^{0}, $K_{L,S}$, n, Λ. Output: momenta of the particles and particle identification (PID) weights

NN based jet reconstruction: starting from particle lists, their momenta, and PID weight build a jet using NN regression algorithm

Work in progress (steps 3 onward)

Software Implementation

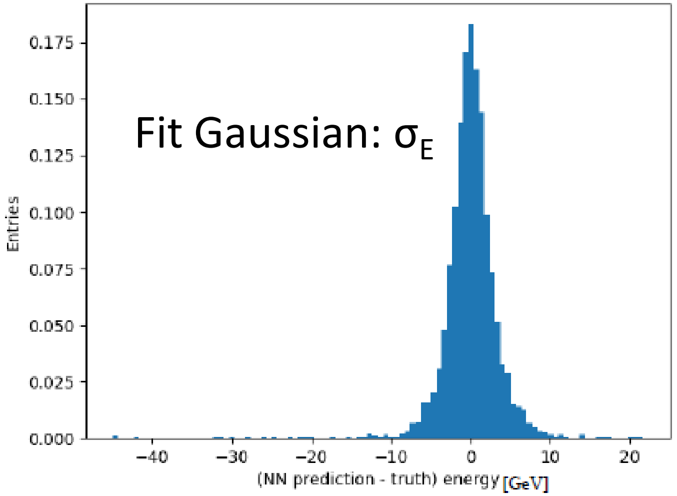
- Input from detector simulation (EDM4HEP) format
- Reading using key4HEP *code*
- Dumping algorithm, input variables for NN training
- NN training using *Tensorflow* on CPU/GPU

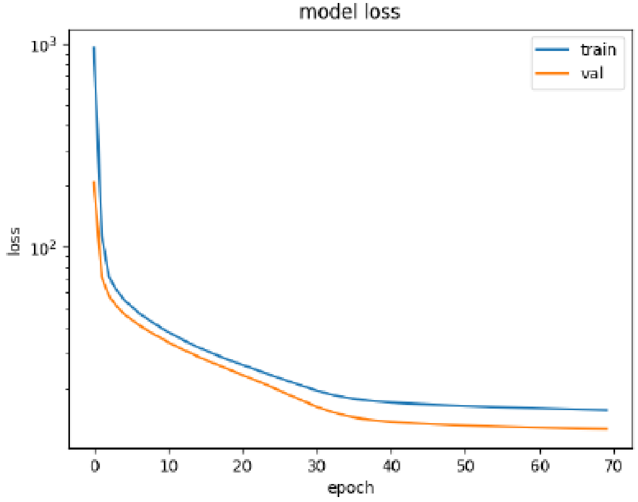
Use Pandora as algorithm collector
Interface via KEY4HEP & EDM4HEP

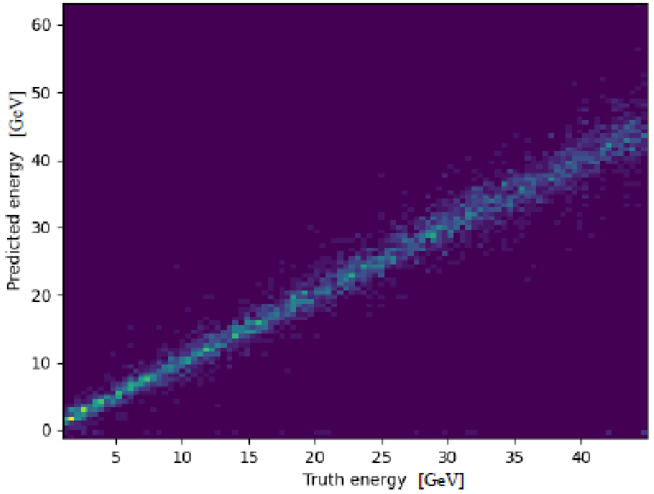
DNN approach: electron energy resolution

TensorFlow & Keras (CPU & GPUs)

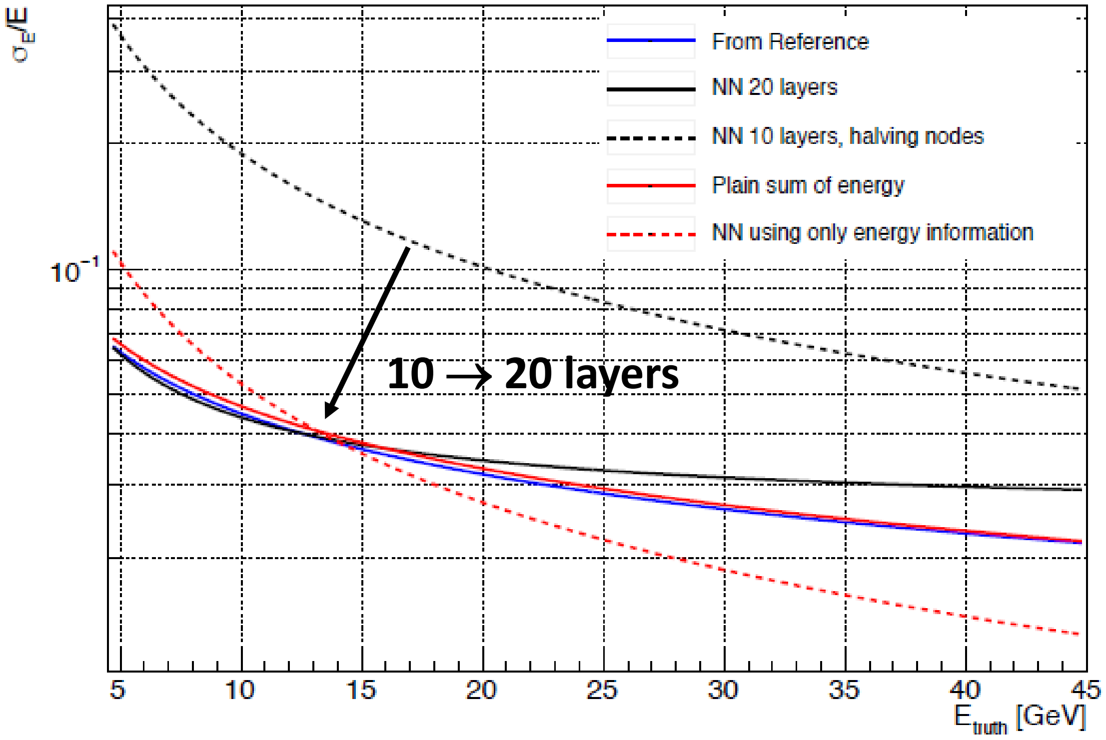
- First attempt: 10 hidden layers
- Halving number of nodes at each layer
- Model loss: MeanSquaredError()
  $$\frac{1}{n}\sum_{i=1}^{n}\left(y_{\text{true}} - y_{\text{pred}}\right)^2$$, optimised with respect to the simulated energy of the incoming electrons
- Adam, a stochastic optimiser, is used as optimiser to minimise the loss *Reference*







Prediction vs Truth E



Physics seems too complex for NN to learn it, hit position seems to add confusion more than information. Puzzled by this we tried different NN architectures.

What NN needs to learn: electron bending in a magnetic field, electron radiation, fiber geometry:
the x, y, z coordinates are the 3D coordinate of the fiber endpoint pointing to the Interaction Point.

CNN approach with proto clustering

- as for DNN, CNN is also not able to reconstruct the electron position in the calo, tried a proto-clustering approach: try to simplify the CNN work by defining the average position of the energy deposits, all fired fiber coordinates are defined respsect to the proto-cluster position.

Clustering ?

