Tomographic Event Reconstruction with MicroBooNE Data

— Application with a Semi-automated $\nu_\mu CC$ Selection —

June 2, 2018

The MicroBooNE Collaboration

Contents

1 INTRODUCTION

Event reconstruction is one of the most challenging tasks in analyzing the data from current and future massive single-phase liquid argon time projection chambers (LArTPCs). The performance of event reconstruction is critical to the potency of the eventual physics result. In this note, we describe a novel event reconstruction method, known as *Wire-Cell tomographic imaging* [1], applied to MicroBooNE data. MicroBooNE is an 85 metric ton single-phase LArTPC and the first of a trio of LArTPCs in the Short Baseline Neutrino (SBN) program. MicroBooNE's principal aims are to understand the MiniBooNE low energy excess, measure neutrino-argon interaction cross sections, and improve the understanding and capability of single-phase LArTPCs through detector R&D.

Wire-Cell tomographic imaging capitalizes on fundamental LArTPC detector information — time, charge, and geometry — to resolve degeneracies inherent in a wire readout. Via the imaging technique, LArTPC signals are reconstructed directly in 3D. We describe various clustering methods developed to spatially group the result of tomographic imaging, 3D charge objects, into physical tracks and showers. In addition, we describe the processing of the scintillation light signal to reconstruct flashes, coincident light signals from multiple PMTs.

A second novel method, the *many-to-many matching of TPC clusters to PMT flashes*, is delineated in this note. The aim of this matching technique is to identify neutrino candidate activity. This method emphasizes the interplay between the scintillation light and ionization charge signals to select beam-spill coincident neutrino candidate TPC activity. In doing so, event classification is vastly simplified owing to the disambiguation of beam-spill identified activity from all other TPC object backgrounds that may occur during an event. Rejecting background TPC objects is especially critical for surface-based single-phase LArTPCs, for which cosmic rays are a significant background to be mitigated for any neutrino physics analysis. The robust reconstruction of 3D charge enabled by tomographic imaging facilitates the power of this neutrino event selection technique.

This note focuses on two novel techniques, *tomographic imaging* and *many-to-many TPC cluster-PMT flash matching*, applied to MicroBooNE data. A detailed description of the principles behind and implementation of tomographic imaging is described in Sec. 2.1. A discussion on the realistic issues and considerations necessary to capitalize on the imaging results to reconstruct track and shower objects via clustering is provided in Sec. 2.2. The process of pairing TPC clusters with PMT flashes (flash construction described in Sec. 2.3) is described in a detailed description of the matching algorithm in Sec. 3.2.

We performed a semi-automated ν_μ CC selection using a small sample of MicroBooNE data to coarsely evaluate the efficacy of these tools. The culmination of the sequential ν_μ CC selection steps — step #1 automated PMT filter (Sec. 3.1), step #2 automated many-to-many TPC cluster-PMT flash matching (Sec. 3.2), and step #3 single-click hand-scan TPC filter (Sec. 3.3). — is summarized in Sec. 3.4. The tools described here are an important step towards realizing the projected capability of high efficiency and high purity neutrino selection with single-phase LArTPCs.

2 EVENT RECONSTRUCTION

In order to fully explain the philosophy behind tomographic event reconstruction, we first review the fundamental information provided by single-phase LArTPCs with wire readout:

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

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

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

Since the charge information provided is the total number of ionization electrons along the wire, the wire readout has an intrinsic ambiguity for charge position determination. Event reconstruction is further complicated by the absence of geometric information. Relative to a pixelated readout with O(n^2) pixels, the O(n) wires afforded by a wire readout scheme results in a loss of information.

However, additional information can be attained by analyzing and processing the fundamental detector information:

(iv) **Sparsity** - the signal (charge hits on each wire plane) is generally expected to be sparse for any physics activity within a short time

(v) **Positivity** - the charge drifting through or collected by a wire plane can only be positive, as in drifting towards the anode wire plane as opposed to leaving the anode wire plane; this is required from the arrangement of external fields

(vi) **Connectivity** - subset of topological information, owing that charged particles are expected to leave continuous energy depositions leading to connected detector hits as the LArTPC detector is fully active

(vii) **Topology** - trajectory type of ionization electrons (long or short, straight or curved, track or electromagnetic shower) following traversal of charged particles across the active LAr medium

We use the above enumerated information to implement a novel event reconstruction technique applied to MicroBooNE data called *Wire-Cell tomographic imaging*. The philosophy of this approach is to use as much information as possible to reconstruct the original charged particle information, relying first on the more robust information (e.g. time, geometry, charge, sparsity, positivity) and in later stages using less robust, derived information (e.g. connectivity, topology). This arrangement delays the introduction of errors in overall event reconstruction due to delaying the introduction of less robust information util later stages. Decoupled from the pattern recognition, the goal of Wire-Cell tomographic imaging is to reconstruct the most exclusive event image independent of the event topology.

2.1 Tomographic Imaging

The single-phase LArTPC with wire readout is a natural application of the *tomography* imaging technique. Since the ionization electrons are uniformly drifting to the anode plane, the wire planes are equivalent to the moving detector in tomography, time is the cross section/slice, one wire plane readout is one view/1D projection, and the charge is the integrated charge/matter effect along the wire/projection direction. As a result, tomographic imaging for single-phase LArTPC is well-motivated and has gradually matured into a new event reconstruction paradigm.

2.1.1 Tiling with 2 or 3 wire planes

The deconvolved TPC signal, which represents the number of ionization electrons going through or collected by the wires, are rebinned into 2 μs time slices (four 0.5 μs time ticks). Detailed procedures on excess noise filtering and TPC signal processing, which converts the raw ADC back into the distribution of ionization electrons through the deconvolution technique, are documented in Ref. [2] and Refs. [3, 4], respectively. During the signal deconvolution [5, 6], a software filter is required to suppress the noise amplification at high frequency, which effectively results in a cut-off frequency (~0.25 MHz) for the signal. According to the

Nyquist theorem [7], if a continuous signal contains no power at frequencies above a certain value (called the Nyquist frequency), the signal can be fully sampled without information loss at a sampling rate higher than twice the Nyquist frequency. The choice of rebinning four time ticks (2 μs) into a single time slice is made to reduce the amount of data and to avoid information loss as guided by the Nyquist theorem. In addition, rebinning is supposed to increase the signal-to-noise ratio and further reduce the loss of inoization charge due to the application of a noise RMS threshold in the signal processing [3].

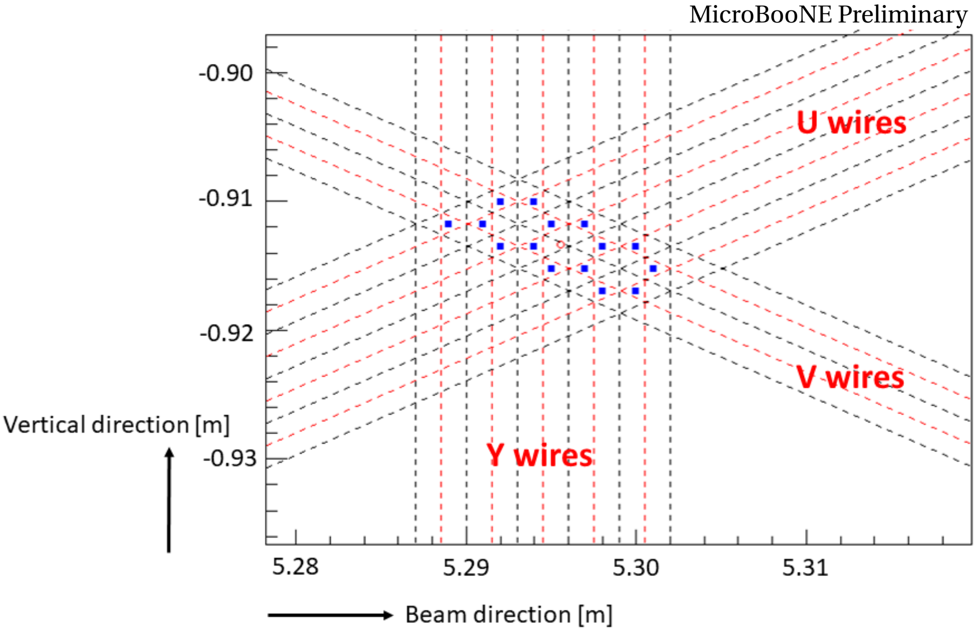


Figure 1: Illustration of blob construction. Each wire is represented by dotted red lines, with the wire boundary represented by dotted black lines. The overlapping region among three crossing wires from different planes is a triangle in the MicroBooNE geometry and is referred to as a *cell*. The cell/triangle is shaped by the dotted black lines (other than the small triangle shaped by the red lines within) with a blue point at the center. The blob pictured consists of 15 cells.

Once the fired wires (or channels) on a particular time slice are identified, the consecutive fired wires are merged together to form a *grouped wire*. The common region going through three grouped wires (one from each wire plane) is referred to as a *blob*. Figure 1 shows an example of blob construction. In this example, four U-plane, two V-plane, and five Y-plane wires are fired. The overlapping region among three single wires (one from each wire plane) has a triangular shape, and is referred to as a *cell*. This blob contains 15 cells.

There are a few critical reasons motivating the merging of consecutive fired wires (i.e. merging cells into a blob). First, when a group of consecutive wires is fired in all three wire planes, the number of cells, $O(n^2)$, increases much faster than the number of fired wires, $O(n)$. The significant increase in the number of cells essentially represents the information loss due to the wire readout. This loss of information is difficult to completely recover at this stage when no topology information is readily available. The merging technique temporarily conceals the loss of information and leaves it for later reconstruction steps. Second, as to be elaborated on in Sec. 2.1.2, matching of the reconstructed ionization charge from the two induction and one collection wire planes is a powerful constraint to reduce ambiguities. For the collection wires, the reconstructed ionization charge is due mostly from ionization electrons collected by these wires, essentially equivalent to the ionization electrons passing through the region of the collection wire. However, this is not the case for the reconstructed ionization charge for the induction wire planes. As discussed in Ref. [3], when a point-like electron cloud passes between two adjacent induction wires, both wires will receive the induction current and the magnitude of the induction current depends on the distance between the electron cloud and the wire (smaller magnitude at further distance). Therefore, even when the ionization electron is passing through an induction wire very closely, the reconstructed charge would be smeared in nearby wires. In addition, the software filter in the wire dimension serves as a smearing of reconstructed charge to nearby wires. Therefore, compared to the reconstructed charge in each individual induction wire, the sum of the charge measured by a group of induction wires is actually a better measurement of the overall charge going through the induction wire plane. Finally, the choice of the blob unit also limits the usage of computer memory; the $O(n^2)$ cells are essentially represented by three groups of $O(n)$ wires.

The described procedure above requires all three wire planes to participate in the tiling process (blob creation), which is powerful in reducing ambiguities. This is because the chance for three randomly fired wires (one from each wire plane) to cross at the same location is much smaller than that for two randomly fired wires (in the case of only two wire planes). However, this choice is not realistic when analyzing MicroBooNE data, where about 10% of channels are unresponsive due to various reasons [2]. The top panel and bottom panel in Figure 2 shows the areas where at least one and two wire planes are unresponsive, respectively. If we require all three wire planes to participate in the tiling, the active area is reduced to only about 70%. On the other hand, if we relax the requirement to only two wire planes (with the third-plane wires in the known unresponsive region), the efficiency increases significantly to about 97%.

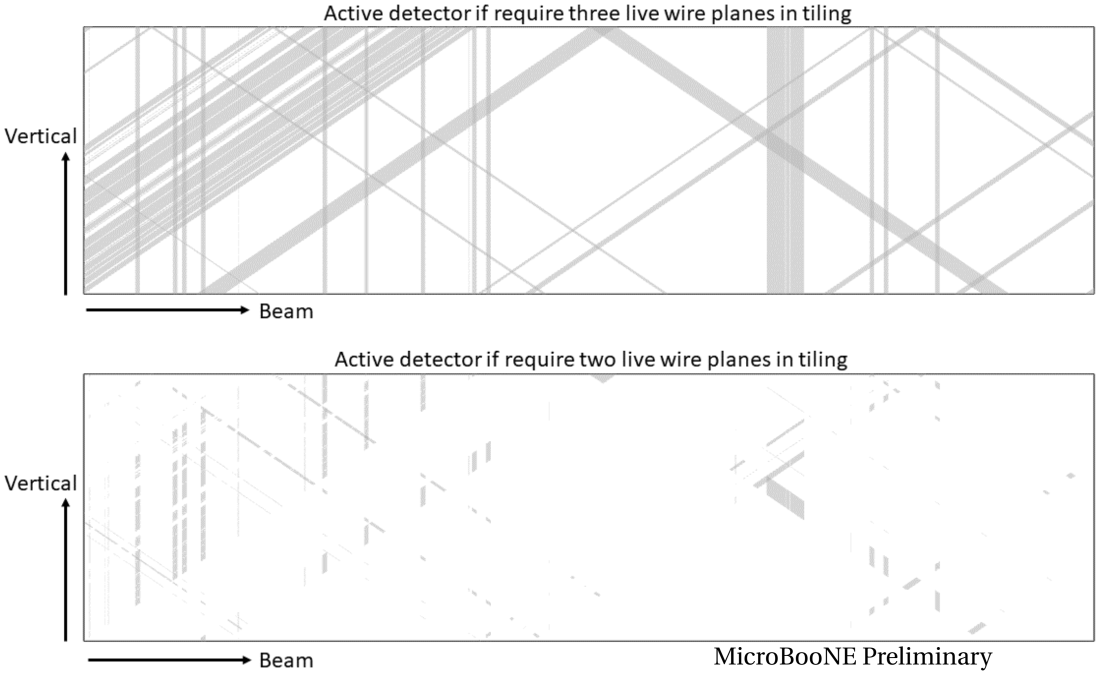


Figure 2: Illustration of the unresponsive regions (gray) in the Y-Z plane requiring three live (top) and two live (bottom) wire planes in tiling. Y (vertical) represents the vertical direction. Z (horizontal) represents the neutrino beam direction. In the top panel, the gray areas comprise about 30% of the entire Y-Z area - at least one of the wire planes is not functional in these regions. The large vertical unresponsive wire region is at z~700 cm. In the bottom panel, the gray areas comprise about 3% of the entire Y-Z area - at least two of the wire planes are not functional in these regions. `https://www.phy.bnl.gov/wire-cell/bee/set/342b1227-74fa-4ea6-88c6-b4f69cf42e15/event/0/`. `https://www.phy.bnl.gov/wire-cell/bee/set/71a3e15b-eed0-47f3-a70e-1d068f96cdc3/event/0/`.

Figure 3 shows the comparison between the 3-plane tiling (left panel) and the 2-plane (means two or more planes) tiling (right panel) at a particular time slice. More blobs are created when only two grouped fired wires are required in the tiling procedure. The efficiency is increased from about 70% to about 97%. At the same time, more fake blobs, formed by random coincidence of grouped fired wires, also show up. The underlying assumption in the 2-plane tiling is that the third-plane unresponsive wires are always fired.

We should emphasize that through the above tiling procedure, a set of blobs, aiming at high efficiency, is created. Figure 4 shows the reconstructed 3D images based on 3-plane (top panel) and 2-plane (bottom panel) tiling. A large amount of ghosts (in general representing the superfluous charge hits due to incorrect explanation of the measured charge by non-local unresponsive wires) are apparent with the 2-plane tiling procedure, though the efficiency is higher. Note that the Figure 4 looks quite busy due to the size of the 3D space points in

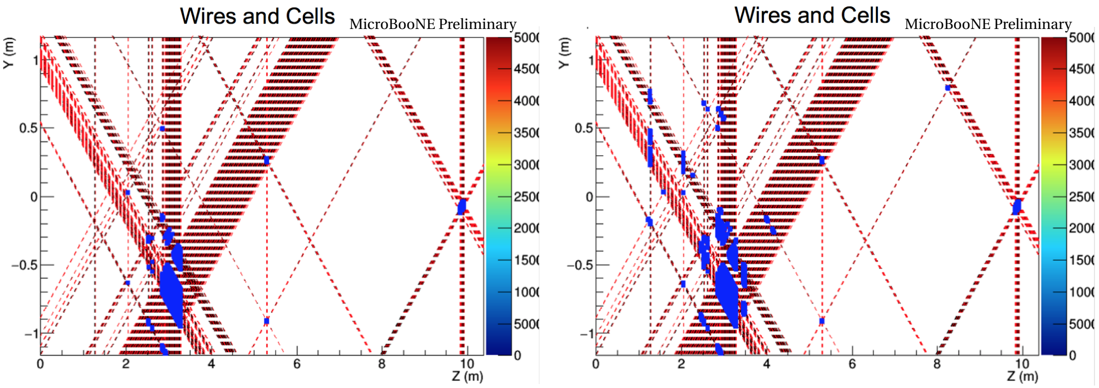


Figure 3: Comparison of 3-plane tiling (left) and 2-plane tiling (right). Fired wires are plotted. The connected cells (blue) form blobs. More blobs are constructed when only two out of three grouped fired wires are required. The efficiency is increased to about 97% (2-plane tiling) from about 70% (3-plane tiling).

the display. In the following sections, we explain various implemented methods to reduce ambiguities, i.e. removing intrinsic fake blobs due to the wire readout ambiguity as well as the ghosts originating from the unresponsive wires.

