Integrating Track/Shower U-Net into Pandora

Andy Chappell

27/04/2020 FD Sim/Reco Meeting

DEEP UNDERGROUND NEUTRINO EXPERIMENT

WARWICK

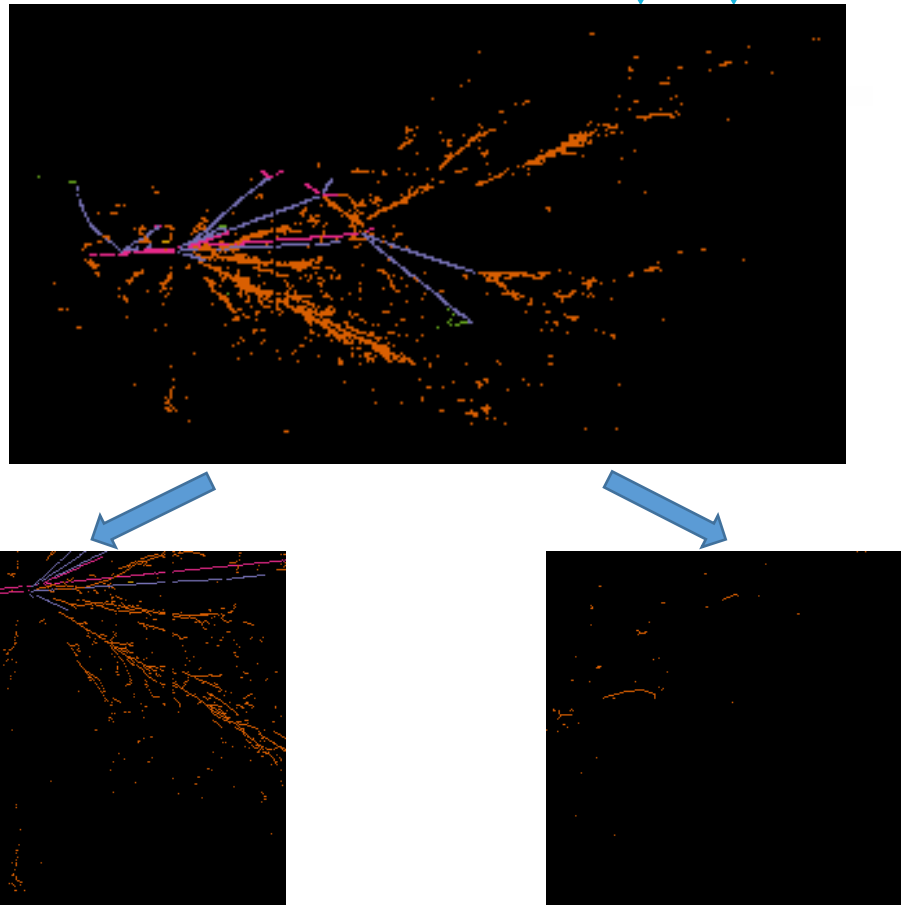
Recent developments

WARWICK

- Track/shower discrimination
  - Investigating sparse network
  - Modification to classification
- Application of the network in Pandora
  - Investigating where network output might benefit downstream pattern recognition

Why sparsification?

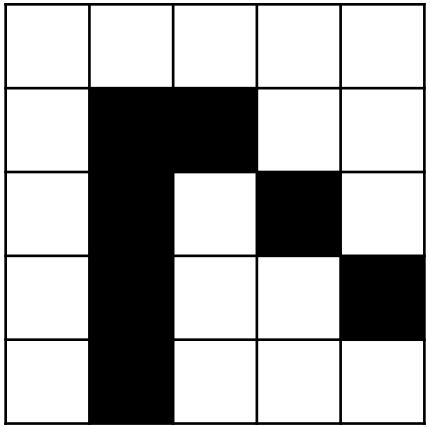
- Our events are mostly empty space
- Even after tiling an image and throwing away empty tiles, in a 20K event training sample, 99.6% of all pixels don't have a hit
- Conventional CNN must still process these pixels



Sparse format

WARWICK

- Existing sparse network implementations exist
  - https://github.com/facebookresearch/SparseConvNet
  - The sparse convolutions only consider active ‘pixels’ in the input
  - Needs PyTorch v1.3+ (see April 21 LArSoft Coordination Meeting)



