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

The MicroBooNE Collaboration
MICROBOONE-NOTE-1083-PUB

E-mail: MICROBOONE_INFO@fnal.gov

Abstract: An accurate and efficient event reconstruction is an imperative element in realizing the full scientific capability of liquid argon time projection chambers (LArTPCs). The current and future neutrino experiments that rely on massive LArTPCs create a need for new ideas and reconstruction approaches. Wire-Cell, proposed in recent years, is a novel tomographic event reconstruction method for LArTPCs. The Wire-Cell 3D imaging approach capitalizes on the charge and sparsity information in addition to the time and geometry information to reconstruct an unambiguous 3D image of the ionization electrons prior to pattern recognition. A second novel method, the many-to-many charge-light matching, then pairs the TPC charge activity to the detected scintillation light signal, thus enabling a powerful rejection of cosmic-ray muons in the MicroBooNE detector. A robust processing of the scintillation light signal and an appropriate clustering of the 3D space points are fundamental to this technique. In this paper, we describe the principles and algorithms of these techniques and their successful application in the MicroBooNE experiment, while addressing several realistic issues. A quantitative evaluation of the performance of these techniques is presented as well. Using these techniques, a 95% efficient pre-selection of the neutrino charged-current events is achieved with a 30-fold reduction of the non-beam-coincident cosmic-ray muons, and about 80% of the selected neutrino charged-current events are well reconstructed with high completeness (at least 70%) and purity (at least 80%).

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

Contents

1 Introduction

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

Event reconstruction is one of the most challenging tasks in analyzing the data from current and future massive single-phase LArTPCs. The performance of event reconstruction is vital to taking full advantage of LArTPC capability for physics measurements. *Wire-Cell*, a novel event reconstruction method [8], has been fully implemented for MicroBooNE. The Wire-Cell 3D imaging capitalizes on the most fundamental LArTPC detector information – time, charge, and geometry – to tomographically reconstruct a topology-agnostic three-dimensional image of the ionization electrons prior to any pattern recognition step. Enabled by the high-performance ionization electron signal processing procedure in MicroBooNE [9–11], the Wire-Cell 3D imaging reduces the degeneracies – where the charge arrives along the wire – inherent in the single-phase LArTPC wire readouts as used by MicroBooNE and numerous other experiments.

Detector defects such as nonfunctional channels (10% of all channels in MicroBooNE) and the numerous cosmic-ray muons (20-30 per TPC readout window) in the MicroBooNE detector pose additional challenges to the overall success of the event reconstruction. We address the first problem by allowing cases of only two functional channels in a region, and for these regions performing a highly refined analysis that relies on information from nearby, fully functional regions. Our method significantly reduces the extent of unusable regions by a factor of ten.

To deal with the high rate of cosmic rays, we developed a second novel method, the many-to-many TPC-charge/PMT-light matching, to distinguish candidate neutrino activity (energy deposited by ionization), which is in coincidence with the beam spill, from the numerous cosmic-rays spanning the entire MicroBooNE detector and TPC readout window. This method relies on the Wire-Cell 3D images and emphasizes the interplay between the scintillation light created by charged particles traversing the LAr, and the ionization charge signals to select the candidate neutrino activity. A robust processing of the scintillation light signals from the photo-detector system and an appropriate clustering in the 3D image of the ionization charge are fundamental to this technique.

In this paper, we describe the principles, algorithms, and evaluation of the Wire-Cell 3D imaging and many-to-many charge-light matching, including the light signal processing and the 3D clustering. The principle and implementation of the Wire-Cell 3D imaging is presented in section 3. The many-to-many charge-light matching to pair the TPC *clusters* to the reconstructed PMT *flashes* is described in section 4 as the final step to select the candidate neutrino activity. Evaluations of the quality of the Wire-Cell 3D imaging and the efficacy of the many-to-many charge-light matching are demonstrated in section 5. A summary of the performance and discussion can be seen in section 6.

2 The MicroBooNE detector

The MicroBooNE detector is the first LArTPC in the SBN Program to measure the neutrino interactions from the on-axis Booster [12] neutrino beam (BNB) at Fermi National Accelerator Laboratory in Batavia, IL. MicroBooNE uses a single-phase LArTPC with a rectangular active volume of the following dimensions: 2.6 m (width, along the drift direction), 2.3 m (height, vertical), and 10.4 m (length, along the beam direction), similar to that illustrated in figure 1. The TPC has an active mass of 85 metric tonnes and is immersed in a single-walled and cylindrical shape cryostat with a 170 tonne liquid argon capacity.

A high voltage of -70 kV is applied on the cathode plane providing a drift field of 273 V/cm. The electrons ionized by any energy deposition from traversing charged particles drift towards the anode wire planes along the electric field at a nominal speed of about 1.10 mm/μs. In this paper, we use X-axis to represent the drift direction, Y-axis to represent the vertical direction, and Z-axis to represent the beam direction, forming a left-handed coordinate system. There are three parallel wire readout planes [13] on the anode side with different wire orientations. The first wire plane facing the cathode is labeled "U", and the second and third plane are labeled "V" and "Y", respectively.

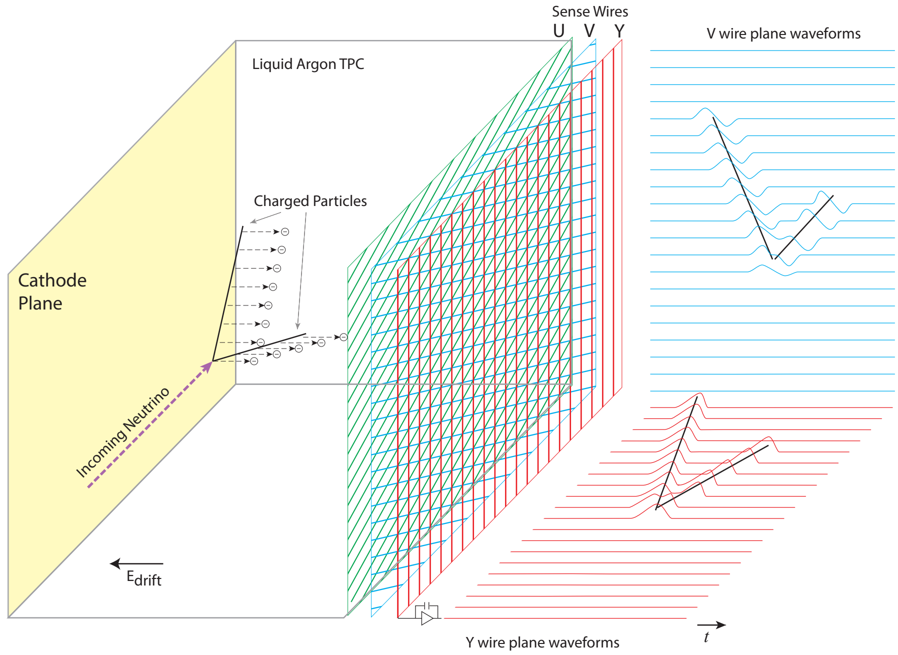

Figure 1: Illustration of single-phase LArTPCs [5]. Each wire plane provides a 2D image of the ionization electrons with respect to a specific wire orientation.

The 3456 wires in the Y plane are oriented vertically and the 2400 wires in the U (V) plane are oriented $+(-)60^\circ$ with respect to the vertical direction. The spacing between adjacent wires and adjacent wire planes are both 3 mm. Different bias voltages, -110 V, 0 V, and 230 V, are applied to the U, V, and Y wire planes, respectively, to ensure all ionization electrons drift through the U and V planes before being collected by the Y plane. The U and V planes are commonly referred to as the induction planes because the ionization electrons induce bipolar electrical signals; the Y plane is referred to as the collection plane and sees unipolar electrical pulses.

The TPC readout is defined with respect to the event trigger and includes three 1.6 ms frames, spanning -1.6 ms to +3.2 ms relative to the trigger (beam spill) time, with a sampling rate of 2 MHz (0.5 μs per tick). Each wire plane provides a 2D image (time versus wire) of the ionization electrons within the full 4.8 ms TPC readout.

Behind the wire planes and external to the TPC, there is an array of 32 8" photomultiplier tubes [14] to detect scintillation light for triggering, timing, and other purposes. The PMT readout includes four 1.6 ms frames with the beam gate window (1.6 μs beam-spill) contained within the second 1.6 ms frame. The sampling rate is 64 MHz (15.625 ns per sample) for each PMT and the signal exists in a dynamic-range-based, paired form for each channel – a high gain (x10) signal and a low gain (x1) signal. The 32 PMTs immediately detect the scintillation light and provide a 2D pattern (signal strength versus PMT position) of the photo-electrons corresponding to each individual TPC activity.

The TPC and PMT readouts cover the full time range (beam time + maximal drift time ~2.3 ms) of the beam neutrino activity as well as cosmic activity that enters the beam spill frame during the relatively "slow" drift of ionization electrons.

3 Wire-Cell 3D Imaging

The single-phase LArTPC with a wire readout scheme is a natural application of the *tomography* technique, which the Wire-Cell 3D imaging strictly follows. Ref. [8] introduces the basic concepts and the key mathematics of Wire-Cell 3D imaging. In this section, we focus more on the realistic issues of the application of the Wire-Cell 3D imaging in MicroBooNE.

The fundamental information provided by such a detector is as follows:

(i) **Time** - when the ionization electrons arrive at the anode wire plane.

(ii) **Geometry** - which wires from each wire plane are fired by the ionization electrons.

(iii) **Charge** - how many ionization electrons are measured by the fired wires for each wire plane.

The "Time" and "Charge" information comes from the time distribution of the deconvolved charge which is obtained via advanced signal processing techniques. In particular, the 2D deconvolution technique [10, 11] significantly improves the signal processing for the induction planes and makes the deconvolved charge consistent across the multiple wire planes. The "Geometry" is the 1D wire position along the wire pitch direction (perpendicular to the wire orientation) and lying in the plane of the wires. Since the wire planes have different wire orientations, each wire plane is taken as a 1D projection of the TPC objects with the summation of the charge available, where the individual charge deposition along the wire is ambiguous.

The Wire-Cell 3D imaging has two major steps to reconstruct the 3D image of the ionization electrons arriving at the anode plane: 1) Reconstruct the 2D image of the ionization electrons on the anode plane in a given time slice, e.g. 2 μs (4 ticks in the TPC readout) considering the intrinsic time smearing of about 1.5 μs after signal processing [10]. The integrated charge within the time slice on each fired wire is used; 2) Connect the 2D images from the previous step in the sequence of time slices. From three wire readout planes, at most three 1D projection views are available, in contrast with the dozens of 1D projection views available in common tomography applications. Relative to a pixelated readout with n^2 pixels, the $O(n)$ wires ($3\times n$ for three wire planes) afforded by a wire readout scheme result in a considerable loss of information, whereas it reduces the heat loads and the cost of the readout system. To recover from the loss of information, additional constraints are thus used:

(iv) **Sparsity** - the distribution of ionization electrons in space are generally expected to be sparse for any physics activity within a short time.

(v) **Proximity** - the ionization electrons are continuous in 3D space for charged particle energy depositions in the LArTPC active volume.

(vi) **Positivity** - the ionization electrons all drift towards (never away from) the wire planes and the number of the drifted ionization electrons can only be positive.

The actual procedure to incorporate the enumerated information is divided into two processes: *tiling* and *solving*, as described in section 3.1 and section 3.2, respectively. In the implementation of Wire-Cell 3D imaging in MicroBooNE, the 10% nonfunctional wires [9] aggravate the wire readout ambiguity, introducing a large number of *ghost* energy depositions. A dedicated deghosting algorithm, discussed in section 3.3, was developed to deal with these.

3.1 Tiling

The 2D image of the ionization electrons in a time slice consists of *cells*, which are the minimal geometric units formed by wires from three planes. Figure 2 shows tens of cells, which are formed by dashed black lines and labeled with blue dots in the center. The wires are represented by solid red lines and the wire (pitch) boundaries are represented by dashed black lines. All cells have equilateral triangular shapes because of the MicroBooNE wire orientations and positioning.

"Time" is used to determine the time slice, whose 2 μs width contains four sampling ticks of 0.5 μs from the TPC readout. The width of the time slice introduces negligible information loss because the software filtering/smearing in the signal processing has a cut-off frequency at about 0.25 MHz to optimize signal-to-noise ratio. "Geometry" is used to determine all possible hit cells within a time slice by finding the intersections of the fired wires. In figure 2, there are 8 fired U wires (2.4 cm wide), 5 fired V wires (1.5 cm wide), and 6 fired Y wires (1.8 cm wide), leading to 55 possible hit cells. The fact that there are fewer knowns (19 fired wires) than unknowns (55 cells) indicates an ambiguity is introduced by the wire readout. Meanwhile, amount of integrated charge in a time slice and the identities of active wires in that time slice are affected by diffusion and long-range induction effects (especially for induction planes) as charge drifts in the TPC, as well as the action of software filters applied to the waveforms [10, 11]. To mitigate the impact from wire ambiguity and charge smearing, a procedure to merge the consecutive hit cells is developed and it is called *tiling*. The group of hit cells after tiling is called *blobs*. The blob in figure 2 is marked by solid blue lines. The connected fired wires are merged as *wire bundles* in the tilling procedure, and a blob is the overlapping area of three wire bundles from each wire plane as shown in figure 2. Note that a cell or a blob is a 3D object and its length along the drift direction is 2 μs, about 2.2 mm.

There are three advantages of the tiling: 1) completely collect the reconstructed charge smeared to the adjacent wires resulting in more consistent charge values across the wire planes, 2) greatly reduce the number of unknowns to be solved in the later stage of *solving*, and 3) significantly limit the computing cost. The charge smearing is different for different wire planes. Obtaining consistent charge measurements across multiple wire planes by the tiling is fundamental to construct and solve the Wire-Cell 3D imaging equation as described in section 3.2.

Figure 2 corresponds to a single track traversing the time slice in a local area. In reality, there could be multiple tracks from cosmic-ray muons or a neutrino interaction traversing the time slice (a fixed X position) at various Y-Z locations as shown in figure 3. The solid red lines represent the fired wires from each wire plane. The resulting blobs are marked in blue or green. One may notice that in figure 3, the green blobs only have two corresponding wire bundles from two wire planes. This is because the fired wires in one wire plane may not be able to provide reasonable signals if they are dead or too noisy. Generally, a 3-plane tiling approach requires the wires from all three wire planes to be functional, and this leads to a nonfunctional region (volume) on the anode plane (in the TPC active volume) as illustrated by only blue blobs in figure 3.

About 10% of channels are nonfunctional in MicroBooNE for various reasons [9]. As a result, about 30% of the anode plane area has at least one nonfunctional wire, as shown in figure 4. To address this issue, we allow for a 2-plane tiling procedure in areas where at least two planes have

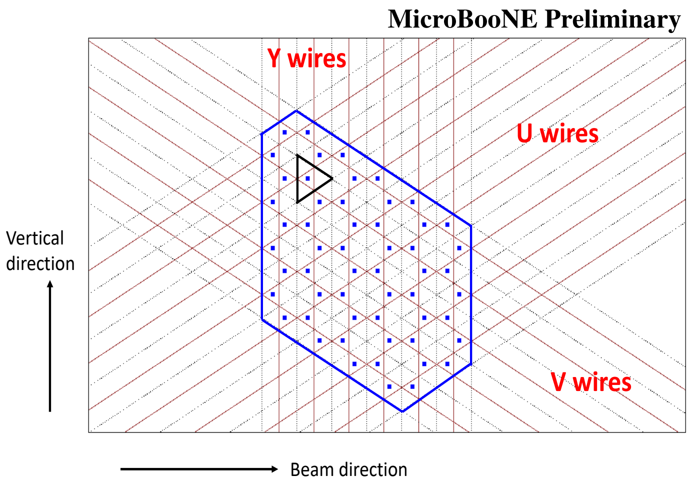

Figure 2: An example of the hit cells and blob constructed by the fired wires with the MicroBooNE detector geometry. Each wire is represented by a solid red line and the wire (pitch) boundaries represented by dashed black lines. All hit cells have equilateral triangular shapes and are marked with blue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

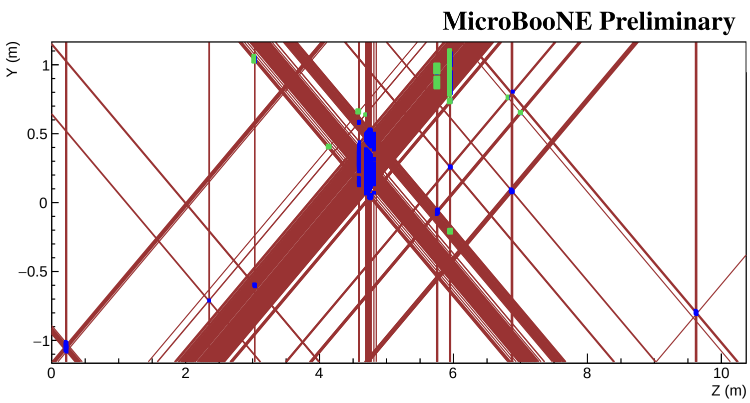

Figure 3: Example of the fired wires and blobs. Blobs are marked in blue or green. Blue blobs correspond to 3-plane tiling requiring all three wire planes to be functional. Green blobs correspond to the additional blobs created in 2-plane tiling requiring at least two wire planes to be functional. Fired wires are represented by solid red lines.

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

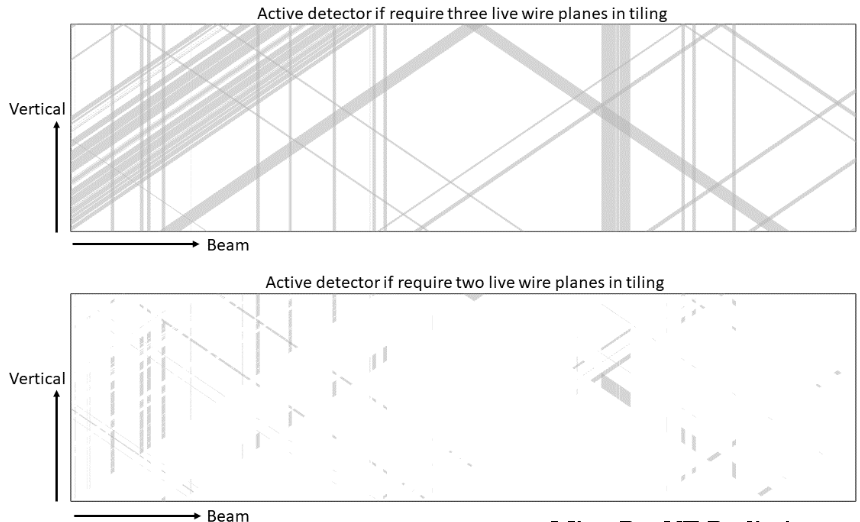

Figure 4: Impact of the nonfunctional wires (grey) on the anode plane. Top: the grey area that has at least one wire nonfunctional is 30%. Bottom: the grey area that has at least two wires nonfunctional is 3%.

While missing 3-plane blobs are recovered with 2-plane tiling, a number of fake blobs, or "ghosts", are created in areas where two functional fired wires cross a third nonfunctional wire assumed to be fired, when in fact no true physical charge was responsible for this pair's readout. Some ghosts would still appear when all three wire planes are functional, but the number of ghosts is significantly increased when tiling is allowed, given the sizable number of "fired" nonfunctional wires.

Using "Time" and "Geometry" information, tiling provides a 3D image of all the possible charge depositions as shown in figure 5. The top panel corresponds to the 3-plane tiling, giving 70% functional volume. The middle panel corresponds to the 2-plane tiling, giving 97% functional volume. Since it is essential to limit the nonfunctional volume in physics measurements, the next task is to remove the ghosts which originate from the wire readout ambiguity and nonfunctional wires in the 2-plane tiling procedure.

