

# Calorimetric classification of track-like signatures in liquid argon TPCs using MicroBooNE data

---

## MicroBooNE Collaboration

P. Abratenko<sup>jj</sup> R. An<sup>o</sup> J. Anthony<sup>d</sup> J. Asadi<sup>ii</sup> A. Ashkenazi<sup>t,gg</sup>  
S. Balasubramanian<sup>mm,l</sup> B. Baller<sup>l</sup> C. Barnes<sup>u</sup> G. Barr<sup>y</sup> V. Basque<sup>s</sup> L. Bathe-Peters<sup>n</sup>  
O. Benevides Rodrigues<sup>ff</sup> S. Berkman<sup>l</sup> A. Bhandari<sup>s</sup> A. Bhat<sup>ff</sup> M. Bishai<sup>b</sup> A. Blake<sup>q</sup>  
T. Bolton<sup>p</sup> L. Camilleri<sup>j</sup> D. Caratelli<sup>l</sup> I. Caro Terrazas<sup>i</sup> R. Castillo Fernandez<sup>l</sup>  
F. Cavanna<sup>l</sup> G. Cerati<sup>l</sup> Y. Chen<sup>a</sup> E. Church<sup>z</sup> D. Cianci<sup>j</sup> J. M. Conrad<sup>t</sup> M. Convery<sup>cc</sup>  
L. Cooper-Troendle<sup>mm</sup> J. I. Crespo-Anadón<sup>f</sup> M. Del Tutto<sup>l</sup> S. R. Dennis<sup>d</sup> D. Devitt<sup>q</sup>  
R. Diurba<sup>v</sup> R. Dorrill<sup>o</sup> K. Duffy<sup>l</sup> S. Dytman<sup>aa</sup> B. Eberly<sup>ee</sup> A. Ereditato<sup>a</sup> J. J. Evans<sup>s</sup>  
R. Fine<sup>r</sup> G. A. Fiorentini Aguirre<sup>dd</sup> R. S. Fitzpatrick<sup>u</sup> B. T. Fleming<sup>mm</sup> N. Foppiani<sup>n</sup>  
D. Franco<sup>mm</sup> A. P. Furmanski<sup>v</sup> D. Garcia-Gamez<sup>m</sup> S. Gardiner<sup>l</sup> G. Ge<sup>j</sup>  
S. Gollapinni<sup>hh,r</sup> O. Goodwin<sup>s</sup> E. Gramellini<sup>l</sup> P. Green<sup>s</sup> H. Greenlee<sup>l</sup> W. Gu<sup>b</sup>  
R. Guenette<sup>n</sup> P. Guzowski<sup>s</sup> L. Hagaman<sup>mm</sup> E. Hall<sup>t</sup> P. Hamilton<sup>ff</sup> O. Hen<sup>t</sup>  
G. A. Horton-Smith<sup>p</sup> A. Hourlier<sup>t</sup> R. Itay<sup>cc</sup> C. James<sup>l</sup> X. Ji<sup>b</sup> L. Jiang<sup>kk</sup> J. H. Jo<sup>mm</sup>  
R. A. Johnson<sup>h</sup> Y.-J. Jwa<sup>j</sup> N. Kamp<sup>t</sup> N. Kaneshige<sup>c</sup> G. Karagiorgi<sup>j</sup> W. Ketchum<sup>l</sup>  
M. Kirby<sup>l</sup> T. Kobilarcik<sup>l</sup> I. Kreslo<sup>a</sup> R. LaZur<sup>i</sup> I. Lepetic<sup>bb</sup> K. Li<sup>mm</sup> Y. Li<sup>b</sup> K. Lin<sup>r</sup>  
B. R. Littlejohn<sup>o</sup> W. C. Louis<sup>r</sup> X. Luo<sup>c</sup> K. Manivannan<sup>ff</sup> C. Mariani<sup>kk</sup> D. Marsden<sup>s</sup>  
J. Marshall<sup>ll</sup> D. A. Martinez Caicedo<sup>dd</sup> K. Mason<sup>jj</sup> A. Mastbaum<sup>bb</sup> N. McConkey<sup>s</sup>  
V. Meddage<sup>p</sup> T. Mettler<sup>a</sup> K. Miller<sup>g</sup> J. Mills<sup>jj</sup> K. Mistry<sup>s</sup> A. Mogan<sup>hh</sup> T. Mohayai<sup>l</sup>  
J. Moon<sup>t</sup> M. Mooney<sup>i</sup> A. F. Moor<sup>d</sup> C. D. Moore<sup>l</sup> L. Mora Lepin<sup>s</sup> J. Mousseau<sup>u</sup>  
M. Murphy<sup>kk</sup> D. Naples<sup>aa</sup> A. Navrer-Agasson<sup>s</sup> R. K. Neely<sup>p</sup> J. Nowak<sup>q</sup> M. Nunes<sup>ff</sup>  
O. Palamara<sup>l</sup> V. Paolone<sup>aa</sup> A. Papadopoulou<sup>t</sup> V. Papavassiliou<sup>w</sup> S. F. Pate<sup>w</sup>  
A. Paudel<sup>p</sup> Z. Pavlovic<sup>l</sup> E. Piasetzky<sup>gg</sup> I. D. Ponce-Pinto<sup>j,mm</sup> S. Prince<sup>n</sup> X. Qian<sup>b</sup>  
J. L. Raaf<sup>l</sup> V. Radeka<sup>b</sup> A. Rafique<sup>p</sup> M. Reggiani-Guzzo<sup>s</sup> L. Ren<sup>w</sup> L. C. J. Rice<sup>aa</sup>  
L. Rochester<sup>cc</sup> J. Rodriguez Rondon<sup>dd</sup> H.E. Rogers<sup>e</sup> M. Rosenberg<sup>aa</sup>  
M. Ross-Lonergan<sup>j</sup> G. Scanavini<sup>mm</sup> D. W. Schmitz<sup>g</sup> A. Schukraft<sup>l</sup> W. Seligman<sup>j</sup>  
M. H. Shaevitz<sup>j</sup> R. Sharankova<sup>jj</sup> J. Sinclair<sup>a</sup> A. Smith<sup>d</sup> E. L. Snider<sup>l</sup> M. Soderberg<sup>ff</sup>  
S. Söldner-Rembold<sup>s</sup> P. Spentzouris<sup>l</sup> J. Spitz<sup>u</sup> M. Stancari<sup>l</sup> J. St. John<sup>l</sup> T. Strauss<sup>l</sup>  
K. Sutton<sup>j</sup> S. Sword-Fehlberg<sup>w</sup> A. M. Szelc<sup>s,k</sup> N. Tagg<sup>x</sup> W. Tang<sup>hh</sup> K. Terao<sup>cc</sup>  
C. Thorpe<sup>q</sup> D. Totani<sup>c</sup> M. Touns<sup>l</sup> Y.-T. Tsai<sup>cc</sup> M. A. Uchida<sup>d</sup> T. Usher<sup>cc</sup>  
W. Van De Pontseele<sup>y,n</sup> B. Viren<sup>b</sup> M. Weber<sup>a</sup> H. Wei<sup>b</sup> Z. Williams<sup>ii</sup> S. Wolbers<sup>l</sup>  
T. Wongjirad<sup>jj</sup> M. Wospakrik<sup>l</sup> N. Wright<sup>t</sup> W. Wu<sup>l</sup> E. Yandel<sup>c</sup> T. Yang<sup>l</sup>  
G. Yarbrough<sup>hh</sup> L. E. Yates<sup>t</sup> G. P. Zeller<sup>l</sup> J. Zennamo<sup>l</sup> C. Zhang<sup>b</sup>

