PROCEEDINGS OF SCIENCE

A Novel Hit-Based Method to Distinguish Tracks and Showers in ProtoDUNE Single Phase

Stefano Vergani[a,b,*]

[a] *for the DUNE Collaboration*
[b] *University of Cambridge, Cambridge CB3 0HE, United Kingdom*

E-mail: sv408@hep.phy.cam.ac.uk

Pandora [1, 2] is a pattern recognition software used in liquid argon time projection chamber (LArTPC) experiments such as MicroBooNE, DUNE, SBND, ICARUS, and ProtoDUNE Single Phase (SP). The output of a LArTPC can be considered a high-resolution 2D image and energy depositions, called hits, from particles in a LArTPC create complicated topologies that are broadly classified into tracks and showers. The event reconstruction is particularly challenging when there are multiple overlapping particles and in order to fully harness the imaging capabilities of those experiments, Pandora needs to separate them. A hit-based approach to this problem is presented, which analyses small regions around each hit in events from DUNE Far Detector (FD) and from those regions it calculates local variables that are used subsequently in a machine learning approach. After this stage, it is given to each hit a probability to belong to a track or shower-like particle. Results will show the performance of separation between tracks and showers. This method is planned to be used for ProtoDUNE SP.

**** The 22nd International Workshop on Neutrinos from Accelerators (NuFact2021) ****
**** 6–11 Sep 2021 ****
**** Cagliari, Italy ****

[*] Speaker

PoS(NuFact2021)179

1. Track and Shower Topologies

The goal of this tool is to identify and separate tracks inside dense topological shower regions, as shown in Figure 1. Although this is a relatively easy task for the human visual system, it is challenging to write automated software able to do that. The approach we have taken consists of finding local properties of small patches from the whole event and trying to deduce global properties of the event.

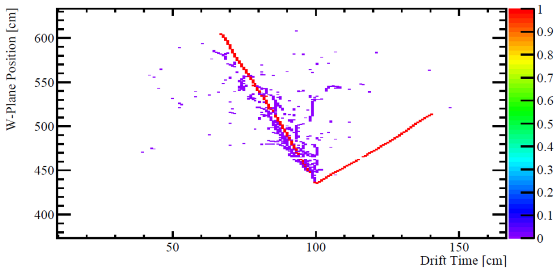


Figure 1: A track-like particle (in red) passing through a dense shower region (in purple). This event characterisation has been made using MC information from a DUNE FD simulated event. The results shown are preliminary.

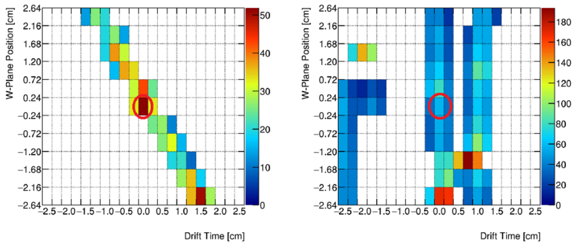
2. The Grid Method

A histogram of 11x21 bins is created around each single hit. Every bin measures 4.8 mm (the wires' pitch in ProtoDUNE-SP [3]) in the Y direction and 2.5 mm in the X direction. The width of the bin and this particular bin configuration have been chosen to be as described before because we could obtain a good resolution, a proportionality between Y and X, and the patch X and Y dimensions are approximately the same (5.25x5.28cm^2). The chosen hit lies in the central bin in such a way that the its relative position in the grid is (0,0) and the coordinates of the other hits are shifted accordingly. The integrated charge deposited by every hit is smeared according to a Gaussian distribution and the final result is shown in Figure 2.

3. Finding Variables

The next step in the process requires to find good variables which characterise each grid to be passed to a Machine Learning (ML) model. In order to do that, Principal Component Analysis (PCA) is used. PCA is an analytical technique commonly used to transform a set of possibly correlated variables into an smaller set of linearly independent variables. In this specific problem, PCA helps us to find the major axis where the variance of the position of the hits is maximally spread. Once the PCA is found for each grid, several variables are evaluated:

- **Average Distance From PCA Major Axis**: calculated for each bin and then averaged. For track-like topologies, this tends to be small. Figure 3 shows examples of grids with low (left) and high (right) values, respectively.