1 PyTorch tensor

Index (y, x, b)	Value
(1, 1, 0)	1
(1, 2, 0)	1
(2, 1, 0)	1
(2, 3, 0)	1
…	…

2 PyTorch tensors

Current status

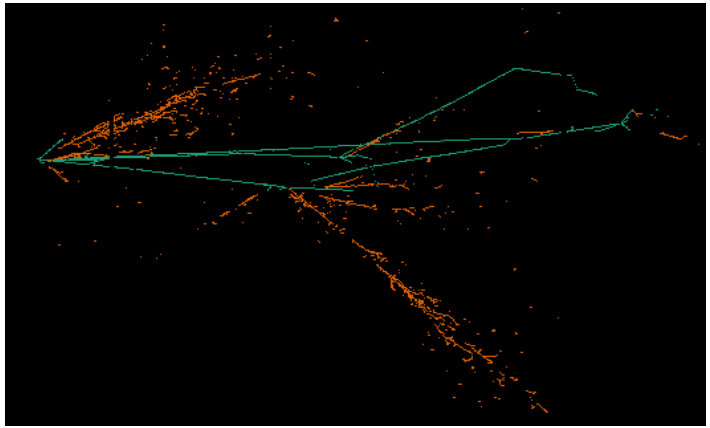
WARWICK

- Promising…
  - In PyTorch on GPU it works and slightly out-performs the dense network
  - Runs about 6x faster on GPU
- But problematic
  - We need this to run on CPU in C++ environment
  - PyTorch models can be converted to TorchScript deployment runtime
  - I can’t get the network through the PyTorch tracer – it doesn’t like the sparse input format

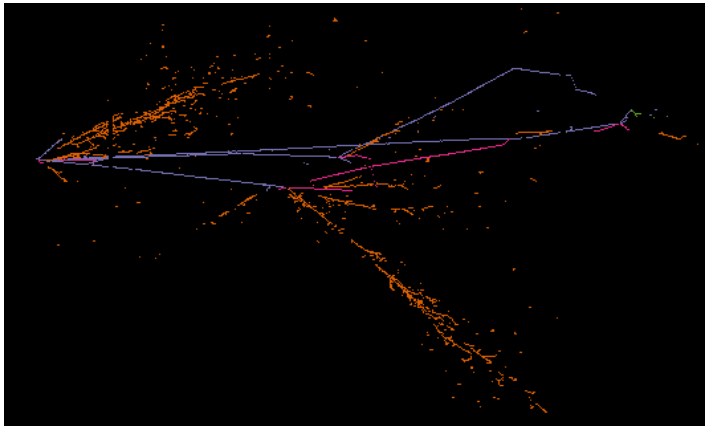
Alternative classification

WARWICK

- Looked at broadening the classification of hits
  - Previously a hit is either track-like or shower-like
  - Expanding definition to MIP, HIP, Michel or shower (see backup)



Binary track/shower classification



Multi-class classification

Confusion matrices

Classification given truth

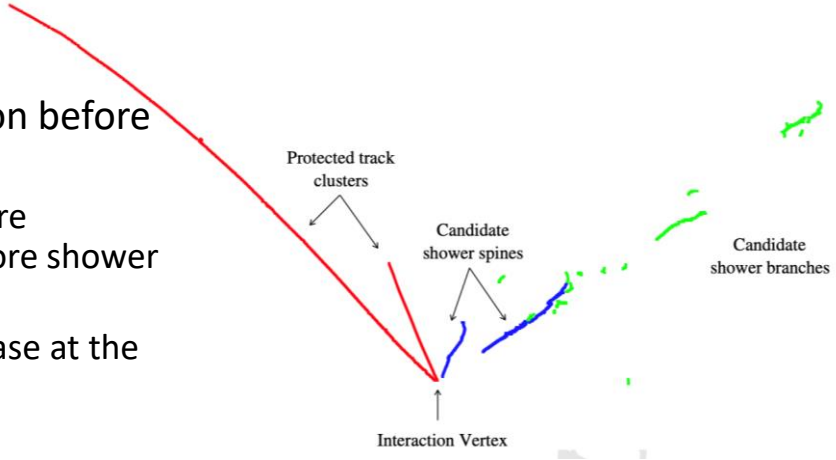
C/T	N	S	M	H	D
N	-	~0	~0	~0	~0
S	-	0.839	0.030	0.075	0.185
M	-	0.003	0.708	0.157	0.011
H	-	0.052	0.239	0.721	0.094
D	-	0.107	0.023	0.047	0.710
	-	1	1	1	1