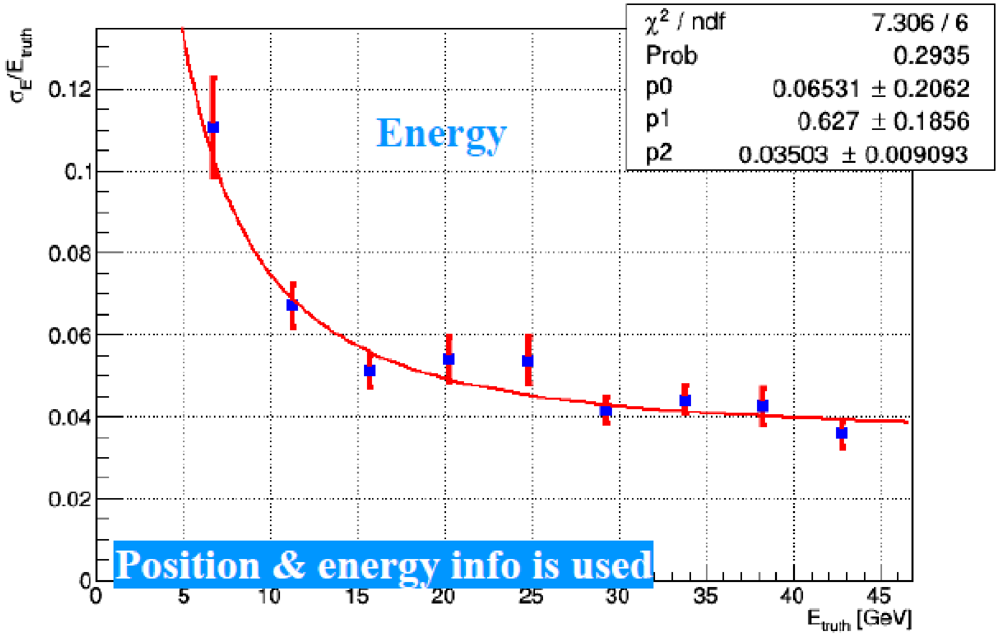
◆ **Clustering seems the obvious way to simplify conceptually the algorithm**

1. Identify energy deposits released by a single particle, collect them, and apply energy regression at cluster level;

2. Preliminary test: hit energy and distance wrt the centroid used as NN input

$$d_i = \text{position}(x,y,z)_i - \frac{\sum_{i=0}^{i=N_{hits}} (\text{position}(x,y,z)_i \cdot \text{energy}_i)}{\sum_{i=0}^{i=N_{hits}} \text{energy}_i}$$

3. Next step: exploit clustering algorithm in the Pandora framework

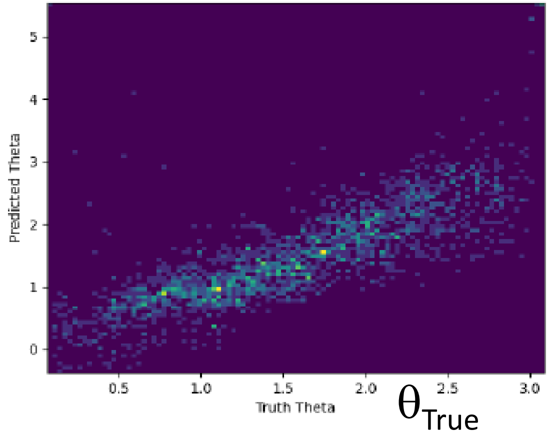


CNN approach angular resolution

- Improvements observed if a proto-clustering is applied

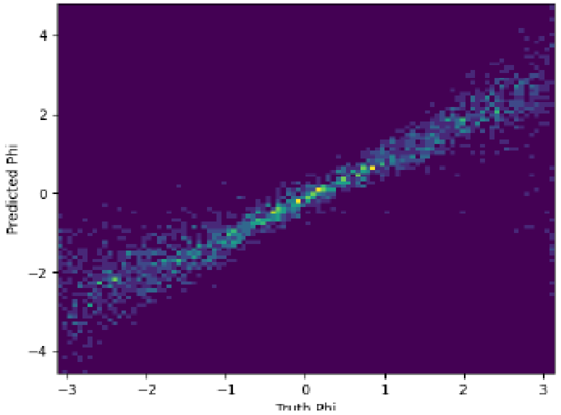
CNN approach
1. VGG-like architecture w/o proto-clustering

θ_{CNN}



θ_{True}

ϕ_{CNN}

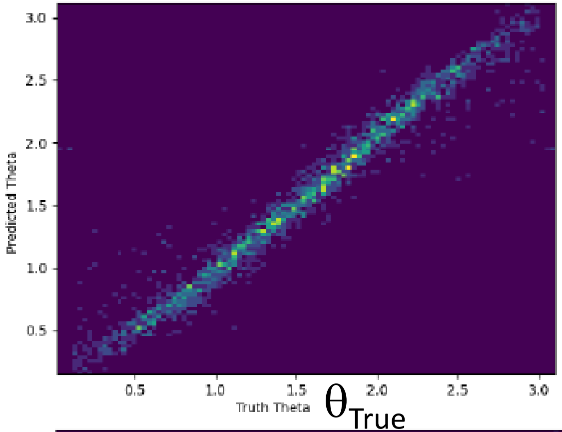


ϕ_{True}

Proto clustering →

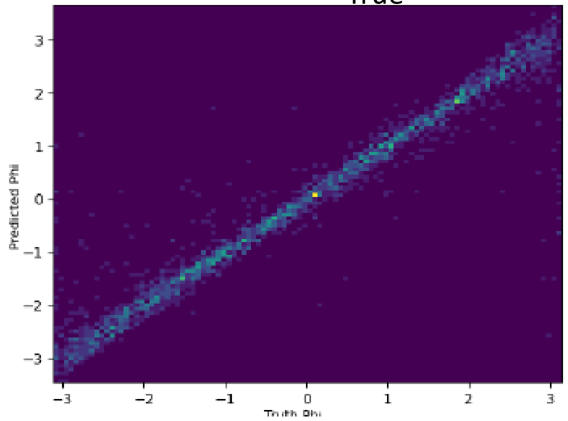
CNN approach
2. VGG-like architecture with proto-clustering

θ_{CNN}



θ_{True}

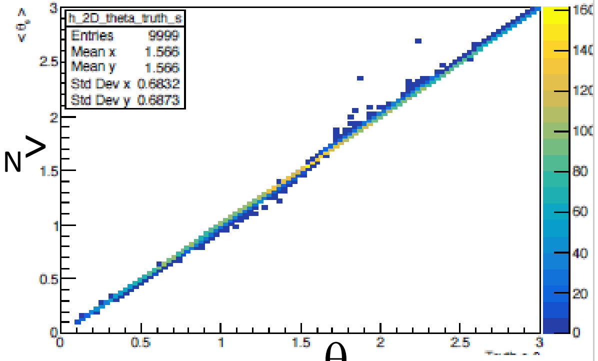
ϕ_{CNN}



ϕ_{True}

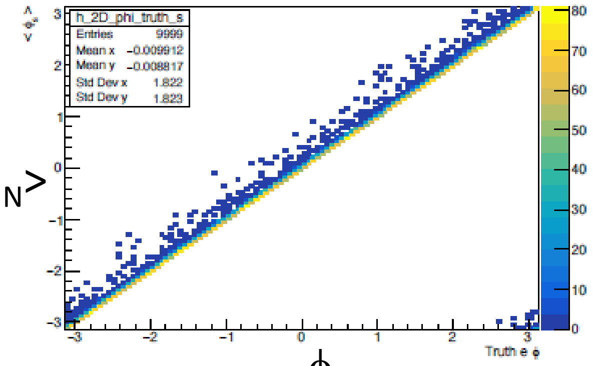
proto-cluster centroid

$<\theta_{CNN}>$



θ_{True}

$<\phi_{CNN}>$



ϕ_{True}

improvement observed, but it doesn't look CNN is adding anything on top of the proto-cluster (it probably deteriorates it)

Thoughts and plans for DR Calorimeter Pandora work

1. we were too much optimistic on the NN ability to solve problems for us :-)
2. we need to change approach, add NN on top of a classical reconstructions where known problems and features are solved (radiation, magnetic field bending, fiber geometry)
3. we need to develop a pre-reconstruction algorithm and use it as inputs to the NN together with row infos
4. this is time consuming, difficult to do it for all particles;
5. we need to re-think to the whole project (objectives, descoping)