3.2 Charge solving

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

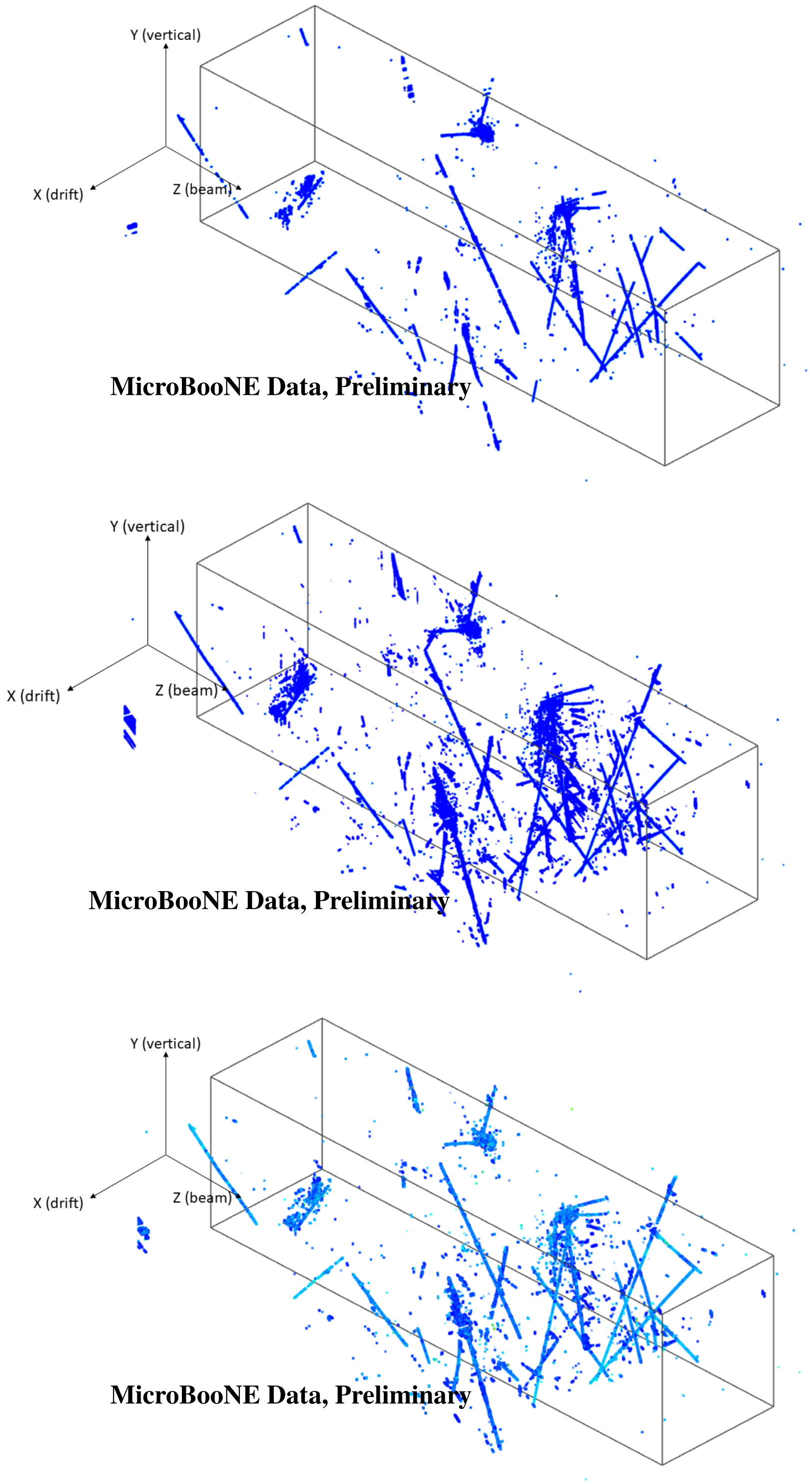

Figure 5: Comparison of the tiling results and the charge solving result from MicroBooNE data (event 41075, run 3493). Only time and geometry information are used in the tiling. Sparsity, positivity, and proximity information are incorporated in the charge solving as described in section 3.2. Top: 3-plane 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.

expressed as follows:

$$y = Ax, \tag{3.1}$$

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

As elaborated in ref. [8], one can find a relevant solution to eq. (3.1) by making it an optimization problem applying additional constraints, e.g.

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

where $||x||_p = (\sum_i |x_i|^p)^{1/p}$ is the ℓ_p-norm of a vector x. Since the physics activities in LArTPC are *sparse*, i.e. most of the true elements of x are zero, the ℓ_0-norm (a count of the nonzero elements) can be used to seek the most sparse or the simplest solution that explains the measurements. The minimization of $||x||_0$ can be achieved by removing the number of unknowns until the linear equation is solvable. For example in figure 3, there are 25 blobs, while only about 10 hits are true[1]. One can remove 15 unknowns out of 25 to solve the equation and find the "best" one satisfying the optimization condition. However, in this case there are $C_{25}^{10} \approx 3.3 \times 10^6$ combinatorial ways to remove the unknowns and the optimization is an NP-hard problem that is extremely expensive in computation. Mathematicians [15] have discovered that another constraint, the ℓ_1-norm, can well approximate the ℓ_0-norm result with a much faster minimization. This ℓ_1 technique, also known as *compressed sensing*, is widely applied in many other fields for signal processing and computational photography. As shown in section 4.3, the compressed sensing technique is also used to perform the many-to-many charge-light matching.

In practice, a chi-square function is constructed to take into account the uncertainties of the measured charge from signal processing [11], and the compressed sensing technique is implemented by adding an ℓ_1-regularization term to the chi-square function:

$$\chi^2 = ||y' - A'x||_2^2 + \lambda||x||_1, \tag{3.3}$$

where the vector y and x are pre-normalized through $V^{-1} = Q^T Q$ (Cholesky decomposition), $y' = Q \cdot y$, $A' = Q \cdot A$, and λ regulates the strength of $||x||_1$. The matrix V is the real symmetric covariance matrix of the charge measurement uncertainties. The ℓ_1-regularized chi-square function is convex with a unique global minimum, enabling fast minimization algorithms such as coordinate

[1]The actual number of unknowns to be removed is the number of zero eigenvalues of matrix $A^T A$.

descent [16]. An implementation of the coordinate descent method can be found in a Wire-Cell git repository [17]. Another constraint, *positivity* of the charge (number of the ionization electrons), can be easily added in the coordinate descent method to help remove the ghosts.

So far we have shown the incorporation of "Charge", "Sparsity", and "Positivity" to seek the unique solution to the equation of the first principle. To further improve the robustness of the ℓ_1-regularization result, *proximity* information is incorporated given the fact that the LArTPC is fully active in 3D space. For those adjacent blobs over different time slices, the regularization strength λ will be applied with an additional scaling factor of a^n to lower the chance of removing the corresponding element in x during the ℓ_1 minimization. n represents the number of the adjacent blobs that are connected to the target blob, and a is a predefined scaling factor. The final chi-square function in the Wire-Cell imaging is transformed to be:

$$\chi^2 = ||y' - A'x||_2^2 + \lambda||\omega \cdot x||_1, \tag{3.4}$$

where λ is an overall regularization strength parameter, and $\omega_i = a^{n_i}$ is the weight for x_i as described in the text. Note that y is a vector of the integrated measured charge for each wire bundle in the tiling and x is a vector of the charge to be solved for each blob. Data events were used to tune λ and a.

The bottom panel of figure 5 shows the result applying the charge solving to the 2-plane tiling result. Compared to the middle panel of figure 5, the ghosts are considerably removed and the 3D voxels (2D blob $\times$ 1D time slice) are associated with different charge values which correspond to the solution x of the equation of the first principle. As elaborated in section 4.3, such 3D charge solving is critical to predict the scintillation light signals for each PMT, allowing for comparison to and matching with the observed light flashes.

3.3 De-ghosting

The amount of ghosts is considerably reduced after the charge solving but the result is still unsatisfactory. The sparsity combined with the proximity is already incorporated in the charge solving to resolve the wire readout ambiguity, however, this procedure is performed in a "local" manner restricted within each time slice or over adjacent time slices. Within the 3D imaging, all connected blobs in 3D space are grouped together as *proto-clusters*. A proto-cluster does not necessarily correspond to a complete TPC activity since there might be true or artifical gaps in the 3D image. The principle of the sparsity of the LArTPC physics activities will be further used in a "global" manner to reconstruct the sparsest 3D images of the TPC activities by removing the less prominent (either shorter or thinner) redundant proto-clusters. Following this philosophy, a dedicated algorithm, *deghosting*, was developed to remove the residual ghosts based on their two main characteristics.

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

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

The area with one nonfunctional wire plane provides much less constraint[3] in the tiling and charge solving. This introduces a large ambiguity of the wire readout and a high probability of the presence

[2] Within a time slice, each wire plane provides a 1D projection view.

[3] The minimal requirement of reconstructing a 2D image is two 1D projection views. Naively, from two wire planes to three wire planes, we have three times more combinations of wire plane pairing to reconstruct the 2D image.

of ghosts. As indicated in the equation of the first principle, the 3D space points are reconstructed by matching the charge for all the functional wire planes. The images are strictly consistent with the original charge measurements in the 2D wire-versus-time views, as shown in figure 6, figure 7, and figure 8, except for the nonfunctional wire regions where any possibility is allowed. As a result, the ghost proto-clusters are mainly present in the nonfunctional regions, explaining a certain amount of the measured charge. The ghost proto-clusters are typically detached from the genuine tracks in 3D space, and in the 2D wire-versus-time views they are most likely redundant in the functional wire planes overlapping a more prominent genuine track.

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

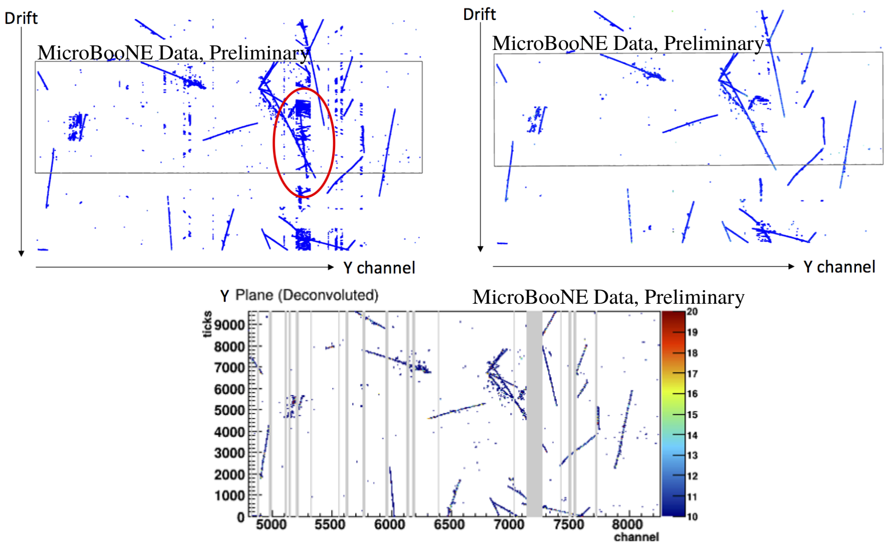

Figure 6: Top left: 2D projection to the Y plane's wire-versus-time view of the reconstructed 3D image without the de-ghosting algorithm. The red circle corresponds to the same volume in the TPC as in figure 7 and figure 8. Top right: with the de-ghosting algorithm. Bottom: Original charge measurement. The vertical axis bin width (time) is 4 ticks 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 grey.

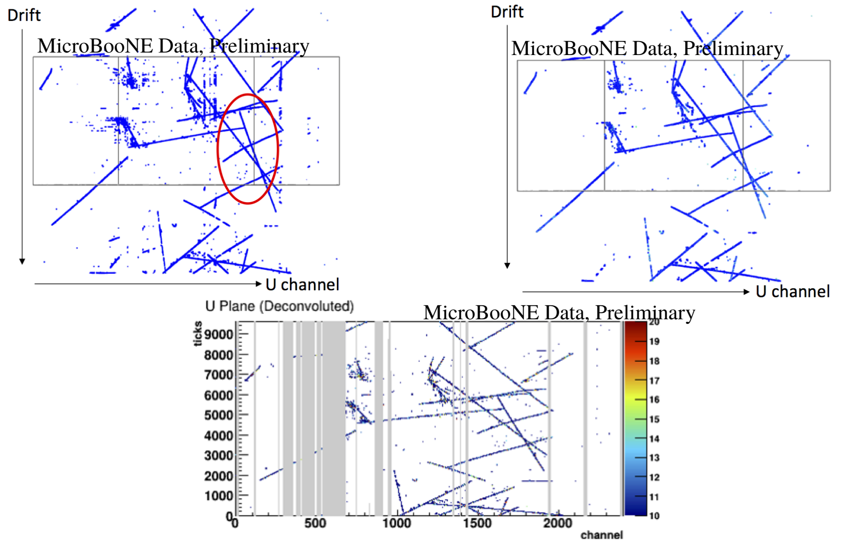

Figure 7: Top left: 2D projection to the U plane's wire-versus-time view of the reconstructed 3D image without the de-ghosting algorithm. The red circle corresponds to the same volume in the TPC as in figure 6 and figure 8. Top right: with the de-ghosting algorithm. Bottom: Original charge measurement. Y-axis bin width (time) is 4 ticks 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 grey.

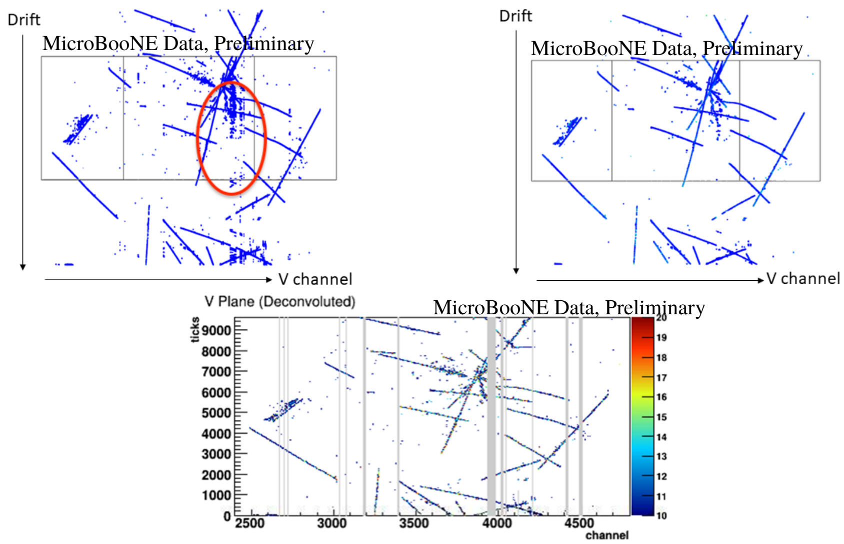

Figure 8: Top left: 2D projection to the V plane's wire-versus-time view of the reconstructed 3D image without the de-ghosting algorithm. The red circle corresponds to the same volume in the TPC as in figure 6 and figure 7. Top right: with the de-ghosting algorithm. Bottom: Original charge measurement. Y-axis bin width (time) is 4 ticks 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 grey.

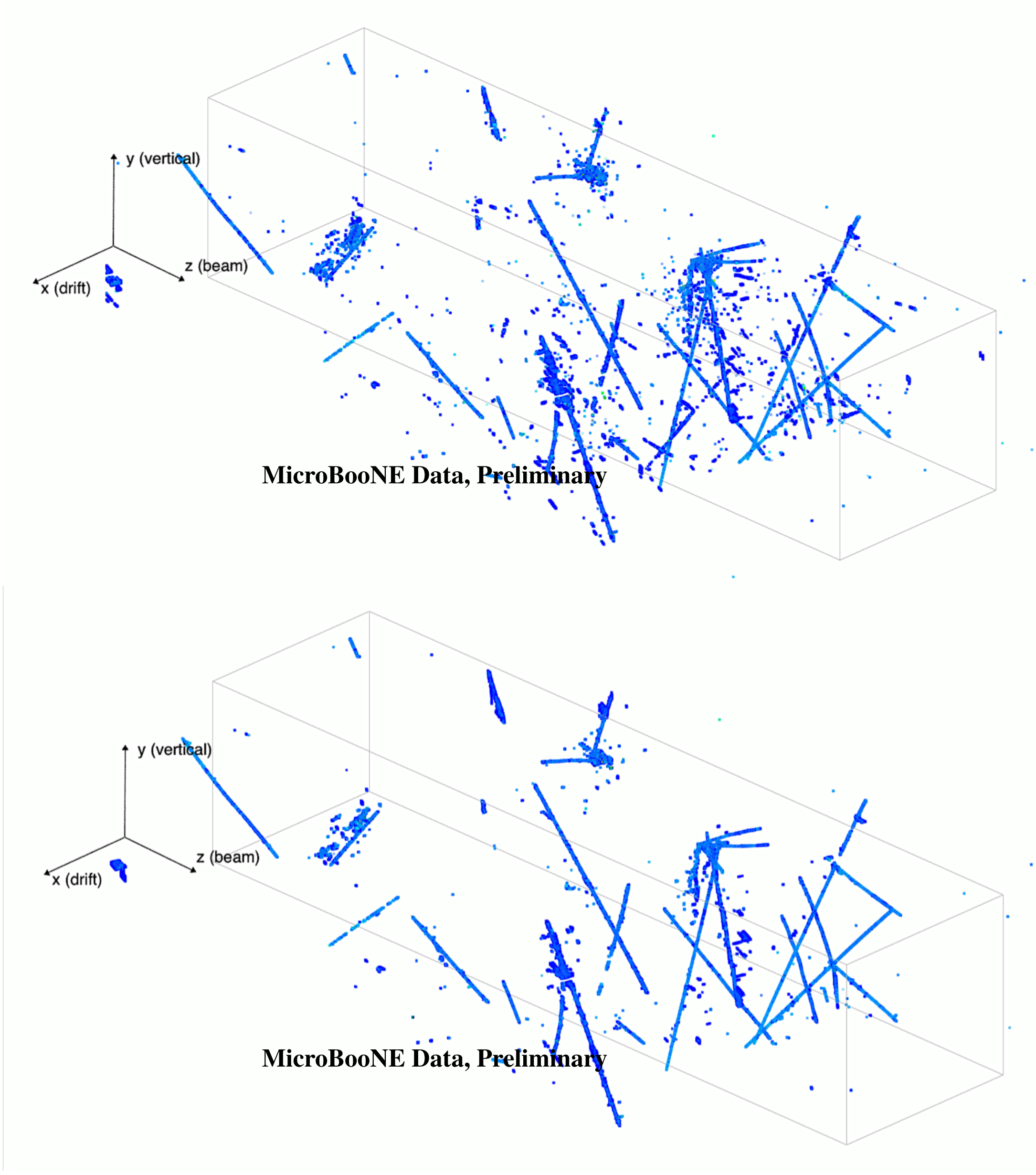

Figure 9: Comparison of the 3D imaging results from MicroBooNE data (event 41075, run 3493) without (top) and with (bottom) the de-ghosting algorithm. Ghosts are significantly reduced after the de-ghosting. Color indicates charge density.

The situation is aggravated by the inefficacy of the noise filtering [9] and the signal processing [10] which may respectively filter out some charge along the isochronous tracks as coherent noise and fail to reconstruct the charge of prolonged tracks due to the bipolar cancellation of the induction plane signals. Consequently one or two of the 2D wire-versus-time views of the charge measurements may have gaps along a track on the functional wire planes. This gap will lead to a gap in the 3D image since the charge measurements across the wire planes can not match. Consequently, the successfully reconstructed charge from the other wire planes corresponding to the gap will probably interplay with the charge measurements from other tracks and be erroneously explained by ghosts. The removal of such ghosts requires a knowledge of the broken tracks, i.e. the bridging of the gaps to connect the separated pieces of the track. This will be further discussed in section 4.1 in the context of an appropriate *clustering* which results in the final clusters based on the proto-clusters.

3.4 Summary

The actual procedure of the application of Wire-Cell 3D imaging in MicroBooNE is iterative, containing multiple rounds of tiling, charge solving, and de-ghosting. An overview is shown in table 1.

Table 1: Overview of the procedure of the Wire-Cell 3D imaging including the 2-plane (≥2 wire planes) tiling, the charge solving, and the de-ghosting.

Step	Description
0	2-plane tiling
1.1	De-ghosting
1.2	1st round of charge solving
1.3	2nd round of charge solving with reweighting for connected blobs
2	Repeat Step 1.1) to 1.3)
3	Repeat Step 1.1) to 1.3)

Since MicroBooNE is a near-surface detector with minimal cosmic shielding, 20-30 cosmic-ray muons per event are input to the Wire-Cell imaging process in the full readout window of 4.8 ms. The time and memory consumption are practical issues to be addressed in the optimization and finalization of the algorithms. Using ~10k MicroBooNE data events, the average time and memory consumption of Wire-Cell 3D imaging are estimated to be about 2 min and less than 2 GB using a single CPU. Most of the memory is used by the tiling to initialize and index the blobs from each time slice. Most of the time is consumed by the charge solving and de-ghosting, which are critical to the quality of the 3D images.

The goal of the Wire-Cell imaging is to reconstruct the 3D image of the ionization electrons independently from the event topology and prior to the application of pattern recognition techniques. The reconstructed 3D image will be fed to the subsequent reconstruction, e.g. the charge-light matching, to distinguish the in-beam neutrino activity candidate from the cosmic backgrounds. The

3D charge associated with each reconstructed space point is another unique feature as used in the prediction of PMT light signals.

Isochronous tracks present a common problem for the LArTPC 3D imaging, as the wire readout ambiguity is dramatically increased, e.g. the hit multiplicity is maximized in the time slice containing the isochronous track. In Wire-Cell 3D imaging, this issue is mitigated by introducing tiling, however, the blobs of the isochronous track are significantly broadened, leading to a much worse 2D spatial resolution only in the Y-Z plane view. Such an example can be found in figure 17. Improvement of spatial resolution can be achieved via trajectory fitting in a later stage. This is beyond the scope of this paper and will be presented in a future publication.

Because of the existence of ∼10% nonfunctional channels, a 2-plane (≥2) tiling strategy is adopted to significantly enhance the volume efficiency at the cost of introducing a large amount of ghosts. Time, geometry, charge, sparsity, positivity, and proximity information are utilized to overcome the wire readout ambiguity and to remove the ghosts. Another round of de-ghosting is performed in the clustering stage as discussed in section 4.1.3. The de-ghosting algorithm has little impact on true charge blobs except for those dot-like charge depositions. This can be seen in the evaluation based on a simulation of ideal tracks as shown in figure 26. Quantitative evaluations of Wire-Cell imaging for various cases are presented in section 5.

4 Matching TPC clusters and PMT flashes

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

In order to accurately and robustly pair the TPC activities to the PMT flashes, a proper 3D clustering is developed to group the corresponding space points[4] for each individual TPC activity into *clusters*. There might be disconnected pieces of a TPC activity for various reasons and the *proto-cluster* from the imaging, which is solely based on proximity, cannot completely represent a TPC object.

