

# 2 Neutrino Event Selection in the MicroBooNE Liquid Argon 3 Time Projection Chamber using Wire-Cell 3-D Imaging, 4 Clustering, and Charge-Light Matching

---

5 P. Abratenko<sup>jj</sup> M. Alrashed<sup>o</sup> R. An<sup>n</sup> J. Anthony<sup>d</sup> J. Asaadi<sup>ii</sup> A. Ashkenazi<sup>s</sup>  
6 S. Balasubramanian<sup>mm</sup> B. Baller<sup>k</sup> C. Barnes<sup>t</sup> G. Barr<sup>x</sup> V. Basque<sup>r</sup> L. Bathe-Peters<sup>m</sup>  
7 O. Benevides Rodrigues<sup>ff</sup> S. Berkman<sup>k</sup> A. Bhanderi<sup>r</sup> A. Bhat<sup>ff</sup> M. Bishai<sup>b</sup> A. Blake<sup>p</sup>  
8 T. Bolton<sup>o</sup> L. Camilleri<sup>j</sup> D. Caratelli<sup>k</sup> I. Caro Terrazas<sup>i</sup> R. Castillo Fernandez<sup>k</sup> F. Cavanna<sup>k</sup>  
9 G. Cerati<sup>k</sup> Y. Chen<sup>a</sup> E. Church<sup>y</sup> D. Cianci<sup>j</sup> J. M. Conrad<sup>s</sup> M. Convery<sup>cc</sup> L. Cooper-Troendle<sup>mm</sup>  
10 J. I. Crespo-Anadón<sup>j,f</sup> M. Del Tutto<sup>k</sup> D. Devitt<sup>p</sup> R. Diurba<sup>u</sup> L. Domine<sup>cc</sup> R. Dorrill<sup>n</sup> K. Duffy<sup>k</sup>  
11 S. Dytman<sup>z</sup> B. Eberly<sup>ee</sup> A. Ereditato<sup>a</sup> L. Escudero Sanchez<sup>d</sup> J. J. Evans<sup>r</sup>  
12 G. A. Fiorentini Aguirre<sup>dd</sup> R. S. Fitzpatrick<sup>t</sup> B. T. Fleming<sup>mm</sup> N. Foppiani<sup>m</sup> D. Franco<sup>mm</sup>  
13 A. P. Furmanski<sup>u</sup> D. Garcia-Gamez<sup>l</sup> S. Gardiner<sup>k</sup> G. Ge<sup>j</sup> S. Gollapinni<sup>hh,q</sup> O. Goodwin<sup>r</sup>  
14 E. Gramellini<sup>k</sup> P. Green<sup>r</sup> H. Greenlee<sup>k</sup> W. Gu<sup>b</sup> R. Guenette<sup>m</sup> P. Guzowski<sup>r</sup> E. Hall<sup>s</sup>  
15 P. Hamilton<sup>ff</sup> O. Hen<sup>s</sup> G. A. Horton-Smith<sup>o</sup> A. Hourlier<sup>s</sup> E.-C. Huang<sup>q</sup> R. Itay<sup>cc</sup> C. James<sup>k</sup>  
16 J. Jan de Vries<sup>d</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>s</sup>  
17 G. Karagiorgi<sup>j</sup> W. Ketchum<sup>k</sup> B. Kirby<sup>b</sup> M. Kirby<sup>k</sup> T. Kobilarcik<sup>k</sup> I. Kreslo<sup>a</sup> R. LaZur<sup>i</sup>  
18 I. Lepetic<sup>aa</sup> K. Li<sup>mm</sup> Y. Li<sup>b</sup> B. R. Littlejohn<sup>n</sup> D. Lorca<sup>a</sup> W. C. Louis<sup>q</sup> X. Luo<sup>c</sup> A. Marchionni<sup>k</sup>  
19 S. Marrocchi<sup>k</sup> C. Mariani<sup>kk</sup> D. Marsden<sup>r</sup> J. Marshall<sup>ll</sup> J. Martin-Albo<sup>m</sup>  
20 D. A. Martinez Caicedo<sup>dd</sup> K. Mason<sup>jj</sup> A. Mastbaum<sup>aa</sup> N. McConkey<sup>r</sup> V. Meddage<sup>o</sup> T. Mettler<sup>a</sup>  
21 K. Miller<sup>g</sup> J. Mills<sup>jj</sup> K. Mistry<sup>r</sup> A. Mogan<sup>hh</sup> T. Mohayaik<sup>k</sup> J. Moon<sup>s</sup> M. Mooney<sup>i</sup> A. F. Moor<sup>d</sup>  
22 C. D. Moore<sup>k</sup> J. Mousseau<sup>t</sup> M. Murphy<sup>kk</sup> D. Naples<sup>z</sup> A. Navrer-Agasson<sup>r</sup> R. K. Neely<sup>o</sup>  
23 P. Nienaber<sup>bb</sup> J. Nowak<sup>p</sup> O. Palamara<sup>k</sup> V. Paolone<sup>z</sup> A. Papadopoulos<sup>s</sup> V. Papavassiliou<sup>v</sup>  
24 S. F. Pate<sup>v</sup> A. Paudel<sup>o</sup> Z. Pavlovic<sup>k</sup> E. Piasetzky<sup>gg</sup> I. D. Ponce-Pinto<sup>j</sup> D. Porzio<sup>r</sup> S. Prince<sup>m</sup>  
25 X. Qian<sup>b</sup> J. L. Raaf<sup>k</sup> V. Radeka<sup>b</sup> A. Rafique<sup>o</sup> M. Reggiani-Guzzo<sup>r</sup> L. Ren<sup>v</sup> L. Rochester<sup>cc</sup>  
26 J. Rodriguez Rondon<sup>dd</sup> H.E. Rogers<sup>e</sup> M. Rosenberg<sup>z</sup> M. Ross-Lonergan<sup>j</sup> B. Russell<sup>mm</sup>  
27 G. Scanavini<sup>mm</sup> D. W. Schmitz<sup>g</sup> A. Schukraft<sup>k</sup> M. H. Shaevitz<sup>j</sup> R. Sharankova<sup>jj</sup> J. Sinclair<sup>a</sup>  
28 A. Smith<sup>d</sup> E. L. Snider<sup>k</sup> M. Soderberg<sup>ff</sup> S. Söldner-Rembold<sup>r</sup> S. R. Soleti<sup>x,m</sup>  
29 P. Spentzouris<sup>k</sup> J. Spitz<sup>t</sup> M. Stancari<sup>k</sup> J. St. John<sup>k</sup> T. Strauss<sup>k</sup> K. Sutton<sup>j</sup>  
30 S. Sword-Fehlberg<sup>v</sup> A. M. Szec<sup>r</sup> N. Tagg<sup>w</sup> W. Tang<sup>hh</sup> K. Terao<sup>cc</sup> C. Thorpe<sup>p</sup> M. Toups<sup>k</sup>  
31 Y.-T. Tsai<sup>cc</sup> S. Tufanli<sup>mm</sup> M. A. Uchida<sup>d</sup> T. Usher<sup>cc</sup> W. Van De Pontseele<sup>x,m</sup> B. Viren<sup>b</sup>  
32 M. Weber<sup>a</sup> H. Wei<sup>b</sup> Z. Williams<sup>ii</sup> S. Wolbers<sup>k</sup> T. Wongjirad<sup>jj</sup> M. Wospakrik<sup>k</sup> W. Wu<sup>k</sup> T. Yang<sup>k</sup>  
33 G. Yarbrough<sup>hh</sup> L. E. Yates<sup>s</sup> H. W. Yu<sup>b</sup> G. P. Zeller<sup>k</sup> J. Zennamo<sup>k</sup> C. Zhang<sup>b</sup>

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

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

36 <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>Fermi National Accelerator Laboratory (FNAL), Batavia, IL 60510, USA  
<sup>l</sup>Universidad de Granada, E-18071, Granada, Spain  
<sup>m</sup>Harvard University, Cambridge, MA 02138, USA  
<sup>n</sup>Illinois Institute of Technology (IIT), Chicago, IL 60616, USA  
<sup>o</sup>Kansas State University (KSU), Manhattan, KS, 66506, USA  
<sup>p</sup>Lancaster University, Lancaster LA1 4YW, United Kingdom  
<sup>q</sup>Los Alamos National Laboratory (LANL), Los Alamos, NM, 87545, USA  
<sup>r</sup>The University of Manchester, Manchester M13 9PL, United Kingdom  
<sup>s</sup>Massachusetts Institute of Technology (MIT), Cambridge, MA, 02139, USA  
<sup>t</sup>University of Michigan, Ann Arbor, MI, 48109, USA  
<sup>u</sup>University of Minnesota, Minneapolis, Mn, 55455, USA  
<sup>v</sup>New Mexico State University (NMSU), Las Cruces, NM, 88003, USA  
<sup>w</sup>Otterbein University, Westerville, OH, 43081, USA  
<sup>x</sup>University of Oxford, Oxford OX1 3RH, United Kingdom  
<sup>y</sup>Pacific Northwest National Laboratory (PNNL), Richland, WA, 99352, USA  
<sup>z</sup>University of Pittsburgh, Pittsburgh, PA, 15260, USA  
<sup>aa</sup>Rutgers University, Piscataway, NJ, 08854, USA, PA  
<sup>bb</sup>Saint Mary's University of Minnesota, Winona, MN, 55987, USA  
<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)

75 ABSTRACT: An accurate and efficient event reconstruction is required to realize the full scientific  
76 capability of liquid argon time projection chambers (LArTPCs). The current and future neutrino  
77 experiments that rely on massive LArTPCs create a need for new ideas and reconstruction ap-  
78 proaches. Wire-Cell, proposed in recent years, is a novel tomographic event reconstruction method  
79 for LArTPCs. The Wire-Cell 3D imaging approach capitalizes on charge, sparsity, time, and ge-  
80 ometry information to reconstruct a topology-agnostic 3D image of the ionization electrons prior  
81 to pattern recognition. A second novel method, the many-to-many charge-light matching, then  
82 pairs the TPC charge activity to the detected scintillation light signal, thus enabling a powerful  
83 rejection of cosmic-ray muons in the MicroBooNE detector. A robust processing of the scintillation  
84 light signal and an appropriate clustering of the reconstructed 3D image are fundamental to this  
85 technique. In this paper, we describe the principles and algorithms of these techniques and their suc-  
86 cessful application in the MicroBooNE experiment. A quantitative evaluation of the performance  
87 of these techniques is presented. Using these techniques, a 95% efficient pre-selection of neutrino  
88 charged-current events is achieved with a 30-fold reduction of non-beam-coincident cosmic-ray  
89 muons, and about 80% of the selected neutrino charged-current events are reconstructed with at  
90 least 70% completeness and 80% purity.

91 KEYWORDS: LArTPC, MicroBooNE, Wire-Cell, 3D imaging, charge-light matching, clustering

---

92	<b>Contents</b>	
93	<b>1 Introduction</b>	<b>1</b>
94	<b>2 The MicroBooNE detector</b>	<b>2</b>
95	<b>3 Wire-Cell 3D Imaging</b>	<b>4</b>
96	3.1 Tiling	5
97	3.2 Charge solving	8
98	3.3 De-ghosting	11
99	3.4 Summary	12
100	<b>4 Matching Charge and Light</b>	<b>16</b>
101	4.1 3D clustering	16
102	4.2 PMT light signal reconstruction	25
103	4.3 Many-to-many charge-light matching	29
104	<b>5 Evaluation of the Wire-Cell 3D imaging and the charge-light matching</b>	<b>36</b>
105	5.1 Imaging performance of ideal tracks	37
106	5.2 Imaging performance of neutrino interactions	40
107	5.3 Final performance in realistic cases	43
108	<b>6 Summary and Discussion</b>	<b>49</b>

---

## 109 1 Introduction

110 The Liquid Argon Time Projection Chamber (LArTPC) [1–4] is a novel detector technology under  
111 rapid development. It is a fully active calorimeter with excellent 3D tracking capability, which can  
112 enable particle identification (PID) of unprecedented power in neutrino detection. This detector  
113 technology has been utilized in many current accelerator neutrino experiments, such as Micro-  
114 BooNE [5] and the Short Baseline Neutrino (SBN) program [6], and it will be used in the future  
115 massive LArTPC experiments, such as DUNE [7].

116 Event reconstruction is one of the most challenging tasks in analyzing the data from current and  
117 future large-scale LArTPCs. A high-performance event reconstruction is vital to take full advantage  
118 of the capability of LArTPCs for physics measurements. Multiple reconstruction approaches are  
119 being developed in MicroBooNE, including the Pandora multi-algorithm pattern recognition [8]  
120 and deep learning with convolutional neural networks [9, 10]. Another novel event reconstruction  
121 method, Wire-Cell, has also been under rapid development for MicroBooNE. The Wire-Cell 3D  
122 imaging [11] capitalizes on the most fundamental LArTPC detector information – time, charge,  
123 and geometry – to tomographically reconstruct a topology-agnostic three-dimensional image of

the ionization electrons prior to any pattern recognition step. The early construction of the 3D image without the involvement of pattern recognition is the primary distinction between Wire-Cell and other reconstruction paradigms [8–10]. This is beneficial because in 3D the particle activities are more separated than in 2D, which reduces the difficulties in clustering and other pattern recognition tasks. Enabled by the high-performance ionization electron signal processing procedure in MicroBooNE [12–14], the Wire-Cell 3D imaging reduces the degeneracies – integrated charge measured along each wire other than pixelated measurement of charge – inherent in the LArTPC wire readouts as used by MicroBooNE and numerous other experiments.

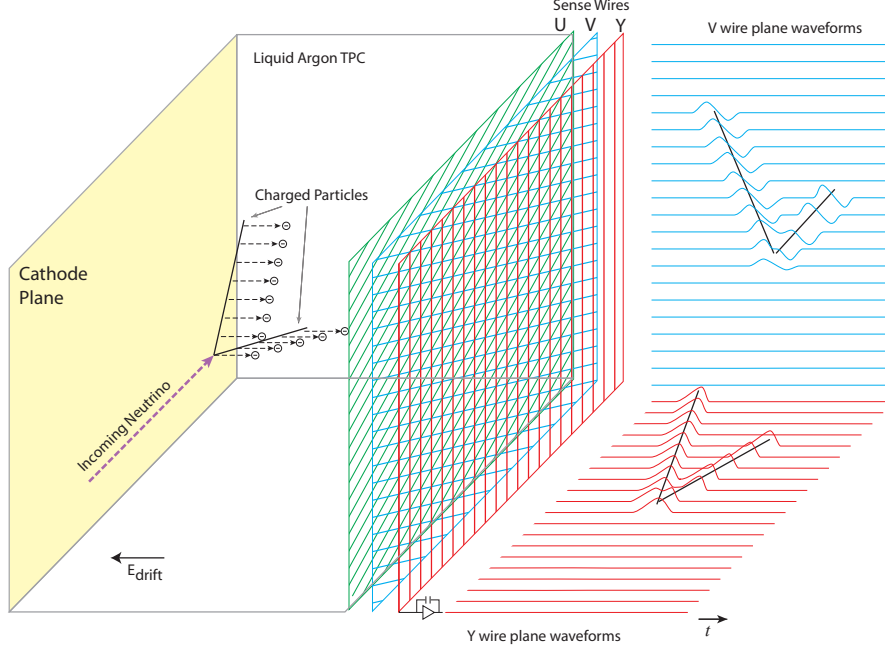
Detector defects such as nonfunctional channels (10% of all wire readouts in MicroBooNE) and the numerous cosmic-ray muons (20–30 per TPC readout window) in the MicroBooNE detector pose additional challenges to the overall success of the event reconstruction. We address the first problem by allowing for the reconstruction in regions where two out of three channels, one from each wire plane, are functional. For these regions, an analysis that also relies on information from nearby fully functional regions is performed. Our method significantly reduces the extent of unusable regions by a factor of ten. To deal with the high rate of cosmic rays, we developed a many-to-many TPC-charge and PMT-light (charge-light) matching method, to distinguish the candidate neutrino activity, which is in coincidence with the beam spill, from the numerous cosmic rays spanning the entire MicroBooNE detector and the TPC readout window. TPC activity hereafter refers to the energy deposition in LArTPC by ionization. It originates from either a cosmic-ray muon or a neutrino interaction. This method relies on the Wire-Cell 3D imaging and emphasizes the interplay between the scintillation light and the ionization charge signals created by charged particles traversing the LAr. A robust processing of the scintillation light signals from the photon detector system and an appropriate clustering, which groups the TPC activities that represent signals initiated by an individual primary particle, are fundamental to this technique.

In this paper, we describe the principles, algorithms, and performance evaluation of the Wire-Cell 3D imaging and the many-to-many charge-light matching, including the light signal processing and the 3D clustering. These techniques provide a solid foundation to reject coincident in-beam cosmic-ray muons [15] with downstream reconstruction techniques (e.g. track trajectory fitting and pattern recognition). The outcome of these tools, e.g. the Wire-Cell 2D and 3D images of the neutrino activities with the surrounding cosmic-ray activities removed, can also improve the performance of other reconstruction paradigms [8–10]. The principle and implementation of the Wire-Cell 3D imaging is presented in section 3. The many-to-many charge-light matching to pair the TPC activities to the reconstructed PMT activities is described in section 4 as the final step to select the candidate neutrino activities. Evaluations of the quality of the Wire-Cell 3D imaging and the efficacy of the many-to-many charge-light matching are demonstrated in section 5. A summary of the performance and discussion is presented in section 6.

## 2 The MicroBooNE detector

The MicroBooNE detector is the first LArTPC in the SBN program to measure neutrino interactions from the on-axis Booster neutrino beam (BNB) [16] at Fermi National Accelerator Laboratory in Batavia, IL. MicroBooNE uses a single-phase (i.e. liquid phase only) LArTPC with a rectangular active volume of the following dimensions: 2.6 m (width, along the drift direction), 2.3 m (height,

vertical), and 10.4 m (length, along the beam direction), as illustrated in figure 1. The TPC has an active mass of 85 metric tonnes and is immersed in a single-walled and cylindrical shaped cryostat with a 170 tonne liquid argon capacity.



**Figure 1:** Illustration of single-phase LArTPCs [5]. Each wire plane provides a 2D image of the ionization electrons with respect to a specific wire orientation.

A high voltage of  $-70$  kV applied on the cathode plane provides a drift field of  $273$  V/cm. The electrons ionized by any energy deposition from traversing charged particles drift towards the anode wire planes along the electric field at a nominal speed of about  $1.10$  mm/ $\mu$ s. In this paper, we use the X-axis to represent the direction away from the readout wire plane and opposite to the ionization charge drift, Y-axis to represent the vertical-up direction, and Z-axis to represent the BNB beam direction. There are three parallel wire readout planes [17] on the anode side with different wire orientations. The first wire plane facing the cathode is labeled “U”, and the second and third plane are labeled “V” and “Y”, respectively. The 3456 wires in the Y plane are oriented vertically and the 2400 wires in the U (V) plane are oriented  $+(-)60^\circ$  with respect to the vertical direction. The spacing between adjacent wires and adjacent wire planes are both 3 mm. Different bias voltages,  $-110$  V,  $0$  V, and  $230$  V, are applied to the U, V, and Y wire planes, respectively, to ensure all ionization electrons drift through the U and V planes before being collected by the Y plane. The U and V planes are commonly referred to as the induction planes and the ionization electrons induce bipolar electrical signals as they pass through the planes; the Y plane is referred to as the collection plane and sees unipolar electrical pulses.

The TPC readout is defined with respect to the event trigger and includes three 1.6 ms frames, spanning -1.6 ms to +3.2 ms relative to the trigger time, with a sampling rate of 2 MHz (0.5  $\mu$ s per time tick). Therefore, each wire plane records a 2D image (time versus wire) of the ionization electrons within the full 4.8 ms TPC readout.

Behind the wire planes and external to the TPC, there is an array of thirty-two 8" photomultiplier tubes (PMTs) [18] to detect scintillation light for triggering, timing, and other purposes. The PMT readout includes four 1.6 ms frames with the beam gate window (1.6  $\mu$ s beam-spill) contained within the second 1.6 ms frame. The sampling rate is 64 MHz (15.625 ns per sample) for each PMT and the signal is recorded in a dynamic-range-based, paired form for each channel – a high gain (x10) signal and a low gain (x1) signal. The 32 PMTs promptly (in a few nanoseconds) detect the scintillation light and provide the intensity and position information of the photo-electrons originating from either a cosmic-ray muon or a neutrino interaction. The TPC and PMT readouts cover the full time range of the beam neutrino activities as well as cosmic-ray activities that enter the beam spill frame during the relatively slow drift of ionization electrons, which has a maximum drift time of 2.3 ms.

### 3 Wire-Cell 3D Imaging

Analysis of the single-phase LArTPC with a wire readout scheme is a natural application of the tomography technique, which the Wire-Cell 3D imaging strictly follows. Ref. [11] introduces the basic concepts and the key mathematics of the Wire-Cell 3D imaging. In this section, we focus more on the realistic issues when applying the Wire-Cell 3D imaging to MicroBooNE data.

The fundamental information provided by a LArTPC is as follows:

- (i) **Time** - when the ionization electrons arrive at the anode wire plane<sup>1</sup>.
- (ii) **Geometry** - the positions of the wires from each plane that have signals from the ionization electrons, i.e. hit wires.
- (iii) **Charge** - the number of ionization electrons measured by the hit wires from each wire plane.

The time and charge information comes from the time distribution of the deconvolved charge, which is obtained via advanced signal processing techniques. In particular, the 2D deconvolution technique [13, 14] significantly improves the signal processing for the induction planes and makes the deconvolved charge consistent across the multiple wire planes. The geometry information is the wire position, along the wire pitch direction (perpendicular to the wire orientation). Since the wire planes have different wire orientations, signals on each wire are taken as a 1D projection of the charge depositions with the summation of the charge available in the proximity of each wire. The position of each individual charge deposition along the wire itself can only be provided by other wire planes.