Optimised APRIL PFA for hadronic jets

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

AIDAinnova WP12

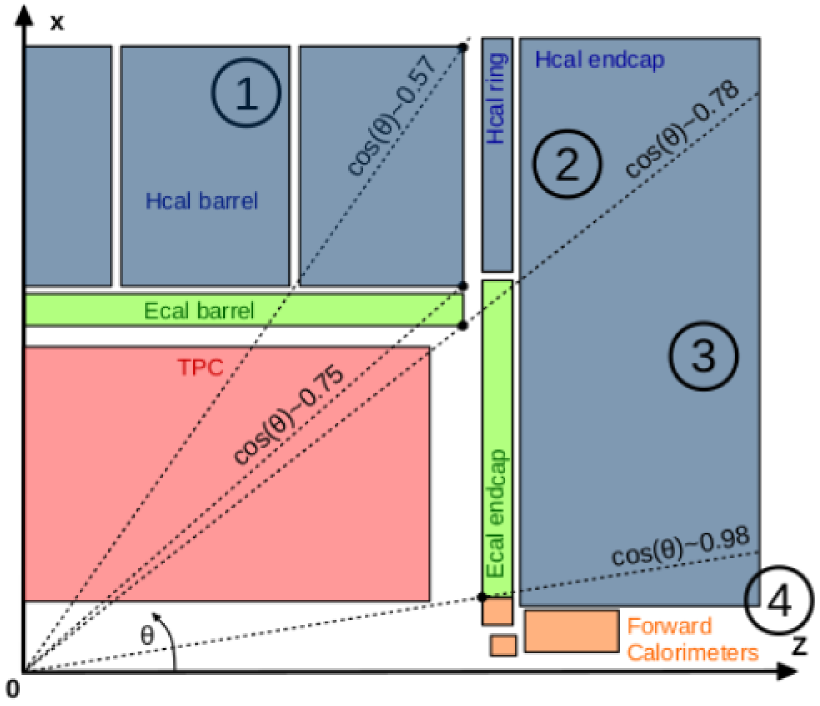
Tanguy Pasquier, Gérald Grenier

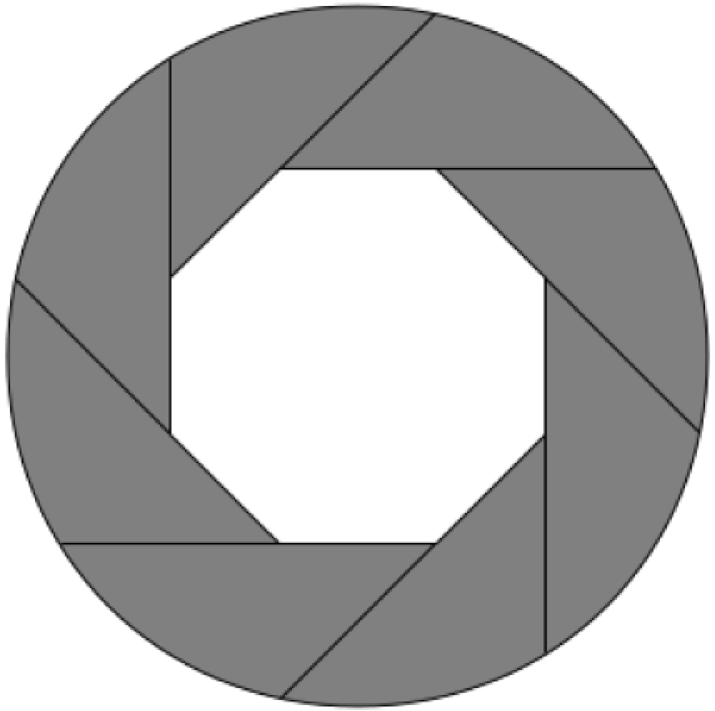
IP2I/Univ Lyon 1

18th March 2024

SDHCAL corrections

- SDHCAL needed theta and phi angle correction $\to E_{\text{rec}}$ too low
- E_{rec} is now corrected with both angles
- Videau geometry taken in consideration
- Created an official SDHCALContent repository for all SDHCAL related plugins ▸ Git repo https://github.com/SDHCAL/SDHCALContent
- Separating detector (SDHCAL, ILD option 2) from PFA (APRIL).





SDHCAL corrections

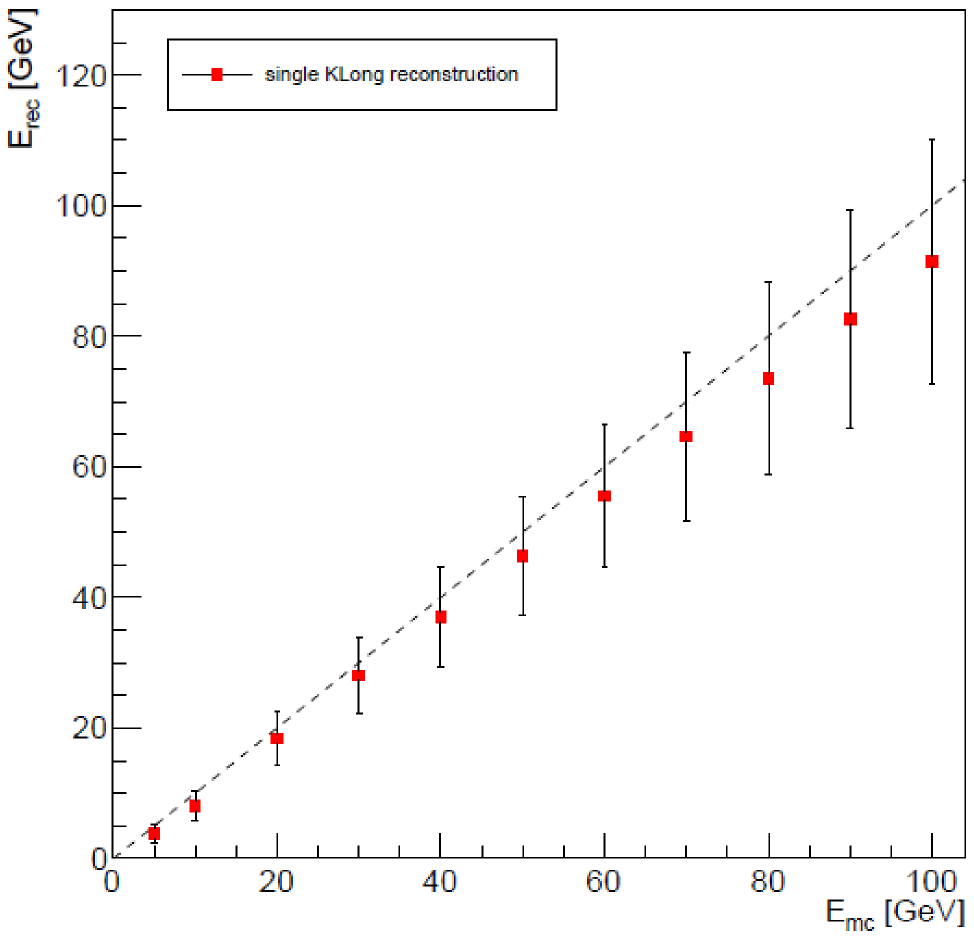


Figure: E_{rec} before correction

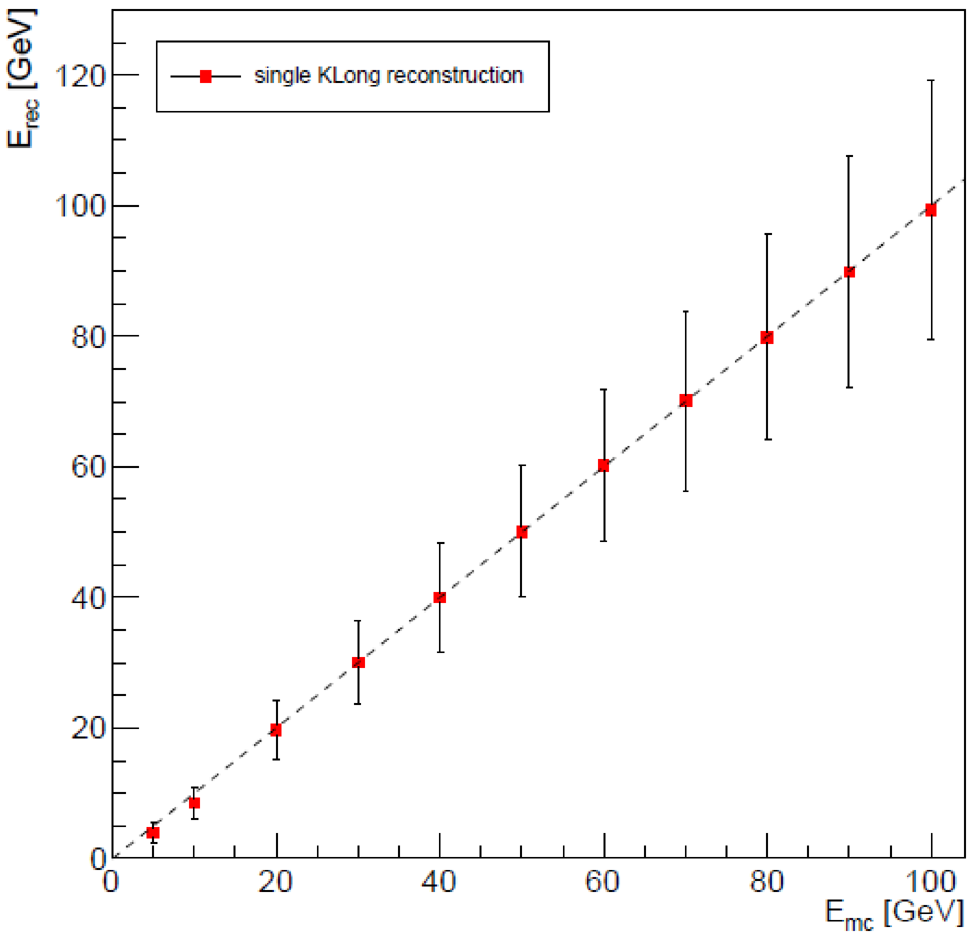


Figure: E_{rec} after correction

APRILContent

- APRILContent uses mlpack nearest neighbours algorithms : update for mlpack v3 to v4.
- Created official APRILContent repository https://github.com/SDHCAL/APRILContent
- Work started to modify names in APRILContent that duplicate names in Pandora LCContent. Will allow to link to both PFA in the same Marlin Processor/ Gaudi Algorithm.