(a) (b)

Figure 2: Two examples of grid are shown. The bin in the middle inside the red circle is the one where the chosen hit lies, different colors represent different intensities of the deposited integrated charged. Figure 2a shows a clear track, whilst 1b is an example of grid in which the topology is not so obvious. The results shown are preliminary.

- **Energy Crossed by PCA Major Axis**: the sum of the energies of every bin crossed by the PCA major axis divided by the total energy deposited in the grid. This value tends to be big for track-like topologies.
- **Deviation Method**: each grid is rotated so that the PCA major axis is parallel to the y-axis. The new grid is then multiplied by a matrix and the result is a single number between -1 and 1. The closer this number is to 1, the more likely it is a track.

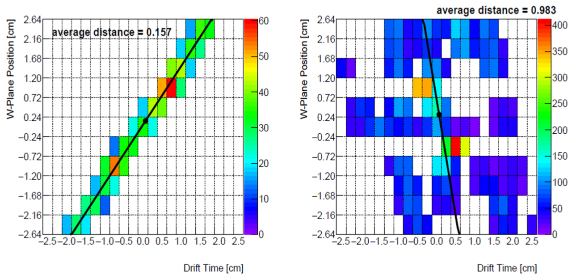


Figure 3: Examples of two different topologies with the calculated average distance from PCA Major Axis, Figure 3a shows a track and 3b shows a shower. The results shown are preliminary.

4. Results and Future Work

The Toolkit for Multivariate Data Analysis with ROOT (TMVA) provides several multivariate classification techniques which have been used for this analysis. Applying the variables calculated in Section 3, simulated 3000 DUNE FD events were used to train the TMVA Boosted Decision Tree model and then subsequently the results have been applied to another set 2800 events chosen differently to avoid over training. A signal efficiency of 78% and background rejection of 92% have been obtained as shown in Figure 4. Results show that separation between tracks and showers in complicated topological regions is possible as in the example in Figure 5. This method will be soon used with ProtoDUNE SP simulation as well as data.

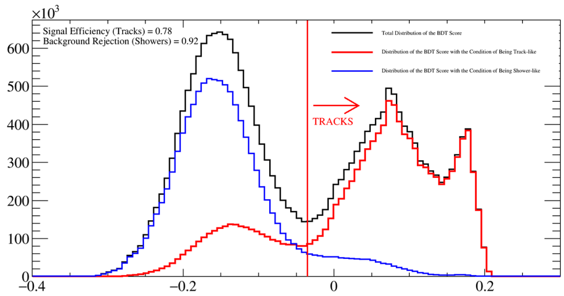


Figure 4: Distribution of the bdt score (in black). In red and in blue the same distribution with the condition of the being track or shower-like. This plot has been calculated from 3244383 grids coming from 3000 DUNE-FD simulated events. The results shown are preliminary.

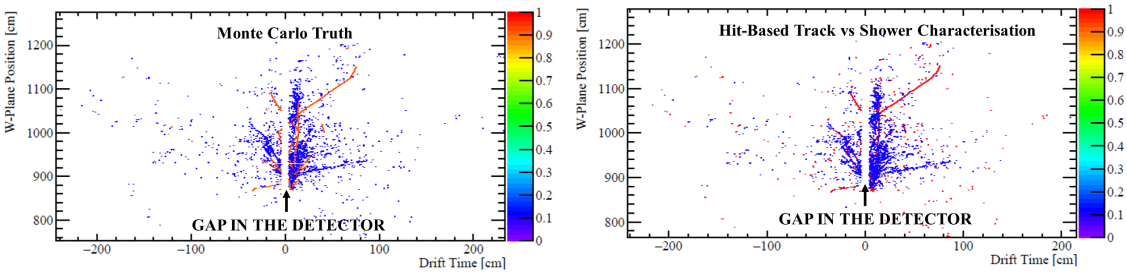


Figure 5: Left shows event from DUNE FD characterised in blue for shower like particles and in red for track like particles with MC true values. Right shows the same event characterised using the algorithm. The gap is a physical gap in the detector. The results shown are preliminary.

References

[1] MicroBooNE collaboration, *The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector*, Eur. Phys. J. C **78** (2018) 82 [1708.03135].

[2] J.S. Marshall and M.A. Thomson, *The pandora software development kit for pattern recognition*, *The European Physical Journal C* **75** (2015) .

[3] B. Abi, R. Acciarri, M.A. Acero, M. Adamowski, C. Adams, D.L. Adams et al., *The single-phase protodune technical design report*, 2017.