

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

34 ^aUniversität Bern, Bern CH-3012, Switzerland

35 ^bBrookhaven National Laboratory (BNL), Upton, NY, 11973, USA

36 ^cUniversity of California, Santa Barbara, CA, 93106, USA

37 ^dUniversity of Cambridge, Cambridge CB3 0HE, United Kingdom
38 ^eSt. Catherine University, Saint Paul, MN 55105, USA
39 ^fCentro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid E-28040,
40 Spain
41 ^gUniversity of Chicago, Chicago, IL, 60637, USA
42 ^hUniversity of Cincinnati, Cincinnati, OH, 45221, USA
43 ⁱColorado State University, Fort Collins, CO, 80523, USA
44 ^jColumbia University, New York, NY, 10027, USA
45 ^kFermi National Accelerator Laboratory (FNAL), Batavia, IL 60510, USA
46 ^lUniversidad de Granada, E-18071, Granada, Spain
47 ^mHarvard University, Cambridge, MA 02138, USA
48 ⁿIllinois Institute of Technology (IIT), Chicago, IL 60616, USA
49 ^oKansas State University (KSU), Manhattan, KS, 66506, USA
50 ^pLancaster University, Lancaster LA1 4YW, United Kingdom
51 ^qLos Alamos National Laboratory (LANL), Los Alamos, NM, 87545, USA
52 ^rThe University of Manchester, Manchester M13 9PL, United Kingdom
53 ^sMassachusetts Institute of Technology (MIT), Cambridge, MA, 02139, USA
54 ^tUniversity of Michigan, Ann Arbor, MI, 48109, USA
55 ^uUniversity of Minnesota, Minneapolis, Mn, 55455, USA
56 ^vNew Mexico State University (NMSU), Las Cruces, NM, 88003, USA
57 ^wOtterbein University, Westerville, OH, 43081, USA
58 ^xUniversity of Oxford, Oxford OX1 3RH, United Kingdom
59 ^yPacific Northwest National Laboratory (PNNL), Richland, WA, 99352, USA
60 ^zUniversity of Pittsburgh, Pittsburgh, PA, 15260, USA
61 ^{aa}Rutgers University, Piscataway, NJ, 08854, USA, PA
62 ^{bb}Saint Mary's University of Minnesota, Winona, MN, 55987, USA
63 ^{cc}SLAC National Accelerator Laboratory, Menlo Park, CA, 94025, USA
64 ^{dd}South Dakota School of Mines and Technology (SDSMT), Rapid City, SD, 57701, USA
65 ^{ee}University of Southern Maine, Portland, ME, 04104, USA
66 ^{ff}Syracuse University, Syracuse, NY, 13244, USA
67 ^{gg}Tel Aviv University, Tel Aviv, Israel, 69978
68 ^{hh}University of Tennessee, Knoxville, TN, 37996, USA
69 ⁱⁱUniversity of Texas, Arlington, TX, 76019, USA
70 ^{jj}Tufts University, Medford, MA, 02155, USA
71 ^{kk}Center for Neutrino Physics, Virginia Tech, Blacksburg, VA, 24061, USA
72 ^{ll}University of Warwick, Coventry CV4 7AL, United Kingdom
73 ^{mm}Wright Laboratory, Department of Physics, Yale University, New Haven, CT, 06520, USA
74 E-mail: 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	Contents	
93	1 Introduction	1
94	2 The MicroBooNE detector	2
95	3 Wire-Cell 3D Imaging	4
96	3.1 Tiling	5
97	3.2 Charge solving	8
98	3.3 De-ghosting	11
99	3.4 Summary	12
100	4 Matching Charge and Light	16
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	5 Evaluation of the Wire-Cell 3D imaging and the charge-light matching	36
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	6 Summary and Discussion	49

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

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

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

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

160 **2 The MicroBooNE detector**

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

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

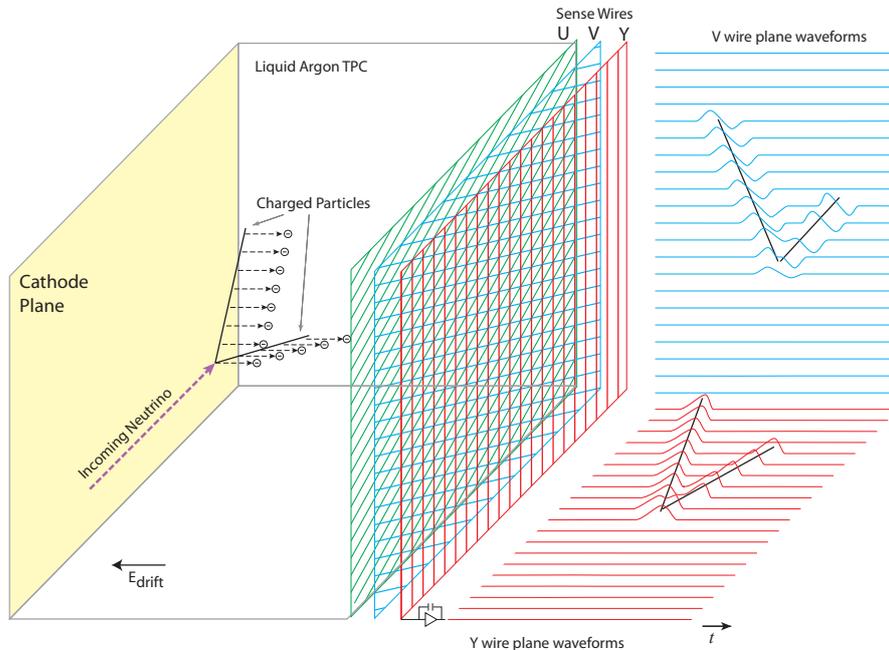


Figure 1: Illustration of single-phase LArTPCs [5]. Each wire plane provides a 2D image of the ionization electrons with respect to a specific wire orientation.

168 A high voltage of -70 kV applied on the cathode plane provides a drift field of 273 V/cm. The
 169 electrons ionized by any energy deposition from traversing charged particles drift towards the
 170 wire planes along the electric field at a nominal speed of about 1.10 mm/ μ s. In this paper, we use
 171 the X-axis to represent the direction away from the readout wire plane and opposite to the ionization
 172 charge drift, Y-axis to represent the vertical-up direction, and Z-axis to represent the BNB beam
 173 direction. There are three parallel wire readout planes [17] on the anode side with different wire
 174 orientations. The first wire plane facing the cathode is labeled “U”, and the second and third plane
 175 are labeled “V” and “Y”, respectively. The 3456 wires in the Y plane are oriented vertically and
 176 the 2400 wires in the U (V) plane are oriented $+(-)60^\circ$ with respect to the vertical direction. The
 177 spacing between adjacent wires and adjacent wire planes are both 3 mm. Different bias voltages,
 178 -110 V, 0 V, and 230 V, are applied to the U, V, and Y wire planes, respectively, to ensure all
 179 ionization electrons drift through the U and V planes before being collected by the Y plane. The U
 180 and V planes are commonly referred to as the induction planes and the ionization electrons induce
 181 bipolar electrical signals as they pass through the planes; the Y plane is referred to as the collection
 182 plane and sees unipolar electrical pulses.

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

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

198 3 Wire-Cell 3D Imaging

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

203 The fundamental information provided by a LArTPC is as follows:

- 204 (i) **Time** - when the ionization electrons arrive at the anode wire plane¹.
- 205 (ii) **Geometry** - the positions of the wires from each plane that have signals from the ionization
206 electrons, i.e. hit wires.
- 207 (iii) **Charge** - the number of ionization electrons measured by the hit wires from each wire plane.

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

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

¹The 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

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

229 (iv) **Sparsity** - the distribution of ionization electrons in space is expected to be sparse, typically
230 occupying less than 10% of the local bounding volume that contains the activities, for any
231 physical signals.

232 (v) **Proximity** - the ionization electrons are read out by consecutive wires because a charged
233 particle ionizes argon atoms continuously in the fully active LArTPC volume.

234 (vi) **Positivity** - the number of the drifted ionization electrons can only be positive.

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

240 3.1 Tiling

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

247 The smallest time unit in the Wire-Cell imaging is a time slice, whose $2 \mu\text{s}$ width contains four
248 sampling ticks from the TPC readout. The width of the time slice introduces negligible information
249 loss because the software filtering in the signal processing has a cut-off frequency at about 0.25 MHz
250 to optimize the signal-to-noise ratio, which in turn smears the time resolution. Geometry is used
251 to determine all possible hit cells within a time slice by finding the intersections of the hit wires.
252 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
253 (1.8 cm wide), leading to 55 possible hit cells. The fact that there are fewer knowns (19 hit wires)
254 than unknowns (55 cells) indicates an ambiguity is introduced by the wire readout. Meanwhile, the
255 amount of integrated charge in a time slice and the identities of active wires in that time slice are
256 affected by diffusion and long-range induction effects (especially for induction planes) as charge
257 drifts in the TPC, as well as the action of software filters applied to the waveforms [13, 14]. To
258 mitigate the impact from wire ambiguity and charge smearing, a procedure to merge the consecutive
259 hit cells is developed, called *tiling*. The groups of hit cells after tiling are called *blobs*. The blob in
260 figure 2 is marked by solid blue lines. The connected hit wires are merged as *wire bundles* in the
261 tiling procedure, and a blob is the overlapping area of three wire bundles from each wire plane as
262 shown in figure 2. Note that a cell or a blob can be taken as a 3D object and its length along the
263 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.

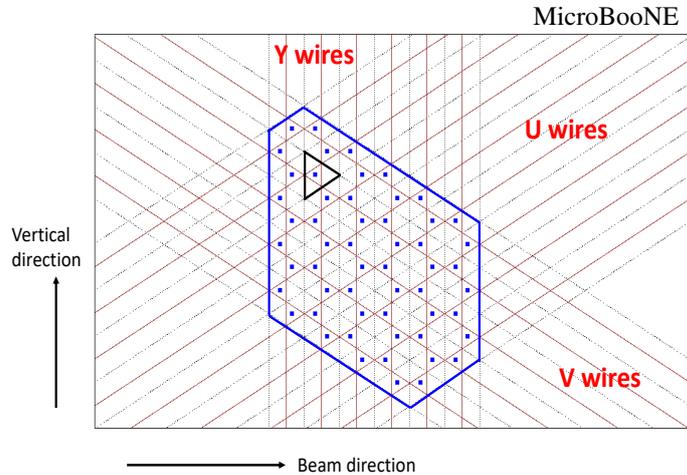


Figure 2: An example of the hit cells and blob constructed by the hit wires with the MicroBooNE detector geometry. Each wire is represented by a solid red line and the wire (pitch) boundaries are represented by dashed black lines. All hit cells have equilateral triangular shapes and are marked with blue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

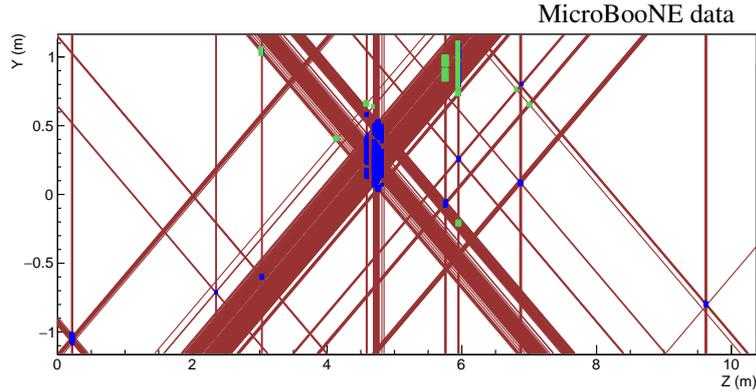


Figure 3: An example event with hit wires and blobs after applying the deghosting algorithm (see section 3.3). Blobs are marked in blue or green. Blue blobs correspond to 3-plane tiling requiring all three wire planes to be functional. Green blobs correspond to the additional blobs created in 2-plane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

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

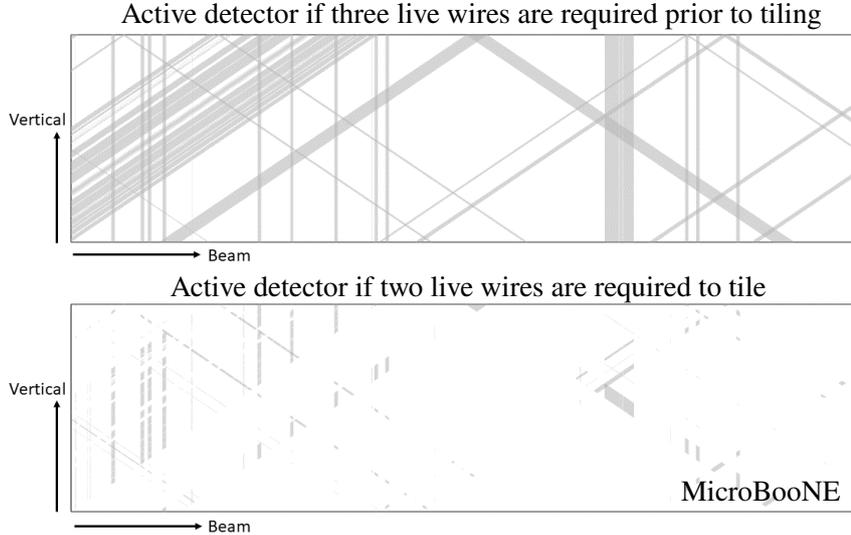


Figure 4: Impact of the nonfunctional wires (gray) on the anode plane. The borders of the two figures correspond to the boundaries of the LArTPC active volume. Top: the gray area that has at least one wire nonfunctional is 30%. Bottom: the gray area that has at least two wires nonfunctional is 3%.