DDMarlinPandora

- Created SDHCAL version of DDMarlinPandora https://github.com/SDHCAL/DDMarlinPandora
- Using preprocessor tag (#ifdef) to switch between Pandora and APRIL compilation.
- SDHCALContent fully compatible with both Pandora LCContent and APRILContent.
- APRILContent and LCContent not fully compilable together.

Next steps

- SDHCALContent
  - Produce single Klong simulation for ILD option 2 with more energy.
  - Fully calibrate the semi digital energy reconstruction with PFA.
- APRILContent and DDMarlinPandora
  - Clean APRILContent plugins names to avoid duplication with Pandora LCContent : goal is to reuse some Pandora plugins (e.g. AHCAL energy corrections) with APRIL.
  - Keep modifying DDMarlinPandora so that it can run both PFA with the same processor : it will need a runtime option to choose the right kind of Pandora CaloHit.
  - Longer term : a Gaudi Algorithm ?

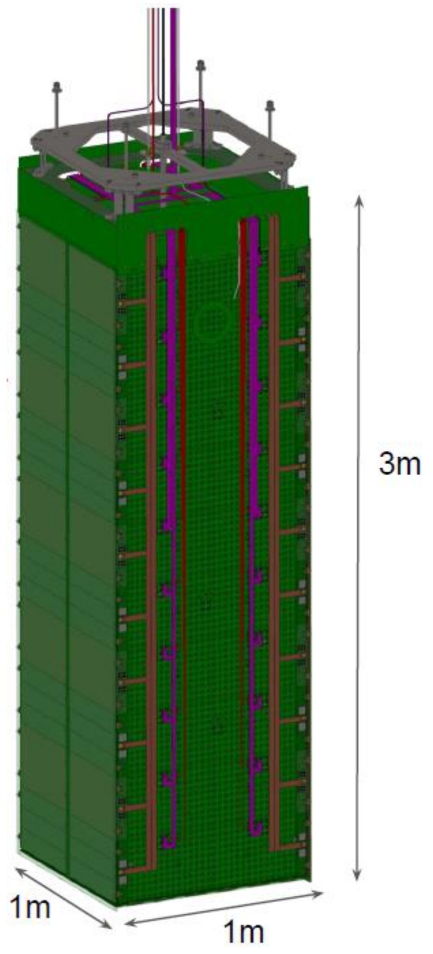
Reconstruction for the DUNE Near Detector (ND)

John Back & John Marshall (Warwick),
Steve Dennis, 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

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

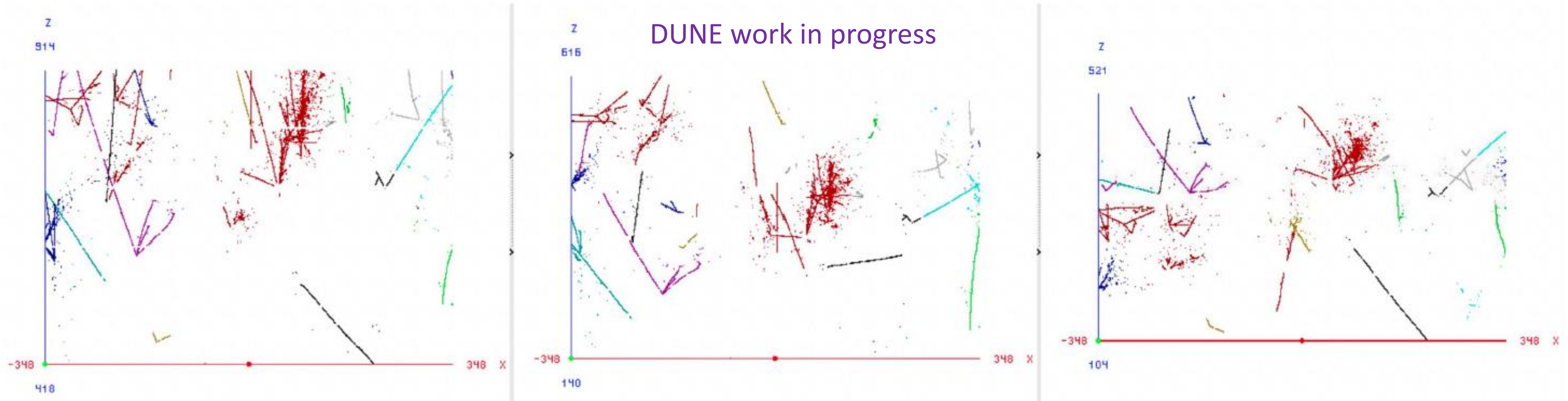
- Expect **~50 ν** interactions per sec for 7x5 ND LAr modules
  - LBNF 120 GeV, 1.2 MW proton beam on graphite target
  - Secondary $\pi \rightarrow \mu\nu$: 7.5×10^{13} protons per beam "spill" (1.2 sec)



Multi-neutrino interactions reco: Slicing

Break up complex spills into independent ν interactions: **1 slice ≅ 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**

Recently updated Pandora's LArContent package:

- Moved slicing object creation from SlicingAlgorithm to LArSlice class
- Allows development of different ND slicing methods (instead of just the current default one)

Pandora 3D ND reco developments (Leigh Whitehead)

- Initial energy hits clustering performed using 3D coords

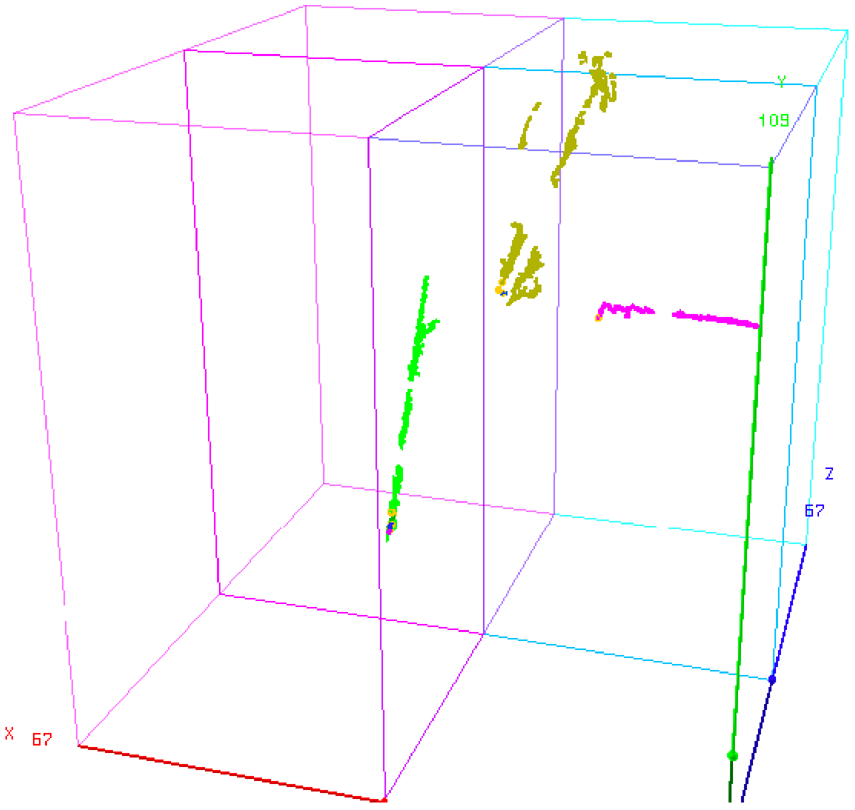
- Cluster merging and refinement
  - Apply existing 2D algorithms to the 3D hit clusters
  - Currently does not use **y** information (to be addressed)

- Clusters projected into 2D
  - Use neutrino reco algorithms based on 2D projections (e.g. MicroBooNE)