Truth given classification

C/T	N	S	M	H	D	
N	-	0.368	0.632	0.000	0.000	1
S	-	0.888	0.084	0.023	0.005	1
M	-	0.002	0.975	0.023	0.000	1
H	-	0.056	0.707	0.232	0.003	1
D	-	0.537	0.305	0.068	0.090	1

- C = Classification, T = Truth
- N = Null, S = Shower, M = MIP, H = HIP, D = Michel

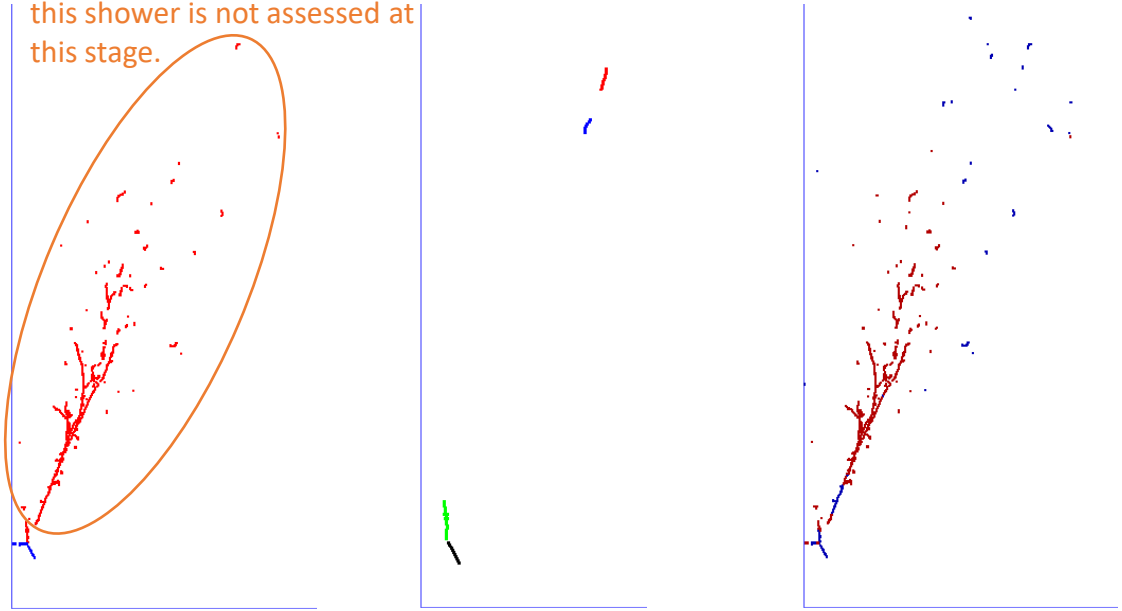
Integrating with Pandora



- Assess the status of the PFO characterisation before shower growing
  - These slides look at status immediately before LArThreeDTrackFragments, the last step before shower algorithms
  - Looking at post LArThreeDTrackFragments case at the moment
- Definition of “correct”
  - At this stage of reco, Pandora is identifying coherent “track-like” trajectories
  - Shower spines, for example, fit into this category
  - But track-like trajectories are protected from shower growing, and for this purpose I’m going to class that as being incorrect
  - A PFO or cluster will be classed as correct if it corresponds to a true track (segment) that will be protected from shower growing, or if it corresponds to a true shower (component) that will be made available to shower growing

Example event 1 - PFOs

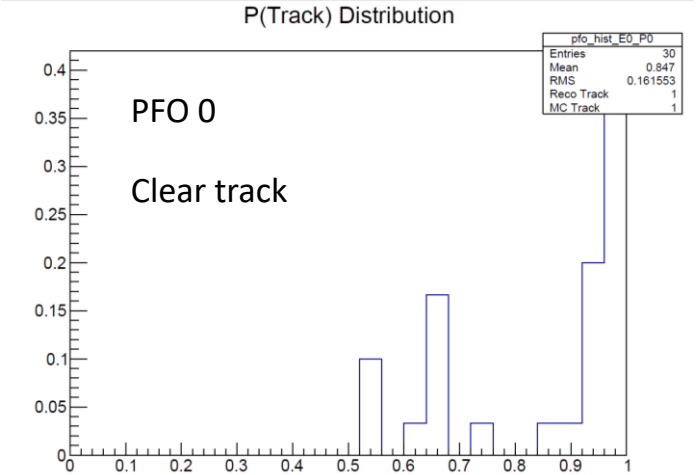
Only attempted to identify track-like PFOs at this stage, so this shower is not assessed at this stage.