2.1.2 Charge matching and solving

Although voxel-based reconstruction is a natural choice in a pixel-readout detector, it is hindered in LArTPCs due to the existence of ambiguities created by the limited number of wire planes. Figure 5 illustrates such a situation with a simplified example consisting of only two wire planes. For simplicity, we only consider wires and cells here. For actual application, we use the merged wires and blobs (cluster of cells) instead. Three cells (H1, H5, H6) are “hit” by the distribution of charge passing through the wire planes in a single time slice. Due to the wire readout, five wires (u1, u2, v1, v2, v3) are fired at this time slice. This naturally leads to 6 possible hit cells, including 3 fake ones (H2, H3, H4) in addition to the 3 true ones (H1, H5, H6). Such ambiguities cannot be resolved with time and geometry information alone.

However, given that any charge deposited inside a cell is measured independently by the associated wires, additional charge equations can be constructed. For the example in Fig. 5,

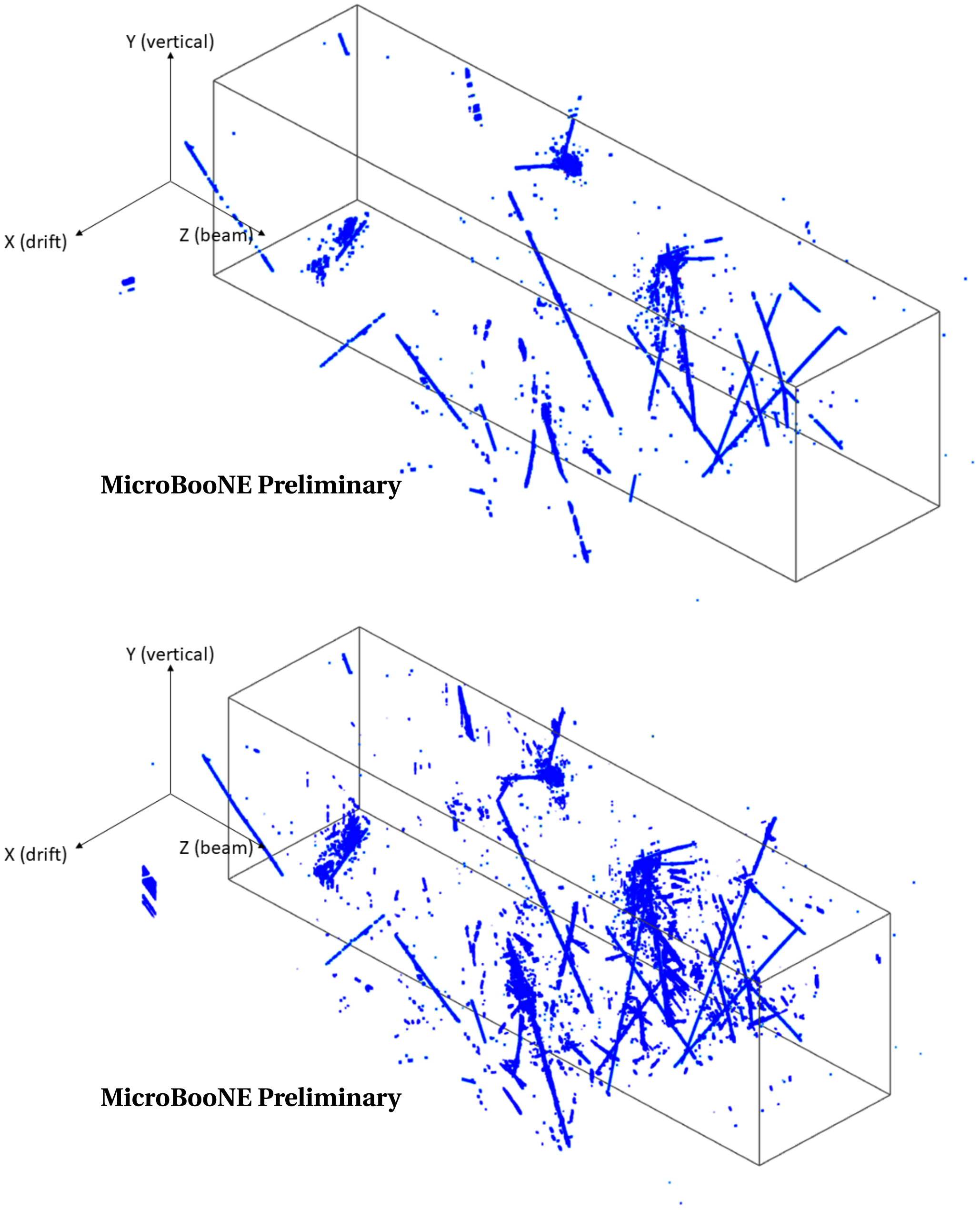


Figure 4: Comparison of 3-plane (top) and 2-plane (bottom) tiling. The volume efficiency for the 3-plane and 2-plane tiling is around 70% and 97%, respectively. The increase in efficiency in the 2-plane tiling is accompanying with large amount ghost activities. Only time and geometry information are used in the tiling. A large amount of fake hits and ghosts are in presence due to the wire readout/geometric ambiguity and unresponsive wires, respectively. `https://www.phy.bnl.gov/wire-cell/bee/set/8773a7c1-4829-498d-a14d-baf841a8bea1/event/0/`. `https://www.phy.bnl.gov/wire-cell/bee/set/018443de-6be6-4bbe-bcdb-c3a15af247af/event/0/`.

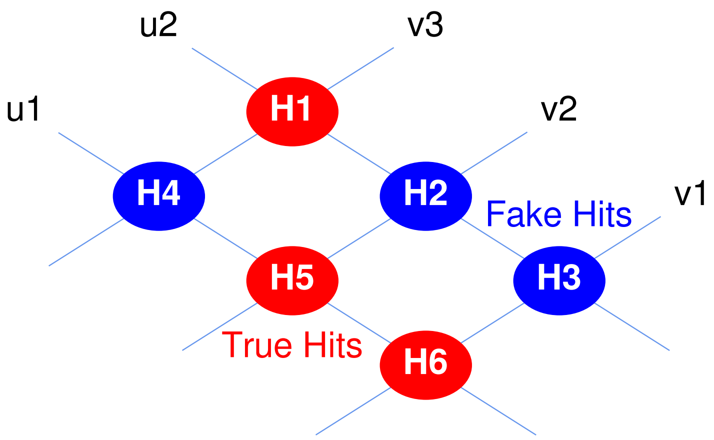


Figure 5: Illustration of the hit ambiguity caused by the wire readout in a simplified two-wire-plane example.

we have:

$$\begin{pmatrix} \mathrm{u1} \\ \mathrm{u2} \\ \mathrm{v1} \\ \mathrm{v2} \\ \mathrm{v3} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \mathrm{H1} \\ \mathrm{H2} \\ \mathrm{H3} \\ \mathrm{H4} \\ \mathrm{H5} \\ \mathrm{H6} \end{pmatrix}, \tag{1}$$

or, more generally, in matrix form:

$$y = Ax, \tag{2}$$

where y is a vector of charge measurements spanning the fired wires from all planes, x is a vector of expected charge in each possible hit cell to be solved, and A is the biadjacency matrix connecting wires and cells, which is determined solely by the wire geometry. In an ideal case, upon solving Eq. 2, the true hit cells will have the deduced charges equal to their true charges, while the fake hit cells will have the deduced charges equal to zero.

The charge equation Eq. 2 does not take into account the uncertainties associated with the charge measurement (e.g. noise contribution or imperfection in signal processing procedures). Considering the charge uncertainties, we can improve Eq. 2 by constructing a χ^2

function:

$$\chi^2 \quad = \quad (y - Ax)^T \cdot V^{-1} \cdot (y - Ax) \tag{3}$$

where V is the covariance matrix representing the uncertainty of the measured charge on each wire. The best solution can be found through minimizing a chi-square function Eq. 3 through the first derivative:

$$\frac{\partial \chi^2}{\partial x} = 0 \rightarrow A^T V^{-1} \left(y - A \cdot x \right) + \left(y - A \cdot x \right)^T V^{-1} A = 0, \tag{4}$$

and the solution can be written as:

$$x = \left(A^T \cdot V^{-1} \cdot A \right)^{-1} \cdot A^T \cdot V^{-1} \cdot y. \tag{5}$$

Therefore, when the matrix $M = A^T V^{-1} A$ is invertible, the best-fit charges of hit cells can be derived directly using Eq. 5. The true hit cells will have the expected charges *close to* the true charges, while the fake hit cells will have the expected charges *close to* zero.

In most occasions, however, the charge equations Eq. 2 are under-determined and do not have a unique solution. Similarly, the χ^2 function constructed through Eq. 3 does not have a unique minimum location, representing a severe loss of information due to the wire-readout system. In such cases, one would find that the matrix $M = A^T V^{-1} A$ in Eq. 5 is non-invertible. For the example in Fig. 5, we have the matrix M:

$$M = \begin{pmatrix} 2 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ 1 & 1 & 2 & 0 & 0 & 1 \\ 1 & 0 & 0 & 2 & 1 & 1 \\ 0 & 1 & 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 & 1 & 2 \end{pmatrix}. \tag{6}$$

The six eigenvalues are 5, 3, 2, 2, 0, 0. The number of zero eigenvalues are two, which prevents the direct inversion of the M. This is easy to understand as there are only five measurements, but we have six number of unknowns in this example. The number of independent measurements is not sufficient to determine all the unknowns.

When the matrix M cannot be inverted, additional constraints by considering the characteristics of typical physics events are necessary to solve the charge equation. One such constraint is that most of the elements in the solution x should be zero and so the solution should be *sparse*. In the example of Fig. 5, the sparse condition implies that out of the 6 possible cells, many are fake. Then, Eq. 2 can be transformed into a constrained linear problem:

$$\text{minimize} \, ||x||_0, \quad \text{subject to: } y = Ax, \tag{7}$$

where $||x||_0$ is the ℓ_0-norm of x, which represents the number of non-zero elements in x. The notation $||\cdot||_p$ defines the ℓ_p-norm of a vector such that $||x||_p = (\sum_i |x_i|^p)^{1/p}$. In other words, we seek the most sparse or simplest solution that explains the measurements. In the rare cases where the sparse condition fails to represent reality, the ability to resolve ambiguities from the wire-readout system would be intrinsically limited by the hardware.

Since the ℓ_0-norm is non-convex, the optimization problem of Eq. 7 is NP-hard [8]. Nonetheless, a practical procedure through combinatorial trials is described as follows:

1. Calculate the number of zero eigenvalues n_0 of the matrix $M = A^T V^{-1} A$.
2. Enumerate the $C_N^{n_0}$ possible reductions of Eq. 2 produced by removing n_0 elements from x of original size N and updating the A and V matrices accordingly.
3. For each reduction, redo step 1. If n_0 is now zero, calculate a solution based on Eq. 5, and record the χ^2 value. Otherwise, go to the next reduction.
4. Accept the solution with the minimal χ^2 value.

As an example, consider Fig. 5 and assume an identity covariance matrix. The six eigenvalues of 6×6 matrix M are 5, 3, 2, 2, 0 and 0. Thus there are two zero eigenvalues. In order to reach a unique solution, two out of the six possible hit cells need to be removed. The best solution is obtained by comparing χ^2 values from the $C_6^2 = 15$ combinations. One can see that this procedure, although sound in principle, becomes computationally intractable when the hit multiplicity is more than a few tens[1] (e.g. $C_{35}^{10} \approx 1.8 \times 10^8$).

Interestingly, the technique of *compressed sensing* [9], originally proposed to recover stable signal from incomplete and inaccurate measurements, solves the above computational issues. Compressed sensing has wide applications in the fields of electrical engineering [10], medicine and biology [11], applied statistics [12], etc. Its application in high-energy and

---

[1] In general, this is the real case even after merging cells to big blobs.

nuclear physics experiments is relatively rare, but has great potential. The key concept is to approximate the NP-hard[2] ℓ_0 problem of Eq. 7 with a ℓ_1 problem:

$$\text{minimize}\, ||x||_1, \quad \text{subject to: } y = Ax, \tag{8}$$

which retains the desired sparsity. The proof of this approximation can be found in Ref. [9]. Equivalently, the χ^2 function defined in

Eq. 3 can be replaced by a ℓ_1-regularized χ^2 [13, 14] as follows:

$$\chi^2 = ||y' - A'x||_2^2 + \lambda ||x||_1, \tag{9}$$

where the vector y and x are then pre-normalized through $V^{-1} = Q^T Q$ (Cholesky decomposition), $y' = Q \cdot y$, and $A' = Q \cdot A$, and λ regulates the strength of the ℓ_1-norm. Since the ℓ_1-regularized χ^2function is convex, fast minimization can now be achieved through algorithms such as coordinate descent [15]. In addition, the *non-negativity* physical constraint (or positivity constraint), defined as $x_i \geq 0$, can also be incorporated during minimization.

As explained in Sec. 2.1.1, we use blob and grouped wires instead of cell and wire, respectively. In this case, Eq. 3 can be updated to:

$$\chi^2 = \left(C \cdot y - G \cdot B\right)^T V_{By}^{-1} \left(C \cdot y - G \cdot B\right), \tag{10}$$

with B representing the vector of unknown charge on all blobs. C is a matrix representing how single wires are merged together as grouped wires. G is the updated geometry matrix which connects the blobs to the grouped wires. $V_{By} \equiv C \cdot V_y \cdot C^T$ is the covariance matrix describing the uncertainty in the grouped wire charges. Similarly, the application of compressed sensing is updated $\lambda ||B||_1$ instead of $\lambda ||x||_1$.

The motivation of the solving stage is to remove fake blobs. The primary new information used is the charge information, as each wire plane should measure the same amount of ionization charge for the charge depositions in a 3D voxel (cell/blob by time slice). A chi-square function is constructed by comparing the measured charge on the grouped wires with the predicted charge based on the blobs. Through the application of compressed sensing, the sparsity and positivity information are incorporated in the charge solving. A penalty term regularized on the absolute value of the unknown with a regularization strength λ is added to the chi-square. From minimizing the chi-square point of view, a larger regularization strength would favor a smaller value of the corresponding ℓ_1-regularized term after charge

---

[2] Non-deterministic polynomial-time hardness

solving, which means the corresponding charge blob will likely be zeroed out. Utilizing this feature, additional information such as *connectivity* can be included through varying the regularization strength for each charge blob to increase the credibility of the charge solving. For example, an isolated charge blob in 3D space will be assigned a larger regularization strength.

Connectivity information is based on the fact that the LArTPC is a fully active detector, and the energy deposition of a charged particle track is continuous. Strictly speaking, the connectivity information is part of the overall pattern recognition. However, this information focuses on the local behavior instead of the global behavior. An algorithm is implemented such that the blobs that are connected to blobs with activity have less chance to be removed. Figure 6 shows the comparison of the 3D imaging results after charge solving for the 3-plane (top panel) and 2-plane (bottom panel) tiling output.

2.1.3 Deghosting

As shown in Fig. 4, with the significant increase in the volume efficiency, the strategy of 2-plane tiling introduces a large amount of ghost activities. Although the charge solving procedure, e.g. varying L1 regularization strength parameters based on connectivity, described in the previous section can reduce some ghosts, the overall performance is still not satisfactory as shown in Fig. 6. Therefore, a set of special algorithms is developed to remove ghosts.

Before explaining the procedure to remove ghosts, we should emphasize how the ghosts are generated. In order to reduce inefficiencies in tiling, we essentially treat the known grouped unresponsive wires as always fired. In this case, the blobs associated with two grouped fired wires and one grouped unresponsive wire would be reconstructed. The 3D images are reconstructed strictly following the three 2D measurements, which is a natural result from the solution to the charge matching equation as shown in Eq. 2. When the 3D images are projected back to each 2D plane (drift time vs. wire number), ghosts only appear in the region covered by the grouped unresponsive wires where the measurements are absent.

The charge hits will be initially clustered[3] based on the connectivity. If the charge cluster is superfluous, i.e. overlapping other bigger clusters, to explain the original measurements in the 2D projective views for the other three or two active wire planes, it will be removed as ghosts. The red circles in Fig. 7 and Fig. 8 correspond to the same volume in TPC. Obviously, the ghosts in Y plane unresponsive region overlap the measurement in U plane active region. This is the foundation to remove ghosts. As a result, the 3D image is most exclusively (in

---

[3] This is the initial clustering solely based on the connectivity. Section 2.2 will describe the additional clustering algorithms that address a couple of realistic issues.