The Wire-Cell 3D imaging uses two major steps to reconstruct the 3D image of the ionization electrons arriving at the anode plane: 1) Reconstruct the 2D image of the ionization electrons on the anode plane in a given time slice, e.g. 2  $\mu$ s (4 ticks in the TPC readout) considering the intrinsic time smearing of about 1.5  $\mu$ s after signal processing [13]. The integrated charge within the time slice on each hit wire is used; 2) Concatenate the 2D images from the previous step in the sequence

---

<sup>1</sup>The absolute starting time of each cosmic-ray muon needs to be corrected by using the light signal information with the charge-light matching technique described in section 4.3

of time slices to form the 3D image. From three wire readout planes, at most three 1D projection views are available within one time slice, in contrast with the dozens or even hundreds of 1D projection views available in common tomography applications, such as those for medical imaging. Compared to a pixelated readout with  $n^2$  pixels, the  $O(n)$  wires ( $3 \times n$  for three wire planes) afforded by a wire readout scheme reduces the heat loads and the cost of the readout system, but result in a considerable loss of information. To recover from the loss of information, additional constraints are used:

- (iv) **Sparsity** - the distribution of ionization electrons in space is expected to be sparse, typically occupying less than 10% of the local bounding volume that contains the activities, for any physical signals.
- (v) **Proximity** - the ionization electrons are read out by consecutive wires because a charged particle ionizes argon atoms continuously in the fully active LArTPC volume.
- (vi) **Positivity** - the number of the drifted ionization electrons can only be positive.

The actual procedure we use to incorporate the above information is divided into two processes: *tiling* and *solving*, as described in section 3.1 and section 3.2, respectively. In the implementation of Wire-Cell 3D imaging in MicroBooNE, the nonfunctional wires [12] aggravate the wire readout ambiguity, and introduce a large number of *ghost* energy depositions. A dedicated de-ghosting algorithm, discussed in section 3.3, is developed to mitigate this effect.

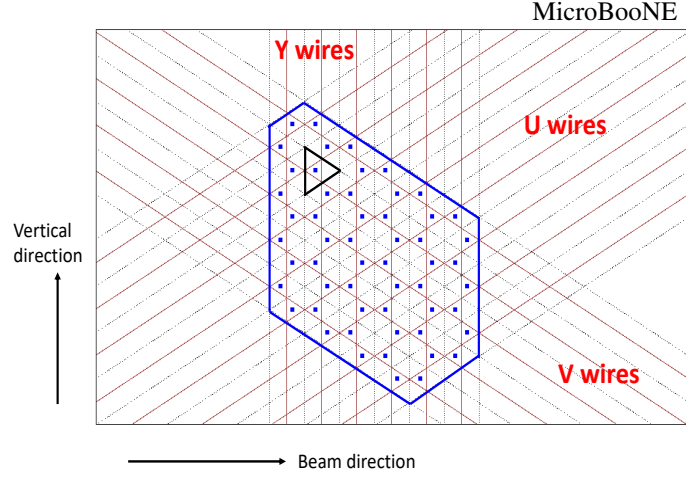
### 3.1 Tiling

The 2D image of the ionization electrons in a time slice consists of *cells*, which are the smallest geometric units formed by wires from three planes. Figure 2 shows tens of cells, for example the black triangle, which is the overlapping area of three wires from the three wire planes. Each wire represents a 2D region centered around the wire location with its width equal to the wire pitch. All cells have equilateral triangular shapes because of the MicroBooNE wire orientations and positioning.

The smallest time unit in the Wire-Cell imaging is a time slice, whose  $2 \mu\text{s}$  width contains four sampling ticks from the TPC readout. The width of the time slice introduces negligible information loss because the software filtering in the signal processing has a cut-off frequency at about 0.25 MHz to optimize the signal-to-noise ratio, which in turn smears the time resolution. Geometry is used to determine all possible hit cells within a time slice by finding the intersections of the hit wires. In figure 2, there are 8 hit U wires (2.4 cm wide), 5 hit V wires (1.5 cm wide), and 6 hit Y wires (1.8 cm wide), leading to 55 possible hit cells. The fact that there are fewer knowns (19 hit wires) than unknowns (55 cells) indicates an ambiguity is introduced by the wire readout. Meanwhile, the amount of integrated charge in a time slice and the identities of active wires in that time slice are affected by diffusion and long-range induction effects (especially for induction planes) as charge drifts in the TPC, as well as the action of software filters applied to the waveforms [13, 14]. To mitigate the impact from wire ambiguity and charge smearing, a procedure to merge the consecutive hit cells is developed, called *tiling*. The groups of hit cells after tiling are called *blobs*. The blob in figure 2 is marked by solid blue lines. The connected hit wires are merged as *wire bundles* in the tiling procedure, and a blob is the overlapping area of three wire bundles from each wire plane as shown in figure 2. Note that a cell or a blob can be taken as a 3D object and its length along the drift direction is the width of the time slice, i.e.  $2 \mu\text{s}$  or about 2.2 mm. In the following sections,



264 3D space points will be used to describe the algorithms and a “space point” is equivalent to a “cell”  
 265 hereafter, which represents a 3D voxel of the space with a finite size. Its charge is deduced by the  
 266 total charge of the blob that contains it, divided by the number of space points within the blob.

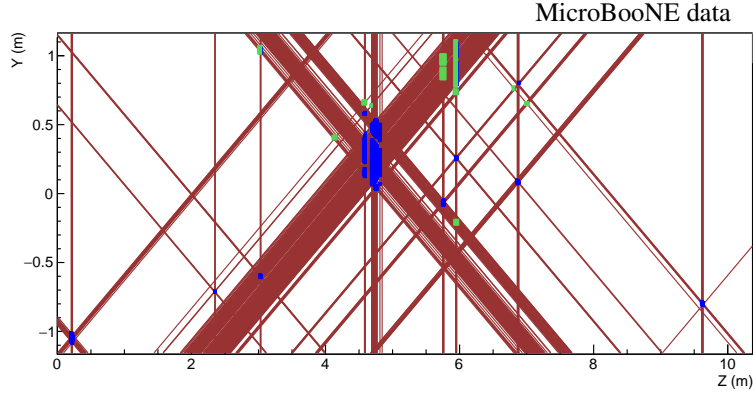


**Figure 2:** An example of the hit cells and blob constructed by the hit wires with the MicroBooNE detector geometry. Each wire is represented by a solid red line and the wire (pitch) boundaries are represented by dashed black lines. All hit cells have equilateral triangular shapes and are marked with blue dots at their centers. An example cell is marked by the black triangle. A blob is formed by the contiguous hit cells and marked by solid blue lines.

267 There are three advantages to the tiling. Firstly, it completely collects the reconstructed charge  
 268 smeared to the adjacent wires, resulting in more consistent charge values across the wire planes.  
 269 Secondly, it greatly reduces the number of unknowns in the later stage of *solving*. Thirdly, it  
 270 significantly reduces the computational cost. The charge smearing is different for different wire  
 271 planes. Obtaining consistent charge measurements across multiple wire planes by the tiling is  
 272 fundamental to construct and solve the Wire-Cell 3D imaging equation as described in section 3.2.

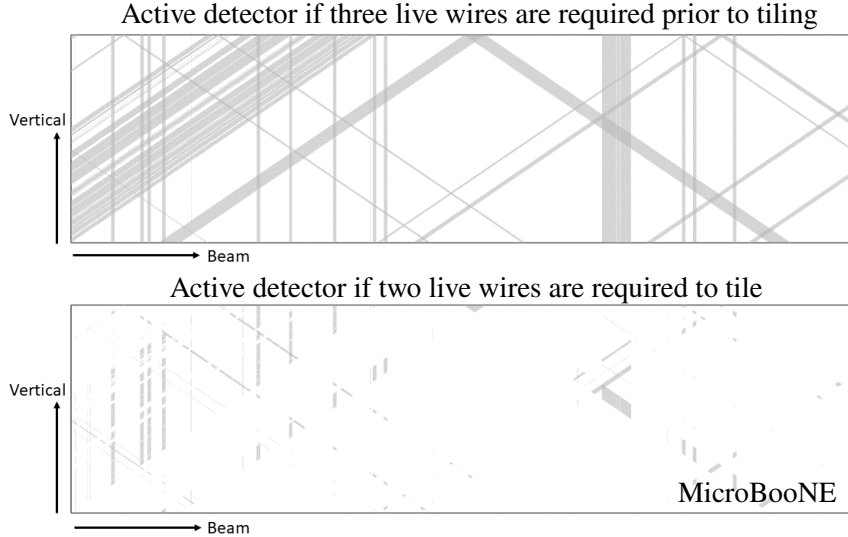
273 Figure 2 corresponds to a single track traversing the time slice in a local area. In reality, there  
 274 could be multiple tracks from cosmic-ray muons or a neutrino interaction traversing the time slice  
 275 (a fixed  $x$  position) at various Y-Z locations as shown in figure 3. The solid red lines represent the  
 276 hit wires from each wire plane. The resulting blobs are marked in blue or green. One may notice  
 277 that in figure 3, the green blobs only have two corresponding wire bundles from two wire planes.  
 278 This is because the hit wires in the third wire plane are not able to provide reasonable signals if  
 279 they are nonfunctional or too noisy. Note that figure 3 is the result after applying the de-ghosting  
 280 algorithm as introduced in section 3.3, so some blobs are determined to be fake and removed.

281 Generally, a 3-plane tiling approach requires the wires from all three wire planes to be func-  
 282 tional. Given that about 10% of channels are nonfunctional in MicroBooNE for various reasons [12],  
 283 this requirement introduces 30% inactive regions on the 2D anode plane as illustrated in the top  
 284 panel of figure 4. To address this issue, we allow for a 2-plane tiling procedure in areas where at least



**Figure 3:** An example event with hit wires and blobs after applying the deghosting algorithm (see section 3.3). Blobs are marked in blue or green. Blue blobs correspond to 3-plane tiling requiring all three wire planes to be functional. Green blobs correspond to the additional blobs created in 2-plane tiling requiring at least two wire planes to be functional. Hit wires are represented by solid red lines.

two planes have functional wires. This means that only the area having two or three nonfunctional wires is regarded as the nonfunctional region. This drastically reduces the nonfunctional volume from 30% to 3% as shown in the bottom panel of figure 4, and an increase of the number of blobs (green blobs) can be seen in figure 3. Outside this 3% nonfunctional region, the 2-plane tiling procedure assumes all the nonfunctional wires are assumed to be hit all the time.



**Figure 4:** Impact of the nonfunctional wires (gray) on the anode plane. The borders of the two figures correspond to the boundaries of the LArTPC active volume. Top: the gray area that has at least one wire nonfunctional is 30%. Bottom: the gray area that has at least two wires nonfunctional is 3%.

While the missing 3-plane blobs are recovered with 2-plane tiling, a number of fake blobs, or

“ghosts”, are created in areas where two functional hit wires cross a third nonfunctional wire, where no true physical charge is responsible for the corresponding wires’ measurements. Some ghosts could still appear when all three wire planes are functional because of the intrinsic ambiguity of the wire-readout scheme, but the number of ghosts is significantly increased when 2-plane tiling is allowed, given the sizable number of nonfunctional wires.

Using the time and geometry information, concatenating the 2D blobs in each time slice from tiling provides a 3D image of all the possible charge depositions, as shown in the example in figure 5. The top panel corresponds to the 3-plane tiling, yielding a 70% functional volume. The middle panel corresponds to the 2-plane tiling, providing a 97% functional volume. Since it is essential to limit the nonfunctional volume in physics measurements to increase the charge collection efficiency and improve the later reconstruction performance, the next task is to remove the ghosts, which originate from the wire readout ambiguity and worsened by nonfunctional wires in the 2-plane tiling procedure.

### 3.2 Charge solving

Charge is one of the most fundamental bases on which to remove the ghosts. A system of linear equations can be constructed by relating the measured charge of a hit wire to the unknown charges of the possible hit cells along this wire. In practice, after the tiling step, blobs and wire bundles are considered here rather than cells and wires. The equation can be expressed as follows:

$$y = Ax, \quad (3.1)$$

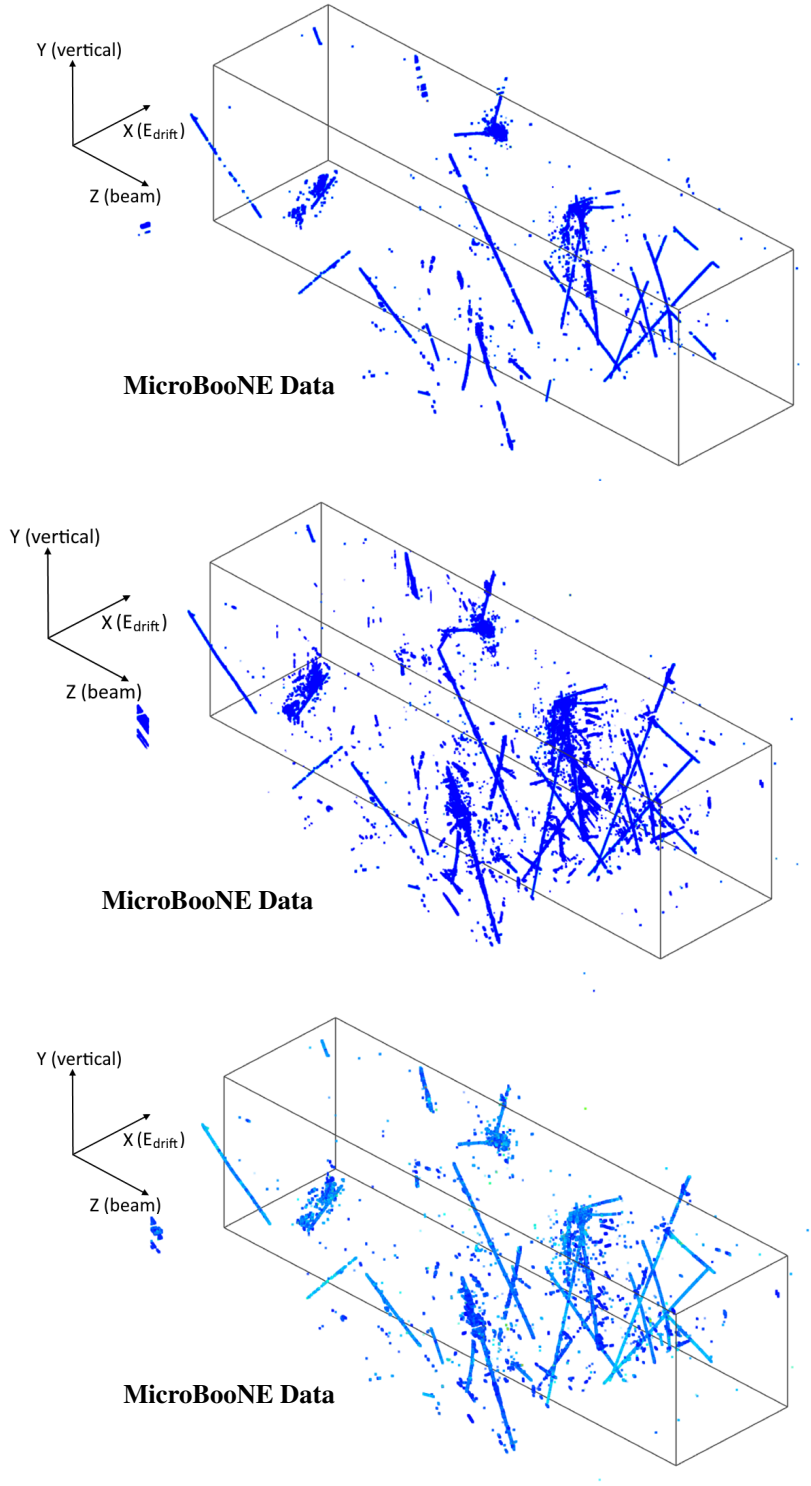
where  $y$  is a vector of the integrated measured charges for the hit wire bundles,  $x$  is a vector of the unknown charges of all blobs, and  $A$  is a matrix with its element  $A_{ij} = 1(0)$  if the blob corresponding to  $x_j$  is (not) on the wire bundle corresponding to  $y_i$ . We call eq. (3.1) the imaging equation of the first principle. In an ideal solution of eq. (3.1), the true hit blob will have charges equal to their truth values, and the fake blobs will have zero charge. However, even if the charges are measured completely and accurately, eq. (3.1) generally has no unique solution. The problem is the result of the fact that there are generally more unknowns than knowns in this system, and this under-determined linear system stems from the wire readout ambiguity. As a consequence, the matrix  $A^T A$  usually does not have full rank and it is not invertible, and the general solution of eq. (3.1),  $x = (A^T \cdot A)^{-1} \cdot A^T \cdot y$ , cannot be used.

As elaborated in ref. [11], one can find an optimized solution to eq. (3.1) by making it an optimization problem after applying additional constraints,

$$\text{minimize } ||x||_p, \quad \text{subject to: } y = Ax, \quad (3.2)$$

where  $||x||_p = (\sum_i |x_i|^p)^{1/p}$  is the  $\ell_p$ -norm of a vector  $x$ . Since the physics activities in LArTPCs are generally sparse, i.e. most of the elements of  $x$  are zero, the  $\ell_0$ -norm (a count of the nonzero elements) can be used to seek the most sparse or the simplest solution that explains the measurements. The minimization of  $||x||_0$  can be achieved by removing the unknowns until the linear equation is solvable. For example in figure 3, there are 25 blobs, while only about 10 hits are true. One can remove 15 unknowns<sup>2</sup> out of the 25 to solve the equation and find the “best” one satisfying

<sup>2</sup>The number of unknowns to be removed is the number of zero eigenvalues of matrix  $A^T A$ .



**Figure 5:** Comparison of the tiling results and the charge solving result from MicroBooNE data (event 41075, run 3493). The solid black box represents the LArTPC active volume with an X-position (converted from the readout time) relative to the neutrino interaction time. Only time and geometry information are used in the tiling. Sparsity, positivity, and proximity information are incorporated in the charge solving as described in section 3.2. Top: 3-plane tiling with 70% active volume. Middle: 2-plane tiling with 97% active volume. Bottom: 2-plane tiling result after the charge solving. The color scale represents the resulting charge values in the charge solving.

the optimization condition. However, in this case there are  $C_{25}^{10} \approx 3.3 \times 10^6$  combinatorial ways to remove the unknowns and in general this optimization is an NP-hard problem that is extremely expensive in computation. Mathematicians [19] have discovered that an alternative constraint, the  $\ell_1$ -norm, can well approximate the  $\ell_0$ -norm result with a much faster minimization. This  $\ell_1$  technique, also known as *compressed sensing*, is widely applied in many other fields for signal processing and computational photography. As shown in section 4.3, the compressed sensing technique is also used to perform the many-to-many charge-light matching.

In practice, a chi-square function is constructed to take into account the uncertainties of the measured charge from signal processing [14], and the compressed sensing technique is implemented by adding an  $\ell_1$ -regularization term to the chi-square function:

$$\chi^2 = \|y' - A'x\|_2^2 + \lambda \|x\|_1, \quad (3.3)$$

where the vector  $y$  and  $x$  are pre-normalized through  $V^{-1} = Q^T Q$  (Cholesky decomposition),  $y' = Q \cdot y$ ,  $A' = Q \cdot A$ , and  $\lambda$  regulates the strength of  $\|x\|_1$ . The matrix  $V$  is the real symmetric covariance matrix of the charge measurement uncertainties. The  $\ell_1$ -regularized chi-square function is convex with a unique global minimum, enabling fast minimization algorithms such as coordinate descent [20]. An implementation of the coordinate descent method can be found in a Wire-Cell git repository [21]. Another constraint, *positivity* of the charge (number of the ionization electrons), is added in the coordinate descent method to help remove the ghosts.

So far we have shown the incorporation of charge, sparsity, and positivity to seek the unique solution to the imaging equation of the first principle. To further improve the robustness of the  $\ell_1$ -regularization result, *proximity* information is incorporated given the fact that the LArTPC is a fully active detector, therefore the measured activities from charged particles are spatially continuous, in contrast to other sampling detectors. For those adjacent blobs over different time slices, the regularization strength  $\lambda$  is applied with an additional scaling factor of  $a^n$  to lower the chance of removing the corresponding element in  $x$  during the  $\ell_1$  minimization.  $n$  represents the number of the adjacent blobs that are connected to the target blob, and  $a$  is a predefined scaling factor. The final chi-square function in the Wire-Cell imaging is transformed to be:

$$\chi^2 = \|y' - A'x\|_2^2 + \lambda \|\omega \cdot x\|_1, \quad (3.4)$$

where  $\lambda$  is an overall regularization strength parameter, and  $\omega_i = a^{n_i}$  is the weight for  $x_i$  as described in the text. The two hyper-parameters  $\lambda$  and  $a$  are tuned by data events. Note that  $y$  is a vector of the integrated measured charge for each wire bundle in the tiling and  $x$  is a vector of the charge to be solved for each blob.

The bottom panel of figure 5 shows the result after applying the charge solving procedure to the 2-plane tiling result in the middle panel. It is clear that the ghosts are further reduced and the 3D voxels are now associated with different charge values, which correspond to the solution  $x$  of the imaging equation of the first principle. As elaborated in section 4.3, such 3D charge solving is critical to predict the scintillation light signals for each PMT, allowing for comparison to and matching with the observed light information.

### 3.3 De-ghosting

The amount of ghosts is considerably reduced after the charge solving but the result is still unsatisfactory. The sparsity combined with the proximity is already incorporated in the charge solving to resolve the wire readout ambiguity; however, this procedure is performed in a “local” manner restricted within each time slice or over adjacent time slices. Within the 3D imaging, all connected blobs in 3D space are grouped together as *proto-clusters*. A proto-cluster does not necessarily group all related TPC activities from a cosmic-ray muon or a neutrino interaction, since there might be true or artificial gaps in the 3D image. The principle of the sparsity of the LArTPC physics activities will be further used in a “global” manner to reconstruct the sparsest 3D images of the TPC activities by removing the less prominent proto-clusters that are redundant to explain the observed 2D-projection measurements from wire planes. Following this philosophy, a dedicated algorithm, *deghosting*, is developed to remove the residual ghosts based on their two main characteristics.

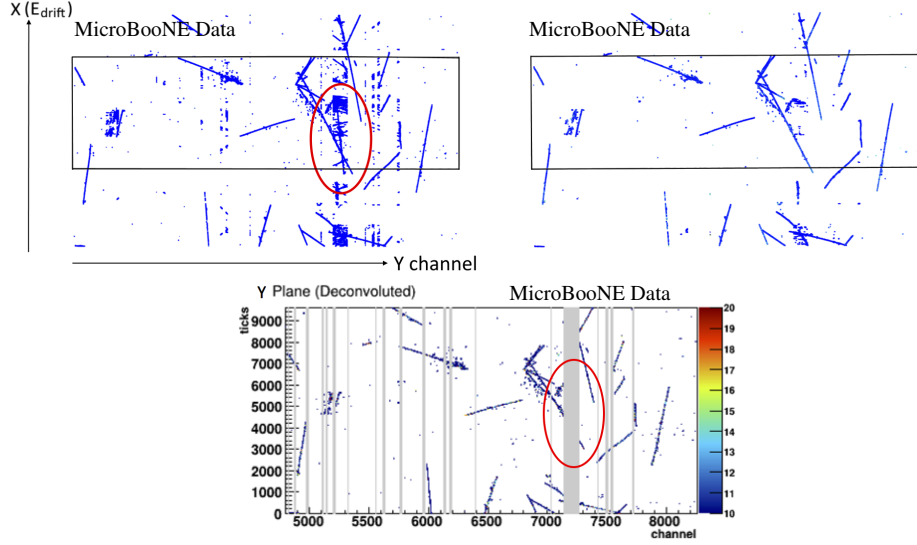
**Position** - the ghosts are mainly present in areas where one wire plane is nonfunctional.

