DUNE FD Sim/Reco Meeting
24th of June 2019

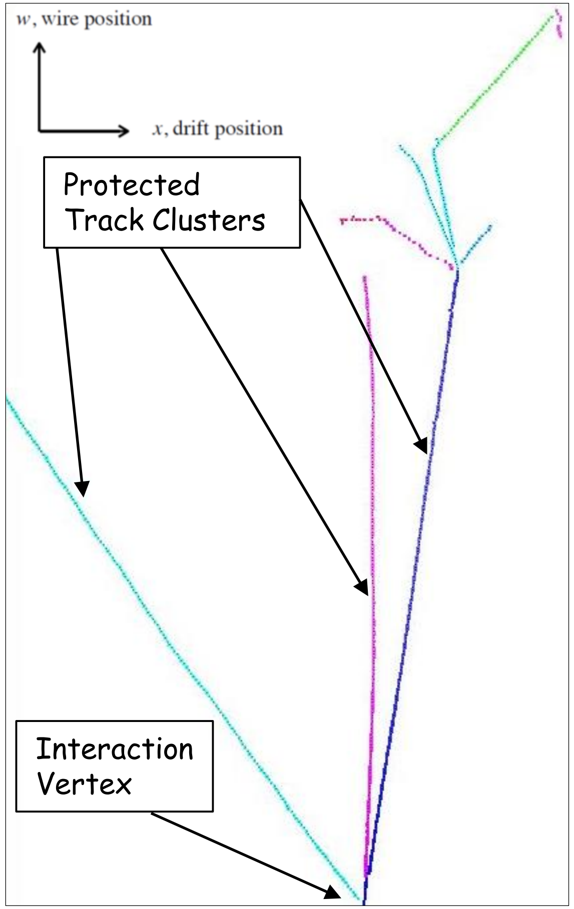
Updates on Vertex Reconstruction in Pandora

J. Ahmed. First Year PhD Student. On behalf of the Pandora Team
Supervisor: Dr John Marshall

Some Figures/Text taken from previous talks by Steven Green and Jack Anthony.

Introduction

- The position of the primary neutrino interaction vertex is interesting in itself (for physics analysis) as well as being a helpful tool to improve the pattern recognition.

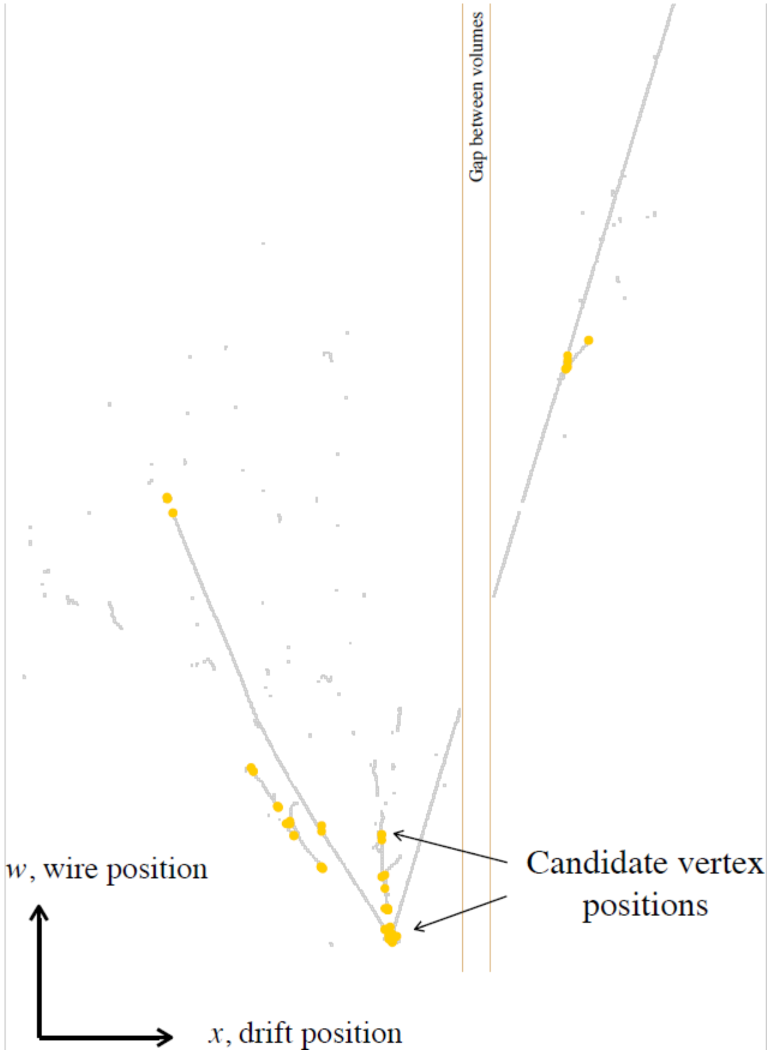


- For example, correct identification of the neutrino interaction vertex helps algorithms to identify individual primary particles and to ensure that they each result in separate reconstructed particles.

- My aim is to improve the vertex reconstruction algorithm that was originally designed and optimised for μBoone and make sure it is suitable for the DUNE Far Detector.

- This is beginning with a modest update/refresh and will then move on to new developments incorporating Deep Learning.

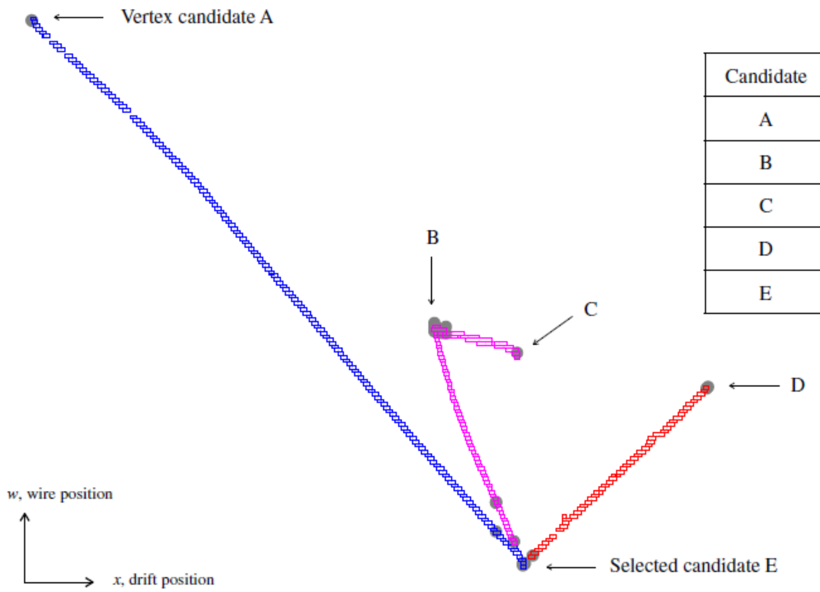
Vertexing in Pandora



The vertex is identified in Pandora at the 2D stage of the reconstruction. The process comprises 2 algorithms:

- The **Candidate Vertex Creation Algorithm** which peppers the event with a large number of 3D candidates. A candidate vertex is created using features of pairs of matched 2D clusters from separate readout planes.
- The **Vertex Selection Algorithm** tries to find the primary interaction vertex out of these candidates. Future plans for the algorithm to try to find any secondary vertices too.

EnergyKickVertexSelectionAlgorithm



Candidate	S	$S_{energy\ kick}$	$S_{asymmetry}$	$S_{beam\ deweight}$
A	1.3E-06	3.5E-06	2.72	0.14
B	3.5E-02	3.1E-02	2.69	0.42
C	2.9E-03	2.4E-03	2.59	0.46
D	1.6E-09	1.1E-09	2.72	0.52
E	2.4E+00	9.0E-01	2.72	0.99

This selection algorithm works by assigning each candidate a value for three types of scores.

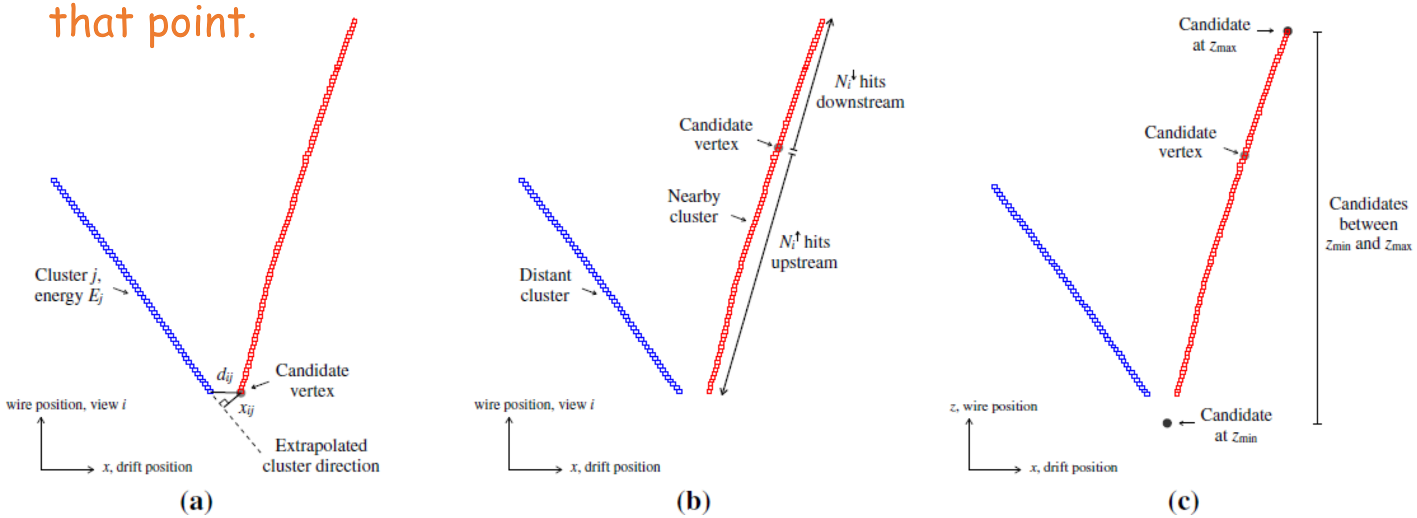
The candidate with the highest product of the three scores is chosen as the primary interaction vertex.

This is an ad hoc approach, using relevant features, and lends itself to later combining these scores in a formal multivariate approach.

