Pandora Deep Learning Workflow

Andy Chappell & Ryan Cross for the Pandora team

24/08/2021 DUNE UK Meeting

DEEP UNDERGROUND NEUTRINO EXPERIMENT

WARWICK

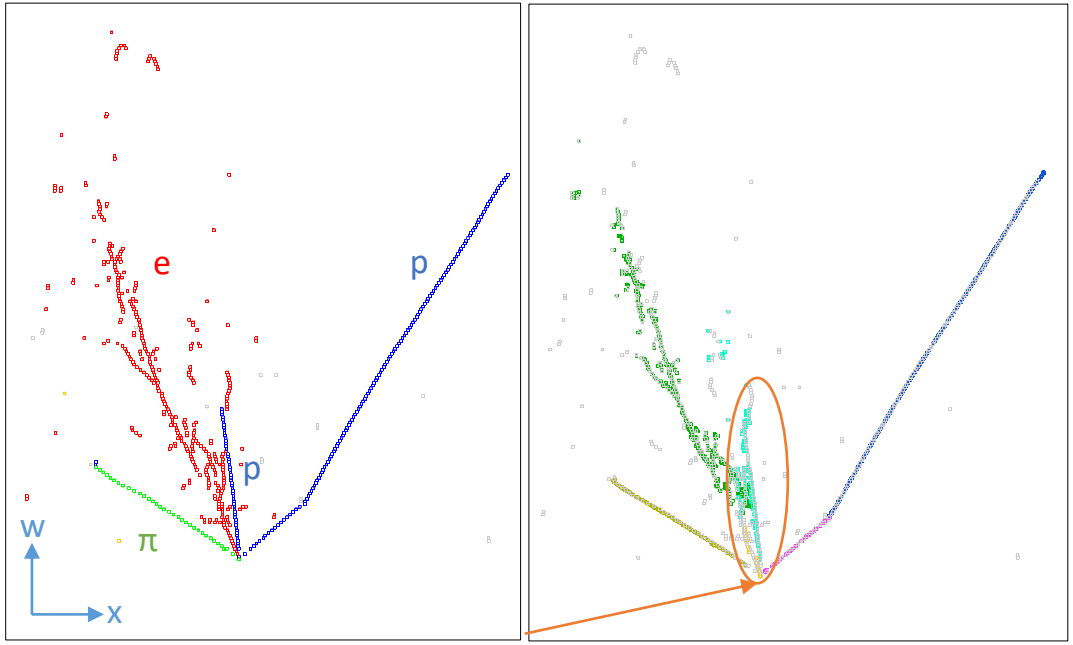
Split reconstruction streams

WARWICK

- Event topologies at DUNEFD are highly variable
  - Simple CCQE topologies are relatively straightforward to reconstruct
  - Complex RES and DIS topologies present significant reconstruction challenges
    - Dense track environments
    - Overlapping track and shower topologies
- Processing all hits simultaneously can lead to both fragmentation and merging of particles
- Aim to reduce these problems by separating track-like and shower-like topologies into separate streams
  - Track-centric algorithms can be applied in a stream containing hits from track-like particles
  - Shower-centric algorithms can be applied in a stream containing hits from shower-like particles

Losing tracks to showers

- CC ν_e event
- Pion and one proton hierarchy reconstructed correctly
- Electron and proton mis-reconstructed
  - Initial stub of electron shower is reconstructed as a particle
  - Bulk of downstream electron shower reconstructed as a separate child particle
  - Proton is merged with part of the electron shower and also added as a child particle of the electron