Given the TPC clusters and PMT flashes, a novel algorithm, *many-to-many charge-light matching*, is developed to match them simultaneously based on the predicted light signals generated from 3D TPC clusters and the measured light signals from PMT flashes. The TPC cluster(s) matched to an in-beam PMT flash is then regarded as a beam neutrino candidate. All the remainders are rejected as cosmic-ray muons. Compared to a previous single-to-single track-light-matching algorithm as described in ref. [18], many-to-many charge-light matching enhances the cosmic rejection power and results in a cleaned-up 3D image of neutrino activity.

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

4.1 3D clustering

Clustering as described in this section aims to group proto-clusters according to their original TPC physics activities into clusters, each of which can represent a complete TPC object from a cosmic-ray muon or a neutrino interaction. This step is an initial separation between neutrino and cosmic activities, and is inevitable to efficiently perform the subsequent many-to-many charge-light matching.

Proto-clustering, which solely relies on proximity, has been performed in the imaging. However, it hardly meets the requirement of carrying out a high performance charge-light matching because of a couple of issues.

Gaps: The presence of gaps compromises the effectiveness of a clustering based on proximity. A gap mainly results from: 1) the ~3% nonfunctional regions as shown in figure 4; 2) removal of isochronous tracks (close to parallel to the wire planes) by the coherent noise filter, as shown

[4]Central points of the voxels in the reconstructed 3D image; for a blob, the central points of constituting cells are used.

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

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

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

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

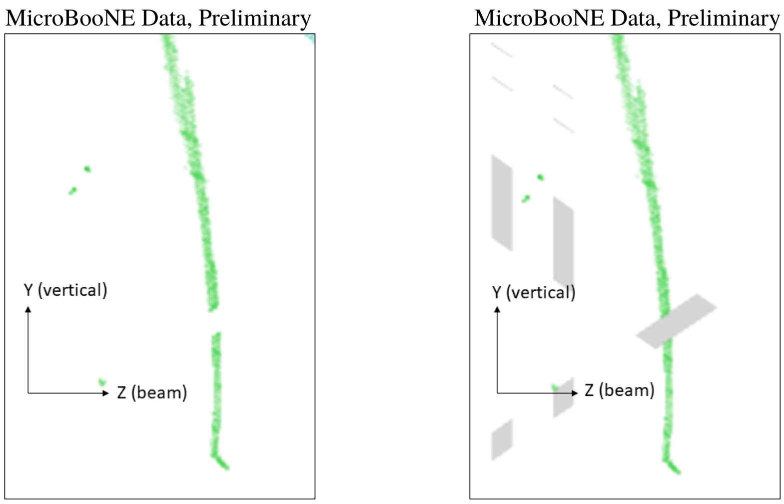

Figure 10: Y-Z view of a cosmic muon with a gap because of the nonfunctional wires (labeled in grey).

4.1.1 Clustering with the presence of gaps

Clustering across gaps mainly relies on two sets of information: **distance** and **directionality**. If two proto-clusters are close to each other along a line, the gap may be bridged and the two proto-clusters are grouped into a single cluster. Many existing tools and algorithms operating on a point cloud can be directly used, as a TPC cluster is a cloud of the reconstructed 3D space points. The distance

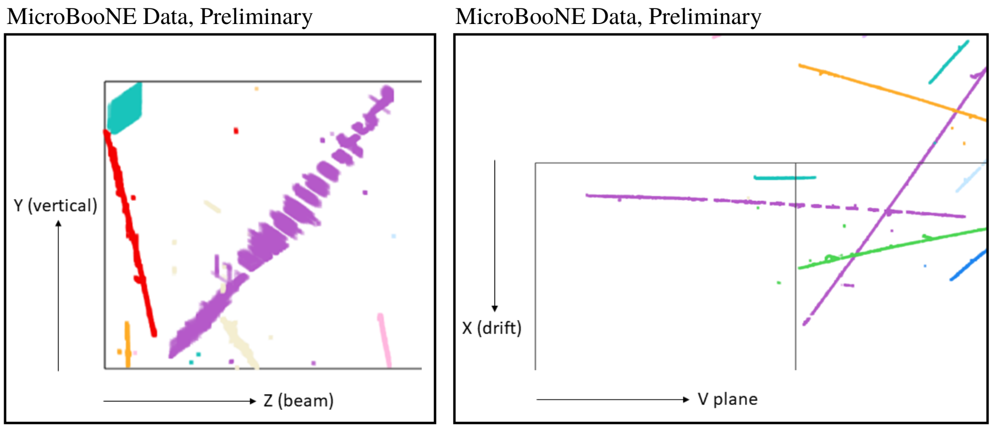

Figure 11: Gaps from the coherent noise filter applied in the signal processing chain. (Left) Y-Z view of a cosmic muon close to parallel to the wire plane. (Right) X-V view of the same event. V plane direction represents the wire pitch direction of the V wire plane.

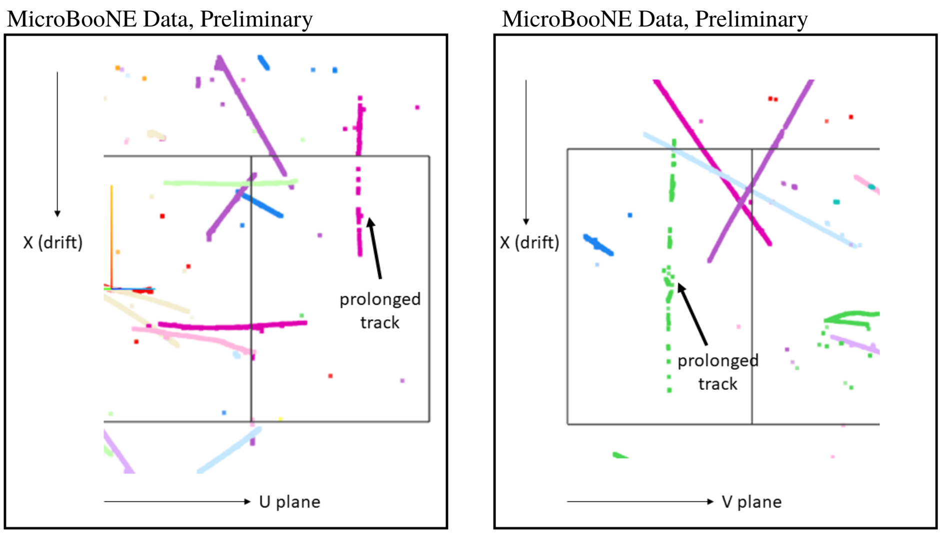

Figure 12: Gaps along prolonged tracks in 2D wire-versus-time(drift) views. (Left) X-U view for a prolonged cosmic muon track (purple) from a MicroBooNE data event. (Right) X-V view for a prolonged cosmic muon track (green) from another MicroBooNE data event. Cluster membership is indicated by uniform color in each plot.

between two clusters (point clouds) is defined as the minimal distance between a pair of space points from respective point clouds. To calculate this distance rapidly, the k-d (k-dimensional) tree based algorithm as implemented in the 'naoflann' package [19] is employed. Once the minimal distance and its direction is obtained, its direction is compared with the directions of the two proto-clusters. These are found using a voting scheme inspired by the Hough transformation [20]. The directional vector, parameterized by a polar angle and an azimuthal angle, is calculated for each point. The most probable value in the 2D distribution of polar-versus-azimuthal is then taken as the primary direction of the point cloud. Given the minimal distance vector and the two proto-cluster directions, the two proto-clusters are grouped (or not) based on the minimum distance and consistency in directions. In practice, a set of criteria were developed and optimized by analyzing hundreds of data events for various topologies and cases.

Figure 13 shows a comparison of the results before and after applying this clustering algorithm. Each cluster is marked by a different color. Separate proto-clusters are successfully grouped into individual clusters. However, one can also see that there are other clustering issues: two connected cosmic-ray muons and an incomplete neutrino cluster. These are dealt with using additional clustering algorithms as introduced in section 4.1.2 and section 4.1.4.

4.1.2 Separation of coincidental overlap clusters

In this section, we describe another algorithm to separate a "coincidental overlap" cluster, and the steps are shown in table 2.

Table 2: Steps of the separation of a 'coincidental overlap' cluster.

Step	Motivation
I	Identification of "coincidental overlap" cluster
II	Find two end points of a primary track
III	Form trajectory of this primary track
IV	Collect space points of this primary track
V	Remove this primary track and repeat this procedure

The first step is to identify the "coincidental overlap" cluster. Principle component analysis (PCA) is performed on each cluster after the bridging of gaps described in section 4.1.1. For a single-track-like cluster, only the primary component (axis) of the PCA has a significantly larger eigenvalue in the data correlation matrix. This is generally not true for a "coincidental overlap" cluster in which two or more tracks are crossing. Once a candidate "coincidental overlap" cluster is identified, the sub-clusters from each individual activity are to be identified and separated one by one.

The separation of each sub-cluster starts with a search for the two end points of the first (arbitrary ordering) primary track (largest distance along the primary PCA axis). The quickhull [21] algorithm operates on the 3D space points of a cluster to obtain the 3D convex hull. The two end points of the current primary track must be contained or in close proximity with the convex hull's vertices. The nearby points around each convex hull's vertex are grouped together to form test clusters. The

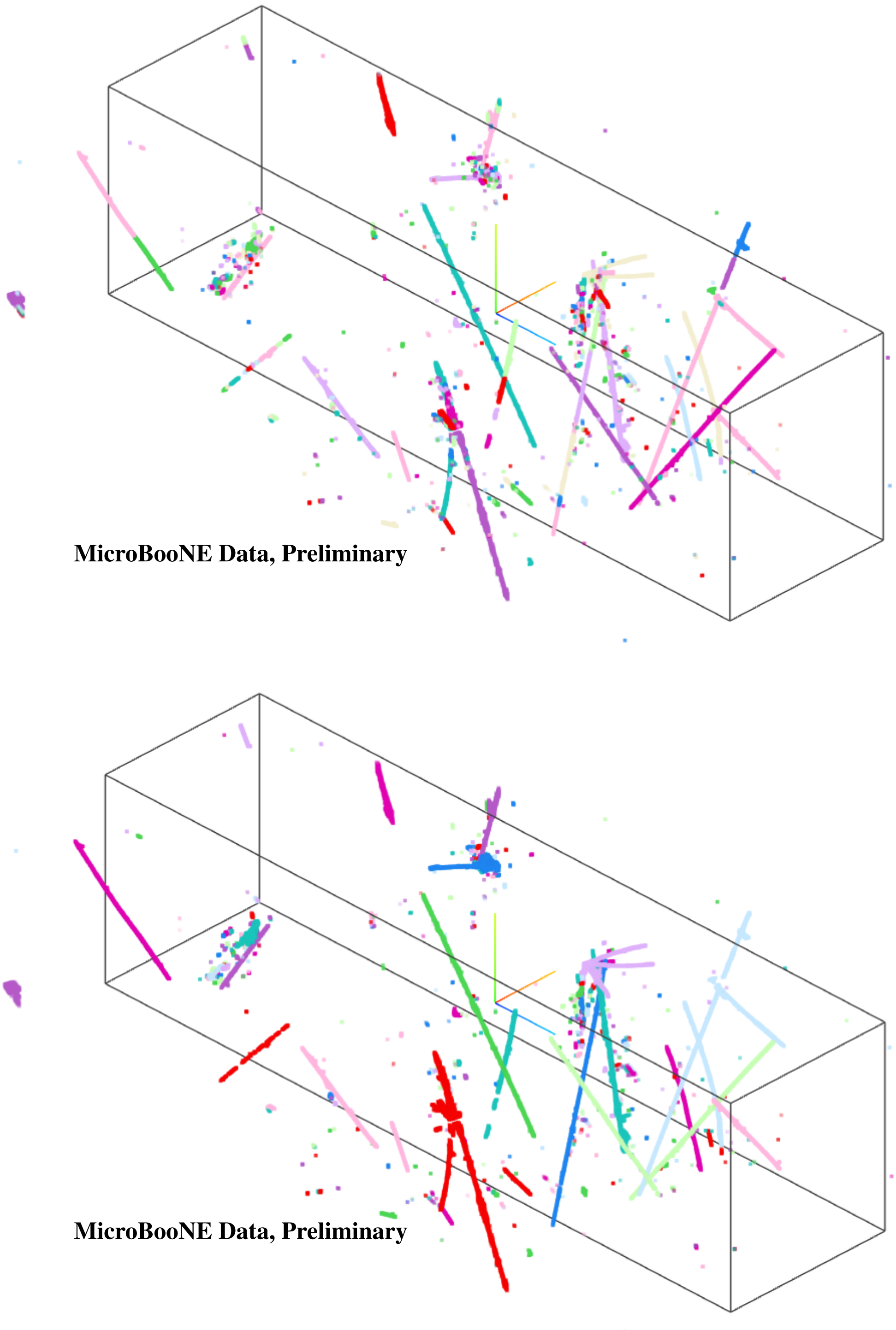

Figure 13: Demonstration of the effectiveness of the algorithm of bridging gaps. Top: proto-clusters solely based on proximity. Bottom: clusters after the application of the algorithm of bridging gaps. Cluster membership is indicated by uniform color.

largest test clusters are used to discover the end points of the primary track, and this requires 1) a small distance to the PCA primary component; 2) a consistent direction of the test cluster with the PCA primary component. Once an end point is found, a Kalman-filter-based technique is used to crawl along this primary track until the other end point is determined. Given the two end points, the trajectory of this primary track is obtained using a graph theory operation, Dijkstra's shortest path. The connected component algorithm from graph theory is then used to collect the space points associated with this trajectory and form a sub-cluster. After removing this sub-cluster of the first (arbitrary ordering) primary track (largest distance along the primary PCA axis), the remaining cluster is further examined and separated until only one primary track is left. Each sub-cluster is taken as an individual cluster in the end.

Figure 14 shows a comparison of the results before and after applying the separation algorithm. Two "coincidental overlap" clusters show up in this event: one case has two cosmic muons crossing each other, the other has a cosmic muon grouped to the neutrino interaction. Figure 15 shows another example, where two cosmic muons cross each another and one of the muons induces an EM shower. After the separation step, part of the EM shower is improperly separated. This could be addressed by the many-to-many charge-light matching, which can further group the clusters into a cluster group after matching the PMT flash. A cluster of a neutrino interaction with multiple final state particles might be identified as "coincidental overlap" as well. There is some protection against over-separating clusters because the neutrino final state particles are mostly forward-going along the beam direction and not fully through-going. Additionally, a dedicated clustering algorithm to group the separate clusters for neutrino interactions is performed later (section 4.1.4).

4.1.3 De-ghosting again

As mentioned in section 3, a de-ghosting algorithm is applied in the imaging stage to remove ghosts following the principle of sparsity of the LArTPC physics activities. This was done prior to clustering. This strategy is initially inefficient since a proto-cluster cannot appropriately represent a complete TPC object. Given the improvements during the clustering stage as described above, the de-ghosting algorithm can be run on the resulting clusters again to remove the residual ghosts.

We present some instructive examples of de-ghosting after clustering has been performed. As shown in the top panel of figure 16, there are some ghosts due to gaps along prolonged tracks. These ghosts cannot be removed during the imaging since they are the only explanation of the charge measurements in functional wire planes. With a bridging of the gaps, the original proto-clusters can be grouped into a bigger cluster, which as a whole can explain the charge measurements in all three wire planes. The ghosts related to gaps in this prolonged track can thus be removed. Another example as shown in figure 17 has a four-track cluster, in which two tracks are ghosts. This cluster is present in the region where there is a nonfunctional wire plane. Since all of these four tracks including the ghosts are connected, the two ghost tracks survive the de-ghosting procedure in the imaging stage. After the application of the algorithm to separate the "coincidental overlap" cluster, the two ghost tracks can be further examined and removed individually.

Figure 18 shows an example of a complex event with a large amount of residual ghosts after imaging. Many tracks including prolonged tracks and isochronous tracks go through the region (area on the left of figure 4) where U wires are mostly nonfunctional. Ghosts with various lengths and

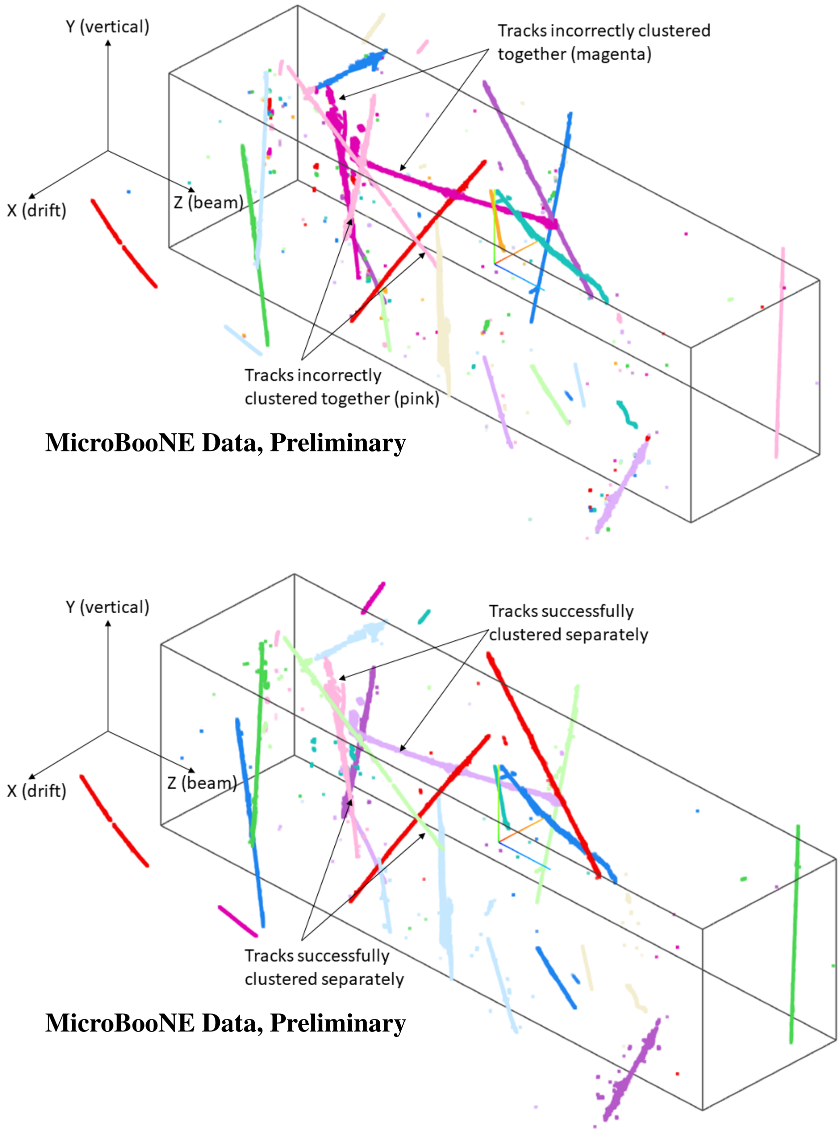

Figure 14: Demonstration of the effectiveness of the clustering algorithm to separate a "coincidental overlap" cluster. The top and bottom panels show the clusters before and after applying this algorithm. Cluster membership is indicated by uniform color.

positions are reconstructed. After bridging the gaps and separating coincidental overlap clusters, the ghosts can be significantly removed by re-running the de-ghosting algorithm.

4.1.4 Clustering for neutrino events

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

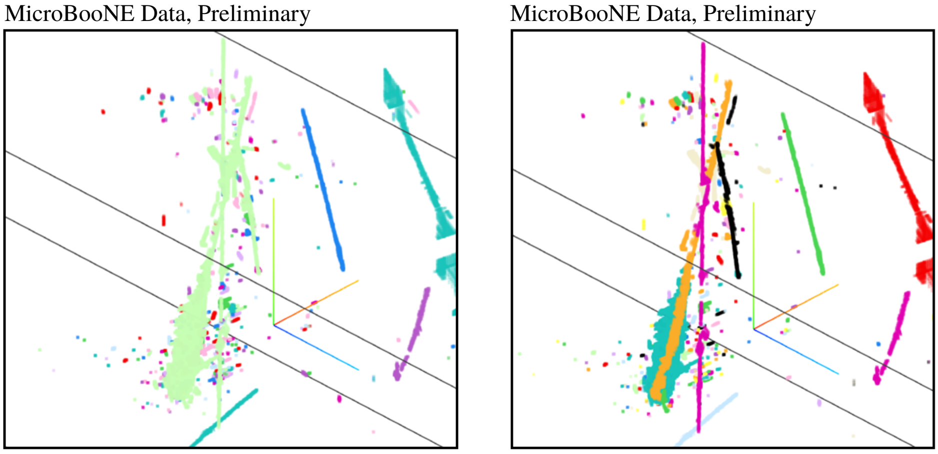

Figure 15: Demonstration of the effectiveness of the clustering algorithm to separate a "coincidental overlap" cluster. The left and right panels show the clusters before and after applying this algorithm. The all-light-green 3D cluster in the left panel is broken into its components in the right panel.

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

- Only clusters within the drift window (maximum drift distance) corresponding to the beam time are considered.
- The direction of each sub-cluster is calculated. End points are examined to ensure that they do not belong to any isolated dot-like clusters, which are ignored due to their small size.
- Each cluster is extended with new space points along the track direction near each end point.
- The extended clusters are examined to find the 'intersection' – connected or close – with other clusters. This 'intersection' is required to be formed by the extended part or the end points of the other clusters.