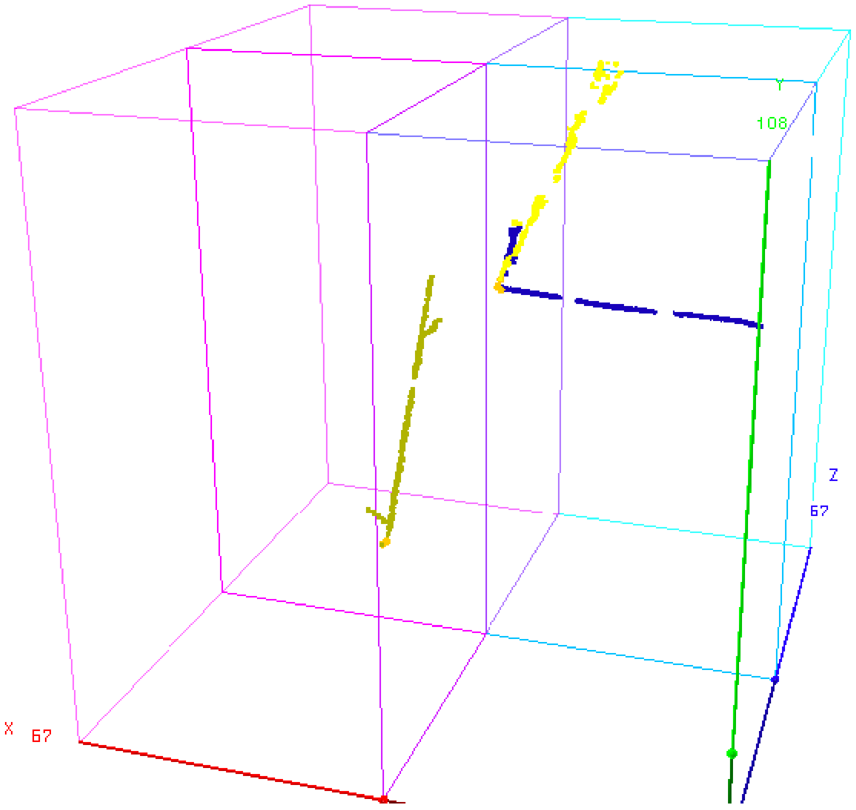
- Create Particle Flow Objects
  - Match 2D hits back to 3D
  - Build final tracks and showers