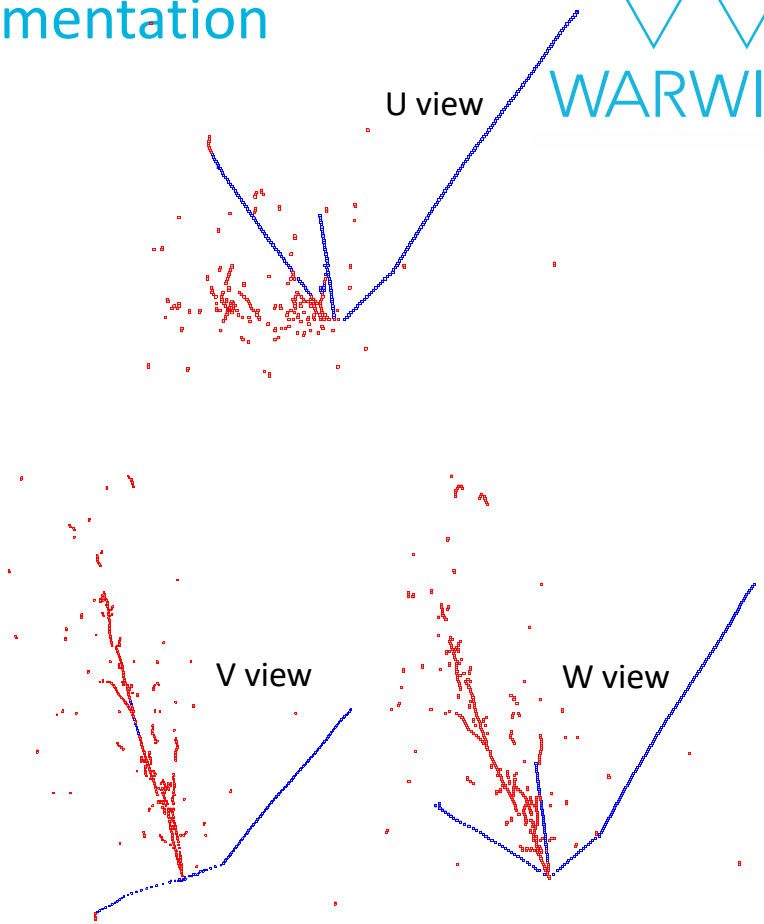
MC

Proton merged into shower

Standard Reco

Hit tagging with semantic segmentation

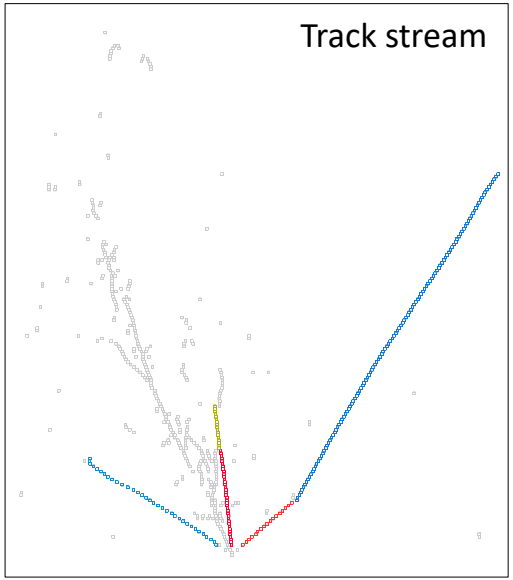
- Neural network tags hits as either track-like (blue) or shower-like (red)
- U and W view hit tagging largely correct
- Overlapping electron shower and proton in V causes mis-classification of proton in this view

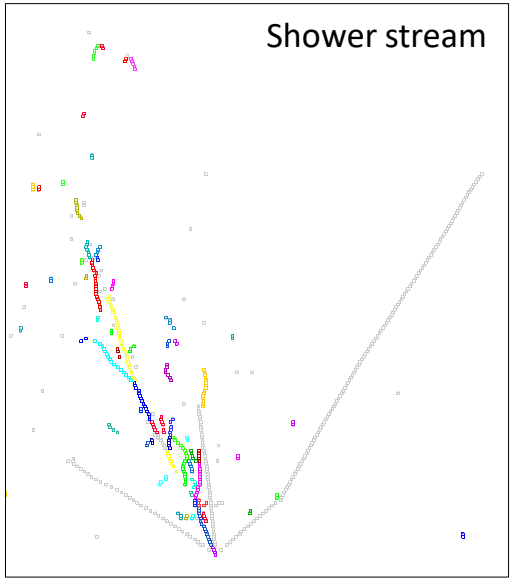


Streamed clustering

WARWICK

- Combining neural network output with multi-algorithm approach allows initial clusters to be split into separate “streams”
- Subsequent reconstruction algorithms operate within the streams
- Gray hits are excluded from the respective stream

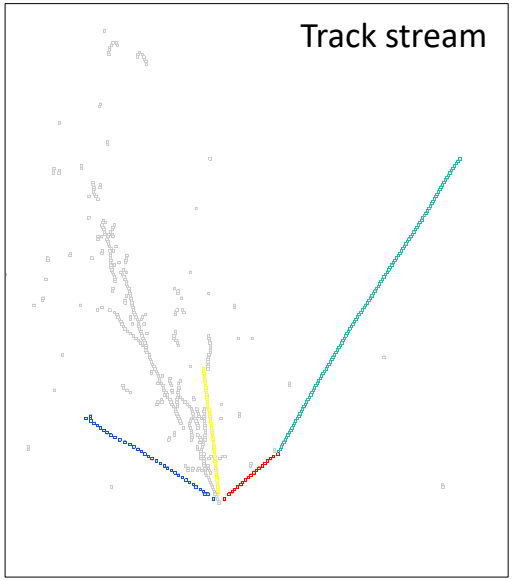


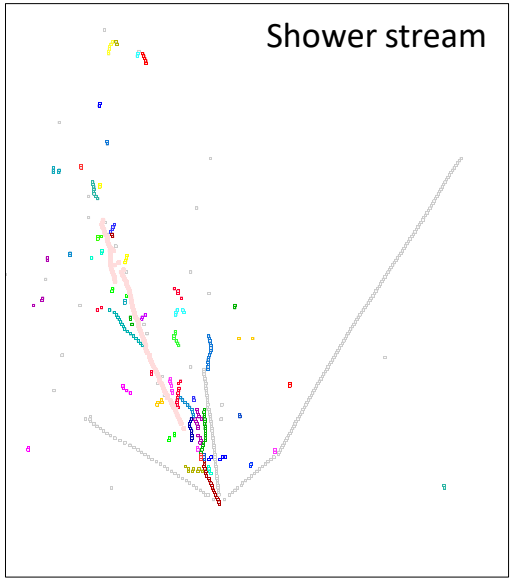


Streamed clustering

WARWICK

- Clean environment allows existing track-centric Pandora algorithms to operate effectively
- Still a high-level of fragmentation in the shower stream before mop-up algorithms, but can now introduce shower-centric algorithms without compromising track reconstruction (see Ryan's work)
- Gray hits are excluded from the respective stream