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

4.2 PMT light signal reconstruction

Proper clusters, which represent individual TPC activities corresponding to either cosmic-ray muons or neutrino interactions, are achieved by a set of advanced clustering algorithms as described in the previous section. Because of the relatively slow drift of ionization charges and spatially distributed, asynchronous activity throughout the TPC, TPC activities are mixed in the time sequence with unknown interaction (start) time. Scintillation light is produced and detected on a dramatically

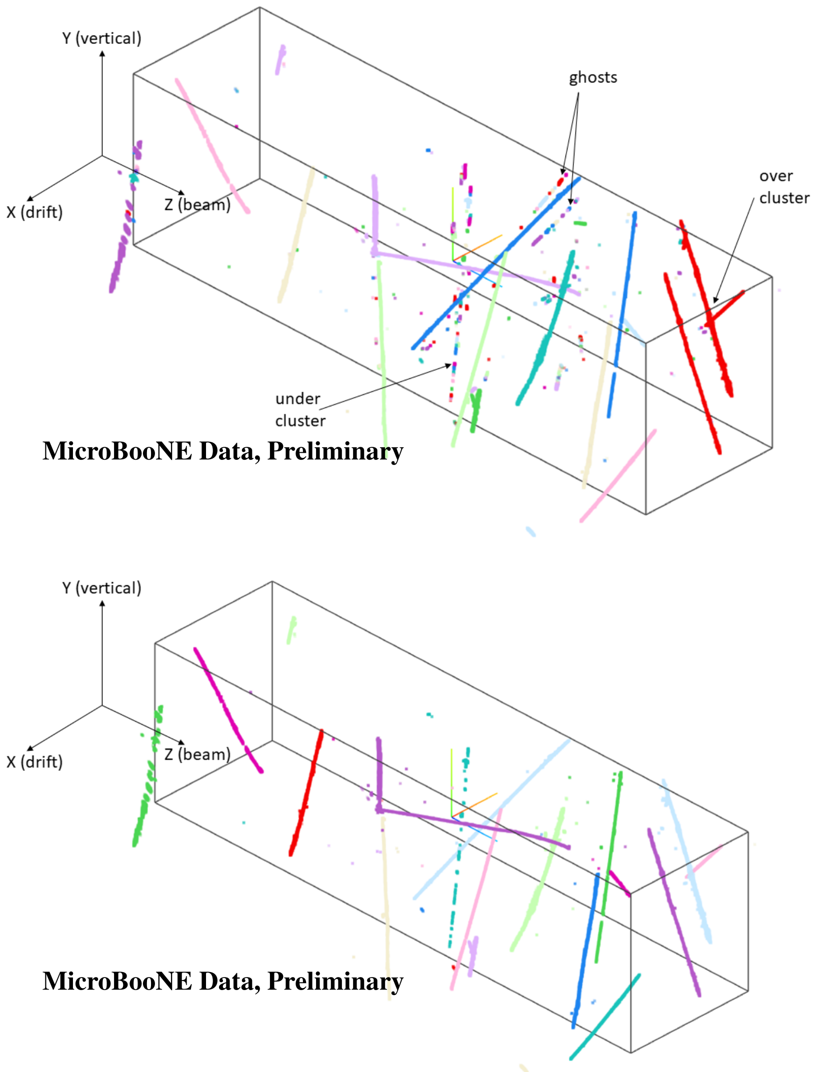

Figure 16: Demonstration of the effectiveness of the de-ghosting algorithm with other advanced clustering algorithms applied. The top and bottom panels show the clusters before and after applying the de-ghosting algorithm after bridging gaps. Color indicates cluster membership.

shorter time scale by spatially distributed PMTs. An offline processing of the light signals from PMTs is thus important to perform the many-to-many charge-light matching to select the neutrino activity corresponding to in-beam PMT signals.

As described in section 1, 32 PMTs are used to detect the scintillation light in MicroBooNE. In the PMT front-end motherboard (FEM), the PMT signal is separated by a splitter into a high-gain (x10) and a low-gain (x1), allowing a wide dynamic range for a 64-MHz 12-bit ADC readout of the PMT pulses [5]. In the PMT readout system, there are two separate readout streams: *beam discriminator* and *cosmic discriminator*. The beam discriminator starts 4 μs before the beam gate. It reads out 1500 consecutive samples (~23.4 μs) of the PMT waveforms. The cosmic discriminator is a self-triggered PMT readout. It reads out 40 consecutive samples (~0.6 μs) of the

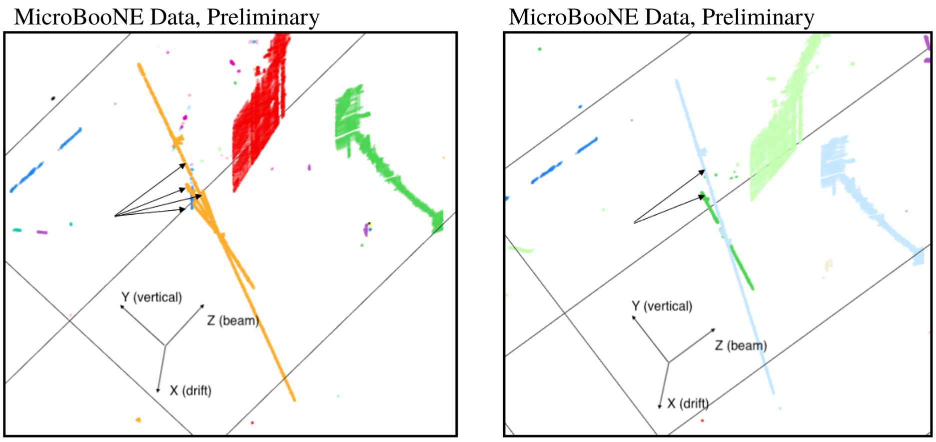

Figure 17: Demonstration of effectiveness of the de-ghosting algorithm with other advanced clustering algorithms applied. The left and right panels show the clusters before and after applying the de-ghosting algorithm following the separation of the "coincidental overlap" cluster. Color indicates cluster membership. The stripy tracks with much worse spatial resolution in Y-Z view correspond to big blobs of isochronous tracks as discussed in section 3.4.

PMT waveforms, which record the light information not only from beam-coincident activities but also activities out-of-time with the beam.

The PMT waveforms are processed offline to reconstruct the time and number of photoelectrons (PE) of a flash, which is a cluster of PMT signals close in time. For the beam discriminator, a deconvolution using Fast Fourier Transformation (FFT) is performed to remove the electronics responses from various RC circuits in the splitter and the shaper. A flash is then formed if the PMT measurements satisfy the multiplicity requirement (>2 PMTs above a threshold of 1.5 PE) and the total integrated PE threshold (>6 total PE) in a ~100 ns window. A flash window lasts 7.3 μs in order to exclude noise and to include the contribution from the late scintillation light. The scintillation light in liquid argon has a prompt and a slow component with decay times of about a few nanoseconds and 1.6 μs[5], respectively.

In the flash window, the time bin with the maximal total PE from all PMTs marks the time of a flash. The PE of each PMT in a flash is integrated over the entire flash window. Though the average time between two adjacent flashes in MicroBooNE is ~100 μs, a procedure is set to end the current flash window and start a new one if the new flash has a larger starting PE[6] and satisfies either of the two requirements: (1) at least 1.6 μs late; (2) a significantly different PMT hit pattern (number of PEs in each PMT) in the first 100 ns of the new flash compared (based on a Kolmogorov-Smirnov test) to the pattern in the last 100 ns of the previous flash. Figure 20 shows an example of two adjacent reconstructed flashes from beam discriminator PMT waveforms.

[5]The two lifetimes correspond to the molecular excimer states excited either in a singlet state or a triplet state.
[6]The starting PE of a flash is calculated as the total PE from all PMTs in the first 100 ns.

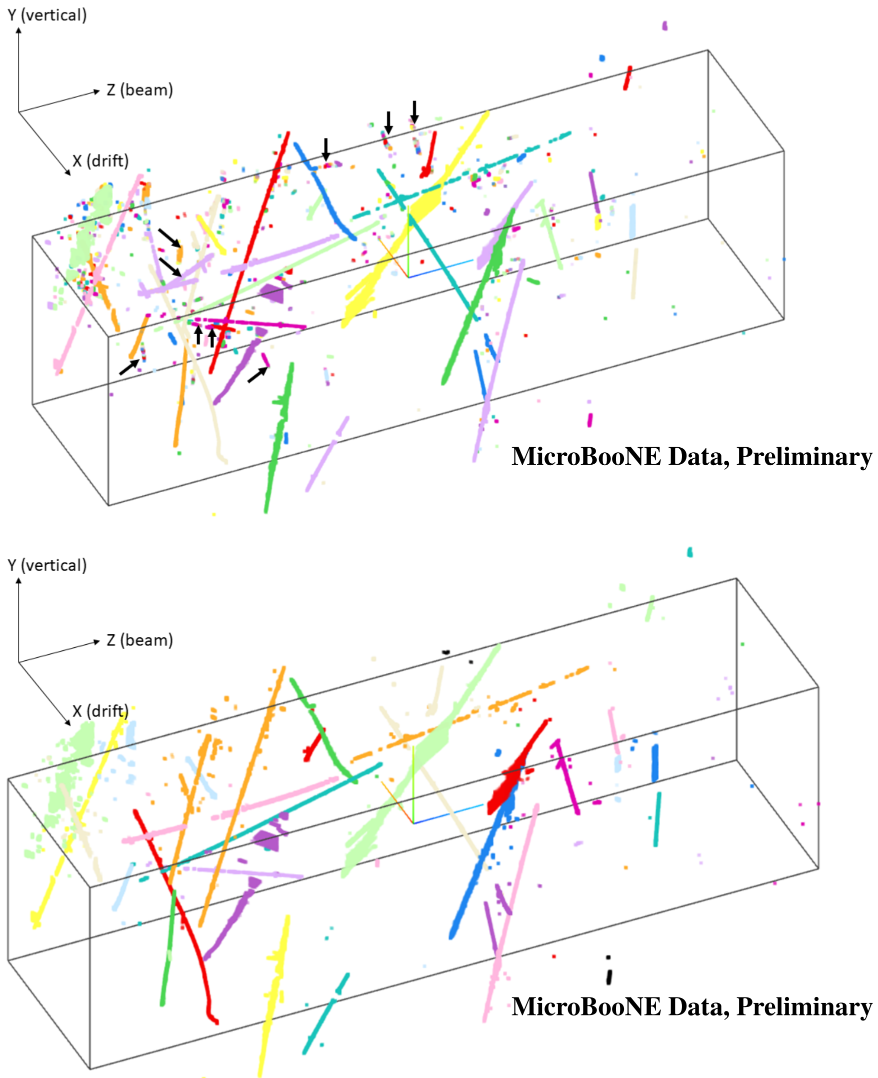

Figure 18: Demonstration of the effectiveness of the de-ghosting algorithm with other advanced clustering algorithms applied. The top and bottom panels show the clusters before and after applying the de-ghosting algorithm. This is a challenging case where multiple tracks go through a region where one wire plane (U plane) is largely nonfunctional. Color represents cluster membership.

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

[7] It is optional to unfold the slow component of the scintillation light, thus identifying Michel electrons.

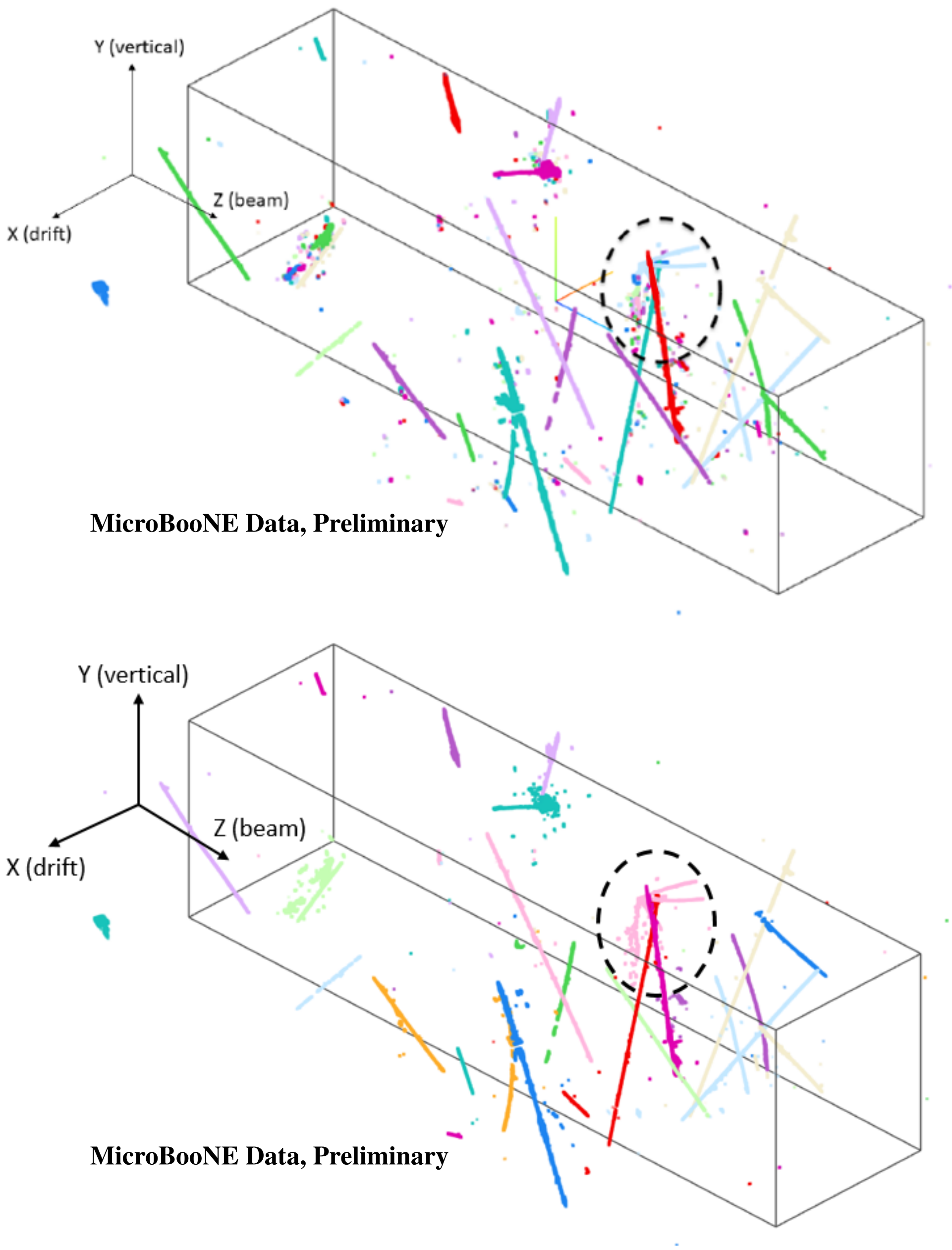

Figure 19: Demonstration of the effectiveness of the clustering algorithm designed for neutrino interaction. Top and bottom panels show the clusters before and after applying such clustering algorithm. The neutrino interaction (the light pink cluster) is in the black dashed circle with multiple particles emitted and two electromagnetic showers (two γ's from a π^0 decay).

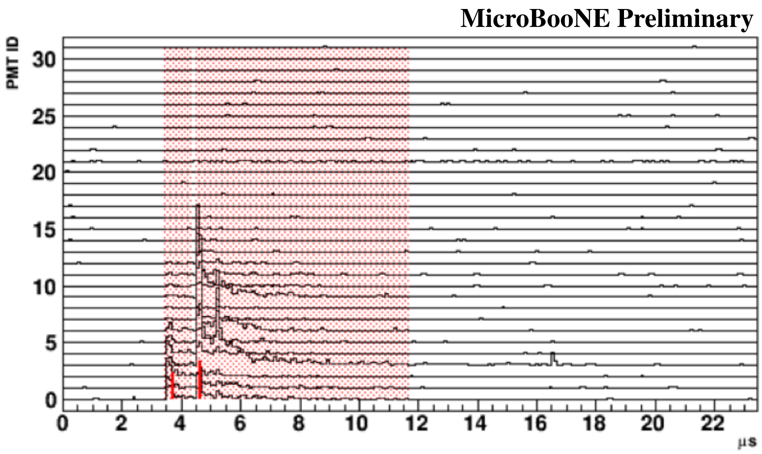

Figure 20: Illustration of two reconstructed flashes from beam discriminator PMT waveforms. The black curves are the deconvolved PE spectra for each PMT. The red lines represent the flash times and the red bands represent the flash windows. For the second flash at about 4.6μs, there is a Michel electron as indicated by the second peak (at about 5.3μs) of its PE spectra[7].

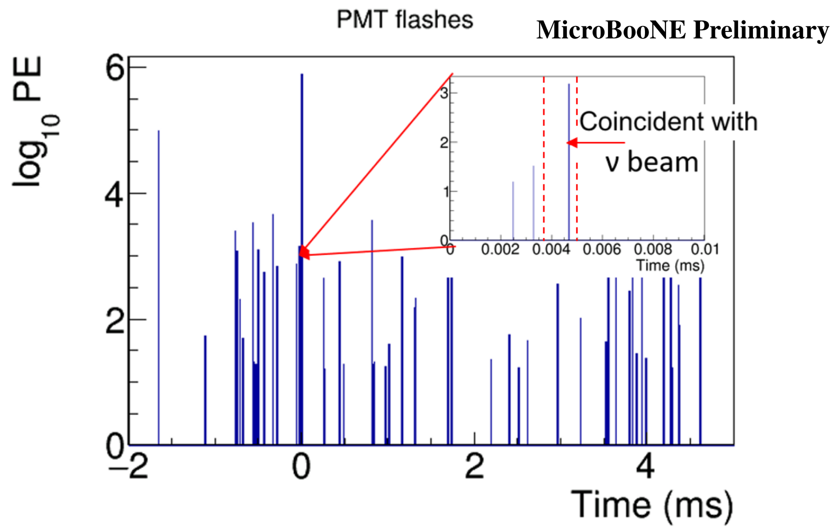

Figure 21: The reconstructed PEs of a flash as a function of flash time. The 6.4 ms PMT readout window is shown relative to the trigger time. The flashes from the beam discriminator (23.6 μs long) are shown as inset. The flash in coincidence with the BNB beam spill (between dashed red lines) is indicated.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

Figure 21 shows the reconstructed PEs and time for each PMT flash from a data event. The flash corresponding to the beam spill is shown in the inset figure near 4μs.

4.3 Many-to-many charge-light matching

Now that the TPC charge activity has been reconstructed and grouped into physically distinct clusters in section 3 and the PMT light measurements have been reconstructed into distinct flashes in section 4.2, it is time to match the 20-30 TPC clusters to the 40-50 PMT flashes. This will allow each matched cluster to be assigned the precise time measurement of the PMTs, allowing the strict BNB time window to reject the vast majority of non-causally-allowed cosmic rays from consideration as neutrino candidates.

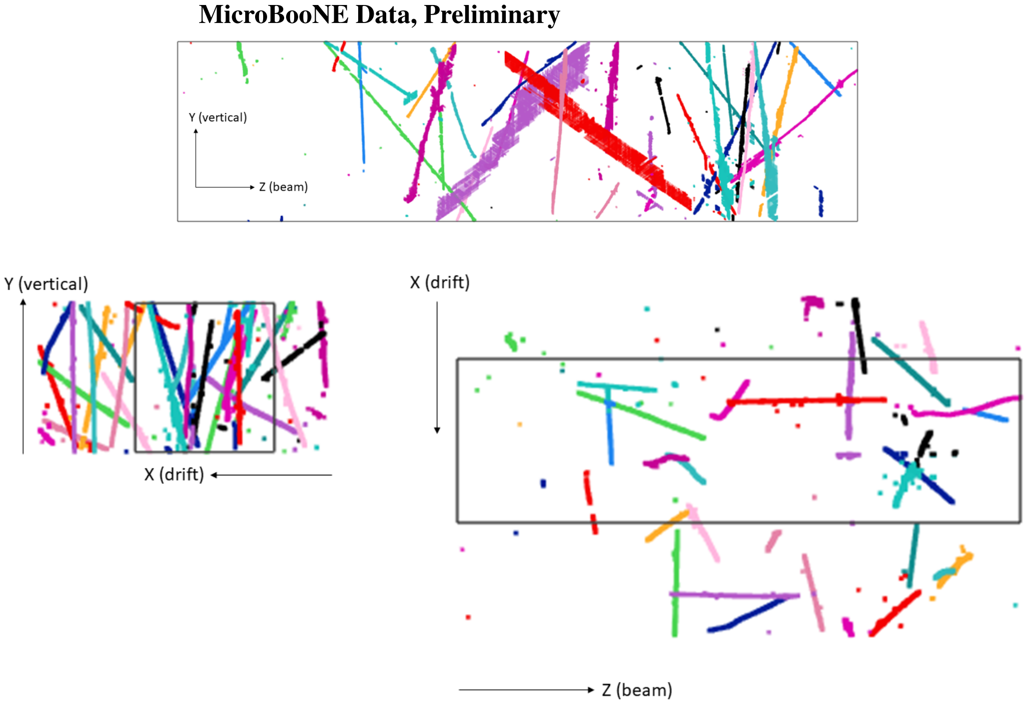

Figure 22: An example of all the TPC clusters from a MicroBooNE data event before matching. Different clusters are labeled in different colors. Top: front (Y-Z) view. Bottom left: side (Y-X) view; the left boundary (anode plane position) of the nominal detector volume (black box) corresponds to the neutrino beam time, and the width of the black box corresponds to the maximum drift window (2.3 ms). Bottom right: top (X-Z) view. The TPC clusters cover the entire 4.8 ms TPC readout window which is about 2 times the maximum drift window.