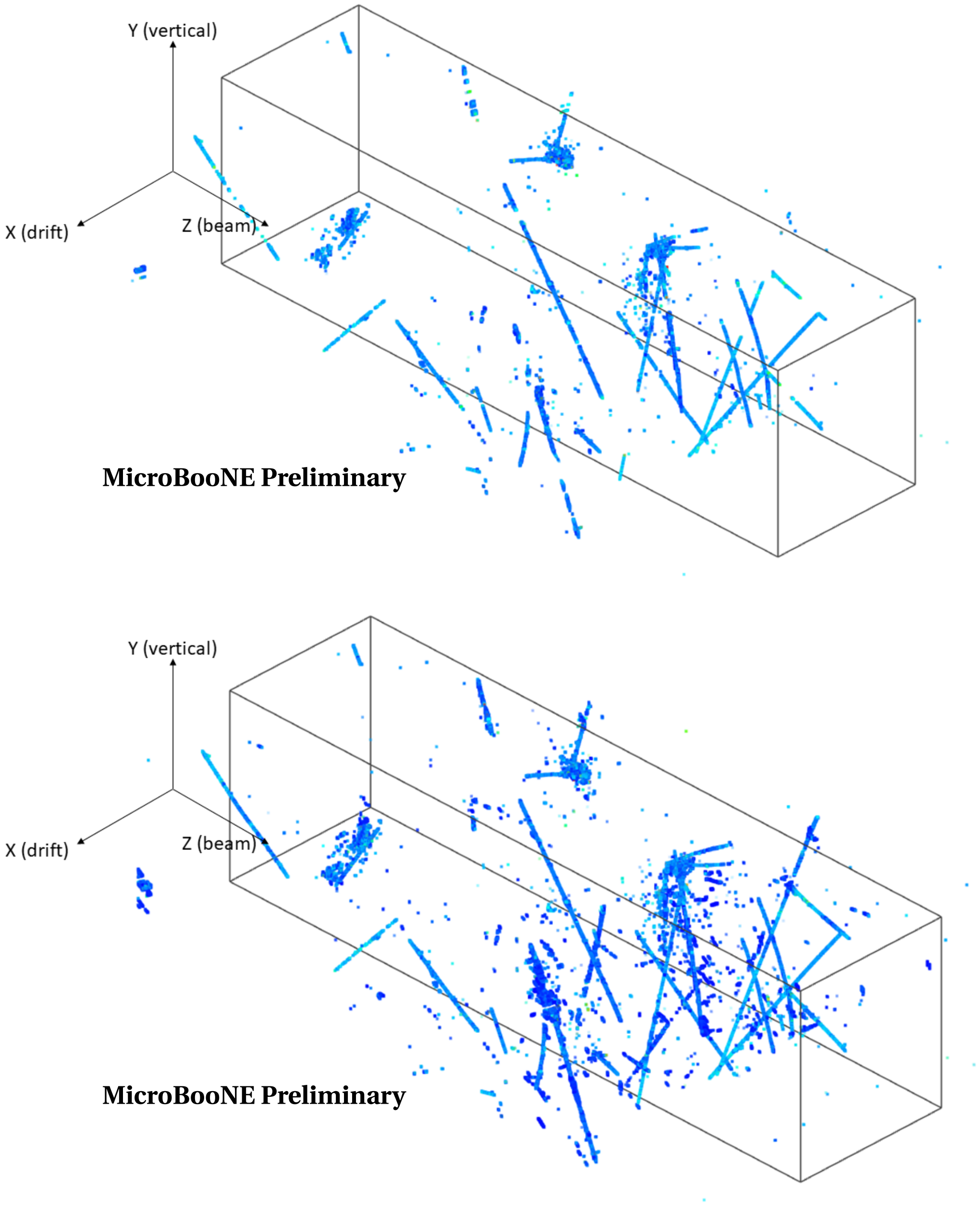


Figure 6: Comparison of the 3-plane (top) and 2-plane (bottom) tiling results after the charge solving. The volume efficiency for the 3-plane and 2-plane tiling is around 70% and 97%, respectively. The increase in efficiency in the 2-plane tiling is accompanied by a large amount ghost activity. The colors represent various charge scales after solving. `https://www.phy.bnl.gov/wire-cell/bee/set/8773a7c1-4829-498d-a14d-baf841a8bea1/event/0/`. `https://www.phy.bnl.gov/wire-cell/bee/set/018443de-6be6-4bbe-bcdb-c3a15af247af/event/0/`.

cluster level) reconstructed based on the three 2D projections (three wire planes). In general, the ghosts are small clusters and disconnected with the major bulk of the TPC activities. After one round of deghosting, another round of solving is needed to recollect the charge carried by the ghosts for the remaining charge hits. In reality, multiple rounds of charge solving and deghosting are performed in the 3D imaging.

Figure 7, Fig. 8, and Fig. 9 compare the 2D projection of the 3D images with the original 2D measurements in each wire plane. Multiple dedicated deghosting algorithms were developed to remove ghosts. Figure 10 show the imaging results with and without deghosting.

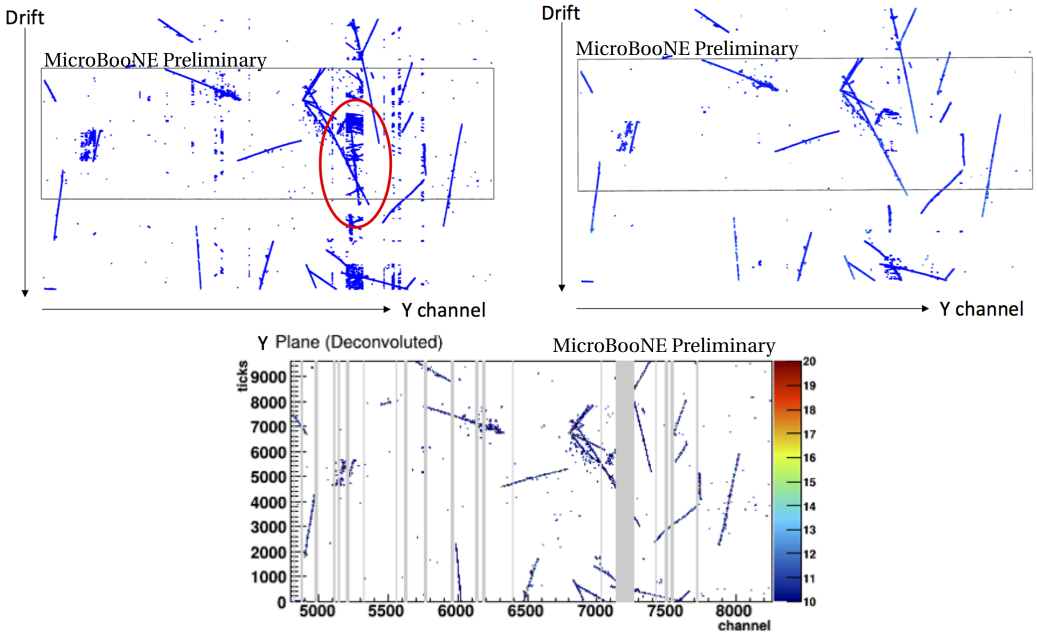


Figure 7: (Top left) Projection to the collection wire plane results from the reconstructed 3D images before applying deghosting algorithm. (Top right) Projection to the collection wire plane results from the reconstructed 3D images after applying deghosting algorithm. (Bottom) Original measurement in the collection wire plane. The grey area represents the unresponsive wires. Comparing the results before and after deghosting, it is obvious that the ghosts are in unresponsive regions as explained. Comparing the result after deghosting with the original measurement, the missing tracks etc. were recovered. The red circle corresponds to the same volume in TPC as indicated in Fig. 8. The ghosts in Y plane unresponsive region overlap the original measurement in U plane active region.

The reduction in the ghosts is apparent. Nevertheless, some ghosts are still visible for the final 3D images. Additional topology information is necessary to further remove these ghosts, e.g. the deghosting in the clustering stage described in the next section.

2.1.4 Time and Memory Performance

MicroBooNE is a surface detector with a large amount of cosmic activity in each readout time window of 4.8 ms. 20-30 cosmics per event are thus input to the wire cell imaging process.

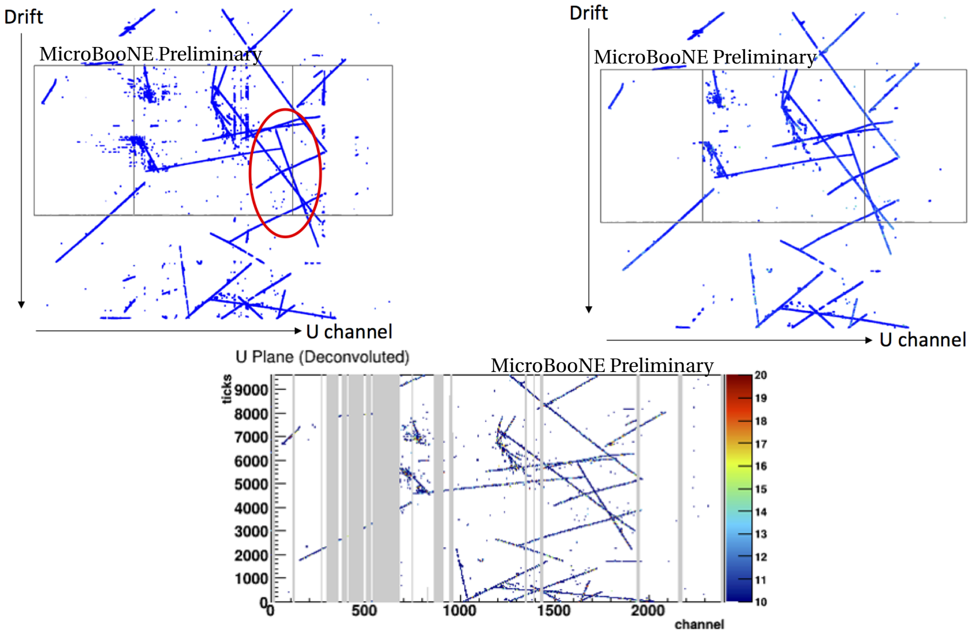


Figure 8: (Top left) Projection to the induction U wire plane results from the reconstructed 3D images before applying deghosting algorithm. (Top right) Projection to the induction U wire plane results from the reconstructed 3D images after applying deghosting algorithm. (Bottom) Original measurement in the induction U wire plane. The grey area represents the unresponsive wires. The red circle corresponds to the same volume in TPC as indicated in Fig. 7. The ghosts in Y plane unresponsive region overlap the original measurement in U plane active region.

MicroBooNE is a high spatial resolution detector and the resolution is about 3 mm by 3 mm by 3 mm. The first 3 mm is determined by the time smearing of the charge extraction [3] and the latter two 3 mm corresponds to the wire pitch. The existing unresponsive channels lead to large ambiguities in the solving process, which in turn results in a large amount of ghosts. Due to all the above mentioned considerations, the time and memory consumption in imaging is crucial to performance.

Using ~10k MicroBooNE data events, the time and memory consumption are shown in Fig. 11. The average time and memory consumption are 190 s and 1.4 GB with the 95% upper limits of 400 s and 1.7 GB, respectively. A benchmark of memory usage is about 1.1 GB initializing the cells as the minimum 2D unit for charge solving. Extra memory is used for indexing and tagging of the 3D space points with any associated charge after solving. A busy event, e.g., an event with a giant electromagnetic shower, will have an extra large memory usage, typically great than 2 GB.

In general, one round of solving takes about 10 s with the compressed sensing technique. However, multiple rounds of charge solving and deghosting are performed in the 3D imaging.

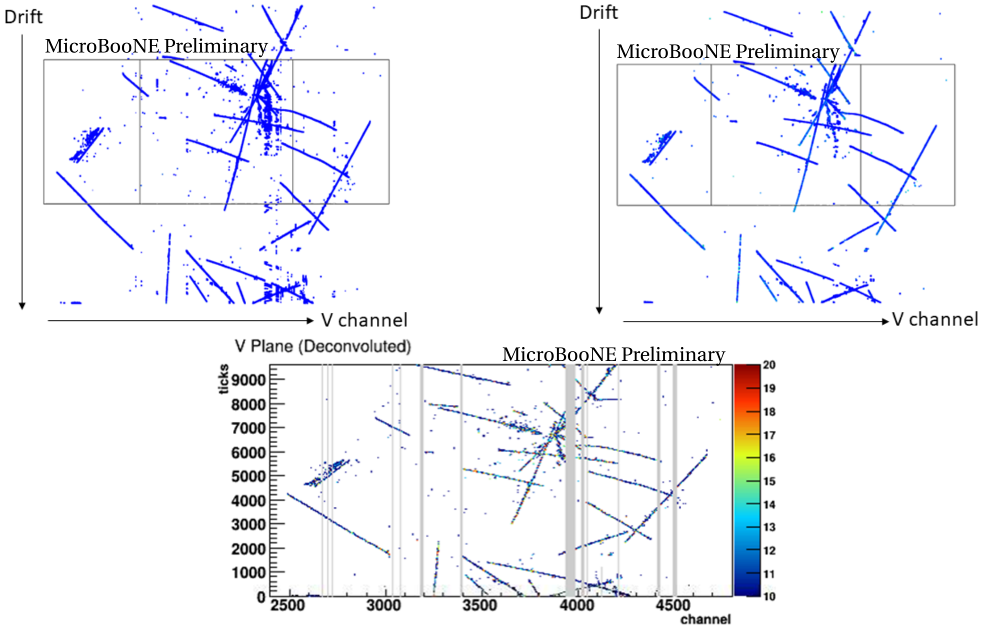


Figure 9: (Top left) Projection to the induction V wire plane results from the reconstructed 3D images before applying deghosting algorithm. (Top right) Projection to the induction V wire plane results from the reconstructed 3D images after applying deghosting algorithm. (Bottom) Original measurement in the induction V wire plane. The grey area represents the unresponsive wires.

Significantly more time is required for solving or deghosting if there's large ambiguity/too many unknowns, typically corresponding to an isochronous track or shower in certain unresponsive regions. In this case, the intrinsic wire readout ambiguity (n^2 v.s. n) will be maximized due to a large number of charge hits arriving at the anode plane at the same time. Worse yet, cells that should have been merged in order to reduce unknowns will be segmented by the unresponsive channels.

2.1.5 Summary

In this section, we describe the implementation of Wire-Cell imaging techniques outlined in Ref. [1] for analysis of MicroBooNE data.

Through the construction of blobs from grouped wires, we reconstruct the most inclusive 3D images that are strictly consistent with all three 2D measurements (Time vs Wire) from the three wire planes. Due to the existence of 10% unresponsive channels, a two-plane tiling procedure is adopted, which allows for the creation of blobs when grouped wires in two of the planes are fired and the corresponding grouped wire in the third plane is known to be unresponsive. This choice significantly enhances efficiency at the cost of the introduction

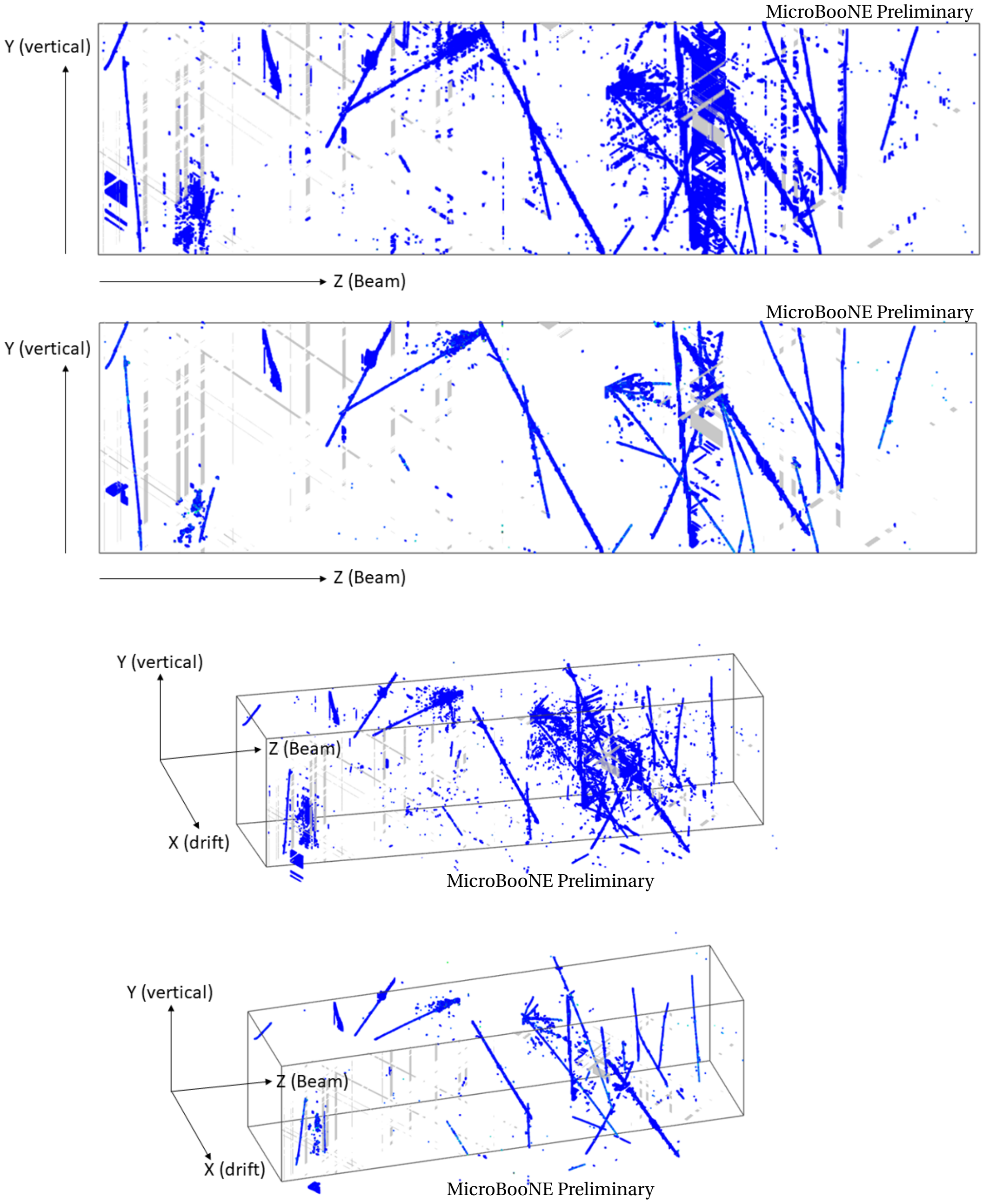


Figure 10: Comparison of the 3D imaging results before (top) and after (bottom) the deghosting algorithm in the Y-Z view and 3D view. Ghosts are significantly reduced after the deghosting procedure. `https://www.phy.bnl.gov/wire-cell/bee/set/018443de-6be6-4bbe-bcdb-c3a15af247af/event/0/`. `https://www.phy.bnl.gov/wire-cell/bee/set/dca28af8-5866-454d-8ab7-c7f6402fdf3c/event/0/`.