Scoring the candidates involves projecting the 3D position into each of the 2D views and examining the local topology about that point.

beamDeweightingScore X energyKickScore X energyAsymmetryScore

Source: Pandora MicroBooNE Paper. Eur. Phys. J. C. (2018) 78: 82



Boosted Decision Trees (BDTs) Approach

- One of the advantages of Pandora's multi-algorithm approach is that individual algorithms can easily be added/removed/replaced, making it easy to plug in Machine Learning (ML)/Deep Learning (DL) modular solutions in areas where they could cause an improvement in performance.

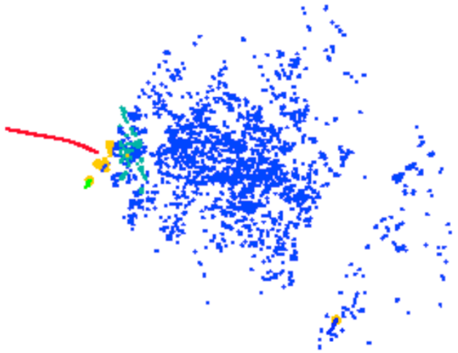
- Pandora supports the use of BDTs as well as Support Vector Machines (SVMs) and other ML/DL approaches.

- New vertex selection algorithm uses BDTs to select the vertex from the candidates similar to the SVM algorithm currently being used in MicroBooNE (originally worked on by **Jack Anthony**) and the BDT algorithm **Steven Green** is working on for ProtoDUNE.

- The problem is split into a **vertex-region-finding algorithm, which evaluates different regions in the event,** and a **vertex-finding algorithm, which chooses the most appropriate vertex in a given region,** since the problems are quite different. This yields a performance benefit.

Example in ProtoDUNE
by Steven Green

Currently



BDT Vertexing

Vertexing in Pandora - BDT Method - Feature Selection

The BDT score is created using variables borrowed from the SVM used for MicroBooNE (work originally by **Jack Anthony**)

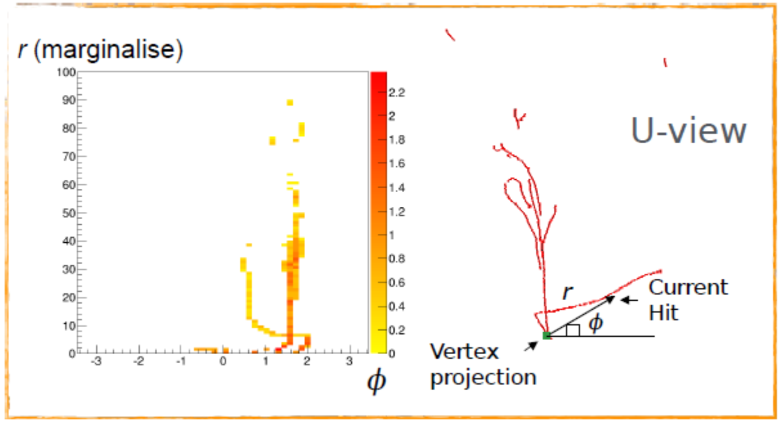
There are two types of variables that go into the BDT, Vertex Features (i.e. related to vertex position), plus some Event-Based Features (i.e. related to the hit distribution only) that allow for different topologies to be treated differently.

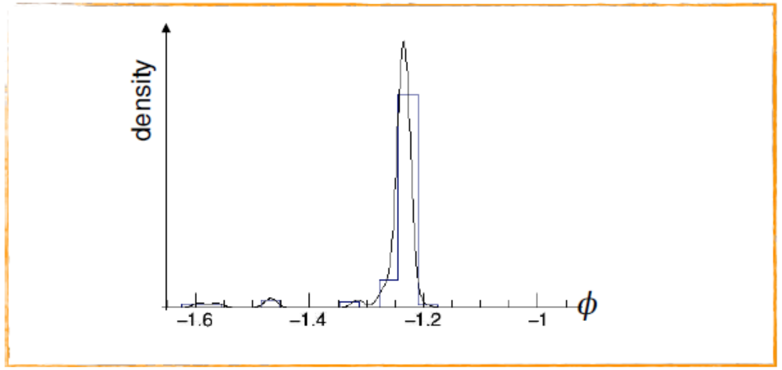
Region-Finding BDT:

- Variables:
  - ➔ Event-Based Features:
    - ➔ Shower Fraction (proportion of shower-cluster-associated hits)
    - ➔ Total Energy
    - ➔ Volume Spanned
    - ➔ Longitudinality (ratio of zspan to xspan)
    - ➔ Number of Hits
    - ➔ Number of Clusters
    - ➔ Number of Vertex Candidates
  - ➔ Vertex-Based Features:
    - ➔ Energy Kick
    - ➔ Local Asymmetry
    - ➔ Beam De-weighting
    - ➔ Global Asymmetry
    - ➔ Shower Asymmetry

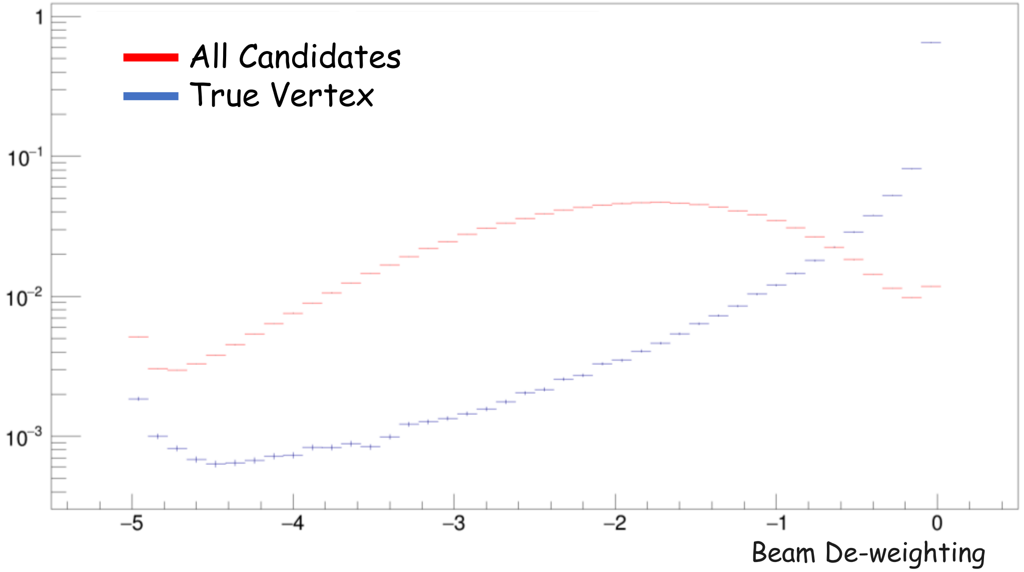
Vertex-Finding BDT:

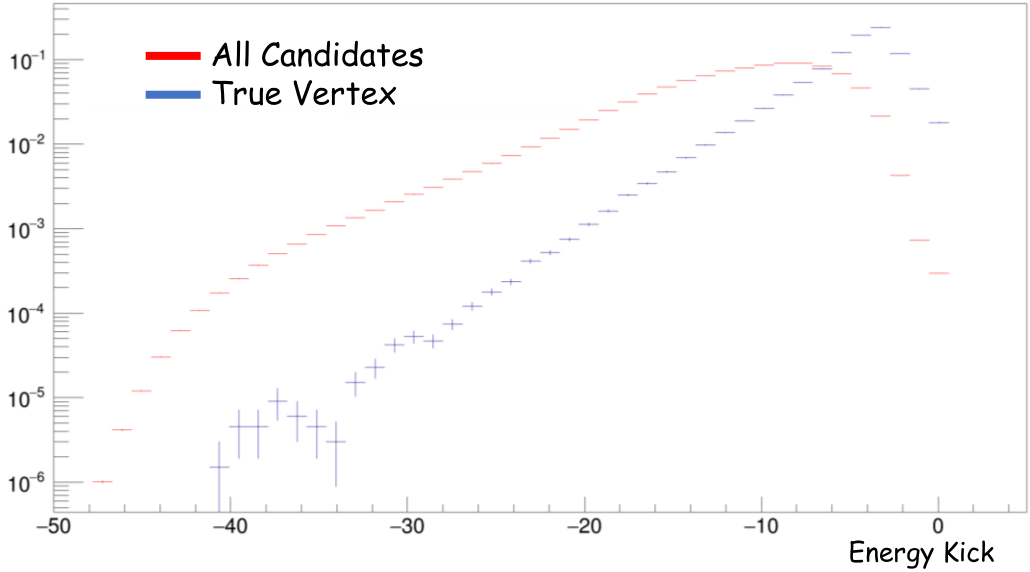
- Same as above plus the vertex-based r/phi feature.