<sup>a</sup> Universität Bern, Bern CH-3012, Switzerland

<sup>b</sup> Brookhaven National Laboratory (BNL), Upton, NY, 11973, USA

<sup>c</sup> *University of California, Santa Barbara, CA, 93106, USA*  
<sup>d</sup> *University of Cambridge, Cambridge CB3 0HE, United Kingdom*  
<sup>e</sup> *St. Catherine University, Saint Paul, MN 55105, USA*  
<sup>f</sup> *Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid E-28040, Spain*  
<sup>g</sup> *University of Chicago, Chicago, IL, 60637, USA*  
<sup>h</sup> *University of Cincinnati, Cincinnati, OH, 45221, USA*  
<sup>i</sup> *Colorado State University, Fort Collins, CO, 80523, USA*  
<sup>j</sup> *Columbia University, New York, NY, 10027, USA*  
<sup>k</sup> *University of Edinburgh, Edinburgh EH9 3FD, United Kingdom*  
<sup>l</sup> *Fermi National Accelerator Laboratory (FNAL), Batavia, IL 60510, USA*  
<sup>m</sup> *Universidad de Granada, E-18071, Granada, Spain*  
<sup>n</sup> *Harvard University, Cambridge, MA 02138, USA*  
<sup>o</sup> *Illinois Institute of Technology (IIT), Chicago, IL 60616, USA*  
<sup>p</sup> *Kansas State University (KSU), Manhattan, KS, 66506, USA*  
<sup>q</sup> *Lancaster University, Lancaster LA1 4YW, United Kingdom*  
<sup>r</sup> *Los Alamos National Laboratory (LANL), Los Alamos, NM, 87545, USA*  
<sup>s</sup> *The University of Manchester, Manchester M13 9PL, United Kingdom*  
<sup>t</sup> *Massachusetts Institute of Technology (MIT), Cambridge, MA, 02139, USA*  
<sup>u</sup> *University of Michigan, Ann Arbor, MI, 48109, USA*  
<sup>v</sup> *University of Minnesota, Minneapolis, Mn, 55455, USA*  
<sup>w</sup> *New Mexico State University (NMSU), Las Cruces, NM, 88003, USA*  
<sup>x</sup> *Otterbein University, Westerville, OH, 43081, USA*  
<sup>y</sup> *University of Oxford, Oxford OX1 3RH, United Kingdom*  
<sup>z</sup> *Pacific Northwest National Laboratory (PNNL), Richland, WA, 99352, USA*  
<sup>aa</sup> *University of Pittsburgh, Pittsburgh, PA, 15260, USA*  
<sup>bb</sup> *Rutgers University, Piscataway, NJ, 08854, USA, PA*  
<sup>cc</sup> *SLAC National Accelerator Laboratory, Menlo Park, CA, 94025, USA*  
<sup>dd</sup> *South Dakota School of Mines and Technology (SDSMT), Rapid City, SD, 57701, USA*  
<sup>ee</sup> *University of Southern Maine, Portland, ME, 04104, USA*  
<sup>ff</sup> *Syracuse University, Syracuse, NY, 13244, USA*  
<sup>gg</sup> *Tel Aviv University, Tel Aviv, Israel, 69978*  
<sup>hh</sup> *University of Tennessee, Knoxville, TN, 37996, USA*  
<sup>ii</sup> *University of Texas, Arlington, TX, 76019, USA*  
<sup>jj</sup> *Tufts University, Medford, MA, 02155, USA*  
<sup>kk</sup> *Center for Neutrino Physics, Virginia Tech, Blacksburg, VA, 24061, USA*  
<sup>ll</sup> *University of Warwick, Coventry CV4 7AL, United Kingdom*  
<sup>mm</sup> *Wright Laboratory, Department of Physics, Yale University, New Haven, CT, 06520, USA*

E-mail: [microboone\\_info@fnal.gov](mailto:microboone_info@fnal.gov)

ABSTRACT: The MicroBooNE liquid argon time projection chamber located at Fermilab is a neutrino experiment dedicated to the study of short-baseline oscillations, the measurements of neutrino cross sections in liquid argon, and to the research and development of this novel detector technology. Accurate and precise measurements of calorimetry are essential to the event reconstruction and are achieved by leveraging the TPC to measure deposited energy per unit length along the particle trajectory, with mm resolution. We describe the non-uniform calorimetric reconstruction performance in the detector, showing dependence on the angle of the particle trajectory. Such non-uniform reconstruction directly affects the performance of the particle identification algorithms which infer particle type from calorimetric measurements. This work presents a new particle identification method which accounts for and effectively addresses such non-uniformity. The newly developed method shows improved performance compared to previous algorithms, illustrated by a 93.7% proton selection efficiency and a 10% muon mis-identification rate, with a fairly loose selection of tracks performed on beam data. The performance is further demonstrated by identifying exclusive final states in  $\nu_\mu CC$  interactions. While developed using MicroBooNE data and simulation, this method is easily applicable to future LArTPC experiments, such as SBND, ICARUS, and DUNE.

---

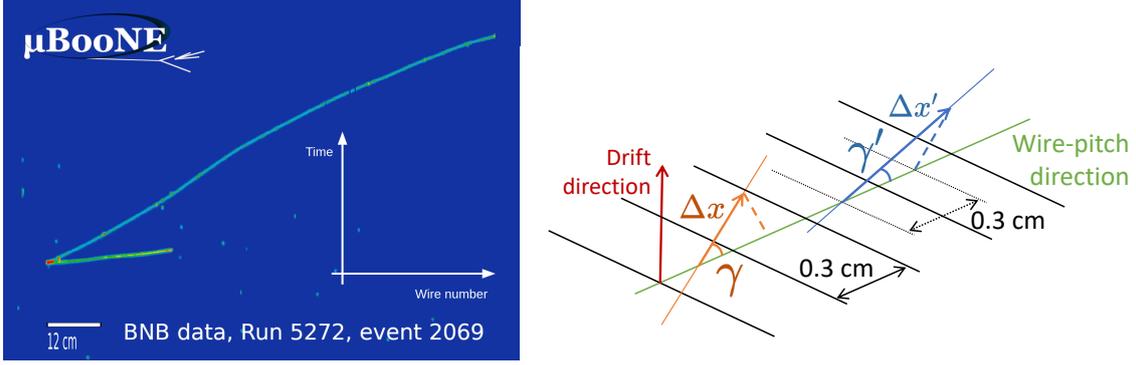
## Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>  | <b>1</b>  |
| <b>2</b> | <b>The MicroBooNE detector and the calorimetric reconstruction of tracks</b>       | <b>3</b>  |
| <b>3</b> | <b>Angular effects in calorimetric reconstruction</b>                              | <b>4</b>  |
| <b>4</b> | <b>Energy deposition profile and particle identification</b>                       | <b>6</b>  |
| <b>5</b> | <b>The likelihood-based method for particle identification</b>                     | <b>6</b>  |
| 5.1      | The $dE/dx$ probability density function   | 6         |
| 5.2      | The likelihood ratio test statistic as PID score                                   | 7         |
| 5.3      | Performance of the particle identification   | 9         |
| <b>6</b> | <b>Applications to physics cases</b>   | <b>10</b> |
| 6.1      | Proton-muon separation for tracks recorded on data                                 | 10        |
| 6.2      | Large collection-plane-local pitch tracks identified with the two induction planes | 12        |
| 6.3      | Exclusive $\nu_\mu$ selection  | 12        |
| <b>7</b> | <b>Conclusions</b>   | <b>15</b> |