of a large amount of ghosts. Additional information including charge, sparsity, positivity, and connectivity is added through the compressed sensing technique in order to remove the ghosts. Special techniques based on initial pattern recognition must be implemented

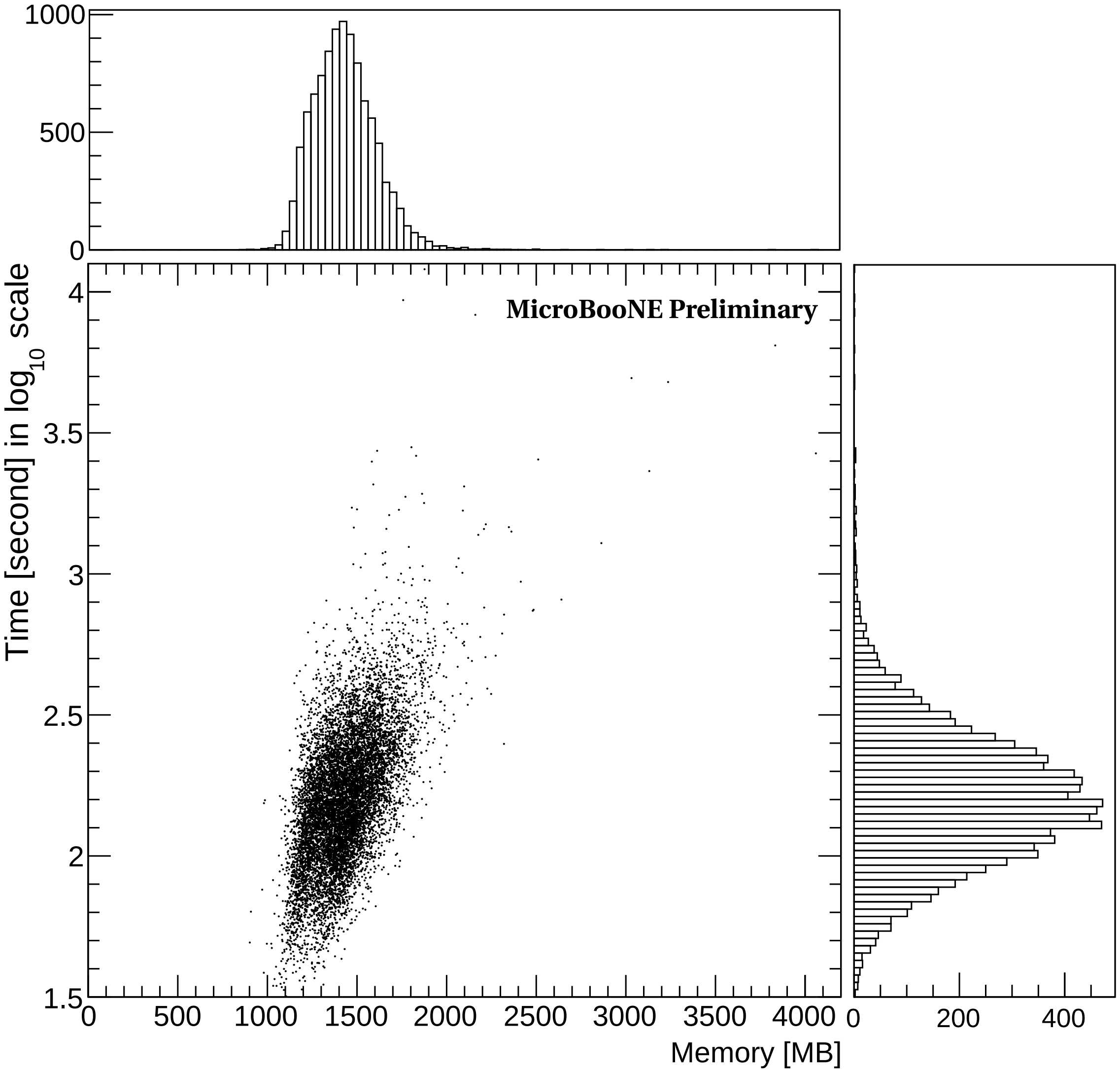


Figure 11: Time v.s. Memory of Wire-Cell tomographic 3D imaging using ~10k events from MicroBooNE data. The average time and memory consumption are 190 s and 1.4 GB with the 95% upper limits of 400 s and 1.7 GB, respectively. More explanations of time and memory usage can be found in the text.

to further remove the ghosts, e.g. the imaging deghosting after the clustering as shown in Sec. 2.2.3.

The evaluation/optimization of the imaging procedure based on simulation and data is ongoing. At present, the tomographic imaging is used to do the semi-automated neutrino selection as shown in this note. The result from the downstream clustering and matching qualitatively implies the efficacy of this tomographic imaging for a variety of event topologies. The robustness of this technique is essential to achieve powerful cosmic removal as well as high efficiency and high purity neutrino selection.

2.2 Clustering

The goal of 3D clustering is to group reconstructed 3D space points according to initial TPC activity, such as a neutrino interaction or a cosmic ray. More specifically, 3D clustering is expected to perform initial separation between cosmic muons and the neutrino interactions into distinct 3D objects. In addition, the proper clustering is inevitable to perform the subsequent many-to-many TPC-flash matching, which matches the reconstructed TPC object to its corresponding scintillation light signals measured by multiple PMTs, i.e. flashes. The TPC-flash matching will correct the t_0 for each cluster, thus rejecting cosmics to the cluster (TPC activity) level that matched to an in-beam flash.

Compared to sampling calorimeters, the detector medium in LArTPC is fully active and charged particle tracks are expected to leave continuous energy depositions resulting in connected space points in the reconstructed 3D images. In the ideal case, the clustering will be straightforward solely based on the connectivity information. Such initial clustering is implemented in the imaging stage. However, several realistic issues arise from the real data as follow.

Gaps: As to be elaborated in Sec. 2.2.1, the presence of gaps due to various realistic detector issues (e.g. unresponsive channels) prevents the effectiveness of this simple clustering strategy.

Random coincidence: For LArTPCs operating on the surface (such as MicroBooNE), the detector is bombarded by large amounts of cosmic muons. Although the cosmic muons generally pass through detector at different times (i.e. t0) and locations, the two 3D images (image of ionization electrons based on the arrival time at the wire plane) of two muons are likely connected as the arrival time of the ionization electrons at the wire plane is determined by both t0 and drifting distance. Two groups of ionization electrons from different muons will lead to connected 3D images if they correspond to the same 2D location on the wire plane and have the same arrival time. So can a neutrino interaction and a cosmic ray. In this case, the reconstructed images of these (separated) cosmic muons or neutrino interaction can appear to be connected, and would lead to over-clustering of space points. The over-clustering further leads to inefficiencies in the TPC-flash matching. Sec. 2.2.2 describes the developed techniques in separating these 'random coincidence' clusters.

Residual ghosts: Due to the existence of 10% unresponsive wires in the MicroBooNE detector, Wire-Cell tomographic imaging utilizes a 2 or 3 wire plane reconstruction strategy to retain a high efficiency in 3D image reconstruction. In this case, the large amount of ghosts appear though the volume efficiency of the 3D imaging is much higher (97% v.s. 70%). The deghosting techniques are applied to the 3D imaging procedure, but they are not completely sufficient in removing these 'ghosts' due to limited information available in this stage. Sec. 2.2.3 describes additional deghosting techniques applied in the clustering stage.

Separated neutrino interaction final states: For neutrino interactions, the π° is a common final state particle. The immediate decay of the π° generally leads to two γs, which start their energy deposition in the detector medium after converting to charged particles ($e^- e^+$ pair and e^- for pair production and Compton scattering, respectively). The reconstructed electromagnetic showers generally appear to be detached from the neutrino interaction vertex and are treated as separated clusters. It is desired that these separated clusters can be grouped together in order to reduce challenges in the later matching of TPC charge and light information. Sec. 2.2.4 describes the developed techniques in clustering separated clusters for neutrino events.

To address these issues, the directionality and distance information will be extracted and utilized. However, due to huge amounts and complexity of the 3D space points from 3D imaging, advanced graph theory algorithms and techniques are necessarily employed for speed and robustness considerations.

2.2.1 Clustering in the presence of gaps

There are four categories of gaps in the reconstructed 3D images from the MicroBooNE detector. The first one is due to the existence of ~10% unresponsive channels. The top panel of Fig. 12 shows the 3% unresponsive regions, where the wires from at least two out of three wire planes are unresponsive. Space points cannot be reconstructed within these unresponsive regions since none or only one wire is active. As shown in the bottom panels of Fig. 12, when a cosmic muon track goes through one of the unresponsive regions, a gap appears in the reconstructed 3D images.

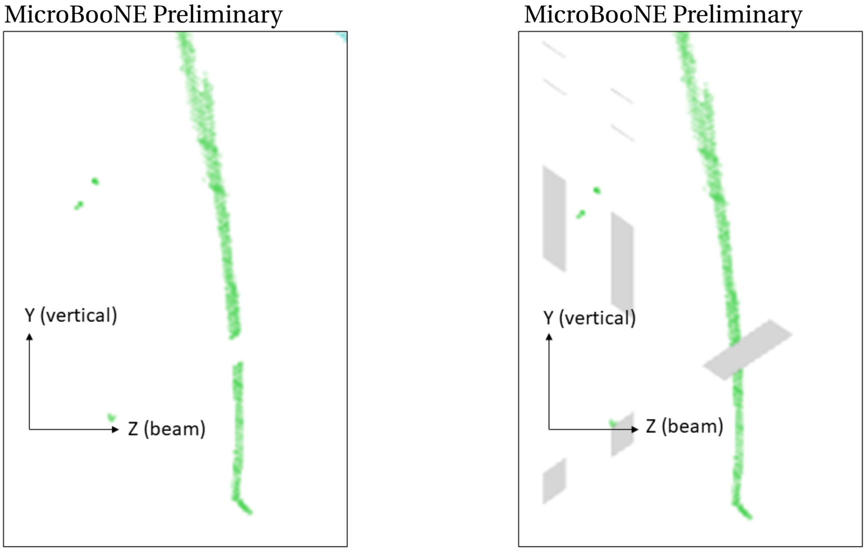


Figure 12: Gaps due to the unresponsive regions as shown in the bottom panel of Fig. 2. (Left) Zoomed-in Y-Z view of a cosmic muon with a gap due to the unresponsive regions. (Right) Same cosmic muon with the unresponsive region overlaid. See `http://www.phy.bnl.gov/twister/bee/set/c43957fb-cc5b-47d3-b44a-4354079b4d30/event/6/`.

The second category of gap is due to the application of coherent noise removal techniques in the noise filtering stage [2]; the goal of coherent noise removal is to reduce the excess of noise introduced by the low-voltage regulator of cold electronics. A naive application of the coherent noise subtraction would lead to deletion of the (coherent) signal for tracks traveling close to parallel to the wire plane. A dedicated signal protection algorithm was developed to avoid this [2]. However, when the signal to noise ratio becomes small, the signal protection algorithm has inefficiencies and gaps appear in the reconstructed images. Figure 13 shows an example of gaps created by the coherent noise removal techniques for a cosmic muon traveling close to parallel to the wire planes. This cosmic muon also travels close to perpendicular to the V-plane wires leading to the smaller density of ionization charge deposition along wire pitch orientation. Further improvement of the coherent noise removal is underway.

As elaborated in Ref. [3], the prolonged track is a difficult event topology for the induction wire plane signal processing. A prolonged track usually leaves a long (in time) ionization charge signal in the induction wire. An extreme case of a prolonged track would be a track traveling completely perpendicular to the wire planes. When the ionization charge signal is long in time, the corresponding measured induced current tends to have smaller magnitude

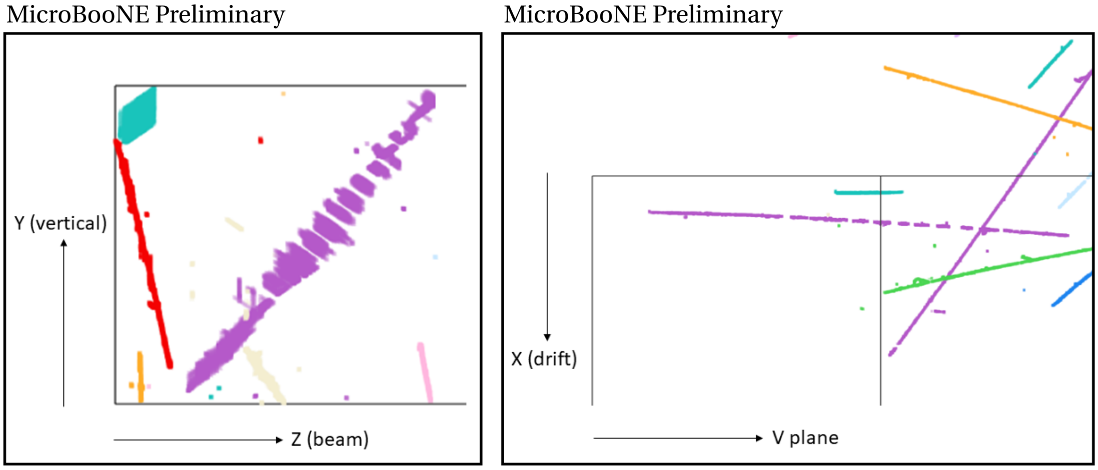


Figure 13: Gaps due to application of coherent noise removal techniques during the excess noise filtering stage. (Left) Y-Z view of a cosmic muon traveling close to parallel to wire plane is shown. (Right), X-V view of the same cosmic muon is shown. X (vertical) represents the electric field direction. V (horizontal) represents the wires in the V wire plane. The gaps are due to deletion of (coherent) signal in the V-plane wires. See `http://www.phy.bnl.gov/twister/bee/set/3f50ab1b-2ad3-4dbd-a8d1-5744e022c37e/event/3/`.

due to the cancellation of the positive and negative induced current, leading to a significantly smaller signal to noise ratio. The charge resolution is expected to be worse for longer signals and the recovery of the ionization charge becomes increasingly more challenging for longer and longer signals. For extremely long signals in the induction wire plane (as shown in Fig. 14), gaps representing failures in recovering the ionization charge start to appear.

The last category consist of observed gaps, which cannot be explained by the previous three reasons. There could be several reasons contributing to these gaps. First, the TPC signal processing [3] requires a signal region of interest (ROI) to be defined. A failure in establishing a proper signal ROI would lead to an inefficiency in reconstructing ionization charge signal. Second, detector non-uniformity (wire pitch, distance between wire planes) can lead to variations in the impulse field response describing the induced current on sense wires. A mismatch between the field response used in TPC signal processing and and the real field response may also lead to gaps. Third, residual excess noise after the noise removal may lead to reduced signal to noise ratio, which may lead to gaps. The number of the gaps in this category tends to be much fewer than those in the previous categories.