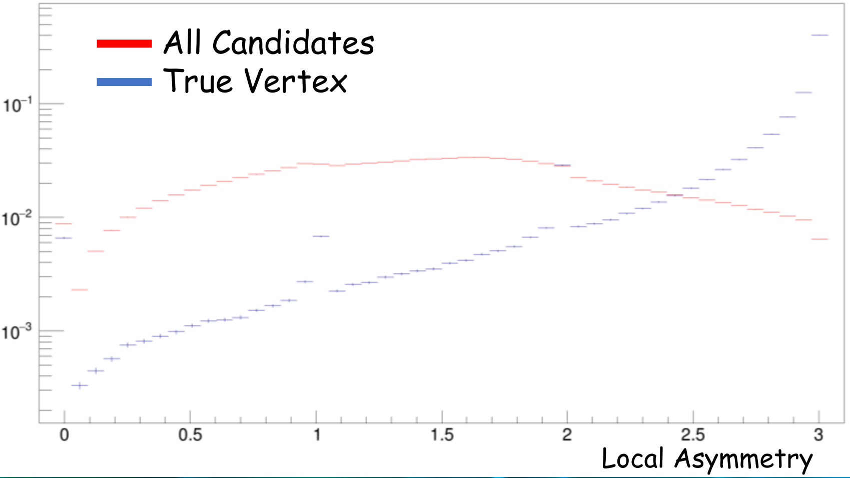


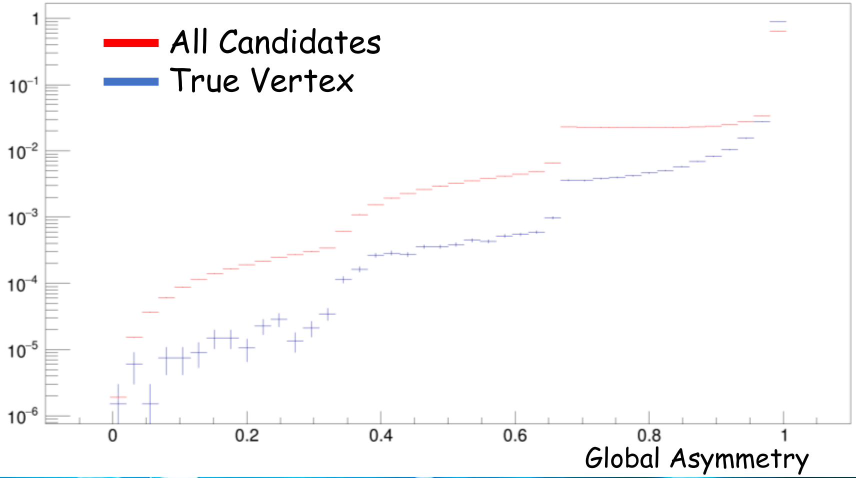


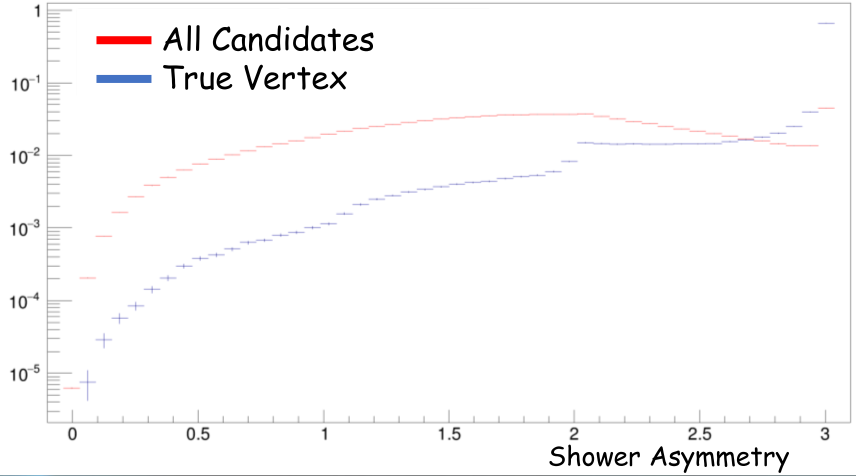
Distributions of Vertex Scores - All Interactions "nue" - MCC11



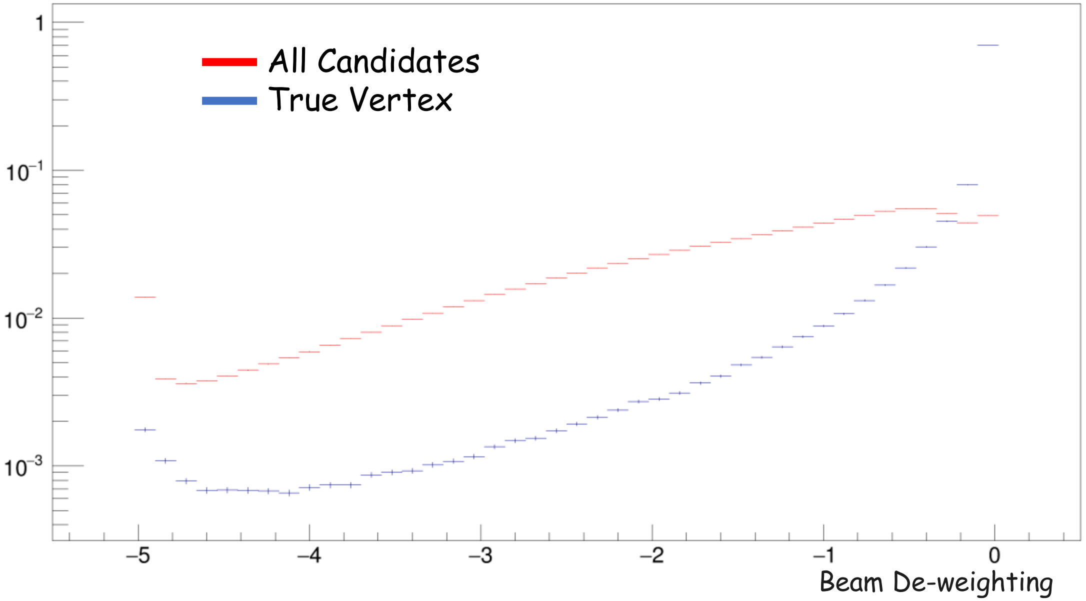


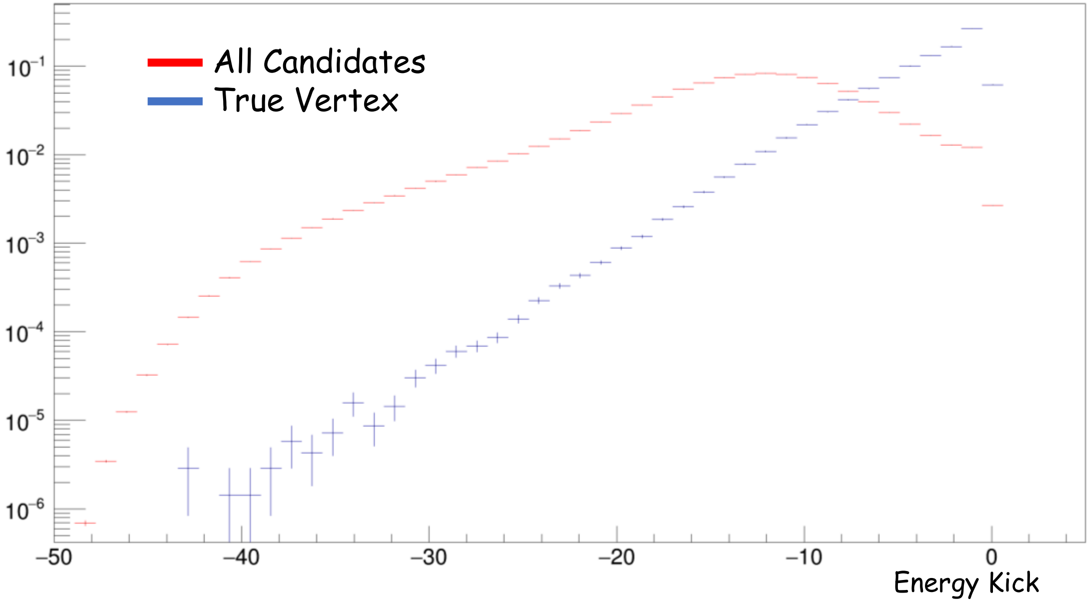


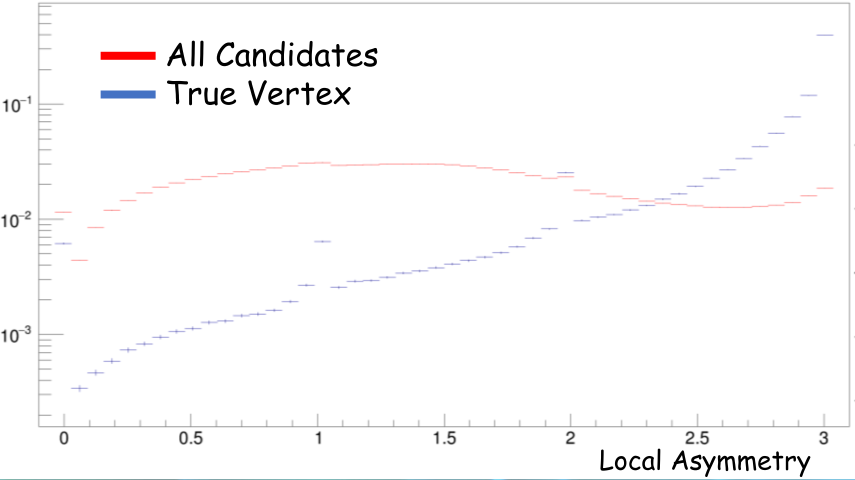


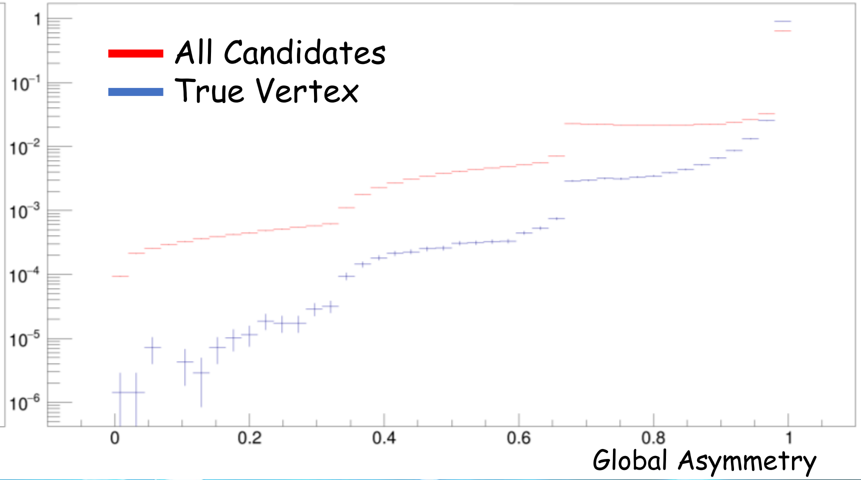


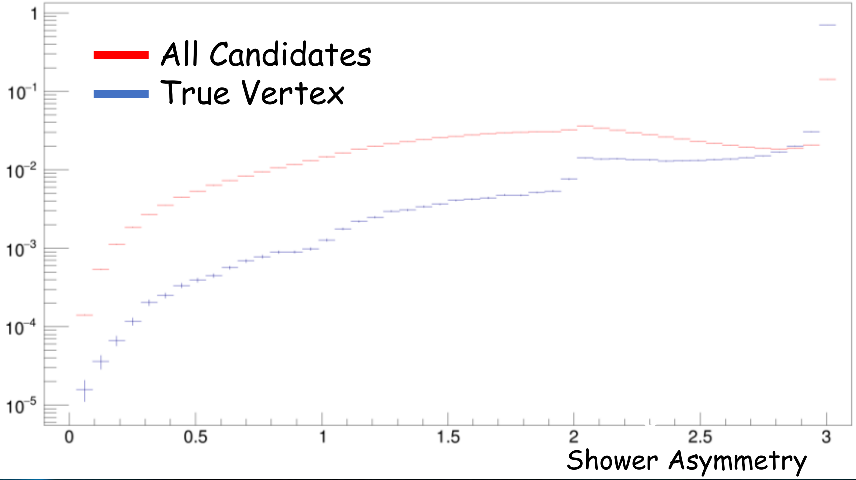
Distributions of Vertex Scores - All Interactions "numu" - MCC11