**Projection** - the ghost proto-clusters, mostly track-like, are generally redundant in all three 2D projection views of wire-versus-time.

The area with one nonfunctional wire plane provides significantly less constraints in the tiling and charge solving. This introduces a large ambiguity in the wire readout and a high probability of the presence of ghosts. As indicated in eq. (3.4), the 3D space points are reconstructed by matching the charge for all the functional wire planes, and the charge that forms a ghost proto-cluster must come from the original measurement from a genuine track. Generally speaking, in one of the wire-versus-time views, the ghost tracks are in the nonfunctional region, and match or coincide with genuine tracks in the other two views. So an effective way to identify ghosts is to check each individual wire-versus-time view to test if a proto-cluster is present as redundant pieces or missing pieces of another more prominent proto-cluster.

Below is an example of a MicroBooNE data event to illustrate the identification of ghosts. Figure 6, figure 7, and figure 8 show the 2D projections of the 3D image and the original charge measurements from the three wire planes: Y, U, and V, respectively. In each figure, the top left is the result before de-ghosting and the top right is the result after de-ghosting, and the bottom is the original charge measurement with the nonfunctional wires marked in gray. The red circles in the three figures correspond to the same 3D volume in the TPC. As can be seen, the ghosts in the Y plane’s (collection plane) nonfunctional region overlap with the measurements in the U plane (induction plane), and those ghost proto-clusters are redundant since other proto-clusters can explain the same measurements in the U plane. In figure 7, the images in the red circle are nearly the same before and after the de-ghosting, and it hints that ghost tracks are redundant in terms of explaining the measured charge. The ghosts in the V plane exhibit similar behavior, as shown in figure 8. Note that after one round of the de-ghosting, another round of the charge solving is needed to reclaim the charge carried by the ghosts. The practical 3D imaging procedure is therefore iterative, and is summarized in section 3.4. Figure 9 shows the imaging results with and without de-ghosting.

The occurrence of ghosts is aggravated by the inefficiency of the noise filtering [12] and the signal processing [13], which may filter out some charges along the isochronous tracks as coherent noise, or fail to reconstruct the charges of prolonged tracks (a long signal along the drift



**Figure 6:** Top left: 2D projection to the Y plane’s wire-versus-time view of the reconstructed 3D image without the de-ghosting algorithm. The black box represents the full detector length in Y and the full cathode-to-anode drift distance in X. The red circle corresponds to the same volume in the TPC as in figure 7 and figure 8. Top right: after the de-ghosting algorithm. Bottom: Original charge measurement. The vertical axis bin width (time) is 4 ticks (2 microseconds), and the color scale represents the number of ionization electrons scaled by a factor 1/500 (comparable to ADC counts from raw waveforms). The nonfunctional wires are marked in gray.

direction) because of the bipolar cancellation of the induction plane signals<sup>3</sup>. Consequently, one or two of the 2D wire-versus-time views of the charge measurements may have gaps along a track even on the functional wire planes. This gap will lead to a separation in the 3D image since the charge measurements across the wire planes can no longer match. Consequently, the successfully reconstructed charges from the other wire planes corresponding to the gap could interplay with the charge measurements from other tracks and be erroneously explained by ghosts. The removal of such ghosts requires a bridging of the gaps to connect the separated pieces of the track. This will be further discussed in section 4.1 in the context of 3D clustering, which results in the final TPC clusters based on the proto-clusters.

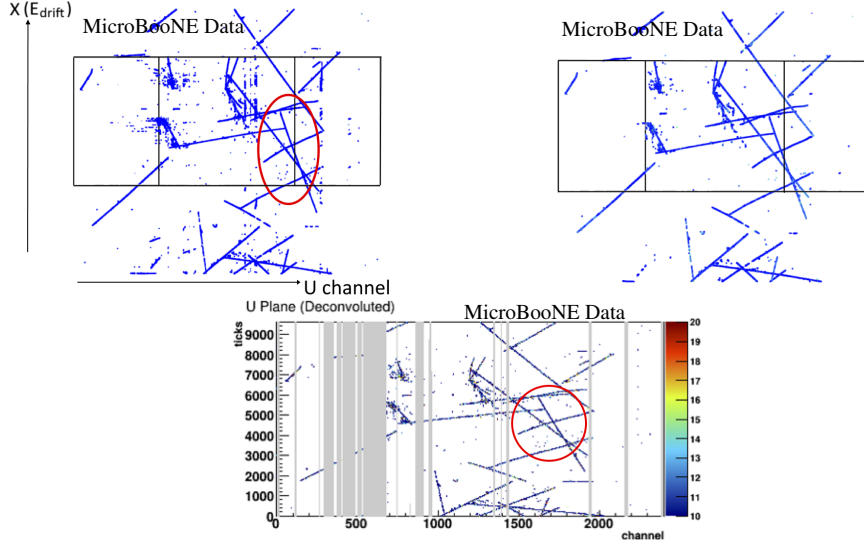
### 3.4 Summary

The actual procedure of the application of Wire-Cell 3D imaging in MicroBooNE is iterative, containing multiple rounds of tiling, charge solving, and de-ghosting. The number of iterations is based on an empirical evaluation based on data events used during the algorithm development. More iterations do not appear to significantly improve the results of the 3D imaging. A summary of procedures is shown in table 1.

Since MicroBooNE is a near-surface detector with limited cosmic ray shielding, 20–30 cosmic-ray muons per event are input to the Wire-Cell imaging process in the full readout window of 4.8 ms.

<sup>3</sup>The recent advancement in TPC signal processing by leveraging the deep learning techniques[22] is expected to reduce this signal inefficiency.





**Figure 7:** Top left: 2D projection to the U plane’s wire-versus-time view of the reconstructed 3D image without the de-ghosting algorithm. The black box represents the full array of U channels and the full cathode-to-anode drift distance in X. The red circle corresponds to the same volume in the TPC as in figure 6 and figure 8. Top right: after the de-ghosting algorithm. Bottom: Original charge measurement. Y-axis bin width (time) is ticks (2 microseconds), and Z-axis value represents the number of ionization electrons scaled by a factor 1/500 (comparable to ADC counts from raw waveforms). The nonfunctional wires are marked in gray.

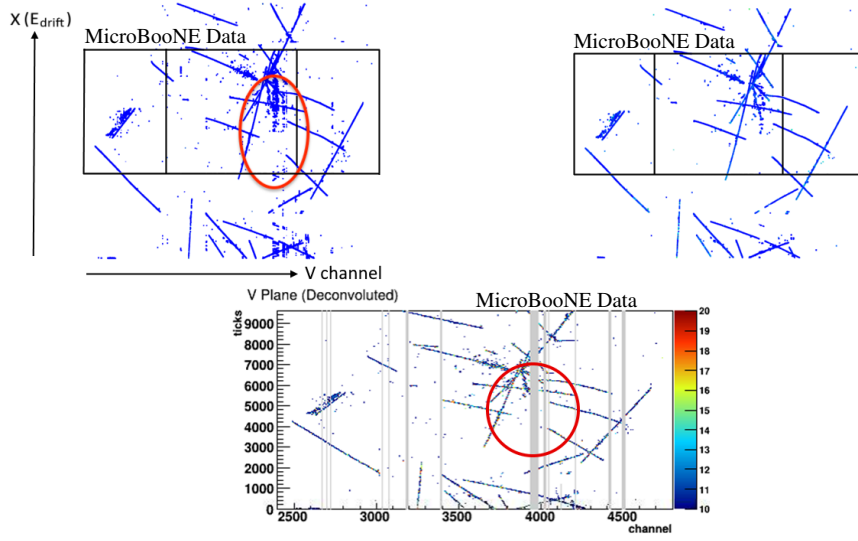
**Table 1:** Summary of the procedures of the Wire-Cell 3D imaging, including the 2-plane ( $\geq 2$  wire planes) tiling, the charge solving, and the de-ghosting.

Step	Description
1	2-plane tiling
2	De-ghosting
3	1st round of charge solving
4	2nd round of charge solving with reweighting for connected blobs
5	Repeat the steps 2, 3, 4
6	Repeat the steps 2, 3, 4 again

The time and memory consumption are practical issues to be addressed in the optimization and finalization of the algorithms. Using  $\sim 10k$  MicroBooNE data events, the average time and memory consumption (a single-threaded program) is estimated to be about 2 minutes and less than 2 GB on an Intel(R) Core(TM) i7-4790K CPU @ 4.00GHz. Most of the memory is used by the tiling to initialize and index the blobs from each time slice. Most of the time is consumed by the charge solving and de-ghosting, which are critical to the quality of the 3D images.

The goal of the Wire-Cell imaging is to reconstruct the 3D image of the ionization electrons independently of the event topology and prior to the application of pattern recognition techniques



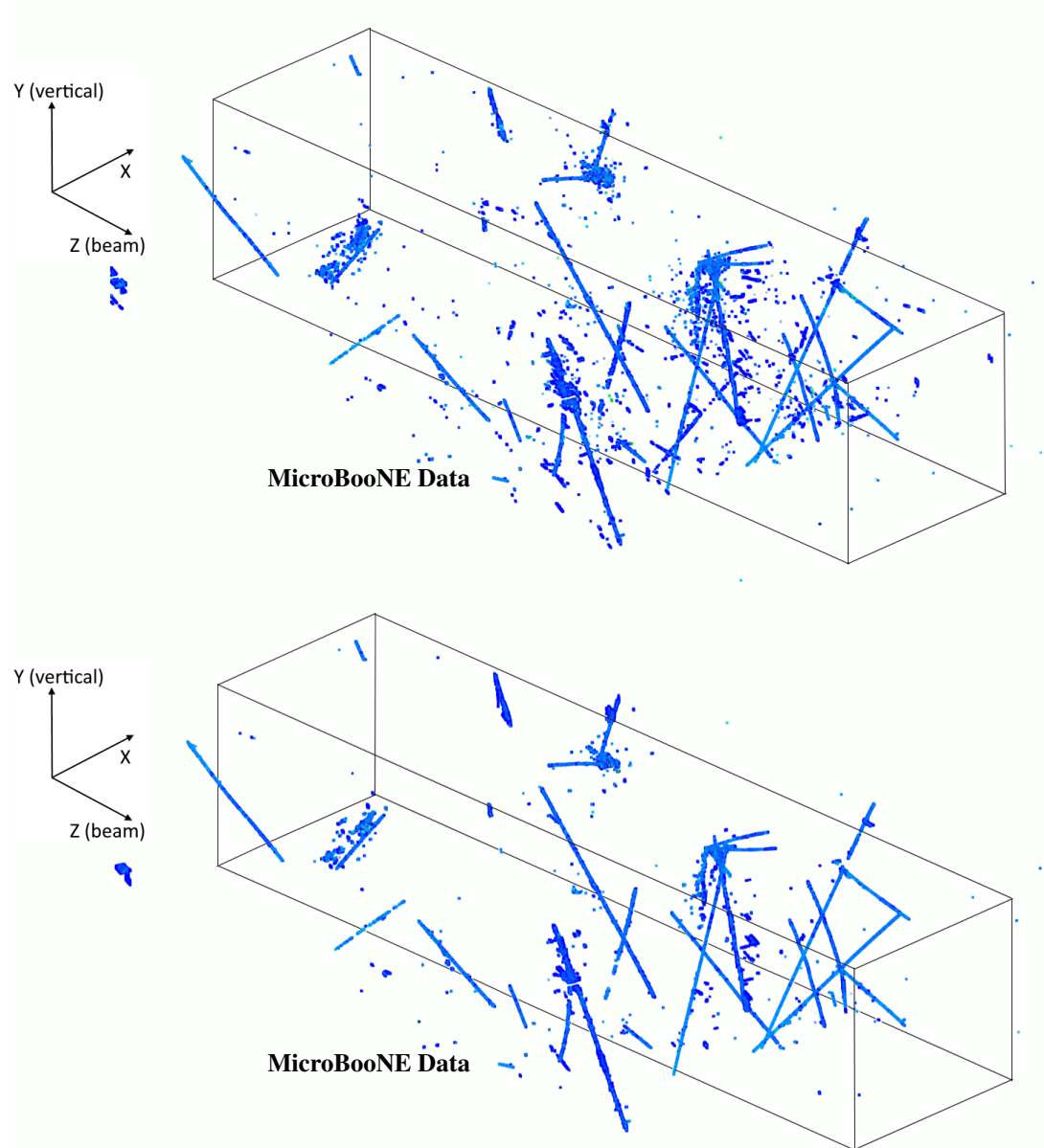


**Figure 8:** Top left: 2D projection to the V plane’s wire-versus-time view of the reconstructed 3D image without the de-ghosting algorithm. The black box represents the full array of V channels and the full cathode-to-anode drift distance in X. The red circle corresponds to the same volume in the TPC as in figure 6 and figure 7. Top right: after the de-ghosting algorithm. Bottom: Original charge measurement. Y-axis bin width (time) is ticks (2 microseconds), and Z-axis value represents the number of ionization electrons scaled by a factor 1/500 (comparable to ADC counts from raw waveforms). The nonfunctional wires are marked in gray.

(such as those presented in Ref. [8]). The reconstructed 3D image is an input to the subsequent reconstruction, e.g. the charge-light matching, to distinguish the in-beam neutrino candidate from the cosmic-ray backgrounds. The 3D charge associated with each reconstructed space point is used in the prediction of PMT light signals.

Isochronous tracks present a common problem in the LArTPC 3D imaging, as the wire readout ambiguity is drastically increased in the time slice containing them. In Wire-Cell 3D imaging, this issue is mitigated by introducing tiling. On the other hand, the blobs of the isochronous tracks are significantly broadened, leading to a much worse 2D spatial resolution within the time slice, i.e. in the nominal Y-Z projection view, or the U-Z and V-Z wire plane views. An example can be found in figure 11. Improvement of spatial resolution can be achieved via trajectory fitting in a later reconstruction stage. This is beyond the scope of this paper and will be presented in a future publication [15].

Because of the existence of nonfunctional channels, a 2-plane ( $\geq 2$ ) tiling strategy is adopted to significantly enhance the image reconstruction efficiency at the cost of introducing more ghosts. Time, geometry, charge, sparsity, positivity, and proximity information is utilized to overcome the wire readout ambiguity and to remove the ghosts. In addition to the de-ghosting steps performed during 3D imaging, another round of de-ghosting is performed in the clustering stage as discussed in section 4.1.3. Quantitative evaluations of Wire-Cell 3D imaging in various cases are presented in section 5.



**Figure 9:** Comparison of the 3D imaging results from MicroBooNE data (event 41075, run 3493) without (top) and with (bottom) the de-ghosting algorithm. Ghosts are significantly reduced after the de-ghosting. The solid black box represents the LArTPC active volume with an X-position (converted from the readout time) relative to the neutrino interaction time. The color scale indicates the charge density.

## 4 Matching Charge and Light

As introduced in section 1, each triggered event in MicroBooNE contains a 4.8 ms TPC readout and a 6.4 ms PMT readout. The Wire-Cell imaging reconstructs a 3D image of the TPC activities, which includes both cosmic-ray muons and a neutrino interaction if present. The PMTs detect the scintillation light on a much shorter timescale than the drifting of the ionization electrons in the TPC, so it can be used to provide the interaction (start) time once it is paired with the corresponding charge signals. The 32 PMTs' waveforms from a cosmic-ray muon or a neutrino interaction are processed to reconstruct a *flash*, which is a group of the PMT signals close in time (e.g. within 100 ns). The detailed definition of a PMT flash can be found in section 4.2. Typically, the cosmic-ray muon rate is 5.5 kHz in the TPC active volume, so there are 20-30 cosmic-ray muons within the 4.8 ms TPC readout window. Within the 6.4 ms PMT readout window, there are 40-50 PMT flashes which correspond not only to the activities inside the TPC but also those outside the TPC but within the LAr volume inside the cryostat.

As described in the previous section, the proto-cluster in the 3D imaging step is solely based on proximity, while a physical signal initiated by a primary particle's interactions could have disconnected pieces, such as from secondary neutral particles or because of imperfect signal processing or reconstruction. In order to accurately and robustly pair the TPC activities to the PMT flashes, an interaction 3D clustering is developed to group the proto-clusters further into a TPC *cluster*, which then represents signals initiated by an individual primary particle such as from a cosmic-ray muon or a neutrino interaction.

Given the TPC clusters and PMT flashes, a novel algorithm, *many-to-many charge-light matching*, is developed to match the clusters and the flashes simultaneously based on the predicted light signals generated by the 3D TPC clusters and the measured light signals from PMT flashes. The TPC cluster(s) matched to an in-beam PMT flash is then regarded as a beam neutrino candidate. All the remainders are rejected as cosmic-ray muons. Compared to a previous single-to-single track-light-matching algorithm as described in ref. [23], many-to-many charge-light matching enhances the cosmic rejection power and results in a cleaned-up 3D image of the neutrino activities.

The algorithms of the 3D clustering and the PMT light reconstruction are delineated in section 4.1 and section 4.2, respectively. The details of the many-to-many charge-light matching procedure are described in section 4.3.

### 4.1 3D clustering

Clustering as described in this section aims to group proto-clusters according to their physics origin into clusters. This step is an initial separation of neutrino and cosmic activities, and is necessary to efficiently perform the subsequent many-to-many charge-light matching.

Proto-clustering, which solely relies on proximity, has been carried out in the 3D imaging step (section 3.3). However, it doesn't meet the requirement of carrying out a high performance charge-light matching because of the following issues:

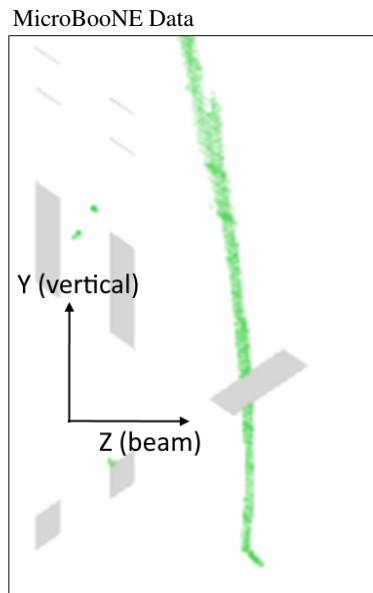
**Gaps:** The presence of gaps compromises the effectiveness of a proto-clustering based on proximity. A gap mainly results from: 1) the  $\sim 3\%$  nonfunctional regions, as shown in figure 10; 2) incorrect removal of parts of the isochronous tracks (close to parallel to the wire planes) by the

coherent noise filter, as shown in figure 11; and 3) failures of the signal processing for parts of the prolonged tracks (a long signal along the drift direction) as shown in figure 12,

**Coincidental overlap:** For LArTPCs operating near the surface (such as MicroBooNE), the detector is bombarded by a large number of cosmic-ray muons. Although the cosmic-ray muons generally pass through the detector at different time and locations, the 3D images from different TPC clusters, e.g. two muons, can appear to be connected when ionization electrons of different activities arrive at the same location of the wire plane at the same time. This leads to an over-clustering of space points, causing mistakes in the charge-light matching.

**Residual ghosts:** The de-ghosting algorithm described in section 3.3 is not completely sufficient because of the incomplete or improper proto-clustering as the two items explained above.

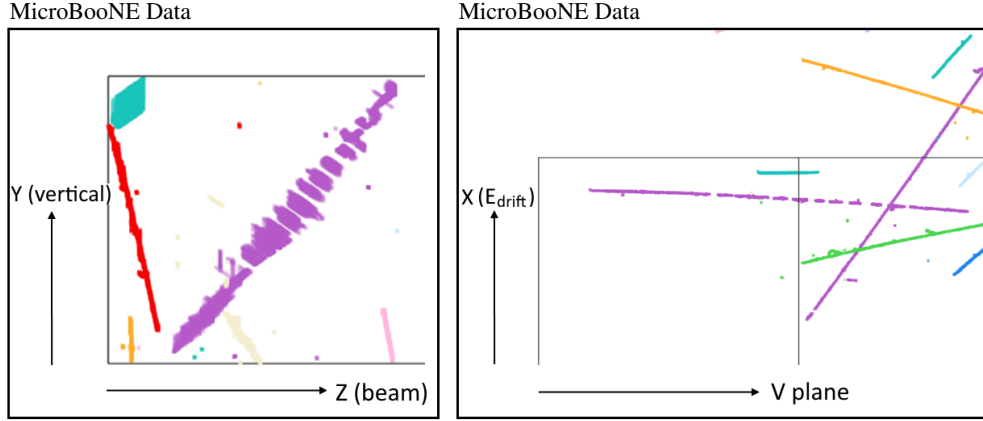
**Separated clusters from a neutrino interaction:** Neutral particles from neutrino interactions with argon nuclei are very likely to travel some distance before depositing their energy. The secondary charged particles from these neutral particles are therefore separated from the neutrino primary vertices. For example, a  $\pi^0$  is a potential final state particle of a neutrino interaction with an argon nucleus. It generally deposits its energy through two electromagnetic (EM) showers from its decay  $\gamma$ 's. The two  $\gamma$ 's are in principle detached from the neutrino primary vertex and other final state particles. A dedicated algorithm is needed to group these separated particles from the primary interaction into a single cluster.



**Figure 10:** Zoomed in Y-Z view of a cosmic muon with a gap because of the nonfunctional regions. The nonfunctional regions are shown in gray.