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

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

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

304 3.2 Charge solving

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

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

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

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

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

321 where $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$ is the ℓ_p -norm of a vector x . Since the physics activities in LArTPCs
 322 are generally sparse, i.e. most of the elements of x are zero, the ℓ_0 -norm (a count of the nonzero
 323 elements) can be used to seek the most sparse or the simplest solution that explains the measurements.
 324 The minimization of $\|x\|_0$ can be achieved by removing the unknowns until the linear equation
 325 is solvable. For example in figure 3, there are 25 blobs, while only about 10 hits are true. One
 326 can remove 15 unknowns² out of the 25 to solve the equation and find the “best” one satisfying

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

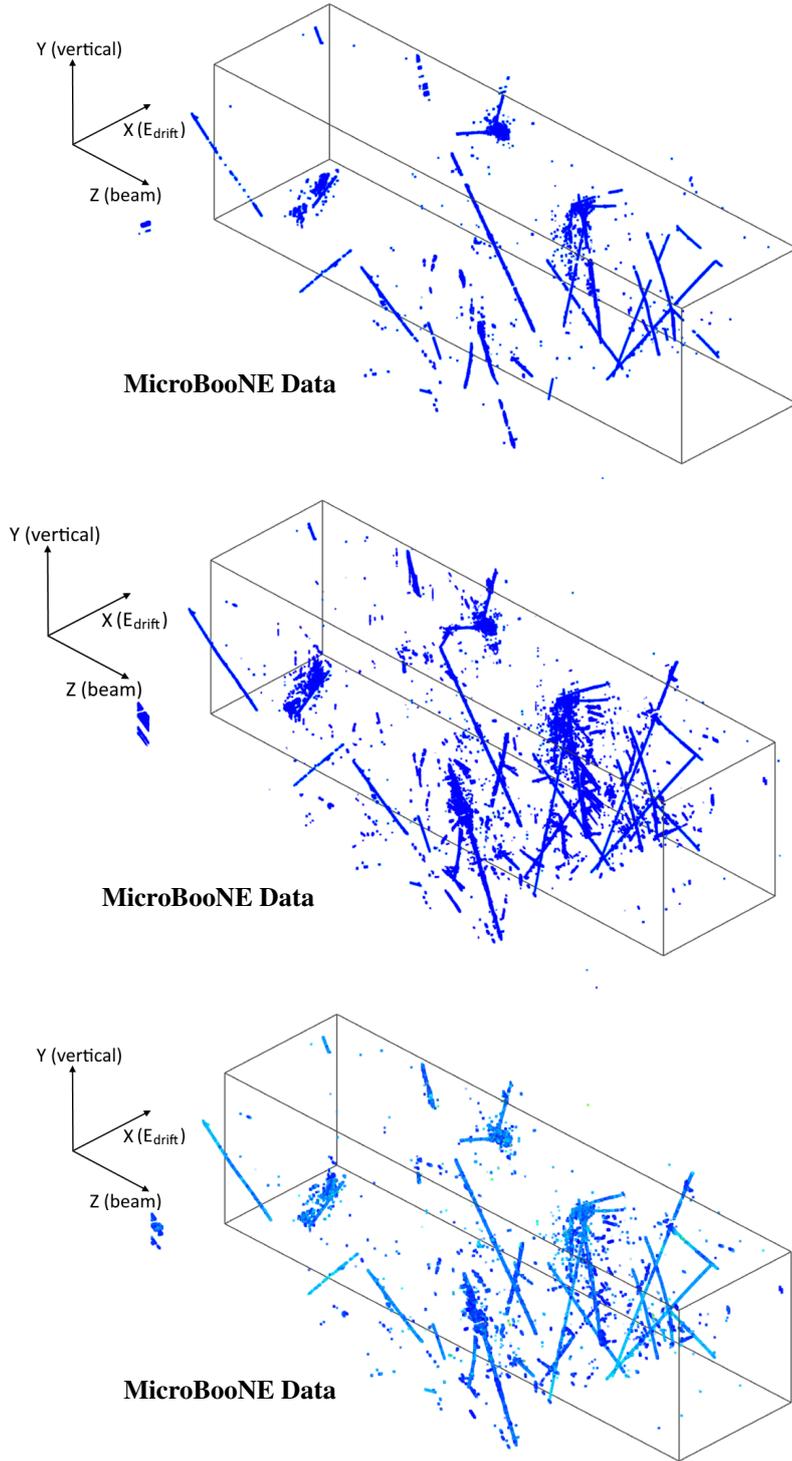


Figure 5: Comparison of the tiling results and the charge solving result from MicroBooNE data (event 41075, run 3493). The solid black box represents the LArTPC active volume with an X-position (converted from the readout time) relative to the neutrino interaction time. Only time and geometry information 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.

327 the optimization condition. However, in this case there are $C_{25}^{10} \approx 3.3 \times 10^6$ combinatorial ways
 328 to remove the unknowns and in general this optimization is an NP-hard problem that is extremely
 329 expensive in computation. Mathematicians [19] have discovered that an alternative constraint,
 330 the ℓ_1 -norm, can well approximate the ℓ_0 -norm result with a much faster minimization. This ℓ_1
 331 technique, also known as *compressed sensing*, is widely applied in many other fields for signal
 332 processing and computational photography. As shown in section 4.3, the compressed sensing
 333 technique is also used to perform the many-to-many charge-light matching.

334 In practice, a chi-square function is constructed to take into account the uncertainties of the
 335 measured charge from signal processing [14], and the compressed sensing technique is implemented
 336 by adding an ℓ_1 -regularization term to the chi-square function:

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

337 where the vector y and x are pre-normalized through $V^{-1} = Q^T Q$ (Cholesky decomposition),
 338 $y' = Q \cdot y$, $A' = Q \cdot A$, and λ regulates the strength of $\|x\|_1$. The matrix V is the real symmetric
 339 covariance matrix of the charge measurement uncertainties. The ℓ_1 -regularized chi-square function
 340 is convex with a unique global minimum, enabling fast minimization algorithms such as coordinate
 341 descent [20]. An implementation of the coordinate descent method can be found in a Wire-Cell git
 342 repository [21]. Another constraint, *positivity* of the charge (number of the ionization electrons), is
 343 added in the coordinate descent method to help remove the ghosts.

344 So far we have shown the incorporation of charge, sparsity, and positivity to seek the unique
 345 solution to the imaging equation of the first principle. To further improve the robustness of the ℓ_1 -
 346 regularization result, *proximity* information is incorporated given the fact that the LArTPC is a fully
 347 active detector, therefore the measured activities from charged particles are spatially continuous,
 348 in contrast to other sampling detectors. For those adjacent blobs over different time slices, the
 349 regularization strength λ is applied with an additional scaling factor of a^n to lower the chance of
 350 removing the corresponding element in x during the ℓ_1 minimization. n represents the number of
 351 the adjacent blobs that are connected to the target blob, and a is a predefined scaling factor. The
 352 final chi-square function in the Wire-Cell imaging is transformed to be:

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

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

357 The bottom panel of figure 5 shows the result after applying the charge solving procedure to
 358 the 2-plane tiling result in the middle panel. It is clear that the ghosts are further reduced and the
 359 3D voxels are now associated with different charge values, which correspond to the solution x of
 360 the imaging equation of the first principle. As elaborated in section 4.3, such 3D charge solving
 361 is critical to predict the scintillation light signals for each PMT, allowing for comparison to and
 362 matching with the observed light information.

363 3.3 De-ghosting

364 The amount of ghosts is considerably reduced after the charge solving but the result is still unsat-
365 isfactory. The sparsity combined with the proximity is already incorporated in the charge solving
366 to resolve the wire readout ambiguity; however, this procedure is performed in a “local” manner
367 restricted within each time slice or over adjacent time slices. Within the 3D imaging, all connected
368 blobs in 3D space are grouped together as *proto-clusters*. A proto-cluster does not necessarily
369 group all related TPC activities from a cosmic-ray muon or a neutrino interaction, since there might
370 be true or artificial gaps in the 3D image. The principle of the sparsity of the LArTPC physics
371 activities will be further used in a “global” manner to reconstruct the sparsest 3D images of the TPC
372 activities by removing the less prominent proto-clusters that are redundant to explain the observed
373 2D-projection measurements from wire planes. Following this philosophy, a dedicated algorithm,
374 *deghosting*, is developed to remove the residual ghosts based on their two main characteristics.

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

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

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

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

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

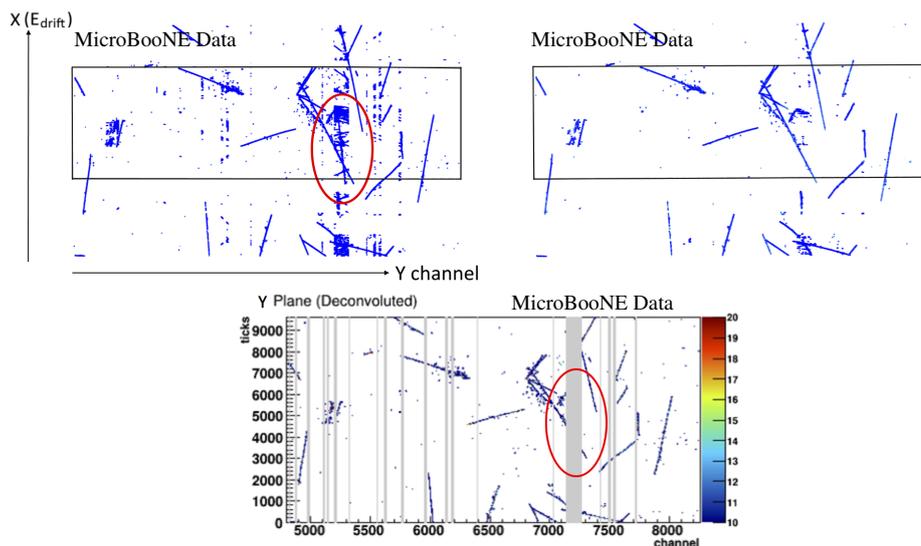


Figure 6: Top left: 2D projection to the Y plane’s wire-versus-time view of the reconstructed 3D image without the de-ghosting algorithm. The black box represents the full detector length in Y and the full cathode-to-anode drift distance in X. The red circle corresponds to the same volume in the TPC 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.

405 direction) because of the bipolar cancellation of the induction plane signals³. Consequently, one
 406 or two of the 2D wire-versus-time views of the charge measurements may have gaps along a track
 407 even on the functional wire planes. This gap will lead to a separation in the 3D image since the
 408 charge measurements across the wire planes can no longer match. Consequently, the successfully
 409 reconstructed charges from the other wire planes corresponding to the gap could interplay with the
 410 charge measurements from other tracks and be erroneously explained by ghosts. The removal of
 411 such ghosts requires a bridging of the gaps to connect the separated pieces of the track. This will
 412 be further discussed in section 4.1 in the context of 3D clustering, which results in the final TPC
 413 clusters based on the proto-clusters.

414 3.4 Summary

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

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

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

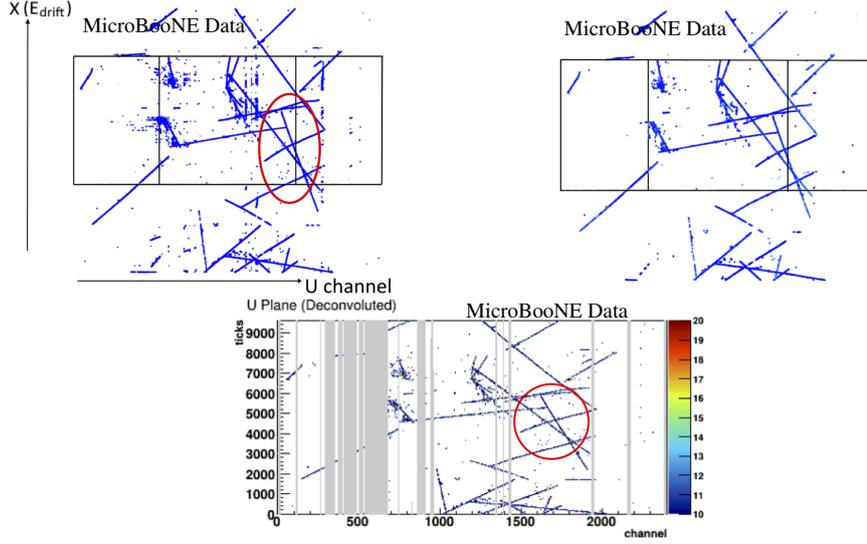


Figure 7: Top left: 2D projection to the U plane’s wire-versus-time view of the reconstructed 3D image without the de-ghosting algorithm. The black box represents the full array of U channels and the full cathode-to-anode drift distance in X. The red circle corresponds to the same volume in the TPC as 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 (≥ 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

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

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

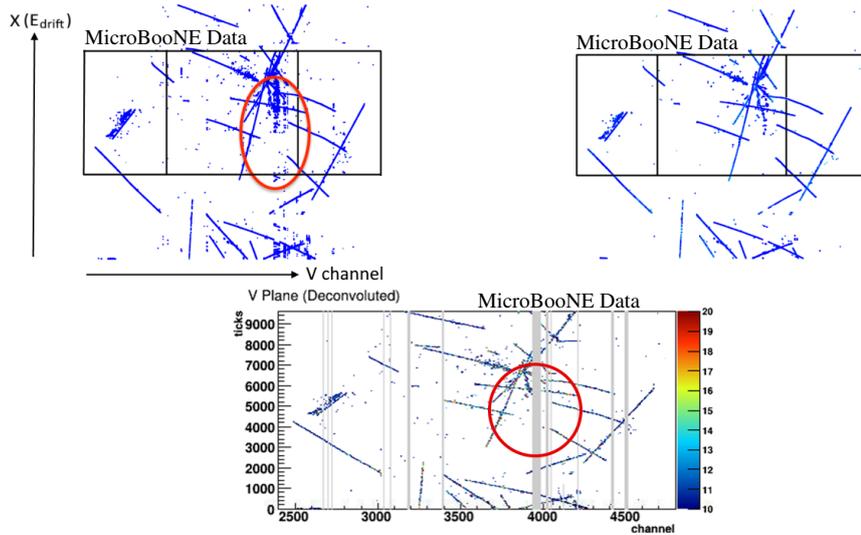


Figure 8: Top left: 2D projection to the V plane’s wire-versus-time view of the reconstructed 3D image without the de-ghosting algorithm. The black box represents the full array of V channels and the full cathode-to-anode drift distance in X. The red circle corresponds to the same volume in the TPC as 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.