Pandora reco of 2x2 simulation neutrinos

MicroBooNE Reconstruction



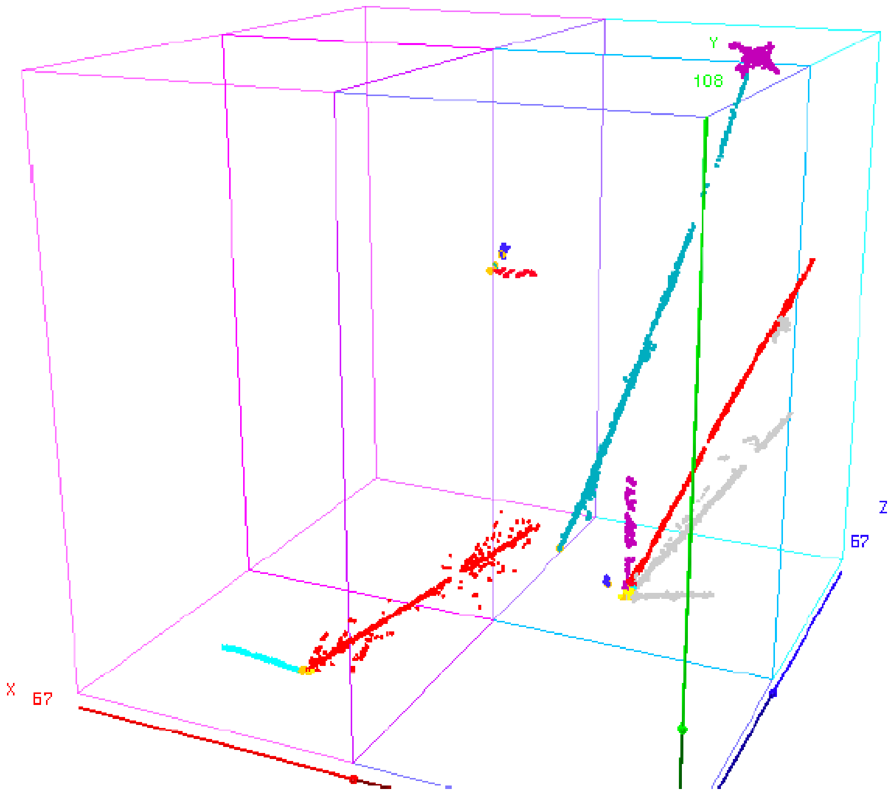
“3D” ND Reconstruction



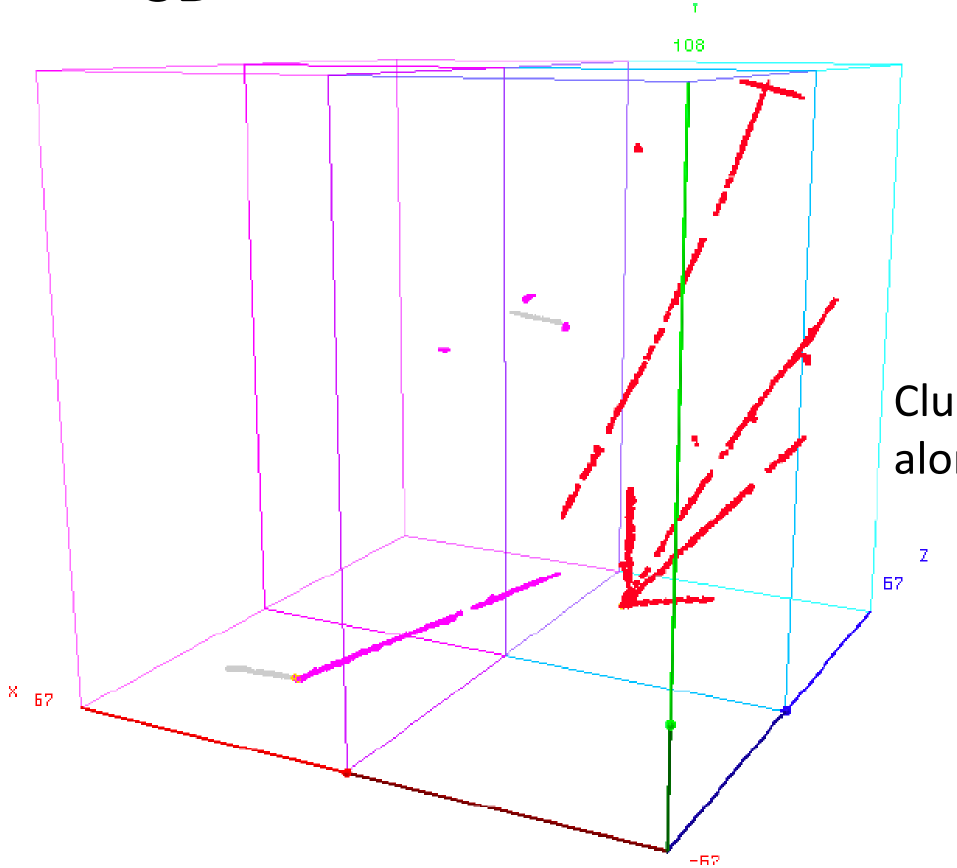
DUNE work in progress

Pandora reco of 2x2 simulation neutrinos

MicroBooNE Reconstruction



"3D" ND Reconstruction



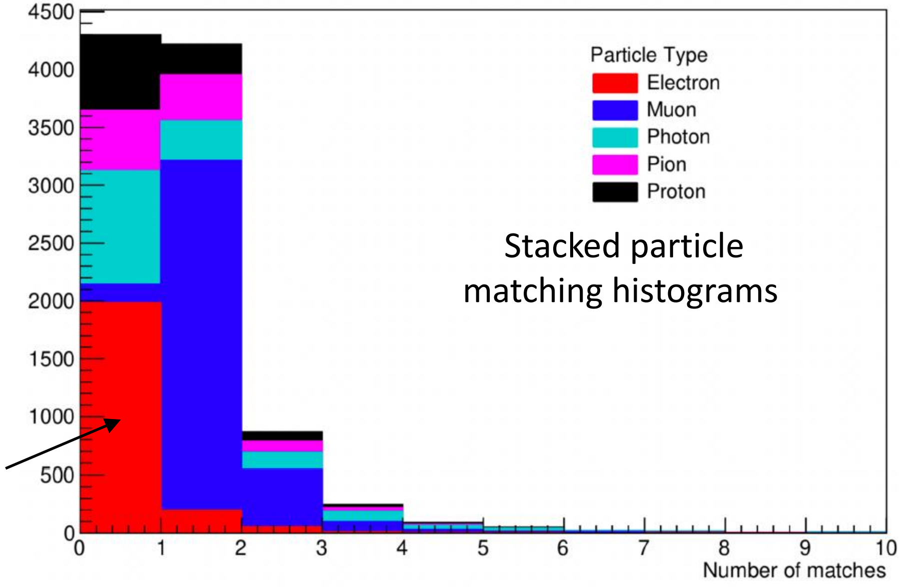
Clusters overlap along y projection

DUNE work in progress

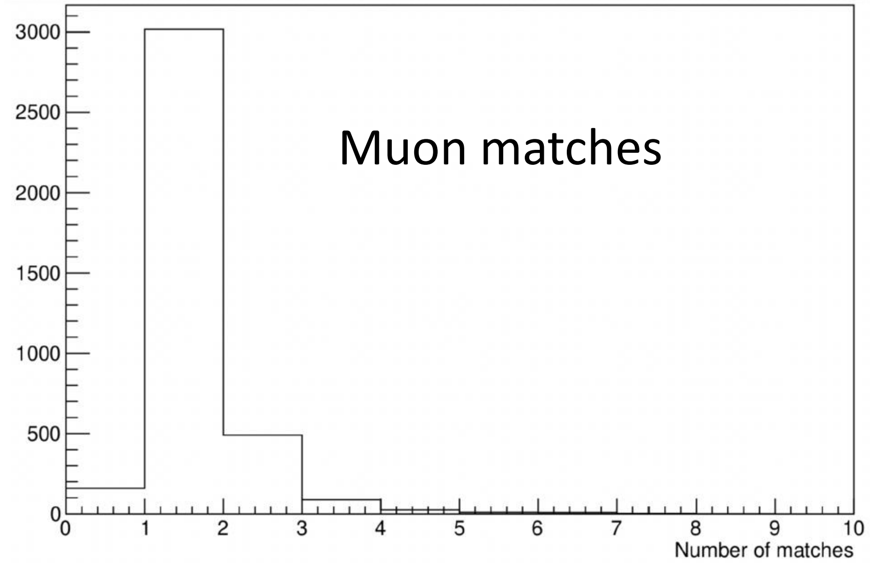
Pandora 2x2 reco event validation: ν_μ events

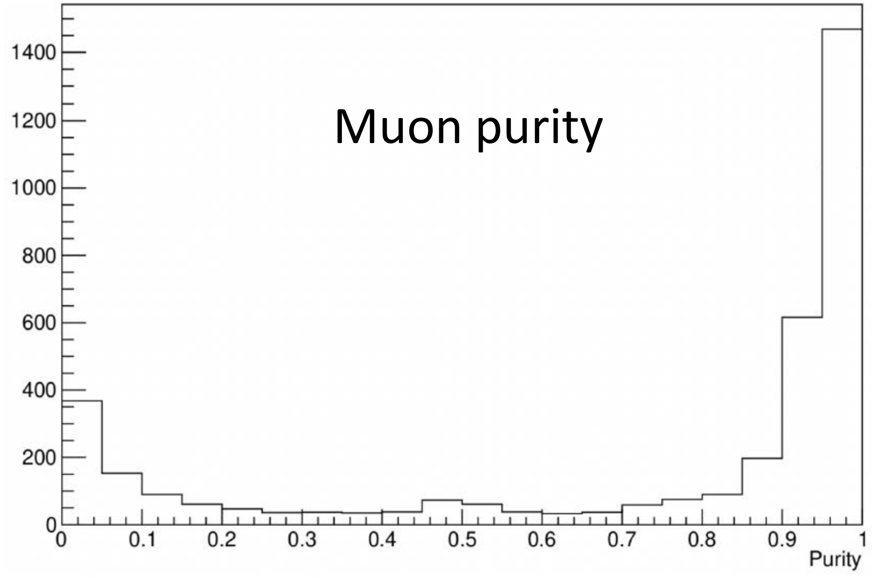
Associating MC truth info with reco objects