#### 4.1.1 Clustering in the presence of gaps

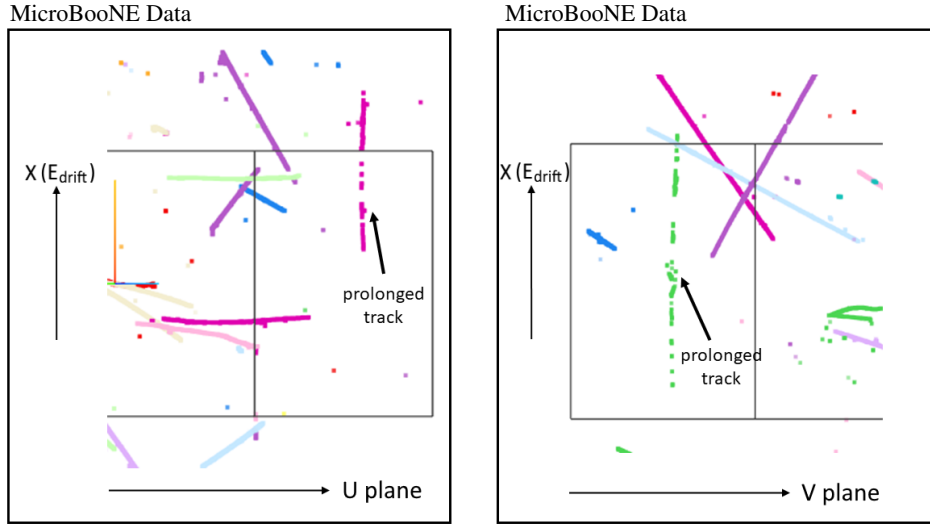
Clustering across gaps mainly relies on two sets of information: distance and directionality. If two proto-clusters are close to each other along a line, the gap may be bridged and the two proto-clusters are grouped into a single cluster. Many existing tools and algorithms operating on a point



**Figure 11:** Gaps along an isochronous track in different zoomed-in 2D views. (Left) Y-Z view of an isochronous track (magenta) from a MicroBooNE data event. (Right) X-V view of the same event. V plane direction represents the wire pitch direction of the V wire plane. The black lines in the inner figures correspond to the boundaries of the 3D LArTPC active volume. Since the 3D boundaries are projected to the 2D visual shown, sometimes edges of the rectangular prism active volume appear in the center of the image. Cluster membership is indicated by uniform color within each plot. Some distant clusters could be marked in the same color because of a finite number of visibly distinctive colors available in the event display.

cloud (a collection of many 3D points) can be directly used, as a TPC cluster is a collection of the reconstructed 3D space points. The distance between two clusters (point clouds) is defined as the minimal distance between a pair of space points, one from each respective point cloud. To calculate this distance rapidly, the k-d (k-dimensional) tree based algorithm as implemented in the “naoflann” package [24] is employed. Once the minimal distance and its direction are obtained, its direction is compared with the directions of the two proto-clusters. These are found using a voting scheme inspired by the Hough transformation [25]. The directional vector, parameterized by a polar angle and an azimuthal angle, is calculated for each point. The most probable value in the 2D distribution of polar-versus-azimuthal is then taken as the primary direction of the point cloud. Given the minimal distance vector and the two proto-cluster directions, the two proto-clusters are grouped (or not) based on their distance and consistency in directions. In practice, a set of criteria are developed and optimized by analyzing hundreds of data events from various topologies.

Figure 13 shows a comparison of the results before and after applying this clustering algorithm. Separate proto-clusters are successfully grouped into individual clusters. Each cluster is marked by a different color. Some distant clusters are properly separated but they are in the same color because of a finite number of visibly distinctive colors available in the event display. In the bottom panel of figure 13, one can find that there are still other clustering issues. For instance, there are two connected cosmic-ray muons and an incomplete neutrino cluster as indicated by the two black



**Figure 12:** Gaps along prolonged tracks in different zoomed-in 2D views. (Left) X-U view for a prolonged cosmic muon track (magenta) from a MicroBooNE data event. (Right) X-V view for a prolonged cosmic muon track (green) from another MicroBooNE data event. The black lines in the inner figures correspond to the boundaries of the 3D LArTPC active volume. Since the 3D boundaries are projected to the 2D visual shown, sometimes edges of the rectangular prism active volume appear in the center of the image. Cluster membership is indicated by uniform color within each plot. Some distant clusters could be marked in the same color because of a finite number of colors available in the event display.

529 circles. These are dealt with using additional clustering algorithms as introduced in section 4.1.2  
 530 and section 4.1.4. The resulting clusters are presented in figure 19.

#### 531 4.1.2 Separation of coincidental overlap clusters

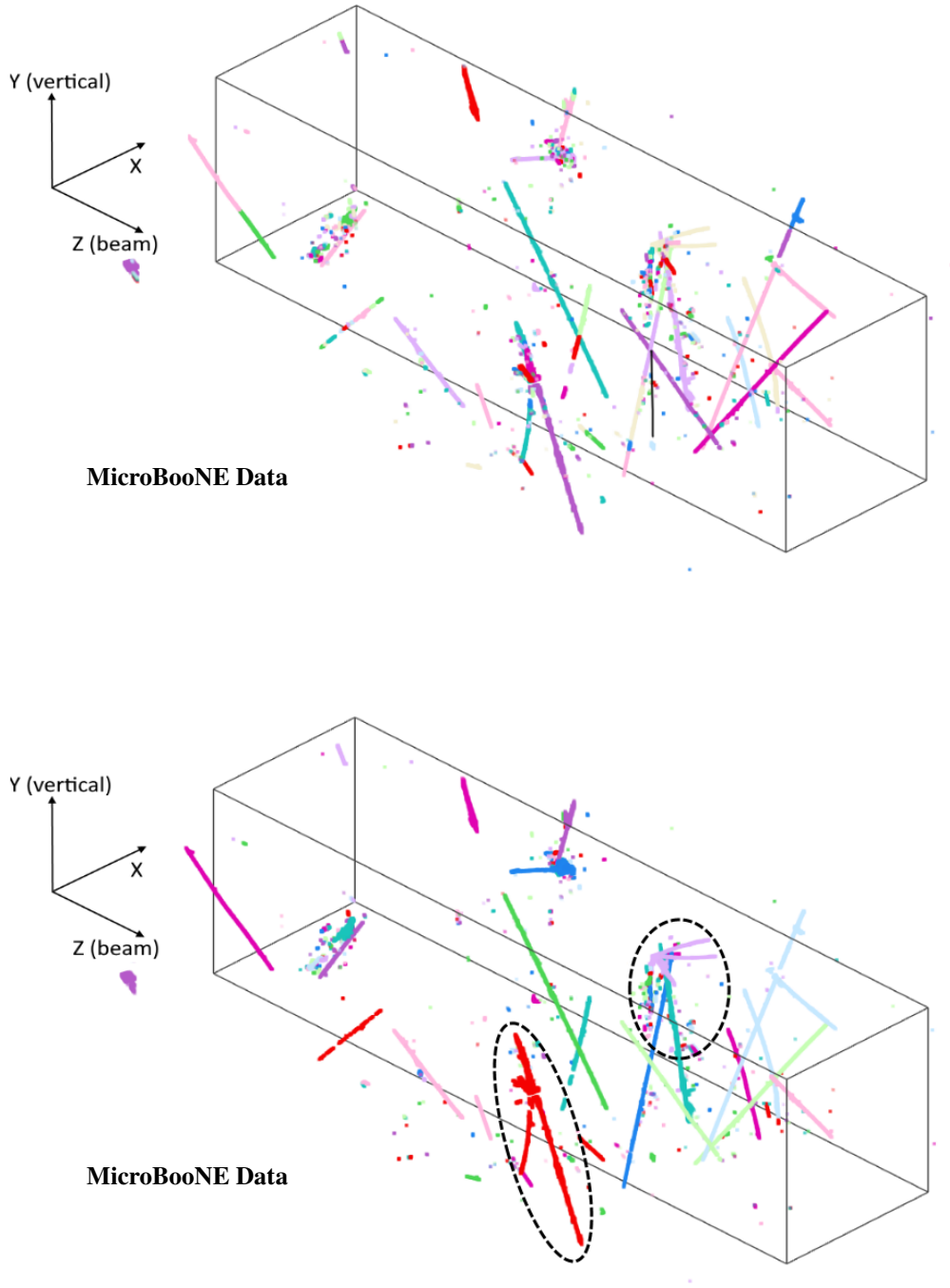
532 In this section, we describe the algorithm to separate a “coincidental overlap” cluster, and the steps  
 are summarized in table 2.

**Table 2:** Steps of the separation of a “coincidental overlap” cluster.

Step	Key operations
1	Identification of “coincidental overlap” cluster
2	Find two end points of a primary track
3	Form trajectory of this primary track
4	Collect space points of this primary track
5	Remove this primary track and repeat this procedure

533

534 The first step is to identify the “coincidental overlap” cluster. Principle component analysis  
 535 (PCA) is performed on each cluster after the bridging of gaps as described in section 4.1.1. For a



**Figure 13:** Demonstration of the effectiveness of the algorithm of bridging gaps. The solid black box represents the LArTPC active volume with an X-position (converted from the readout time) relative to the neutrino interaction time. Top: proto-clusters solely based on proximity. Bottom: clusters after the application of the algorithm of bridging gaps. The two circles indicate remaining clustering issues, e.g. over-clustering of cosmic-ray muons and under-clustering of neutrino interactions. Cluster membership is indicated by uniform color.



single-track-like cluster, only the primary component (axis) of the PCA has a significantly larger eigenvalue in the data correlation matrix. This is generally not true for a “coincidental overlap” cluster in which two or more tracks are crossing. Once a candidate “coincidental overlap” cluster is identified, the sub-clusters representing different physical interactions are to be identified and separated one by one.

The separation of each sub-cluster starts with identifying the two end points of a primary track in this cluster. A primary track is the one that best matches one of the primary PCA axes, i.e. the longest along this primary PCA axis. Firstly, the quickhull [26] algorithm operates on the 3D space points of a coincidental overlap cluster to obtain the 3D convex hull, which is the smallest convex shape that contains all the space points. The two end points of the current primary track must be contained or in close proximity with the convex hull’s vertices. Secondly, the nearby points around each convex hull’s vertex are grouped together to form test clusters. The largest test clusters are used to discover the end points of the primary track, and this requires 1) a small distance to the PCA primary component; 2) a consistent direction of the test cluster with the PCA primary component. In general, such end points can always be found for a prominent cluster. Once an end point is identified, a Kalman-filter-based technique is used to crawl along this primary track until the other end point is determined. Given the two end points, the trajectory of this primary track is obtained using a graph theory operation, the Dijkstra’s shortest path [27]. The connected component algorithm from graph theory is then used to collect the space points associated with this trajectory and form a sub-cluster. After removing this sub-cluster from the current primary track, the remaining cluster is further examined and sub-clusters are removed until only one primary track is left. Each removed sub-cluster is taken as an individual cluster in the end.

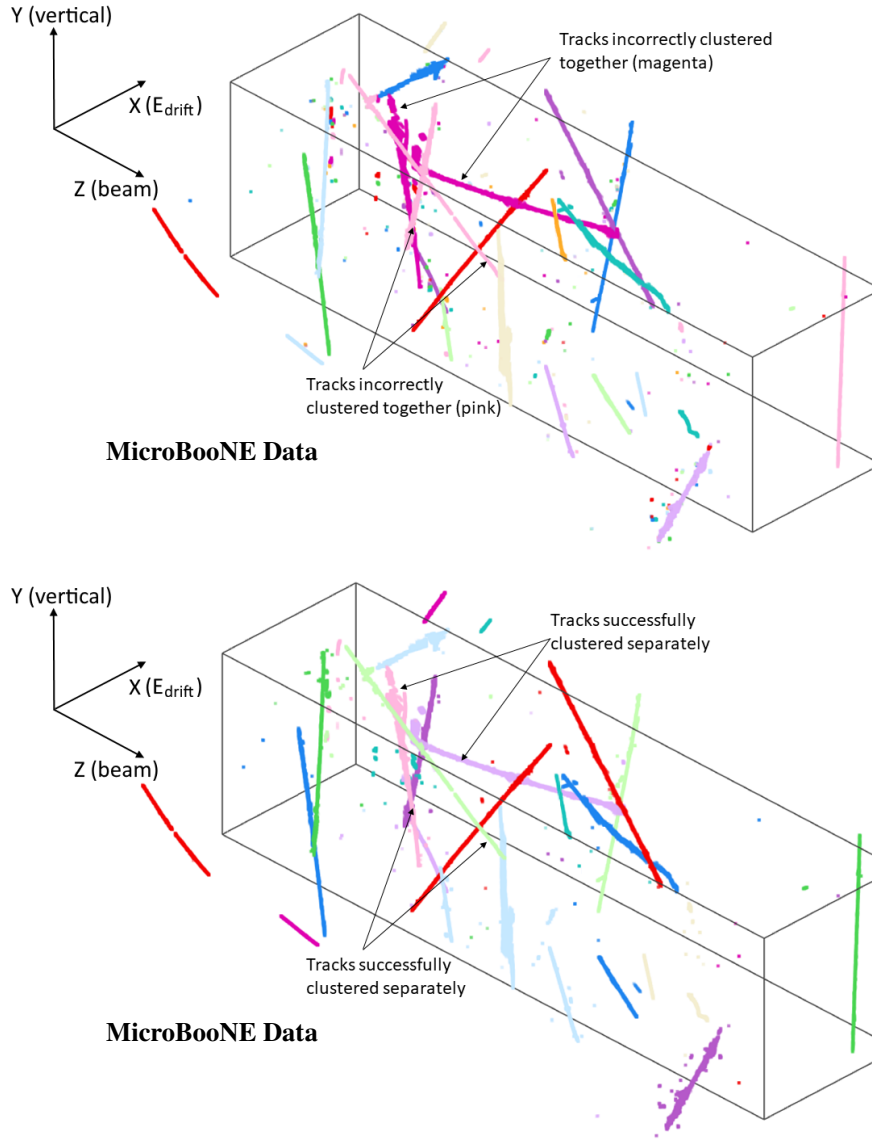
Figure 14 shows a comparison of the results before and after applying the separation algorithm. Two “coincidental overlap” clusters show up in this event: one case has two cosmic-ray muons crossing each other, the other has a cosmic-ray muon grouped to a neutrino interaction. Figure 15 shows another example, where two cosmic-ray muons cross each other and one of the muons induces an EM shower. After the separation step, part of the EM shower is improperly separated. This could be addressed by the many-to-many charge-light matching later, which can further group the clusters that as a whole match the same PMT flash. Note that a cluster of a neutrino interaction with multiple final state particles might be incorrectly identified as a “coincidental overlap”. There is a protection against over-separating clusters because the neutrino final state particles are mostly forward-going along the beam direction while most cosmic-ray muons are pointing downward. Additionally, a dedicated clustering algorithm to group the separate clusters from the same neutrino interaction is performed later as described in section 4.1.4.

### 4.1.3 Further de-ghosting

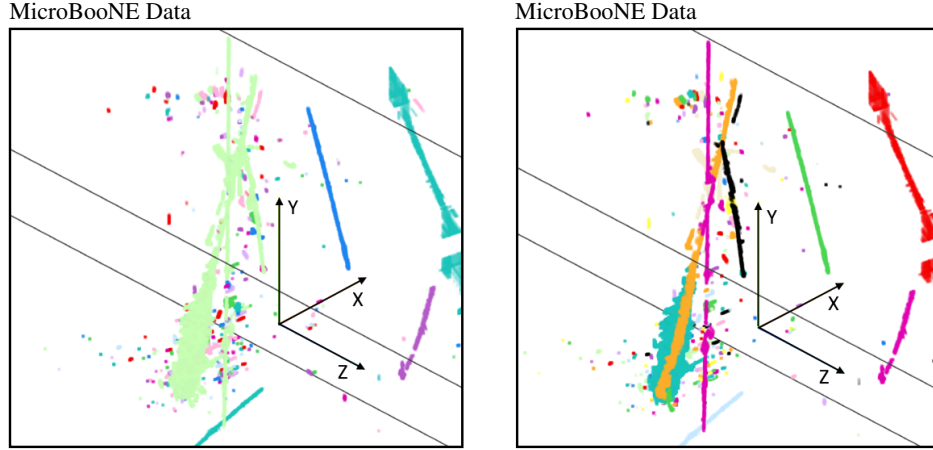
As mentioned in section 3, a de-ghosting algorithm is applied in the 3D imaging stage to remove ghosts. This is done prior to clustering. This strategy is initially inefficient since a proto-cluster cannot appropriately represent complete TPC activities initiated by a primary particle’s interaction. Given the improvements during the clustering stage as described above, the de-ghosting algorithm is run on the resulting clusters again to remove the residual ghosts.

We present some instructive examples of de-ghosting after clustering has been performed. As shown in the top panel of figure 16, there are some ghosts due to gaps along the prolonged tracks.





**Figure 14:** Demonstration of the effectiveness of the clustering algorithm to separate a “coincidental overlap” cluster. The solid black box represents the LArTPC active volume with an X-position (converted from the readout time) relative to the neutrino interaction time. The top and bottom panels show the clusters before and after applying this algorithm. Cluster membership is indicated by uniform color.



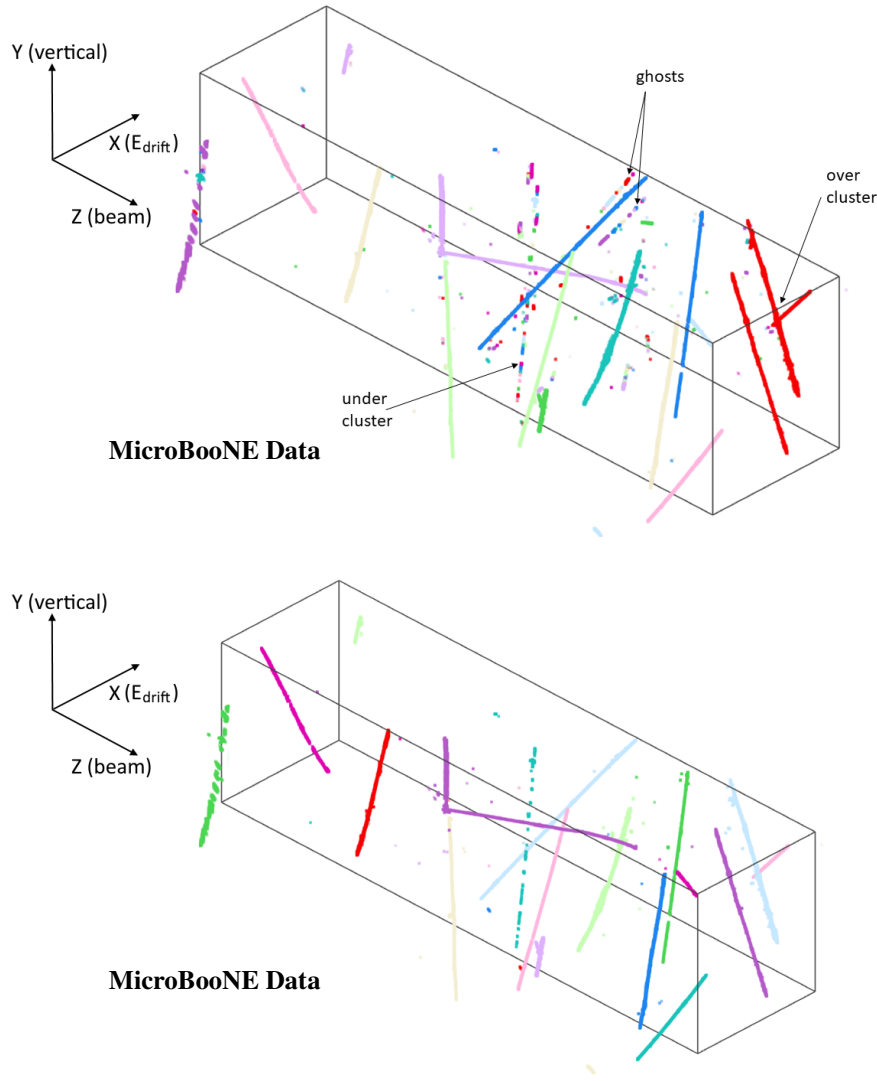
**Figure 15:** Demonstration of the effectiveness of the clustering algorithm to separate a “coincidental overlap” cluster. Black lines in each subfigure correspond to the boundaries of the LArTPC active volume. The left and right panels show the clusters before and after applying this algorithm. The all-light-green 3D cluster in the left panel is broken into its components in the right panel.

These ghosts cannot be removed during the 3D imaging since they are the only explanation of the charge measurements in functional wire planes. With a bridging of the gaps, the original proto-clusters are grouped into a larger cluster, which as a whole can explain the charge measurements in all three wire planes. The ghosts related to gaps in this prolonged track can thus be removed. Another example as shown in figure 17 has a four-track cluster, in which two tracks are ghosts. This cluster is present in the region where there is a nonfunctional wire plane. Since all of these four tracks including the ghosts are connected, the two ghost tracks survive the de-ghosting procedure in the 3D imaging stage. After the application of the algorithm to separate the “coincidental overlap” cluster, the two ghost tracks are identified and removed individually.

Figure 18 shows an example of a complex event with a large number of residual ghosts after imaging. The ghosts are indicated by black arrows in the top panel of this figure. Many tracks including prolonged tracks and isochronous tracks go through the region (area on the left of figure 4) where U wires are mostly nonfunctional. Ghosts with various lengths and positions are reconstructed. After bridging the gaps and separating coincidental overlap clusters, the number of ghosts is significantly reduced by re-running the de-ghosting algorithm.

#### 4.1.4 Clustering for neutrino events

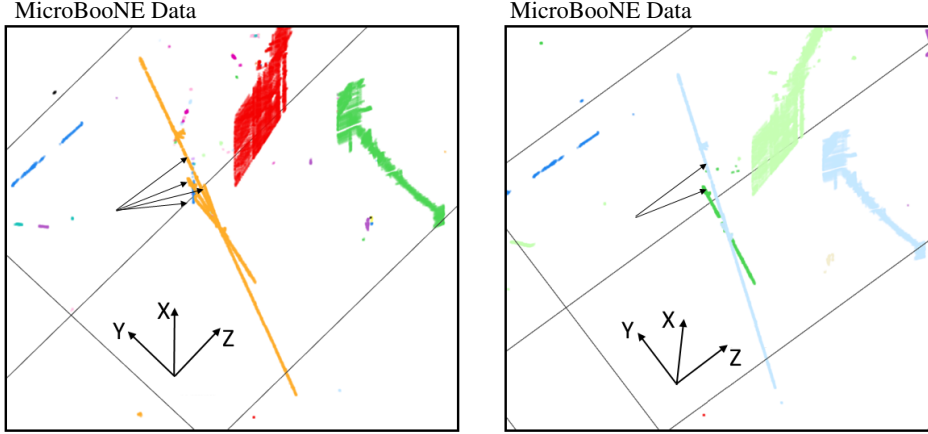
In this section, we describe a dedicated clustering algorithm to group separate clusters from the same neutrino interaction. Generally, the neutral particles from neutrino interactions, such as a neutron or  $\pi^0$ , can lead to clusters that are detached from the primary neutrino interaction vertex. These clusters are truly separated in 3D space and should be identified and grouped properly. In order to do so, the major task is to find the common vertex based on the direction of each sub-cluster.