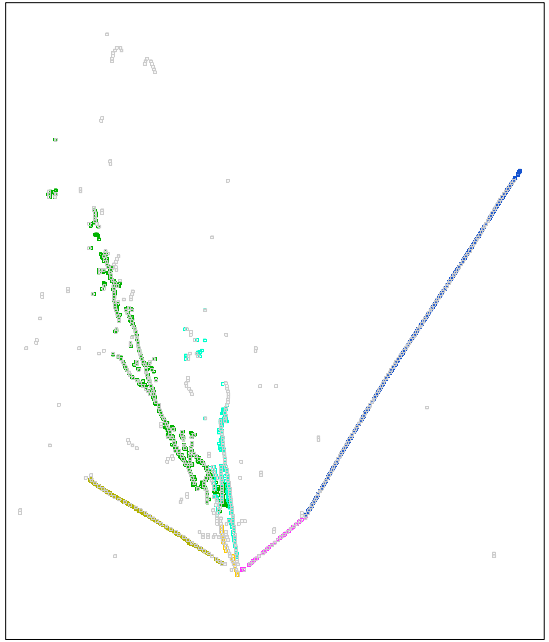




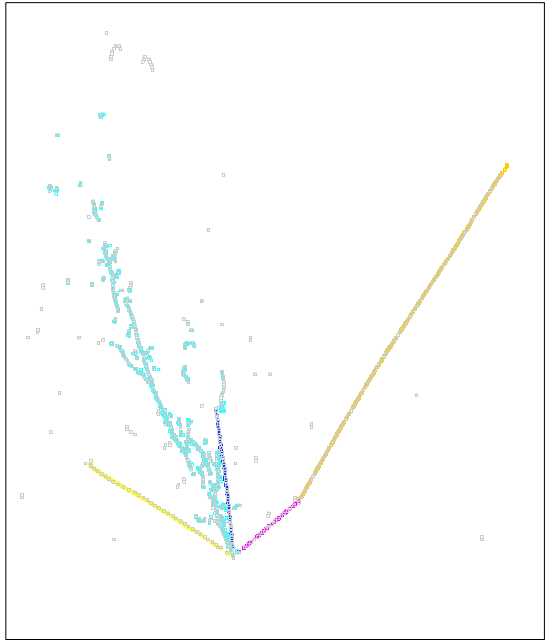
Final reconstruction

WARWICK

- Separation of track-like and shower-like hits precludes the possibility of merging the proton and electron
- Electron correctly reconstructed back to the vertex
- Also permits hits beyond the proton to be correctly associated with the electron



Standard Reco



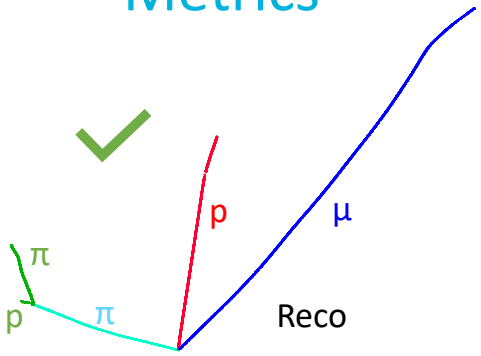
Streamed Reco

Metrics

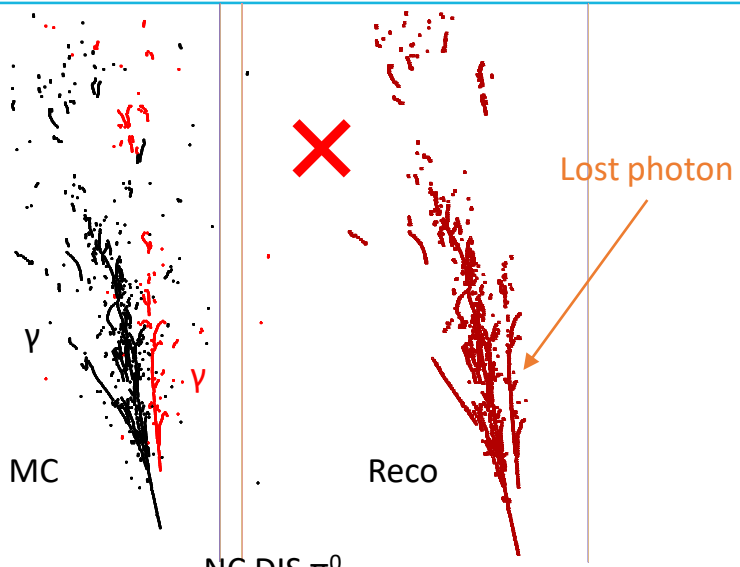
WARWICK

- Generated metrics for 100K events (even split between ν_e and ν_μ)
- Matching between reconstructed particles and MC particles is performed on a folded basis
  - Hits from child particles are folded back to the parent primary particle and considered as a single object for matching purposes (more later)
- Purity – the fraction of reconstructed hits that are shared with the matched MC particle
- Completeness – the fraction of MC hits that are shared with the matched reconstructed particle

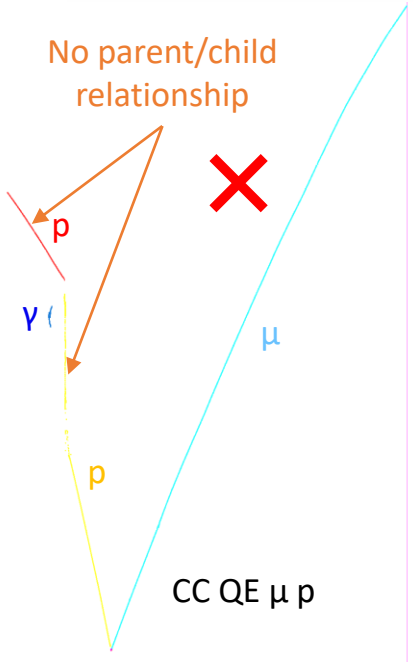
Metrics



CC DIS μ p π^+



NC DIS π^0



CC QE μ p

2 PF particles matched to 1 folded MC object => segmented

- Good match definition
  - A 1-to-1 match between a reconstructed particle and an MC particle exists (N-to-1 matches “segmented”)
  - Purity > 50% and completeness > 10% otherwise “below threshold”
- Correct event fraction
  - An event is correct if **all** primary MC particles have a good match
- Correct leading lepton fraction
  - An event is correct if the leading lepton has a good match