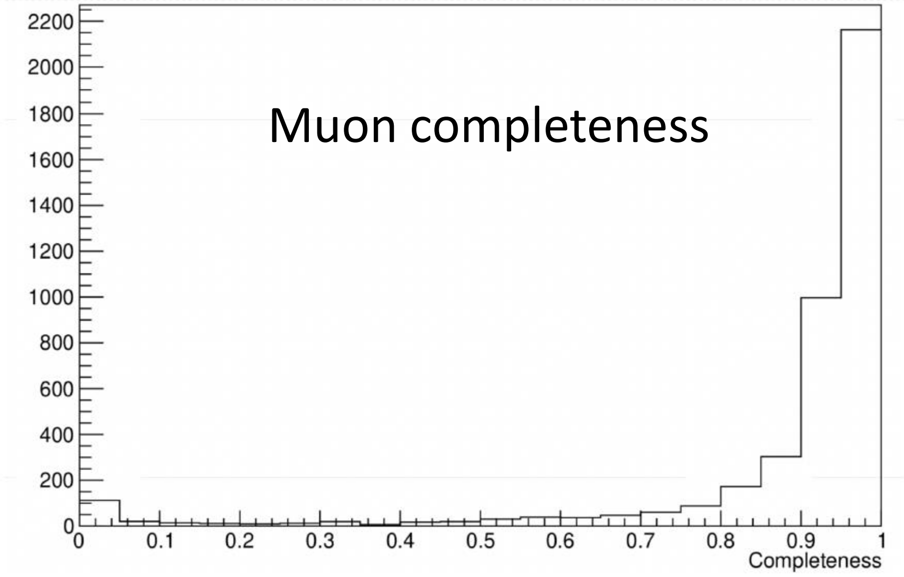
DUNE work in progress



low E
δ rays



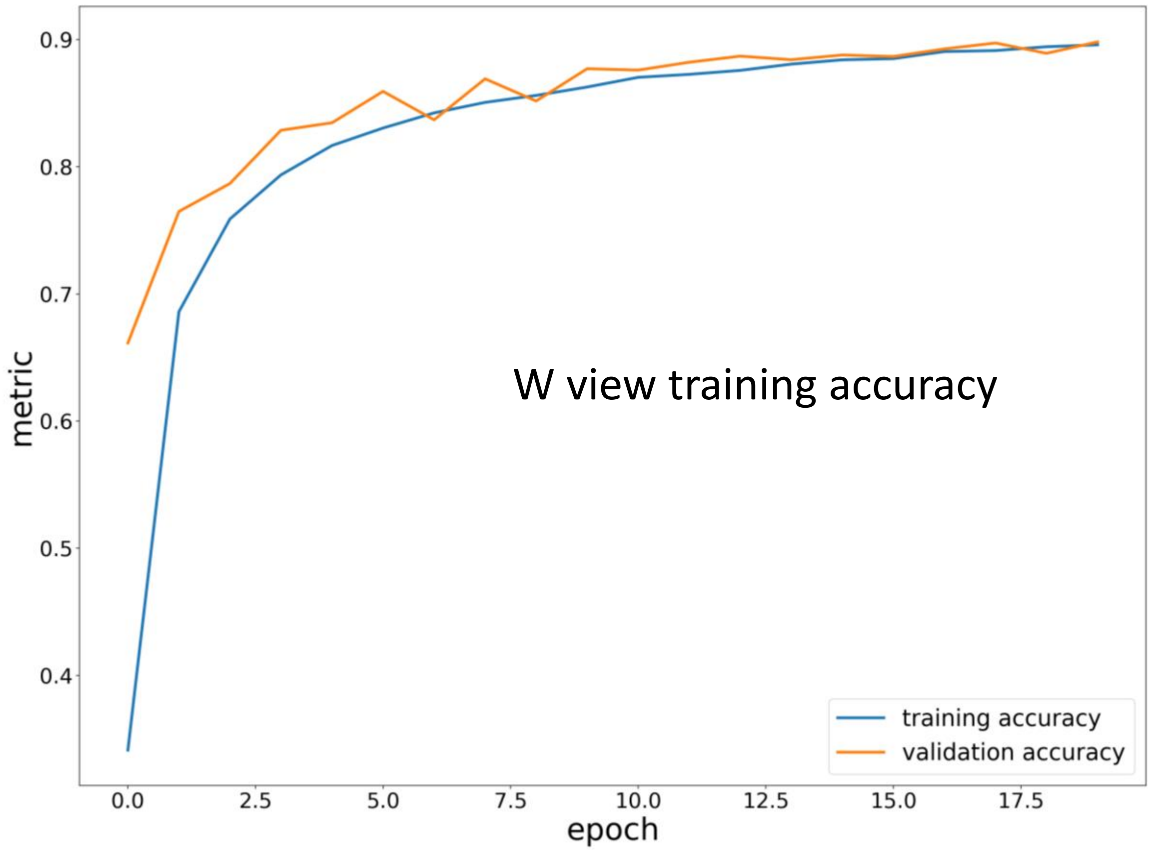


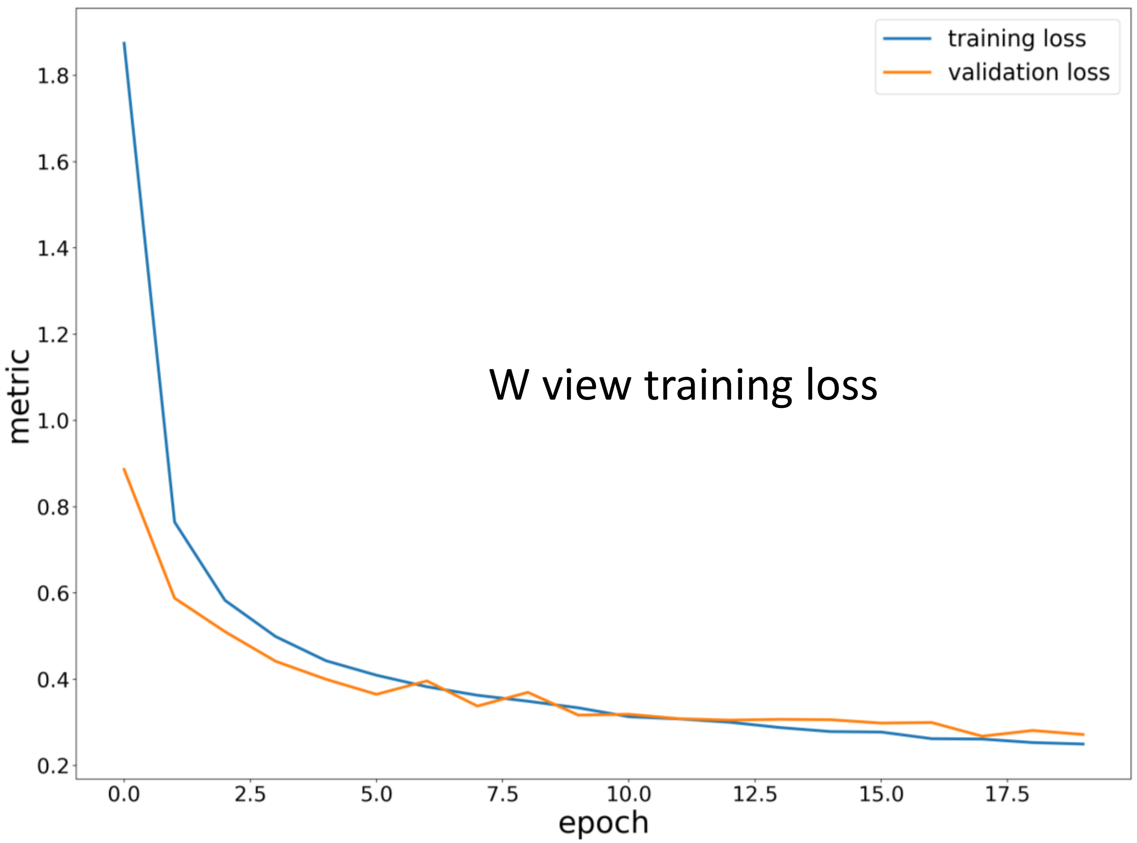


DUNE Near Detector Deep Learning vertexing: pytorch

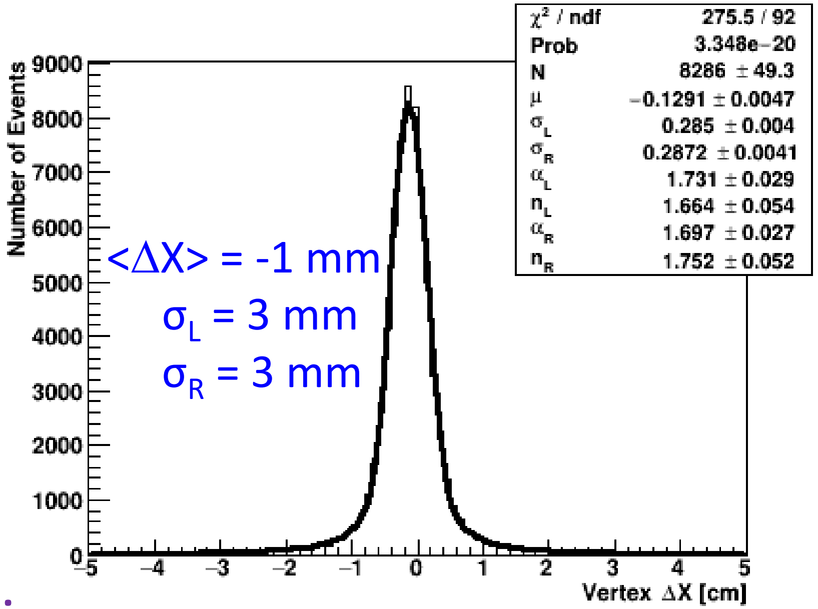
Using "**U, V, W**" 2D projection views for **50k** ν_μ training LArTPC simulation images

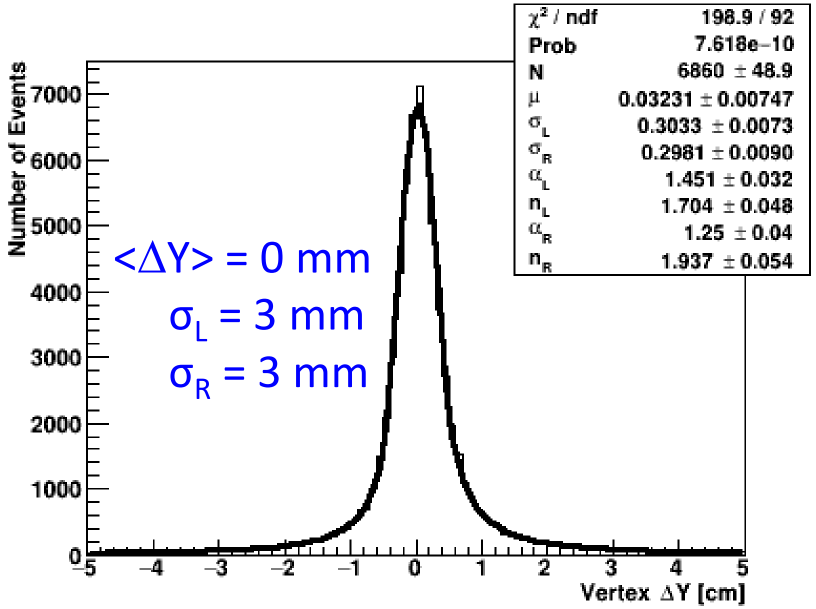
Two training passes: **whole** neutrino images, then x0.5 **zoomed-in** interaction vertex region images





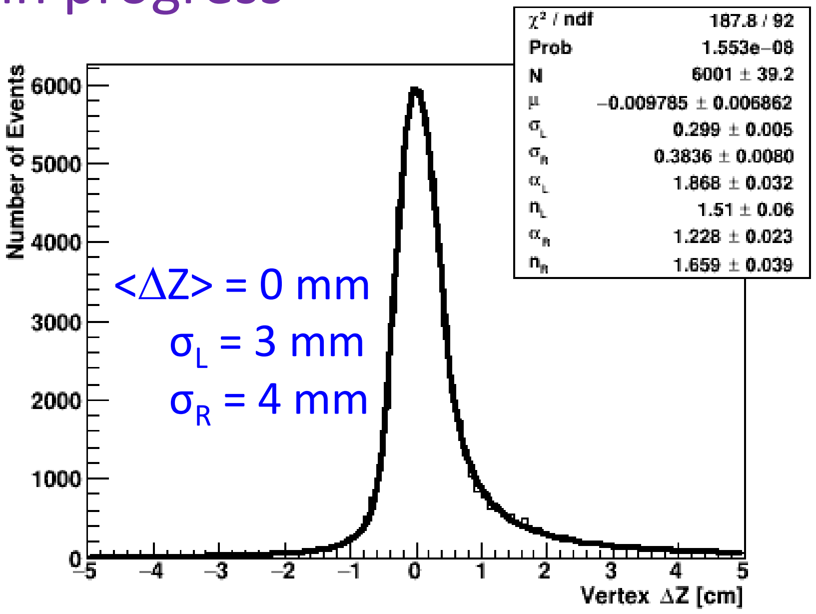
Vertex reco – true MC residuals (cm)