**Figure 16:** Demonstration of the effectiveness of the de-ghosting algorithm with other advanced clustering algorithms applied. The solid black box represents the LArTPC active volume with an X-position (converted from the readout time) relative to the neutrino interaction time. The top and bottom panels show the clusters before and after applying the de-ghosting algorithm after bridging gaps. Color indicates cluster membership.

The operations to obtain the primary direction, find extreme points, associate nearby points, and calculate the direction, are the same as those introduced in the previous sections. The main steps are described below:

- Only clusters within the drift window that corresponds to the beam time are considered.
- The direction of each sub-cluster is calculated. End points are examined to ensure that they do not belong to any isolated dot-like (less than 1-cm length) clusters, which are ignored because of their small size.
- Each cluster is extended with virtual space points along the track direction near each end



**Figure 17:** Demonstration of effectiveness of the de-ghosting algorithm with other advanced clustering algorithms applied. The black lines inside each subfigure correspond to the boundaries of the LArTPC active volume. The left and right panels show the clusters before and after applying the de-ghosting algorithm following the separation of the “coincidental overlap” cluster. Color indicates cluster membership. The large clusters with much worse spatial resolution in Y-Z view correspond to big blobs of isochronous tracks as discussed in section 3.4.

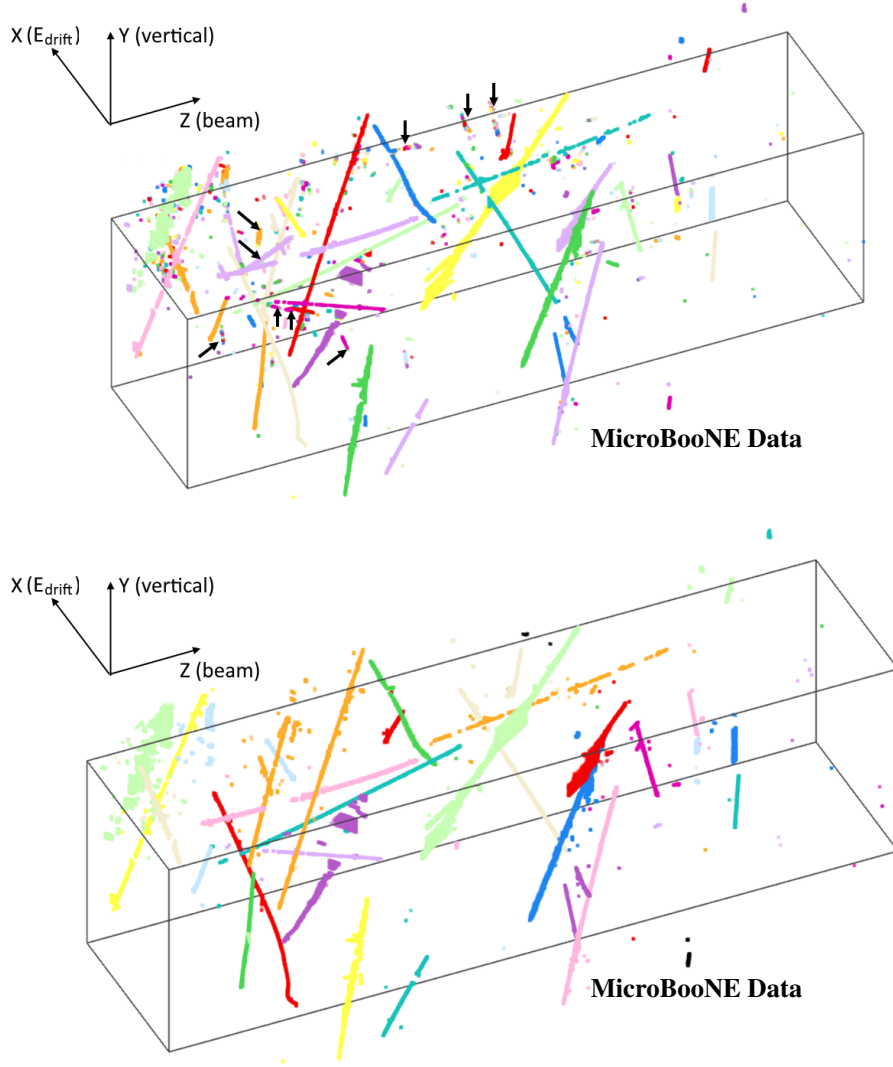
point.

- The extended clusters are examined to find the “intersection” point with other clusters. This “intersection” is required to be formed by the extended part or the end points of the other clusters.

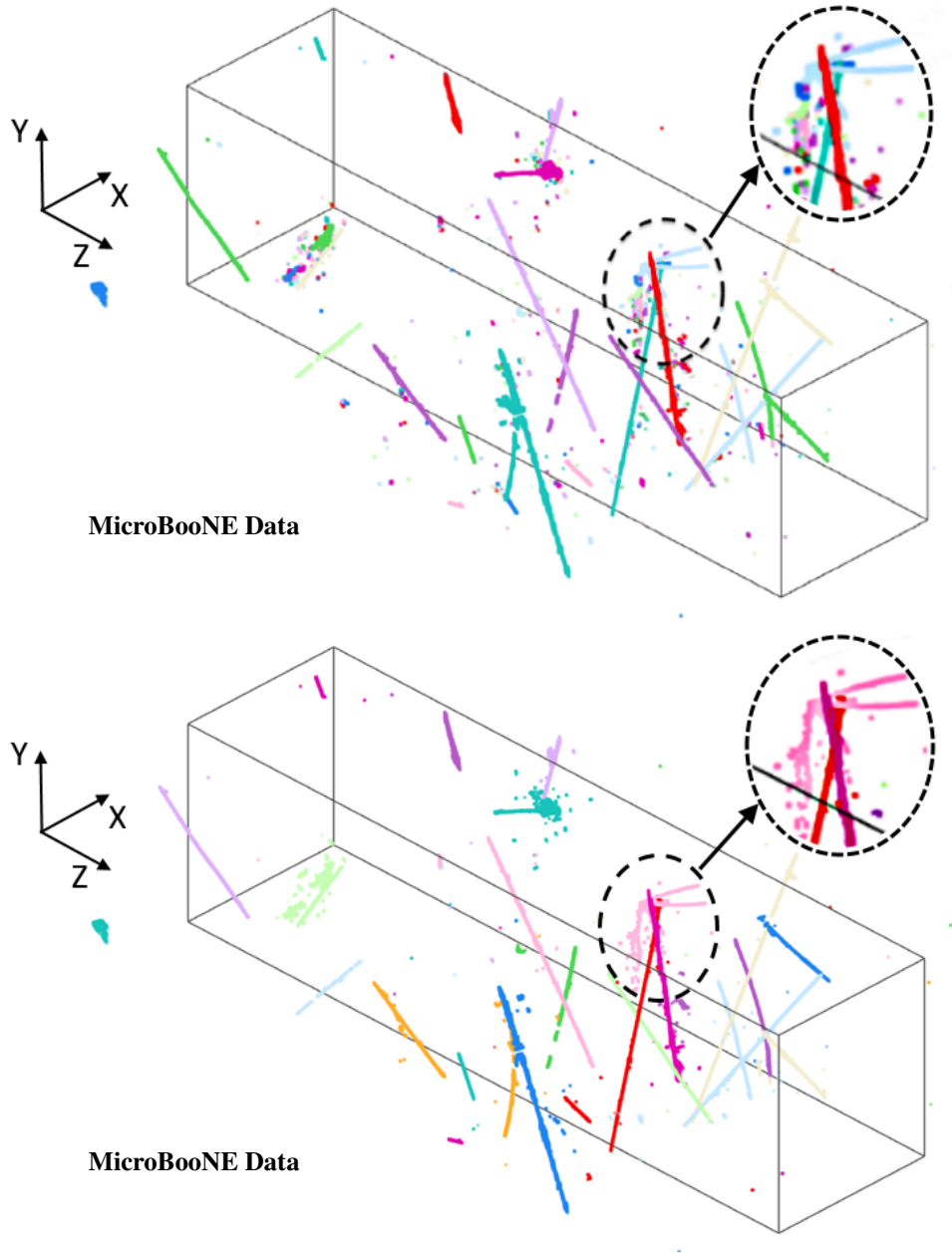
The “intersection” is not necessarily the primary neutrino vertex, as the separated clusters from the secondary interaction vertex are also expected to be grouped together. An under-clustering issue may arise for neutrino interactions, but this is expected to be addressed by the charge-light matching step later when a many-to-many matching strategy is adopted. Figure 19 shows an example of a complex neutrino interaction. Two  $\gamma$ ’s from a  $\pi^0$  decay and a detached charged particle are clustered properly.

## 4.2 PMT light signal reconstruction

TPC clusters, which represent grouped TPC activities corresponding to either cosmic-ray muons or a neutrino interaction, are formed by the clustering algorithms as described in the previous section. Because of the asynchrony of the TPC readout system with the PMT readout system, TPC activities are mixed in the time sequence with an unknown interaction (start) time. Scintillation light is produced and detected on a much shorter time scale by the spatially distributed PMTs. An offline processing of the light signals from PMTs is thus important to perform the many-to-many charge-light matching to select the neutrino activities corresponding to the in-beam PMT signals that coincide with the beam spill.



**Figure 18:** Demonstration of the effectiveness of the de-ghosting algorithm with other advanced clustering algorithms applied. The solid black box represents the LArTPC active volume with an X-position (converted from the readout time) relative to the neutrino interaction time. The dot-like clusters and superfluous clusters off the main trajectories are generally ghost tracks. The top and bottom panels show the clusters before and after applying the de-ghosting algorithm. The example ghosts in the top panel are indicated by the black arrows. This is a challenging case where multiple tracks go through a region where one wire plane (U plane) is largely nonfunctional. Color represents cluster membership.



**Figure 19:** Demonstration of the effectiveness of the clustering algorithm designed for neutrino interactions. The solid black box represents the LArTPC active volume with an X-position (converted from the readout time) relative to the neutrino interaction time. Top and bottom panels show the clusters before and after applying the clustering algorithm. The neutrino interaction (the light pink cluster) is in the black dashed circle with multiple particles emitted and two electromagnetic showers (two  $\gamma$ 's from a  $\pi^0$  decay).

As described in section 1, 32 PMTs are used to detect the scintillation light in MicroBooNE. In the PMT front-end motherboard (FEM), the PMT signal is separated by a splitter into high-gain (x10) and low-gain (x1) amplifiers, allowing a wide dynamic range for a 64-MHz 12-bit ADC readout of the PMT pulses [5]. In the PMT readout system, there are two separate readout streams: *beam discriminator* and *cosmic discriminator*. The beam discriminator starts 4  $\mu\text{s}$  before the beam gate. It reads out 1500 consecutive samples ( $\sim 23.4 \mu\text{s}$ ) of the PMT waveforms. The cosmic discriminator is a self-triggered PMT readout. It reads out 40 consecutive samples ( $\sim 0.6 \mu\text{s}$ ) of the PMT waveforms, which record the light information not only from beam-coincident activities but also activities out-of-time with the beam.

The PMT waveforms are processed offline to reconstruct the time and number of photoelectrons (PE) of a flash, which is a group of PMT signals close in time. For the beam discriminator, a deconvolution using the Fast Fourier Transformation (FFT) is performed to unfold the electronics responses from various RC circuits in the splitter and the shaper. A flash is then formed if the PMT measurements satisfy the multiplicity requirement ( $>2$  PMTs above a threshold of 1.5 PE) and the total integrated PE threshold ( $>6$  total PE) in a 100 ns window. A flash window lasts 7.3  $\mu\text{s}$  in order to exclude noise and to include the contribution from the late scintillation light. The scintillation light in liquid argon has a prompt and a slow component with decay times of about a few nanoseconds and 1.6  $\mu\text{s}$ <sup>4</sup>, respectively.

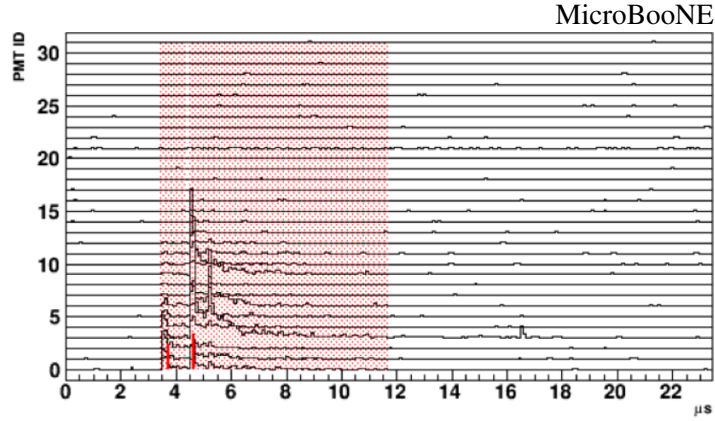
Within the flash window, the time bin with the maximal total PE from all PMTs marks the starting time of a flash. The PE of each PMT in a flash is integrated over the entire flash window. Though the average time between two adjacent flashes in MicroBooNE is  $\sim 100 \mu\text{s}$ , a procedure is set to end the current flash window and start a new one if the new flash has a large starting PE, calculated as the total PE from all PMTs in the first 100 ns. and satisfies either of the two requirements: (1) the new flash is at least 1.6  $\mu\text{s}$  later than the preceding one; (2) a significantly different PMT hit pattern (number of PEs in each PMT) in the first 100 ns of the new flash compared to the pattern in the last 100 ns of the preceding flash using a Kolmogorov-Smirnov test [28]. Figure 20 shows an example of two adjacent reconstructed flashes from beam discriminator PMT waveforms.

For the cosmic discriminator, the readout window is shorter than the slow component of the scintillation light. The light yield ratio of the slow to the prompt component is about 3:1 for the minimum ionizing particles. The integrated PE of a cosmic discriminator is scaled by a factor of two to take into account the slow component portion of the scintillation light not fully recorded by the readout window. Because of the inefficiency of the cosmic discriminator, the data from the cosmic discriminator is ignored when the beam discriminator data is present, and the cosmic discriminator performance is calibrated by the beam discriminator data.

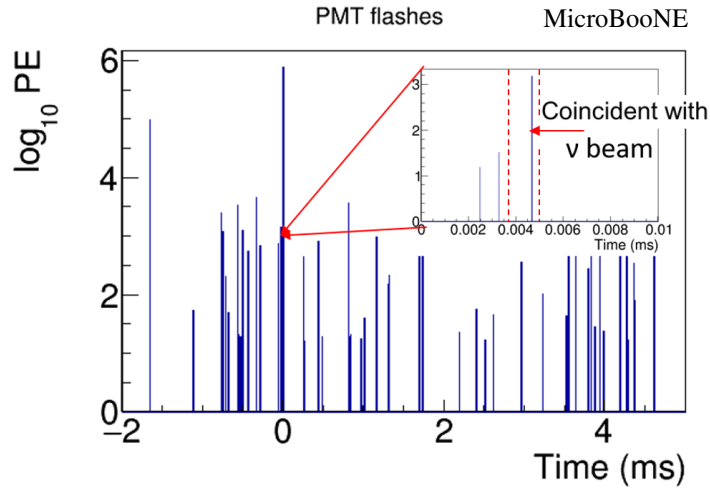
Figure 21 shows the reconstructed PEs and time for each PMT flash from a data event. The flash corresponding to the neutrino interaction is shown in the inset figure between the dashed red lines that indicate the beam spill window. One can see that about 50 flashes are reconstructed in this event and it is challenging to match the TPC clusters to these many PMT flashes. On the other hand, if a robust charge-light matching is developed, each TPC cluster's starting time measured by the PMTs can be used to reject the overwhelming cosmic-ray muon background in the neutrino selection.

---

<sup>4</sup>The two lifetimes correspond to the molecular excimer states excited either in a singlet state or a triplet state.



**Figure 20:** Illustration of two reconstructed flashes from beam discriminator PMT waveforms. The black curves are the deconvolved PE spectra for each PMT. The red lines represent the flash times and the red bands represent the flash windows. For the second flash at about  $4.6\mu\text{s}$ , there is a Michel electron as indicated by the second peak (at about  $5.3\mu\text{s}$ ) of its PE spectra.



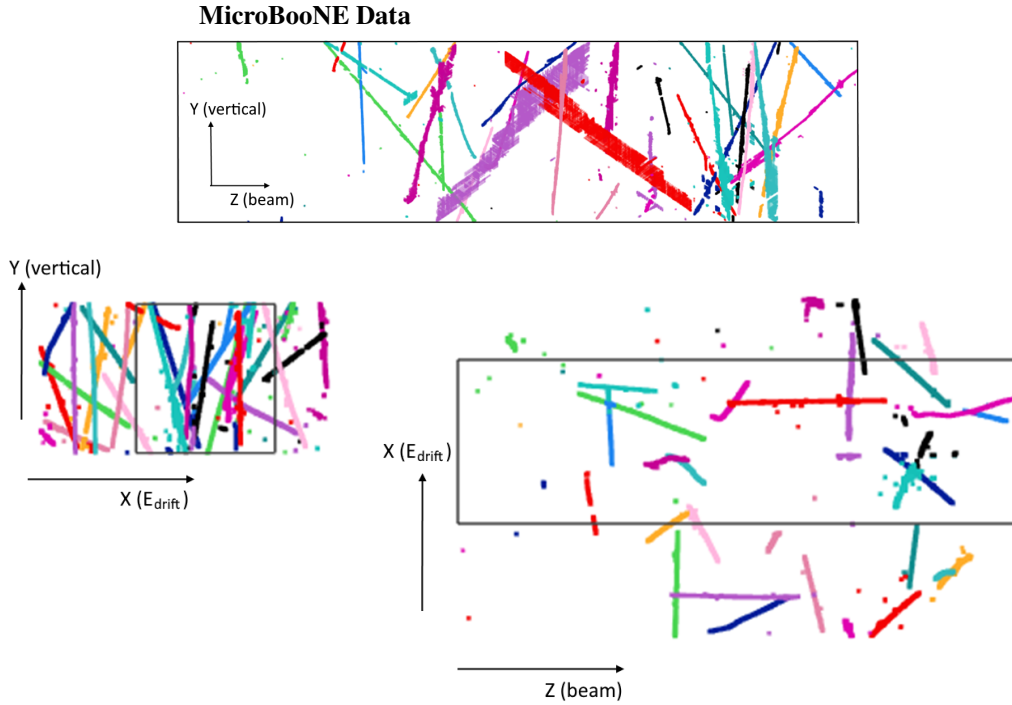
**Figure 21:** The reconstructed PEs of a flash as a function of flash time. The 6.4 ms PMT readout window is shown relative to the trigger time. The flashes from the beam discriminator ( $23.6\mu\text{s}$  long) are shown as inset. The flash in coincidence with the BNB beam spill (between dashed red lines) is indicated. In general, there are 40–50 reconstructed PMT flashes in each BNB event.

### 4.3 Many-to-many charge-light matching

Now that the TPC charge activities have been reconstructed and grouped into physically distinct clusters in section 3 and the PMT light measurements have been reconstructed into distinct flashes in section 4.2, the next step is to match the 20–30 TPC clusters to the 40–50 PMT flashes for each recorded event. This will allow each matched cluster to be assigned the precise starting time measured by the PMTs, and enable using the short BNB time window to reject the vast majority of cosmic-ray muons as neutrino candidates.

As an example shown in figure 22, there are many TPC clusters spanning the entire readout window with unknown electron drift start time. The X-position is assigned by a direct conversion





**Figure 22:** An example of all the TPC clusters from a MicroBooNE data event before charge-light matching. Different clusters are labeled in different colors, but each cluster is labeled in the same color in different 2D views. The borders represented by the black lines are the boundaries of the LArTPC active volume. Top: front (Y-Z) view. Bottom left: side (Y-X) view. Bottom right: top (X-Z) view. The X-position of the black box corresponds to the starting time of the neutrino interaction, and the X-position shift of cosmic-ray muon clusters will be corrected after the charge-light matching. The entire readout time window, i.e. the X-axis range, is about 2 times the TPC width.

from wire readout time relative to the trigger time. More PMT flashes are generally recorded than the number of TPC clusters since PMTs sense not only the activity inside the TPC but also that outside the TPC where LAr is present within the cryostat. On the other hand, a TPC cluster does not necessarily have a corresponding PMT flash since the light collection system (e.g. the cosmic discriminators) has inefficiencies, especially for clusters either with low visible energy or near the cathode (far from the PMTs). Also, as mentioned in section 4.1, the resulting clusters after the application of the clustering algorithm may still have an under-clustering issue, which is intended to be addressed in this matching stage by allowing several TPC clusters to match to a single PMT flash. In summary, there are two requirements in the matching algorithm:

- (A) One TPC cluster can match to zero or at most one PMT flash.
- (B) One PMT flash can match to zero, one, or multiple TPC clusters. These multiple clusters that as a whole match the same PMT flash form a *cluster bundle*.

The “match” is defined as a good agreement between the predicted and measured light signals, considering the signal intensity of each individual PMT as well as the hit pattern of all 32 PMTs. Assuming a TPC cluster to be associated with a PMT flash, a prediction of the PE distribution for

the 32 PMTs can be made. The electron drift start time of the TPC cluster is shifted from the default BNB beam time to the measured time of the PMT flash. This enables a correction of the X-position of the TPC cluster. Then, the charge associated with each space point in the TPC cluster is used to predict the PMT light signals based on a photon library [29]. The TPC volume ( $2.56 \text{ m} \times 2.32 \text{ m} \times 10.36 \text{ m}$ ) is divided into  $75 \times 75 \times 400$  voxels. Millions of optical photons of 128 nm wavelength from scintillation are generated and emitted with a  $4\pi$  angular distribution in each voxel, and the propagation of these photons is simulated with realistic optical photon processes of absorption and scattering in Geant4. The PMT acceptance of optical photons emitted at different locations in the TPC volume is calculated and recorded in the photon library. With this photon library, the PEs from each of the PMTs for a given TPC cluster can be predicted by applying the PMT acceptance to the charge of each space point. An overall scaling factor is applied to take into account the calibrated scintillation light yield per unit deposited energy.