---

## 1 Introduction

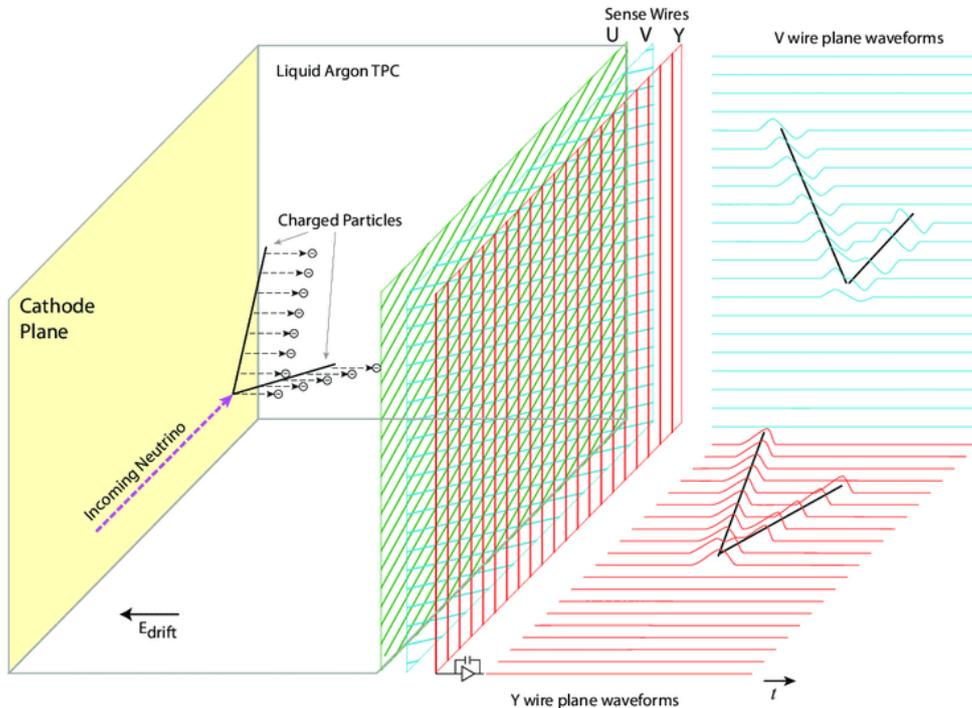
Liquid argon time projection chambers (LArTPCs) are powerful neutrino detectors which enable the study of topological and calorimetric signatures of particles produced in neutrino interactions with millimeter spatial resolution [6, 10, 12]. These particles are reconstructed as track-like (or simply tracks) if they deposit energy predominantly through ionization, or shower-like (or simply showers), if the main energy loss mechanism is bremsstrahlung and pair production. Tracks are primarily associated with the reconstruction of muons, protons, and pions, while showers are associated with electrons and photons. At typical MicroBooNE energies, between hundreds of MeV to few GeV, hadrons do not produce hadronic showers. However, thanks to the high resolution, if a LArTPC like MicroBooNE were to be operated at larger energies, the different particles produced in hadronic showers could be reconstructed individually. Energy lost by final state charged particles results in the ionization of argon atoms. Trails of ionization electrons are detected by multiple wire planes to provide a 3D image of the particles' propagation through the detector. Particle identification (PID), determination of the type of particle given its calorimetric measurement, is performed by studying the profiles of each ionization electron trail. The left panel of figure 1 shows the collection plane projection of a muon neutrino charged



**Figure 1:** Left: Example of a raw image of a  $\nu_\mu CC$  interaction in MicroBooNE with two tracks in the final state, as recorded on the collection plane. The deposited charge (color scale) is shown as a function of wire number (x-axis) and time (y-axis). The ionization profiles of the two particles are used to identify them as one proton and one muon. Right: A sketch of the relevant directions and angle in the calorimetric reconstruction. The orange and blue arrows represent two possible particle trajectories with different gamma angles and different local pitch values, represented by their lengths. Black solid lines represent wires, spaced 0.3 cm apart. The dashed lines are perpendicular to the wire-pitch direction, and make evident the connection between the angle  $\gamma$  and  $\Delta x$ , through eq. (2.1).

current ( $\nu_\mu CC$ ) interaction, in which two track-like particles are produced. These tracks are classified as one proton and one muon, the two most common track-like particles in MicroBooNE, differentiated by the different amounts of energy deposited per unit length at any given point in their trajectories.

After a brief discussion of the MicroBooNE detector and reconstruction in section 2, in section 3 we illustrate the angular dependence of LArTPCs' calorimetric reconstruction. In section 4, a review of the particle identification principles is presented. In section 5 we address the central topic of the paper: a new method to perform particle identification that more easily accounts for angular dependencies in calorimetric reconstruction. The performance of this method is discussed in section 6, where the identification of different muon neutrino interaction final states is presented.



**Figure 2:** In this schematic representation of the MicroBooNE LArTPC (see text for details), a neutrino interacts in the detector, producing two charged particles in the final state. The ionization is drifted towards the anode, and induces waveforms on the wire planes, which are displayed for the V and Y planes. Figure taken from [3].

## 2 The MicroBooNE detector and the calorimetric reconstruction of tracks

The MicroBooNE detector is an 85 ton active mass liquid argon time projection chamber (LArTPC) [3]. The drift direction ( $x$ ), vertical direction ( $y$ ), and the beam direction ( $z$ ) measure 2.56 m, 2.33 m, and 10.36 m, respectively. The nominal electric field inside the TPC is 273.9 V/cm resulting in an electron drift velocity of 0.11 cm/ $\mu$ s. The drifted ionization charge from particle interactions is read out by three wire planes spaced 0.3 cm apart, with a 0.3 cm wire spacing, oriented vertically for the collection plane (Y plane), at +60 deg for the first induction planes (U plane), and at -60 deg for the second induction plane (V plane). The digitized charge on the readout wires is noise-filtered [4] and deconvolved [7, 8]. The high signal-to-noise ratio of MicroBooNE's TPC and cold electronics allow for the accurate measurement of charge on all three wire-planes, an essential ingredient in reliable particle identification. The data is represented as a set of three two-dimensional images (one for each wire plane), with wire number plotted along the horizontal axis and drift time plotted along the vertical axis. Each provides a two-dimensional projection of the charge deposited

in the event. An example is given in the left panel of figure 1.

The position and amount of charge deposited is characterized with the hit-finding process: charge depositions on a given wire at a given time, called hits, are extracted through a Gaussian fit to the waveforms [13]. Next, the Pandora multi-algorithm pattern recognition framework [5] groups nearby hits into clusters. Clusters on the three planes are subsequently matched to reconstruct three dimensional particles. Each cluster is a projection on a given plane of the charge deposited by a particle. The deposited charge is corrected for detector effects to provide a spatially uniform response. For each hit in a cluster the charge deposited is converted to deposited energy, providing a local measurement of the energy  $\Delta E$  along a three-dimensional distance  $\Delta x$  extracted from the reconstructed trajectory [9]. The conversion is performed with a multiplicative factor specific to the plane, and correcting for the recombination of electron and ions [2, 9]. The three-dimensional distance  $\Delta x$  is called *local pitch*, and is computed as

$$\Delta x = 0.3 \text{ cm} / \cos(\gamma), \quad (2.1)$$

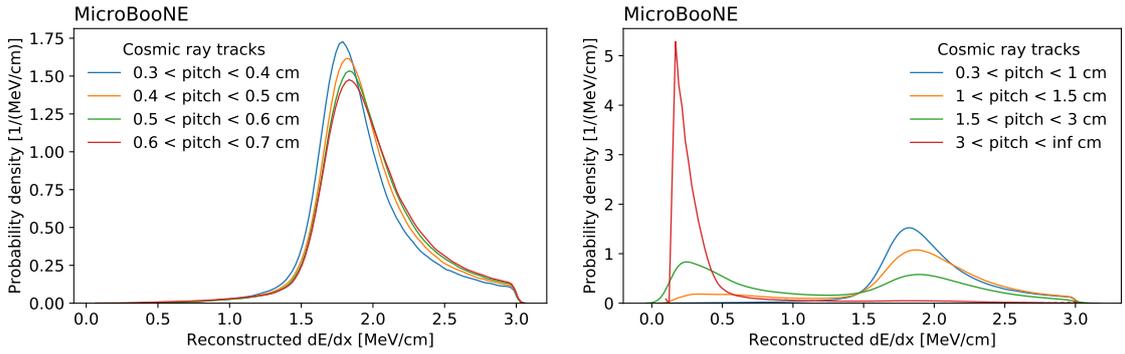
where 0.3 cm is the wire spacing and  $\gamma$  is the three-dimensional angle between the local direction of the track and the vector that connects adjacent wires (also called wire-pitch direction), as illustrated in the right panel of figure 1. The angle  $\gamma$  ranges between 0 and 90 degrees, while  $\Delta x$  takes values between 0.3 cm and infinity. Combined with the measurement of  $\Delta E$ , the ionization density  $dE/dx$  can be estimated for every hit.

### 3 Angular effects in calorimetric reconstruction

When measuring calorimetric information in a LArTPC, the fact that charge is drifted along a particular direction (drift direction) and projected on wire planes with different orientations makes the calorimetric reconstruction angle-dependent. Both  $dE/dx$  and the precision with which it is measured depend on the direction of the ionization trace left by the particle, even in a "perfect detector", absent of detector effects and angle-dependent detector response non-uniformity. The dependence appears primarily through the angle  $\gamma$  illustrated in the bottom plot of figure 1. Because the angle  $\gamma$  relates directly to the local pitch through a bijective function (eq. (2.1)),  $\gamma$  and local pitch can be used interchangeably, and local pitch will be used in the rest of the article.