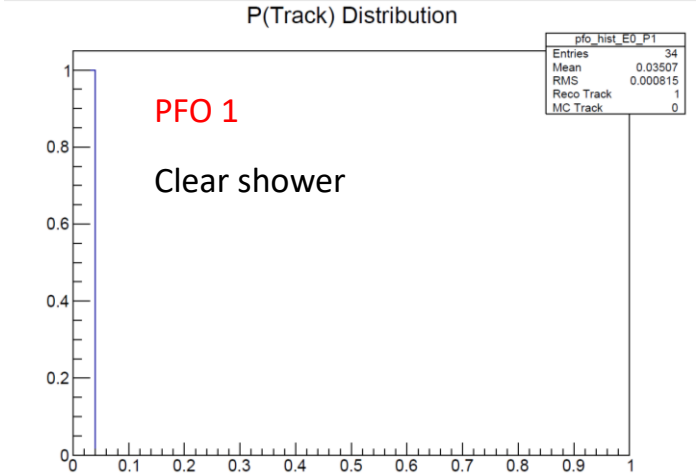


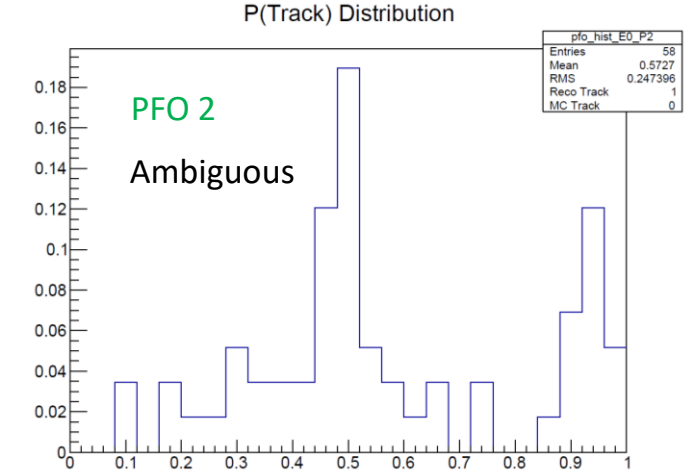
MC PFOs Network

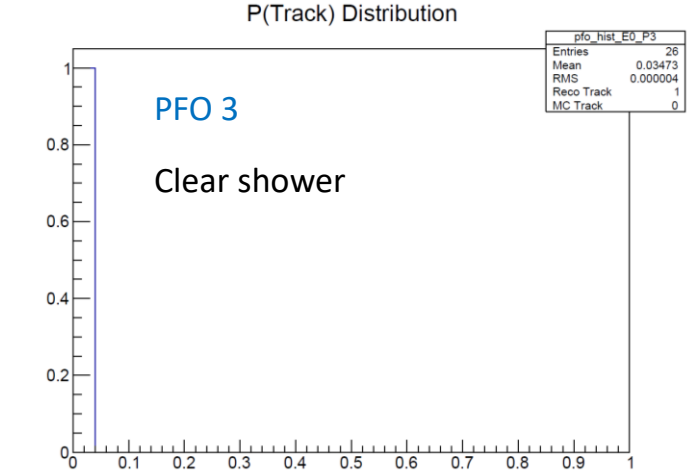
- PFO 0
- PFO 1
- PFO 2
- PFO 3
- 1, 2 and 3 misclassified

Note: Network doesn't care about hit reconstructability at inference time, so network display has extra hits


P(Track) Distribution
pfo_hist_E0_P0
Entries 30
Mean 0.847
RMS 0.161553
Reco Track 1
MC Track 1
PFO 0
Clear track