Interestingly, such a many-to-many matching problem is very similar to the charge solving problem as introduced in section 3.2. There are more unknowns than knowns in this system, and the imaging equation of the first principle as shown in eq. (3.1) can be used to relate the predicted light signals from all possible TPC clusters to the measured signals from PMT flashes. Hypothetical pairs of TPC clusters and PMT flashes are created and tested, in order to find the most compatible ones and eliminate the rest. Again, the compressed sensing technique is utilized to perform this many-to-many matching by minimizing an  $\ell_1$ -regularized chi-square function. In practice, a set of matching algorithms are developed to pre-select, fit the  $\ell_1$ -regularized chi-square, and re-examine the hypothetical TPC-PMT pairs.

**Pre-selection:** A pre-selection of the hypothetical TPC-PMT pairs is important to reduce the number of unknowns in the  $\ell_1$ -regularized chi-square fitting, allowing for a more robust minimization. Two major tests, time range compatibility and PMT hit pattern compatibility, are performed to remove the incompatible TPC-PMT pairs. For the time range compatibility, the TPC cluster is required to be fully contained within the maximum drift window corresponding to the PMT flash time<sup>5</sup>. For example, as shown in figure 22, X-positions (along the drift) of the space points in any TPC cluster have an overall shift because of the unknown electron drift start time, but the in-beam activities must be contained in the nominal detector volume (black box) which is relative to the beam time. For the PMT hit pattern compatibility, the pairs with highly incompatible predicted and measured light signals are ruled out. A Kolmogorov-Smirnov test (K-S test) and a chi-square test, which inspect the hit pattern and the absolute normalization of the 32 PMTs' signals, respectively, are combined to discriminate the incompatible pairs. Specifically, to enable a many-to-one TPC-PMT matching, the TPC clusters paired to the same PMT flash are jointly tested to maintain the many-to-one potential. The most compatible TPC-PMT pair is used as a basis and the other ones are added individually to check the change in compatibility. The pairs which significantly reduce the compatibility are ruled out.

**Chi-square fitting:** Given the passing candidate TPC-PMT pairs after the pre-selection, a chi-square function incorporating a  $\ell_1$ -regularization term is constructed to compare the predicted and

---

<sup>5</sup>A precise cut can be applied since the space charge effects [30] are insignificant along the drift direction.

730 measured light signals:

$$\chi^2 = \sum_i \sum_j \chi_{ij}^2 + \chi_{p1}^2 + \chi_{p2}^2 + \chi_{p3}^2, \quad (4.1)$$

$$\chi_{ij}^2 = \frac{(M_{ij} - \sum_k a_{ik} \cdot P_{ikj} - b_i \cdot M_{ij})^2}{\delta M_{ij}^2}, \quad (4.2)$$

$$\chi_{p1}^2 = \sum_i \frac{(\sum_k a_{ik} - 1)^2}{c_1^2}, \quad (4.3)$$

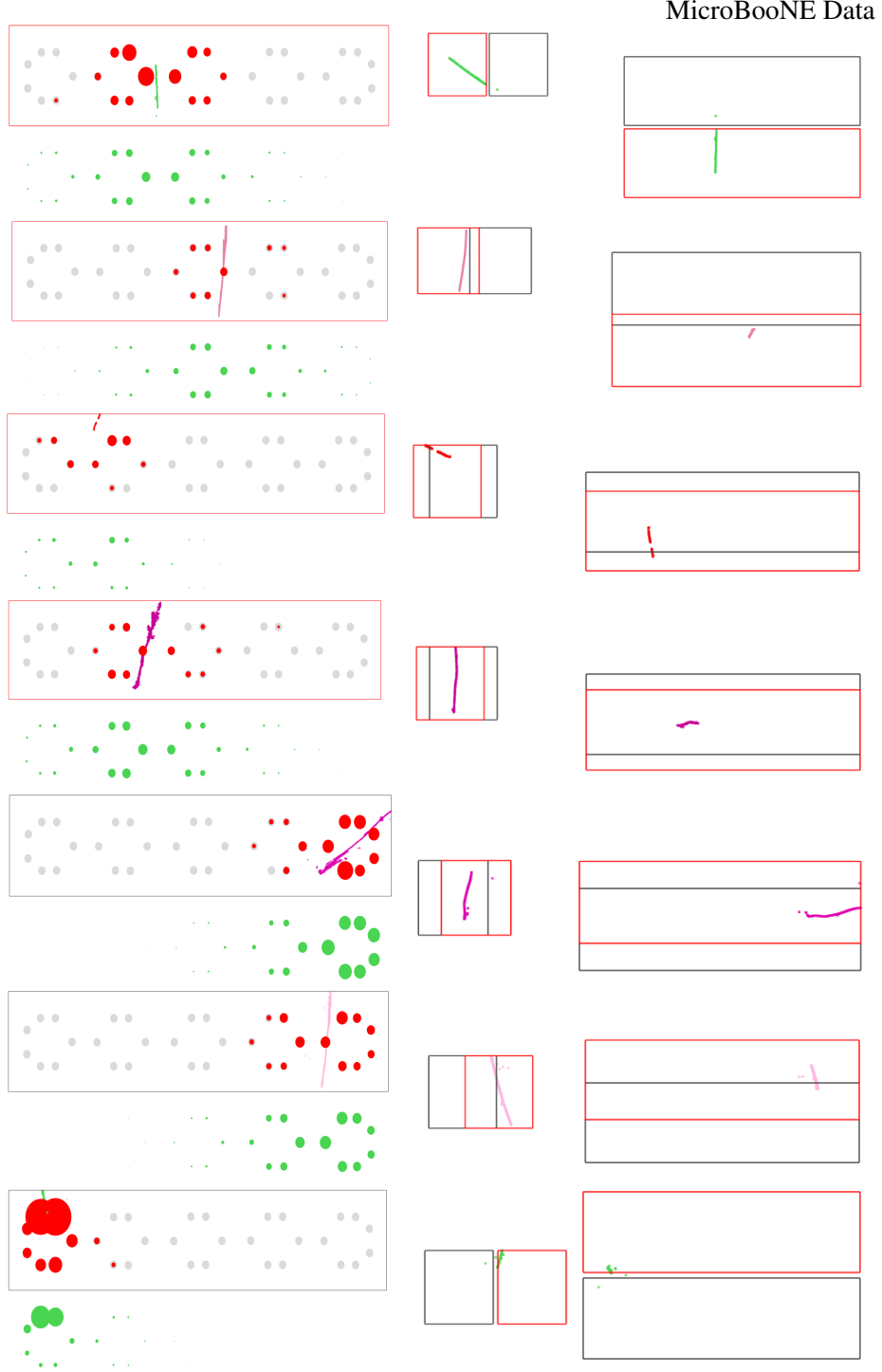
$$\chi_{p2}^2 = \sum_i \frac{b_i^2}{c_2^2}, \quad (4.4)$$

$$\chi_{p3}^2 = \lambda \cdot \sum_i \sum_k a_{ik}. \quad (4.5)$$

731 For the input TPC-PMT pairs, the index  $i$  runs through all PMT flashes,  $j$  runs through all hit  
 732 PMTs of each flash, and  $k$  runs through all the TPC clusters.  $M_{ij}$  and  $\delta M_{ij}$  represent the measured  
 733 PE and its uncertainty of the  $j$ -th PMT in the  $i$ -th flash, respectively. The uncertainties from light  
 734 yield and charge measurements are conservatively assigned.  $P_{ikj}$  represents the predicted PE of  
 735 the  $j$ -th PMT in the  $i$ -th flash from the  $k$ -th TPC cluster. The  $a_{ik}$ 's, which represent the credibility  
 736 of a correct match between the  $k$ -th TPC cluster and the  $i$ -th PMT flash pair, are the parameters  
 737 of interest in the fit. All  $a_{ik}$ 's are constrained to be non-negative. A well-matched TPC-PMT pair  
 738 will have  $a_{ik}$  close to 1, while a bad match will have  $a_{ik}$  close to zero.  $\chi_{p1}^2$  applies the constraints  
 739 that each TPC cluster should only be used once, i.e. matched to at most one PMT flash. The  
 740 introduction of the  $b_i$  term is to take into account the possibility that some of the PMT flashes may  
 741 not be associated with any TPC clusters, in which case  $b_i$  is close to 1, though the  $\chi_{p2}^2$  term gives  
 742 the constraint that  $b_i$  is preferred to be close to 0. The  $\chi_{p3}^2$  term represents the application of the  
 743 compressed sensing technique which prefers a best-fit solution where most of  $a_{ij}$  terms are zero.  $\lambda$   
 744 is the regularization strength.  $c_1$  and  $c_2$  are two hyper-parameters to regularize the corresponding  
 745 penalty terms, and the values are 0.01 and 0.025, respectively, tuned by real data. After the fitting,  
 746 the most incompatible TPC-PMT pairs with extremely small  $a_{ik}$  values are eliminated from further  
 747 consideration. Naturally, PMT flashes that do not match any TPC clusters are eliminated as well.  
 748 The remaining TPC-PMT pairs go into the second round fitting to further approach the best solution,  
 749 with the unnecessary  $\chi_{p2}^2$  and other  $b$  related terms removed.

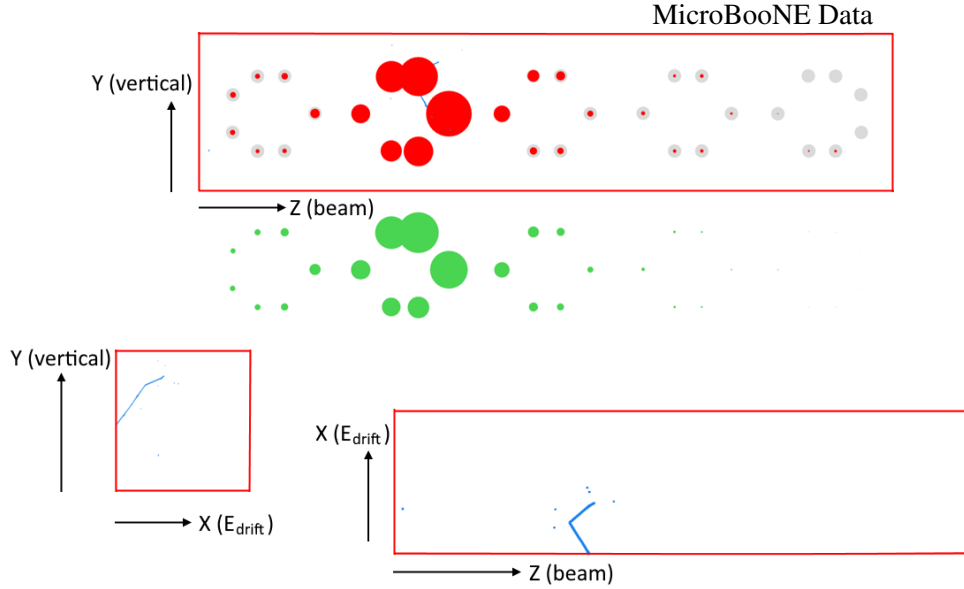
750 **Re-examination:** After the two rounds of chi-square fitting, for each TPC cluster, the most probable  
 751 TPC-PMT pair with the largest  $a_{ik}$  is selected for further examination. The hit pattern compatibility  
 752 test as introduced in “Pre-selection” is performed. Since many-to-one TPC-PMT matching is  
 753 allowed in this procedure, the biggest TPC cluster that pairs to a PMT flash is defined as the  
 754 principle component. Then, another TPC cluster that is paired to the same PMT flash is added to  
 755 the hit pattern compatibility test. If the test result becomes worse, this cluster is removed. Otherwise,  
 756 it is added to the many-to-one TPC-PMT pairs, i.e. the cluster bundle. After the re-examination  
 757 of all selected TPC-PMT pairs, the unmatched TPC clusters will be tested against the unmatched  
 758 PMT flashes to check if any possible pairings are missed.

759 Figure 23 shows an example of 7 matched pairs out of a total of 31 matched pairs from one  
 760 MicroBooNE data event. After the many-to-many matching, the in-beam, flash-matched TPC

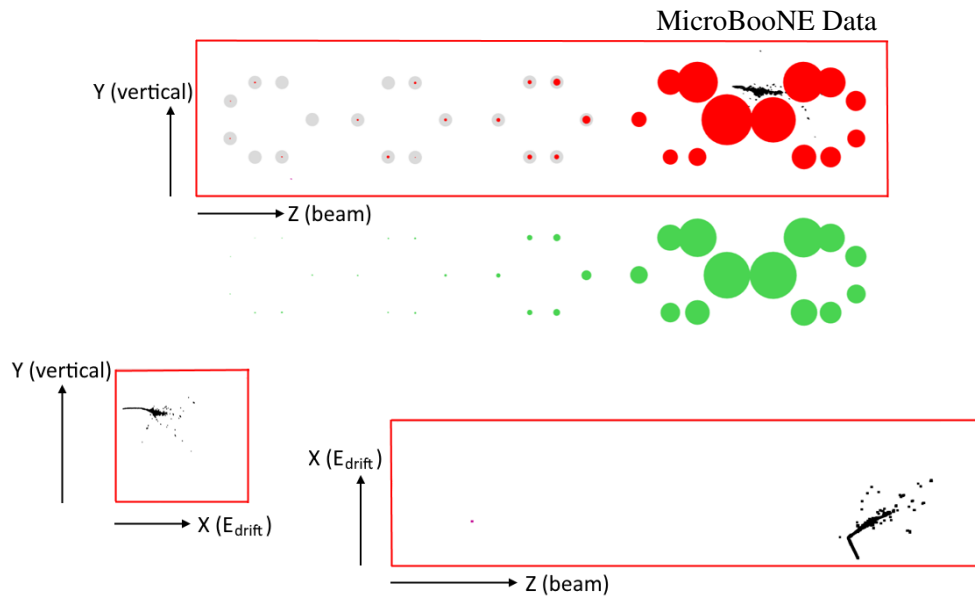


**Figure 23:** Selected 7 matched pairs out of the 31 pairs from a data event. From left to right, they are the front (Y-Z), side (Y-X), and top (X-Z) views of the detector, respectively. The black or red boxes correspond to the LArTPC active volume. The gray solid circles in the front view represent PMTs in different locations. The red solid circles represent the measured PE in the PMTs. The green solid circles represent the predicted PE based on the matched TPC cluster(s). The area of the circle is proportional to the number of PEs. The black box has no X-position shift, and it corresponds to the starting time of the neutrino interaction. The red box corresponds to the time of the matched PMT flash, i.e. the starting time of the cosmic-ray muon, and the X-position shift is corrected.

clusters are taken to be neutrino interaction candidates, and the remainders are rejected as cosmic-ray background. Figure 24 and figure 25 demonstrate successfully matching muon and electron neutrino clusters to their respective in-beam flashes. The performance of the matching algorithm is evident from these event displays and quantitative evaluations are provided in section 5. On average, the charge-light matching consumes about 30 seconds per event with less than 1.5 GB memory on an Intel(R) Core(TM) i7-4790K CPU @ 4.00GHz.



**Figure 24:** A muon neutrino event is shown with its matched flash. The red boxes correspond to the LArTPC active volume. The gray solid circles in Y-Z view represent the PMTs in different locations. The red solid circles represent the measured PE in the PMTs. The green solid circles represent the predicted PE based on the TPC cluster(s). The area of the red or green circle is proportional to the number of PEs.



**Figure 25:** An electron neutrino event is shown with its matched flash. The red boxes correspond to the LArTPC active volume. The gray solid circles in Y-Z view represent the PMTs in different locations. The red solid circles represent the measured PE in the PMTs. The green solid circles represent the predicted PE based on the TPC cluster(s). The area of the red or green circle is proportional to the number of PEs.

## 5 Evaluation of the Wire-Cell 3D imaging and the charge-light matching

In this section, the quantitative evaluations of the Wire-Cell 3D imaging and the many-to-many charge-light matching are presented. The performance of these three-dimensional approaches to reconstruct neutrino activities is demonstrated as well. The intrinsic problem with 3D imaging stems from the wire readout ambiguity, and this is worsened by nonfunctional wires. As a consequence, ghost tracks appear in the final 3D image and cannot be completely removed, despite the dedicated algorithms described in section 3 and section 4. On the other hand, a true hit, which is a space point associated with true energy depositions, might be discarded in the charge solving and the de-ghosting steps. Two major metrics are used to evaluate the quality of the 3D imaging result as follows:

**Purity** of the 3D image – the number of the reconstructed hits overlapping true TPC hits divided by the total number of the reconstructed hits.

**Completeness** of the 3D image – the number of the true hits overlapping the reconstructed hits divided by the total number of the true hits. The true hits are required to be within the TPC active volume and are weighted by their true deposited (visible) energy.

The 3D metrics are relevant to understand the performance of the subsequent Wire-Cell reconstruction. For example, the cosmic-ray background rejection and the pattern recognition are expected to operate on the 3D images in order to maximize the potential capability of LArTPCs.

Given the numerous cosmic-ray muons in the TPC, the 3D clustering and the many-to-many charge-light matching are applied to properly group the neutrino interaction and match it to the in-beam flash. The clustering and charge-light matching may fail to select the neutrino interaction or suffer from both the over-clustering and under-clustering issues. The correctness of the matching and the efficiency of selecting neutrino interactions after matching are evaluated as well. These two metrics are defined below and evaluated from simulation:

**Correctness** of the charge-light matching – the fraction of all in-beam neutrino candidates that true neutrino interactions. The incorrectly-matched candidates have no neutrino interactions but do have cosmic-ray muon activities with extremely low completeness values as defined above.

**Efficiency** of selecting neutrino interactions – the fraction of the events with neutrino interactions that have correct in-beam matches.

The development and optimization of the Wire-Cell 3D reconstruction techniques described in previous sections are based on  $\sim 1500$  data events. The evaluations in this section are carried out using the MicroBooNE detector simulation. The MicroBooNE simulation has incorporated a realistic detector response model which is in good agreement with data. A data-driven noise model and long-range wire responses [12–14] are implemented in addition to the capability to overlay real data from cosmic rays with a simulated neutrino interaction. The MicroBooNE full detector simulation software *LArSoft* [31] and *uboonecode* [32] are used to simulate the BNB neutrino charged current (CC) and neutral current (NC) interactions in the cryostat that contains the rectangularly shaped TPC active volume, as described in section 1. The GENIE neutrino generator [33] and the Geant4 simulation toolkit [34, 35] are incorporated into the MicroBooNE simulation software.

Three different Monte Carlo (MC) samples are used to perform the evaluations:

1. Ideal tracks – lines of charge deposition corresponding to minimum ionizing particles (MIPs) to demonstrate the intrinsic performance of the 3D imaging and the impact from nonfunctional wires and the signal processing chain. See section 5.1.
2. Neutrino only – full detector simulation of a neutrino interaction without cosmic-ray muons to demonstrate the performance of the 3D imaging on the complex topology of neutrino interaction final states. See section 5.2.
3. Neutrino overlay – full detector simulation of a neutrino interaction mixed with real cosmic-ray data to demonstrate the final performance after the Wire-Cell 3D imaging, clustering, and charge-light matching. This sample is used to show the correctness and the neutrino efficiency after the matching step. See section 5.3.

By comparing the purity and the completeness results between sample B and sample C, the impact from cosmic-ray muons and the performance of clustering and charge-light matching on the neutrino interaction will be shown and discussed. In the neutrino-only or neutrino-overlay samples, the  $\nu_\mu$  or  $\nu_e$  energy spectra are from the BNB beam flux simulations. Only the neutrino interactions with their primary vertices in the TPC active volume are considered. Neutrino interactions outside the active volume are largely or completely invisible because the ionization electrons outside the active volume cannot drift and be collected by the wire planes. Evaluation of the performance on cosmic-ray only data is not specifically performed. The coincident in-beam cosmic-ray activities is expected to be selected in this case, and they will be further rejected by dedicated cosmic-ray muon taggers in the later reconstruction chain [15, 36], which is out of the scope of this paper.

## 5.1 Imaging performance of ideal tracks

About twenty one-meter-long ideal tracks (lines of charge depositions corresponding to MIPs) in each event are simulated in the MicroBooNE TPC. The angular distribution is uniform in  $4\pi$ . The start position distribution is uniform in the TPC active volume. The number of hit cells on the anode plane per unit time is close to the real data, mimicking the numerous cosmic-ray muons traversing the MicroBooNE detector.

Three scenarios of the simulation are constructed to study the performance of the 3D imaging as well as the impact from the nonfunctional wires and signal processing (SP):

**Perfect SP:** The true charge deposition on each wire is only convoluted with the smearing effects from the diffusion during the charge drift and the software filters used in the signal processing. In this perfect signal processing procedure, there is no bias or failure of the charge extraction.

**Dead + perfect SP:** Nonfunctional wires are added based on data observations and perfect signal processing is applied.

**Dead + real SP:** Nonfunctional wires are added and realistic signal processing is applied. For a prolonged track which leaves a long signal in each individual wire readout, the realistic signal processing may fail to reconstruct the charge for the induction plane wires because of the bipolar signal cancellation. See ref. [13] for more details. This results in gaps in the 2D wire-versus-time views of the charge measurement as mentioned in section 3 and section 4.1.

The results of reconstructed tracks by the 3D imaging are categorized into 4 types:

**Good** – tracks are well reconstructed with at least 99% completeness.

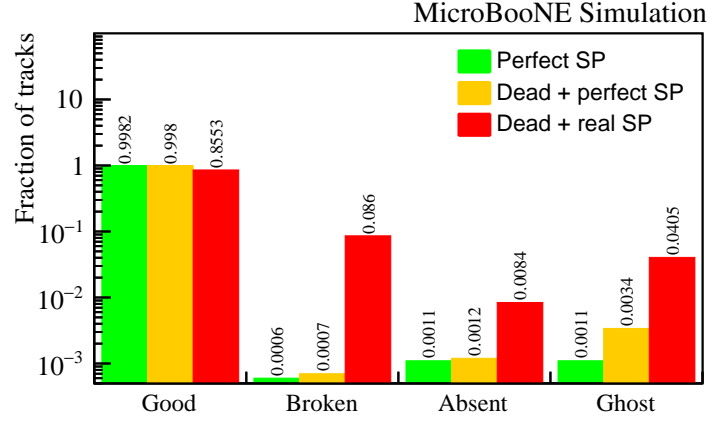


848 **Broken** – tracks have gaps and are broken into separate segments.

849 **Absent** – tracks completely fail to be reconstructed.

850 **Ghost** – tracks have no overlap with any true track.