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

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

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

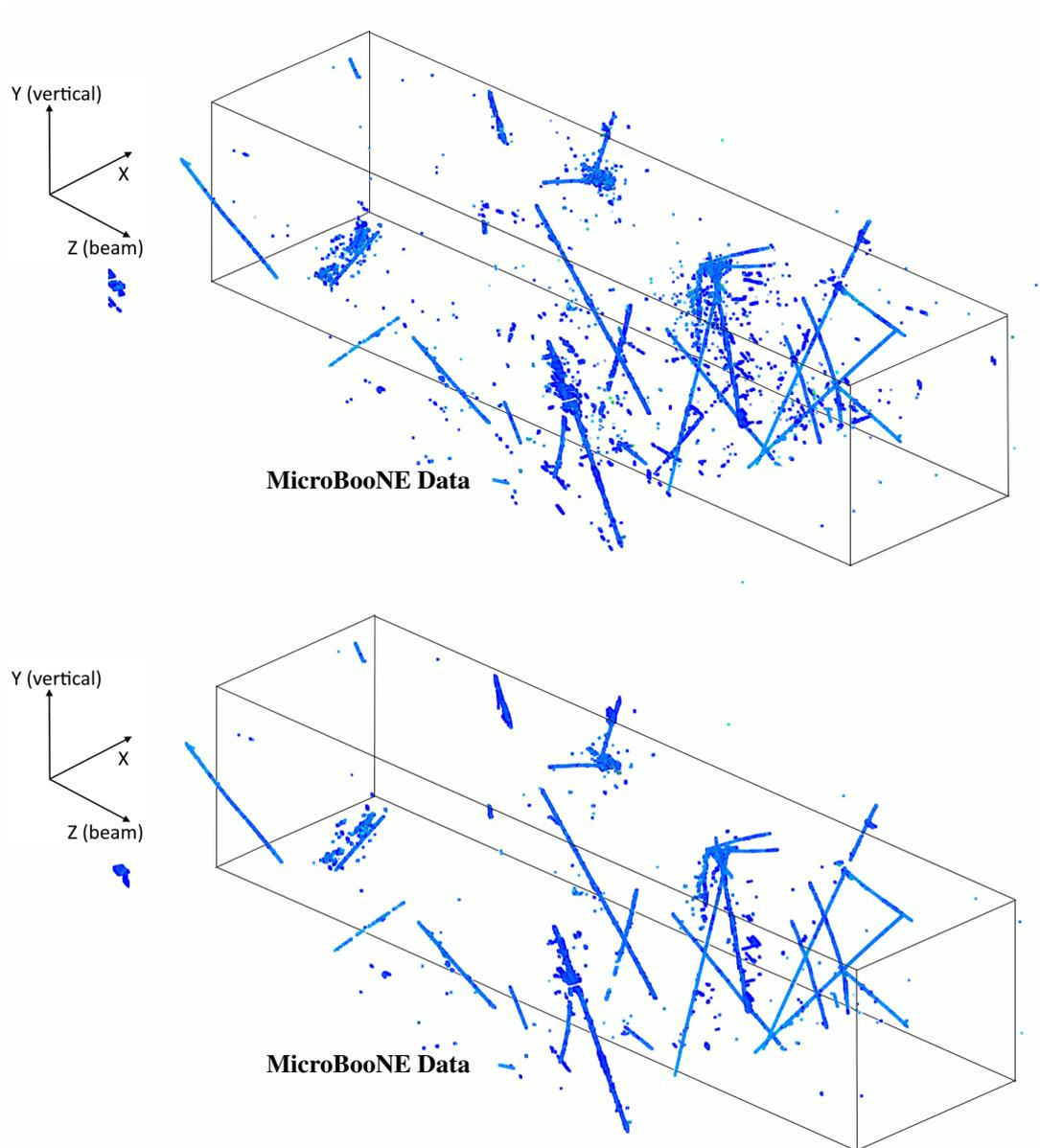


Figure 9: Comparison of the 3D imaging results from MicroBooNE data (event 41075, run 3493) without (top) and with (bottom) the de-ghosting algorithm. Ghosts are significantly reduced after the de-ghosting. The solid black box represents the LArTPC active volume with an X-position (converted from the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

449 4 Matching Charge and Light

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

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

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

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

479 4.1 3D clustering

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

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

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

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

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

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

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

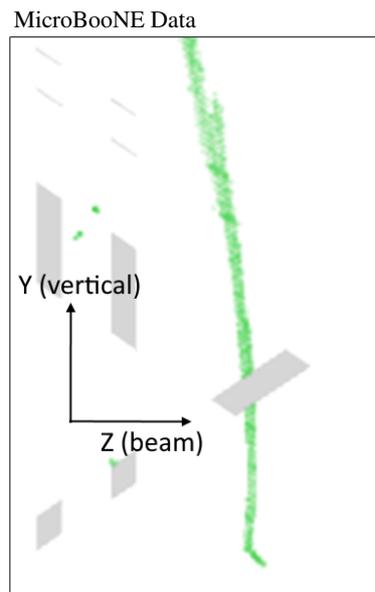


Figure 10: Zoomed in Y-Z view of a cosmic muon with a gap because of the nonfunctional regions. The nonfunctional regions are shown in gray.

507 4.1.1 Clustering in the presence of gaps

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

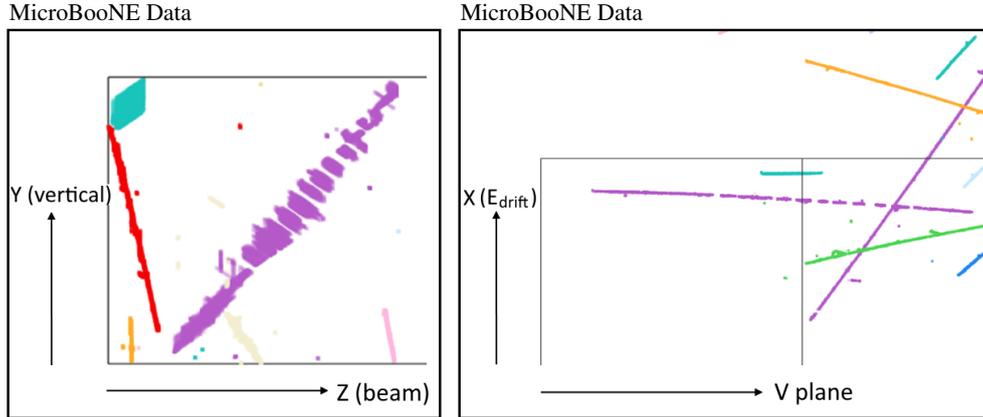


Figure 11: Gaps along an isochronous track in different zoomed-in 2D views. (Left) Y-Z view of an isochronous track (magenta) from a MicroBooNE data event. (Right) X-V view of the same event. V plane direction represents the wire pitch direction of the V wire plane. The black lines in the inner figures 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.

511 cloud (a collection of many 3D points) can be directly used, as a TPC cluster is a collection of
 512 the reconstructed 3D space points. The distance between two clusters (point clouds) is defined as
 513 the minimal distance between a pair of space points, one from each respective point cloud. To
 514 calculate this distance rapidly, the k-d (k-dimensional) tree based algorithm as implemented in the
 515 “naoflann” package [24] is employed. Once the minimal distance and its direction are obtained,
 516 its direction is compared with the directions of the two proto-clusters. These are found using a
 517 voting scheme inspired by the Hough transformation [25]. The directional vector, parameterized by
 518 a polar angle and an azimuthal angle, is calculated for each point. The most probable value in the
 519 2D distribution of polar-versus-azimuthal is then taken as the primary direction of the point cloud.
 520 Given the minimal distance vector and the two proto-cluster directions, the two proto-clusters are
 521 grouped (or not) based on their distance and consistency in directions. In practice, a set of criteria
 522 are developed and optimized by analyzing hundreds of data events from various topologies.

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

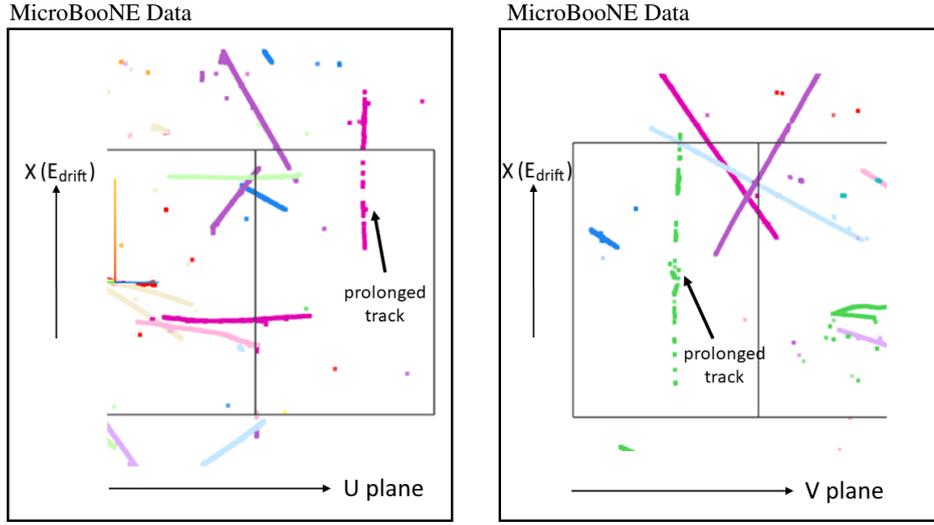


Figure 12: Gaps along prolonged tracks in different zoomed-in 2D views. (Left) X-U view for a prolonged cosmic muon track (magenta) from a MicroBooNE data event. (Right) X-V view for a prolonged cosmic muon track (green) from another MicroBooNE data event. The black lines in the inner figures correspond 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

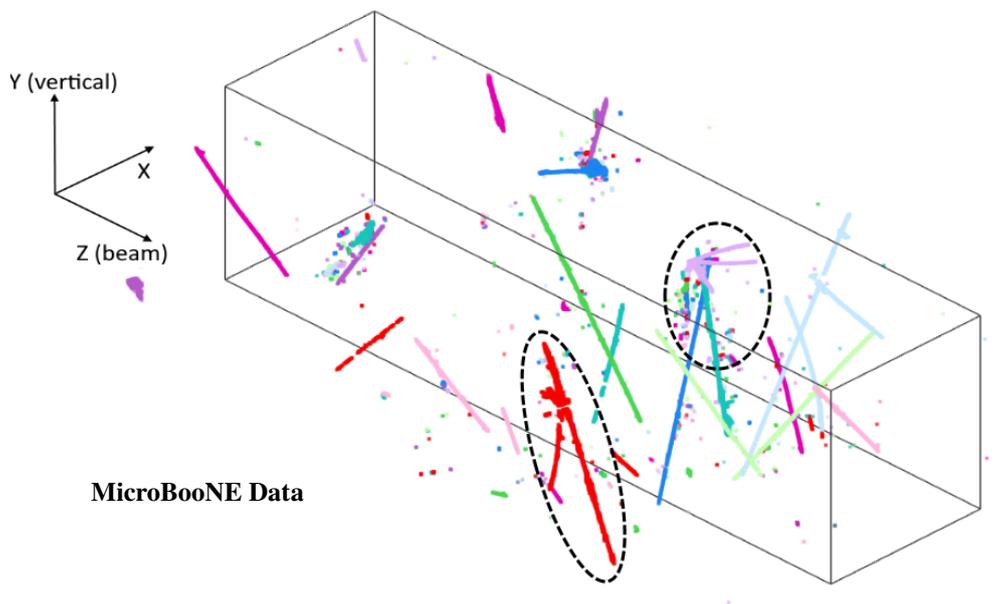
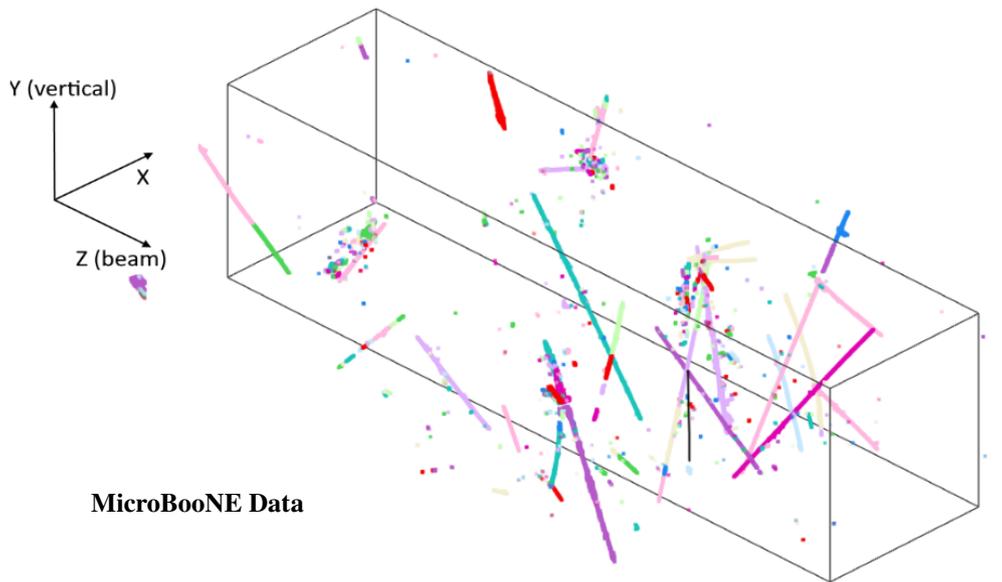


Figure 13: Demonstration of the effectiveness of the algorithm of bridging gaps. The solid black box represents the LArTPC active volume with an X-position (converted from the readout time) relative to the neutrino interaction time. Top: proto-clusters solely based on proximity. Bottom: clusters after 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.