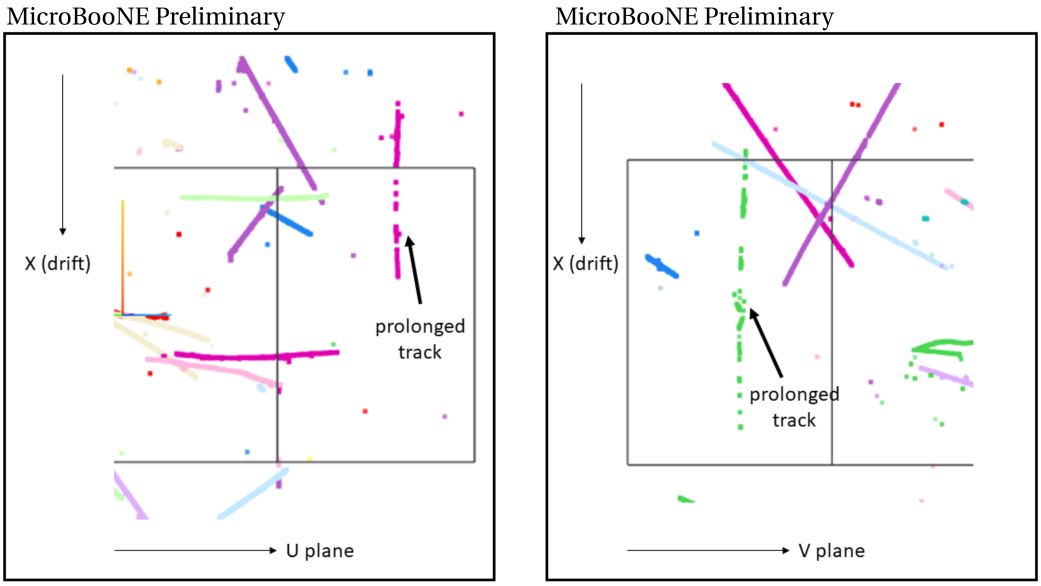


Figure 14: Gaps due to prolonged track topology. (Left) X-U view for a prolonged cosmic muon track (purple). (Right) X-V view for a prolonged cosmic muon track (green). The prolonged tracks lead to long signal along the drift direction (X-axis). See `http://www.phy.bnl.gov/twister/bee/set/3f50ab1b-2ad3-4dbd-a8d1-5744e022c37e/event/125/`. See `http://www.phy.bnl.gov/twister/bee/set/3f50ab1b-2ad3-4dbd-a8d1-5744e022c37e/event/77/`.

Clustering across gaps mainly relies on two sets of information: distance and directionality. 3D space points are constructed within each blob. The distance between any two space points can be easily calculated. A common operation on the distance is to find those space points closest to the edge that have the closest distance to each other from two groups of space points. Such an operation can be achieved by looping through all possible space point pairs. However, this is computationaly intensive given that thousands of space points are reconstructed. Therefore, we employed the ‘nanoflann’ package [16] to do a k-d tree-based operation. The k-d tree data structure (k-dimensional tree) is commonly used in computer science for organizing points in a space with k dimensions. For our purpose, it is limited to three and two dimensions for 3D space points and projected 2D space points, respectively. The k-d tree is a binary search tree, which is very useful for searches in range and of nearest neighbor. Compared to an exclusive search, the average cost of a search using k-d tree is significantly reduced to O(log n) from O(n).

The directionality information is calculated through a voting scheme inspired by a similar technique as that used in the Hough transformation. Given a prechosen space point, the direction towards any other space points of interests can be calculated. The direction vectors for every point, parameterized using polar angle θ and azimuthal angle ϕ, is then filled into a 2D histogram. The peak of the 2D histogram, representing the most favored angles by the space points of interests, corresponds to the final directional vector for a group of space points.

For each pair of clusters, a pair of space points with the closest distance to each other are identified using the k-d tree searches. Within each cluster, the directionality information w.r.t. the picked space point can be calculated. Given this pair of space points with its distance and directional information, one can merge them into a single cluster when the distance and direction information satisfy some pre-specified criteria for each type of gap.

Although the basic technique of clustering is quite straightforward, the actual implementation [17], which is tuned based on a subset of MicroBooNE data (~1500 events), is much more complicated. Given the various event topologies, it is extremely challenging to achieve proper clustering balancing the under-clustering and over-clustering within one loop of initial clusters. Instead, a progressive approach is taken. Table 1 shows the actual functions and their respective purposes.

Table 1: Summary of actual implementation of functions to mitigate gaps in clustering. The actual name of the functions in Ref. [17] and the motivation behind the implementation is listed.

Step	Motivation
I	Clustering over known unresponsive regions
II	Clustering over known (large) unresponsive regions
III	Clustering over gaps between two close clusters
IV	Clustering over smaller gaps for regular tracks
V	Clustering over larger gaps for regular tracks
VI	Clustering over gaps for parallel and prolonged tracks
VII	Further clustering for prolonged tracks
VIII	Further clustering for parallel tracks
IX	Further clustering for regular tracks
X	Clustering over known unresponsive regions

Figure 15 shows a comparison of the results before and after applying these 3D clustering techniques in mitigating the gaps. Space points within the same clusters are labeled with a common color. Due to the limited number of color choices, different clusters may sometimes share the same color in the plot. Using the web application, ‘Bee’, the color of each cluster can be changed randomly to help differentiate separated clusters. Comparing the top and bottom panels of Fig. 15, it can be seen that many separated track segments are properly clustered together, leading to a significant reduction in the total number of clusters. However, there are two cases, where separated cosmic muon tracks are clustered together as they appear to be connected. In addition, the electromagnetic shower (presumably from a π°) is not clustered together with the main cluster of a neutrino interaction. These are dealt with using additional clustering techniques to be described in the following sections.

2.2.2 Separation of 'random coincidence' clusters

Due to the large amount of cosmic muons going through the MicroBooNE detector, the ionization electrons from different cosmic muons can likely arrive at the same location of the anode wire plane simultaneously (e.g. see the red cross near the bottom of left panel in Fig. 16), since the imaging is to reconstruct the 3D image of ionization electrons based on the arrival time at the wire plane. The previous clustering strategy based on connectivity would naturally cluster these separated cosmic muon tracks together leading to an over-clustering situation. The over-clustering of different cosmic muons, which generally go through the detector at separated times, would lead to difficulties when performing matching with the light signal. The over-clustering of the cosmic muon and neutrino interactions could also lead to inefficiencies in selecting neutrino candidates. In this section, we describe the developed techniques to separate 'random coincidence' clusters.

Table 2 summarizes the various steps in separating 'random coincidence' clusters. First, the 'random coincidence' cluster candidates need to be identified. Principle component analysis (PCA) is performed on each of the cluster candidates. For clusters containing a single-track, the eigenvalue of the second component of PCA would be much smaller than that of the first component. This is generally not true for 'random coincidence' clusters, in which two or more tracks are crossing. A cut is applied on the ratio of eigenvalues between the second and first PCA component. Given this algorithm, the cluster of neutrino interactions can be identified as well. In this case, a special algorithm is developed to protect against separating the neutrino-interaction cluster. The algorithm is based on the fact that neutrino interaction must be contained by the default detector boundary, as the neutrino beam timing is known.

Once the 'random coincidence' cluster candidates are identified, the second step is to find the starting and ending points of the first track. First, the quickhull [18] algorithm operates on the 3D space points of the chosen cluster in order to compute the 3D convex hull and its associated points. These points are further merged according to their distances between them. The remaining points are then ranked along the first component of the PCA analysis after examining the number of space points nearby. The examination of the number of nearby

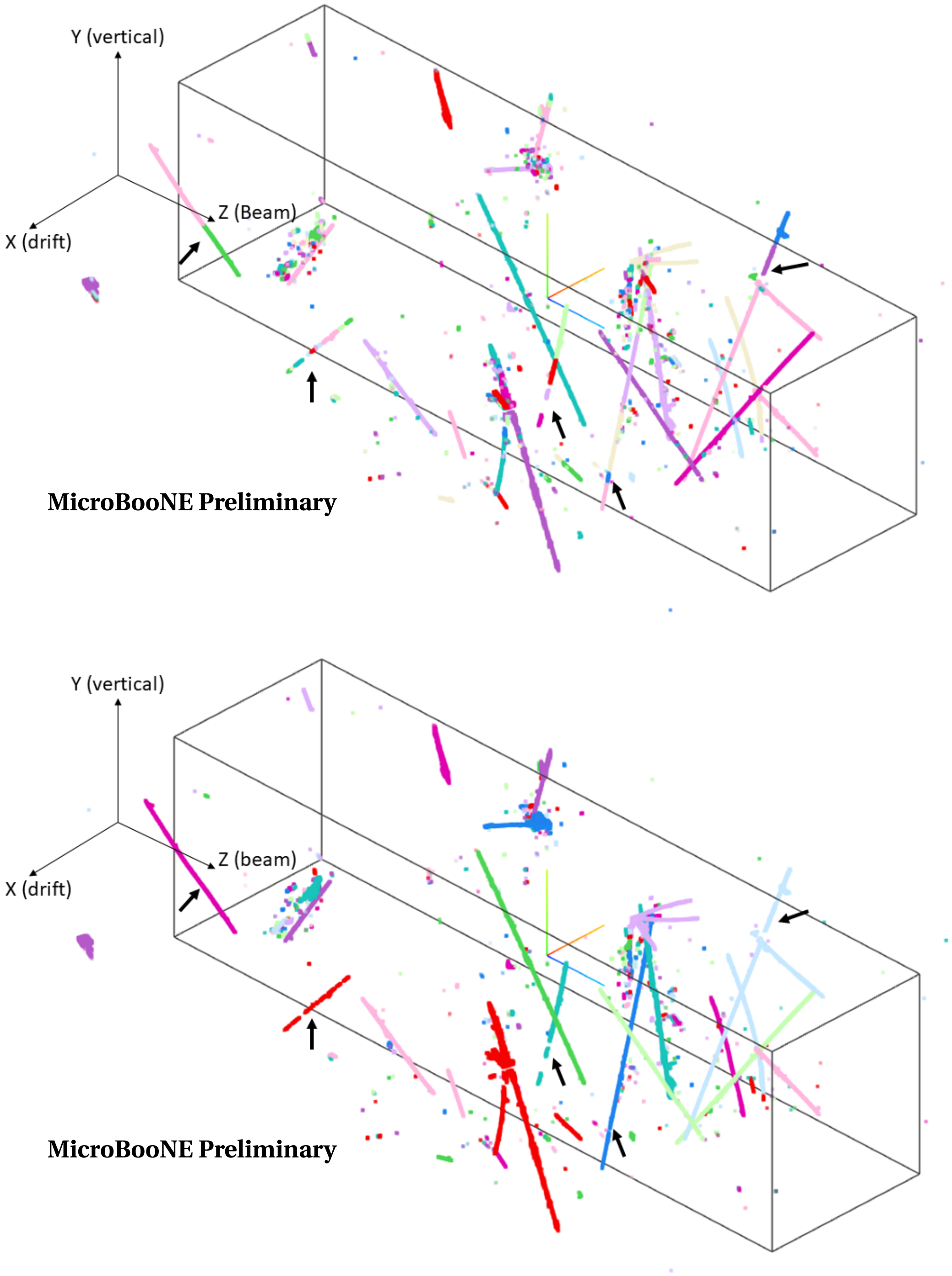


Figure 15: (Top) initial clusters based solely on connectivity. (Bottom) updated clusters after the application of 3D clustering techniques in mitigating gaps. The arrows indicate the significant improvements on some example broken tracks before and after the 3D clustering. See `http://www.phy.bnl.gov/twister/bee/set/3130ffcf-467f-4880-b6ec-c45daba71cd2/event/0/`. See `http://www.phy.bnl.gov/twister/bee/set/59dc5ed3-4624-432d-9bc2-942ade5b8c2d/event/0/`.

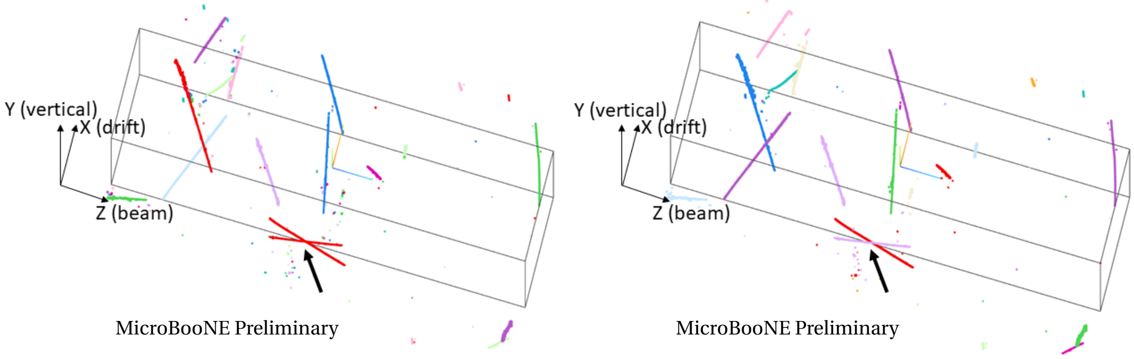


Figure 16: Illustration of a cluster with two 'random coincidence' cosmic muons (the red cross near the bottom of the left panel). The left and right panel shows the clusters before and after applying separation algorithms described in this section. This single cluster is separated into two separated ones. See `http://www.phy.bnl.gov/twister/bee/set/c6840c91-3653-41c0-9e2e-7f34bd1f6c8d/event/1/`. See `http://www.phy.bnl.gov/twister/bee/set/4b10f009-e7d3-4705-a7d9-254e27ede521/event/8/`.

Table 2: Overview of steps in separating 'random coincidence' clusters.

Step	Motivation
I	Identification of 'random coincidence' clusters
II	Find starting/ending points of the first track
III	Form trajectory of the first track
IV	Collect space points for the first cluster
V	Remove the first cluster and its associated activities
VI	Repeat step I for the remaining part

space points is to avoid the small isolated clusters inside the main cluster, which is a result of over-clustering from earlier steps. With this ranking, the two most extreme points are saved. The starting point is further decided from these extreme points considering i) the local directionality calculated based on a voting technique with respect to the first component of PCA and ii) distance to the main PCA component. These conditions are applied to find the starting point of a track which is mostly consistent with the first component of PCA. Once the starting point is found, the ending point is obtained through a Kalman-filter based

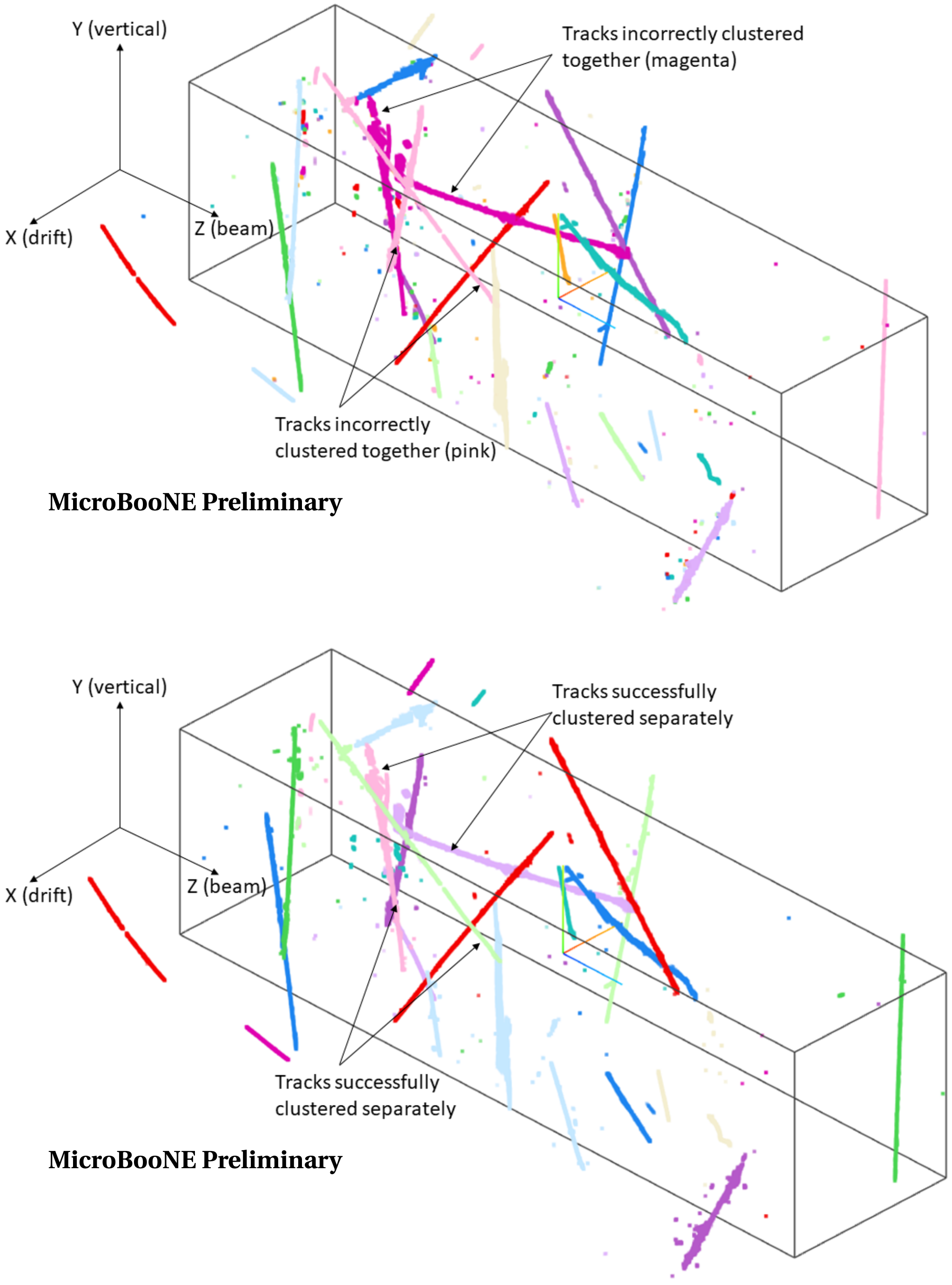


Figure 17: Demonstration of the effectiveness of cluster separation algorithm. The left and right panels show the clusters before and after applying the cluster separation algorithm. See text for more discussions. See `http://www.phy.bnl.gov/twister/bee/set/9cc45a0c-e8fb-4401-ba39-5bc4b7631c4e/event/0/`. See `http://www.phy.bnl.gov/twister/bee/set/4b10f009-e7d3-4705-a7d9-254e27ede521/event/12/`.

technique as illustrated in Table 3. The Kalman-filter based technique is designed to be able to accommodate gaps in the images as introduced in previous section. The step size in Kalman filter is about 5 cm.