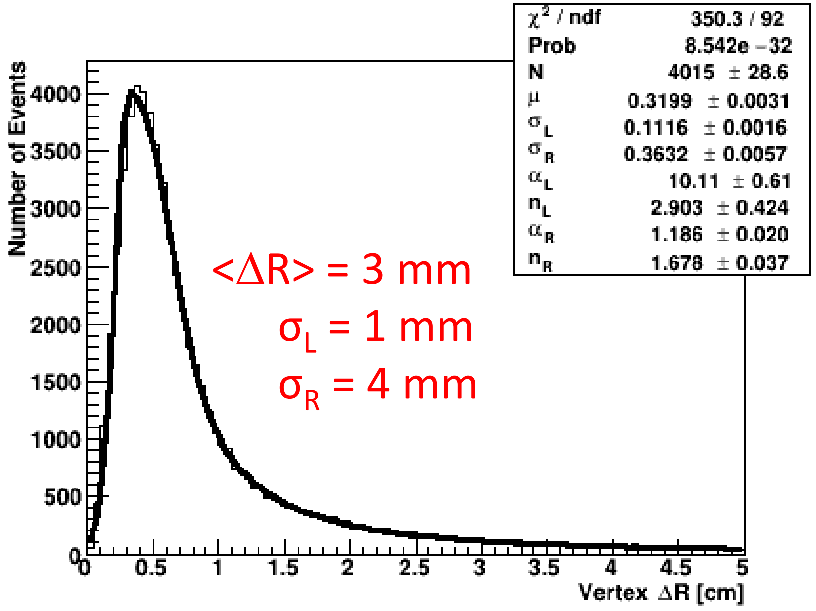




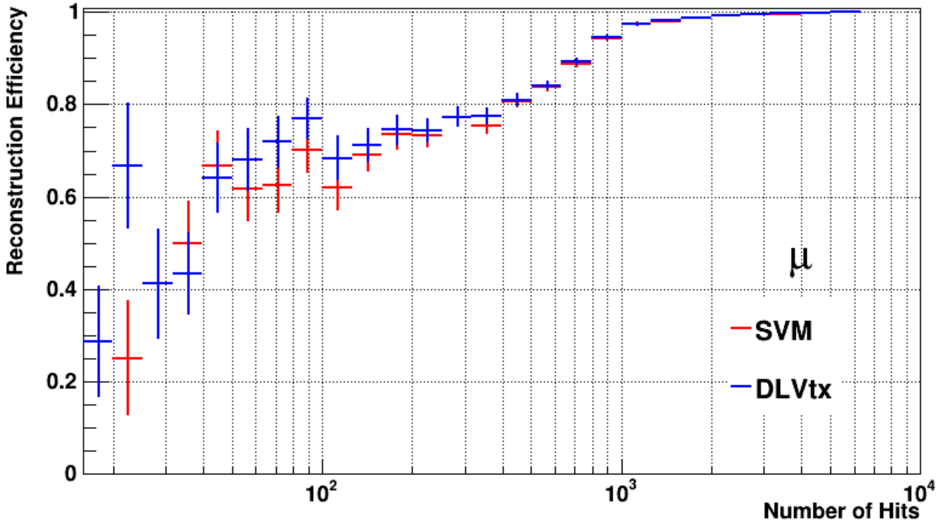
Gaussian fits with μ, σ_L, σ_R & power law tails

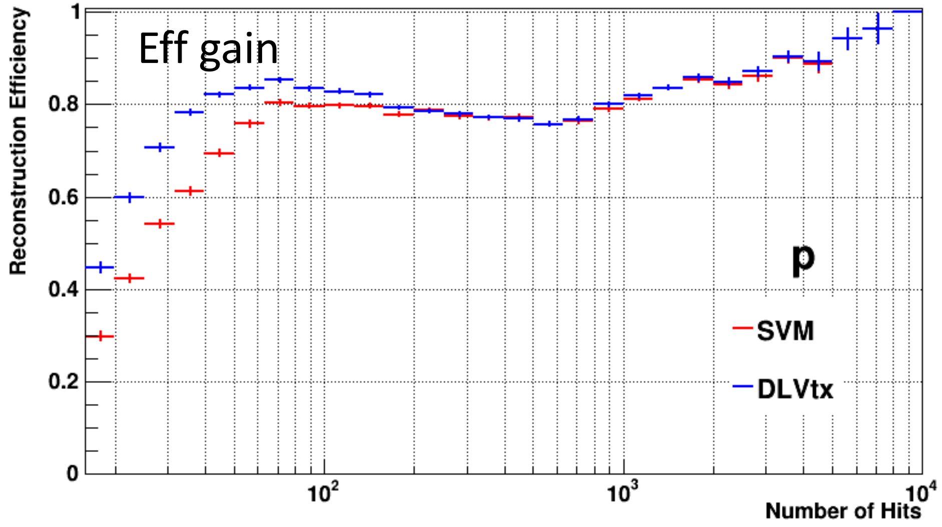
DUNE work in progress



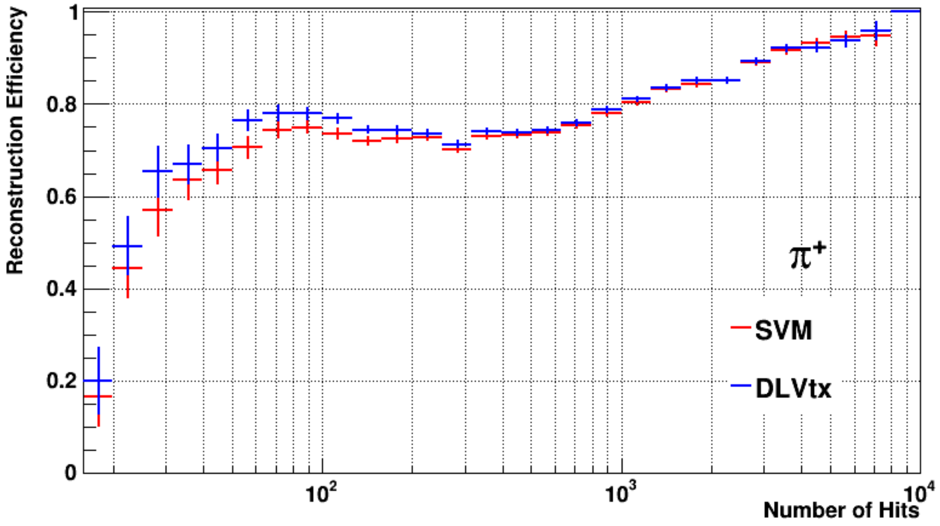


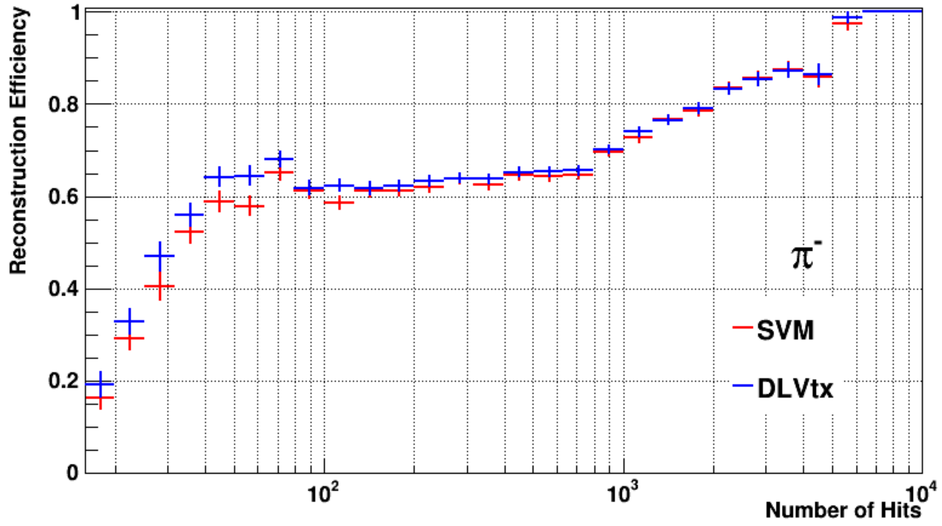
Pandora hit reconstruction efficiencies





DUNE work in progress





SVM: MicroBooNE support vector machine, DLVtx: pytorch ND deep learning vertexing

Summary

- Dual readout calorimeters
  - NN (TensorFlow) jet reconstruction & PID
  - Rescoping: Use traditional (physics based) reco methods 1st,
    then use NN to improve performance

- APRIL
  - Developing reconstruction algorithms for SDHCAL (ILD option 2)
  - Energy calibration studies
  - SDHCALContent, APRILContent & SDHCAL/DDMarlinPandora git repositories

- DUNE
  - Near Detector Pandora reconstruction: LArRecoND package
  - 2x2 LArTPC prototype
  - Deep Learning vertexing trained for ND
  - 3D reco developments to be added soon: PR10