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

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

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

570 4.1.3 Further de-ghosting

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

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

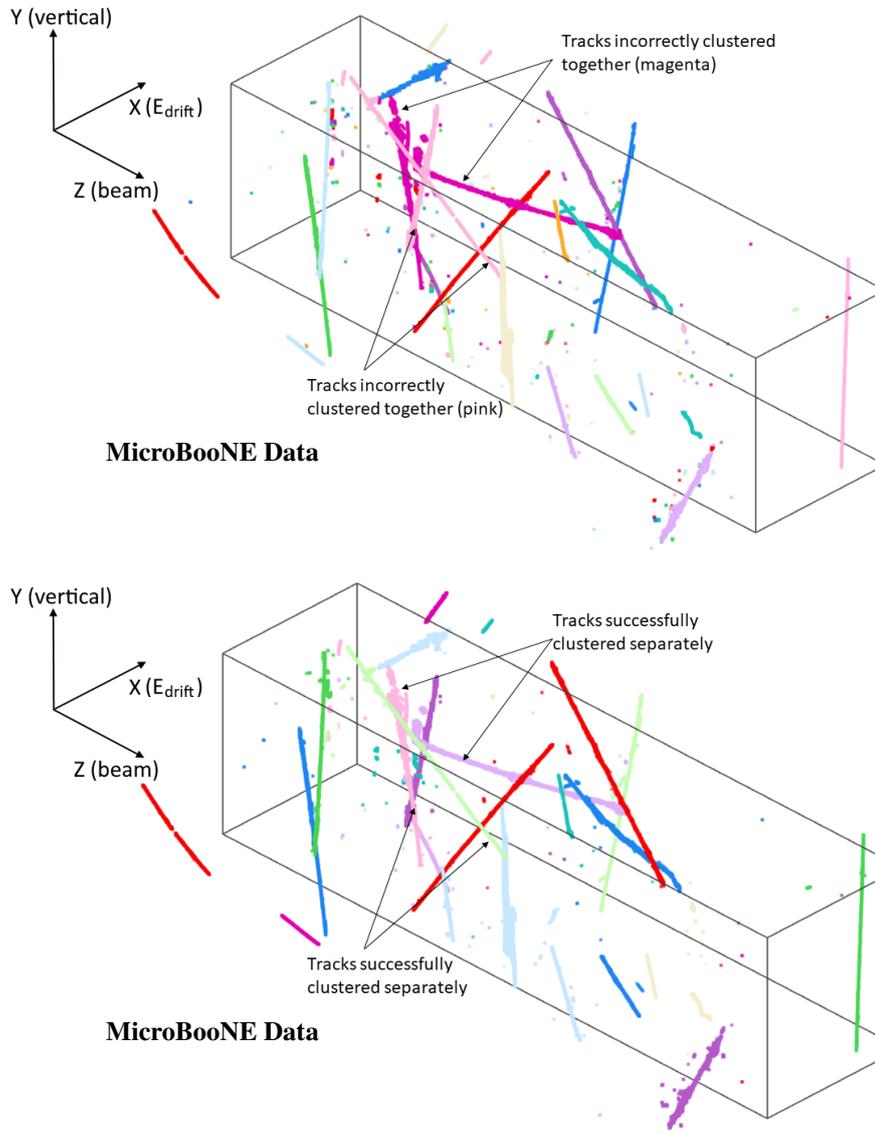


Figure 14: Demonstration of the effectiveness of the clustering algorithm to separate a “coincidental overlap” cluster. The solid black box represents the LArTPC active volume with an X-position (converted from the readout time) relative to the neutrino interaction time. The top and bottom panels show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

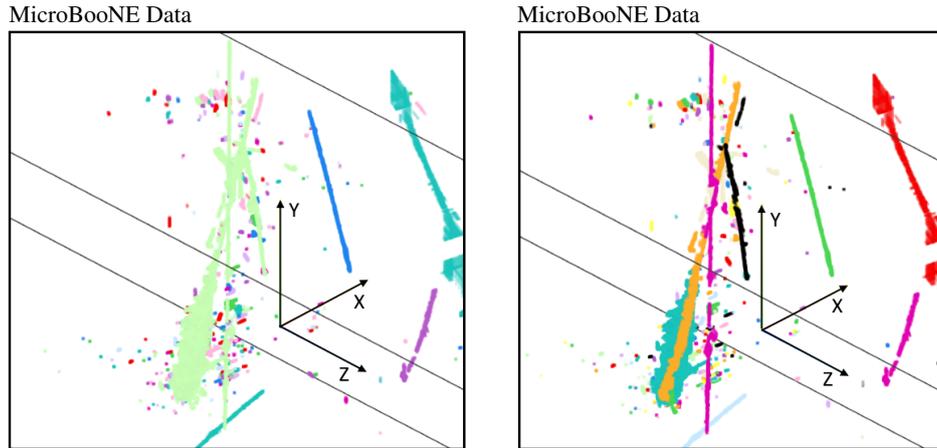


Figure 15: Demonstration of the effectiveness of the clustering algorithm to separate a “coincidental overlap” cluster. Black lines in each subfigure correspond to the boundaries of the LArTPC active volume. The left and right panels show the clusters before and after applying this algorithm. The all-light-green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

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

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

593 4.1.4 Clustering for neutrino events

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

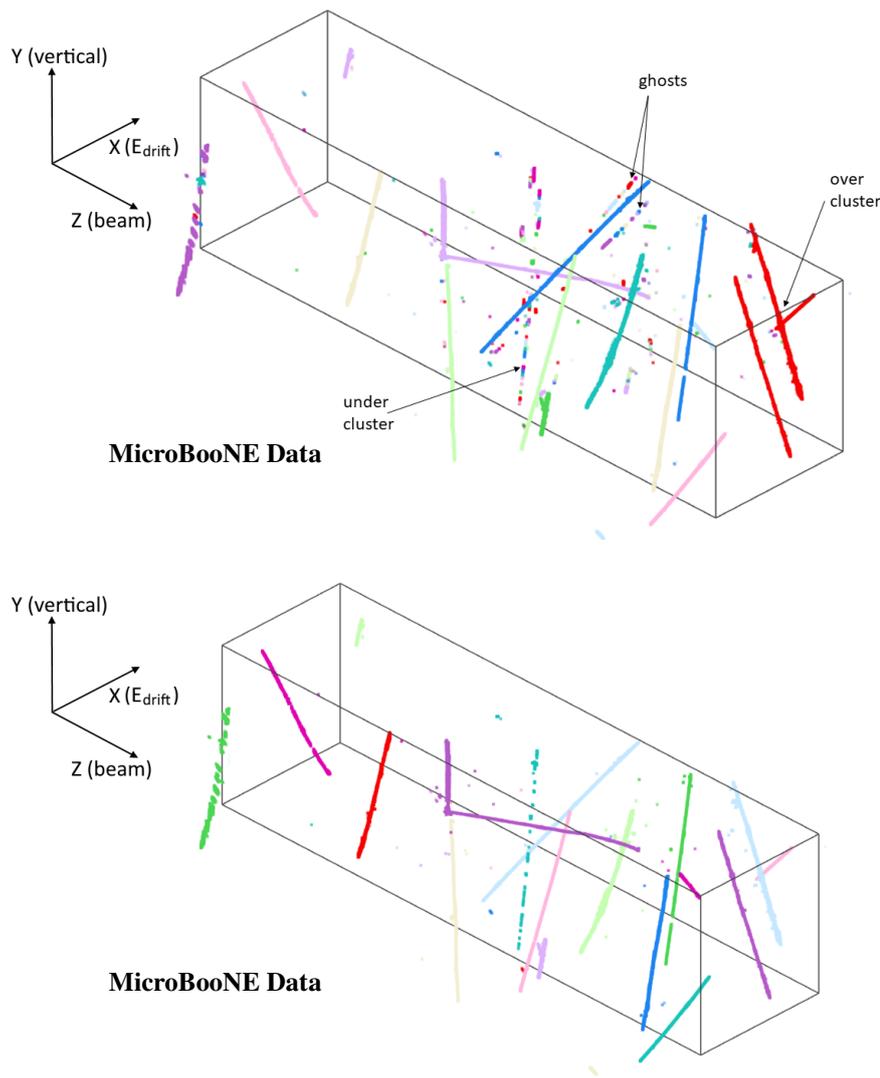


Figure 16: Demonstration of the effectiveness of the de-ghosting algorithm with other advanced clustering algorithms applied. The solid black box represents the LArTPC active volume with an X-position (converted from the readout time) relative to the neutrino interaction time. The top and bottom panels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

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

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

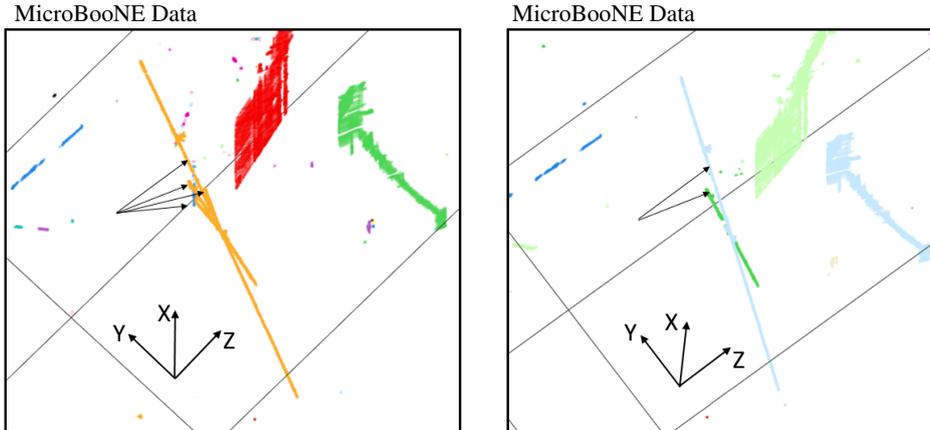


Figure 17: Demonstration of effectiveness of the de-ghosting algorithm with other advanced clustering algorithms applied. The black lines inside each subfigure correspond to the boundaries of the LArTPC active volume. The left and right panels show the clusters before and after applying the de-ghosting 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.

607 point.

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

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

617 4.2 PMT light signal reconstruction

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

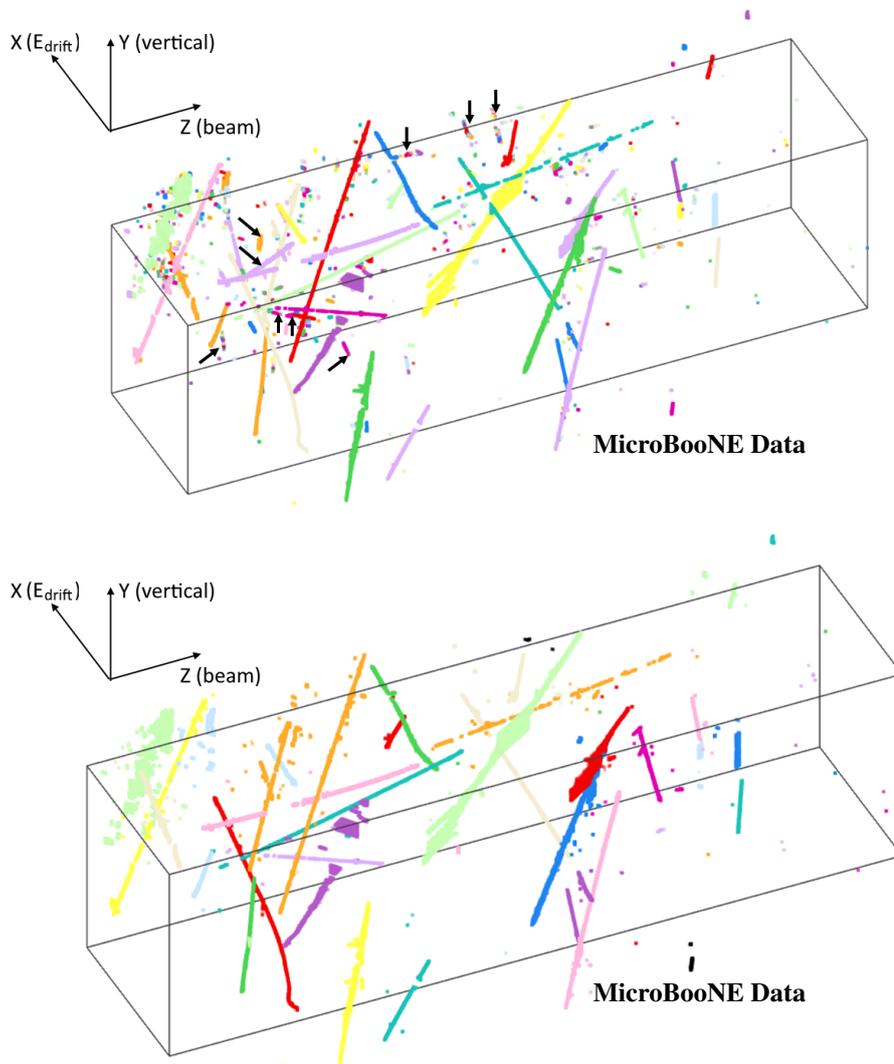


Figure 18: Demonstration of the effectiveness of the de-ghosting algorithm with other advanced clustering algorithms applied. The solid black box represents the LArTPC active volume with an X-position (converted from the readout time) relative to the neutrino interaction time. The dot-like clusters and 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.