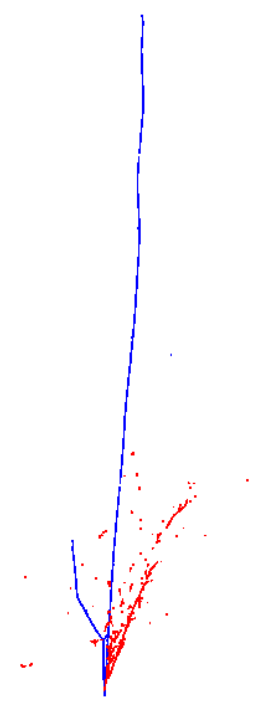

P(Track) Distribution
pfo_hist_E0_P1
Entries 34
Mean 0.03507
RMS 0.000815
Reco Track 1
MC Track 0
PFO 1
Clear shower


P(Track) Distribution
pfo_hist_E0_P2
Entries 58
Mean 0.5727
RMS 0.247396
Reco Track 1
MC Track 0
PFO 2
Ambiguous

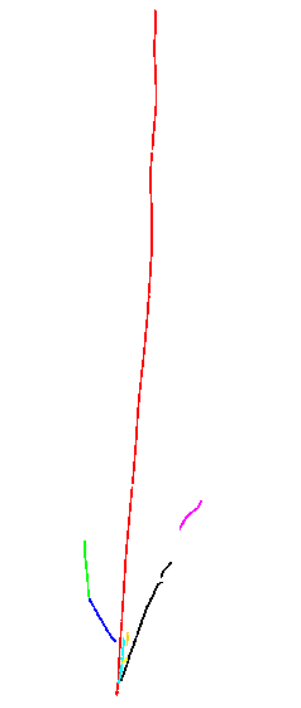

P(Track) Distribution
pfo_hist_E0_P3
Entries 26
Mean 0.03473
RMS 0.000004
Reco Track 1
MC Track 0
PFO 3
Clear shower