As shown in figure 22, there are many TPC clusters spanning over the entire readout window with unknown electron drift start time. The X-position is just a direct conversion from wire readout time. More PMT flashes are generally recorded since PMTs sense not only the activities inside the TPC but also those outside the TPC within the cryostat. Note that a TPC cluster does not necessarily have a corresponding PMT flash since the light collection system (e.g. the cosmic discriminators) has inefficiencies, especially for clusters either with low visible energy or near the cathode (far from the PMTs). Also, as mentioned in section 4.1, the resulting clusters after the application of the advanced clustering algorithm may still have an under-clustering issue which is intended to be solved in the matching stage, so a collection of TPC clusters are permitted to match to a single PMT flash. As a result, two requirements need to be satisfied in the matching algorithm:

(A) One TPC cluster can match to zero or at most one PMT flash.

(B) One PMT flash can match to zero, one, or multiple TPC clusters.

The “match” is defined as a good agreement between the predicted and measured (reconstructed) light signals, considering the signal strengths for each individual PMT as well as the hit pattern of the 32 PMTs. Assuming a TPC cluster to be associated with a PMT flash, a prediction of the PE distribution for the 32 PMTs can be made. The electron drift start time of the TPC cluster is shifted from the default BNB beam time to the measured time of the PMT flash. This enables a correction of the X-position of the TPC cluster. Then the solved charge associated with each space point in 3D images is used to predict the PMT light signals based on a *photon library* [22]. To reflect the position dependence of photoelectrons (PEs) on each 8" PMT, a photon library is pre-generated based on Geant4 to estimate the PMT acceptance of optical photons emitted at different locations in the TPC volume. The TPC volume is divided into 75 (~2.5 m width) × 75 (~2.3 m height) × 400 (~10.4 m length) voxels. Millions of optical photon tracks (with a wavelength of 128 nm for LAr scintillation light) with a 4π angular distribution were generated in each voxel and simulated with realistic optical photon processes of absorption and scattering. Then, the PEs from each of the PMTs for a given TPC cluster can be predicted by applying the PMT acceptance for the 3D charge for each space point. An overall scaling factor is also applied to take into account the scintillation light yield per unit deposited energy.

As one can imagine, such a many-to-many matching problem is very similar to the charge solving problem as introduced in section 3.2. There are more unknowns than knowns in this system, and the first principle equation is to relate the predicted light signals from all possible TPC clusters to the measured signals from PMT flashes. Hypothetical pairs of TPC clusters and PMT flashes are created and tested, in order to find the most compatible ones and eliminate the rest. Again, the compressed sensing technique is utilized to perform this many-to-many matching by minimizing a ℓ_1-regularized chi-square function.

In practice, a set of matching algorithms were developed to pre-select, fit (ℓ_1-regularized chi-square), and re-examine the hypothetical TPC-PMT pairs.

Pre-selection: A pre-selection of the hypothetical TPC-PMT pairs is important to reduce the number of unknowns in the ℓ_1-regularized chi-square fitting, allowing for a robust minimization. Two major tests, time range compatibility and PMT hit pattern compatibility, are performed to remove the incompatible TPC-PMT pairs. For the time range compatibility, the TPC cluster is required to be fully contained within the maximum drift window corresponding to the PMT flash time[8]. For example, as shown in figure 22, X positions (along the drift) of the space points in any TPC cluster have an overall shift because of the unknown electron drift (start) time, but the in-beam activity must be contained in the nominal detector volume (black box) which corresponds to the beam time. For the PMT hit pattern compatibility, the pairs with highly incompatible predicted and measured light signals are ruled out. A Kolmogorov-Smirnov test (K-S test) and a chi-square test, which inspect the hit pattern and the absolute normalization of the 32 PMTs' signals, respectively, are combined to discriminate the incompatible pairs. Particularly, to enable a one-to-many PMT-TPC matching (see requirement (B) in this section), the TPC clusters paired to the same PMT flash are jointly tested to maintain the many-to-one potential. The most compatible TPC-PMT pair is used as a base and the other ones are added individually to check the change in compatibility. The pairs which significantly reduce the compatibility are ruled out.

[8] A precise cut can be applied since the space charge effects, i.e. the squeezed drift electric field and the biased reconstructed position, are insignificant along the drift direction.

Chi-square fitting: Given the passing candidate TPC-PMT pairs, a chi-square function incorporating a ℓ_1-regularization term is constructed to compare the predicted and measured light signals:

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

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

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

$$\chi_{p2}^2 = \sum_i \frac{b_i^2}{c_2^2}, \tag{4.4}$$

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

For the input TPC-PMT pairs, the index i runs through all PMT flashes (a group of PMT signals close in time), j runs through all fired PMTs of each flash, and k runs through all the TPC clusters. M_{ij} and δM_{ij} represent the measured PE and its uncertainty of the j-th PMT in the i-th flash, respectively. The uncertainties from light yield and charge measurements are conservatively assigned. P_{ikj} represents the predicted PE of the j-th PMT in the i-th flash from the k-th TPC cluster. The a_{ik}'s, which represent the possibility of the k-th TPC cluster to the i-th PMT flash pair, are the parameters of interest in the fit. All a_{ik}'s are constrained to be non-negative. A TPC-PMT pair of good match will have an a_{ik} close to 1, while a bad match will have a_{ik} close to zero. χ_{p1}^2 applies the constraints that each TPC cluster should only be used once – matched to at most one PMT flash. The introduction of the b_i term is to take into account the possibility that some of the PMT flashes may not be associated with any TPC clusters in which case b_i is close to 1, though the χ_{p2}^2 term gives the constraint that b_i is preferred to be close to 0. The χ_{p3}^2 term represents the application of the compressed sensing technique which prefers a best-fit solution where most of a_{ij} terms are zero. λ is the regularization strength. c_1 and c_2 are two constants to regularize the corresponding penalty terms, and the values are 0.01 and 0.025, respectively, tuned by real data. After the first round fitting, the most incompatible TPC-PMT pairs with extremely small a_{ik} values are eliminated from further consideration. Naturally, PMT flashes that do not match to any TPC clusters are eliminated as well. The remaining TPC-PMT pairs go into the second round fitting to further approach the best solution, with the unnecessary b term removed.

Re-examination: After the two rounds of chi-square fitting, the hit pattern compatibility test as introduced in "Pre-selection" is again performed on the passing TPC-PMT pairs. Note that the power of the K-S test cannot be easily incorporated into the chi-square fitting. Instead, for each TPC cluster, the most probable TPC-PMT pair, with the largest a_{ik}, is then taken as the final result. For each PMT flash, the pairs of the TPC clusters that are detrimental to the compatibility are removed. As the last step, all the unmatched TPC clusters will be tested against the unmatched PMT flashes to keep the possible pairing.

Figure 23 shows an example of 6 final matched pairs from MicroBooNE data. In this event, there are a total of 31 matched pairs. After the many-to-many matching, the in-beam, flash-matched TPC activities are taken to be neutrino interaction candidates, and the remainder are rejected as

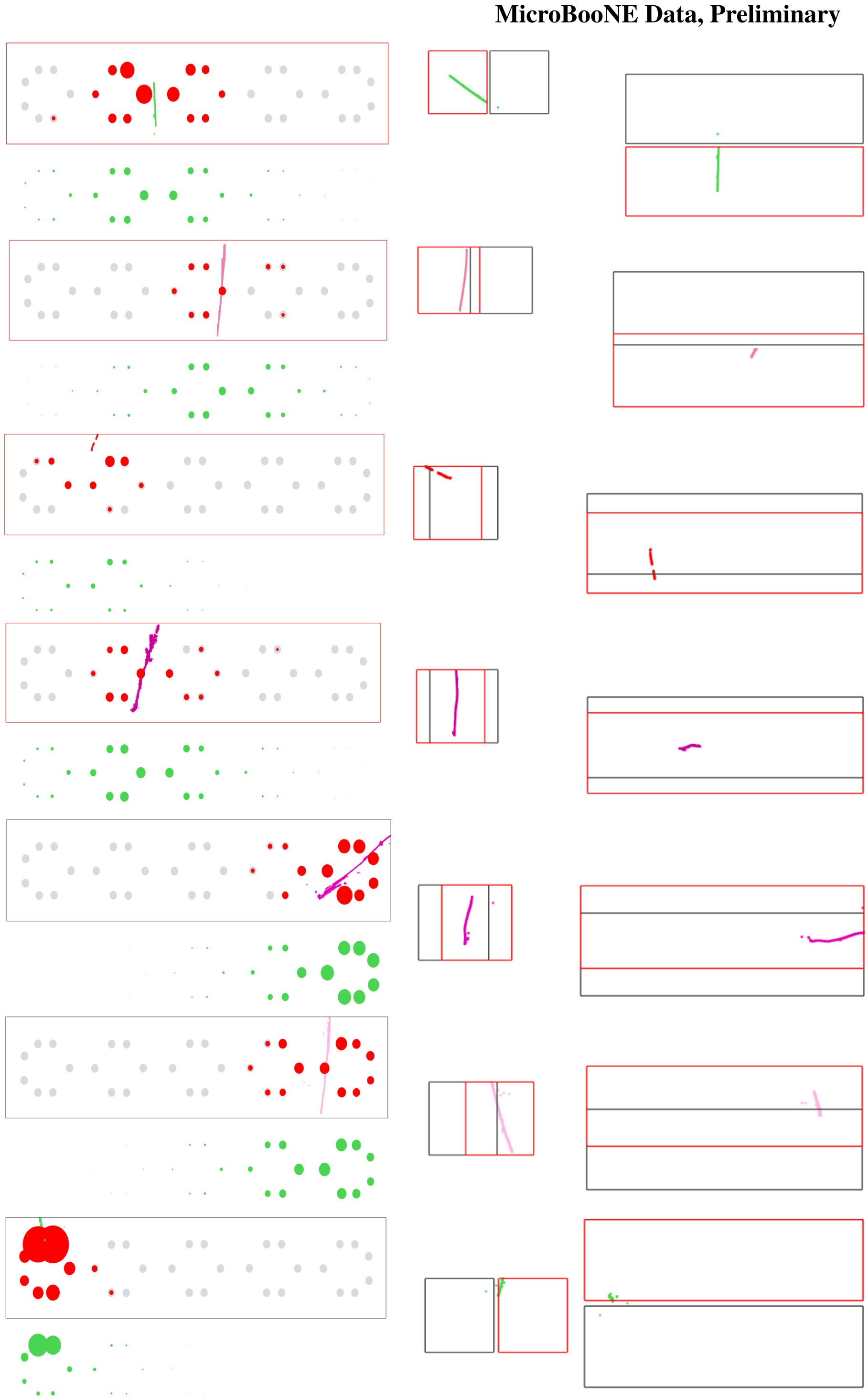

Figure 23: Selected 6 matched pairs from a data event, in which there are a total of 31 matched pairs. The red solid circles represent the measured PE for this flash. The green solid circles represent the predicted PE based on the matched TPC cluster(s), which are shown projected into the Y-Z plane. The area of the circle is proportional to the number of PEs. The black box represents the drift window relative to the beam time. The red box represents the drift window relative to the matched PMT flash time.

cosmic background. Figure 24 and figure 25 demonstrate successfully matching muon and electron neutrino clusters to their respective in-beam flashes. The performance of the matching algorithm is self-evident from these event displays and quantitative evaluations can be seen in section 5. More event displays can be seen in appendix A.

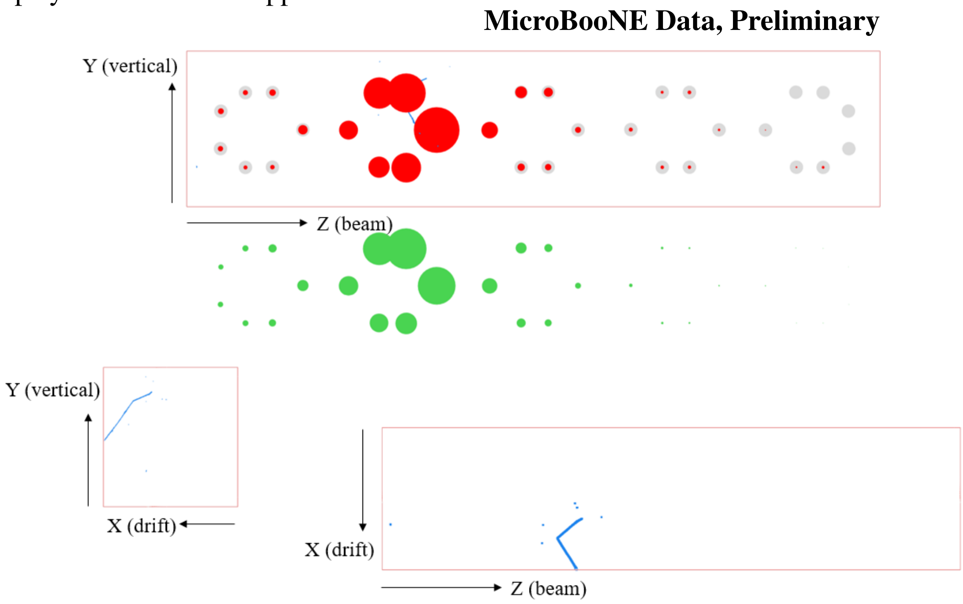

Figure 24: A muon neutrino event is shown with its matched flash. The area of the solid red circle is proportional to the measured PE. The green solid circles in the top panel represent the predicted PMT signals based on the two TPC clusters. The red box in the bottom panel represents the drift window relative to the matched PMT flash time.

4.4 Summary

In this section, we first describe the 3D clustering techniques which follow the initial tiling, charge solving, and de-ghosting algorithms. The challenges of gaps, coincidental overlap clusters, and separate clusters of neutrino interactions are addressed in the clustering stage. After the other clustering algorithms are applied, the de-ghosting algorithm is applied again, removing further ghosts not accessible in the first application. Many advanced techniques, such as the quick convex hull, Hough transformation, k-d tree, principle component analysis, Kalman filter, and various graph theory operations on point clouds are used or referenced in the clustering algorithms. At this point, we introduce a robust signal processing of the scintillation light signals. The PMT waveforms from beam discriminator and cosmic discriminator are both processed to reconstruct flashes. Given the TPC clusters and PMT flashes, we explain the many-to-many charge-light matching procedures, which match multiple TPC clusters (charge) to the corresponding PMT flashes (light). The many-to-many matching strategy considering all clusters (and not merely beam-coincident clusters) performs an overall optimization of the TPC-PMT pairing, boosting the ultimate performance of the matching result. After matching, only the in-beam flash-matched activity is considered as a potential neutrino candidate. This results in an about 30-fold reduction of the cosmic activity in each event. The matching (including clustering and light reconstruction) requires many fewer computing resources

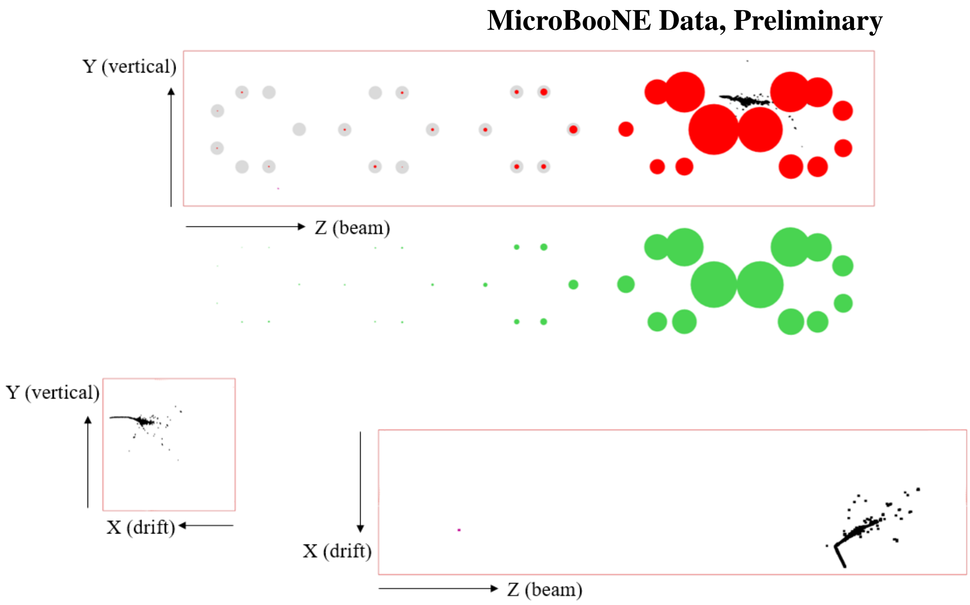

Figure 25: An electron neutrino event is shown with its matched flash. The area of the solid red circle is proportional to the measured PE. The green solid circles in the top panel represent the predicted PMT signals based on the two TPC clusters. The red box in the bottom panel represents the drift window relative to the matched PMT flash time.

than the imaging. On average, processing time is 30 seconds per event, consuming less than 1.5 GB memory for a single CPU.

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

In this section, the quantitative evaluations of the Wire-Cell 3D imaging and the many-to-many charge-light matching are presented. The performance of these three-dimensional approaches to reconstruct neutrino activities is demonstrated as well. As indicated in section 3, the intrinsic problem with the imaging stems from the wire readout ambiguity. This gets worsened by nonfunctional wires. As a consequence, ghost tracks (clusters) appear in the final 3D image and cannot be completely removed, even though compressed sensing algorithm is used with charge, sparsity, positivity, and proximity information. On the other hand, the true hits[9] might be discarded in the charge solving and the de-ghosting, since the final 3D image is expected to approximate the "best" solution based on the incomplete and inaccurate information of this under-determined system. Two major metrics are used to evaluate the quality of the imaging result as follows:

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

Completeness of the 3D image – the fraction of the true hits overlapping the reconstructed 3D images divided by the total number of the true hits. The true hits are required to be within the TPC active volume and are weighted by the deposited (visible) energy[10].

It is emphasized that the 3D imaging result is consistent with the original measurements in the 2D wire-versus-time views of the functional wires. The 3D metrics are relevant to understand the performance as the subsequent Wire-Cell reconstruction, for example, the beam-coincident cosmic rejection and the pattern recognition, are expected to operate on the 3D images, in order to maximize the potential capability of LArTPCs.

Given the numerous cosmic-ray muons in the TPC, the many-to-many matching (including the 3D clustering) is applied to properly group the neutrino activity and correctly match it to the in-beam flash. For neutrino activity with multiple final-state particles, the clusters from initial 3D clustering are expected to be merged into a cluster group in the matching stage. The clustering and matching may fail to select the neutrino activity or suffer both the over-clustering (impurity) and under-clustering (incompleteness) issues. The correctness of the matching and the efficiency of selecting neutrino interactions after matching are evaluated as well. These two metrics are defined as below:

Correctness of the matching – the fraction of all in-beam neutrino activity candidates which are indeed in-beam neutrino activity. The incorrect matches have no neutrino activity but cosmic activity with an extremely low completeness value as defined above.

Efficiency of neutrino interactions – the fraction of the events with neutrino interactions which have correct matches.

The development and optimization of these Wire-Cell 3D reconstruction techniques were based on ~1500 data events. All event displays in section 3 and section 4 are from MicroBooNE data. The evaluations in this section are carried out using the MicroBooNE detector simulation. The MicroBooNE simulation has incorporated a realistic TPC modeling which is in good agreement with data. A data-driven noise model and long-range wire responses [9–11] were implemented in

[9] One hit corresponds to one finite-size space point in a 3D image. "space point" is defined in section 4.

[10] There might be a large number of dot-like isolated true hits with very low energy depositions. The "size" of such a true hit is not well defined and the energy weighting represents a more sensible choice to evaluate completeness.

addition to the capability to overlay real cosmic data. The MicroBooNE full detector simulation softwares *LArSoft* [23] and *uboonecode* [24] were used to simulate the BNB neutrino charged current (CC) and neutral current (NC) interactions in the cryostat which contains the rectangularly shaped TPC active volume as introduced in section 1. The GENIE neutrino generator [25] and the Geant4 simulation toolkit [26, 27] were incorporated into the MicroBooNE simulation softwares.

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

A Ideal tracks – lines of charge depositions corresponding to minimum ionizing particles (MIPs) to demonstrate the intrinsic performance of the 3D imaging and the impacts from nonfunctional wires and the signal processing chain. See section 5.1.

B Neutrino only – full detector simulation of neutrino interactions without cosmic-ray muons – to demonstrate the performance of the 3D imaging on the complex topology of neutrino interaction final states. See section 5.2.

C Neutrino overlay – full detector simulation of neutrino interactions mixed with real cosmic data – to demonstrate the final performance after Wire-Cell 3D imaging, clustering, and matching. This sample is used to show the correctness and the neutrino efficiency after the matching. See section 5.3.

By comparing the 3D purity and the 3D completeness results between sample B and sample C, the impact from cosmic-ray muons and the performance of clustering and matching on the neutrino activity will be shown and discussed. In neutrino-only or neutrino-overlay samples, the ν_μ or ν_e energy spectra correspond to the BNB neutrino spectrum. Only the neutrino interactions with primary vertices in the TPC active volume are considered. The neutrino interactions outside the active volume are largely or completely invisible because the ionization electrons cannot drift and be collected by the wire planes. Evaluation of the matching performance on cosmic data or a full simulation of cosmic-ray muons is not specifically performed. The coincident in-beam cosmic activity is expected to be selected in this case, and they will be further rejected by dedicated cosmic-ray muon taggers.