The problem with metrics

WARWICK

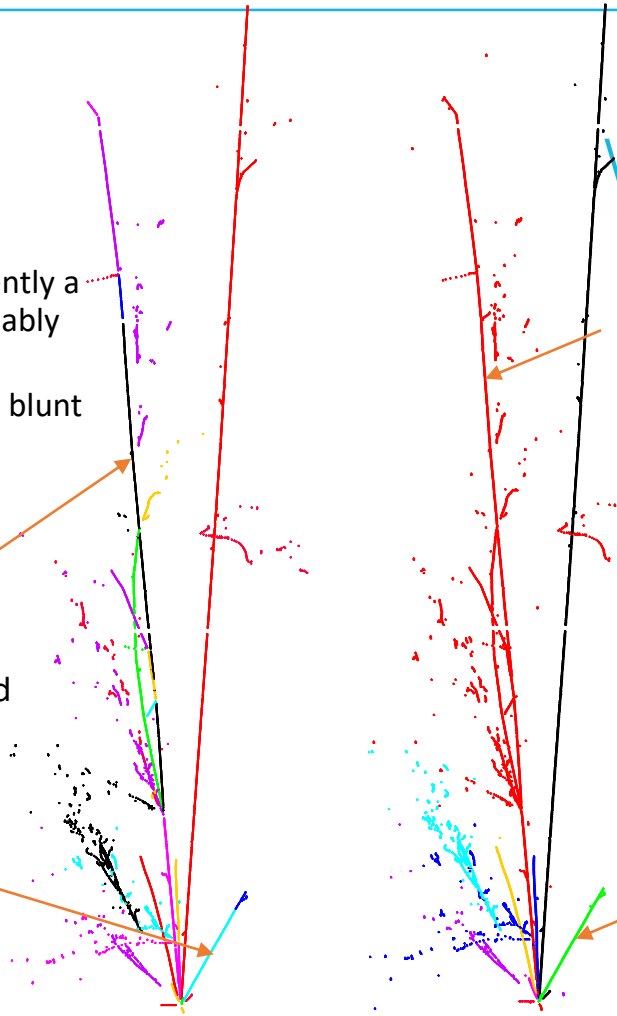
- Defining useful metrics to assess reconstruction performance is a hard problem
  - We can often see the improvements when hand-scanning events, but the limitations in the metrics we've used historically (correct event fraction, purity and completeness) for high stats assessment are increasingly evident
  - Going from 4/8 primaries to 7/8 primaries doesn't change the correct event fraction
  - Going from 8/8 to 7/8 does
- Incorporating some new metrics
  - Some primaries are more important than others => How often do we get the leading lepton right?
  - How much of an event do we get right => Fraction of primaries with good matches
    - Not convinced this is the best way to define this, but it's what we have for now
- Also see Isobel's talk for CP-sensitivity-driven performance assessment

The problem with metrics

- Folding
  - MC particle hierarchies are very detailed – frequently a level of detail well beyond what you could reasonably expect reconstruct
  - This is why we fold back to primaries, but this is a blunt instrument and can hide both problems and improvements…

Numerous interactions of ~6 GeV proton leads to many MC particles that Pandora will (quite reasonably I think) reconstruct as a single track

Charged pion scatter yields two MC particles that Pandora will reconstruct as one track



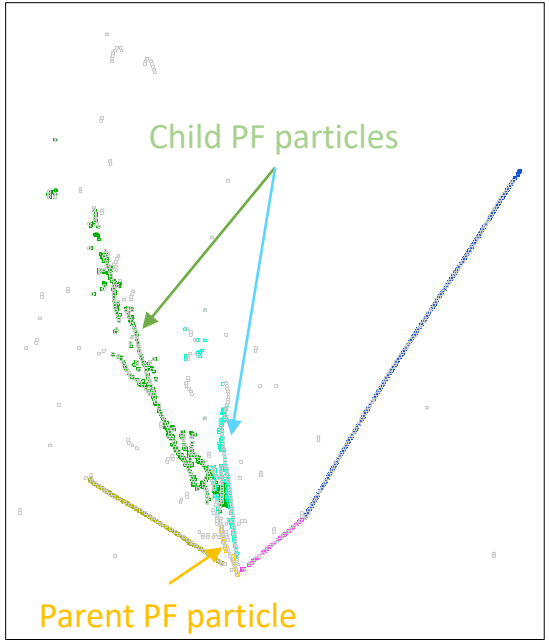
We combine the various proton segments, but a lot of hierarchy detail is now obscured

Sensible

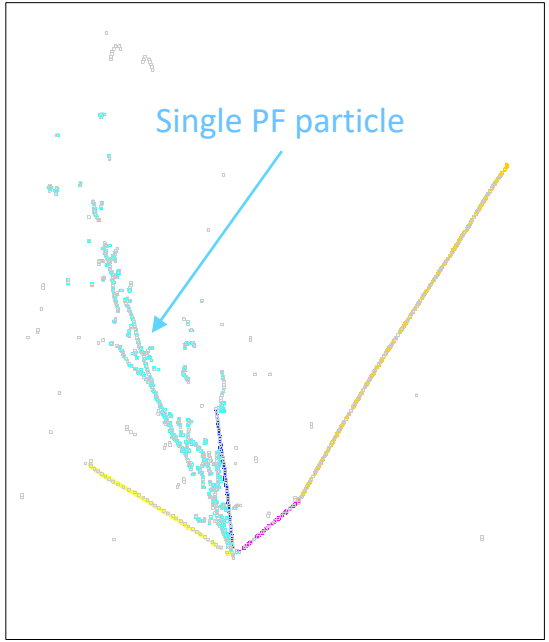
The Folding Problem

WARWICK

- In a fold to primaries regime, the two “electron” hierarchies both qualify as good matches
- One is clearly superior to the other, but you won’t see it in the metrics
- Currently working on a “smart” folding option to find a middle ground between coarse fold to primaries and excessively detailed unfolded hierarchy