CK

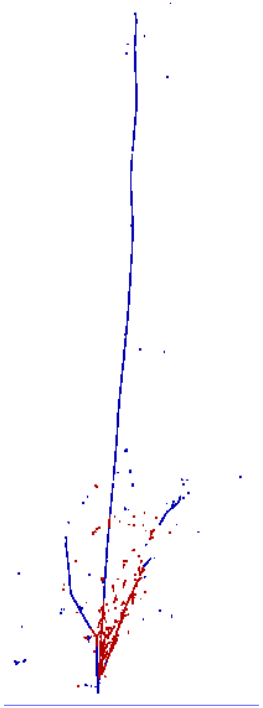
Example event 2 - PFOs



MC

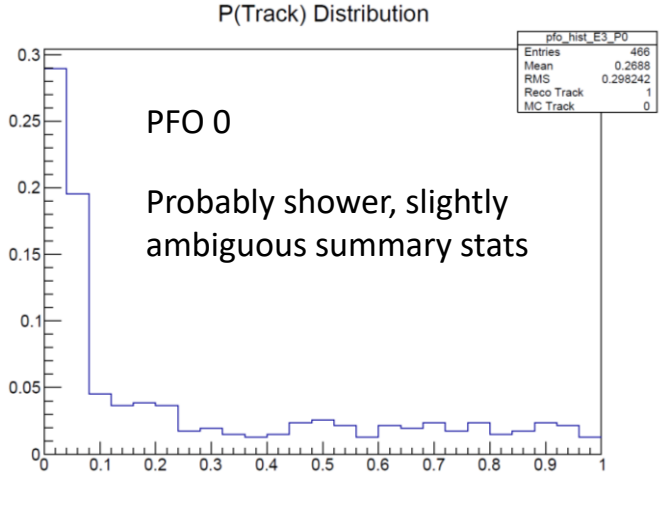


PFOs

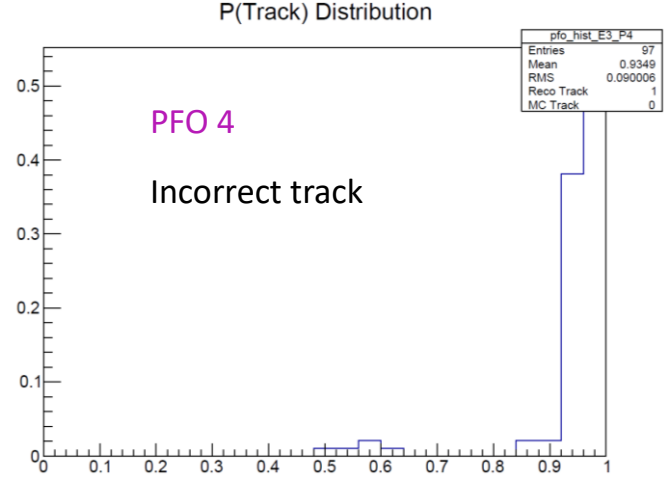


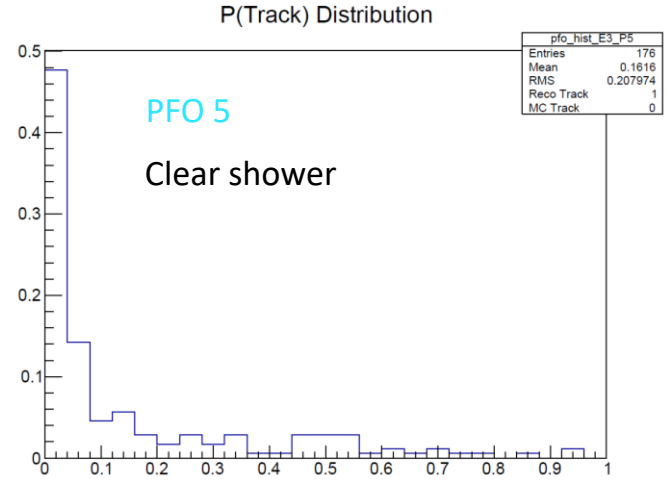
Network

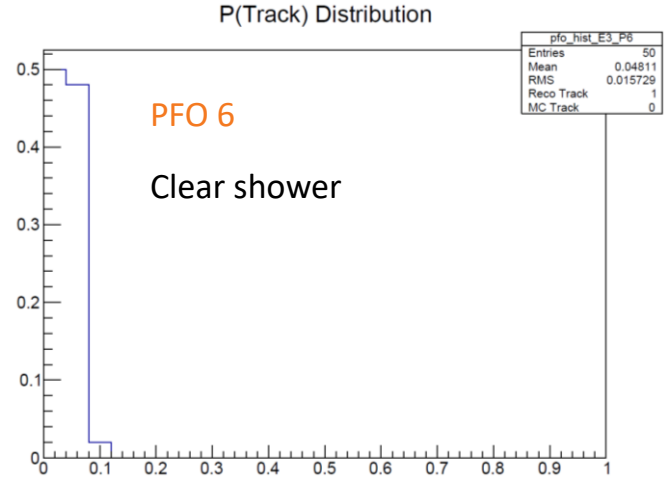
- PFO 0, PFO 1, PFO 2, PFO 3, PFO 4, PFO 5, PFO 6, PFO 7
- All true tracks identified Network agrees
- Remaining track-like trajectories identify overall skeleton of event
- Network has potential to identify shower spines to be made available to shower growing algorithms