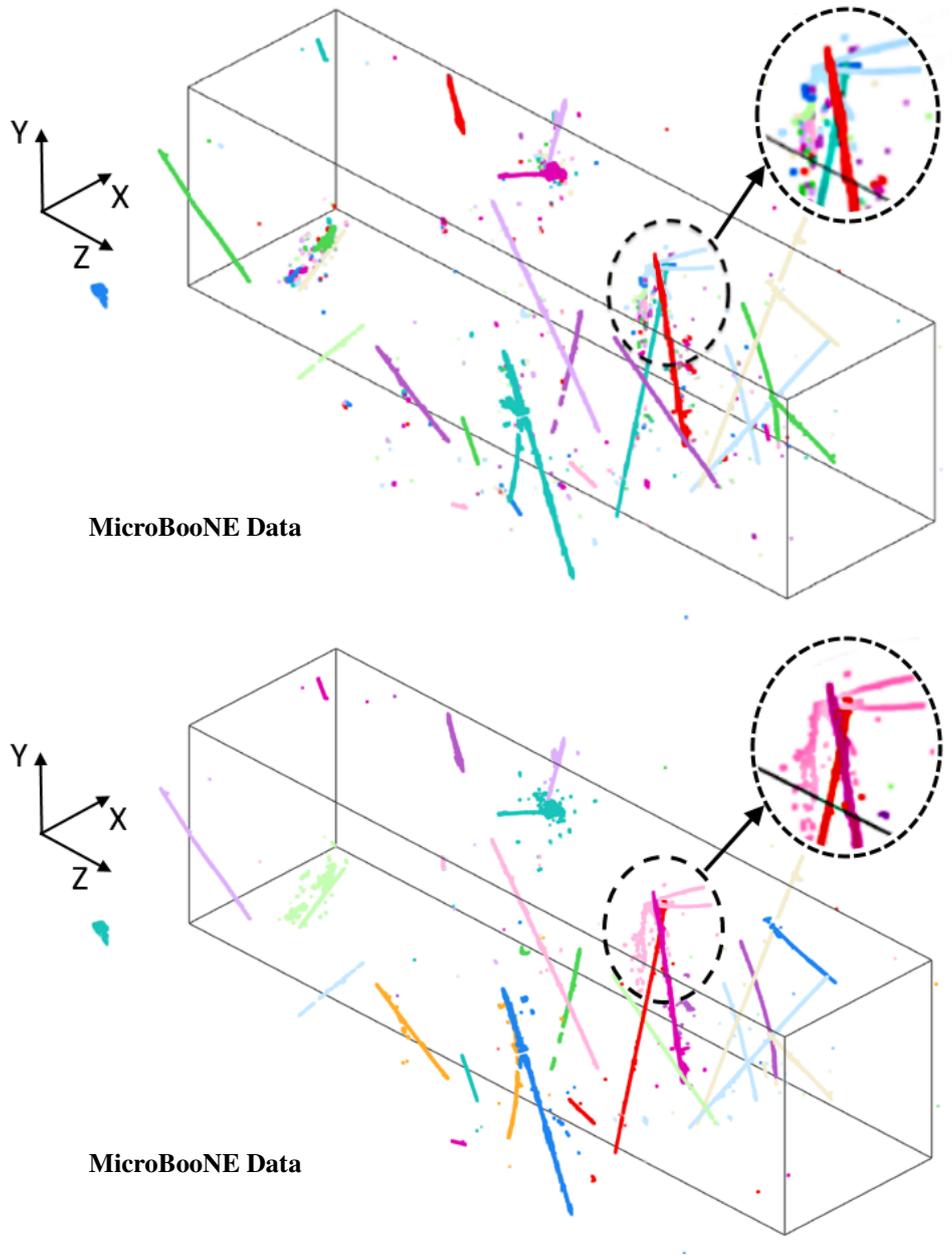


Figure 19: Demonstration of the effectiveness of the clustering algorithm designed for neutrino interactions. The solid black box represents the LArTPC active volume with an X-position (converted from the readout time) relative to the neutrino interaction time. Top and bottom panels show the clusters 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 γ 's from a π^0 decay).

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

635 The PMT waveforms are processed offline to reconstruct the time and number of photoelectrons
636 (PE) of a flash, which is a group of PMT signals close in time. For the beam discriminator, a
637 deconvolution using the Fast Fourier Transformation (FFT) is performed to unfold the electronics
638 responses from various RC circuits in the splitter and the shaper. A flash is then formed if the
639 PMT measurements satisfy the multiplicity requirement (>2 PMTs above a threshold of 1.5 PE)
640 and the total integrated PE threshold (>6 total PE) in a 100 ns window. A flash window lasts
641 7.3 μs in order to exclude noise and to include the contribution from the late scintillation light. The
642 scintillation light in liquid argon has a prompt and a slow component with decay times of about a
643 few nanoseconds and 1.6 μs ⁴, respectively.

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

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

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

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

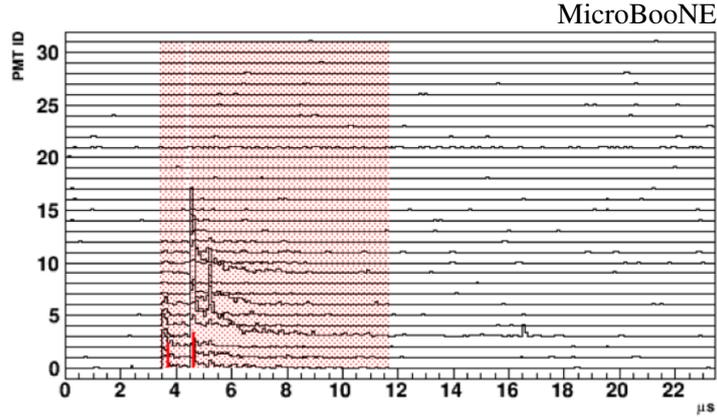


Figure 20: Illustration of two reconstructed flashes from beam discriminator PMT waveforms. The black curves are the deconvolved PE spectra for each PMT. The red lines represent the flash times and the red bands represent the flash windows. For the second flash at about $4.6\mu\text{s}$, there is a Michel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

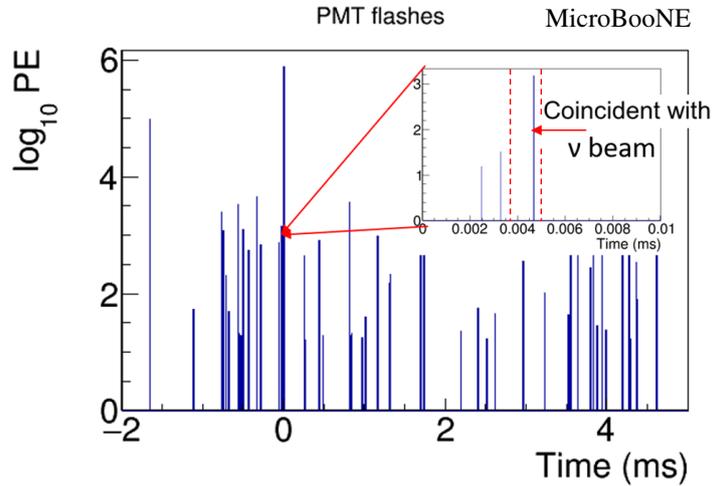


Figure 21: The reconstructed PEs of a flash as a function of flash time. The 6.4 ms PMT readout window is shown relative to the trigger time. The flashes from the beam discriminator ($23.6\mu\text{s}$ long) are shown as inset. The flash in coincidence with the BNB beam spill (between dashed red lines)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

667 4.3 Many-to-many charge-light matching

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

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

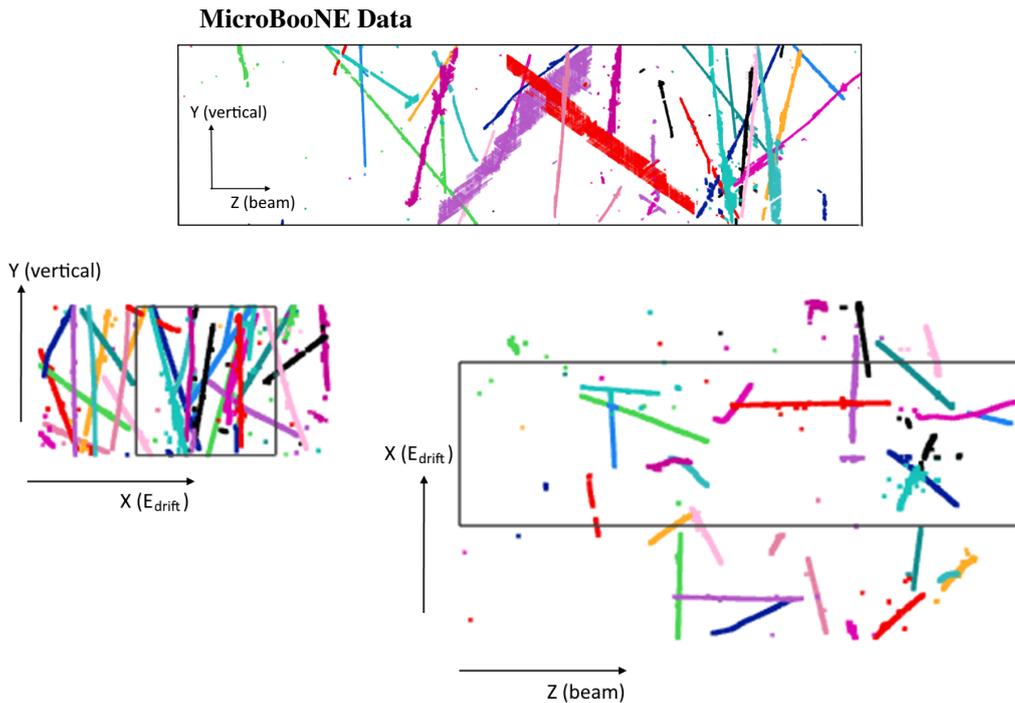


Figure 22: An example of all the TPC clusters from a MicroBooNE data event before charge-light matching. Different clusters are labeled in different colors, but each cluster is labeled in the same color in different 2D views. The borders represented by the black lines are the boundaries of the LArTPC 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.

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

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

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

691 the 32 PMTs can be made. The electron drift start time of the TPC cluster is shifted from the default
692 BNB beam time to the measured time of the PMT flash. This enables a correction of the X-position
693 of the TPC cluster. Then, the charge associated with each space point in the TPC cluster is used to
694 predict the PMT light signals based on a photon library [29]. The TPC volume ($2.56 \text{ m} \times 2.32 \text{ m} \times$
695 10.36 m) is divided into $75 \times 75 \times 400$ voxels. Millions of optical photons of 128 nm wavelength
696 from scintillation are generated and emitted with a 4π angular distribution in each voxel, and the
697 propagation of these photons is simulated with realistic optical photon processes of absorption and
698 scattering in Geant4. The PMT acceptance of optical photons emitted at different locations in the
699 TPC volume is calculated and recorded in the photon library. With this photon library, the PEs from
700 each of the PMTs for a given TPC cluster can be predicted by applying the PMT acceptance to the
701 charge of each space point. An overall scaling factor is applied to take into account the calibrated
702 scintillation light yield per unit deposited energy.

703 Interestingly, such a many-to-many matching problem is very similar to the charge solving
704 problem as introduced in section 3.2. There are more unknowns than knowns in this system, and the
705 imaging equation of the first principle as shown in eq. (3.1) can be used to relate the predicted light
706 signals from all possible TPC clusters to the measured signals from PMT flashes. Hypothetical
707 pairs of TPC clusters and PMT flashes are created and tested, in order to find the most compatible
708 ones and eliminate the rest. Again, the compressed sensing technique is utilized to perform this
709 many-to-many matching by minimizing an ℓ_1 -regularized chi-square function. In practice, a set of
710 matching algorithms are developed to pre-select, fit the ℓ_1 -regularized chi-square, and re-examine
711 the hypothetical TPC-PMT pairs.

712 **Pre-selection:** A pre-selection of the hypothetical TPC-PMT pairs is important to reduce the number
713 of unknowns in the ℓ_1 -regularized chi-square fitting, allowing for a more robust minimization. Two
714 major tests, time range compatibility and PMT hit pattern compatibility, are performed to remove
715 the incompatible TPC-PMT pairs. For the time range compatibility, the TPC cluster is required to
716 be fully contained within the maximum drift window corresponding to the PMT flash time⁵. For
717 example, as shown in figure 22, X-positions (along the drift) of the space points in any TPC cluster
718 have an overall shift because of the unknown electron drift start time, but the in-beam activities must
719 be contained in the nominal detector volume (black box) which is relative to the beam time. For
720 the PMT hit pattern compatibility, the pairs with highly incompatible predicted and measured light
721 signals are ruled out. A Kolmogorov-Smirnov test (K-S test) and a chi-square test, which inspect
722 the hit pattern and the absolute normalization of the 32 PMTs' signals, respectively, are combined to
723 discriminate the incompatible pairs. Specifically, to enable a many-to-one TPC-PMT matching, the
724 TPC clusters paired to the same PMT flash are jointly tested to maintain the many-to-one potential.
725 The most compatible TPC-PMT pair is used as a basis and the other ones are added individually to
726 check the change in compatibility. The pairs which significantly reduce the compatibility are ruled
727 out.

728 **Chi-square fitting:** Given the passing candidate TPC-PMT pairs after the pre-selection, a chi-
729 square function incorporating a ℓ_1 -regularization term is constructed to compare the predicted and

⁵A 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 δM_{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. χ_{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 χ_{p2}^2 term gives
 742 the constraint that b_i is preferred to be close to 0. The χ_{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. λ
 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 χ_{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

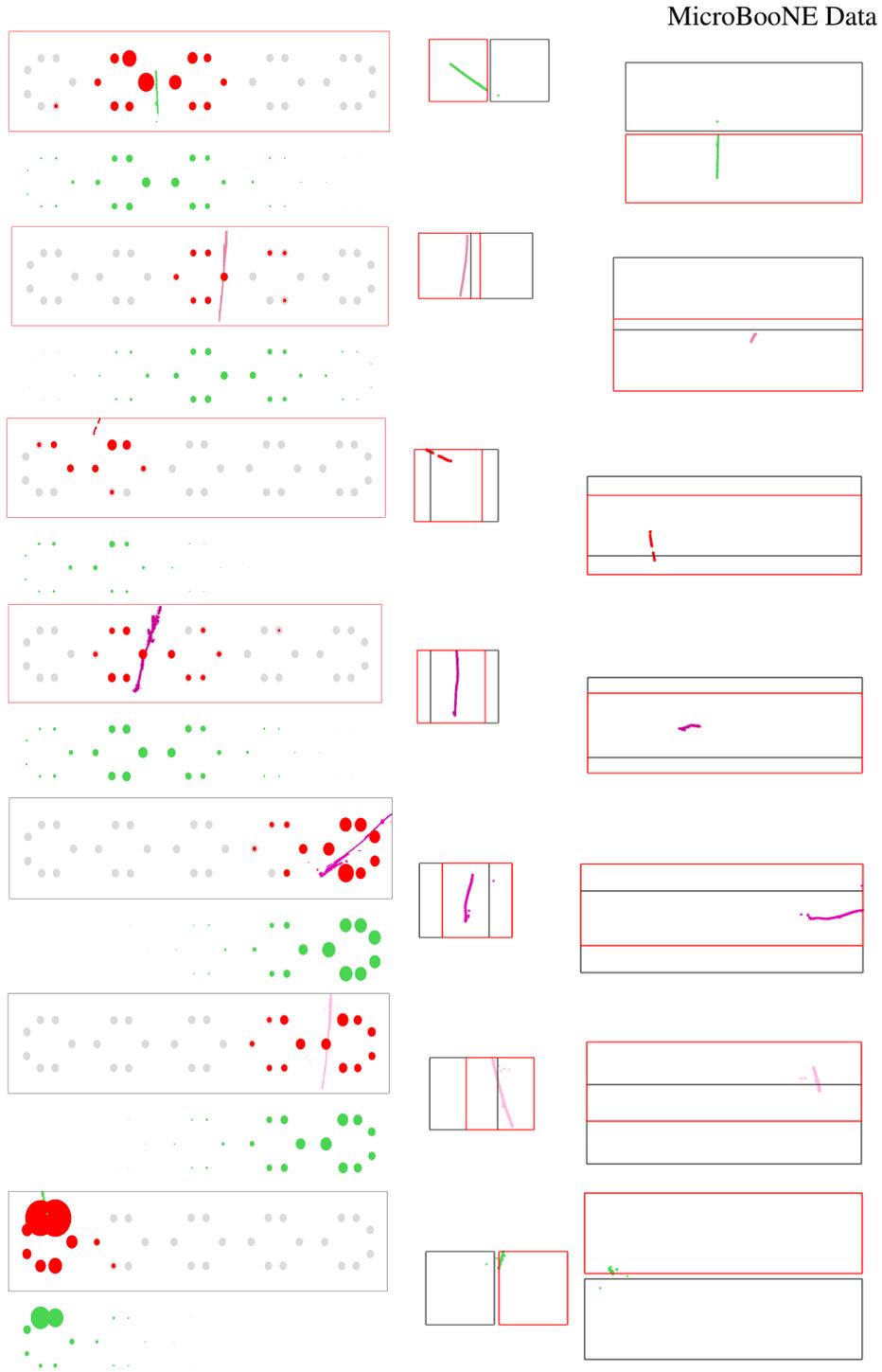


Figure 23: Selected 7 matched pairs out of the 31 pairs from a data event. From left to right, they are the front (Y-Z), side (Y-X), and top (X-Z) views of the detector, respectively. The black or red boxes correspond to the LArTPC active volume. The gray solid circles in the front view represent PMTs 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.