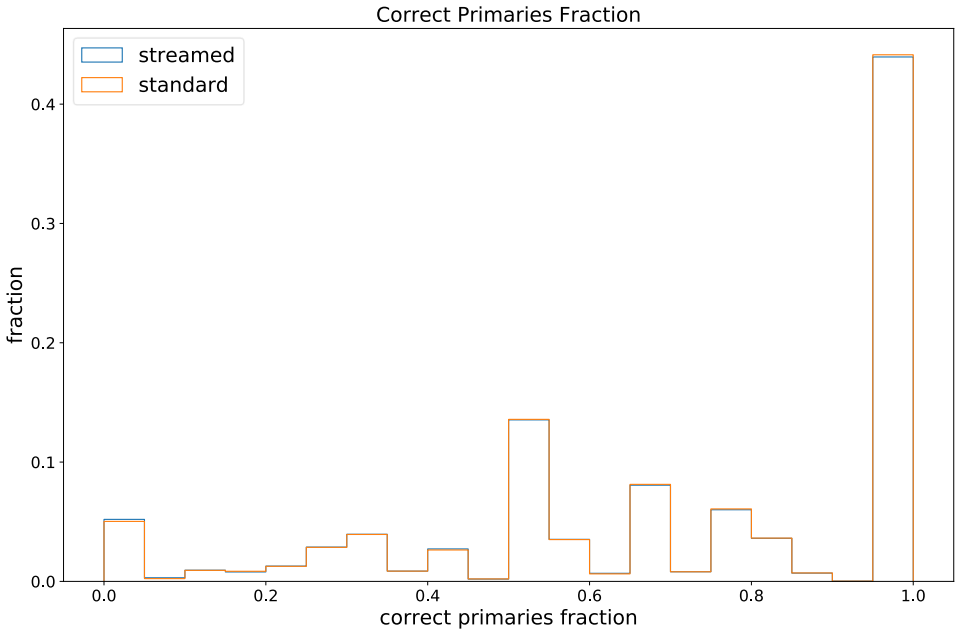
Standard Reco



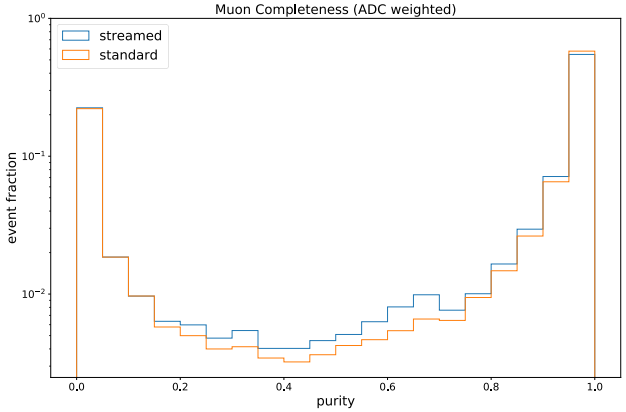
Streamed Reco

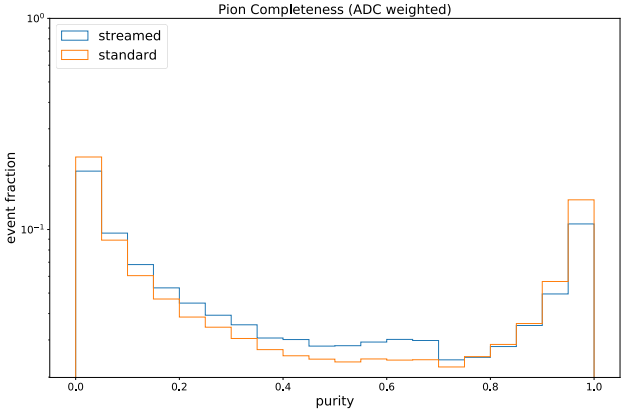
Metrics

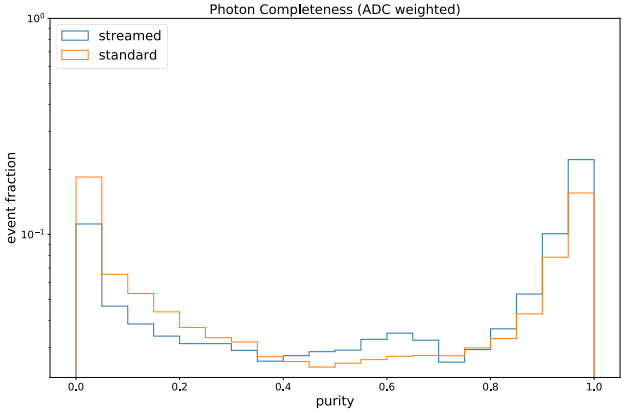
Workflow	Correct event fraction	Correct leading lepton fraction
Standard	45%	89%
Streamed	45%	90%