5.1 Imaging performance on ideal tracks

About twenty one-meter-long ideal tracks (lines of charge depositions corresponding to MIPs) in each event were simulated in the MicroBooNE TPC. The angular distribution is uniform in 4π. The hit multiplicity on the anode plane per unit time is close to the real case, mimicking the numerous cosmic-ray muons traversing the MicroBooNE detector.

Three scenarios of the simulation were constructed to study the performance of the imaging and the impacts from the nonfunctional wires and the signal processing:

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

Dead + perfect SP: Nonfunctional wires are taken into account and perfect signal processing is applied.

Dead + real SP: Nonfunctional wires are taken into account and real signal processing is applied. For a prolonged track which leaves a long signal in each individual wire reading it out, the real signal processing may fail to reconstruct the charge for the induction plane wires/channels due to the bipolar signal cancellation. One can see more details in ref. [10]. This results in gaps in the 2D wire-versus-time views of the charge measurement as mentioned in section 3 and section 4.1.

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

Good – the tracks which are well reconstructed with at least 99% completeness.
Broken – the tracks which have gaps and are broken into pieces.
Absent – the tracks which completely fail to be reconstructed.
Ghost – the reconstructed tracks which have no overlap with any true tracks.

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

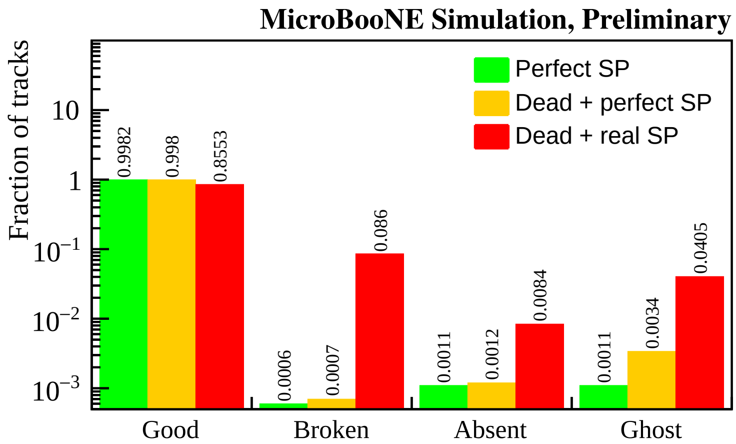

Figure 26: The fraction of (good, broken, absent, ghost) reconstructed tracks from the Wire-Cell 3D imaging for different scenarios. For good, broken, and ghost tracks, the fraction is weighted by their lengths and normalized to the total length of the true tracks. See text for the definitions of each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

lengths and normalized to the total length of the true tracks. In addition, the summation of the fractions of these three scenarios is very close to 100% which means the ghost tracks explain the missing parts of the broken tracks. The result of "Dead + Perfect SP" is roughly the same as the result of "Perfect SP" and almost all the tracks are well reconstructed with negligible ghost tracks. This shows that the nonfunctional wire issue is properly addressed in the 3D imaging and a ~97% active volume efficiency has been achieved. The impact of the nonfunctional wires on the quality of the 3D image will be further discussed in section 5.2 and section 5.3. The fraction of the ghost tracks in the scenario of "Dead + Perfect SP" is three times that of "Perfect SP" due to the presence of the nonfunctional wires but is still negligible. In the scenario of "Dead + Real SP", there is a dramatic increase of both broken tracks and ghost tracks. The broken tracks result from the gaps, which as above mentioned are attributed to the failure of the signal processing e.g. for the prolonged tracks. In this simulation of ideal tracks, there is a certain amount of prolonged tracks since they are generated with a 4π uniform angular distribution. This is not true for the real beam neutrino interactions, in which the final-state particles are mostly forward-going.

A large amount of ghost tracks appear when including both the nonfunctional wires and the gaps from the signal processing. The gaps indicate a failure in the charge matching across the multiple functional wire planes. The measured charge corresponding to the gap in other wire planes can probably be erroneously explained by the ghosts lying in the nonfunctional region. It is clear the ghost tracks are almost exclusively situated in the non-functional region as shown in figure 27.

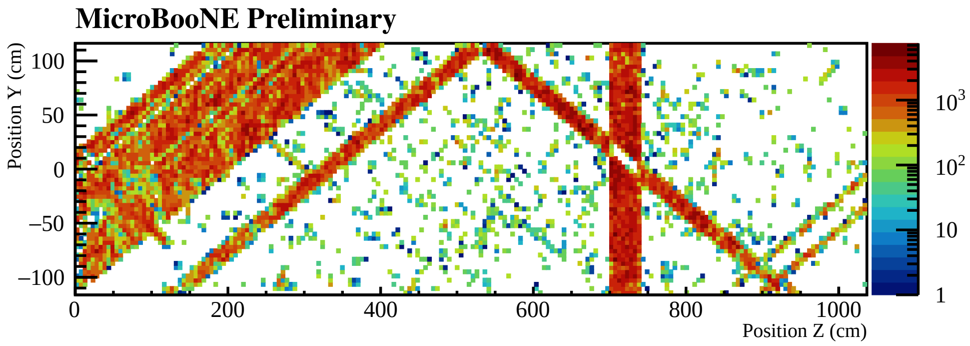

Figure 27: The position (Y/vertical versus Z/beam) distribution of the ghost tracks in the scenario of "Dead + real SP". Color scale (Z-axis value) represents the count of space points in ghost tracks. The bands correspond to the nonfunctional regions as shown in figure 4.

The purity for each event is calculated by dividing the total length of the non-ghost tracks by the total length of all the reconstructed tracks. The distribution of purity scores is presented in figure 28. For "Dead + real SP", 96.4% of the events have at least 90% purity. Figure 29 shows the distribution of completeness values for all true tracks. For the scenario of "Dead + real SP", 86.5% (and 93.0%) of the true tracks have at least 99% (and 80%) completeness. The extremely low completeness corresponds to the prolonged tracks, especially the tracks close to normal to the wire planes. One can see that signal processing is important to retain the good quality of the 3D imaging result. Overall, both the completeness and the purity in these cases are well understood.

5.2 Imaging performance on neutrino interactions

It is important to validate the robustness of the Wire-Cell 3D imaging using the neutrino interactions which are of great physics interest. The performance of the imaging on the neutrino interactions is presented in this section. Unlike the simulated ideal tracks in section 5.1, the topology of neutrino interaction final states is much more sophisticated than the single-track-like cosmic-ray muons. A neutrino interaction most often has multiple final-state particles, which are mostly forward-going along the beam direction. The neutrino-only samples without cosmic-ray muons were used. In order to evaluate the performance of the imaging, the charge-light matching including clustering was bypassed and all the 3D space points reconstructed in the imaging were taken as the neutrino activity. This is equivalent to perfect clustering and matching.

In neutrino interactions with argon nuclei, there are generally multiple final-state particles. On one hand, there is very limited phase space for the final-state particles to make tracks in the orientations we call prolonged or isochronous. These types of tracks may cause a failure in the signal processing chain and result in gaps along the track. On the other hand, with the complexity of the neutrino interactions, other failure modes may arise. Apart from that, a heavy ionization particle

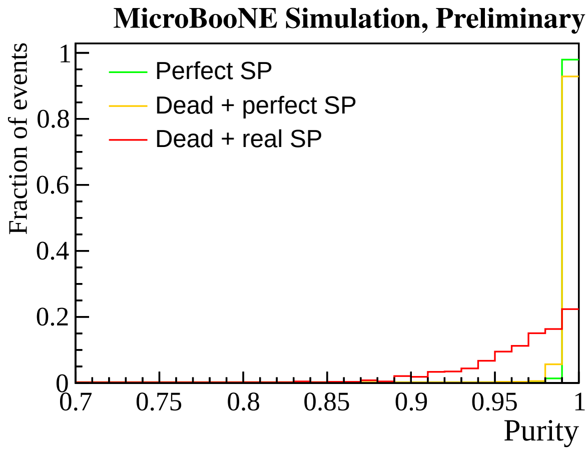

Figure 28: Purity of the reconstructed tracks in the event level for different scenarios. Ghosts dramatically arise with the presence of both nonfunctional wires and the real imperfect signal processing. The histograms are normalized separately for each category.

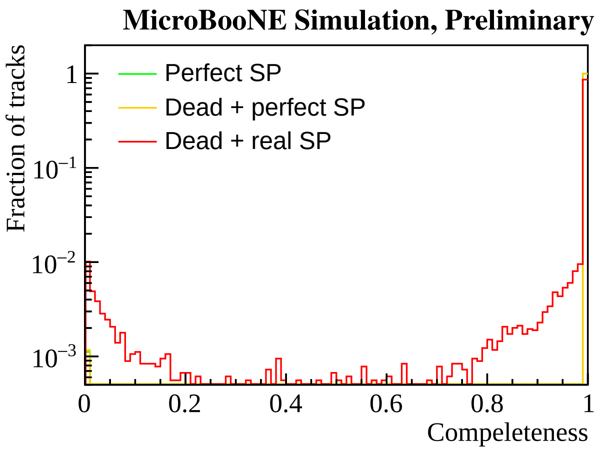

Figure 29: Completeness for each true track for different scenarios. The inefficiency of the signal processing for prolonged tracks leads to the extremely low completeness and the inefficiency does not seem to have a gradual change in the phase space. The histograms are normalized separately for each category.

(HIP) may avoid such failures in the signal processing chain since it has a significantly higher signal-to-noise ratio. Some of the particles like neutrons, γ's from pion decays, and primary or secondary electrons would pass through the liquid argon with low-visible-energy (<MeV) depositions via e.g. nuclear recoil, Compton scattering, and Bremsstrahlung radiation. These low energy depositions are likely to be suppressed due to the thresholding in the signal processing or removed in the imaging as they resemble the dot-like ghosts. As a result, the completeness will be overall biased and smeared to lower values with respect to the high-value portion of figure 29. The thresholding in the signal processing is primarily to suppress fake signals from noise fluctuations but is also related to the number of nonfunctional wires. A lower thresholding in the signal processing would

create more fake charge and can interplay with other true charge, leading to ghost tracks in the nonfunctional region.

Figure 30 shows two 2D snapshots of the 3D event displays for a $1\mu 1p$ ν_μ CC interaction and a $1e1p$ ν_e CC interaction, respectively. The red points represent the truth space points and the blue ones represent the reconstructed space points in the 3D imaging. The image of the reconstructed points are blurred because of the charge diffusion in the drift and the software filter smearing in the signal processing. Generally speaking, the reconstructed 3D image has both good completeness and purity compared to the truth 3D image in these two examples; even the short tracks belonging to the EM shower and some of the small dots are reconstructed.

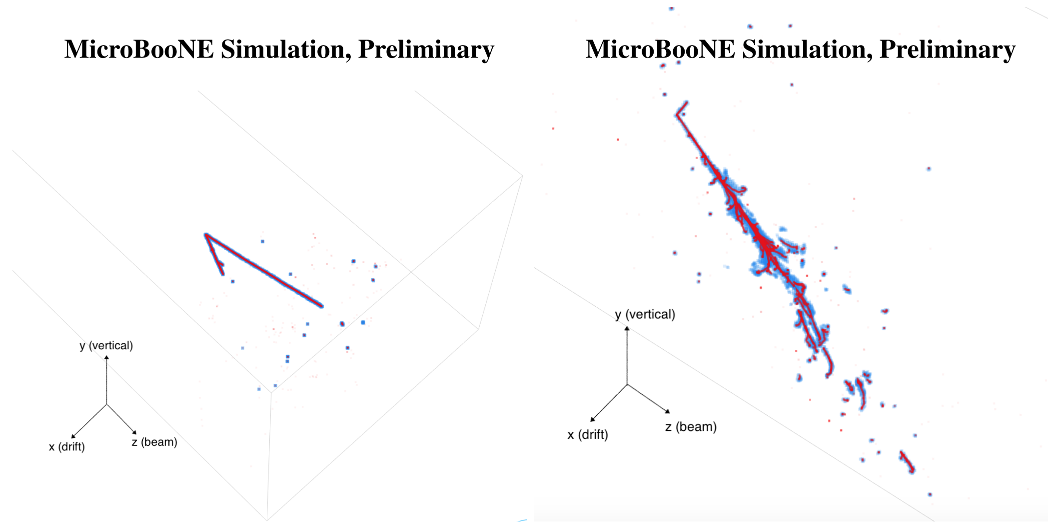

Figure 30: Left: $1\mu 1p$ ν_μ CC interaction. Right: $1e1p$ ν_e CC interaction. Blue: reconstructed 3D image. Red: truth trajectories. The voxel size and opacity are tuned for event display.

The quantitative evaluations of the 3D purity and the 3D completeness for BNB ν_μ CC, ν_e CC, and ν NC interactions in the TPC[11] are shown in figure 31. The completeness and purity are reasonably good. The results are summarized in table 3. The purity is high in the neutrino-only cases in which there are no cosmic-ray muons. For neutrino energy less than 400 MeV, the purity performance, e.g. the fraction of events with greater than 90% purity, is worsened by about 10%. This is due to the inefficiency of de-ghosting for low-visible-energy events. For example, the lower purity for the NC interactions mainly corresponds to the events with visible energy less than 100 MeV in which case the 3D image consists of a bunch of dot-like or very short tracks. The completeness has no extremely low values but is considerably smeared compared to the result in figure 29. This is attributed to the complex topologies of the final-state particles and a sizable amount of low energy deposition as explained in the above text. Since the primary electrons of ν_e CC interactions lead to EM showers through significant Bremsstrahlung radiation, the peaking completeness is obviously biased down to 90% or so. Such a bias is not critical to the track/shower identification and can be corrected in the shower energy reconstruction. The NC interactions do not have a primary particle leading to an EM shower but the hadronic processes generate protons, neutrons, or pions.

[11]The primary neutrino interaction vertices are within the TPC active volume.

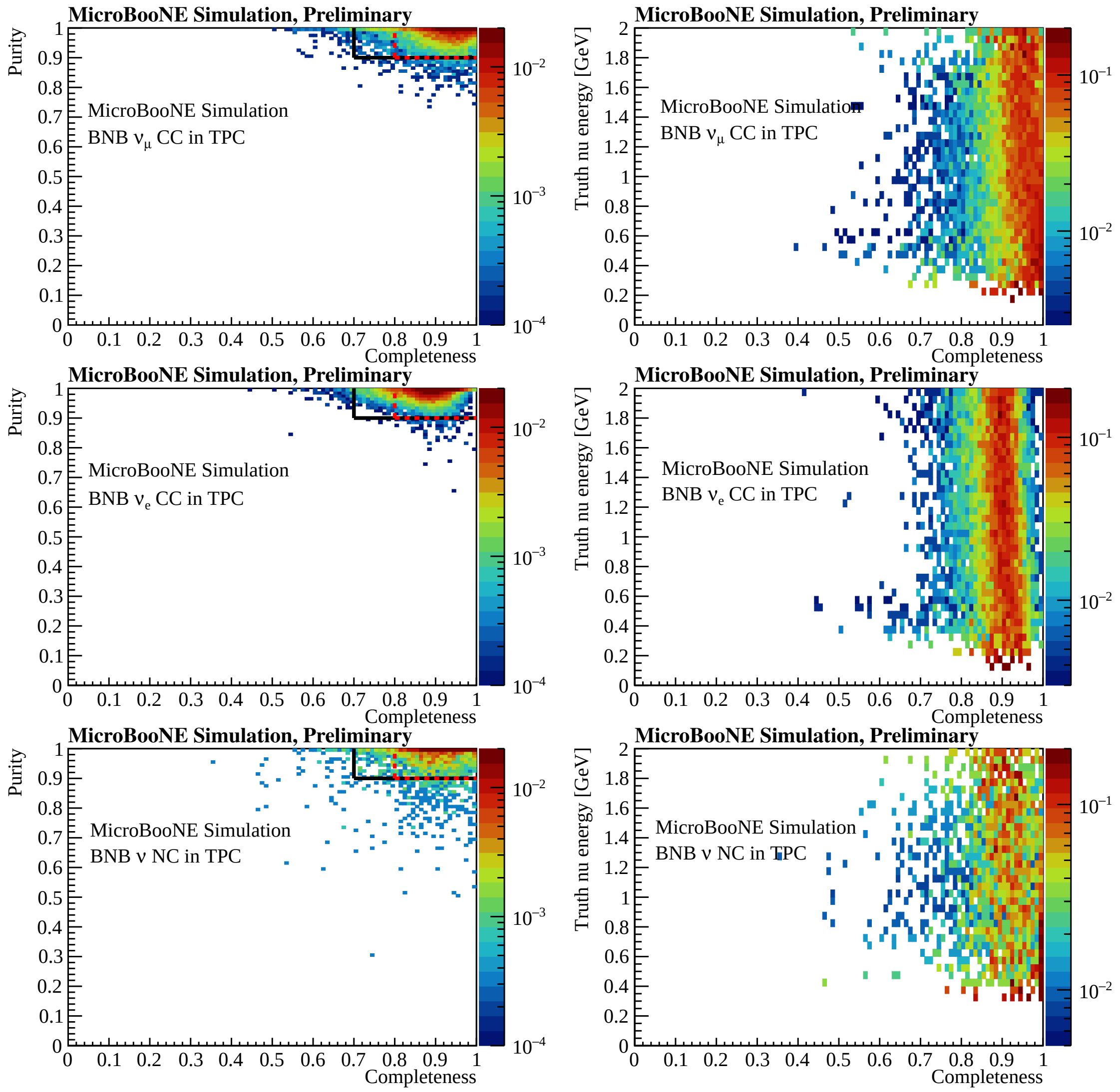

Figure 31: Completeness and purity of the imaging results for BNB ν_μ CC, ν_e CC, and NC interactions in the TPC. Color scale (Z-axis value) represents the fraction of events. There are no cosmic-ray muons in this simulation with ideal clustering and matching. Left: purity v.s. completeness for each neutrino interaction; the color scale (Z-axis value) represents the fraction of events. Right: the true neutrino energy v.s. the completeness; the distribution is normalized for each row of true neutrino energy bin; the color scale (Z-axis value) represents the fraction of events for each row. The integrated fraction of the events within the solid black and dashed red boxes can be found in table 3.

Table 3: Fraction of the events that correspond to the completeness and purity values within the 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 800 MeV. See text for more discussion on energy dependence. All neutrino interactions are simulated within the TPC active volume, without cosmic-ray muons.

	BNB ν_μ CC	BNB ν_e CC	BNB NC
Purity >90% and			
Completeness > 80%	88.6%	89.2%	80.7%
Completeness > 70%[a]	93.3%	96.7%	87.0%

[a] About 10% bias in the peaking completeness of ν_e CC and NC interactions.

These particles would travel with low energy deposition in the liquid argon as explained above, introducing a highly smeared completeness distribution. The “ 100%” completeness peak for low energy neutrino NC interactions as seen in the bottom right panel of figure 31 mainly corresponds to quasi-elastic scattering with a single low energy (short) proton emitted. The top right plots of these figures demonstrate a dependence of the completeness on the truth neutrino energy since a higher energy neutrino interaction is usually accompanied with a stronger hadronic production. The integrated completeness performance, e.g. the fraction of events with greater than 80% or 70% completeness, is uniform across different energy regions.

5.3 Final performance in the realistic case

In this section the neutrino-overlay samples were used to demonstrate the final performance of the Wire-Cell 3D imaging, clustering, and charge-light matching. Neutrino interactions were simulated and mixed with real cosmic data. The charge-light matching, including the clustering and light signal reconstruction, was applied on the 20-30 TPC clusters and 40-50 PMT flashes to select the in-beam neutrino activity, leading to a cosmic-ray muon reduction of approximately 30-fold. The efficiency and correctness of the matching and the quality of the 3D images of the selected neutrino candidates are key to the downstream reconstruction.

Figure 32 is a 2D snapshot of the 3D event display for a ν_μ CC interaction with the final states $1\mu 2\pi^0$ surrounded by a large amount of cosmic-ray muons traversing the TPC. Figure 33 is another example, showing one of the most challenging cases. The top panel shows the 3D images of all TPC activities including cosmic-ray muons and a neutrino interaction. The bottom panel shows the reconstructed 3D image of the selected (matched) in-beam TPC activities and the truth trajectories of the neutrino interaction final-state particles. In this example, there are an electron EM shower and two protons connected to the neutrino primary vertex. A π^0 is also created and decays into two γ's. The two γ's deposit energy through electrons from Compton scattering or electrons and positrons from pair production. A proper clustering of the two detached γ's is difficult considering the surrounding cosmic-ray muons. In this example, there is also a ghost track which crosses one proton track in the 2D snapshot but it is actually detached from the proton track in the 3D space.