761 clusters are taken to be neutrino interaction candidates, and the remainders are rejected as cosmic-
 762 ray background. Figure 24 and figure 25 demonstrate successfully matching muon and electron
 763 neutrino clusters to their respective in-beam flashes. The performance of the matching algorithm is
 764 evident from these event displays and quantitative evaluations are provided in section 5. On average,
 765 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.

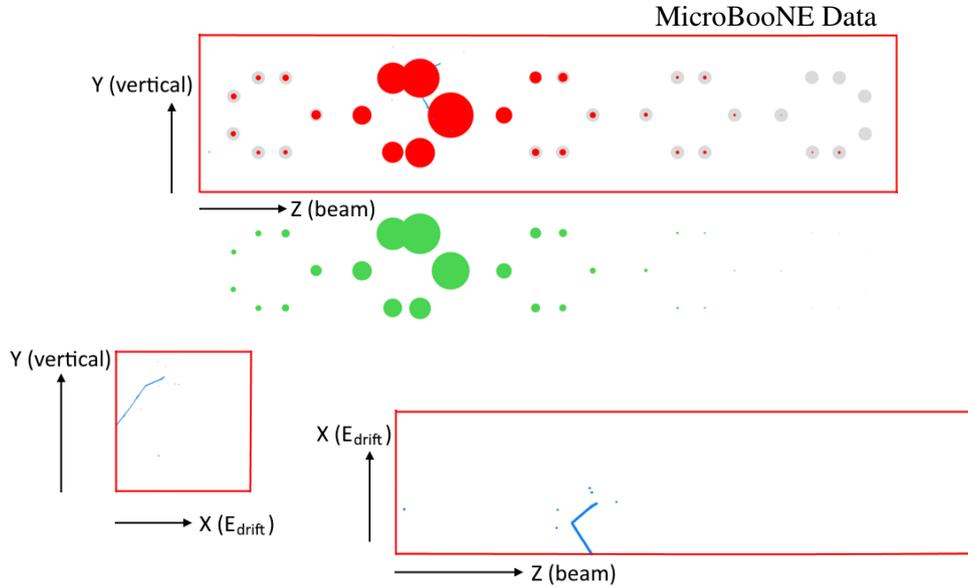


Figure 24: A muon neutrino event is shown with its matched flash. The red boxes correspond to the LArTPC active volume. The gray solid circles in Y-Z view represent the PMTs in different locations. The red solid circles represent the measured PE in the PMTs. The green solid circles represent the predicted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

766

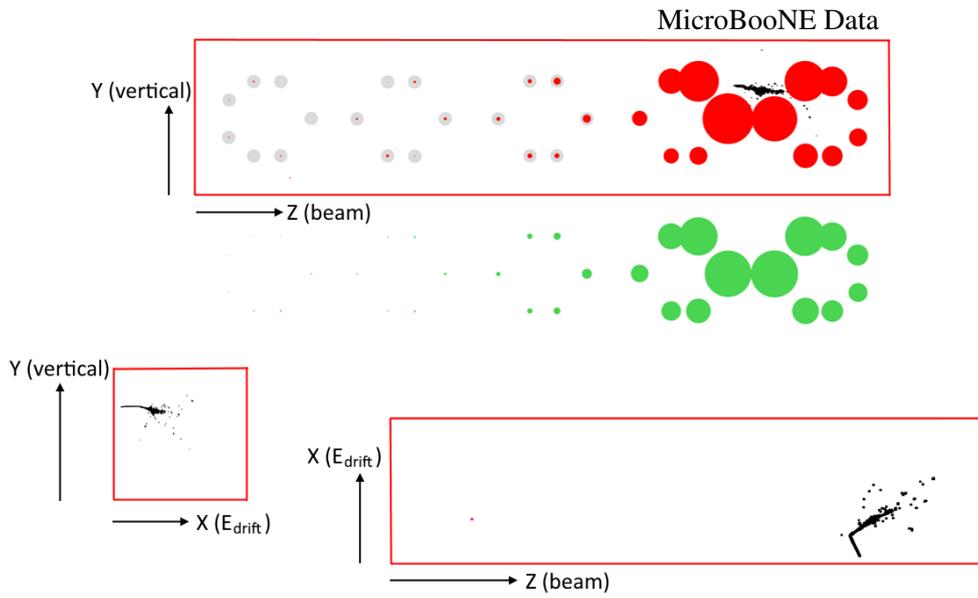


Figure 25: An electron neutrino event is shown with its matched flash. The red boxes correspond to the LArTPC active volume. The gray solid circles in Y-Z view represent the PMTs in different locations. The red solid circles represent the measured PE in the PMTs. The green solid circles represent the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

767 5 Evaluation of the Wire-Cell 3D imaging and the charge-light matching

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

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

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

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

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

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

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

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

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

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

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

828 5.1 Imaging performance of ideal tracks

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

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

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

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

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

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

847 **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

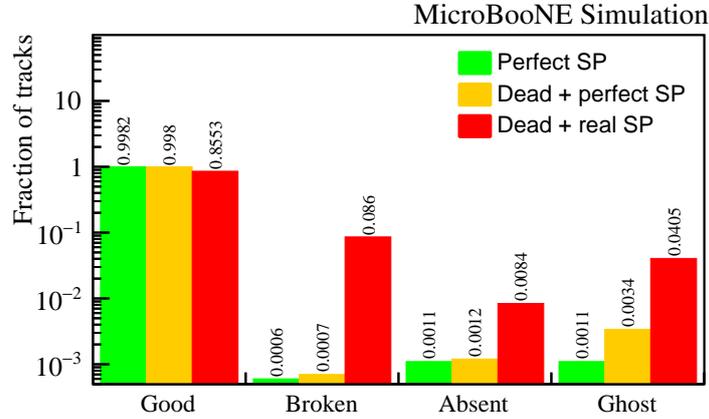


Figure 26: The fraction of (good, broken, absent, ghost) reconstructed tracks from the Wire-Cell 3D imaging for different scenarios. For good, broken, and ghost tracks, the fraction is weighted by their lengths and normalized to the total length of true tracks. See text for definitions of each category.

852

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

871

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

872 functional region as shown in figure 27. In one wire plane, the realistic signal processing, which
 873 may fail to extract the charges, could introduce a gap in the 3D image no matter the signal processing
 874 in the other two wire planes is successful or not. The measured charges originating from the TPC
 875 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.

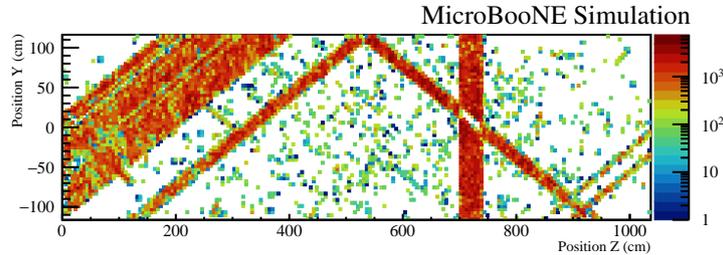


Figure 27: The position (Y/vertical versus Z/beam) distribution of the ghost tracks in the scenario of “Dead + real SP”. Color scale (Z-axis value) represents the count of space points in ghost tracks. The bands correspond to the nonfunctional regions as shown in figure 4.

876
 877 The purity for each event is calculated by dividing the total length of the non-ghost tracks by
 878 the total length of all the reconstructed tracks. The distribution of purity scores is presented in
 879 figure 28. For “Dead + real SP”, 96.4% of the events have at least 90% purity. Figure 29 shows
 880 the distribution of the completeness for all simulated tracks. For the scenario of “Dead + real SP”,
 881 86.5% (and 93.0%) of the simulated tracks have at least 99% (and 80%) completeness. The low
 882 completeness values correspond to the prolonged tracks, especially those with directions close to
 883 normal to the wire planes. This emphasizes again that good signal processing is important to retain
 the good quality of the 3D imaging result.

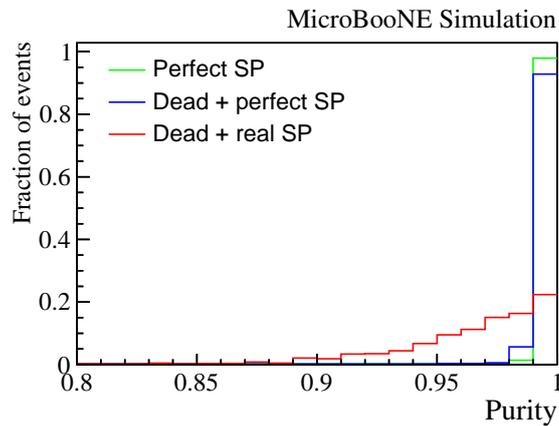


Figure 28: Distribution of the purity of the reconstructed tracks from each event for different scenarios. The number of ghosts significantly increases with the presence of both nonfunctional wires and the real imperfect signal processing. The histograms are normalized separately for each category.

884

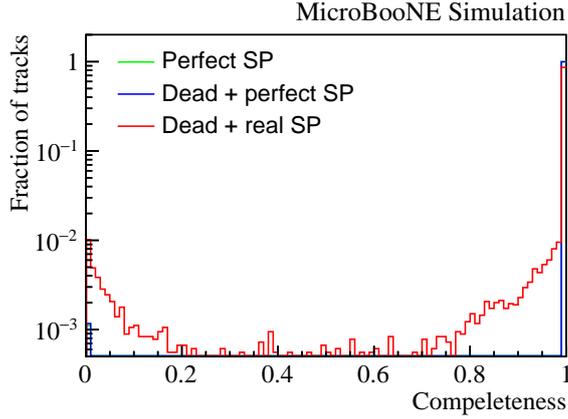


Figure 29: Distribution of the completeness of each simulated track for different scenarios. The distributions are normalized for each category, respectively. The results of “Perfect SP” and “Dead + perfect SP” are basically the same in which case the green line is covered by the blue line. The inefficiency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

885 5.2 Imaging performance of neutrino interactions

886 Unlike the simulated ideal tracks in section 5.1, the topology of a neutrino interaction’s final state
 887 particles could be much more complicated than the single-track-like cosmic-ray muons. Neutrino-
 888 only samples without cosmic-ray muons are used in this case. In order to evaluate the performance
 889 of the 3D imaging, the clustering as well as the charge-light matching steps are bypassed and all the
 890 3D space points reconstructed in the 3D imaging are taken as neutrino activities. This is equivalent
 891 to performing perfect clustering and charge-light matching.

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

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

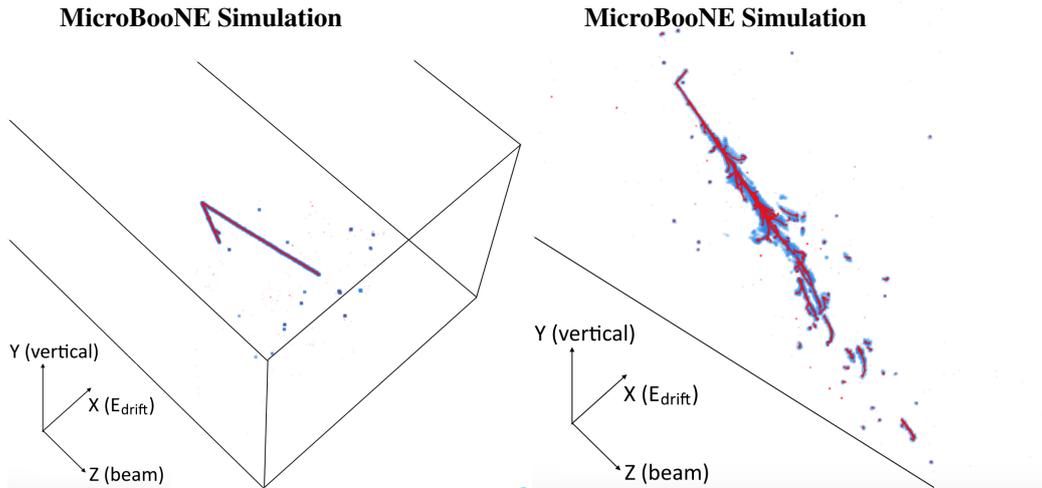


Figure 30: Left: $1\mu 1p \nu_\mu$ CC interaction. Right: $1e 1p \nu_e$ CC interaction. Blue: reconstructed 3D image. Red: truth trajectories. The voxel size and opacity are tuned for better illustration.