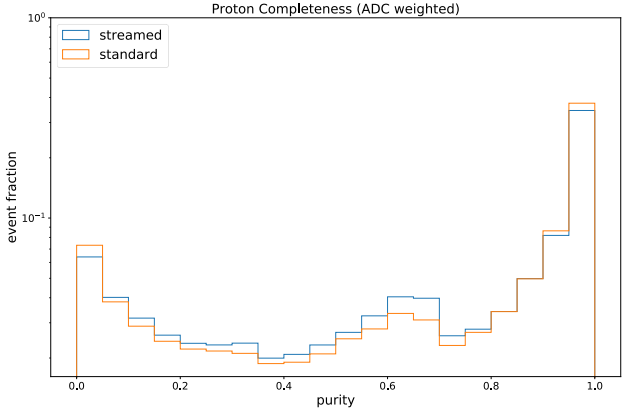


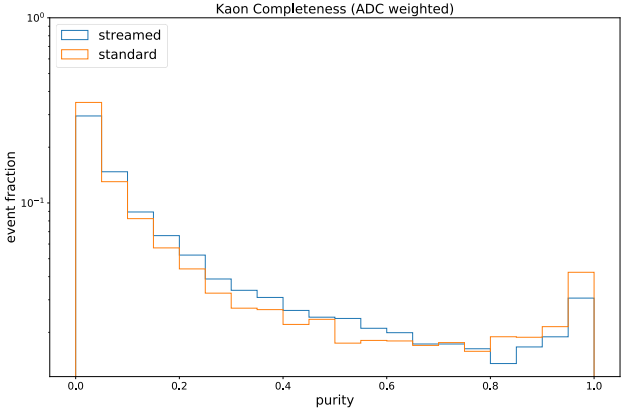
Completeness

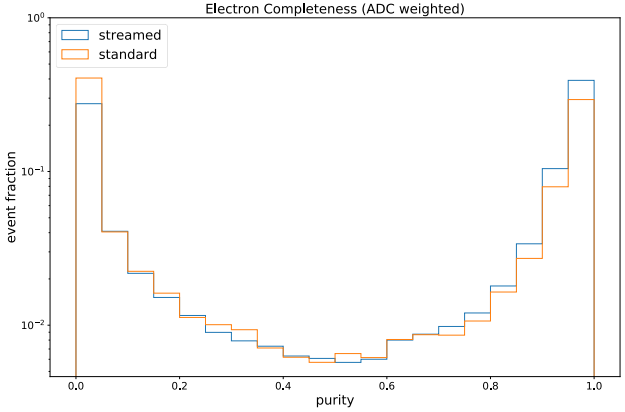






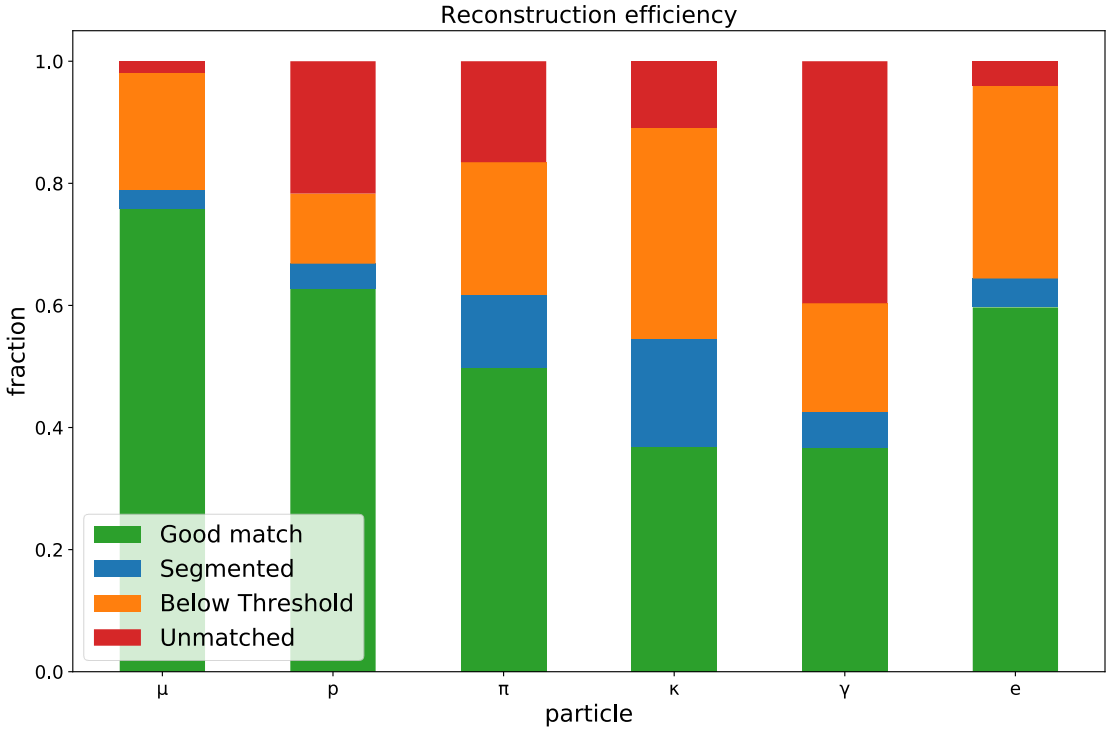






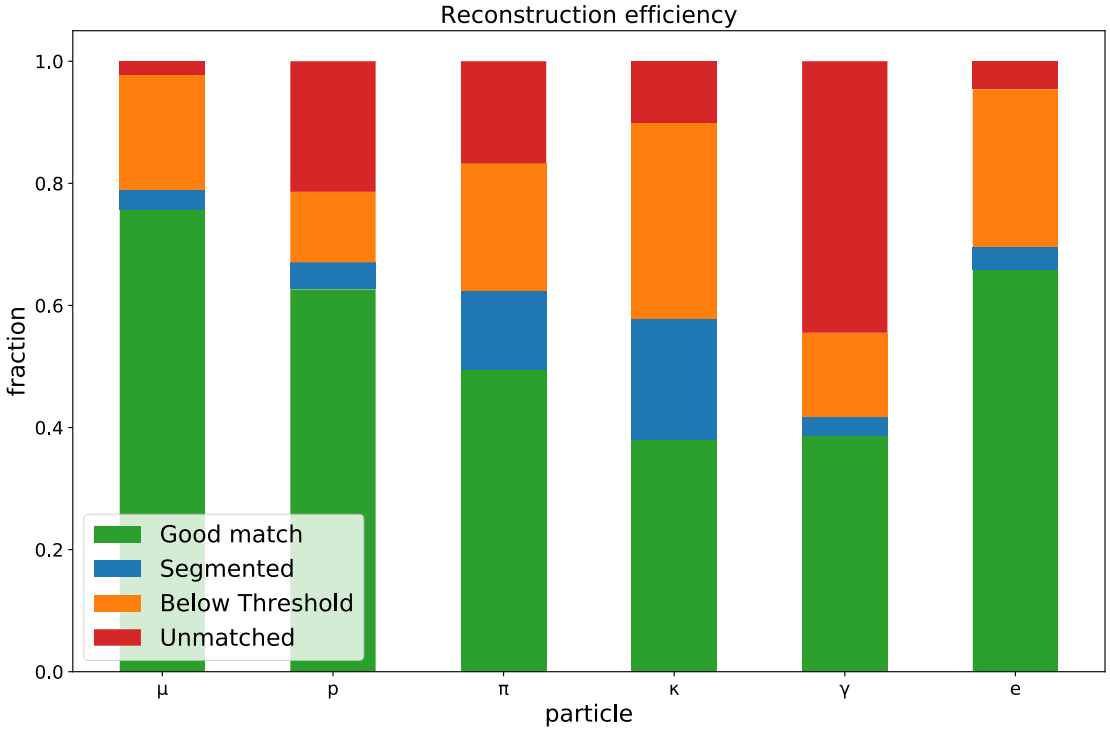
Particle reconstruction efficiency

Standard



Particle reconstruction efficiency

Streamed



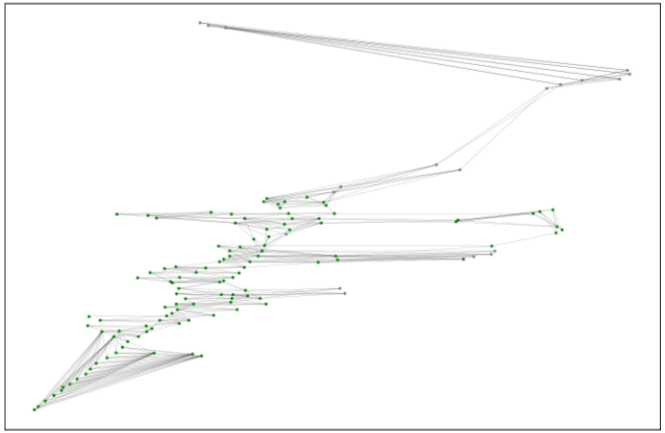
- Tracks essentially unchanged
- Photons, and in particular, electrons see improvements in the fraction of good matches