BDT Vertex Selection Algorithm – Training Set Production

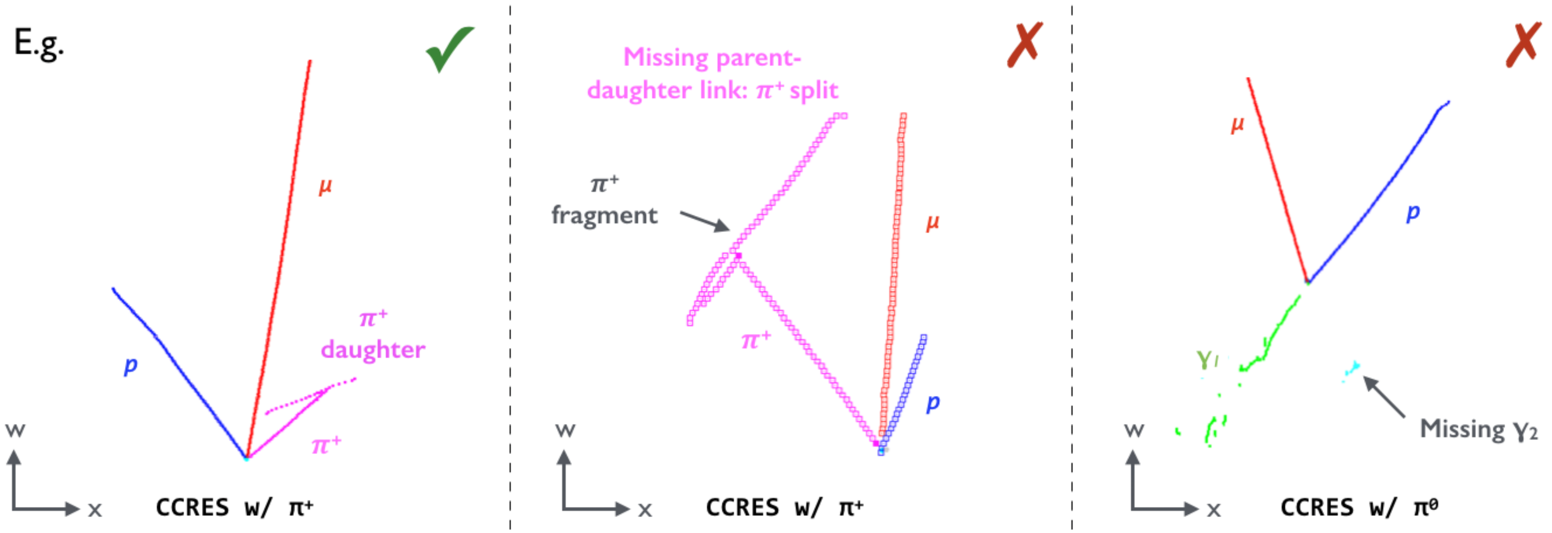
- The reconstruction is run over the training samples with the BDT Vertex Selection algorithm in training mode.
- Two training sets are produced: one for the region-finding BDT and one for the vertex-finding BDT.
- The training is done offline using the python module sklearn, then the output is packaged as an XML file to be read by Pandora.
- For more information on sklearn see https://scikit-learn.org. For scikit-learn BDT description and examples see https://scikit learn.org/stable/modules/tree.html.
- The BDT was trained and then tested using MCC11 events with an equal number of nue and numu events.

Pandora Reconstruction Metrics

- To assess performance for simulated DUNE events, similar performance metrics were used as at MicroBooNE.

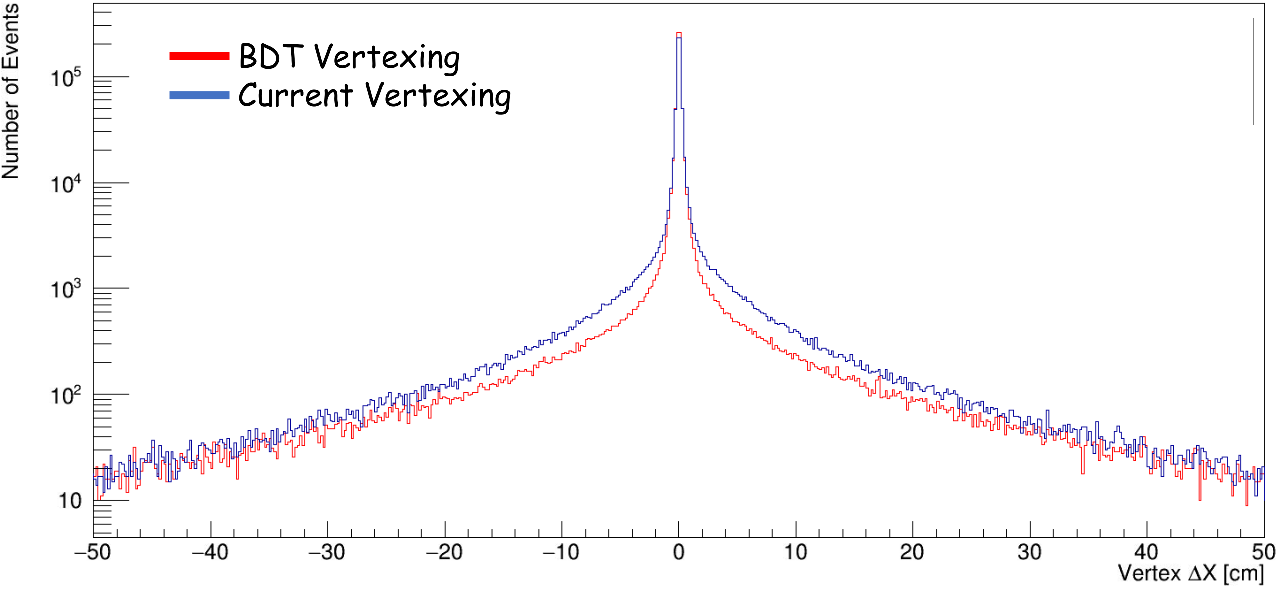
- Examine fraction of events deemed "correct" by very strict pattern-recognition metrics:
  - Consider exclusive final-states where all true particles pass simple quality cuts (e.g. nHits)
  - Correct means exactly one reconstructed primary particle is matched to each true primary particle

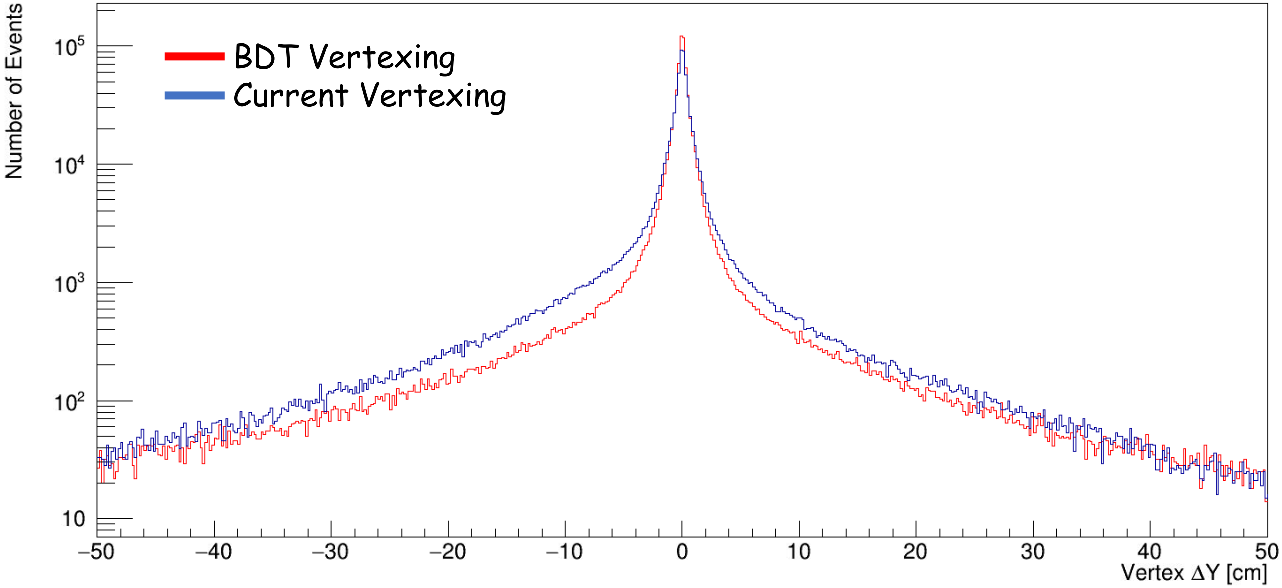
- For vertexing, in specific, main two performance metrics are "dR68%" and "%>50cm".
  - "dR68%" is the value of deltaR where 68% of events have a deltaR less than that.
  - "%>50cm" is the percentage of events that have a deltaR larger than 50cm