851 Based on thousands of simulated events, the fractions of each category of reconstructed tracks are shown in figure 26. For “good”, “broken”, and “ghost” tracks, the fraction is weighted by their

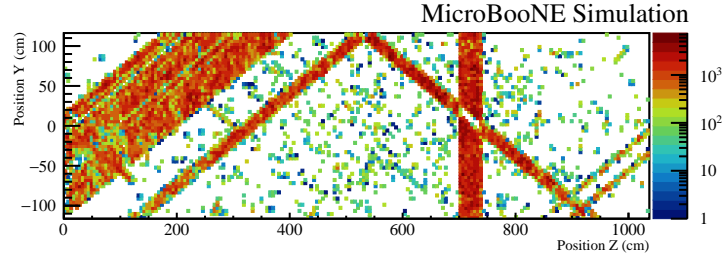


**Figure 26:** The fraction of (good, broken, absent, ghost) reconstructed tracks from the Wire-Cell 3D imaging for different scenarios. For good, broken, and ghost tracks, the fraction is weighted by their lengths and normalized to the total length of true tracks. See text for definitions of each category.

852 lengths and normalized to the total length of true tracks. Therefore, the sum of the fractions for  
 853 these three categories could be less than 100% when there are gaps in the reconstructed tracks  
 854 because the signal processing has inefficiency for events with a certain topology, e.g. for prolonged  
 855 tracks. It could also be greater than 100% because of the occurrence of ghost tracks in some places  
 856 where there are no true charge depositions. Note that the sum of the fractions of “good”, “broken”,  
 857 and “ghost” tracks is very close to 100%, which indicates the ghost tracks explain the missing parts  
 858 of the broken tracks. The result of “Dead + Perfect SP” is very similar to the result of “Perfect  
 859 SP” and almost all the tracks are well reconstructed. This shows that the nonfunctional wire issue  
 860 is properly addressed in the 3D imaging and a 97% active volume efficiency has been achieved.  
 861 The impact of the nonfunctional wires on the quality of the 3D image will be further discussed in  
 862 section 5.2 and section 5.3. The fraction of the ghost tracks in the scenario of “Dead + Perfect SP”  
 863 is three times larger than that of “Perfect SP” because of the presence of the nonfunctional wires,  
 864 but it is still negligible. In the scenario of “Dead + Real SP”, there is a large increase of both broken  
 865 tracks and ghost tracks. The broken tracks come from the gaps, which as mentioned previously are  
 866 attributed to the failure of signal processing for the prolonged tracks. In this simulation of ideal  
 867 tracks, there are a certain number of prolonged tracks since they are generated with a  $4\pi$  uniform  
 868 angular distribution. The situation is better for the beam neutrino interactions, in which case the  
 869 final state particles are mostly forward-going.

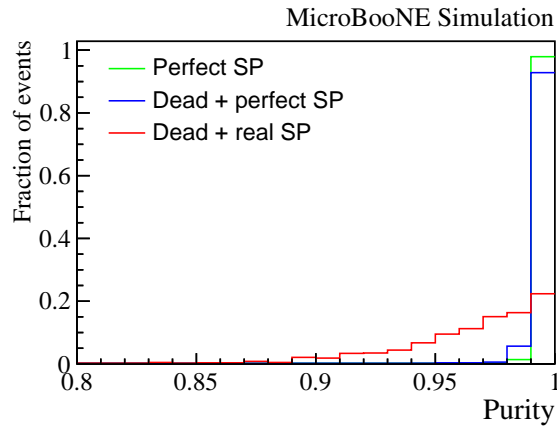
871 With the realistic signal processing, more ghost tracks appear almost exclusively in the non-

functional region as shown in figure 27. In one wire plane, the realistic signal processing, which may fail to extract the charges, could introduce a gap in the 3D image no matter the signal processing in the other two wire planes is successful or not. The measured charges originating from the TPC activities along this gap, if any in the other two wire planes, therefore tend to be explained by ghosts lying in a nonfunctional region where a 2-plane tiling is allowed.

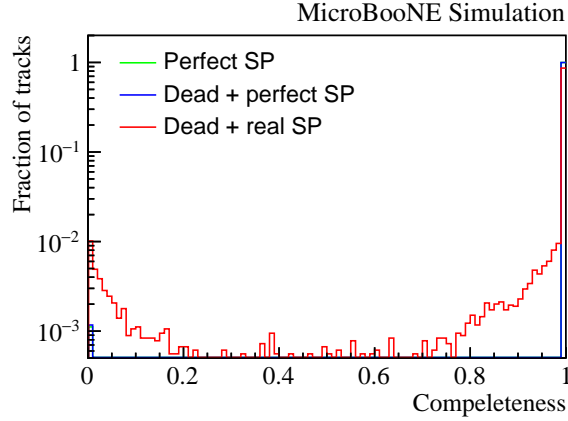


**Figure 27:** The position (Y/vertical versus Z/beam) distribution of the ghost tracks in the scenario of “Dead + real SP”. Color scale (Z-axis value) represents the count of space points in ghost tracks. The bands correspond to the nonfunctional regions as shown in figure 4.

The purity for each event is calculated by dividing the total length of the non-ghost tracks by the total length of all the reconstructed tracks. The distribution of purity scores is presented in figure 28. For “Dead + real SP”, 96.4% of the events have at least 90% purity. Figure 29 shows the distribution of the completeness for all simulated tracks. For the scenario of “Dead + real SP”, 86.5% (and 93.0%) of the simulated tracks have at least 99% (and 80%) completeness. The low completeness values correspond to the prolonged tracks, especially those with directions close to normal to the wire planes. This emphasizes again that good signal processing is important to retain the good quality of the 3D imaging result.



**Figure 28:** Distribution of the purity of the reconstructed tracks from each event for different scenarios. The number of ghosts significantly increases with the presence of both nonfunctional wires and the real imperfect signal processing. The histograms are normalized separately for each category.



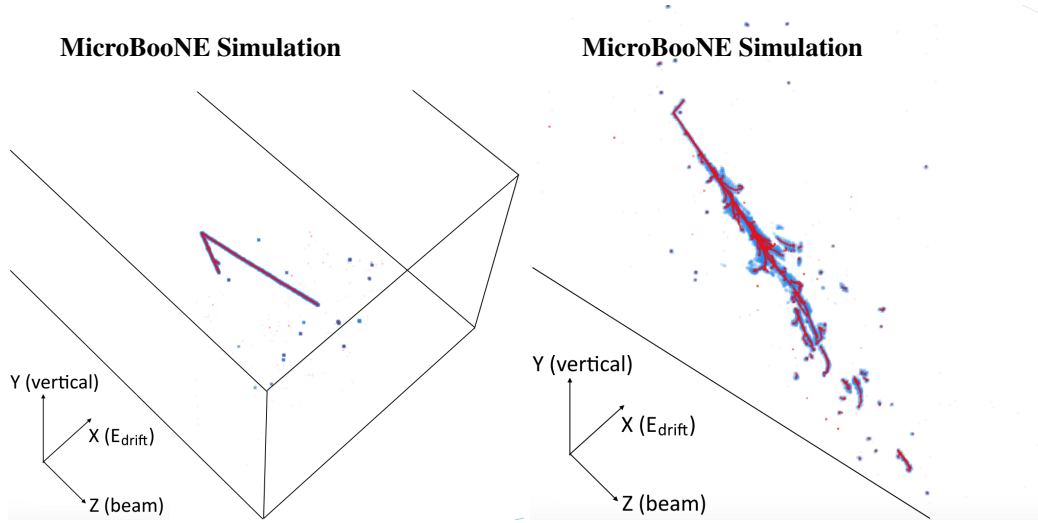
**Figure 29:** Distribution of the completeness of each simulated track for different scenarios. The distributions are normalized for each category, respectively. The results of “Perfect SP” and “Dead + perfect SP” are basically the same in which case the green line is covered by the blue line. The inefficiency of the signal processing for prolonged tracks leads to very low completeness values.

## 5.2 Imaging performance of neutrino interactions

Unlike the simulated ideal tracks in section 5.1, the topology of a neutrino interaction’s final state particles could be much more complicated than the single-track-like cosmic-ray muons. Neutrino-only samples without cosmic-ray muons are used in this case. In order to evaluate the performance of the 3D imaging, the clustering as well as the charge-light matching steps are bypassed and all the 3D space points reconstructed in the 3D imaging are taken as neutrino activities. This is equivalent to performing perfect clustering and charge-light matching.

When a neutrino interacts with an argon nucleus, there are generally multiple final state particles. On one hand, there is a very limited phase space for the final state particles to be in the prolonged or isochronous directions, in which case the 3D image may have gaps. Note that a highly ionizing particle (HIP) may avoid such failures in the signal processing since it has a significantly higher signal-to-noise ratio. On the other hand, with the complexity of the neutrino interactions, other failure modes may arise. Some of the particles like neutrons,  $\gamma$ ’s from pion decays, and primary or secondary electrons could yield low-energy (sub-MeV) depositions via nuclear recoil, Compton scattering, or Bremsstrahlung radiation, respectively. These low energy depositions are likely to be suppressed because of the thresholding in the signal processing or removed in the 3D imaging as they resemble the dot-like ghosts. As a result, the completeness distribution will be biased and smeared to lower values compared with that in figure 29. The thresholding in the signal processing is primarily to suppress fake signals from noise fluctuations. A lower thresholding in the signal processing would create more fake charges, which can interplay with true charges and lead to ghost tracks in the nonfunctional region.

Figure 30 shows two 2D snapshots of the 3D event displays. The left is a  $\nu_\mu$  CC interaction producing a muon and a single proton ( $1\mu 1p$ ) in the final state. The right is a  $\nu_e$  CC interaction producing an electron EM shower and a single proton ( $1e 1p$ ) in the final state. The red points represent the space points from Monte-Carlo truth and the blue ones represent the reconstructed



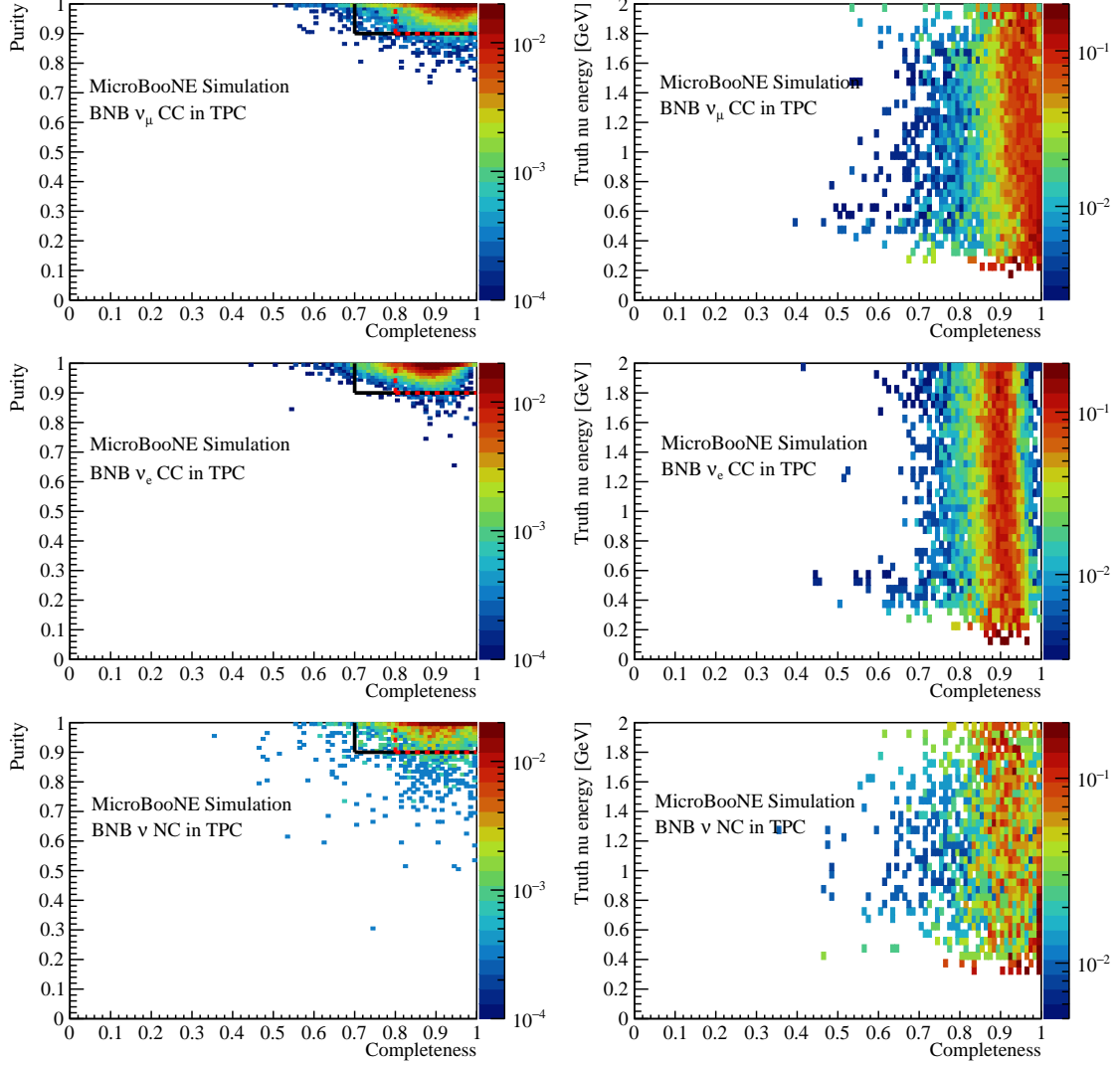
**Figure 30:** Left:  $1\mu 1p \nu_\mu$  CC interaction. Right:  $1e 1p \nu_e$  CC interaction. Blue: reconstructed 3D image. Red: truth trajectories. The voxel size and opacity are tuned for better illustration.

space points in the 3D imaging. The image of the reconstructed points are blurred because of the charge diffusion during the drift and the software filter smearing in the signal processing. Generally speaking, the reconstructed 3D image has both good completeness and purity compared to the truth 3D image in these two examples; even the short tracks belonging to the EM shower and isolated energy depositions are reconstructed. The quantitative evaluations of the purity and the completeness for BNB  $\nu_\mu$  CC,  $\nu_e$  CC, and NC interactions in the TPC are shown in figure 31. The results are summarized in table 3.

**Table 3:** Fraction of the events that correspond to the completeness and purity values within the black or red boxes as shown in figure 31. These numbers are the overall performance for the integrated BNB neutrino flux which has an average neutrino energy of about 800 MeV. See text for more discussions on the energy dependence. All neutrino interactions are simulated within the TPC active volume, without cosmic-ray muons.

Scenario: neutrino only	BNB $\nu_\mu$ CC	BNB $\nu_e$ CC	BNB NC
Purity > 90% and			
Completeness > 80%	88.6%	89.2%	80.7%
Completeness > 70%	93.3%	96.7%	87.0%

The purity is high in the neutrino-only cases in which there are no cosmic-ray muons. For neutrino energy less than 400 MeV, the purity performance, e.g. the fraction of events with greater than 90% purity, is reduced by about 10% compared to that in higher energy regions. This is due to the inefficiency of de-ghosting for low-energy events. The lower purity for NC interactions mainly corresponds to the events with visible energy less than 100 MeV, in which case the 3D image consists of many dot-like or very short tracks. Unlike figure 29, figure 31 has no ultra-



**Figure 31:** Two-dimensional distributions of the completeness and the purity of the 3D image for BNB  $\nu_\mu$  CC,  $\nu_e$  CC, and NC interactions in the TPC. There are no cosmic-ray muons in this simulation. Left: purity vs. completeness for each neutrino interaction. The color scale (Z-axis value) represents the fraction of events. Right: true neutrino energy vs. completeness; the distribution is normalized for each row of the true neutrino energy bin. The color scale (Z-axis value) represents the fraction of events in each row. The integrated fraction of the events within the solid black and dashed red boxes can be found in table 3.

low completeness events because a final state of a single prolonged track can rarely happen for a neutrino interaction. For  $\nu_e$  CC interactions, since primary electrons lead to EM showers through significant Bremsstrahlung radiation, the peak completeness is biased down to  $\sim 90\%$  because of the inefficiency for isolated low energy depositions in the shower. Such a bias is not critical to the track versus shower identification and can be corrected in the shower energy reconstruction. NC interactions generally generate protons, neutrons, or pions. These particles could yield low energy depositions during their travel in the liquid argon as explained previously, introducing a much smeared completeness distribution. The 100% completeness peak for the low-energy neutrino NC interactions as seen in the bottom right panel of figure 31 mainly corresponds to elastic scattering with a single low-energy proton emitted.

A dependence of the completeness on the true neutrino energy is indicated by the right panel of figure 31. A high energy neutrino is more likely to produce multiple energetic hadrons, introducing distant or isolated low-energy depositions via nuclear recoils, de-excitation of argon nuclei, pion decays, etc. These low-energy TPC activities are more likely to be suppressed in the signal processing or 3D imaging as discussed previously.

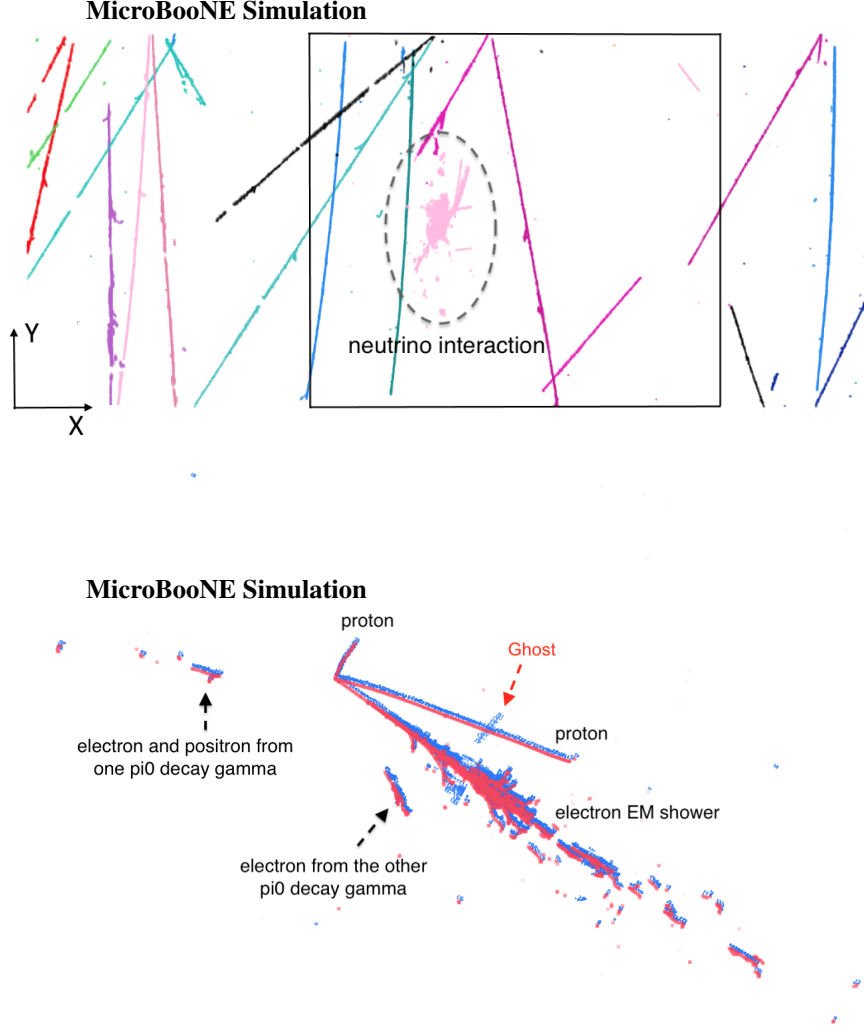
### 5.3 Final performance in realistic cases

In this section, the neutrino-overlay samples are used to demonstrate the final performance of the Wire-Cell 3D imaging, clustering, and charge-light matching. Neutrino interactions are simulated and mixed with real cosmic-ray data. The clustering, light signal reconstruction, and charge-light matching are applied on the 20–30 TPC clusters and 40–50 PMT flashes to select the in-beam neutrino activities. The efficiency and correctness of the charge-light matching and the quality of the 3D images of the selected neutrino candidate clusters are keys to the downstream reconstructions.

Figure 32 shows an example of one of the most challenging cases. The top panel shows the X-Y projection of all TPC activities including cosmic-ray muons and a neutrino interaction. The bottom panel shows the reconstructed 3D image of the matched in-beam TPC activities and the truth trajectories of the neutrino interaction’s final state particles. In this example, there are two protons and an electron EM shower connected to the neutrino interaction vertex. A  $\pi^0$  is also created and decays into two  $\gamma$ ’s. The two  $\gamma$ ’s deposit energy through electrons from Compton scattering or pair production into electrons and positrons. A proper clustering of the two detached  $\gamma$ ’s is difficult considering the surrounding cosmic-ray muons. In this example, there is also a ghost track which crosses one proton track in the 2D projection, but it is actually detached from the proton track in the 3D space. It resides in the nonfunctional wire region and originates from part of a cosmic-ray muon track.