910 space points in the 3D imaging. The image of the reconstructed points are blurred because of the
 911 charge diffusion during the drift and the software filter smearing in the signal processing. Generally
 912 speaking, the reconstructed 3D image has both good completeness and purity compared to the
 913 truth 3D image in these two examples; even the short tracks belonging to the EM shower and
 914 isolated energy depositions are reconstructed. The quantitative evaluations of the purity and the
 915 completeness for BNB ν_μ CC, ν_e CC, and NC interactions in the TPC are shown in figure 31. The
 916 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 ν_μ CC	BNB ν_e CC	BNB NC
Purity >90% and			
Completeness > 80%	88.6%	89.2%	80.7%
Completeness > 70%	93.3%	96.7%	87.0%

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

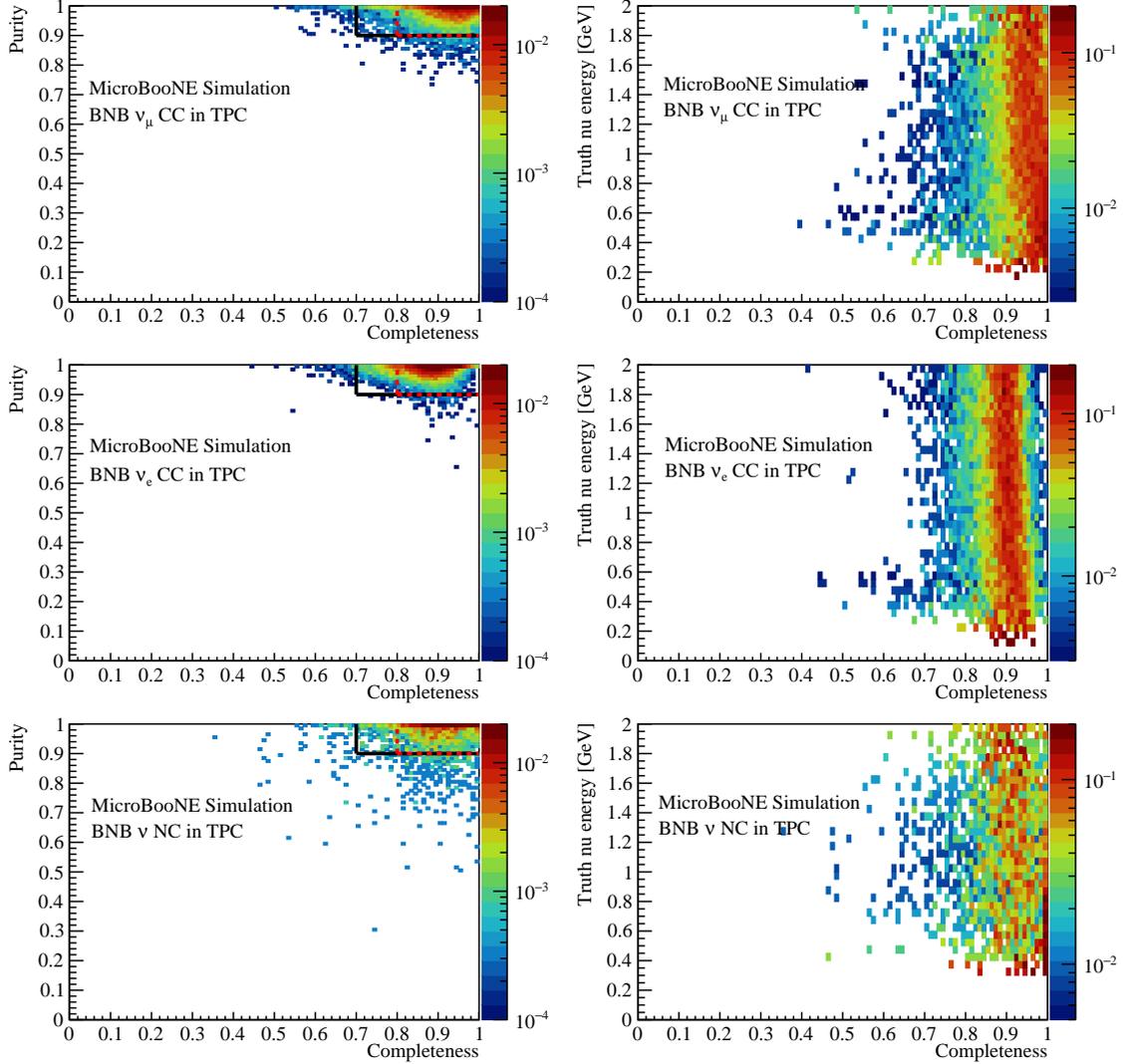


Figure 31: Two-dimensional distributions of the completeness and the purity of the 3D image for BNB ν_μ CC, ν_e CC, and NC interactions in the TPC. There are no cosmic-ray muons in this simulation. Left: purity vs. completeness for each neutrino interaction. The color scale (Z-axis value) represents 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.

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

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

938 5.3 Final performance in realistic cases

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

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

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

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

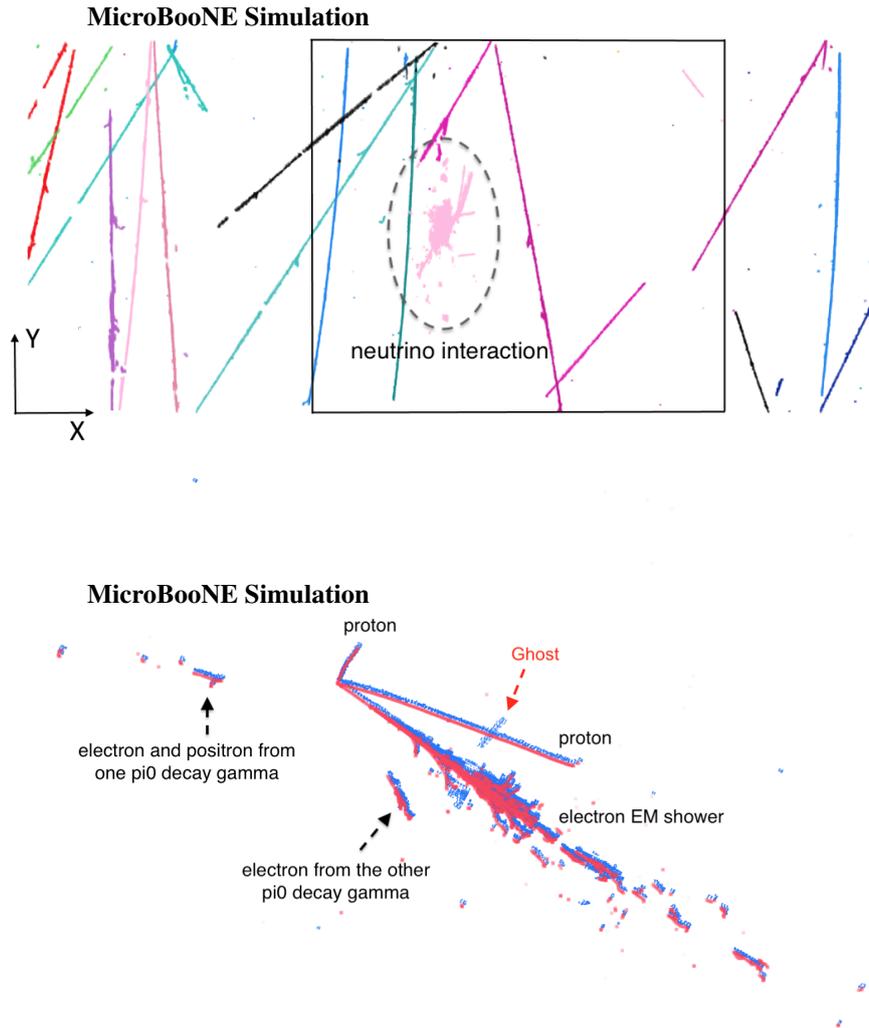


Figure 32: Event display of a $1e2p1\pi^0 \nu_e$ CC interaction. Top: side view of the full TPC readout; each cluster is labeled in one color. The black box corresponds to the LArTPC active volume with an X-position (converted from the readout time) relative to the neutrino interaction time. Bottom: 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 ν_μ CC, ν_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.

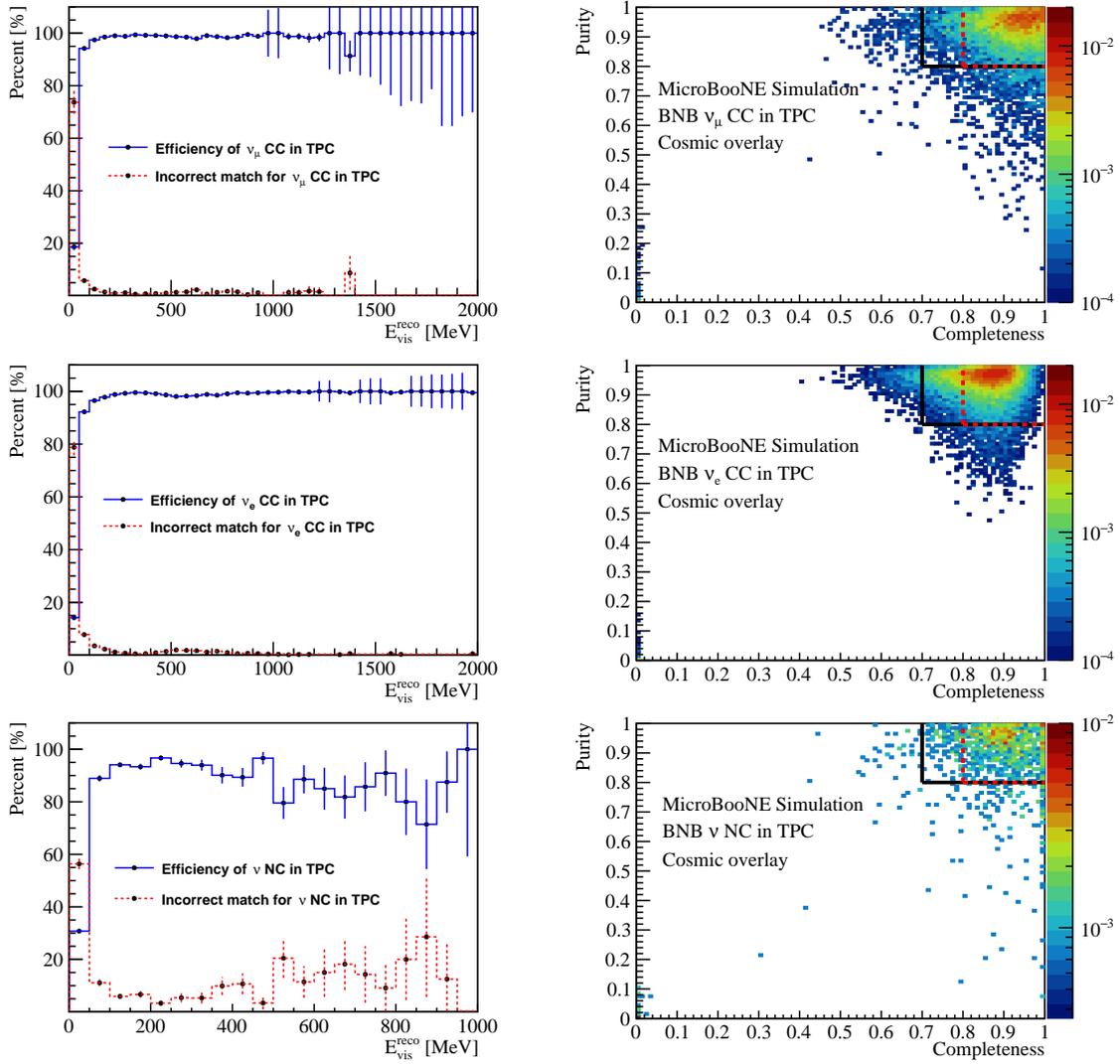


Figure 33: 3D imaging and charge-light matching performance for BNB ν_{μ} CC, ν_e CC, and NC interactions in the TPC. The neutrino interactions are simulated and overlaid with real data from cosmic rays. The clustering and charge-light matching steps are applied to select the neutrino interaction. Left: 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.

973 Right panel of figure 33 presents the completeness and purity of the selected TPC activities
 974 for BNB ν_μ CC, ν_e CC, and NC interactions, respectively. The results of completeness and purity
 975 are summarized in table 4. The events with extremely low completeness and purity values as
 976 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 ν_μ CC	BNB ν_e CC	BNB NC
Purity >80% and Completeness > 80%	73.0%	67.7%	56.0%
Completeness > 70%	80.2%	83.4%	66.5%

977 of the completeness and the purity can be attributed to the numerous cosmic-ray muons that traverse
 978 the detector. Direct comparisons of the completeness and the purity are independently performed,
 979 and the distributions can be seen in figure 34. The scenarios of “neutrino-only” and “neutrino +
 980 cosmic” correspond to figure 31 and figure 33, respectively. In the scenario of “neutrino + cosmic”,
 981 the neutrino activities suffer not only an over-clustering issue with the cosmic-ray activities (or its
 982 related ghost tracks) but also an under-clustering issue since part of the detached activities from
 983 the neutrino primary cluster may be grouped to cosmic-ray muons. These two issues introduce a
 984 smearing of both the completeness and purity distributions. The typical values of the completeness
 985 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 ν_μ CC	BNB ν_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%