Table 3: Kalman-filter based technique to search for the ending point.

Step	Action
I	Calculate local direction with starting point based on voting
II	Make a guess of the next position based on direction and previous point
III	Use k-d tree to find the closest point, and check for compatibility
IV	If not found, keep extending the test position until reaching the limit
V	If found, update the directionality based on previous direction and the vector between the current position and previous position
VI	Repeat step II until reaching limit

With the starting and ending points identified, the trajectory of the first track is obtained using a graph theory operation, the Dijkstra's Shortest Paths, to find the shortest path between the starting and ending points using BOOST c++ libraries [19]. The step size in this operation is much finer than the previous Kalman filter step. With the determined trajectory consisting of a set of 3D space points including the starting and ending point, the nearby space points close to the trajectory can be collected. The space points are collected based on the distance in the 3D space as well as that in the three 2D projections. The graph theory operation, connected components, is used to achieve this.

After the removal of the first cluster, the remaining clusters are examined again for cluster separation. With this repetition, the 'random coincidence' clusters with three or four separated cosmic muons can be properly separated. Figure 17 shows a comparison of the clusters before and after applying the cluster separation algorithm. Two clusters (one with two cosmic muons passing through each other, the other one with one cosmic muon and one neutrino interaction close to each other) are properly separated. Figure 18 shows another example. In this event, a hard scattering is presumably induced by a cosmic muon generating multiple final state tracks including a giant electromagnetic shower. Another presumably 'random coincidence' cosmic muon track passed through one of the final state tracks from the hard scattering. After the application of the track separation algorithm, different tracks are separated into different pieces as expected. Note that, part of the giant shower would likely be improperly clustered, however, such mild under-clustering issue could be addressed by the many-to-many TPC-flash matching using compressed sensing technique as introduced in Sec. 3.2. The downstream pattern recognition is also able to deal with this.

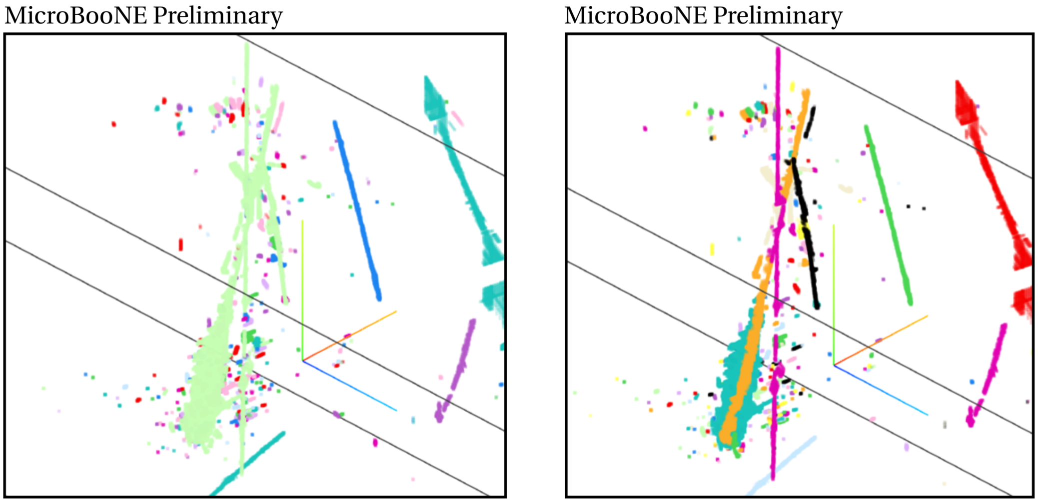


Figure 18: Demonstration of effectiveness of cluster separation algorithm. The left and right panels show the clusters before and after applying the cluster separation algorithm. See text for more discussions. See `http://www.phy.bnl.gov/twister/bee/set/0318a50a-7d24-41a3-bc23-64a1ad398bb0/event/1/`. See `http://www.phy.bnl.gov/twister/bee/set/8d45a3bb-03bc-4839-886c-d3ca46bc27f9/event/38/`.

2.2.3 Image deghosting

As mentioned in Section 2.1, the principle of the Wire-Cell imaging is to reconstruct the most exclusive 3D images based on the measurements from 2D projections. All possible 3D space points are constructed based on geometry and time information. Some of the space points are then removed based on constraints from charge matching (among different wire planes) and sparsity (through implementation of compressed sensing). Since the 2D measurements from each wire plane serve as the constraints, the projection of reconstructed 3D space points back to each wire plane is automatically consistent with the original measurement.

For an isochronous track, a large blob (merged cells) covering a wide range in Y-Z view may show up, e.g. the purple cluster in the left panel of Fig. 13. This type of ghost can be removed once the track hypothesis is asserted through the application of pattern recognition and is generally not a problem. There is another type of ghost, which is more challenging to remove. These ghosts are the results of unresponsive channels or inefficiencies from the TPC signal processing or a combination of both.

Although the deghosting algorithm in the imaging step can remove a large amount of ghosts, it cannot remove all. As shown in the top panel of Fig. 19, ghosts are created when

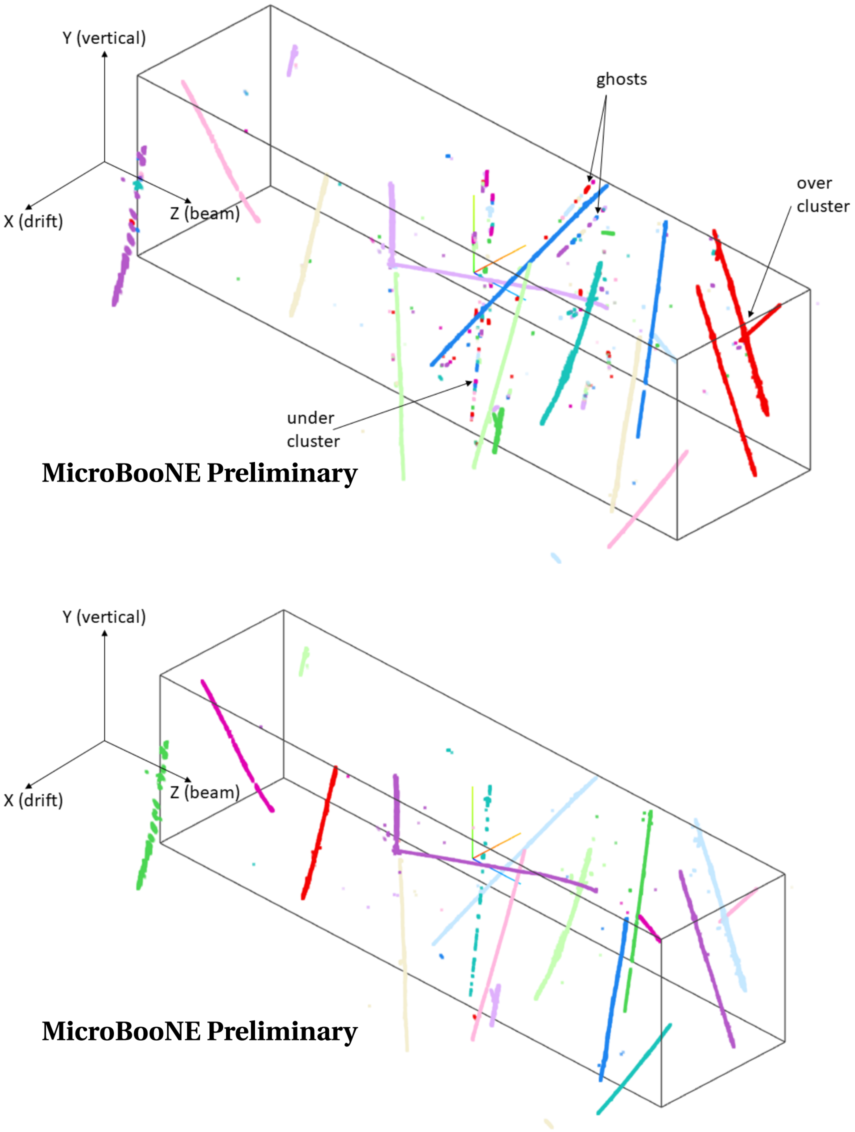


Figure 19: Demonstration of the effectiveness of the deghosting algorithm applied during clustering stage. The top and bottom panels show the clusters before and after applying deghosting algorithm as well as bridging of gaps. See `http://www.phy.bnl.gov/twister/bee/set/b59f1a8b-62a7-4e5b-9cd8-f37ebffb9c87/event/1/`. See `http://www.phy.bnl.gov/twister/bee/set/3f50ab1b-2ad3-4dbd-a8d1-5744e022c37e/event/35/`.

there are inefficiencies in the TPC signal processing or unresponsive channels. However, these ghosts cannot be removed, as they are the only ones which can explain the observed signals in the other two wire planes without TPC signal processing inefficiencies. Another situation is shown in Fig. 20. In a detector region where only two wire planes are active, if there are two real tracks going through, the most exclusive image reconstruction may reconstruct

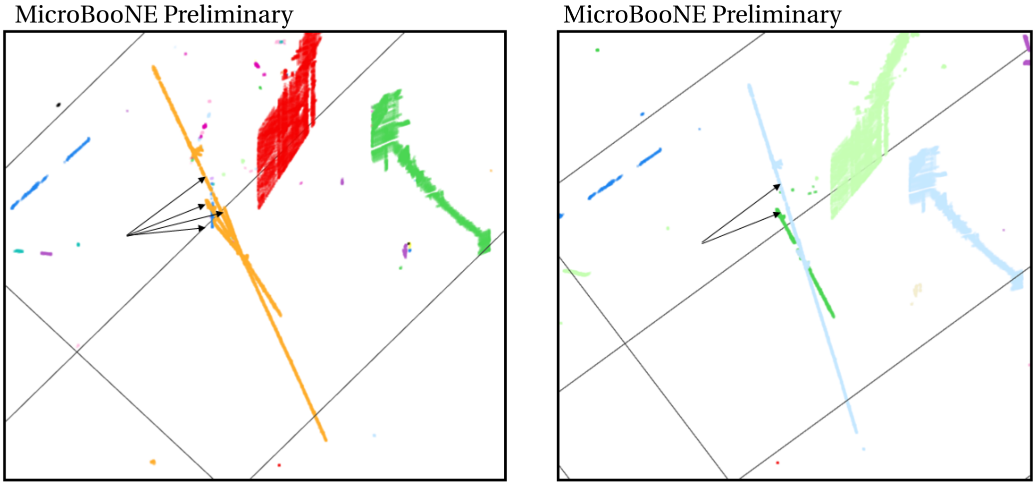


Figure 20: Demonstration of effectiveness of the deghosting algorithm applied during clustering stage. The left and right panels show the clusters before and after applying deghosting algorithm. The four tracks (where two are ghosts) are reduced to two tracks after deghosting. See `http://www.phy.bnl.gov/twister/bee/set/3540665a-6062-4706-805c-cecd5f7e485a/event/0/`. See `http://www.phy.bnl.gov/twister/bee/set/3f50ab1b-2ad3-4dbd-a8d1-5744e022c37e/event/5/`.

four tracks with two being ghosts. If the original two tracks are crossing each other, the four tracks would also be reconstructed to be connected. Since the clustering during the imaging step is based on connectivity, the two ghost tracks would be clustered together with the two real tracks. As a group, these four tracks would survive the deghosting step in the imaging.

With more proper clustering, e.g., bridging of gaps or separation of "random coincidence", the initial clusters in the imaging stage will be connected or separated. Each of the newly created cluster will be re-examined through the same deghosting procedure in the imaging, and the ghosts will be further removed. Two examples are described below. Figure 19 shows an example where ghosts originating from a prolonged track are removed. In the imaging stage, one wire plane has inefficient signal processing and this results in a broken track in this wire plane view. However, the other two wire plane views both have continuous tracks and some ghosts in the unresponsive region are the only solution to explain the charge measurement in this two views. After the final clustering, the pieces of the broken track are grouped together, and a virtual trajectory bridging the gaps can be formed. Since the gaps are filled with additional space points, the ghosts in the unresponsive region are superfluous to explain the charge measurements. Figure 20 shows an example where two

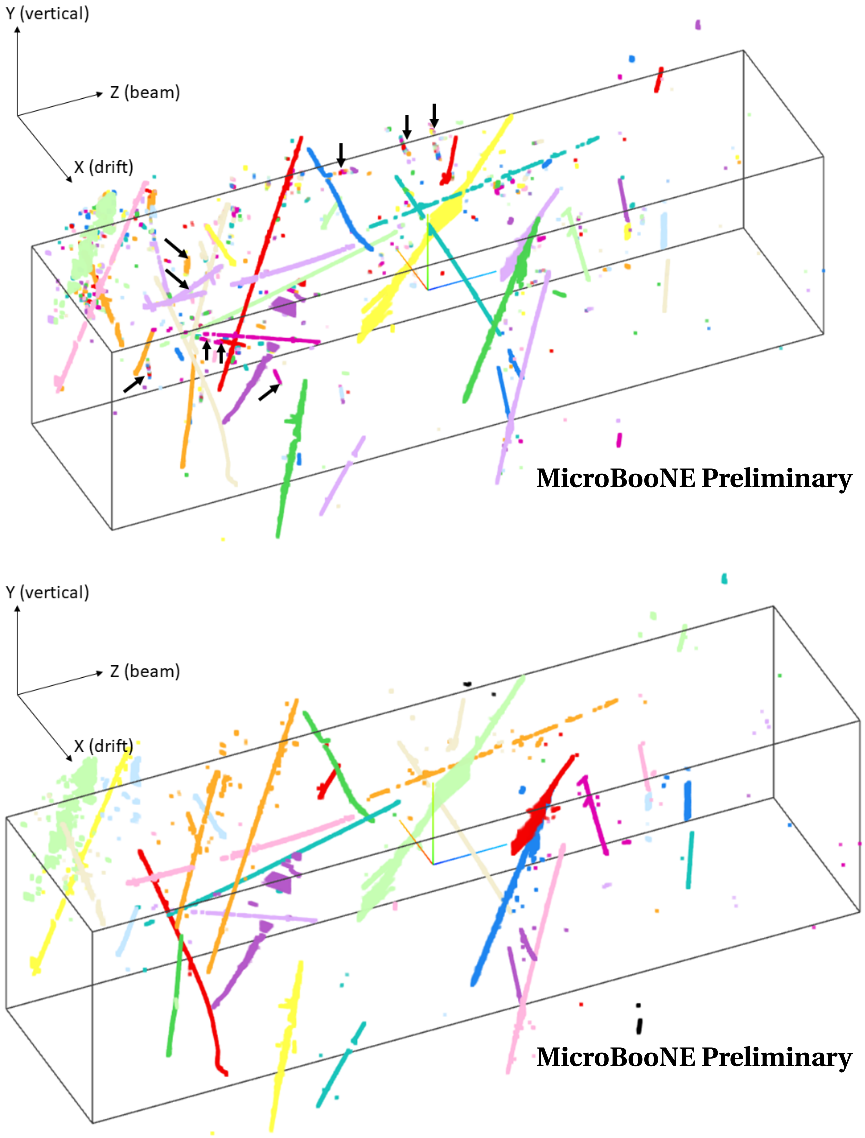


Figure 21: A more complex example demonstrating the effectiveness of the deghosting algorithm applied during clustering stage. The top and bottom panels show the clusters before and after applying deghosting algorithm. Multiple tracks including a prolonged track go through a region where only two wire planes are active. Multiple ghosts with various topology are constructed. They are largely removed after deghosting. See `http://www.phy.bnl.gov/twister/bee/set/3540665a-6062-4706-805c-cecd5f7e485a/event/1/`. See `http://www.phy.bnl.gov/twister/bee/set/3f50ab1b-2ad3-4dbd-a8d1-5744e022c37e/event/12/`.

crossing tracks generate two connected ghost tracks. The single cluster is separated into four clusters. The two clusters corresponding to the two fake tracks are then re-examined through the deghosting procedure and are removed, while the four tracks from a single cluster are examined as a whole in the imaging stage and cannot be removed. Figure 21 shows a more complex situation. Multiple tracks including a prolonged track go through a region where U wire plane is largely inactive. Many ghosts with various lengths are reconstructed. Some of the ghosts appear to be connected with real tracks and are clustered together. After the clustering and track separation, the deghosting algorithm can be run much more efficiently.