Deep Learning Shower Growing - Concept

With the deep learning based track-shower ID network, and track/shower streams, we can implement a more intelligent approach to growing the initial shower clusters, to help improve the final reconstruction output.

The initial way of working on this was a Graph Neural Network, which would function similarly to the existing shower growing, but hopefully perform better:

- Take an event, turn it into a graph.
- Pick a cluster as input, run the network.
- Receive an output of "should merge" or "don't merge" for every cluster in the event.



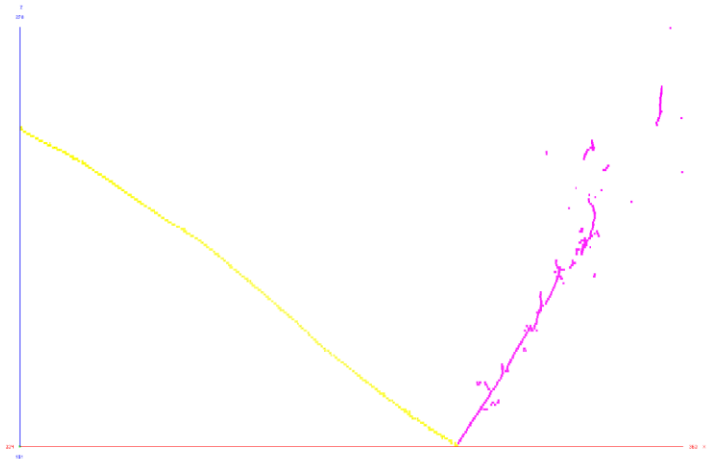
Initial investigation into if shower growing needed improving: here

Deep Learning Shower Growing - Current Progress

Currently, a basic version of both the network (in Python for training), and the usage of the network (a `libtorch` enabled version in C++ inside Pandora) have been completed.

Work is being done to move through various test datasets, that ramp up the difficulty and add additional needs to the network (one shower to grow, multiple to grow, full neutrino events).

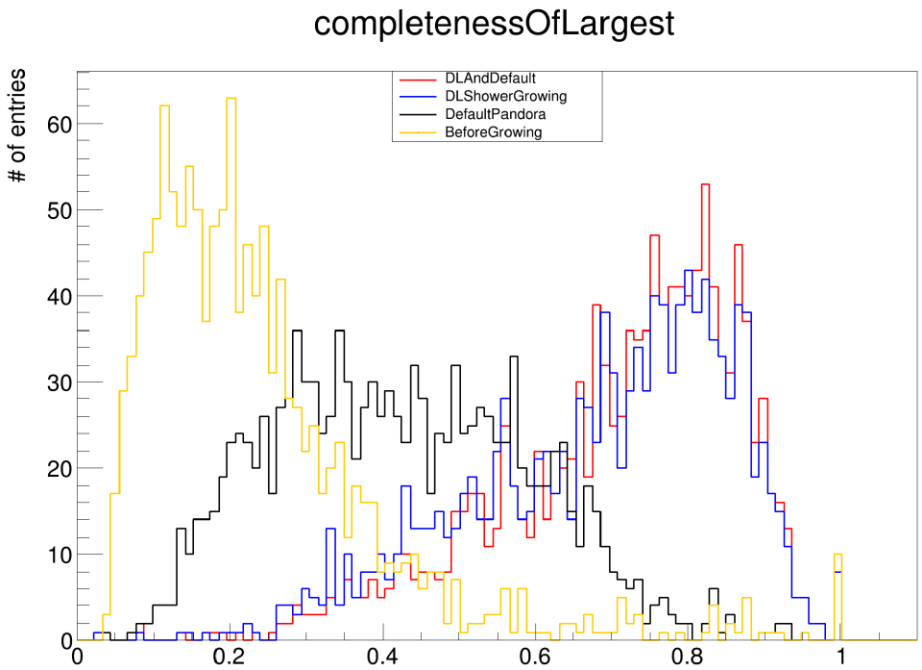
Results shown here are for a small (~400) sample of single track, single shower events. That is, 1 muon and 1 electron particle gun, from a single point with a flat distribution in angle between them.



A more thorough explanation of network here

Deep Learning Shower Growing - Results

The main result to show is this, a comparison of the largest showers completeness, before and after shower growing. This compares before shower growing (common for all 3 results), against the current shower growing, the DL based one, and both running in series. The final performance follows.



Deep Learning Shower Growing - Remarks

The shown results (and full results set linked in the footnote) show that the new deep learning based shower growing is able to grow single showers effectively, and do so in a way that results in improved completeness for analysers. This table shows the end performance, based on the largest shower, for the test dataset.

Mode	Completeness	Purity
Existing	86.2%	98.4%
DL-based	90.6%	94.9%
Both	90.9%	94.8%

There is lots of work to be done to extend from a single shower (multiple runs, stopping criteria, input cluster selection), and parts of that are being worked on to test in a two shower dataset, which will also help demonstrate how the network deals with two close showers.

Its also encouraging that this work was done in a somewhat basic way, not fully utilising the track/shower streaming, so its possible that deeper integration will enable even better performance (either through purer clusters or more intelligent ways of picking input clusters).