Measured  $dE/dx$ , even with a perfect detector, is angle-dependent because of intrinsic statistical fluctuations in particle energy loss, which impact the probability density function of measurements when averaged over different travel distances (different local pitch). A  $dE/dx$  distribution, typically described by a Landau function for small local pitch values [14], becomes narrower at larger local pitch and its most probable value moves to higher  $dE/dx$  values while its average remains constant. This general geometrical effect applies to all LArTPCs and is shown on figure 33.8 in [16].

The precision of  $dE/dx$  measurements also depends on the local pitch since the shape of the signals induced on the wires by the drifting charge appears very different at lower local pitch ( $< 0.7$  cm) compared to larger local pitch ( $> 0.7$  cm) [7], impacting hit reconstruction and making measurements more precise at lower local pitch. Figure 3 illustrates these



**Figure 3:** Normalized distributions of  $dE/dx$  for different local pitch in the low-local pitch regime (left) and over the entire local pitch spectrum (right), as measured on the collection plane in a sample of cosmic muons tracks in the MicroBooNE data.

effects as measured with cosmic ray tracks in MicroBooNE’s data. By requiring the tracks to cross the detector, this selection results in a clean set of relativistic muons, which are minimum ionizing particles, with a constant average  $dE/dx$ . The left plot shows the shape of the  $dE/dx$  distribution for different small local pitch values, where the peak of the distributions shifts towards larger  $dE/dx$  values for larger local pitch, varying by 4% in the low local pitch range (between 0.3 cm and 0.7 cm). The width of the distributions increases at larger local pitch because the finite precision introduced by the detector smears out the distribution more than the predicted shrinking induced by geometrical effects. The right plot shows analogous distributions integrated over a wider range of local pitch values, illustrating a major change in the shape of the distributions at very large local pitch. The distributions for larger local pitch values show a second peak at smaller  $dE/dx$  values, mainly populated by particles traveling parallel to the drift direction. These tracks induce long signals in time, for which the Gaussian fit performed by the hit-finding process is not sufficient. Therefore, such signals are fit by a sum of Gaussian shapes, for which the overall number of hits, their positions, and their amplitudes are free parameters. The width is fixed to reduce the degeneracy of the problem. The total deposited charge is therefore segmented into multiple hits, leading to an underestimation of the hit charge and resulting in smaller  $dE/dx$  values. These small and non-physical values of  $dE/dx$  encountered at large local pitch are not correlated with the true  $dE/dx$ , bringing no additional information and reducing the particle identification performance.

Only considering the dependence on the angle  $\gamma$ , which is the polar angle with respect to the wire-pitch direction, is an approximation, as it encapsulates most, but not all the angular dependence. A more complete analysis will consider the additional dependence on the relative azimuthal angle, that, together with  $\gamma$ , uniquely describes the 3D trajectory. Nonetheless, this approximation captures most of the angular dependence, significantly improving particle identification performance.

## 4 Energy deposition profile and particle identification

Typical particle identification methods condense calorimetric information into a score used to distinguish different particle species. The score is typically obtained by starting from the measured  $dE/dx$  profile as a function of residual range - the distance of a given energy deposition within the track from the endpoint of the track itself. This profile is compared with the expectation for different particle hypotheses to choose the hypothesis that best matches the data. The hypothesized  $dE/dx$  profile is computed by integrating the Bethe-Bloch function for a given particle mass and charge. In performing this comparison, the intrinsic statistical nature of  $dE/dx$  must be accounted for, as do the angular dependencies (described in the previous section) which also affect the shape of the  $dE/dx$  distribution. As local pitch values are different on different readout planes, combining the three wire plane measurements ensures that the calorimetric information provided by the LArTPC in the entire  $4\pi$  solid angle is fully leveraged for the purpose of performing particle identification.

## 5 The likelihood-based method for particle identification

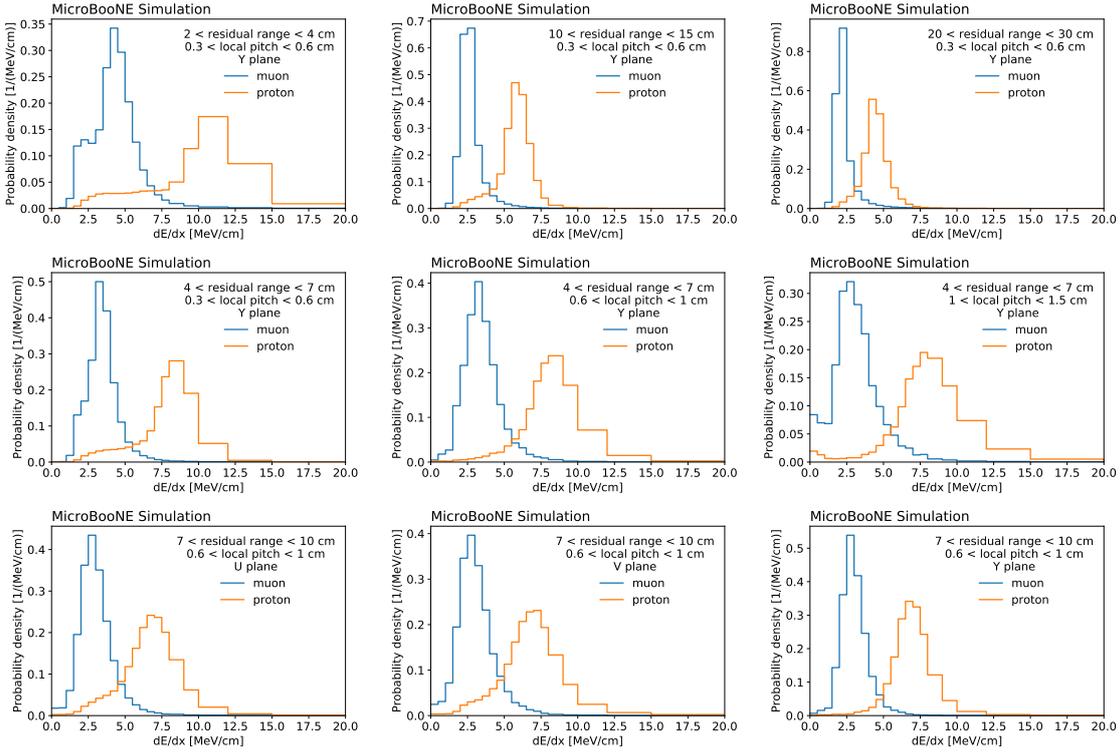
The newly-developed method of particle identification accounts for angular dependencies in calorimetric reconstruction and compensates for distortions at large local pitch by efficiently combining the three wire plane measurements.

This method computes the likelihood for different particle hypotheses given the experimentally measured  $dE/dx$  profile. The likelihood is based on a model which is computed through an accurate description of the expected  $dE/dx$  probability density function.

### 5.1 The $dE/dx$ probability density function

The  $dE/dx$  probability density function (PDF) for each particle type is the basic ingredient of the likelihood calculation. In principle, the average  $dE/dx$  as a function of the residual range could be estimated by integrating the Bethe-Bloch function before applying detector reconstruction. However, a complete characterization of the  $dE/dx$  distribution requires an analytic description of the intrinsic fluctuations of the ionization energy loss and effect of the detector reconstruction. This is challenging given the very long computational time required and because there is no straightforward way to derive an analytic description of the detector reconstruction. The  $dE/dx$  PDF is instead estimated from the MicroBooNE simulation which incorporates all the described effects, as demonstrated in the data/simulation comparisons in section 6.

The PDF is estimated through a three dimensional histogram of  $dE/dx$ , residual range, and local pitch. The histogram is normalized so that for each combination of values of residual range and local pitch, the integral of the  $dE/dx$  distribution sums to one, providing an estimate of the conditional PDF  $p(dE/dx|\text{residual range, local pitch})$  that is not informed by the underlying kinematics of tracks. This procedure is repeated for each plane and for two particle species, namely muons and protons. The histograms are filled with hits associated with well-reconstructed tracks produced in simulated neutrino interactions. These tracks are required to be complete, meaning that more than 90% of the true deposited



**Figure 4:** Expected  $dE/dx$  distributions for muon (blue) and proton (orange) hits. The top row shows distributions on the collection plane, with a fixed value of local pitch, and three different values of the residual range. The middle row shows distributions on the collection plane, with a fixed value of the residual range, and three different values of local pitch. The bottom row shows distributions on the three wire planes, with a fixed value of residual range and local pitch. As expected, the peak  $dE/dx$  value reduces at higher residual range, passing from 11 MeV/cm to 5.75 MeV/cm, and to 4.25 MeV/cm across the three bins under consideration for protons.

charge is reconstructed. They are also required to be pure, meaning that more than 90% of their reconstructed charge was deposited by a single particle. The tracks are also required to be contained within a fiducial volume, where both the start and end points are at least 20 cm away from the boundaries of the TPC.