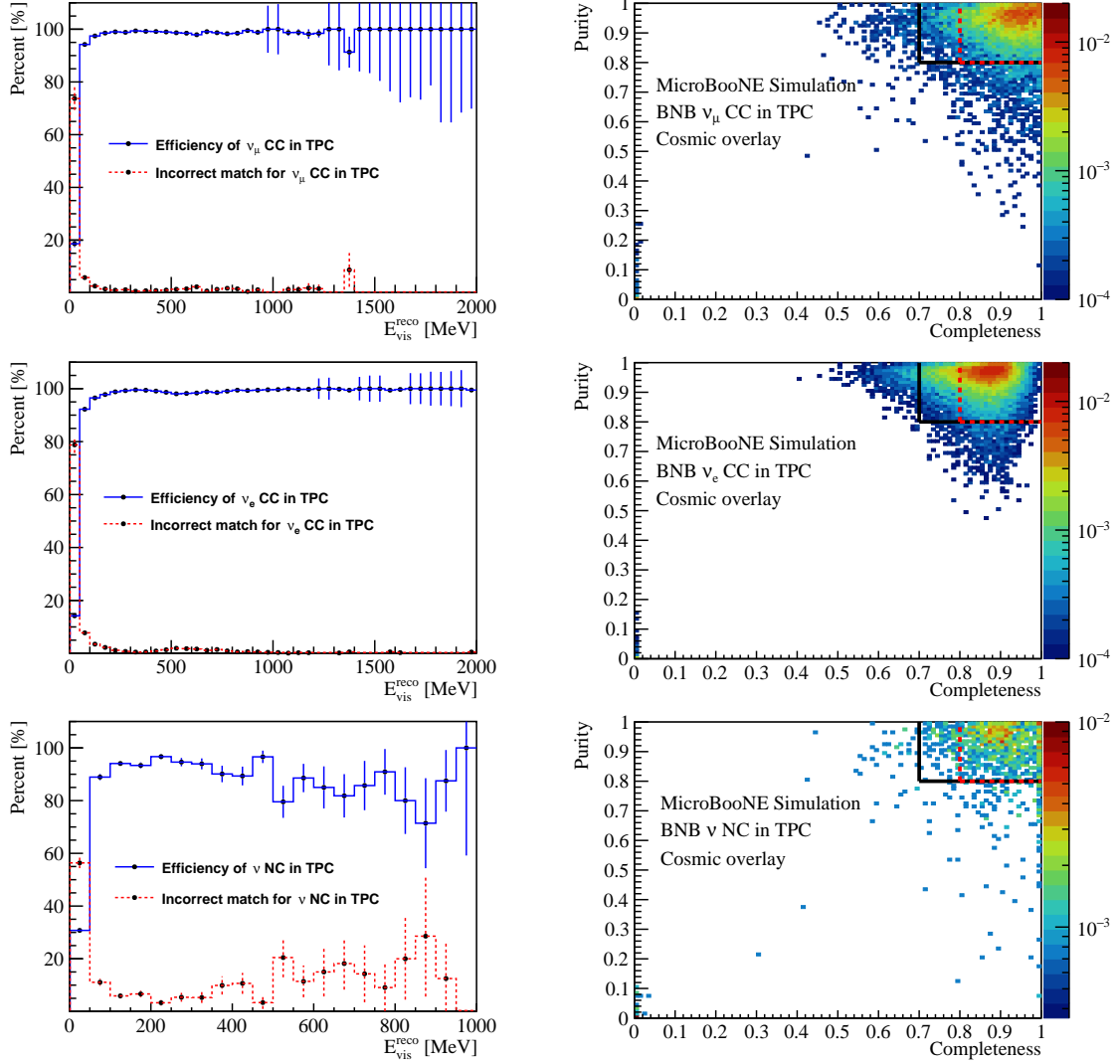
Without any pattern recognition or topological reconstruction at this stage, the completeness is a more critical metric than the purity. There is little chance to fix the incompleteness issue in the downstream analysis chain once the charge is already lost. However, the purity can be further improved. For example, the ghost track in figure 32 can be removed by checking the directionality, or by particle identification using  $dE/dx$  information, in which case this ghost track will be regarded as a cosmic-ray muon background.

Left panel of figure 33 shows the efficiency and incorrectness of the charge-light matching procedure. The overall efficiency to select the neutrino CC interactions in the TPC active volume is 95%, and the overall incorrectness values of charge-light matching are 4.6%, 3.8%, and 28.7%



**Figure 32:** Event display of a  $1e2p1\pi^0 \nu_e$  CC interaction. Top: side view of the full TPC readout; each cluster is labeled in one color. The black box corresponds to the LArTPC active volume with an X-position (converted from the readout time) relative to the neutrino interaction time. Bottom: the charge-light matching result – the in-beam flash matched TPC activities; the blue points are the reconstructed 3D space points and the red ones are the true space points corresponding to the neutrino interaction. The voxel size and opacity are tuned for event display.

965 for BNB  $\nu_\mu$  CC,  $\nu_e$  CC, and NC interactions, respectively. The efficiency plus incorrectness is  
 966 100% in this figure except for the first bin with low visible energy  $<50$  MeV in which case some  
 967 of events fail to match in-beam TPC activities to any PMT flash. A neutrino interaction, close to  
 968 the TPC boundary or with a significant number of neutral particles in the final states, tends to have  
 969 a large portion of its charges escaping the active TPC volume, which then become invisible to the  
 970 wire readout plane. However, the light signals originating from this neutrino interaction can still  
 971 be collected if there is any charge deposition outside the TPC but still in the liquid argon volume.  
 972 Such inconsistent TPC activities and PMT signals may result in incorrect matches or no matches.



**Figure 33:** 3D imaging and charge-light matching performance for BNB  $\nu_\mu$  CC,  $\nu_e$  CC, and NC interactions in the TPC. The neutrino interactions are simulated and overlaid with real data from cosmic rays. The clustering and charge-light matching steps are applied to select the neutrino interaction. Left: efficiency and incorrectness of charge-light matching as a function of the simply reconstructed visible energy (a simple conversion from the reconstructed visible charge using a constant conversion factor); binomial statistics is used to calculate the efficiency uncertainty while Poisson statistics (large error bars in the plot) is used where the efficiency is 100%, mainly for the low statistic bins. Right: purity vs. completeness for each selected neutrino interaction. The color scale (Z-axis value) represents the fraction of events. The integrated fraction of the events within the solid black and dashed red boxes can be found in table 4.



Right panel of figure 33 presents the completeness and purity of the selected TPC activities for BNB  $\nu_\mu$  CC,  $\nu_e$  CC, and NC interactions, respectively. The results of completeness and purity are summarized in table 4. The events with extremely low completeness and purity values as shown in the bottom left corner in each sub-figure of the right panel correspond to the incorrect charge-light matches as discussed previously. Comparing figure 31 and figure 33, the degradation

**Table 4:** Fraction of the events that correspond to the completeness and purity values within the black or red boxes as shown in figure 33. These numbers are the overall performance for the integrated BNB neutrino flux which has an average neutrino energy of about 800 MeV. All neutrino interactions are simulated within the TPC active volume, with cosmic-ray data (beam-off) overlaid.

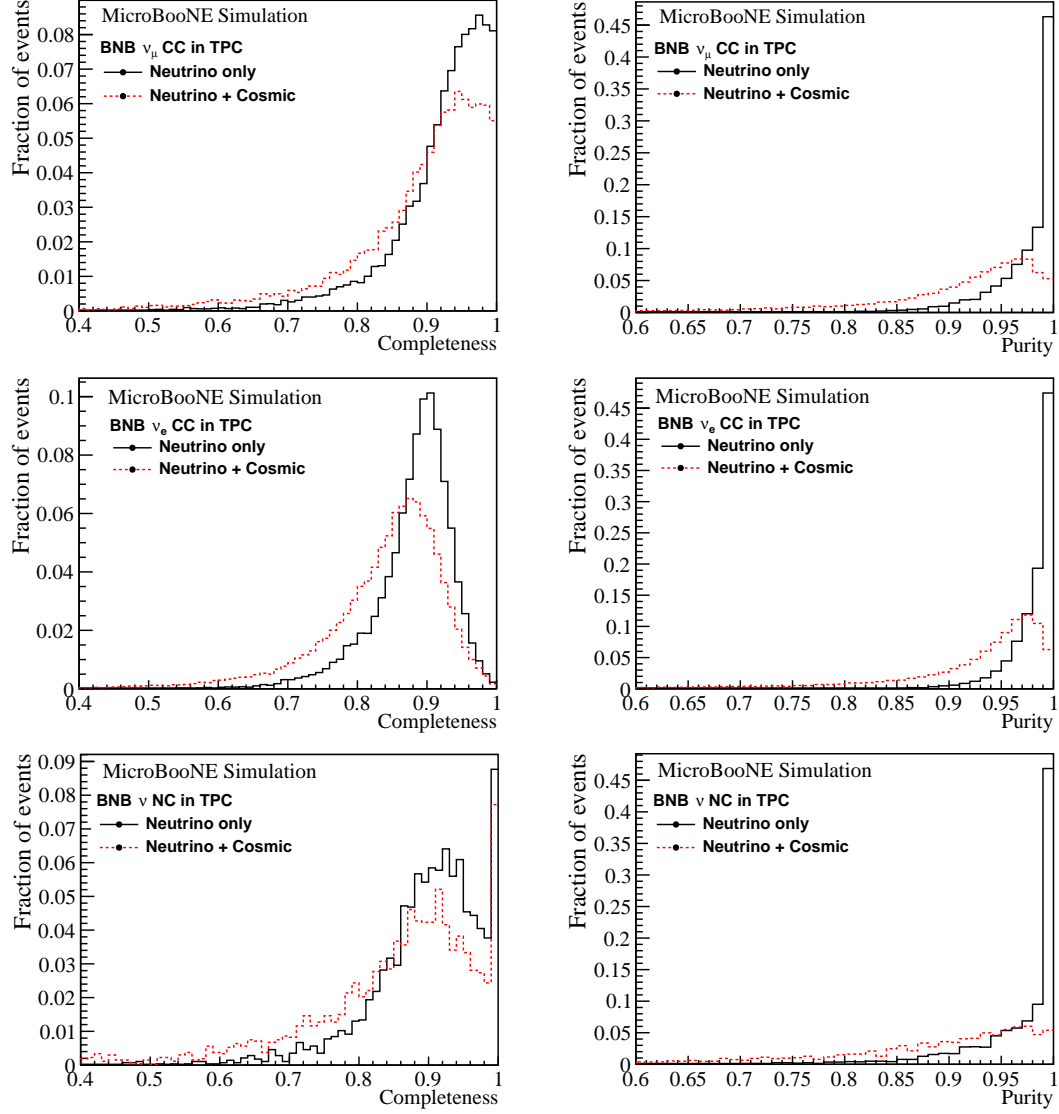
Scenario: neutrino + cosmic	BNB $\nu_\mu$ CC	BNB $\nu_e$ CC	BNB NC
Purity >80% and Completeness > 80%	73.0%	67.7%	56.0%
Completeness > 70%	80.2%	83.4%	66.5%

of the completeness and the purity can be attributed to the numerous cosmic-ray muons that traverse the detector. Direct comparisons of the completeness and the purity are independently performed, and the distributions can be seen in figure 34. The scenarios of “neutrino-only” and “neutrino + cosmic” correspond to figure 31 and figure 33, respectively. In the scenario of “neutrino + cosmic”, the neutrino activities suffer not only an over-clustering issue with the cosmic-ray activities (or its related ghost tracks) but also an under-clustering issue since part of the detached activities from the neutrino primary cluster may be grouped to cosmic-ray muons. These two issues introduce a smearing of both the completeness and purity distributions. The typical values of the completeness and the purity for different scenarios and interaction types are summarized in table 5.

**Table 5:** Summary of typical completeness and purity values corresponding to the distributions as shown in figure 34. Independent comparisons of completeness and purity are performed. The numbers are given as the fraction of the corresponding events. All neutrino interactions are simulated within the TPC active volume.

	Scenario	BNB $\nu_\mu$ CC	BNB $\nu_e$ CC	BNB NC
Completeness >80% (>70%)	Neutrino + cosmic	84.5% (93.4%)	74.9% (92.5%)	73.1% (87.9%)
	Neutrino only	92.8% (97.8%)	90.8% (98.5%)	90.2% (97.4%)
Purity >80%	Neutrino + cosmic	84.4%	89.2%	72.6%
	Neutrino only	99.4%	99.8%	97.1%

About 85% of events have at least 80% completeness for BNB  $\nu_\mu$  CC interactions. About 90% of events have at least 70% completeness for BNB  $\nu_e$  CC or NC interactions. The degradation of purity in the scenario of “neutrino + cosmic” is more severe than the degradation of completeness, as expected. However, as explained previously, the completeness is more critical in the 3D imaging and charge-light matching stage, since the purity can be further improved in the later analysis chain.



**Figure 34:** Independent comparisons of completeness and purity distributions for the two scenarios of “neutrino-only” and “neutrino + cosmic”. The 3D clustering and charge-light matching steps are applied in the scenario of “neutrino + cosmic” to select the neutrino interaction. Top: BNB  $\nu_\mu$  CC interactions in the TPC. Middle: BNB  $\nu_e$  CC interactions in the TPC. Bottom: BNB  $\nu$  NC interactions in the TPC; the “ $\sim 100\%$ ” completeness peak value mainly corresponds to the NC quasi-elastic scattering with a single low-energy (short) proton emitted and it can also be seen in the bottom right panel of figure 31 for low-energy neutrino NC interactions. See text for more details.

992 In the optimization of the clustering and charge-light matching algorithms, the completeness is thus  
993 considered more favorably than the purity. In reality, the final purity performance corresponding to  
994 the scenario of “neutrino + cosmic” is still very good and 80%-90% events have at least 80% purity  
995 for the CC interactions.

996 In summary, the quantitative evaluations of the Wire-Cell 3D imaging, clustering, and charge-  
997 light matching have been presented in this section. These techniques result in a high performance  
998 selection of the neutrino activities in the MicroBooNE LArTPC with a clean removal of the 20–30  
999 cosmic-ray muons within a TPC readout. The quality (completeness and purity) of the 3D images  
1000 of the selected in-beam neutrino activities is very good considering the complexities from the wire  
1001 readout ambiguity, nonfunctional wires, non-perfect signal processing, and numerous cosmic ray  
1002 muons.

## 6 Summary and Discussion

This article describes the principle and algorithms of the Wire-Cell 3D imaging, clustering, and many-to-many charge-light matching applied in the MicroBooNE LArTPC. The 3D imaging tomographically reconstructs the 3D image of the ionization electrons using the fundamental information of charge, time, and geometry of the LArTPC detector. Other characteristics of the LArTPC physics activities such as sparsity, positivity, and proximity are utilized as additional constraints to improve the 3D imaging performance. The many-to-many charge-light matching with 3D clustering and light signal reconstruction is developed to pair the TPC clusters and PMT flashes to identify the neutrino interaction among numerous cosmic-ray muons. Several realistic issues, e.g. the nonfunctional wires, the gaps due to inefficient signal processing, detached neutrino activities, and coincidentally connected clusters, are properly addressed. Using the MicroBooNE detector simulation, the realistic performance of the reconstruction techniques is evaluated.

In spite of the effort, there are some limitations in the 3D imaging as shown in the event displays in this paper. For example, prolonged tracks, which often develop gaps in the signal processing stage, cannot be entirely fixed via the bridging of the gaps as implemented. Similar issues occur for the isochronous tracks that often develop gaps because of the insufficient coherent noise filtering. Subsequent pattern recognition techniques, e.g. particle-level clustering and trajectory fitting, may further address this problem. Isochronous tracks present another common problem for the LArTPC 3D imaging, as the wire readout ambiguity is drastically increased in the time slice containing them. In Wire-Cell 3D imaging, this issue is mitigated by introducing tiling, however, the blobs of the isochronous track are significantly broadened, leading to a much worse spatial resolution in the Y-Z plane view. This issue can be further mitigated via trajectory fitting in a later stage.

Generally speaking, the 3D event reconstruction techniques as presented in this paper are adequately accurate and efficient, and can successfully select neutrino interaction activities. About 95% of the neutrino CC interactions in the TPC active volume are selected, with a 30-fold reduction of non-beam-coincident cosmic-ray muons. Good completeness and purity of the resulting 3D image of the selected neutrino activities have been achieved. Greater than 80% of the selected neutrino CC interactions have a reconstructed 3D image of at least 70% completeness and 80% purity. These techniques will benefit the downstream pattern recognition and neutrino selection, and they are important steps towards realizing the full capability of single-phase LArTPCs. In particular, the Wire-Cell based neutrino selection and analyses take full advantage of these tools to further reject cosmic muons and select neutrinos [15] and demonstrate a very promising high efficiency and high purity neutrino selection in LArTPCs. Other analyses using techniques such as deep learning with convolutional neural networks [9, 10] and Pandora multi-algorithm pattern recognition [8] can also benefit from the outcome of the Wire-Cell 3D reconstruction tools, as it provides a clean and intact 3D image of the in-beam neutrino activities with the surround cosmic ray muons removed.

## Acknowledgments

This document is 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 is provided by the Albert Einstein Center for Fundamental Physics, Bern, Switzerland.

## References

- [1] C. Rubbia. *The liquid argon time projection chamber: A new concept for neutrino detector*, . CERN-EP/77-08, 1977.
- [2] H. H. Chen, P. E. Condon, B. C. Barish and F. J. Sciulli. *A Neutrino detector sensitive to rare processes. I. A Study of neutrino electron reactions*, . FERMILAB-PROPOSAL-0496, 1976.
- [3] W. J. Willis and V. Radeka, *Liquid argon ionization chambers as total absorption detector*, *Nucl. Instrum. Meth.* **120** (1974) 221.
- [4] D. R. Nygren, *The Time Projection Chamber: A New 4 pi Detector for Charged Particles*, *eConf C740805* (1974) 58.
- [5] MICROBooNE collaboration, R. Acciarri et al., *Design and Construction of the MicroBooNE Detector*, *JINST* **12** (2017) P02017, [[arXiv: 1612.05824](#)].
- [6] P. A. Machado, O. Palamara and D. W. Schmitz, *The Short-Baseline Neutrino Program at Fermilab*, *Ann. Rev. Nucl. Part. Sci.* **69** (2019) , [[arXiv:1903.04608](#)].
- [7] DUNE collaboration, B. Abi et al., *Volume I. Introduction to DUNE*, *JINST* **15** (2020) T08008, [[arXiv:2002.02967](#)].
- [8] MICROBooNE collaboration, R. Acciari 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** (2018) 82, [[arXiv:1708.03135](#)].
- [9] MICROBooNE collaboration, R. Acciarri et al., *Convolutional neural networks applied to neutrino events in a liquid argon time projection chamber*, *JINST* **12** (2017) P03011, [[arXiv:1611.05531](#)].
- [10] MICROBooNE collaboration, C. Adams et al., *Deep neural network for pixel-level electromagnetic particle identification in the MicroBooNE liquid argon time projection chamber*, *Phys. Rev.* **D99** (2019) 092001, [[arXiv:1808.07269](#)].
- [11] X. Qian, C. Zhang, B. Viren and M. Diwan, *Three-dimensional Imaging for Large LArTPCs*, *JINST* **13** (2018) P05032, [[arXiv:1803.04850](#)].
- [12] MICROBooNE collaboration, R. Acciarri et al., *Noise Characterization and Filtering in the MicroBooNE Liquid Argon TPC*, *JINST* **12** (2017) P08003, [[arXiv:1705.07341](#)].
- [13] MICROBooNE collaboration, C. Adams et al., *Ionization electron signal processing in single phase LArTPCs. Part I. Algorithm Description and quantitative evaluation with MicroBooNE simulation*, *JINST* **13** (2018) P07006, [[arXiv:1802.08709](#)].

- [14] MICROBooNE collaboration, C. Adams et al., *Ionization electron signal processing in single phase LArTPCs. Part II. Data/simulation comparison and performance in MicroBooNE*, *JINST* **13** (2018) P07007, [[arXiv:1804.02583](#)].
- [15] MICROBooNE collaboration, P. Abratenko et al., *Cosmic Ray Background Rejection with Wire-Cell LArTPC Event Reconstruction in the MicroBooNE Detector*, [arXiv:2101.05076](#).
- [16] MINIBOO NE collaboration, A. Aguilar-Arevalo et al., *The Neutrino Flux prediction at MiniBooNE*, *Phys. Rev. D* **79** (2009) 072002, [[arXiv:0806.1449](#)].
- [17] MICROBooNE collaboration, R. Acciarri et al., *Construction and Assembly of the Wire Planes for the MicroBooNE Time Projection Chamber*, *JINST* **12** (2017) T03003, [[arXiv: 1609.06169](#)].
- [18] MICROBooNE collaboration, J. Conrad, B. J. P. Jones, Z. Moss, T. Strauss and M. Touns, *The Photomultiplier Tube Calibration System of the MicroBooNE Experiment*, *JINST* **10** (2015) T06001, [[arXiv:1502.04159](#)].
- [19] E. J. Candès, J. K. Romberg and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, *Communications on Pure and Applied Mathematics* **59** (8) (2006) 1207–1223, [[arXiv:math/0503066](#)].
- [20] J. Friedman, T. Hastie and R. Tibshirani, *Regularization paths for generalized linear models via coordinate descent*, *Journal of Statistical Software* **33** (1) (2010) 1–22.
- [21] Wire-Cell git repository for sparse signal regression, <https://github.com/BNLIF/wire-cell-ress/>.
- [22] H. Yu et al., *Augmented Signal Processing in Liquid Argon Time Projection Chambers with a Deep Neural Network*, [arXiv:2007.12743](#).
- [23] MICROBooNE collaboration, D. Caratelli, *Neutrino identification with scintillation light in MicroBooNE*, *JINST* **15** (2020) C03023.
- [24] J. L. Blanco and P. K. Rai, “nanoflann: a C++ header-only fork of FLANN, a library for nearest neighbor (NN) with kd-trees.” <https://github.com/jlblancoc/nanoflann>, 2014.
- [25] P. Hough, “Method and means for recognizing complex patterns.” US Patent 3069654, 1962.
- [26] quickhull. <https://github.com/akuukka/quickhull>, 2017.
- [27] E. W. Dijkstra, *A note on two problems in connexion with graphs*, *Numerische mathematik* **1** (1959) 269–271.
- [28] F. J. Massey Jr., *The kolmogorov-smirnov test for goodness of fit*, *Journal of the American Statistical Association* **46** (1951) 68–78.
- [29] Ben Jones, *Sterile Neutrinos in Cold Climates*, FERMILAB-THESIS-2015-17, Massachusetts Institute of Technology, 2015.
- [30] MICROBooNE collaboration, P. Abratenko et al., *Measurement of Space Charge Effects in the MicroBooNE LArTPC Using Cosmic Muons*, *JINST* **15** (2020) P12037, [[arXiv:2008.09765](#)].
- [31] LArSoft release v08.05.00.10, <https://cdcv.s.fnal.gov/redmine/projects/larsoft/wiki/ReleaseNotes08050010>.
- [32] Uboonecode release v08.00.00.28, <https://cdcv.s.fnal.gov/redmine/projects/uboonecode/wiki/ReleaseNotes08000028>.
- [33] C. Andreopoulos, C. Barry, S. Dytman, H. Gallagher, T. Golan, R. Hatcher et al., *The GENIE Neutrino Monte Carlo Generator: Physics and User Manual*, [arXiv:1510.05494](#).

- 1121 [34] GEANT4 collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, *Nucl. Instrum. Meth.*  
1122 [A506](#) (2003) 250–303.
- 1123 [35] J. Allison et al., *Recent developments in Geant4*, *Nucl. Instrum. Meth.* **A835** (2016) 186–225.
- 1124 [36] MICROBooNE collaboration, P. Abratenko et al., *High-performance Generic Neutrino Detection in a*  
1125 *LArTPC near the Earth’s Surface with the MicroBooNE Detector*, [arXiv:2012.07928](#).