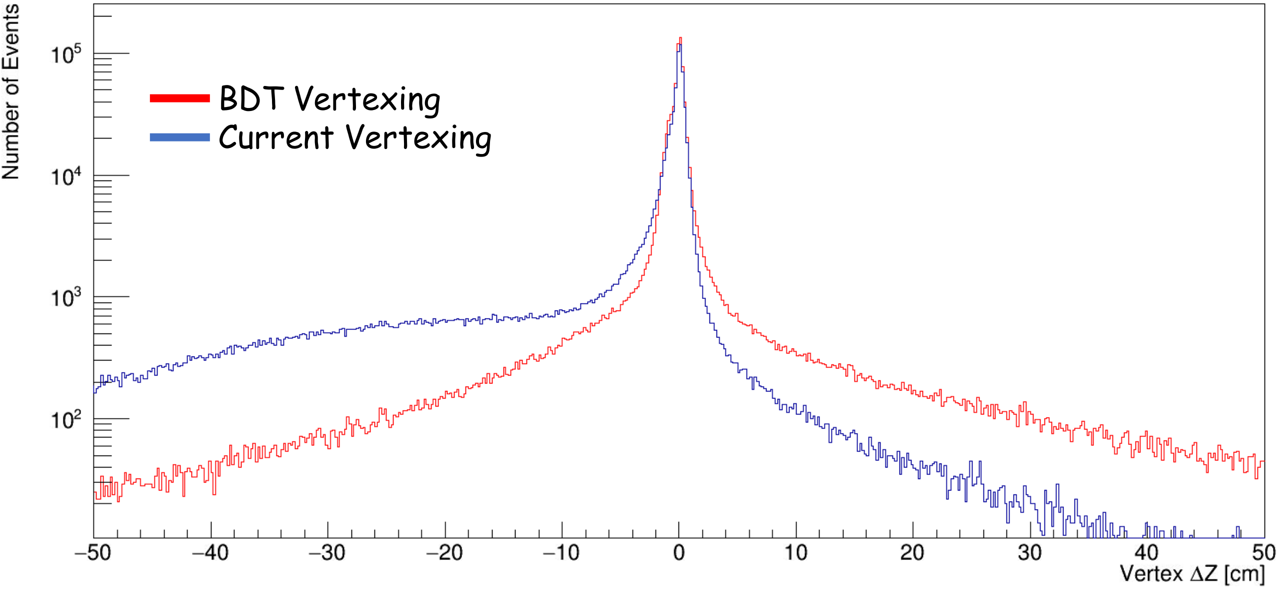


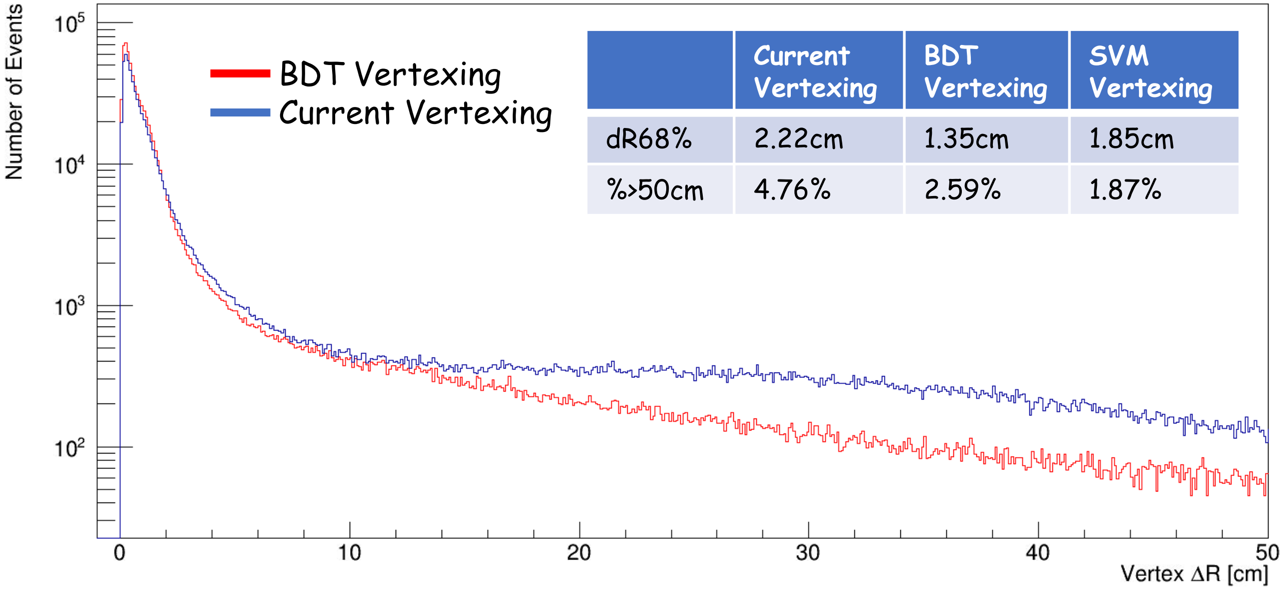
For more details see: Pandora MicroBooNE Paper. Eur. Phys. J. C. (2018) 78: 82

(True-Reco) Vertex Positions Distributions: All Interactions "nue"