Full results for test dataset here.

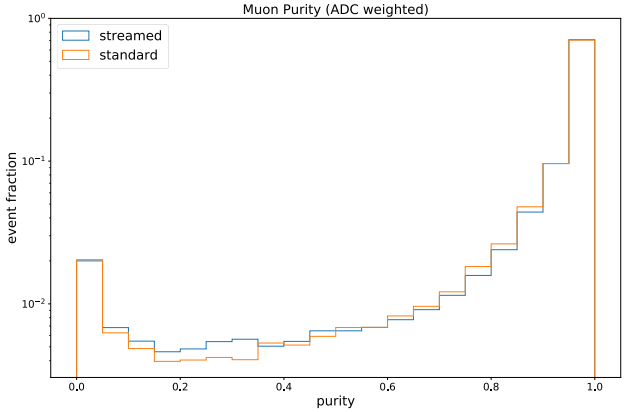
Summary

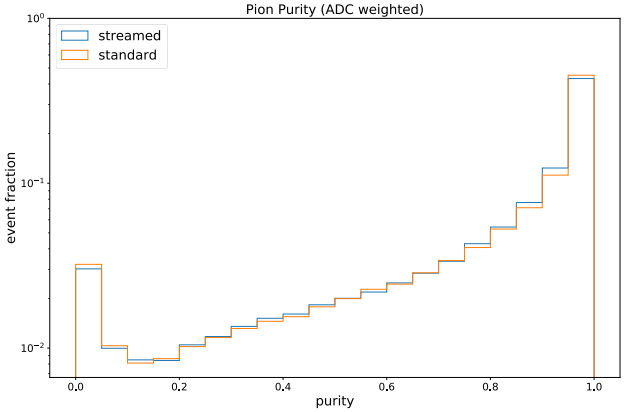
WARWICK

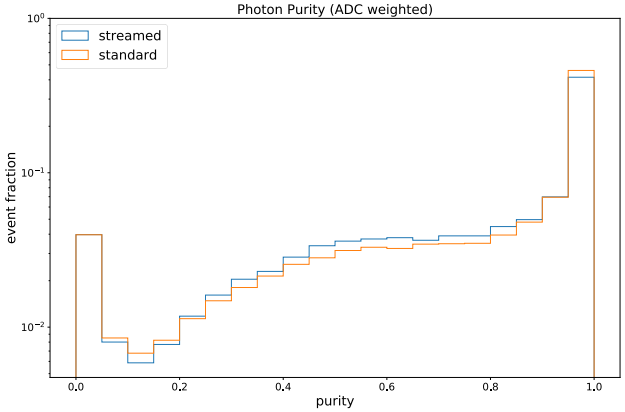
- Pandora can now seamlessly integrate deep learning into the multi-algorithm approach
- Splitting reconstruction into streams that separately process track-like and shower-like topologies shows promise and allows for algorithms targeting each topology
- GNN in development to grow shower clusters
- New metrics under development to better reflect changes in reconstruction performance

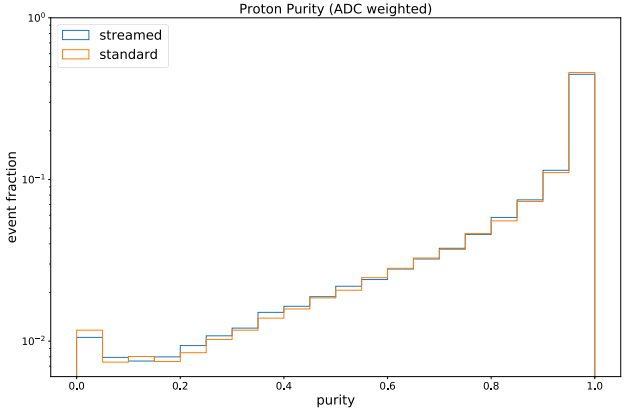
Backup

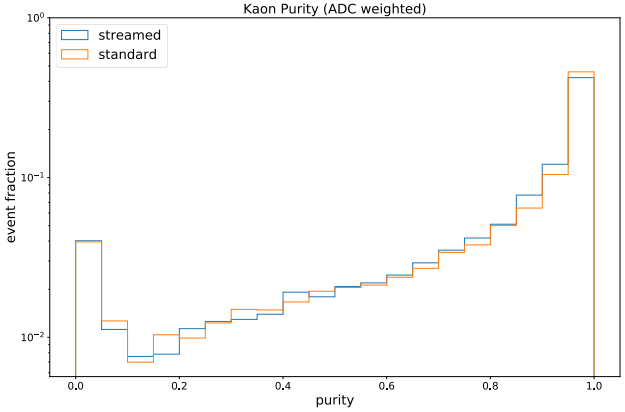
Purity

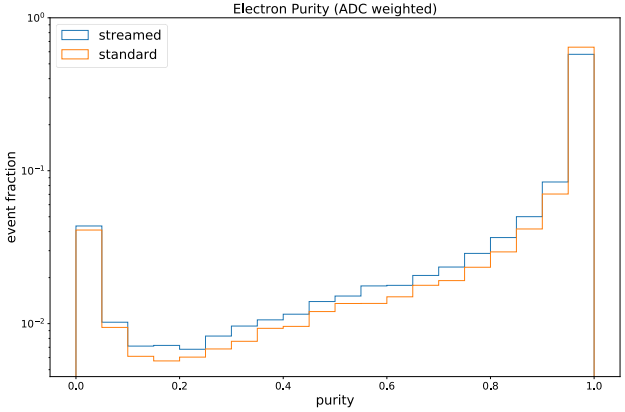

Muon Purity (ADC weighted)
streamed
standard
event fraction
purity


Pion Purity (ADC weighted)
streamed
standard
event fraction
purity


Photon Purity (ADC weighted)
streamed
standard
event fraction
purity


Proton Purity (ADC weighted)
streamed
standard
event fraction
purity


Kaon Purity (ADC weighted)
streamed
standard
event fraction
purity


Electron Purity (ADC weighted)
streamed
standard
event fraction
purity