2.2.4 Clustering for neutrino events

The clustering algorithm described in the previous sections essentially conclude the clustering based on connectivity. In this section, we describe an additional clustering algorithm designed for neutrino interaction events in which case the final state particles (such as neutron or π°) can lead to detached reconstructed clusters from the primary neutrino interaction vertex. These clusters are reconstructed to be separated. However, for certain applications (such as matching with the light signal with the reconstructed TPC object), it is desirable to merge these separated clusters together. In the following, we describe the detailed steps and the major task is to find the common vertex for certain clusters:

- Each cluster is examined to find the ones that are fully or partially contained inside the TPC volume w.r.t the beam spill time. This selection is designed for neutrino interactions, as the DAQ trigger initiated by the beam ensures that the neutrino-related activity is reconstructed inside the TPC volume.

- For each contained cluster, the two most extreme points are identified. The distance between the two extreme points are the longest among all possibilities. Both points are further examined to ensure that they do not belong to any isolated small clusters. If one of the points is from a small cluster, the entire cluster is divided into small clusters using a graph theory operation, connecting components. The two extreme points for each small sub-cluster are then saved.

- The local direction starting from each extreme point based on the neighboring points is calculated. New space points within certain ranges are added along the inversed local direction. For clusters with relatively small size (difficult to calculate a robust direction), additional space points are added within certain range from the center of the cluster.

- The existing or newly created space points from all clusters are examined to find the common vertex. Take two gammas from π° decay as one example, each group of space points is used to calculate a line along its main axis. When the distance between the two lines are smaller than a predefined cut value, an "intersection" is defined. This "intersection" is required to be a candidate of the interaction vertex (edge point of one or more clusters, not the middle point of any track). If this is true, these two clusters are merged into a bigger one. The above algorithm can also work for one big cluster and one small cluster. In this case, the small cluster would define a line according to its direction. The closest distance would be between the line and the points in the big cluster.

Figure 22 shows a successful clustering of a complicated neutrino interaction event. Two γ's from a π° decay and a detached charged particle track from an energetic neutrino interaction are clustered.

2.2.5 Summary

In this section, we describe the 3D clustering technique after the 3D imaging. The presence of gaps significantly increases the challenges in clustering. Many advanced techniques, such as graph theory operations (Dijkstra's shortest distance, connected components, minimal spanning tree), Kalman-filter, 3D and 2D k-d trees, are used to obtain a robust result in a quick and efficient way. The development of the current chain of 3D clustering relies on a hand scan of ~1500 data events, which contains several hundred neutrino events with a large variety of event topologies. We estimate the clustering mistakes (compared to hand scan) on the 2-3% level per event (0.1% level per cluster). A full evaluation of the robustness of these clustering techniques is on-going. This technique hooks the TPC-flash matching up to the tomographic imaging and is the foundation of successful many-to-many TPC-flash matching. Many more event displays of the semi-automated neutrino selection also qualitatively implies the efficacy of this clustering technique.

2.3 Light signal reconstruction

Capitalizing on of the interplay between scintillation light and ionization charge signals has the potential to significantly improve ν selection performance and mitigate backgrounds. In the subsections to follow, we describe the reconstruction of optical *flashes* — time coincident light signals among multiple channels.

MicroBooNE has an array of 32 PMTs positioned inside the cryostat external to the TPC on a plane parallel to the anode. From the cryostat, the PMT signal is split from the HV at the splitter. After the splitter the signal exists in a redundant, paired form for each channel – a high gain signal (x10) and a low gain signal. Each signal is subsequently shaped and digitized at 64 MHz.

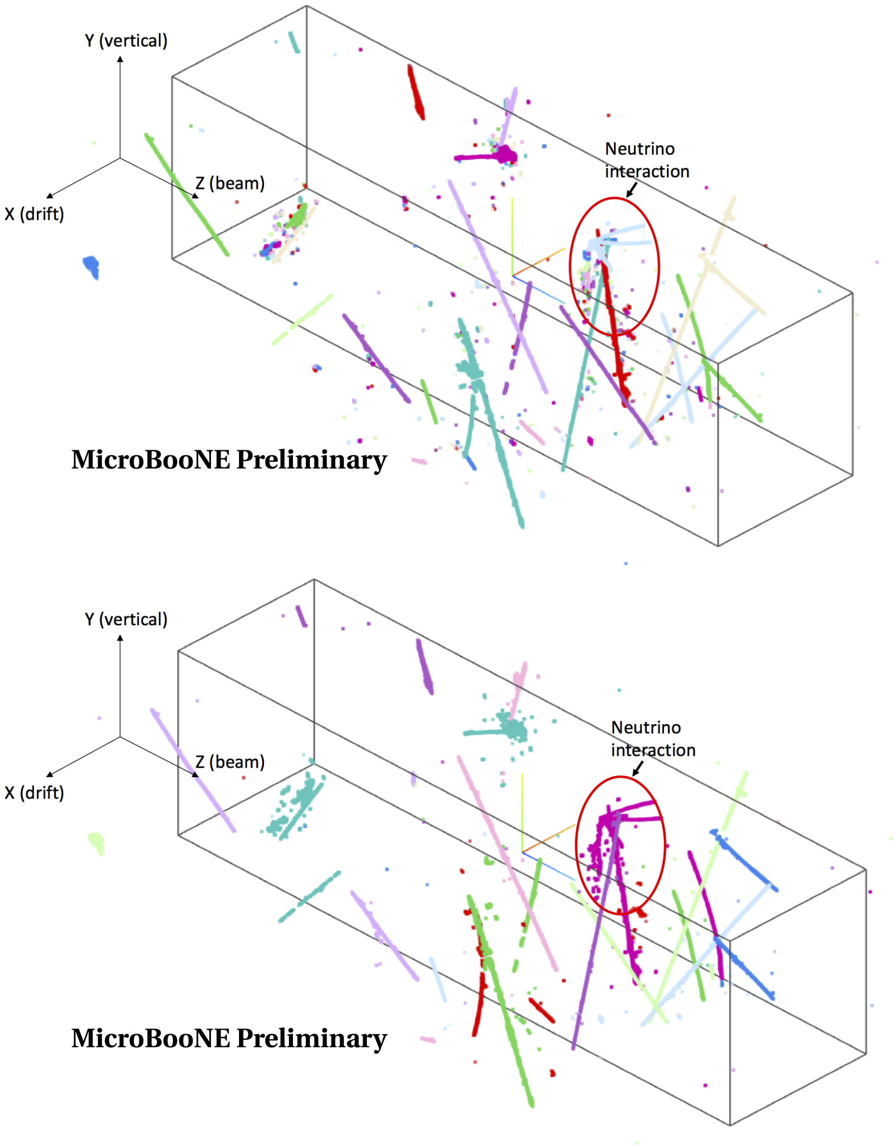


Figure 22: Demonstration of the effectiveness of the clustering algorithm designed for neutrino interactions. Top and bottom panels shows the clusters before and after applying such clustering algorithm. The two electromagnetic showers (presumably from two γ's from π° decay) and one detached track (presumably from an energetic neutron re-interaction) are labeled as separated cluster on the top panel. In the bottom panel, they are included into the main neutrino interaction cluster. See `http://www.phy.bnl.gov/twister/bee/set/59dc5ed3-4624-432d-9bc2-942ade5b8c2d/event/0/`. See `http://www.phy.bnl.gov/twister/bee/set/9dcd67d9-15f6-46a5-a856-c8517e86f343/event/63/`.

PMT signals are recorded in two forms for any given event. The first is called the beam discriminator. The beam discriminator starts to record upon receipt of the beam-gate signal at the trigger board. A 1500 sample readout is recorded for all 32 channels. The second form in which PMT signals are recorded is called the cosmic discriminator. When a signal threshold is exceeded for a particular channel, a 40 sample readout is recorded. The cosmic discriminators span 6.4 ms, fully covering the 4.8 ms spanned by the TPC readout. A comparison of the different intervals of activity for the different PMT signals is provided in Figure 23. A description of flash reconstruction for both cosmic discriminator and beam discriminator readouts is provided below.

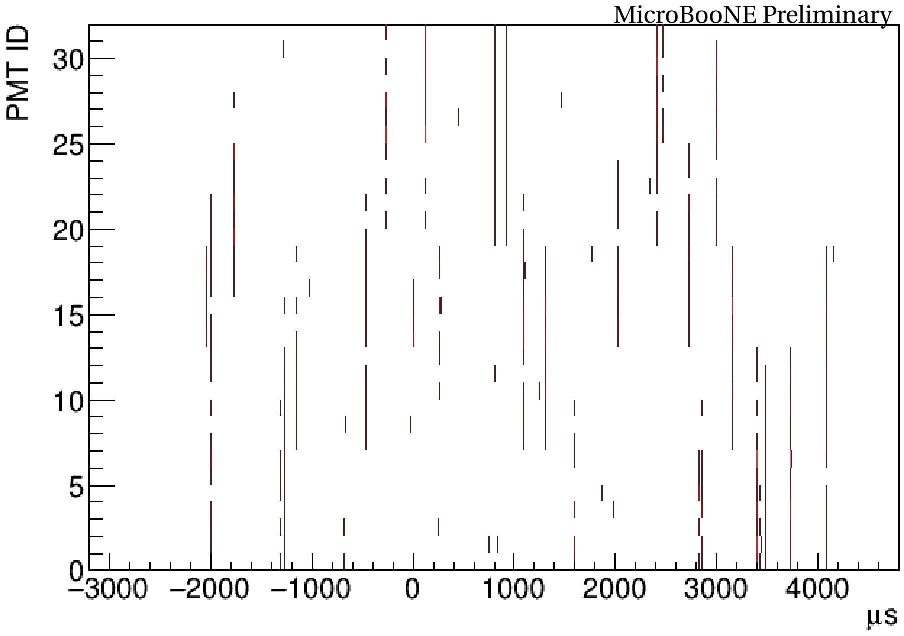


Figure 23: Reconstructed PMT "flashes" from cosmic and beam discriminators in the full event readout.

2.3.1 Cosmic discriminator flash construction

Some cosmic discriminators carry redundant information, e.g. high gain and low gain cosmic discriminators for the same PMT fire on the same scintillation light signal. If the high gain cosmic discriminator signal saturates the ADC, information from the low gain cosmic discriminator is used. It is imperative to understand for a given channel which fired high gain cosmic discriminator corresponds to which low gain cosmic discriminator to make this replacement when the high gain saturates. This process of pairing high gain and low gain cosmic discriminators is called *merging*.

After merging, cosmic discriminators are processed to form a collection of optical hits, or"COphit", for an event. The baseline is taken as the first bin in the 40-sample cosmic discriminator. The peak (maximum ADC minus baseline) and integral (integration of baseline-subtracted ADC) are calculated for each cosmic discriminator. Cosmic discriminators with an unstable baseline most likely re-fired on late light[4]. A small integral value (smaller than 20k ADC) or a baseline value outside of a prescribed range (above 2100 ADC or below 2000 ADC) are indicators of a late light fired cosmic discriminator. If the baseline is identified to be bad, we use the ADC peak value to deduce the integral in order to get a more precise value. The number of photoelectrons (PE) is calculated by taking the ratio of the ADC integral to the calibrated PMT gain. A factor of two is then multiplied on the PE to take into account the late light contribution, which is not recorded by the 40-sample cosmic discriminator[5].

The time difference between time adjacent cosmic discriminators is plotted in Fig. 24 and consistent with expectation. A band of "COphit" are grouped together to form a *flash* if they fire within 100 ns of each other.

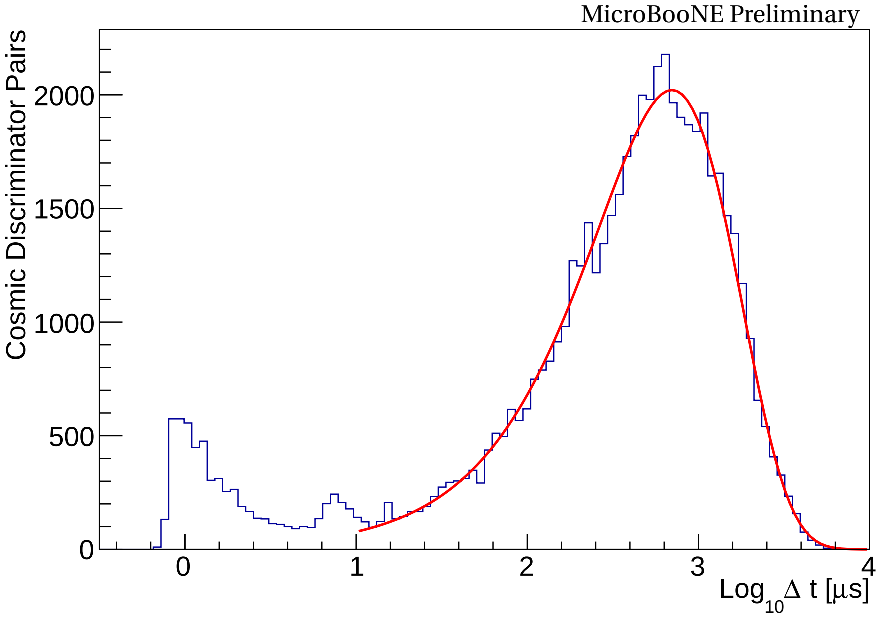


Figure 24: Time difference distribution among time adjacent cosmic discriminators. The first peak is due to cosmic discriminators re-firing on late light. The peak at ~7.4μs is believed to be due to an instrumental effect brought on by PMT power supply switching. The final peak is from random light and is well described by a Poisson distribution. By fitting this final peak, we obtain a decay constant of 1.7±0.2 μs which is consistent with the argon triplet state time constant of ~1.6μs.

.

---

[4] Argon scintillation light has two components: a fast component described by an exponential with $\tau \approx 6$ ns and a slow component described by an exponential with $\tau \approx 1.6$ μs. We refer to the latter as late light.

[5] The factor of two scaling approximates late light that would otherwise be unaccounted for due to the discriminated cosmic discriminator readout. This scaling for the cosmic discriminator is meant to put these flashes on somewhat equal footing as flashes produced by the beam discriminator.

2.3.2 Beam discriminator flash construction

Like the cosmic discriminators, the beam discriminators too carry redundant high and low gain information. The channel-wise beam discriminators are merged if the high gain saturates. Only the portion of the discriminator that saturated is swapped out with scaled low-gain bins.

After merging, a simple baseline correction and subtraction are performed. Next, we deconvolve the overall response (RC-CR^4 from the shaper and RC from the splitter). The deconvolved results are rebinned by six (a single time bin is now 93.75 ns).

The algorithm scans through the deconvolved waveform bin-by-bin to determine if a flash should be made. For a flash to form in the beam discriminator, at a minimum the following two criteria must be met. First, the total summed PE across all channels in a given time bin (93.75 ns) must meet or exceed 6 PE, where channels with less than 0.2 PE are ignored. And second, three or more channels must contain at least 1.5 PE. Typically only one flash may be initiated within 78 time bins ($\sim$7.3 μs). However, the creation of a new flash may override the hold-out time from another flash if the following conditions are met:

- the time difference between a given time bin and the last flash's time bin is greater than 4 time bins (shaping time 62.5 ns)
- AND the summed total PE across all channels in a given time bin exceeds the last flash PE
- AND EITHER the time difference between a given time bin and the last flash's time bin is greater than 15 time bins (1.4 μs), OR the PE distribution across different channels at a particular time bin is adequately dissimilar to the same previous time bin's PE distribution across different channels as determined by a KS shape test.

A flash's total duration is $\sim$7.3μs (corresponding to 80 bins), where the max PE bin marks the flash "time". The second bin preceding the flash "time" mark the flash's "begin time" and the 78th bin from the "begin time" marks the flash's "end time". The flash "time" is used for the PMT filter and matching code. Figure 25 step-by-step shows the results from the light signal processing and flash formation.