The  $dE/dx$  PDF is visualized through three series of examples in figure 4, where in each row only one of the three parameters (residual range, local pitch, and plane, respectively) is varied, while keeping the other two fixed. The PDF changes considerably, showing, for example, a reduction of the  $dE/dx$  of the peak value at higher residual range and an increase of the width at higher local pitch, justifying the need for such a construction as a function of these three variables.

## 5.2 The likelihood ratio test statistic as PID score

Using the PDF previously constructed, the likelihood of any particle hypothesis can be computed for each reconstructed track. Interactions of neutrinos in the GeV energy range

in liquid argon lead to comparable rates of muons, charged pions, and protons, making the classification between these particle species important. However, this paper will focus on the binary classification problem of distinguishing muons from protons. As pions and muons have very similar masses, the calorimetric separation of these two particle species is not addressed in this paper, and pion tracks will appear as muon-like by means of this algorithm. Kaons are instead very rare (approximately 0.1% of the events in MicroBooNE are predicted to contain a kaon) and are omitted in this work.

The likelihood for a track is computed starting from the single-hit-likelihood:

$$\mathcal{L}_{\text{hit}}(\text{type}|\text{plane}, dE/dx, \text{rr}, \text{local pitch}) = p(dE/dx|\text{type}, \text{plane}, \text{rr}, \text{local pitch}), \quad (5.1)$$

where  $p$  stands for the PDF,  $\text{type}$  refers to muon or proton and  $\text{rr}$  stands for residual range. The local pitch is measured locally, and it is generally different for each hit associated with the same track, because of changes in track trajectory due to multiple Coulomb scattering. The plane is included because the PDFs are significantly different for the different planes. The single-plane-likelihood is computed by taking the product of the single-hit-likelihood for each hit on a given plane:

$$\mathcal{L}_{\text{plane}}(\text{type}|\text{plane}, \{dE/dx\}_{i=1,\dots,N}, \{\text{rr}\}_{i=1,\dots,N}, \{\text{local pitch}\}_{i=1,\dots,N}) = \prod_{i=1}^N \mathcal{L}_{\text{hit}}(\text{type}|\text{plane}, dE/dx_i, \text{rr}_i, \text{local pitch}_i), \quad (5.2)$$

where  $i = 1, \dots, N$  indexes the hits on the plane under consideration. The three-plane-likelihood, which is the likelihood for the entire track, is then computed as the product of the single-plane-likelihoods for the three wire planes.

The likelihood defined this way is an approximation as it neglects correlations between the charge measured on different wires and planes. However, fluctuations of the hit charge are in general correlated among wires on different planes, as they record the same charge through different projections. Moreover, the induced charge on neighboring wires and correlated noise introduce additional correlations between the charge recorded on different wires on the same plane [7, 8]. Modeling these correlations is in general complex, as they depend on the geometry on a track-by-track basis. Neglecting such correlations and using an approximation of the likelihood makes the method less optimal, and may result in a loss of separation power. A possible discrepancy between the data and the simulation for the values of  $dE/dx$ , which are the inputs of the PID method, could introduce a systematic bias. However, as shown in the plots in section 6 where the data and the simulation are compared, there is no evidence that this effect is important. In fact, all the correlation and noise effects introduced previously are reproduced in the simulation [7, 8], and a dedicated correction of the  $dE/dx$  distribution in angular bins is applied on top of the overall calibration [9], making the simulation precise and accurate.

The likelihood is then used to compute the likelihood ratio test statistic, which is employed to perform the classification task:

$$T(dE/dx, \text{rr}, \text{local pitch}) = \mathcal{L}(\text{muon}|dE/dx, \text{rr}, \text{local pitch})/\mathcal{L}(\text{proton}|dE/dx, \text{rr}, \text{local pitch}), \quad (5.3)$$

where the indices running on wires and planes have been omitted here for simplicity, and either a single-plane-likelihood (eq. (5.2)) or the three-plane-likelihood can be considered.

The binary classification problem of distinguishing protons from muons has the likelihood ratio as the most powerful statistical test, as proven by the Neyman-Pearson lemma. It provides the largest classification efficiency for any given value of the mis-identification rate.

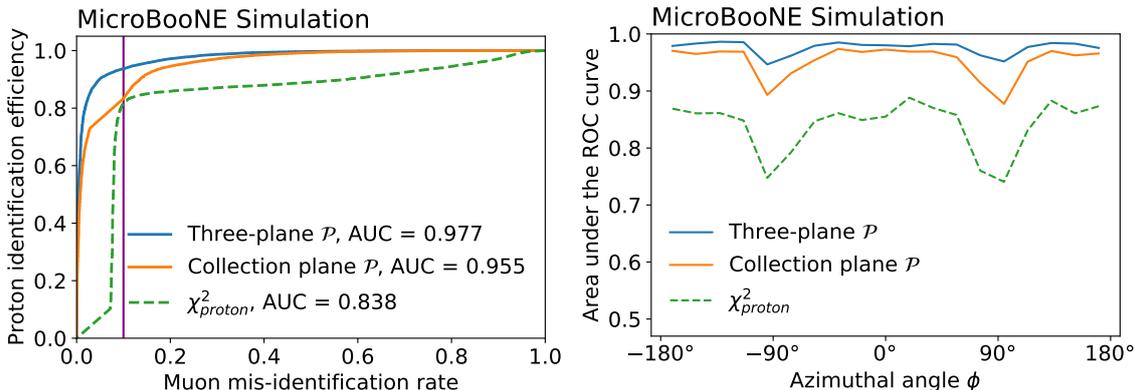
For computational purposes, in the rest of article, instead of  $T$ , the PID score  $\mathcal{P}$  will be considered, defined as:

$$\mathcal{P} = \frac{2}{\pi} \arctan(\log(T)/100). \quad (5.4)$$

Computing the logarithm of  $T$  is convenient as it reduces to a sum of log-likelihoods rather than a product of likelihoods. This bijective non-linear transformation of  $T$  does not change the separation power of the method, but it constrains the value of the PID score  $\mathcal{P}$ , otherwise unbounded, between -1 and 1, making it easier to display.

### 5.3 Performance of the particle identification

The performance of  $\mathcal{P}$  is evaluated on a test sample of more than 20000 simulated protons and muons, selected in the same manner as in section 5.1. The sample contains inclusive neutrino interactions from the Booster Neutrino Beam (BNB) simulated with GENIE v3.0.6 [11], using a tailored MicroBooNE tune [15]. A receiver operating characteristic (ROC)



**Figure 5:** Comparison of the proton/muon separation power of different PID scores. The left plot shows the ROC curves on the entire sample, and the right plot shows the area under the curve (AUC) in bins of the track angle  $\phi$ . The blue curves refer to the proposed PID score  $\mathcal{P}$  using three planes, the orange refer to the collection plane only  $\mathcal{P}$ , whereas the green curves show the  $\chi^2$  test with respect to the proton hypothesis. The purple vertical line on the left plot slices the curves at muon mis-identification rate 0.1, comparing the proton identification efficiencies of the three methods at the same working point.

curve is calculated from the test statistics distributions for the two particle types and shown in the left plot of figure 5 for the three-plane  $\mathcal{P}$ , for the collection plane only  $\mathcal{P}$ , and for the  $\chi^2$  test with respect to the proton hypothesis, which represents the previous state of the art [9]. The latter quantity, computed by comparing the data with the expectation