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

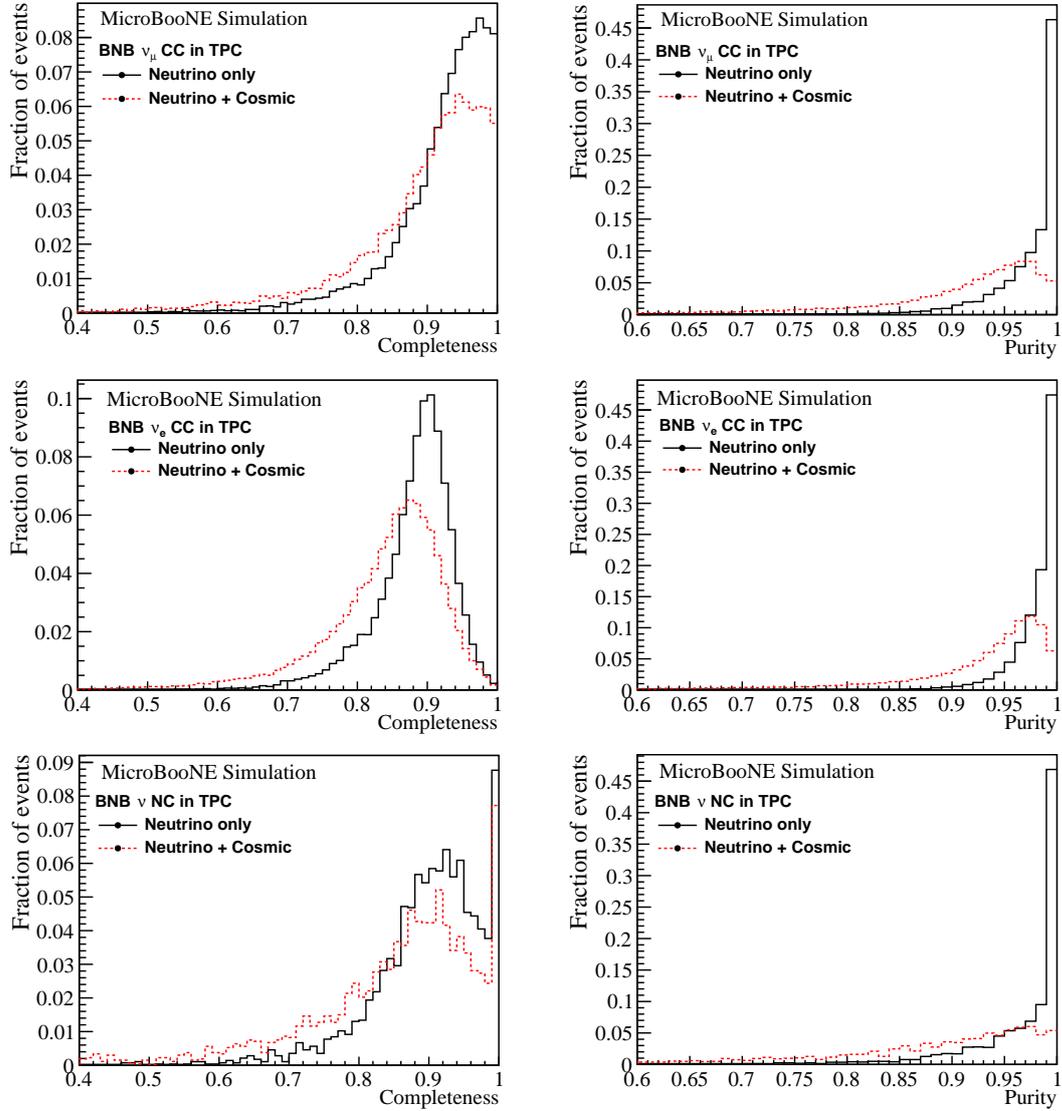


Figure 34: Independent comparisons of completeness and purity distributions for the two scenarios of “neutrino-only” and “neutrino + cosmic”. The 3D clustering and charge-light matching steps are applied in the scenario of “neutrino + cosmic” to select the neutrino interaction. Top: BNB ν_μ CC interactions in the TPC. Middle: BNB ν_e CC interactions in the TPC. Bottom: BNB ν NC interactions in the TPC; the “~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.

1003 6 Summary and Discussion

1004 This article describes the principle and algorithms of the Wire-Cell 3D imaging, clustering, and
1005 many-to-many charge-light matching applied in the MicroBooNE LArTPC. The 3D imaging tomo-
1006 graphically reconstructs the 3D image of the ionization electrons using the fundamental information
1007 of charge, time, and geometry of the LArTPC detector. Other characteristics of the LArTPC physics
1008 activities such as sparsity, positivity, and proximity are utilized as additional constraints to improve
1009 the 3D imaging performance. The many-to-many charge-light matching with 3D clustering and light
1010 signal reconstruction is developed to pair the TPC clusters and PMT flashes to identify the neutrino
1011 interaction among numerous cosmic-ray muons. Several realistic issues, e.g. the nonfunctional
1012 wires, the gaps due to inefficient signal processing, detached neutrino activities, and coinciden-
1013 tally connected clusters, are properly addressed. Using the MicroBooNE detector simulation, the
1014 realistic performance of the reconstruction techniques is evaluated.

1015 In spite of the effort, there are some limitations in the 3D imaging as shown in the event displays
1016 in this paper. For example, prolonged tracks, which often develop gaps in the signal processing
1017 stage, cannot be entirely fixed via the bridging of the gaps as implemented. Similar issues occur for
1018 the isochronous tracks that often develop gaps because of the insufficient coherent noise filtering.
1019 Subsequent pattern recognition techniques, e.g. particle-level clustering and trajectory fitting, may
1020 further address this problem. Isochronous tracks present another common problem for the LArTPC
1021 3D imaging, as the wire readout ambiguity is drastically increased in the time slice containing them.
1022 In Wire-Cell 3D imaging, this issue is mitigated by introducing tiling, however, the blobs of the
1023 isochronous track are significantly broadened, leading to a much worse spatial resolution in the Y-Z
1024 plane view. This issue can be further mitigated via trajectory fitting in a later stage.

1025 Generally speaking, the 3D event reconstruction techniques as presented in this paper are
1026 adequately accurate and efficient, and can successfully select neutrino interaction activities. About
1027 95% of the neutrino CC interactions in the TPC active volume are selected, with a 30-fold reduction
1028 of non-beam-coincident cosmic-ray muons. Good completeness and purity of the resulting 3D
1029 image of the selected neutrino activities have been achieved. Greater than 80% of the selected
1030 neutrino CC interactions have a reconstructed 3D image of at least 70% completeness and 80%
1031 purity. These techniques will benefit the downstream pattern recognition and neutrino selection,
1032 and they are important steps towards realizing the full capability of single-phase LArTPCs. In
1033 particular, the Wire-Cell based neutrino selection and analyses take full advantage of these tools
1034 to further reject cosmic muons and select neutrinos [15] and demonstrate a very promising high
1035 efficiency and high purity neutrino selection in LArTPCs. Other analyses using techniques such
1036 as deep learning with convolutional neural networks [9, 10] and Pandora multi-algorithm pattern
1037 recognition [8] can also benefit from the outcome of the Wire-Cell 3D reconstruction tools, as it
1038 provides a clean and intact 3D image of the in-beam neutrino activities with the surround cosmic
1039 ray muons removed.

1040 Acknowledgments

1041 This document is prepared by the MicroBooNE collaboration using the resources of the Fermi
1042 National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP
1043 User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under Contract
1044 No. DE-AC02-07CH11359. MicroBooNE is supported by the following: the U.S. Department of
1045 Energy, Office of Science, Offices of High Energy Physics and Nuclear Physics; the U.S. National
1046 Science Foundation; the Swiss National Science Foundation; the Science and Technology Facilities
1047 Council (STFC), part of the United Kingdom Research and Innovation; and The Royal Society
1048 (United Kingdom). Additional support for the laser calibration system and cosmic ray tagger is
1049 provided by the Albert Einstein Center for Fundamental Physics, Bern, Switzerland.

1050 References

- 1051 [1] C. Rubbia. *The liquid argon time projection chamber: A new concept for neutrino detector*, .
1052 CERN-EP/77-08, 1977.
- 1053 [2] H. H. Chen, P. E. Condon, B. C. Barish and F. J. Sciulli. *A Neutrino detector sensitive to rare*
1054 *processes. I. A Study of neutrino electron reactions*, . FERMILAB-PROPOSAL-0496, 1976.
- 1055 [3] W. J. Willis and V. Radeka, *Liquid argon ionization chambers as total absorption detector*, *Nucl.*
1056 *Instrum. Meth.* **120** (1974) 221.
- 1057 [4] D. R. Nygren, *The Time Projection Chamber: A New 4 pi Detector for Charged Particles*, *eConf*
1058 **C740805** (1974) 58.
- 1059 [5] MICROBooNE collaboration, R. Acciarri et al., *Design and Construction of the MicroBooNE*
1060 *Detector*, *JINST* **12** (2017) P02017, [[arXiv: 1612.05824](#)].
- 1061 [6] P. A. Machado, O. Palamara and D. W. Schmitz, *The Short-Baseline Neutrino Program at Fermilab*,
1062 *Ann. Rev. Nucl. Part. Sci.* **69** (2019) , [[arXiv:1903.04608](#)].
- 1063 [7] DUNE collaboration, B. Abi et al., *Volume I. Introduction to DUNE*, *JINST* **15** (2020) T08008,
1064 [[arXiv:2002.02967](#)].
- 1065 [8] MICROBooNE collaboration, R. Acciari et al., *The Pandora multi-algorithm approach to automated*
1066 *pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector*, *Eur. Phys.*
1067 *J. C* **78** (2018) 82, [[arXiv:1708.03135](#)].
- 1068 [9] MICROBooNE collaboration, R. Acciarri et al., *Convolutional neural networks applied to neutrino*
1069 *events in a liquid argon time projection chamber*, *JINST* **12** (2017) P03011, [[arXiv:1611.05531](#)].
- 1070 [10] MICROBooNE collaboration, C. Adams et al., *Deep neural network for pixel-level electromagnetic*
1071 *particle identification in the MicroBooNE liquid argon time projection chamber*, *Phys. Rev.* **D99**
1072 (2019) 092001, [[arXiv:1808.07269](#)].
- 1073 [11] X. Qian, C. Zhang, B. Viren and M. Diwan, *Three-dimensional Imaging for Large LArTPCs*, *JINST*
1074 **13** (2018) P05032, [[arXiv:1803.04850](#)].
- 1075 [12] MICROBooNE collaboration, R. Acciarri et al., *Noise Characterization and Filtering in the*
1076 *MicroBooNE Liquid Argon TPC*, *JINST* **12** (2017) P08003, [[arXiv:1705.07341](#)].
- 1077 [13] MICROBooNE collaboration, C. Adams et al., *Ionization electron signal processing in single phase*
1078 *LArTPCs. Part I. Algorithm Description and quantitative evaluation with MicroBooNE simulation*,
1079 *JINST* **13** (2018) P07006, [[arXiv:1802.08709](#)].

- 1080 [14] MICROBooNE collaboration, C. Adams et al., *Ionization electron signal processing in single phase*
1081 *LArTPCs. Part II. Data/simulation comparison and performance in MicroBooNE*, *JINST* **13** (2018)
1082 [P07007](#), [[arXiv:1804.02583](#)].
- 1083 [15] MICROBooNE collaboration, P. Abratenko et al., *Cosmic Ray Background Rejection with Wire-Cell*
1084 *LArTPC Event Reconstruction in the MicroBooNE Detector*, [arXiv:2101.05076](#).
- 1085 [16] MINIBooNE collaboration, A. Aguilar-Arevalo et al., *The Neutrino Flux prediction at MiniBooNE*,
1086 *Phys. Rev. D* **79** (2009) 072002, [[arXiv:0806.1449](#)].
- 1087 [17] MICROBooNE collaboration, R. Acciarri et al., *Construction and Assembly of the Wire Planes for the*
1088 *MicroBooNE Time Projection Chamber*, *JINST* **12** (2017) T03003, [[arXiv: 1609.06169](#)].
- 1089 [18] MICROBooNE collaboration, J. Conrad, B. J. P. Jones, Z. Moss, T. Strauss and M. Touns, *The*
1090 *Photomultiplier Tube Calibration System of the MicroBooNE Experiment*, *JINST* **10** (2015) T06001,
1091 [[arXiv:1502.04159](#)].
- 1092 [19] E. J. Candès, J. K. Romberg and T. Tao, *Stable signal recovery from incomplete and inaccurate*
1093 *measurements*, *Communications on Pure and Applied Mathematics* **59** (8) (2006) 1207–1223,
1094 [[arXiv:math/0503066](#)].
- 1095 [20] J. Friedman, T. Hastie and R. Tibshirani, *Regularization paths for generalized linear models via*
1096 *coordinate descent*, *Journal of Statistical Software* **33** (1) (2010) 1–22.
- 1097 [21] Wire-Cell git repository for sparse signal regression,
1098 <https://github.com/BNLIF/wire-cell-ress/>.
- 1099 [22] H. Yu et al., *Augmented Signal Processing in Liquid Argon Time Projection Chambers with a Deep*
1100 *Neural Network*, [arXiv:2007.12743](#).
- 1101 [23] MICROBooNE collaboration, D. Caratelli, *Neutrino identification with scintillation light in*
1102 *MicroBooNE*, *JINST* **15** (2020) C03023.
- 1103 [24] J. L. Blanco and P. K. Rai, “nanoflann: a C++ header-only fork of FLANN, a library for nearest
1104 neighbor (NN) with kd-trees.” <https://github.com/jlblancoc/nanoflann>, 2014.
- 1105 [25] P. Hough, “Method and means for recognizing complex patterns.” US Patent 3069654, 1962.
- 1106 [26] quickhull. <https://github.com/akuukka/quickhull>, 2017.
- 1107 [27] E. W. Dijkstra, *A note on two problems in connexion with graphs*, *Numerische mathematik* **1** (1959)
1108 269–271.
- 1109 [28] F. J. Massey Jr., *The kolmogorov-smirnov test for goodness of fit*, *Journal of the American Statistical*
1110 *Association* **46** (1951) 68–78.
- 1111 [29] Ben Jones, *Sterile Neutrinos in Cold Climates*, FERMILAB-THESIS-2015-17, Massachusetts
1112 Institute of Technology, 2015.
- 1113 [30] MICROBooNE collaboration, P. Abratenko et al., *Measurement of Space Charge Effects in the*
1114 *MicroBooNE LArTPC Using Cosmic Muons*, *JINST* **15** (2020) P12037, [[arXiv:2008.09765](#)].
- 1115 [31] LArSoft release v08.05.00.10,
1116 <https://cdcv.s.fnal.gov/redmine/projects/larsoft/wiki/ReleaseNotes08050010>.
- 1117 [32] Uboonecode release v08.00.00.28,
1118 <https://cdcv.s.fnal.gov/redmine/projects/uboonecode/wiki/ReleaseNotes08000028>.
- 1119 [33] C. Andreopoulos, C. Barry, S. Dytman, H. Gallagher, T. Golan, R. Hatcher et al., *The GENIE*
1120 *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](#).