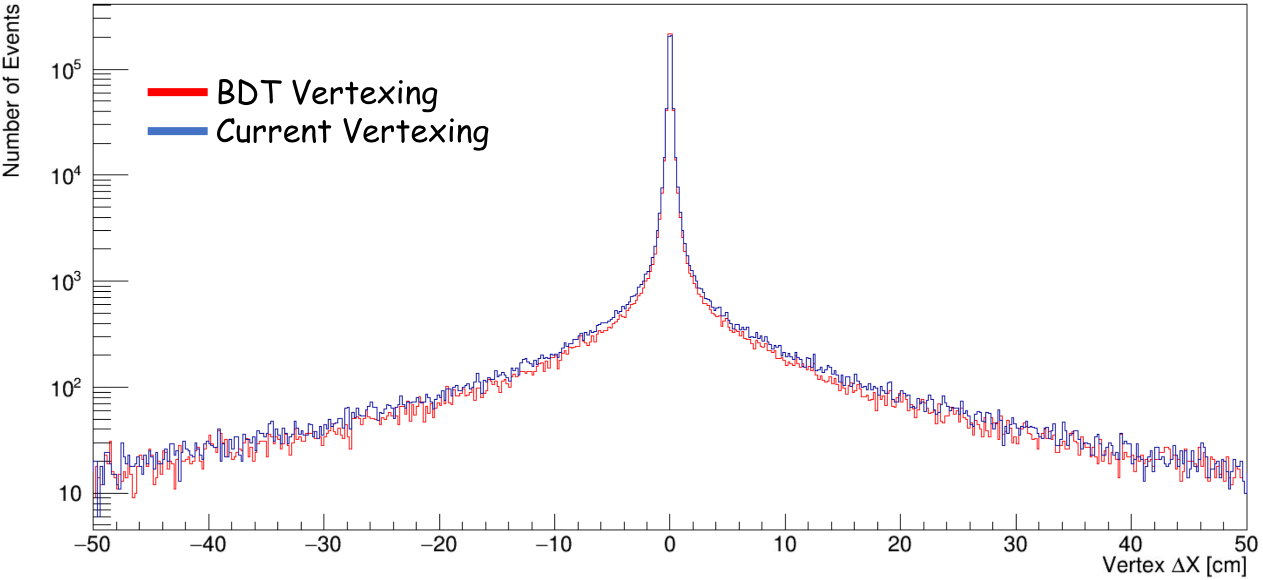


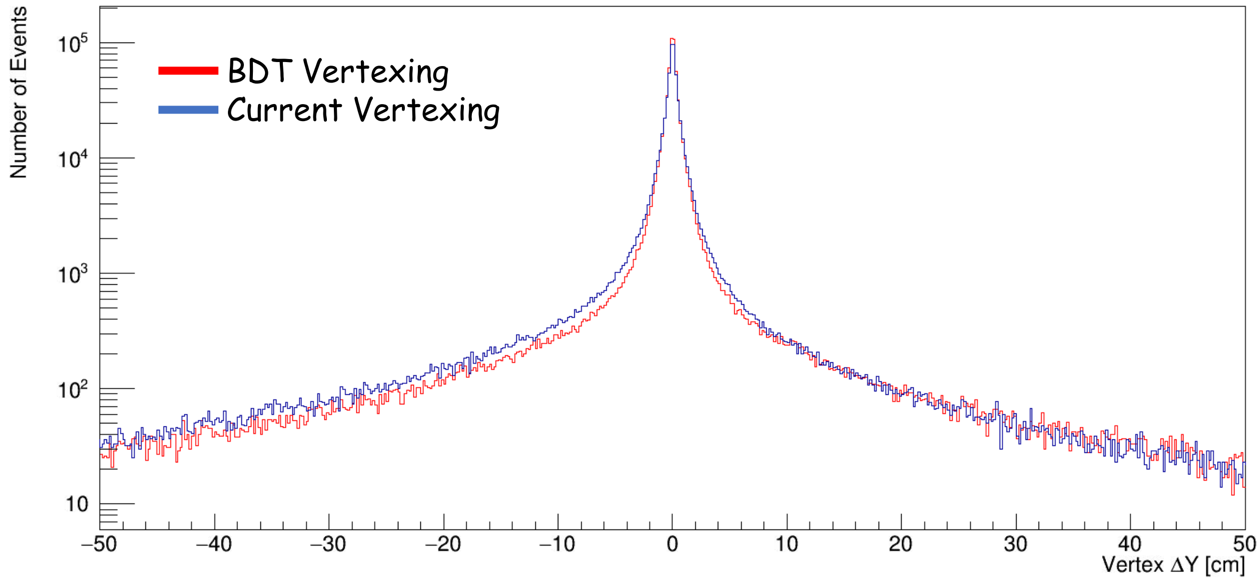


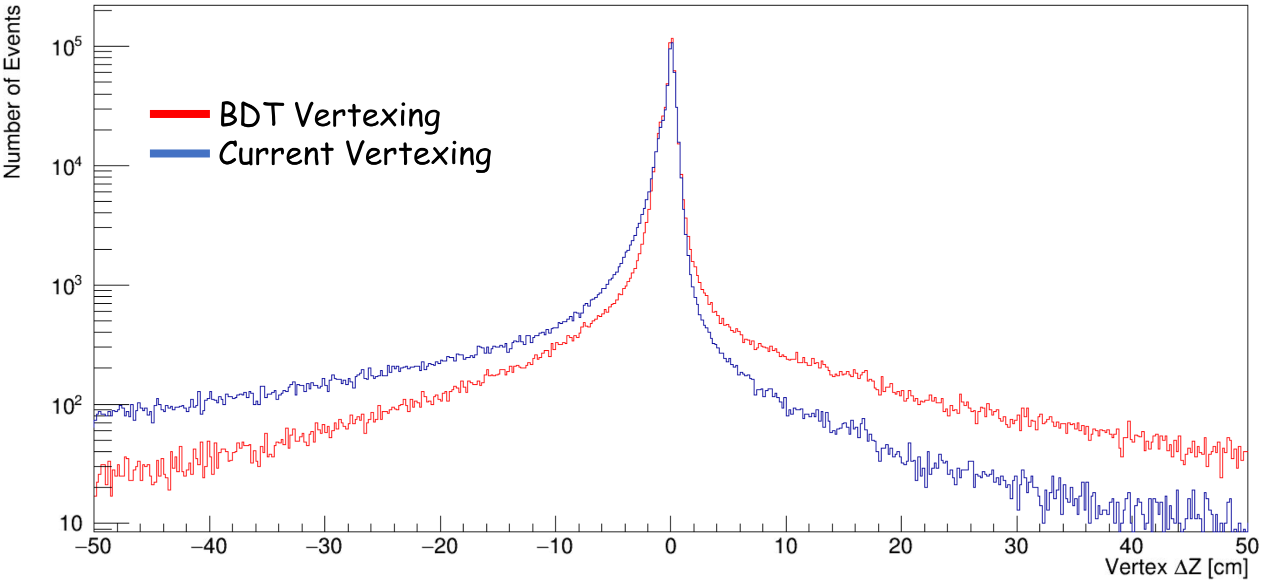


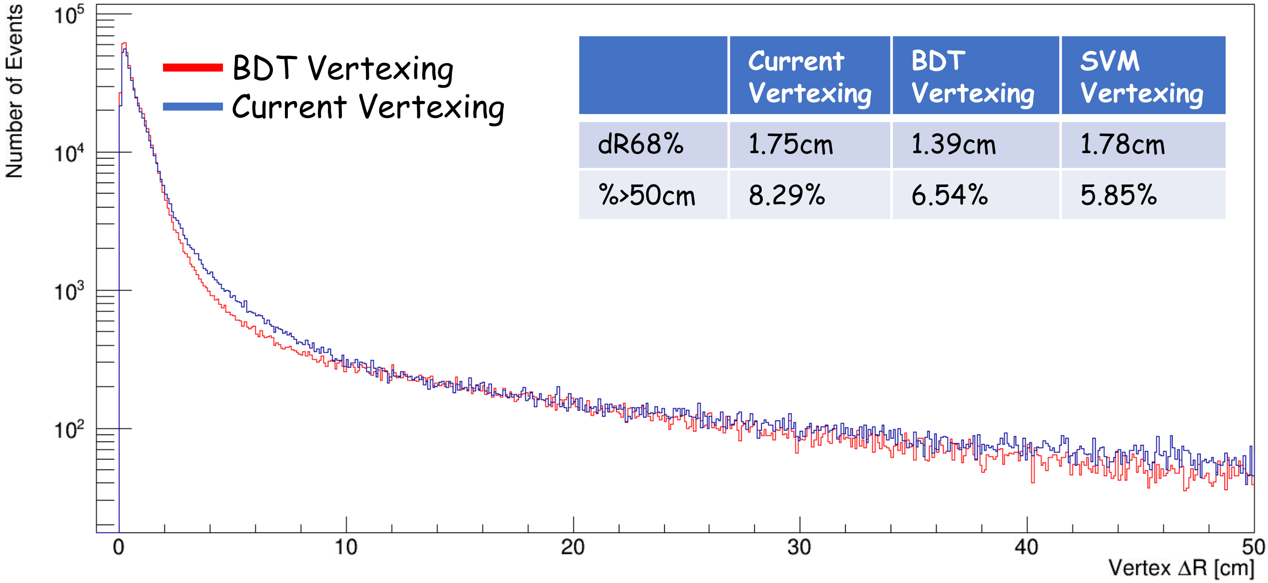
	Current Vertexing	BDT Vertexing	SVM Vertexing
dR68%	2.22cm	1.35cm	1.85cm
%>50cm	4.76%	2.59%	1.87%

(True-Reco) Vertex Positions Distributions. All Interactions "numu"



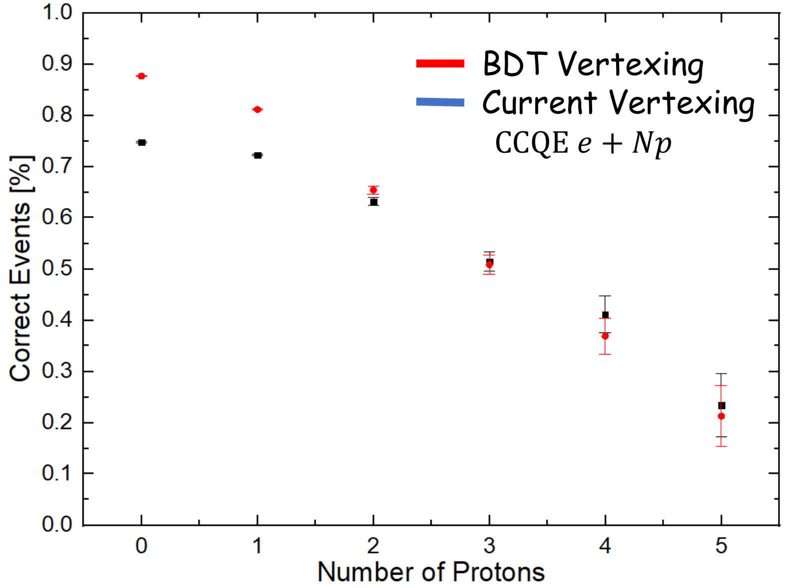


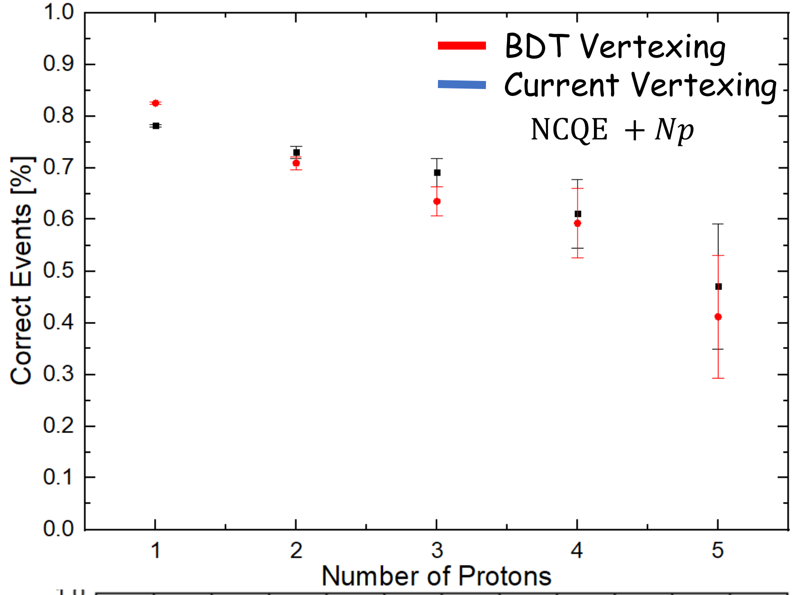


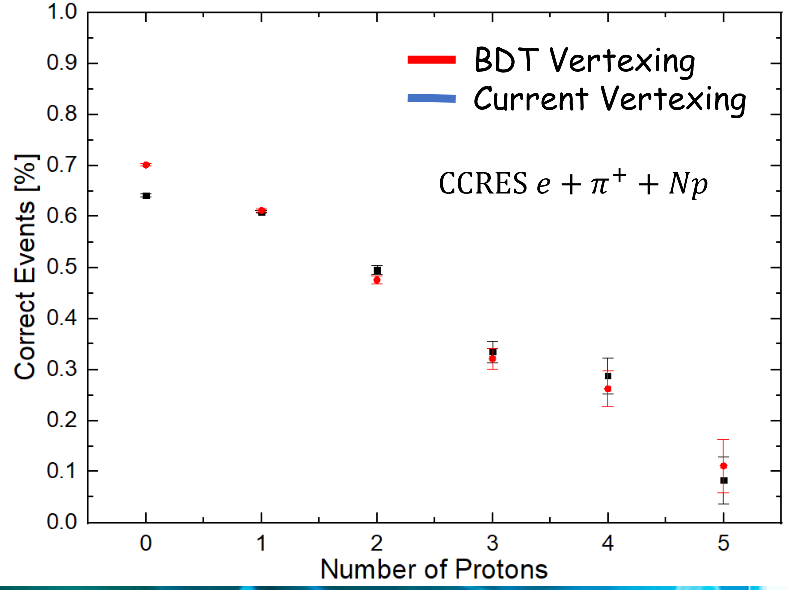


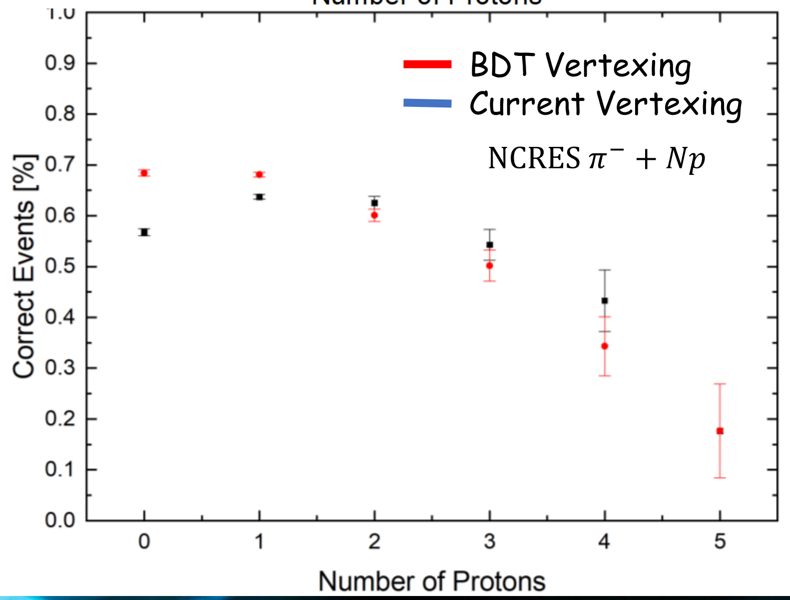
	Current Vertexing	BDT Vertexing	SVM Vertexing
dR68%	1.75cm	1.39cm	1.78cm
%>50cm	8.29%	6.54%	5.85%