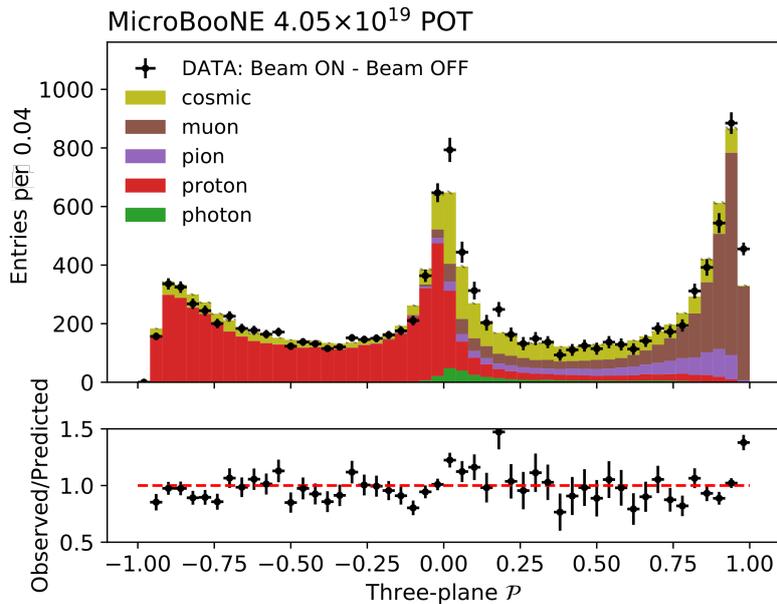
from Bethe-Bloch theory, has been used in several previous MicroBooNE analyses and it is shown here as a reference for comparison. The ROC curves show the proton efficiency as a function of the muon mis-identification rate, which are bounded between 0 and 1. For a given method, every possible cut value between -1 and 1 corresponds to a point on the ROC curve. The performance is quantified at the working point with 10% of the muons mis-identified as protons: the three-plane  $\mathcal{P}$  provides 93.7% efficiency at selecting protons compared to 83.4% for the collection plane only  $\mathcal{P}$  and 81.6% for the the  $\chi^2$  test with respect to the proton hypothesis. An overall measure of the separation power is defined using the area under the ROC curve (AUC). When this metric is equal to 1, the variable allows perfect separation at any working point, whereas a value of 0.5 represents a random guess. The three-plane  $\mathcal{P}$  scores a AUC of 0.977 compared to 0.955 for the collection-plane only  $\mathcal{P}$  and 0.838 for the the  $\chi^2$  test with respect to the proton hypothesis. The robustness of the quoted performance is tested against detector systematic uncertainties. The performance is evaluated on a series of simulations with a modified detector response to assess the detector systematic uncertainty. This leads to an uncertainty on the proton efficiency of 1.2% at 10% muon mis-id. The uncertainty on the AUC is 0.002 units for the nominal value of 0.976. This uncertainty is dominated by the modeling of electron-ion recombination. The statistical uncertainty on the efficiency and AUC determination is negligible. The right plot of figure 5 shows the AUC in bins of the azimuthal angle  $\phi$  of the track, which describes the direction of the track on the plane orthogonal to the beam direction:  $\phi = 0^\circ, \pm 180^\circ$  refer to the drift direction and  $\phi = \pm 90^\circ$  refer to the vertical direction. Both plots illustrate an overall improvement of the separation power with respect to the  $\chi^2$  test with respect to the proton hypothesis, and a mitigation of the dependence of the performance on the track angle. Combining the three planes improves the separation power in every angular region, especially for vertical tracks ( $\phi \sim \pm 90^\circ$ ), where the collection plane is the least effective.

## 6 Applications to physics cases

The following analyses were developed using data collected by MicroBooNE with the BNB during winter and spring 2016. This data amounts to  $4.8 \times 10^{19}$  protons on target (POT). This data, in which neutrino interactions are present, is labeled as DATA Beam ON. The prediction comes from a combination of the simulation of neutrino interactions and data collected out of the beam windows, labeled as DATA Beam OFF. Even in events where a neutrino interaction is present,  $\mathcal{O}(10)$  cosmic rays cross the detector on average. Instead of being simulated, cosmic ray waveforms are acquired out of the beam window and overlaid to simulated neutrino interactions.

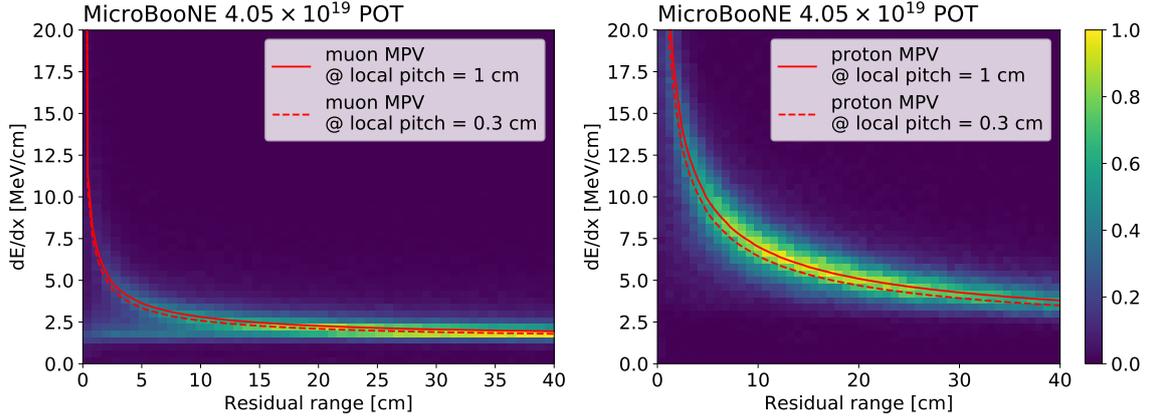
### 6.1 Proton-muon separation for tracks recorded on data

The first test performed is to verify if the result obtained in the simulation in section 5.3 holds also with neutrino data. Tracks are selected by requiring track-score  $> 0.5$ , a measurement of the likeliness of a reconstructed particle to be a track, with values ranging from 0 for shower-like particles to 1 for track-like particles. Track-score is provided by the Pandora reconstruction. Tracks are also required to be reconstructed within 5 cm from the



**Figure 6:** Distribution of the three-plane  $\mathcal{P}$  for neutrino-induced tracks selected in data and simulation. The data (black cross) shows the difference between the DATA Beam ON and the DATA Beam OFF, in order to remove the contribution from non-beam events. The simulation (stacked histogram) is normalized to the same number of events observed in the data, and it is broken down for different particle species. In the case the particle selected in the simulation is an overlaid cosmic, it is assigned to the category "cosmic". The uncertainties shown on the data points are the expected statistical uncertainties from Poisson counting.

vertex, and to be contained within a fiducial volume, defined as the set of points that are at least 20 cm apart from every side of the TPC. Figure 6 shows the distribution of the PID score for these tracks, comparing the data (black cross) with the simulation (stacked colored histogram). Protons, reconstructed with a low  $\mathcal{P}$ , populate the left side of the distribution. These are well separated from lighter particles, such as muons and pions, which populate the region at larger values of  $\mathcal{P}$ . Tracks associated with cosmic rays are distributed along the whole spectrum, as they can be induced by cosmic muons or by protons kicked out of the argon nuclei. A peak at  $\mathcal{P} \sim 0$  is also present. These are short tracks, for which there is too little information to discriminate between the two hypotheses. In fact,  $\log(T)$  is additive for each hit: the longer the track, the more hits, the larger  $\log(T)$  and thus  $\mathcal{P}$  can be. The simulation reproduces the shape of the data, confirming the performance studied in the simulation. Two additional plots (figure 7) illustrate that  $\mathcal{P}$  correctly identifies the Bragg peaks, in good agreement with the theoretical prediction. Fully contained tracks, with track-score  $> 0.8$  and collection plane local pitch  $< 1$  cm are selected in beam data events. The 2D distributions  $dE/dx$  vs residual range on the collection plane, for muon-like tracks with  $\mathcal{P} > 0.5$ , and proton-like tracks with  $\mathcal{P} < -0.5$ , are plotted on the



**Figure 7:** Collection-plane  $dE/dx$  vs residual range profile for well reconstructed, contained, and low local pitch tracks in data, identified as muon (proton) candidates on the left (right) plot. The profiles are compared to the most probable values (MPV) as expected by the theory for the extremes of the range of local pitch under consideration (red lines). The two plots are normalized to the maximum value in order to share a common color scale.