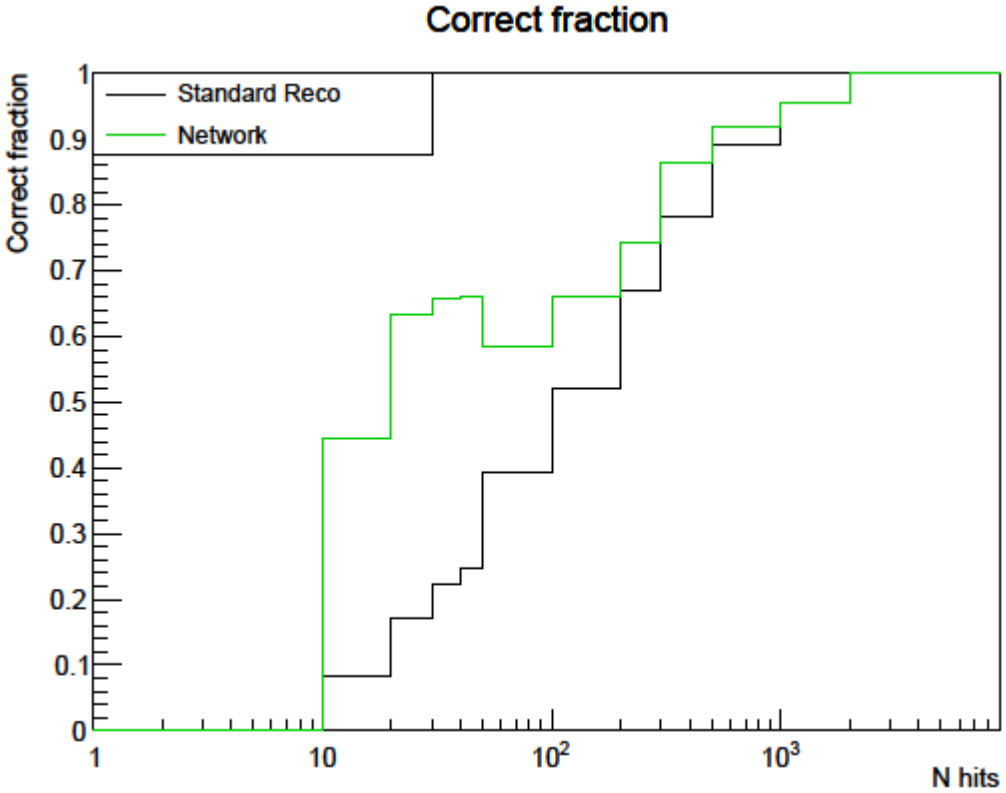
PFO 7 also clear shower





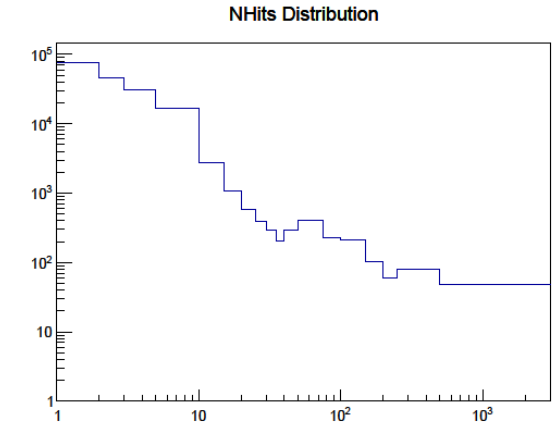


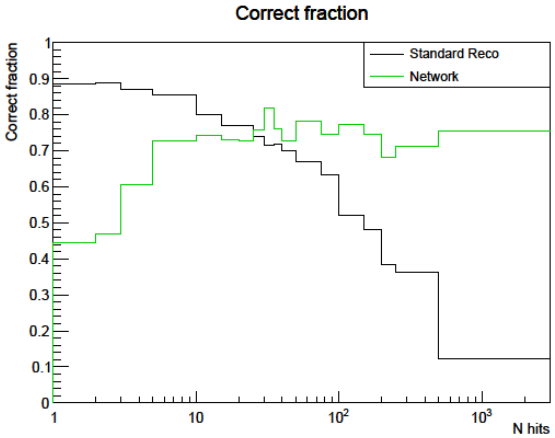
Performance by hit count



- Network able to correctly tag many true showers with coherent trajectories and make them available to shower growing
- Standard Reco: 47.6% correct
  +Network: 70.7% correct

Available clusters





- Clusters not allocated to a PFO (green hits) at this stage go forward to shower growing
- Are there track-like clusters that should be protected from shower-growing?
  - Yes, but not that many
- Network only useful here for clusters with more than 25 hits
  - There aren't many of these
- Standard Reco: 87.6% correct
  +Network: 87.7% correct

Summary and future work

WARWICK

- Network looks promising with respect to PFOs
  - Many PFOs corresponding to true showers were previously unavailable to Pandora's shower growing algorithms. Network can tag many of these cases, and rarely tags PFOs corresponding to true tracks as shower-like
- Most "available" clusters correspond to true showers
  - Network offers little to such clusters at this point in the algorithm chain
- Currently based on simple summary statistics from network
  - Mean/RMS from PFO/cluster probability distributions
  - More sophisticated methods may yield improved performance
- Look at influence on final reconstruction

Backup

WARWICK

MC truth

WARWICK

- MIP – Muons and pions
- HIP – Protons, kaons and nuclei
- Michel
- Showers (and EM activity) – e/γ