Correct Event Fraction Plots

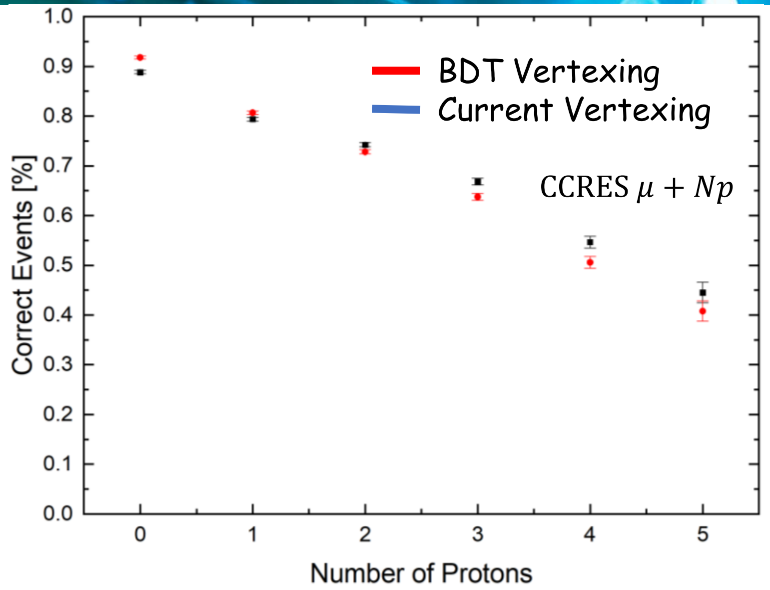


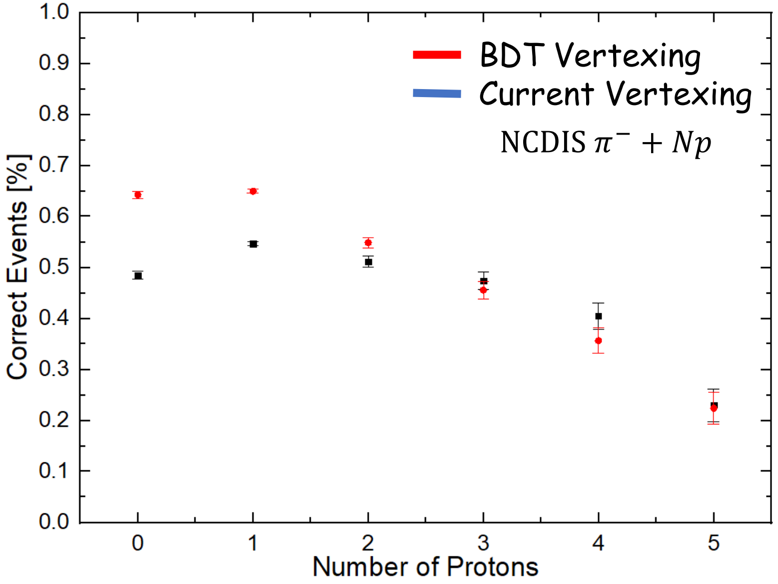


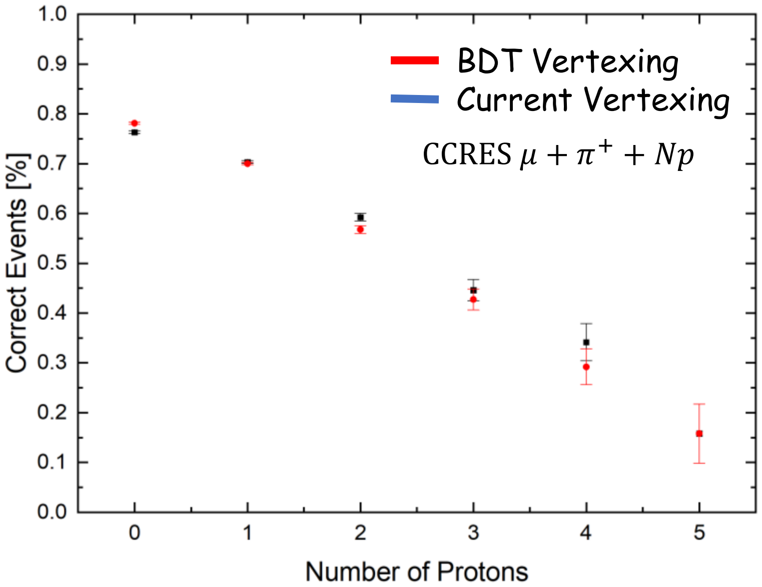


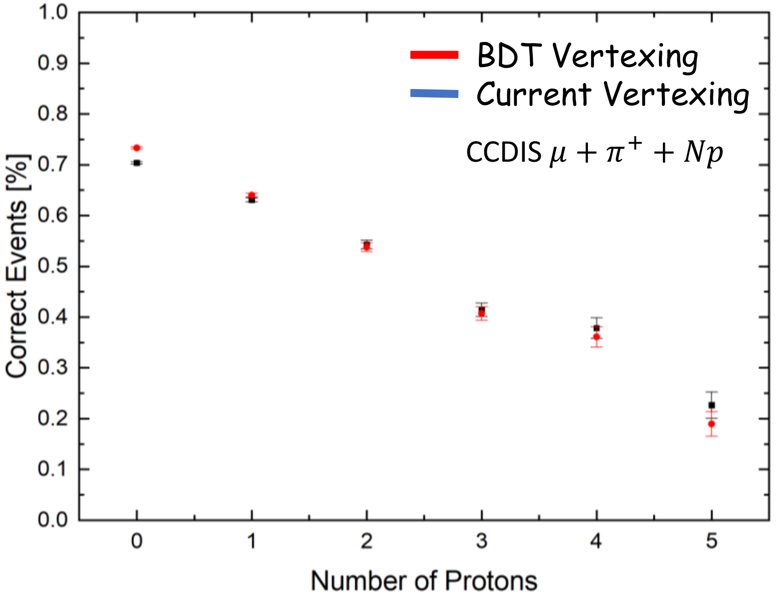


Correct Event Fraction Plots





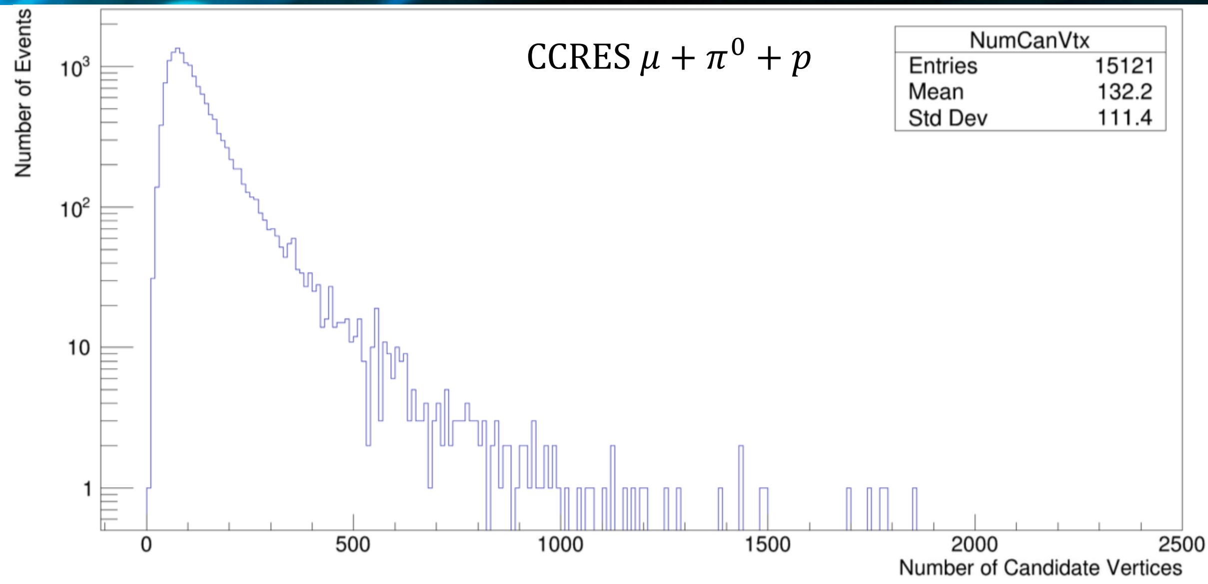


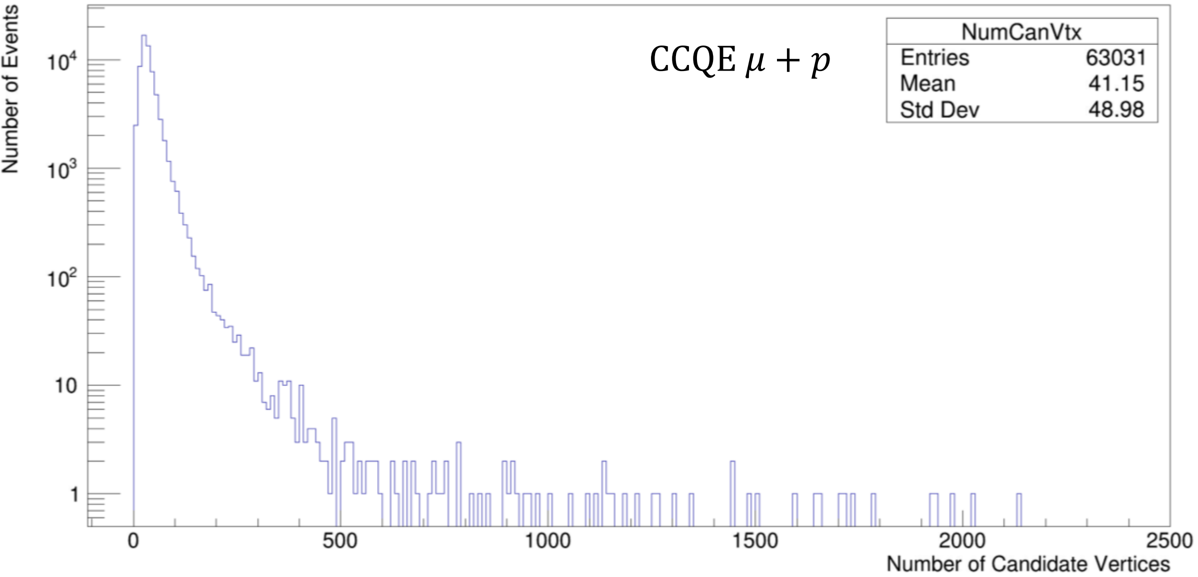


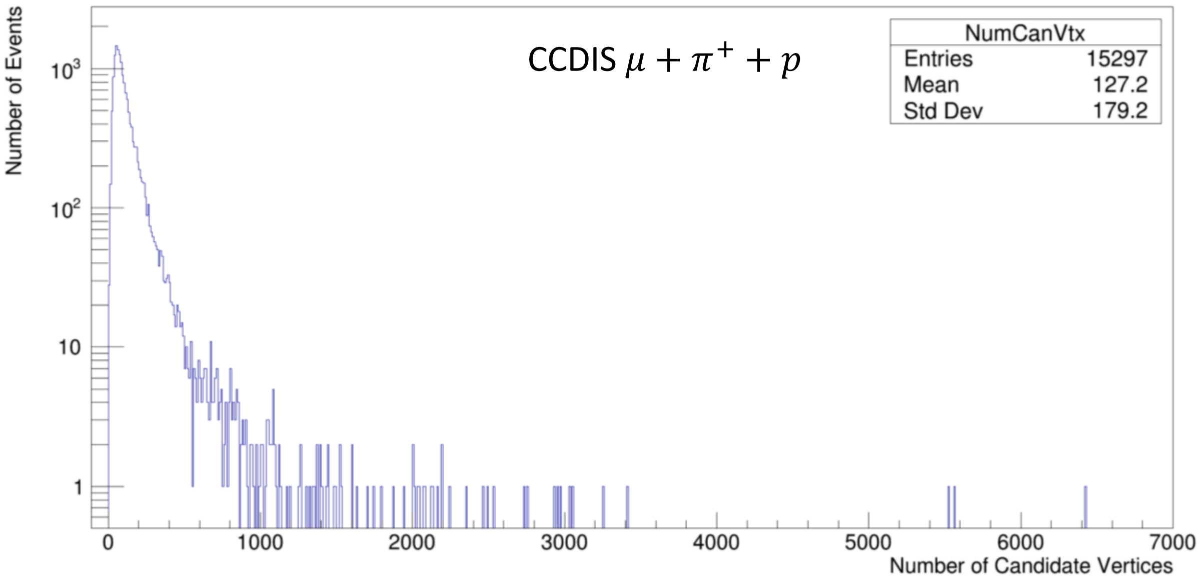
CandidateVertexCreationAlgorithm

- Currently clear there are too many candidates for events in DUNE FD.

  - The algorithm can give four candidates from every pairing of 2D clusters between two views, and some of these clusters (at this stage in the reconstruction) will be very small, especially in shower-like regions.

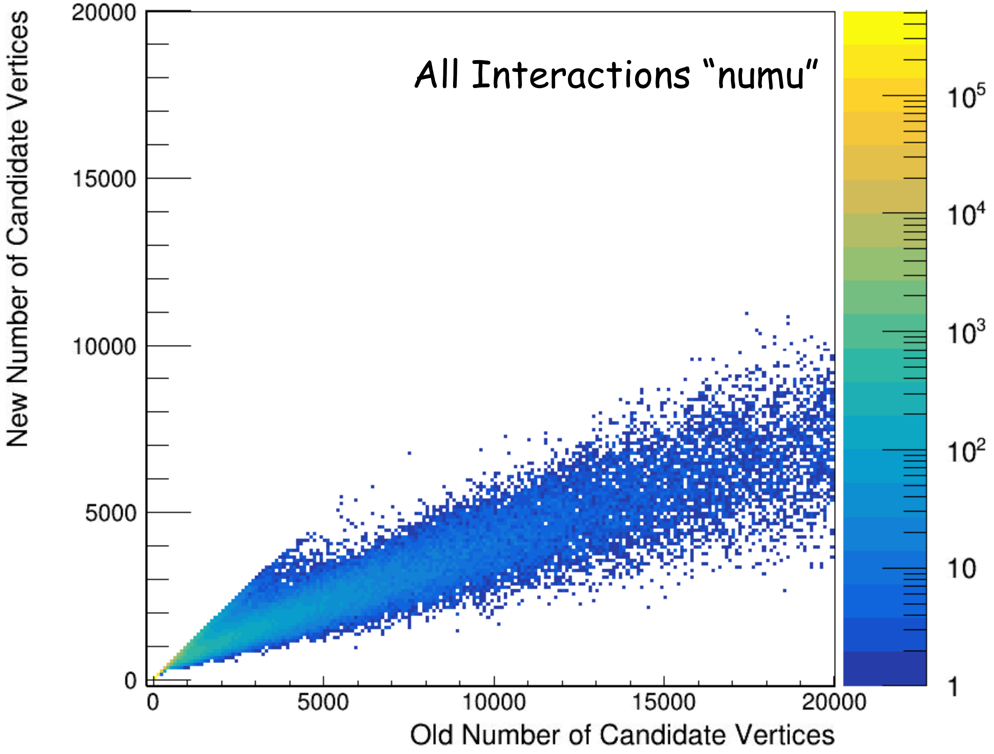


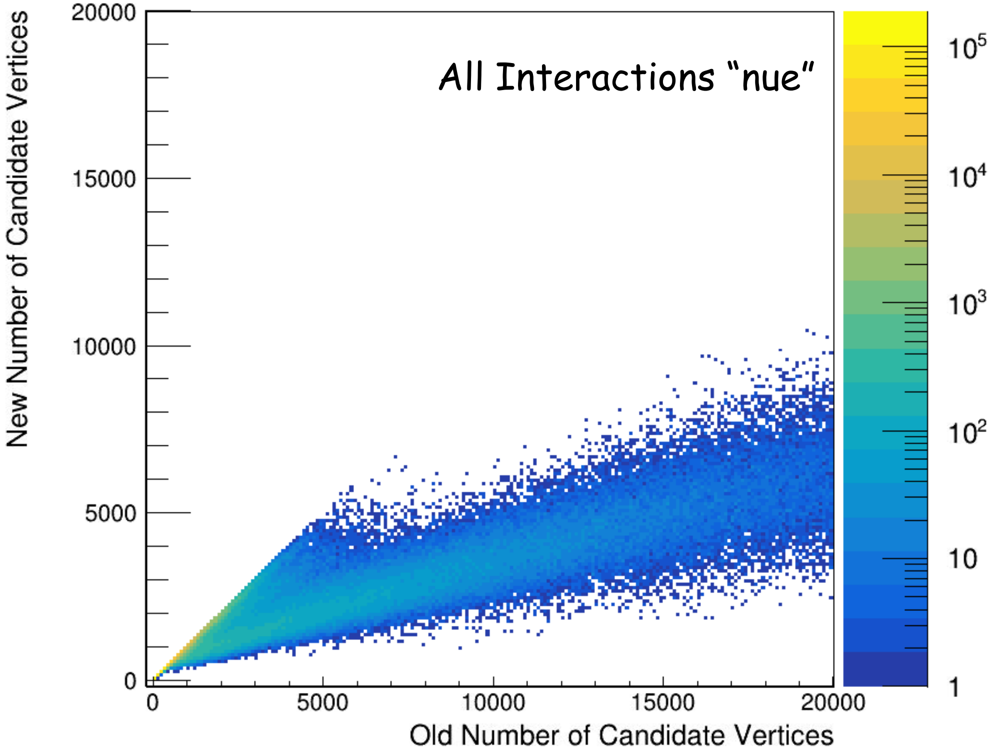




Candidate Vertex Creation Algorithm Modifications

- The Candidate Vertex Creation Algorithm was modified so that events which had a large number of candidates will now have a reduced number created in exceptionally showery regions.
- This reduces the mean number of candidates per event without having any significant impact on the performance metrics.
- Further work will follow to continue to reduce the number of created candidates.





Conclusions/Summary

- Vertexing in Pandora has been tuned for DUNE FD events for the very first time, using a BDT.
- This new developed BDT vertexing algorithm provides a significant benefit to the quality of the vertex reconstruction for DUNE FD events compared to the current method (especially for nue).

All Interactions "numu"	Current Vertexing	BDT Vertexing
dR68%	1.75cm	1.39cm
%>50cm	8.29%	6.54%

All Interactions "nue"	Current Vertexing	BDT Vertexing
dR68%	2.22cm	1.35cm