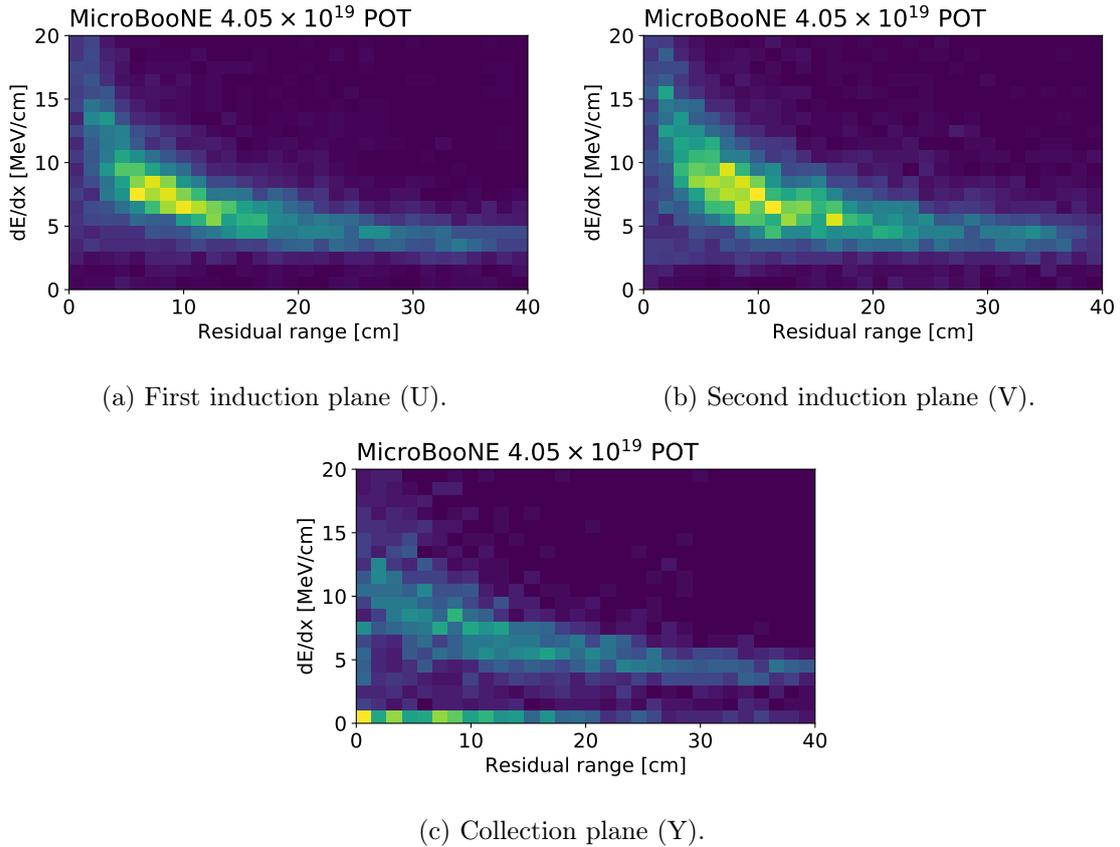
left and right of figure 7, respectively. The two Bragg peaks are clearly visible and distinct. This is possible because of the track local pitch requirement: selecting hits with small local pitch ensures  $dE/dx$  is measured properly, resulting in physical and meaningful values. The solid and dashed red lines show the theoretical prediction of the most probable value of the  $dE/dx$  distribution for the extremes of the range of local pitch under consideration. The core of the data distribution lies between the two bands, demonstrating good calorimetric reconstruction for small local pitch.

## 6.2 Large collection-plane-local pitch tracks identified with the two induction planes

The following example illustrates the efficacy of combining the calorimetric measurements performed with the three wire planes. Figure 8 shows the 2D distribution of  $dE/dx$  and residual range measured on the U, V, and Y planes, for proton candidate tracks with large collection plane local pitch. Proton candidates are required to be fully contained, and to have track-score  $> 0.8$ . Proton-likeness is required through  $\mathcal{P} < -0.5$ . The collection plane local pitch is required to be larger than 1 cm: such tracks lie on the plane orthogonal to the beam, traveling in directions where the calorimetric reconstruction is subject to large distortion. For this set of tracks, only the induction planes exhibit the expected Bragg peak: combining the three planes recovers the separation power by correctly classifying protons whose calorimetric reconstruction is not accurate on one or more views.

## 6.3 Exclusive $\nu_\mu$ selection

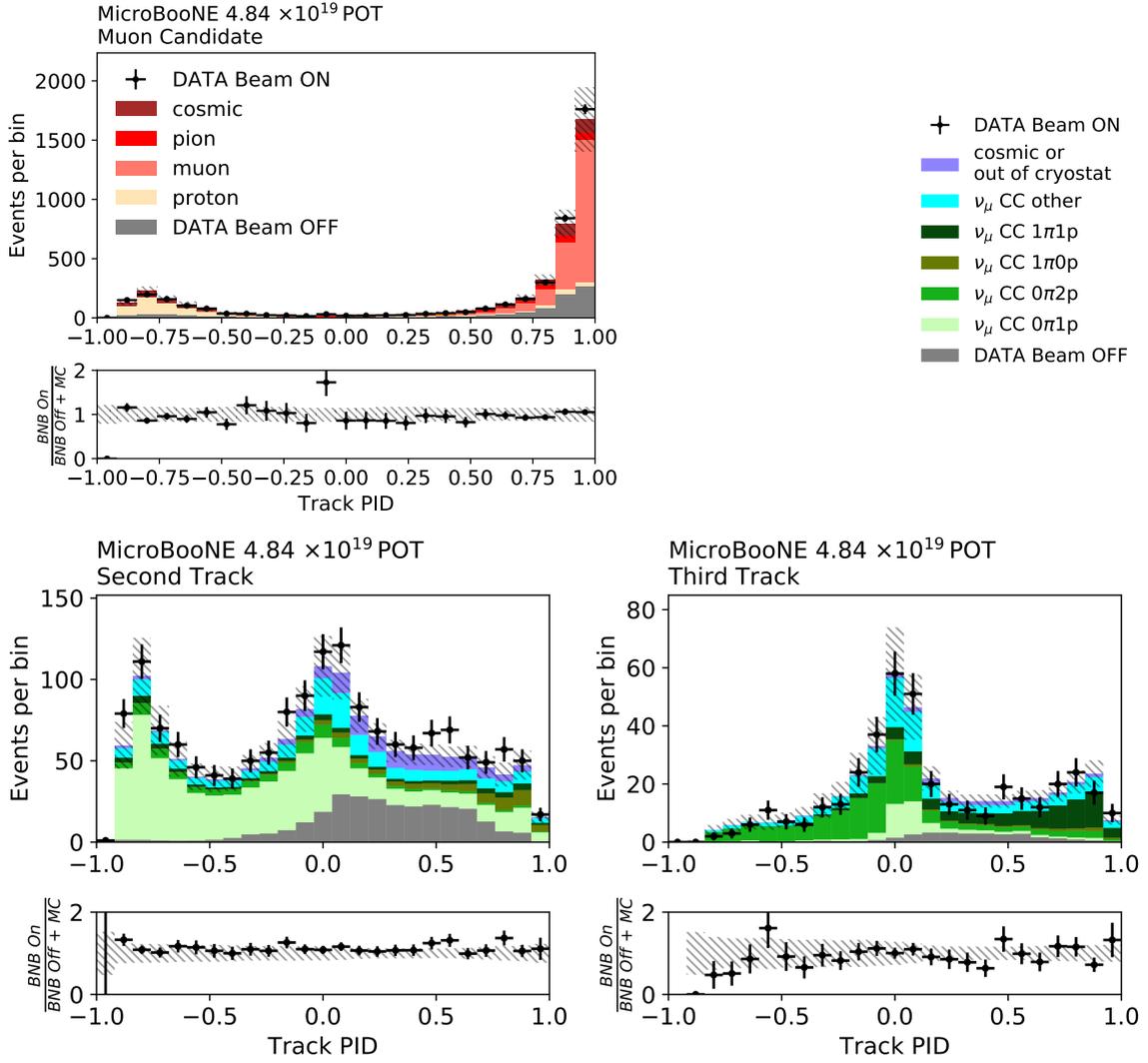
To further illustrate the separation power of  $\mathcal{P}$ , a general  $\nu_\mu CC$  selection is performed, with the selected events divided into different exclusive channels. Events are selected similarly to the procedure in [1], adding a containment requirement for all tracks reconstructed in



**Figure 8:** 2D distribution of  $dE/dx$  and residual range measured on the three wire planes for tracks identified as proton candidates in the data, with large collection plane local pitch.

the event, by requiring the start and end points of each track to lie inside the fiducial volume, as described in section 6.1. First, a muon candidate is chosen among the tracks attached to the vertex that are longer than 10 cm, by selecting the one with the largest  $\mathcal{P}$  value. The top plot of figure 9 shows the distribution of  $\mathcal{P}$  for muon candidates, showing a good separation between muon and proton tracks. Selecting only events with a candidate with  $\mathcal{P} > 0.2$  rejects most of the proton background, ensuring a pure selection of  $\nu_\mu CC$  interactions.