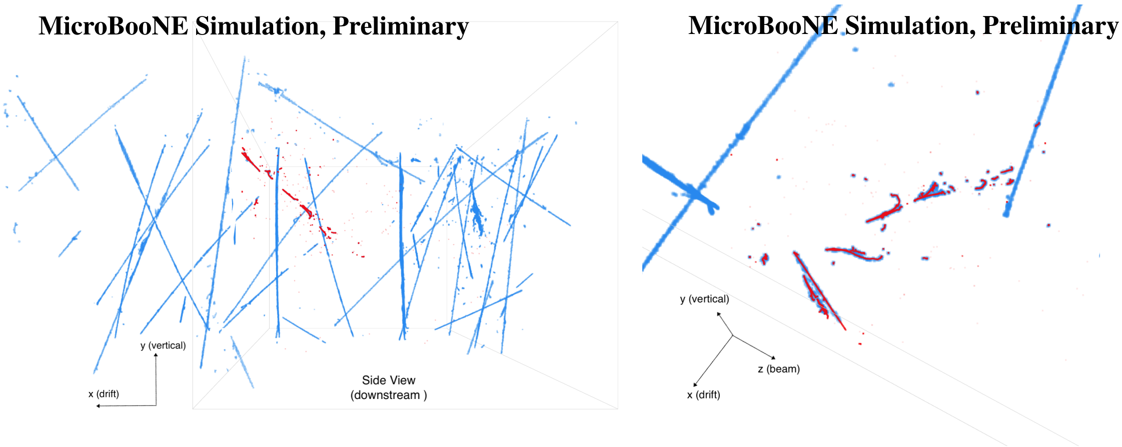

Figure 32: Event display of a $1\mu 2\pi^0$ ν_μ CC interaction with 4 evident final state EM showers. Left: side view of the full TPC readout. Right: zoom-in of the neutrino activity in TPC. Blue: reconstructed 3D image of all cosmic-ray muons and the neutrino interaction final states. Red: true trajectories of the final-state particles of the neutrino interaction.

It resides in the nonfunctional wire region and presumably originates from a piece of a cosmic-ray muon track.

Without any pattern recognition or topological reconstruction at this stage, the 3D completeness is a more critical metric than the 3D purity in the real case. There is little chance to fix the incompleteness issue in the downstream pattern recognition but the purity of the selected TPC activity can be further improved. For example, the ghost track in figure 33 can be removed by checking the directionality, which is irrelevant to the neutrino primary vertex, and/or by particle identification using dE/dx information in which case this ghost track will be regarded as a muon irrationally generated.

Figure 34 presents the 3D completeness and 3D purity of the selected TPC activity for BNB ν_μ CC, ν_e CC, and NC interactions, respectively. The correctness and the neutrino efficiency of the matching are evaluated and presented in the left panel of this figure. Incorrect matching corresponds to an extremely low completeness value as shown in the bottom left corner of the right panel of figure 34. One can see that the neutrino efficiency and the incorrectness of the matching are very high/low except for the low-visible-energy region. A neutrino interaction close to the TPC boundary very likely escapes the detector volume with a vast amount of energy (charge) that is then invisible to the wire readout plane. However, the light signal can still be collected if there is any charge deposition in the cryostat (liquid argon). The inconsistent TPC activity and PMT signal may result in an incorrect match or no match. The performance for NC interactions is not as good as that for CC interactions since the NC final-state particles are much more likely to introduce low visible energies in the TPC. It is supposed that the efficiency plus incorrectness is 100% for each bin but this is not true for the first low energy bin <50 MeV. This is because some of the events have no match as explained above. The overall incorrectness of the matching procedure are 4.6%, 3.8%, and 28.7% for BNB ν_μ CC, ν_e CC, and NC interactions, respectively. The results of completeness and purity are summarized in table 4.

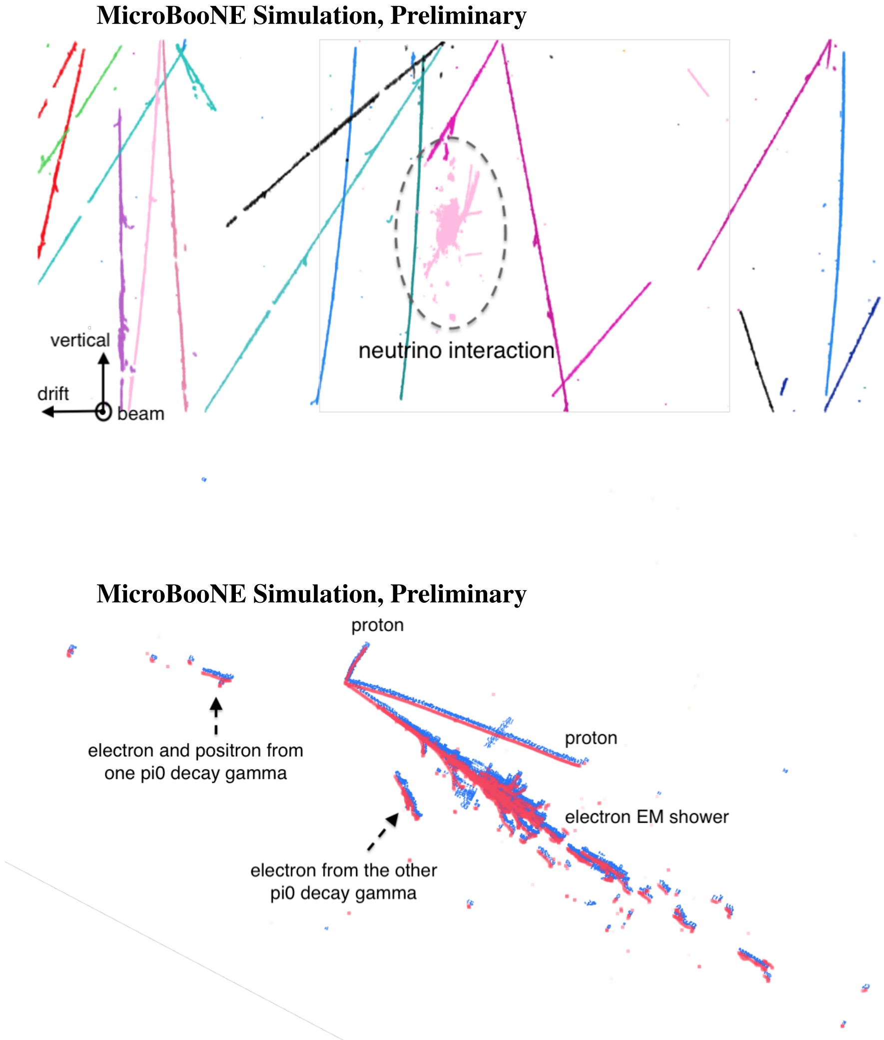

Figure 33: Event display of a $1e2p1\pi^0$ ν_e CC interaction. Left: side view of the full TPC readout; each cluster is labeled in one color. Right: the matching result – the in-beam flash matched TPC activity; the blue points are the reconstructed 3D space points and the red ones are the true space points corresponding to the neutrino interaction. The voxel size and opacity are tuned for event display.

Comparing figure 31 and figure 34, the degradation of the completeness and the purity can be attributed to the numerous cosmic-ray muons that traverse the detector. However, the overall performance is quite good considering the challenging task of selecting neutrino activity out of 20-30 cosmic-ray muons. Direct comparisons of the completeness and the purity are independently performed, and the distributions can be seen in figure 35. The scenarios of “neutrino-only” and “neutrino + cosmic” correspond to figure 31 and figure 34, respectively. In the scenario of “neutrino + cosmic”, the neutrino activity could not only be over-clustered with the cosmic activity (or its related ghost tracks) but also be under-clustered since part of the detached activity off the neutrino primary cluster may be grouped to the cosmic-ray muons. These two issues introduce a smearing

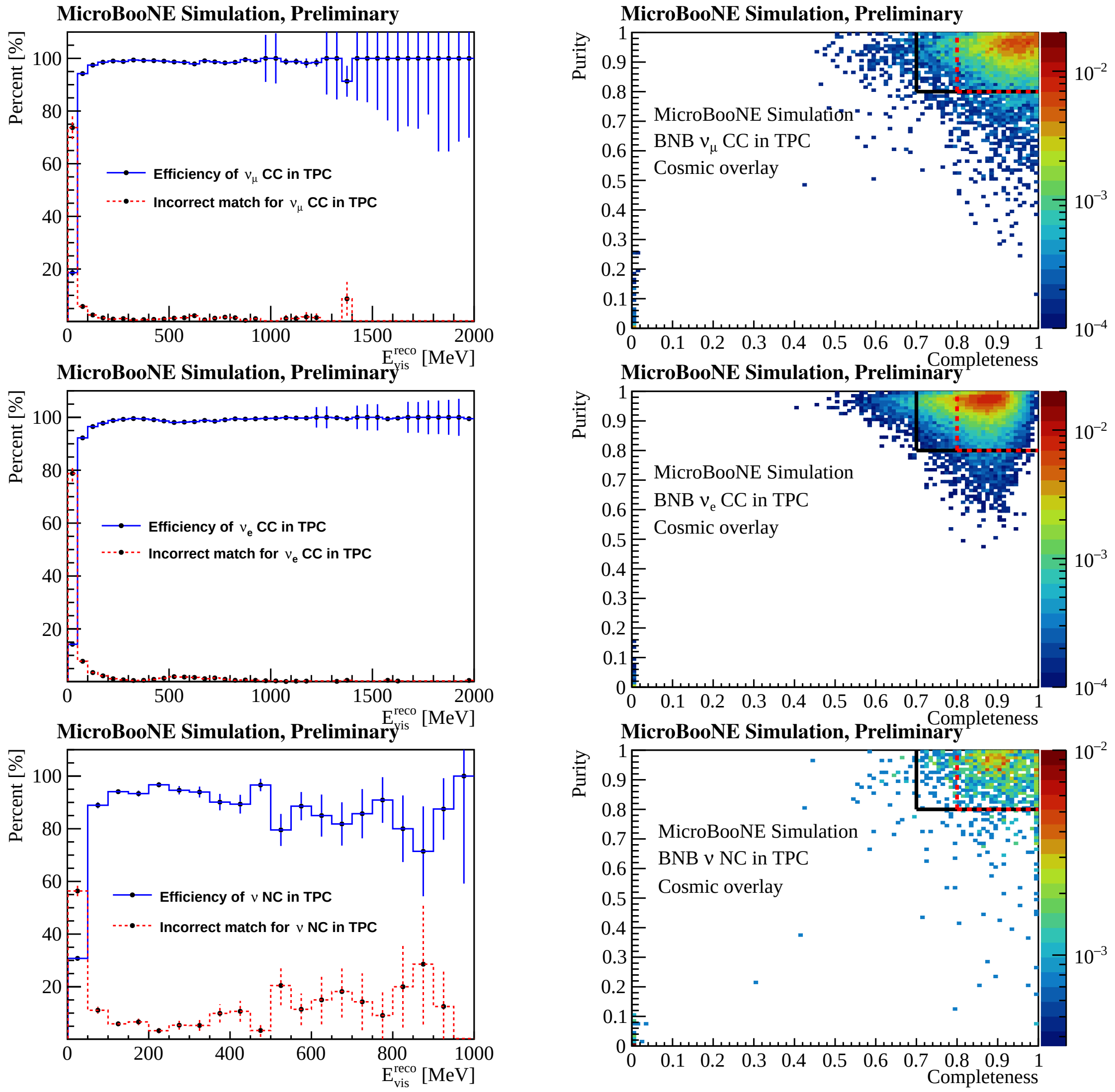

Figure 34: Imaging and matching performance for BNB ν_μ CC , ν_e CC, and NC interactions in the TPC. The simulation of neutrino interactions were overlaid with cosmic data. The clustering and matching were applied to select the neutrino activity. Left: efficiency and incorrectness of the matching as a function of the simply reconstructed visible energy (a simple conversion from the reconstructed visible charge using a constant conversion factor); binomial statistics were used to calculate the efficiency uncertainty while Poisson statistics (large error bars in the plot) were used where the efficiency is 100%, mainly for the low statistic bins. Right: purity v.s. completeness for each selected TPC activity; 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.

Table 4: Fraction of the events that correspond to the completeness and purity values within the boxes as shown in figure 34. These numbers are the overall performance for the integrated BNB neutrino flux which has an average neutrino energy of 800 MeV. All neutrino interactions are simulated within the TPC active volume, with cosmic data (beam-off) overlaid.

	BNB ν_μ CC	BNB ν_e CC	BNB NC
Purity >80% and			
Completeness > 80%	73.0%	67.7%	56.0%
Completeness > 70%[a]	80.2%	83.4%	66.5%

[a] About 10% bias in the peaking completeness of ν_e CC and NC interactions.

of both the completeness and the purity distributions. The typical numbers of the completeness and the purity for different scenarios and interaction types are summarized in table 5.

Table 5: Summary of the typical numbers of the completeness and the purity corresponding to the distributions as shown in figure 35. 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	Neutrino + cosmic	84.5% (93.4%)	74.9% (92.5%)	73.1% (87.9%)
>80% (>70%[a])	Neutrino	92.8% (97.8%)	90.8% (98.5%)	90.2% (97.4%)
Purity >80%	Neutrino + cosmic	84.4%	89.2%	72.6%
	Neutrino	99.4%	99.8%	97.1%

[a] About 10% bias in the peaking completeness of ν_e CC and NC interactions.

About 85% of events have at least 80% completeness for BNB ν_μ CC interactions. About 90% of events have at least 70% completeness for BNB ν_e CC or NC interactions. The degradation of the purity in the scenario of "neutrino + cosmic" is more severe than the degradation of completeness as expected, as the surrounding cosmic muons bring in a large amount of neutrino-unrelated TPC activities. As explained regarding the ghost tracks in figure 33, the completeness is more critical in the imaging and matching stage. The purity can be further improved in the downstream reconstruction chain. In the optimization of the clustering and matching, the completeness thus has a greater weight. In reality, the final purity performance corresponding to the scenario of "neutrino + cosmic" is still reasonably good and 80%-90% events have at least 80% purity for CC interactions.

In summary, the quantitative evaluations of the Wire-Cell 3D imaging, clustering, and matching have been presented in this section. These techniques result in a high performance selection of the neutrino activities in the MicroBooNE LArTPC. Meanwhile, a cosmic-ray population reduction of approximately 30-fold has been achieved. The correctness and the efficiency of the selected neutrino activities are very high except for the low-visible-energy events less than 50 MeV. The

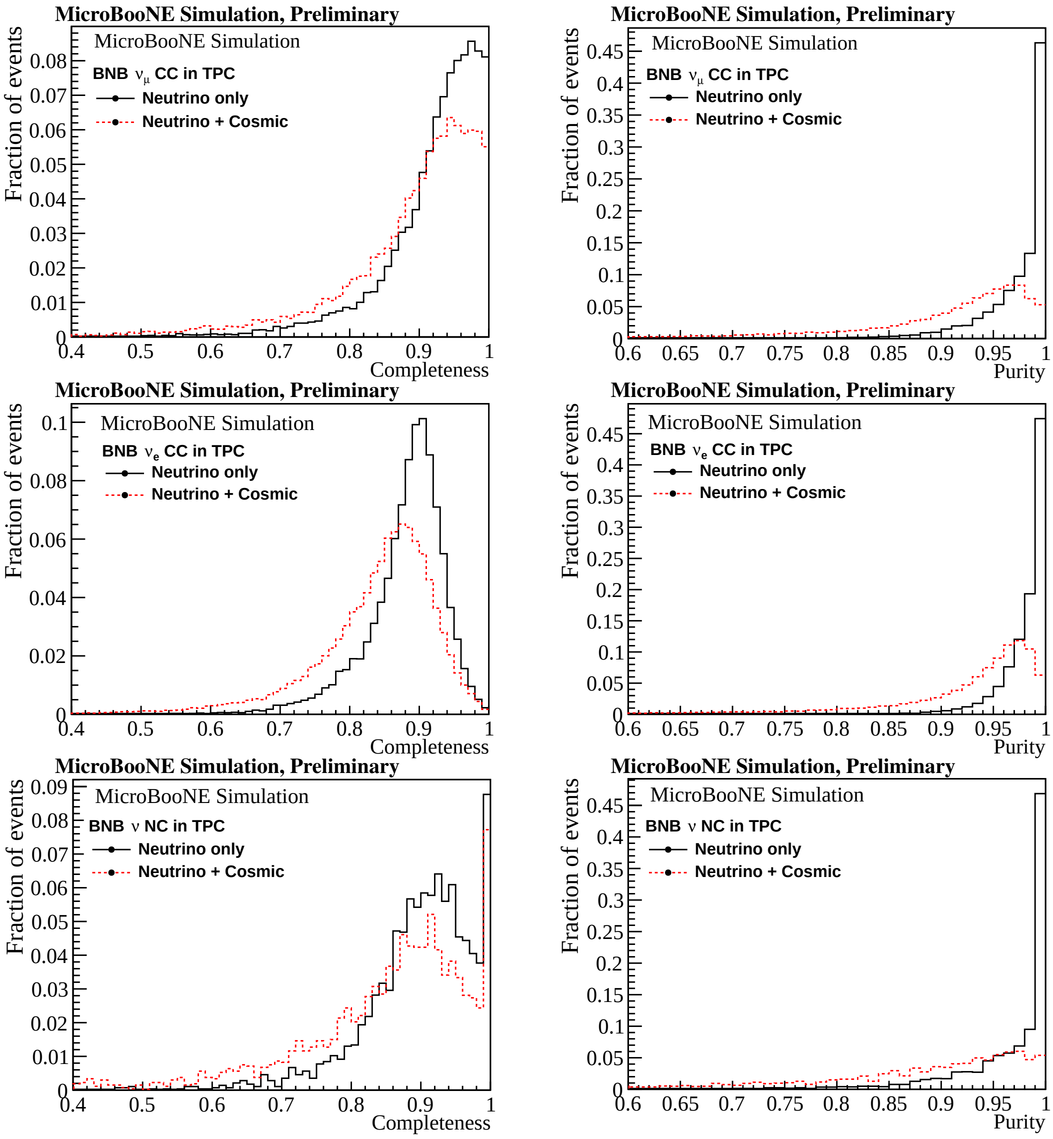

Figure 35: Independent comparisons of completeness and purity distributions for two scenarios of "neutrino-only" and "neutrino + cosmic". The charge-light matching and clustering have to be applied in the scenario of "neutrino + cosmic" to select the neutrino activity. Top: BNB ν_μ CC interactions in the TPC. Middle: BNB ν_e CC interactions in the TPC. Bottom: BNB ν NC interactions in the TPC; the "∼100%" completeness peak 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.

quality (3D completeness and 3D purity) of the 3D images of the selected in-beam neutrino activities are reasonably good considering the wire readout ambiguity, the nonfunctional wires, the inefficient signal processing, and the numerous cosmic-ray muons.

6 Summary and Discussion

This paper describes the principle and algorithms of the Wire-Cell 3D imaging, clustering, and many-to-many charge-light matching applied in the MicroBooNE LArTPC. The 3D imaging tomographically reconstructs the 3D image of the ionization electrons using the fundamental information in charge, time, and geometry. Other characteristics of the LArTPC physics activity such as sparsity, positivity, and proximity are utilized as additional constraints to improve the imaging performance. The many-to-many charge-light matching with 3D clustering and light signal reconstruction is developed to pair the TPC clusters and PMT flashes identifying the neutrino activity among numerous cosmic-ray muons. Several realistic issues, e.g. the nonfunctional wires, the gaps due to inefficient signal processing or noise filtering, detached neutrino activity, and coincidently connected clusters, are properly addressed. Using the latest MicroBooNE detector simulation, we evaluated the realistic performance of the described reconstruction techniques. In addition to the event displays shown in section 3 and section 4, a few more event displays of the selected in-beam TPC activity from both beam neutrino and cosmic data (coincident in-beam TPC activity) are shown in Appendix A.

Beyond what has been evaluated, there are some other limitations in the 3D imaging as one can see in the event displays in this paper. For example, prolonged tracks, which often develop gaps in the signal processing stage, cannot be entirely fixed via the bridging of the gaps as implemented. It is the same for the isochronous tracks that often develop gaps due to the MicroBooNE coherent noise filter. Subsequent pattern recognition, i.e. particle-level clustering, may further address this problem. Isochronous tracks present another common problem for the LArTPC 3D imaging, as the wire readout ambiguity (hit multiplicity) is dramatically increased in the time slice containing the isochronous track. In Wire-Cell 3D imaging, this issue is mitigated by introducing tiling, however, the blobs of the isochronous track are significantly broadened, leading to a much worse 2D spatial resolution in the Y-Z plane view. Some improvement of spatial resolution can be achieved via trajectory fitting in a later stage.

Generally speaking, the 3D event reconstruction techniques as presented in this paper are adequately accurate and efficient, and successfully select the neutrino activity. About 95% of the neutrino CC interactions in the TPC active volume are selected, with a 30-fold reduction of non-beam-coincident cosmic-ray muons. Good 3D completeness and 3D purity of the resulting 3D image of the selected neutrino activity have been achieved. Greater than 80% of the selected neutrino CC interactions have a clean and intact reconstructed 3D image of at least 70% completeness and 80% purity. These techniques will benefit the downstream pattern recognition and neutrino selection, and it is believed they are important steps towards realizing the projected capability of single-phase LArTPCs. For those data events without neutrino interactions, these techniques can also provide a solid foundation (high quality 3D image) to reject the coincident in-beam cosmic-ray muons. In particular, the Wire-Cell based neutrino selection and analyses take full advantage of these tools to further reject cosmic muons and select neutrinos [28] and demonstrate a very promising high efficiency & high purity neutrino selection in LArTPCs. Other analyses using techniques such as deep learning with convolutional neural networks [29, 30] and Pandora multi-algorithm pattern recognition [31] can also benefit from the outcome of the Wire-Cell 3D reconstruction tools, as the Wire-Cell 3D image of the neutrino activity remains largely intact with the surrounding cosmic activity cleaned up.