%>50cm	4.76%	2.59%

- This will be officially released soon. Feature branches are ready and are feature/larpandoracontent_v03_15_00 in dunetpc and larpandoracontent. The training output file is in dune_pardata.
- Plans to continue developing algorithms to improve the vertex reconstruction incorporating aspects of Deep Learning.
- Working on efficiency improvements for vertex reconstruction, also on to do list.

The End

Thank you for your attention!

Pandora Team

Pandora is an open project and new contributors would be extremely welcome.
We'd love to hear from you and we will always try to answer your questions.

Pandora SDK Development	John Marshall (John.Marshall@warwick.ac.uk) Mark Thomson (thomson@hep.phy.cam.ac.uk)
LAr TPC algorithm development	John Marshall (John.Marshall@warwick.ac.uk) Andy Blake (a.blake@lancaster.ac.uk)
DUNE FD Integration	Lorena Escudero (escudero@hep.phy.cam.ac.uk)
ProtoDUNE Integration	Steven Green (sg568@hep.phy.cam.ac.uk)
MicroBooNE Integration	Andy Smith (asmith@hep.phy.cam.ac.uk)
Graduate Students	MicroBooNE : Joris Jan de Vries, Jack Anthony ProtoDUNE : Stefano Vergani DUNE : Jhanzeb Ahmed, Mousam Rai, Ryan Cross

https://github.com/PandoraPFA

https://pandorapfa.slack.com

UNIVERSITY OF CAMBRIDGE

Lancaster University

WARWICK
THE UNIVERSITY OF WARWICK

Backup Slides

Candidate Reduction Plots

Histograms of number of candidate vertices for each event for MCC11 “numu” and “nue” events. Blue is old Pandora. Red is from the new Pandora. The mean for the blue distribution is ~644 and for the red distribution is ~348 for “numu”. The mean for the blue distribution is ~1999 and for the red distribution is ~968 for “nue”.