Among the  $\nu_\mu CC$  candidates, events with one additional reconstructed track (two-track events) are selected. If correctly reconstructed, they result predominantly from  $\nu_\mu CC$  interactions with either one proton and no pions ( $\nu_\mu CC0\pi1p$ ) or one pion and no protons ( $\nu_\mu CC1\pi0p$ ) in the final state. In general, the former case predominantly (but not solely) results from quasi-elastic interactions while the latter is largely produced by the decay of a  $\Delta$  resonance. The PID score of the second track (bottom left plot in figure 9) separates the two cases, with  $\nu_\mu CC0\pi1p$  populating the left side while the  $\nu_\mu CC1\pi0p$  are located at positive values, because pions are indistinguishable from muons with this variable. With a similar methodology, events with two additional reconstructed tracks (three-track events), are selected. Events with two protons and no pions in the final state ( $\nu_\mu CC0\pi2p$ ), predicted



**Figure 9:** PID score distributions for the muon candidate track (top), the second track in events with two reconstructed tracks (bottom left), and the third track in events with three reconstructed tracks (bottom right). The DATA Beam ON (black cross) is compared with the prediction based on the sum of the simulation of neutrino interactions (stacked histogram) and DATA Beam OFF (gray bars). The selections are based on reconstructed quantities, while the prediction is broken down into different categories based on truth information. In the first plot the different colors correspond to different particle types, while in the other two they correspond to different final states. The uncertainties shown on the data points are the expected statistical uncertainties from Poisson counting, while the hashed patches on the stacked histogram illustrates systematic uncertainties on the prediction related to the simulation of the neutrino flux and interaction model.

to be mainly induced by meson-exchange current interactions and final state effects, can be distinguished from events with one proton and one pion in the final state ( $\nu_\mu\text{CC}1\pi1p$ ),

produced by a resonance decay. Because the presence of a proton, identified by a large negative  $\mathcal{P}$ , is common to the two cases, the track with the largest PID score among the two additional tracks (bottom right plot in figure 9) is used to discriminate between  $\nu_\mu CC0\pi2p$ , on the left, and  $\nu_\mu CC1\pi1p$ , on the right. These examples demonstrate the potential of the PID score in selecting exclusive final states that can enrich contributions from different interaction modes, eventually leading to precise and detailed neutrino cross-section measurements.

## 7 Conclusions

The capability to perform precise calorimetric measurements is one of the most important factors that make LArTPCs powerful tools in the study of neutrino interactions. This work illustrates a detailed study of the performance of the calorimetric reconstruction using MicroBooNE data, the longest-operating LArTPC in a neutrino beam producing a large dataset of GeV scale neutrino interactions. The first important observation is that the calorimetric reconstruction performed by LArTPCs exhibits angular dependencies. Since the charge is drifted in a specific direction, and projected onto wires oriented in three different directions, both the  $dE/dx$  distribution that a perfect detector would measure and the precision and accuracy with which  $dE/dx$  is actually measured depend on the track direction. These effects are intrinsic to any wire-based readout LArTPC. When not properly accounted for, they result in non-uniform and sub-optimal particle identification performance. This work proposes a new likelihood-based method to perform particle identification which properly accounts for angular dependencies and mitigates their impact by effectively combining the calorimetric measurements performed on the three wire planes. It does so through the calculation of a likelihood derived from the detailed MicroBooNE simulation. The resulting PID score shows greater separation power between proton and muon-induced tracks, with smaller dependence on the track angle with respect to the previous PID method. This is quantified as a 94% proton selection efficiency with a 10% muon mis-identification rate. This methodology allows a highly effective separation of different final states, with only minor angular dependency, as demonstrated by selecting exclusive final states originated by different  $\nu_\mu CC$  interaction modes. Part of the future MicroBooNE analyses will incorporate this method in the event selection strategy. This methodology can be exported to other present and future LArTPCs, making it an important ingredient to address the ambitious liquid argon neutrino physics program moving forward.

## Acknowledgments

This research is supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under the Award Number DE-SC0007881.

This document was prepared by the MicroBooNE collaboration using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under Contract No. DE-AC02-07CH11359. MicroBooNE is supported by the

following: the U.S. Department of Energy, Office of Science, Offices of High Energy Physics and Nuclear Physics; the U.S. National Science Foundation; the Swiss National Science Foundation; the Science and Technology Facilities Council (STFC), part of the United Kingdom Research and Innovation; and The Royal Society (United Kingdom). Additional support for the laser calibration system and cosmic ray tagger was provided by the Albert Einstein Center for Fundamental Physics, Bern, Switzerland. N.F. dedicates this paper to Federico Tonielli, a physicist, peer, and friend who inspired and taught many young physicists but disappeared too early for the contribution he could have made to science.

## References

- [1] P. Abratenko et al. First Measurement of Inclusive Muon Neutrino Charged Current Differential Cross Sections on Argon at  $E_\nu \sim 0.8$  GeV with the MicroBooNE Detector. *Phys. Rev. Lett.*, 123(13):131801, 2019. doi: 10.1103/PhysRevLett.123.131801.
- [2] R. Acciarri et al. A Study of Electron Recombination Using Highly Ionizing Particles in the ArgoNeuT Liquid Argon TPC. *JINST*, 8:P08005, 2013. doi: 10.1088/1748-0221/8/08/P08005.
- [3] R. Acciarri et al. Design and Construction of the MicroBooNE Detector. *JINST*, 12(02):P02017, 2017. doi: 10.1088/1748-0221/12/02/P02017.
- [4] R. Acciarri et al. Noise Characterization and Filtering in the MicroBooNE Liquid Argon TPC. *JINST*, 12(08):P08003, 2017. doi: 10.1088/1748-0221/12/08/P08003.
- [5] R. Acciarri et al. 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(1):82, 2018. doi: 10.1140/epjc/s10052-017-5481-6.
- [6] R. Acciarri et al. The Liquid Argon In A Testbeam (LArIAT) Experiment. *JINST*, 15(04):P04026, 2020. doi: 10.1088/1748-0221/15/04/P04026.
- [7] C. Adams et al. Ionization electron signal processing in single phase LArTPCs. Part I. Algorithm Description and quantitative evaluation with MicroBooNE simulation. *JINST*, 13(07):P07006, 2018. doi: 10.1088/1748-0221/13/07/P07006.
- [8] C. Adams et al. Ionization electron signal processing in single phase LArTPCs. Part II. Data/simulation comparison and performance in MicroBooNE. *JINST*, 13(07):P07007, 2018. doi: 10.1088/1748-0221/13/07/P07007.
- [9] C. Adams et al. Calibration of the charge and energy loss per unit length of the MicroBooNE liquid argon time projection chamber using muons and protons. *JINST*, 15(03):P03022, 2020. doi: 10.1088/1748-0221/15/03/P03022.
- [10] C. Anderson et al. The ArgoNeuT Detector in the NuMI Low-Energy beam line at Fermilab. *JINST*, 7:P10019, 2012. doi: 10.1088/1748-0221/7/10/P10019.
- [11] C. Andreopoulos et al. The GENIE Neutrino Monte Carlo Generator. *Nucl. Instrum. Meth. A*, 614:87–104, 2010. doi: 10.1016/j.nima.2009.12.009.
- [12] M. Antonello et al. Precise 3D track reconstruction algorithm for the ICARUS T600 liquid argon time projection chamber detector. *Adv. High Energy Phys.*, 2013:260820, 2013. doi: 10.1155/2013/260820.
- [13] B. Baller. Liquid argon TPC signal formation, signal processing and reconstruction techniques. *JINST*, 12(07):P07010, 2017. doi: 10.1088/1748-0221/12/07/P07010.
- [14] L. D. Landau. On the energy loss of fast particles by ionization. *J. Phys.*, 8(4):201–205, 1944. URL <https://cds.cern.ch/record/216256>.
- [15] MicroBooNE. Neutrino Interaction Model and Uncertainties for MicroBooNE Analyses. upcoming publication.
- [16] P. A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020. doi: 10.1093/ptep/ptaa104.