Acknowledgments

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

A Appendix: selected event displays from MicroBooNE data

Event displays from the MicroBooNE beam-on and beam-off data are presented in this appendix. The links to "Bee" [32] – a WebGL based Wire-Cell 3D event display tool – are also attached. In the following event displays, the red solid circles in the front view represent the measured PMT signals for the in-beam flash. The green solid circles represent the predicted PMT signals based on the matched 3D image. The area of the solid circle is proportional to the number of photoelectrons.

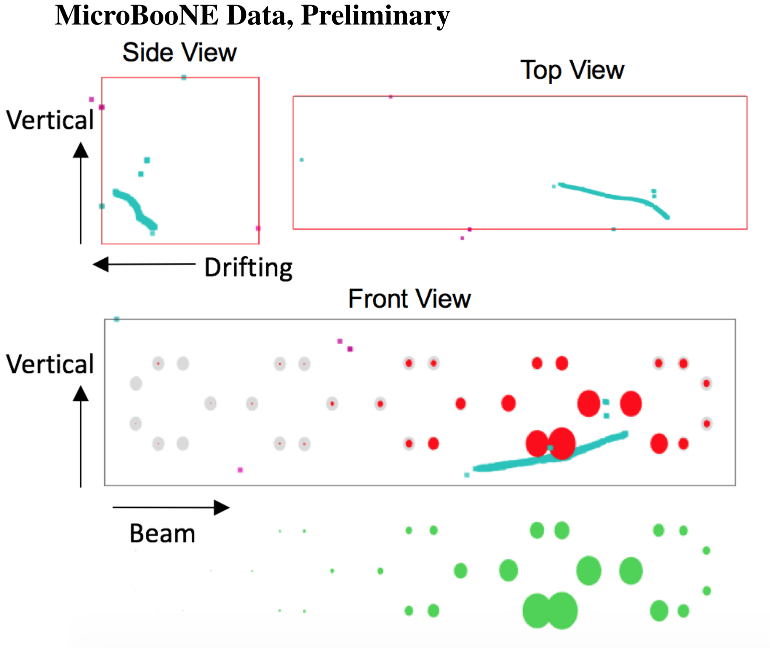

Figure 36: Example of a fully contained ν_μ CC candidate.

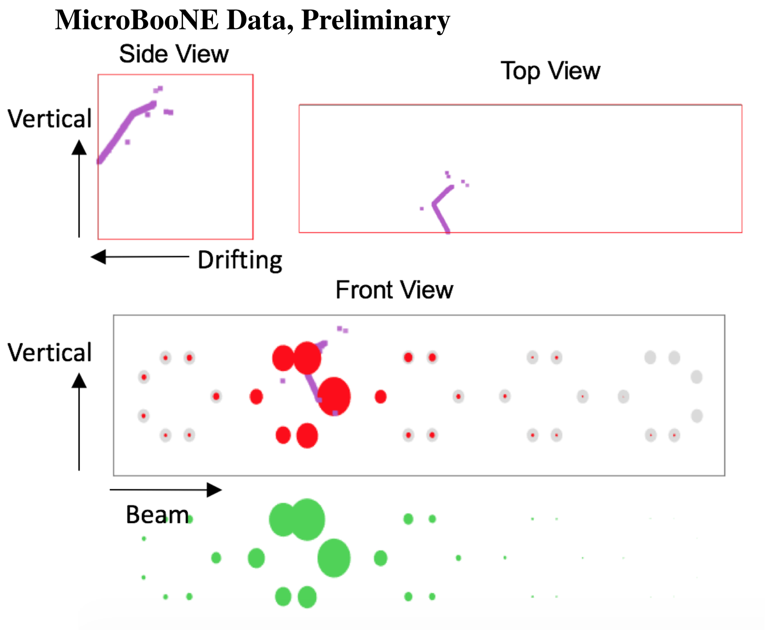

Figure 37: Example of a partially contained ν_μ CC candidate.

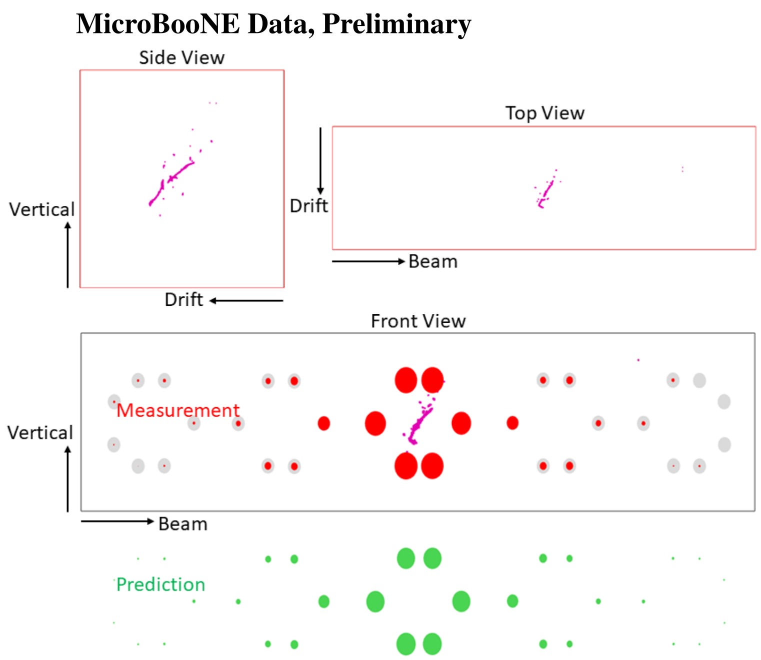

Figure 38: Example of a ν_e CC candidate.

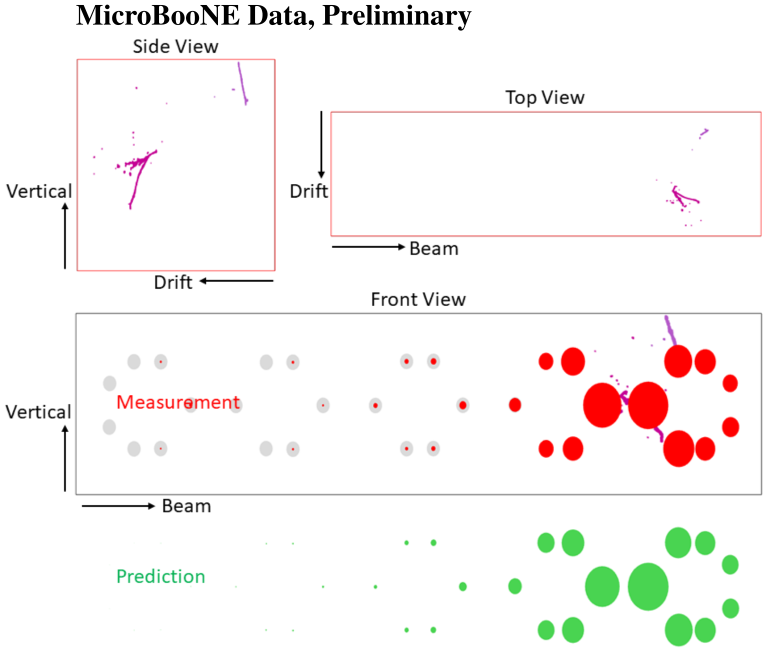

Figure 39: Example of a ν_e CC candidate over-clustered with a cathode-side stopped muon.

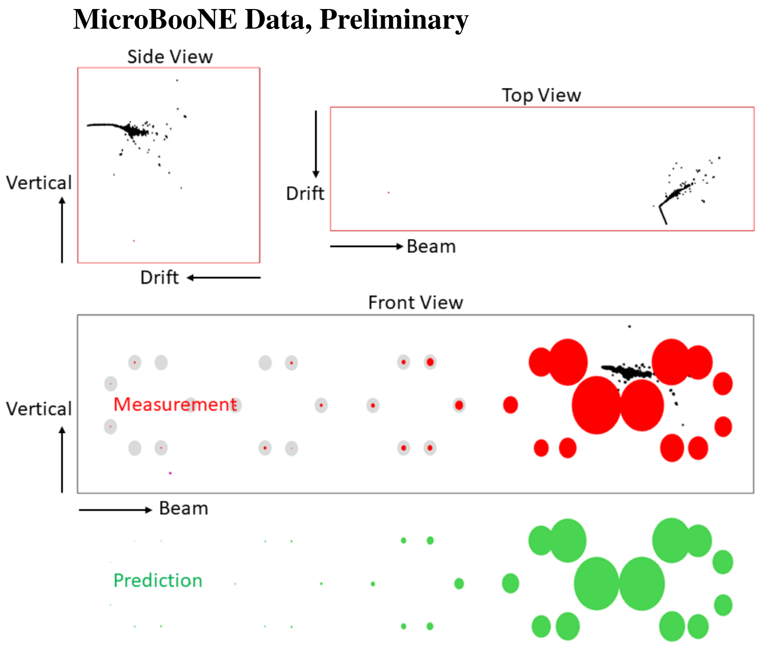

Figure 40: Example of a $1e1p$ ν_e CC candidate.

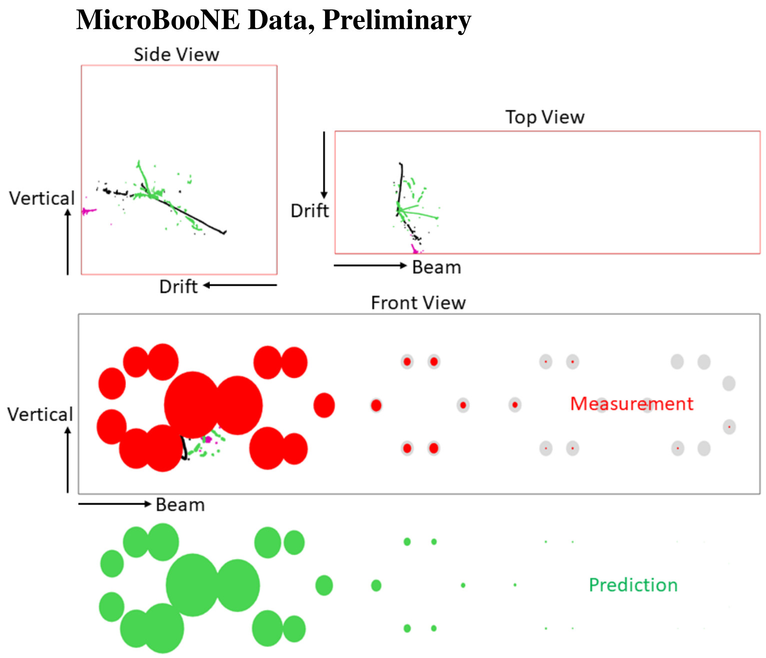

Figure 41: Example of a multiple π^0 event.

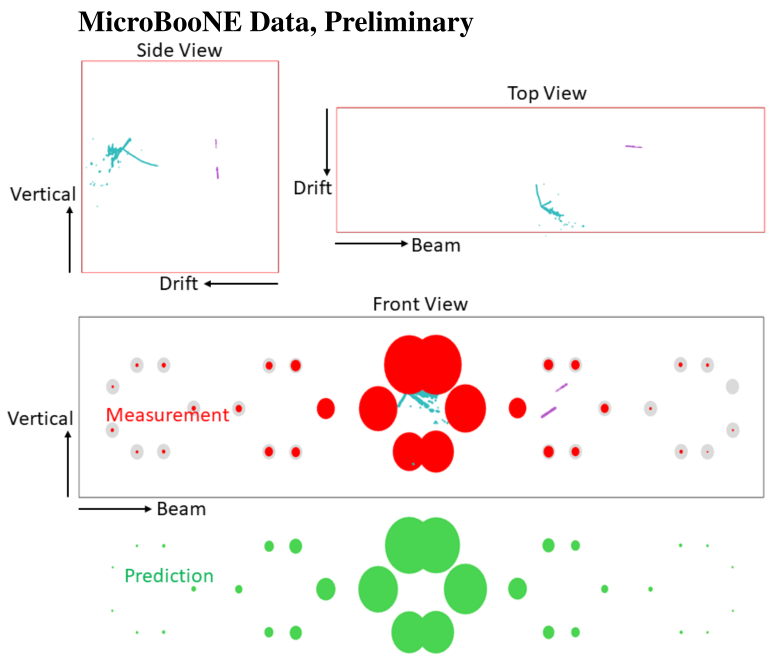

Figure 42: Example of two EM showers (e^+e^-?) connected to the primary vertex.

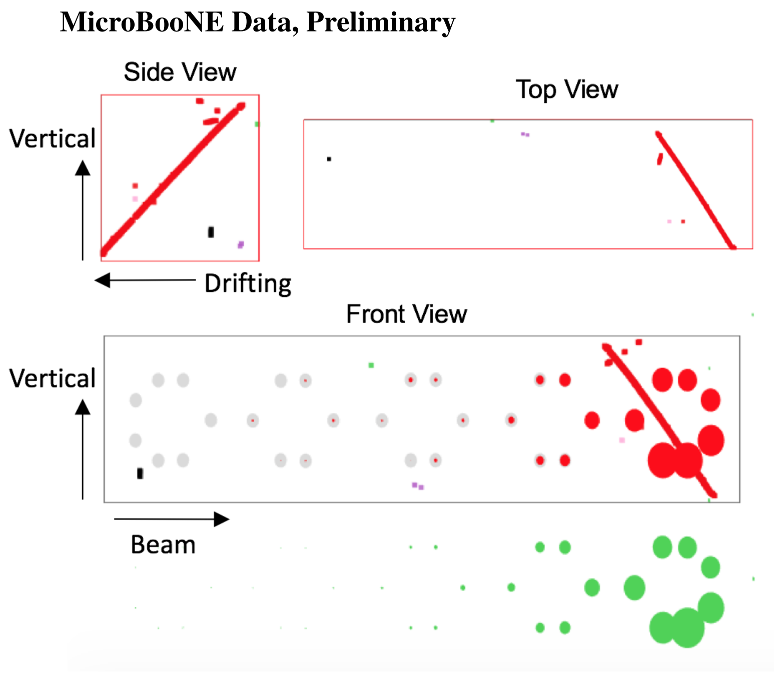

Figure 43: Example of a through-going muon from beam-off data. The upper end in the side view is on the space charge boundary.

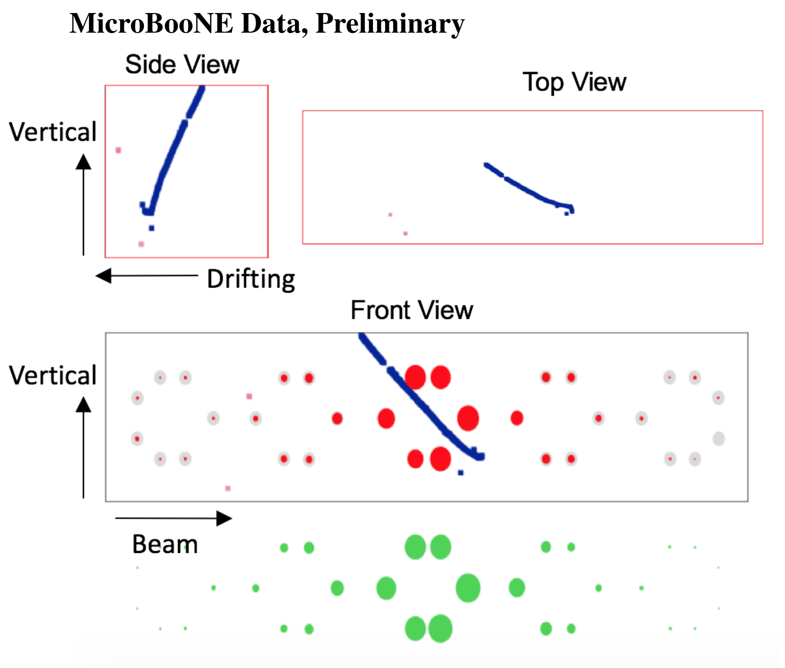

Figure 44: Example of a stopped muon with a Michel electron from beam-off data.

References

[1] C. Rubbia. *The liquid argon time projection chamber: A new concept for neutrino detector*, . CERN-EP/77-08, 1977.

[2] H. H. Chen, P. E. Condon, B. C. Barish and F. J. Sciulli. *A Neutrino detector sensitive to rare processes. I. A Study of neutrino electron reactions*, . FERMILAB-PROPOSAL-0496, 1976.

[3] W. J. Willis and V. Radeka, *Liquid argon ionization chambers as total absorption detector*, *Nucl. Instrum. Meth.* **120** (1974) 221.

[4] D. R. Nygren, *The Time Projection Chamber: A New 4 pi Detector for Charged Particles*, *eConf* **C740805** (1974) 58.

[5] MicroBooNE collaboration, R. Acciarri et al., *Design and Construction of the MicroBooNE Detector*, *JINST* **12** (2017) P02017, [arXiv: 1612.05824].

[6] P. A. Machado, O. Palamara and D. W. Schmitz, *The Short-Baseline Neutrino Program at Fermilab*, *Ann. Rev. Nucl. Part. Sci.* **69** (2019) , [arXiv:1903.04608].

[7] DUNE collaboration, B. Abi et al., *The DUNE Far Detector Interim Design Report Volume 1: Physics, Technology and Strategies*, arXiv:1807.10334.

[8] X. Qian, C. Zhang, B. Viren and M. Diwan, *Three-dimensional Imaging for Large LArTPCs*, *JINST* **13** (2018) P05032, [arXiv:1803.04850].

[9] MicroBooNE collaboration, R. Acciarri et al., *Noise Characterization and Filtering in the MicroBooNE Liquid Argon TPC*, *JINST* **12** (2017) P08003, [arXiv:1705.07341].

[10] MicroBooNE collaboration, C. Adams et al., *Ionization Electron Signal Processing in Single Phase LArTPCs I. Algorithm Description and Quantitative Evaluation with MicroBooNE Simulation*, arXiv:1802.08709.

[11] MicroBooNE collaboration, C. Adams et al., *Ionization electron signal processing in single phase LArTPCs. Part II. Data/simulation comparison and performance in MicroBooNE*, *JINST* **13** (2018) P07007, [arXiv:1804.02583].

[12] MiniBooNE collaboration, A. Aguilar-Arevalo et al., *The Neutrino Flux prediction at MiniBooNE*, *Phys. Rev. D* **79** (2009) 072002, [arXiv:0806.1449].

[13] MicroBooNE collaboration, R. Acciarri et al., *Construction and Assembly of the Wire Planes for the MicroBooNE Time Projection Chamber*, *JINST* **12** (2017) T03003, [arXiv: 1609.06169].

[14] MicroBooNE collaboration, J. Conrad, B. J. P. Jones, Z. Moss, T. Strauss and M. Toups, *The Photomultiplier Tube Calibration System of the MicroBooNE Experiment*, *JINST* **10** (2015) T06001, [arXiv:1502.04159].

[15] E. J. Candès, J. K. Romberg and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, *Communications on Pure and Applied Mathematics* **59 (8)** (2006) 1207–1223, [arXiv:math/0503066].

[16] J. Friedman, T. Hastie and R. Tibshirani, *Regularization paths for generalized linear models via coordinate descent*, *Journal of Statistical Software* **33 (1)** (2010) 1–22.

[17] Wire-Cell git repository for sparse signal regression, https://github.com/BNLIF/wire-cell-ress/.

[18] MicroBooNE collaboration, D. Caratelli, *Neutrino identification with scintillation light in MicroBooNE*, *JINST* **15** (2020) C03023.

[19] J. L. Blanco and P. K. Rai, “nanoflann: a C++ header-only fork of FLANN, a library for nearest neighbor (NN) wih kd-trees.” https://github.com/jlblancoc/nanoflann, 2014.

[20] P. Hough, “Method and means for recognizing complex patterns.” US Patent 3069654, 1962.

[21] quickhull. https://github.com/akuukka/quickhull, 2017.

[22] Ben Jones, Sterile Neutrinos in Cold Climates, FERMILAB-THESIS-2015-17.

[23] LArSoft release v08.05.00.10, https://cdcvs.fnal.gov/redmine/projects/larsoft/wiki/ReleaseNotes08050010.

[24] Uboonecode release v08.00.00.28, https://cdcvs.fnal.gov/redmine/projects/uboonecode/wiki/ReleaseNotes08000028.

[25] C. Andreopoulos, C. Barry, S. Dytman, H. Gallagher, T. Golan, R. Hatcher et al., *The GENIE Neutrino Monte Carlo Generator: Physics and User Manual*, arXiv:1510.05494.

[26] GEANT4 collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, *Nucl. Instrum. Meth.* **A506** (2003) 250–303.

[27] J. Allison et al., *Recent developments in Geant4*, *Nucl. Instrum. Meth.* **A835** (2016) 186–225.

[28] Cosmic Ray Background Rejection with Wire-Cell LArTPC Event Reconstruction in MicroBooNE, MicroBooNE collaboration, MicroBooNE-NOTE-1804-PUB.

[29] MicroBooNE collaboration, R. Acciarri et al., *Convolutional neural networks applied to neutrino events in a liquid argon time projection chamber*, *JINST* **12** (2017) P03011, [arXiv:1611.05531].

[30] MicroBooNE collaboration, C. Adams et al., *Deep neural network for pixel-level electromagnetic particle identification in the MicroBooNE liquid argon time projection chamber*, *Phys. Rev.* **D99** (2019) 092001, [arXiv:1808.07269].

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

[32] Wire-Cell Bee 3D event display, https://github.com/WireCell/wire-cell-bee.