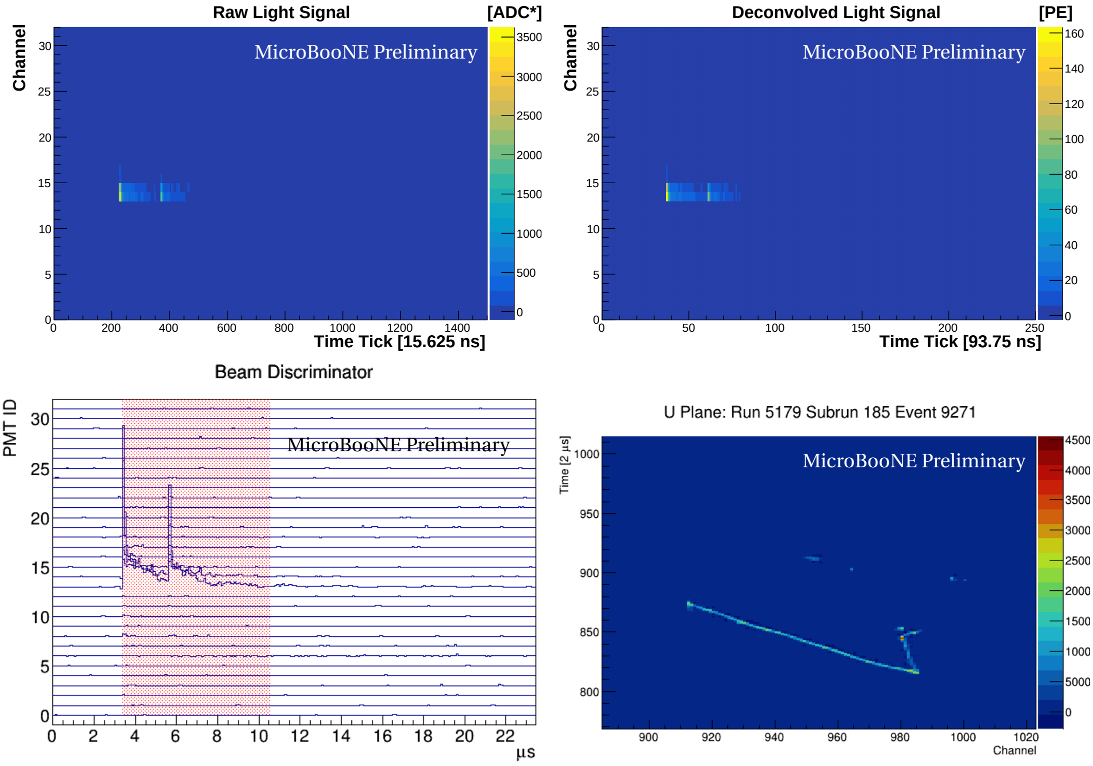


Figure 25: Illustration of the light signal processing. The top left panel shows the baseline subtracted PMT waveform [unit ADC*] for 32 channels. The top right panel displays the deconvolved spectrum [unit photoelectrons] for 32 channels. The bottom left panel depicts the flash (red band) reconstructed in the beam discriminator. The second vertical blue line indicates that the light signal from a Michel electron is tagged by a dedicated algorithm (not described). The bottom right figure shows the TPC object, in this case a muon decaying to a Michel electron, corresponding to the flash as measured by the TPC U plane.

3 SEMI-AUTOMATED ν_μ CC SELECTION ON MICROBOONE DATA

With the tools described in Section 2, we performed a semi-automated scan of BNB (on beam) and external BNB (off beam) events to gauge the efficacy of these tools. All steps, apart from the final step, are automated.

- **Automated** TPC noise filtering and signal processing
  The noise filtering and signal processing are not described in this note. Details, however, can be found at References [2], [3], and [4].

- **Automated** 3D imaging (Sec. 2.1)
  The 3D image of the event is reconstructed. Note that this is different from showing the "reconstructed 3D tracks", which could suffer from reconstruction biases. The 3D image is faithfully consistent with each 2D projective measurement [1]. This step alone

is a vast improvement over traditional 2D-based hand-scan tools because the event topology and detector boundary are much more clear in this view.

- **Automated** 3D clustering (Sec. 2.2)
  In this step, the 3D image is automatically clustered into spatially separated 3D objects. This is a pivotal step in that millions of space points are condensed into tens of "3D objects" that represent physical tracks and showers.
- **Automated** light reconstruction (Sec. 2.3)
  In the light reconstruction we make full use of all the light information collected. A robust and unbiased light (flash) reconstruction is crucial so that it can match the TPC object's predicted light pattern. Typically tens of flashes are reconstructed for each event.
- **Automated** PMT filter (Sec. 3.1)
  The optical filter relies exclusively on the timing of flashes in each event. Only events for which there is a flash coincident with the beam spill are kept for further analysis.
- **Automated** cluster-flash matching (Sec. 3.2)
  At this stage, the tens of TPC clusters are matched to the tens of light flashes. Each TPC object is then placed at the correct t_0 (i.e. the correct drift location with respect to the boundary).
- **Automated** in-beam matched cluster selection (Sec. 3.2)
  The in-beam cluster selection is now trivial. We only need to scan for the TPC cluster that has a corresponding flash coincident with the beam spill.
- **Single key hand-scan** on the matched cluster in 3D (Sec. 3.4)
  With the stroke of a single key "\" in the Bee event display, the beam-spill coincident TPC cluster is identified. At this point, we have an extremely *clean* event that is available to be hand-scanned in 3D. We check if the measured light pattern matches the predicted light pattern and if the 3D topology is like a neutrino. We use the detector boundary, evidence of multiple Coulomb scattering, Michel electron-like topology, to discard likely background events. Note, no calorimetric handles have been employed.

In the sub-sections to follow, we describe the sequential steps employed in the semi-automated ν_μ CC selection on MicroBooNE data:

1. PMT filter (Sec. 3.1)

Trigger Type	Trigger Fraction	PMT Filter Fraction
BNB (beam on)	0.046	0.302 (0.327)
External BNB (beam off)	0.041	0.240 (0.236)
Ratio	1.12	1.26 (1.39)

Table 4: "Trigger Fraction" corresponds to a 5E19 POT data sample. "PMT Filter Fraction" provides two numbers: one corresponds to a 5E19 POT data sample and the other (in parenthesis) corresponds to the small (4k BNB + 3k external BNB) data samples that were used in the semi-automated scan.

2. Flash-cluster signal matching (Sec. 3.2)
3. TPC filter (Sec. 3.3)

The performance of this selection on 3.96 k event beam on (BNB) and 3.15 k event beam off (external BNB) samples is described in Sec. 3.4.

3.1 PMT Filter (Automated)

The first step in the analysis sequence for neutrino selection is the PMT filter. Using flash timing, we veto events which do not contain a beam discriminator flash within the beam-spill. The beam-spill is expected between 3 μs and 5 μs in the beam discriminator.

For the 3.96 k BNB (beam on) and 3.15 k external BNB (beam off) event scan samples, the event passing rate is 32.7% and 23.6%, respectively. We infer that the PMT filter excludes some late light cosmic triggered events because the ratio of on beam to off beam events after the PMT filter relative to the ratio after the software trigger increases, as provided in Table 4.

3.2 Many-to-Many Cluster-Light Matching (Automated)

The second step for neutrino selection is the matching of TPC clusters to PMT flashes. We perform a “many-to-many” matching between TPC clusters and PMT flashes in an event to identify the potential neutrino-candidate and to disambiguate this activity from other activity that may appear in the event (e.g. cosmic rays).

In a single-phase LArTPC, it takes O(ms) for ionization electrons to drift to the anode readout plane. While the TPC can provide a high precision position measurement, it is a relatively slow detector. For LArTPCs operated at the Earth’s surface, cosmic muons that arrive at the detector readout randomly compared to the beam-induced activity can be a major source of background for a neutrino analysis. Since the TPC only measures the arrival time of ionization electrons on the anode wire plane relative to the trigger time, a direct conversion of the drift time to drift distance leads to incorrect reconstructed positions along the drift direction. For example, as shown in Fig. 26, some of the TPC objects can be reconstructed outside the expected detector active volume (black box).

In MicroBooNE, since the TPC trigger time is very close to that of the BNB beam pulse time, the TPC signal for a real accelerator neutrino-induced interaction is expected to be fully contained[6] inside the default drift volume of the detector. Cosmic muons entering into the active volume are expected to attach to a boundary of the active volume. However, since the true event time of the cosmic muons can be different from the TPC trigger time, an offset along the drift direction will be present in the reconstruction. For example, a stopped cosmic muon, which would in reality pierce the boundary of the active volume, could appear to be fully contained inside the active volume. This situation may lead to potential backgrounds in a neutrino selection. Therefore, determining the true time for each cosmic activity is critical in rejecting this kind of background.

The cluster-light matching results in a significant reduction of the TPC activities on the object (cluster) level, basically from 20-30 objects per event to a single object that is matched with an in-beam flash. On the event level, the cluster-light matching results in a small reduction in the data after the PMT filter. For the 3.96 kBNB and 3.15 k external BNB samples, the overall passing rate is 31.7% and 23.4%, respectively, after the matching step.

The sub-sections to follow describe the principal steps in the many-to-many TPC cluster/PMT flash matching algorithm. This algorithm attempts to determine the best match for every TPC cluster and every PMT flash.

3.2.1 Hypothesis on matched flash-cluster pair

Each TPC flash is examined against every TPC cluster to test if they can be a potential matched pair. For a matched pair, the TPC cluster should be contained inside the expected active volume given the TPC flash time (red box as shown in Fig. 28). Once the potential pair is confirmed, one can predict the expected number of photo-electrons (PEs) on each of the PMTs given the reconstructed ionization charge information from 3D imaging (described in Sec. 2.1).

Under the assumption that the number of scintillation photons generated is proportional to the number of ionization electrons, the PEs recorded on each PMT can be predicted. The

---

[6] Approximately half of the LAr in the MicroBooNE cryostat is external to the TPC active volume. Therefore, there are neutrino interactions that are partially contained or even entirely external to the TPC active volume. However, any neutrino event, even if entering from the outside into the TPC active volume, is still coincident with the beam pulse time. Therefore a partially contained, external originating neutrino event is consistent with the default drift volume.

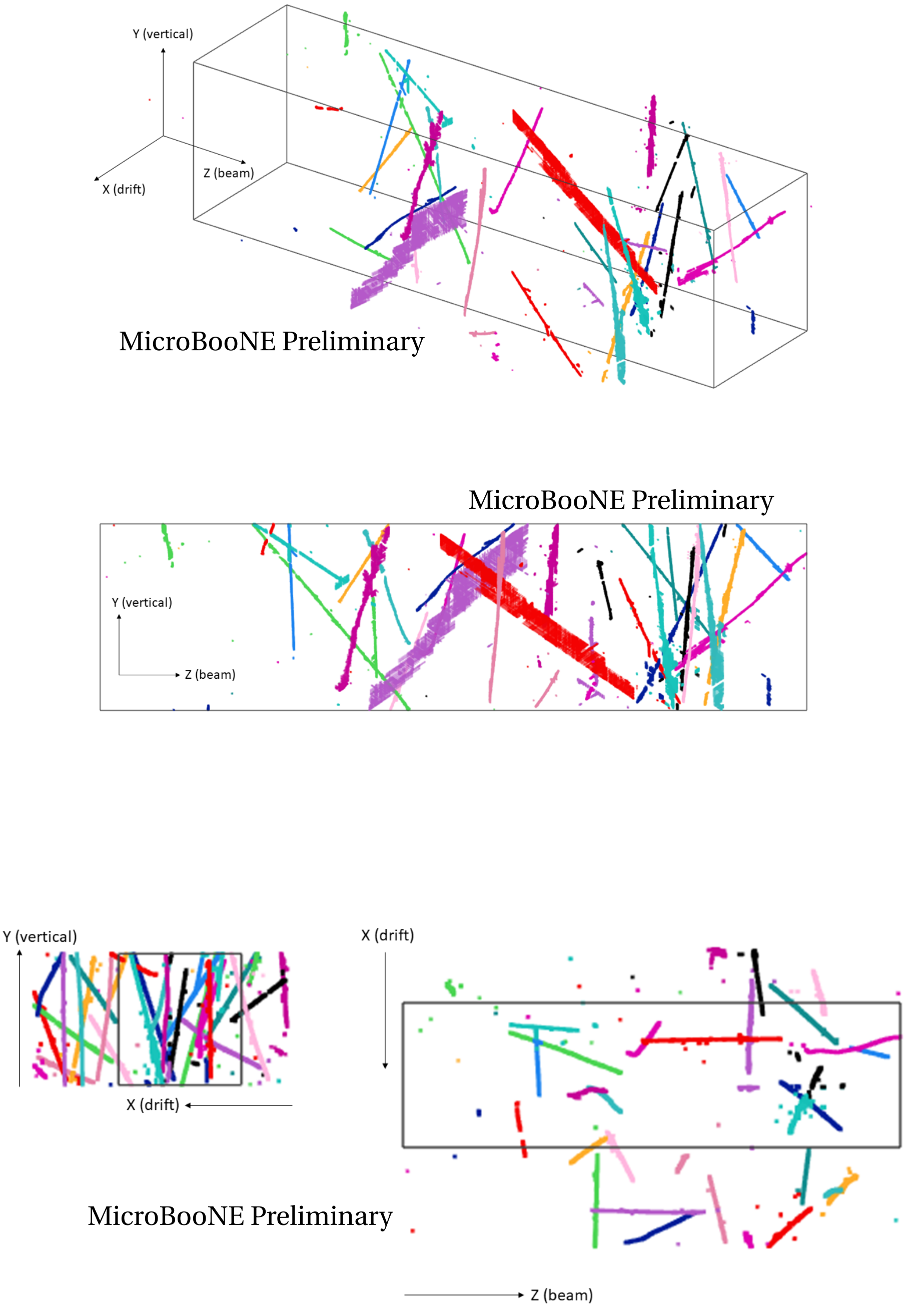


Figure 26: An example event showing the performance of the matching algorithm. Different clusters are represented by different colors. (Top) 3D view of the event. (Middle) Y-Z projection view. (Bottom left) Y-X projection. (Bottom right) X-Z projection. The original Bee link can be found at `https://www.phy.bnl.gov/twister/bee/set/ef826cc4-d428-479c-91a5-9c4e3497ef99/event/0/`.

PE prediction requires the usage of the "photon library", which is a pregenerated simulation. In this simulation, the entire detector is divided into 75x75x400 voxels. For each of the 32

PMTs, the visibility (or acceptance of an optical photon) in each voxel is recorded in the photon library.

The position of the TPC cluster(s) along the drift direction is determined by the matched PMT flash time. The pair of TPC cluster(s) and the PMT flash time is referred to as *bundle*. This step takes advantage of a unique feature of the tomographic imaging result — the event image with 3D charge information is reconstructed independent of the event topology information and prior to the application of pattern recognition techniques.

3.2.2 Iterative examination of bundles

At this stage, every bundle is examined to determine if the predicted PMT light distribution is highly consistent with the measured PMT light distribution. Level of consistency is based on the Kolmogorov-Smirnov test (K-S test) and the calculation of a chi-square. The K-S test focuses on the shape comparison of the predicted and measured PE distributions, which is less sensitive to the absolute normalization of the prediction. The inclusion of the K-S test is important, as the predicted light is made by the visible TPC activity inside the active volume. Although by definition the invisible TPC activity cannot generate visible TPC signals, the scintillation photons can be collected by the PMTs. In this case, one expects the measured PEs to be higher than the predicted PEs. The K-S test somewhat reduces the impact of this mismatch between prediction and measurement.

In the first round examination of all bundles, for bundles that clearly correspond to an incorrect match (for example, the predicted PEs are large, while the measured PEs are close to zero), the bundle is deleted. The highly consistent bundles are recorded for later usage.

After a first round examination of all bundles, the remaining bundles are categorized according to the TPC cluster. If one or more bundles correspond to the same TPC cluster, these bundles are identified as highly consistent and the other bundles are deleted.

In the final round of bundle examination, all the remaining bundles are categorized according to PMT flash. If one or more bundles corresponding to the same PMT flash are identified to be highly consistent, the other bundles are examined under the assumption that the highly consistent bundle is present. With this assumption, the predicted PEs from TPC clusters from different bundles are added together before being compared to the measured PEs to test their group compatibility. If a TPC cluster cannot pass any group compatibility tests, the bundle associated with this TPC cluster is deleted.

3.2.3 L1 Fits

The purpose of the previous step is to remove obviously mismatched bundles. The removal of these clearly incorrect bundles reduces the number of unknowns involved in the fitting procedure described next. The fitting procedure relies on a compressed sensing technique and takes into account all possible bundles simultaneously. In principle, we know a TPC cluster can only be associated with one PMT flash and one PMT flash can be associated with multiple TPC clusters. With this information in hand, we can write a chi-square function:

$$
\begin{aligned}
\chi^2 &= \sum_i \sum_j \chi^2_{ij} + \chi^2_{p1} + \chi^2_{p2} + \chi^2_{p3} && (11)\\
\chi^2_{ij} &= \frac{(M_{ij} - \sum_k a_{ik} \cdot P_{ikj})^2 - b_{ij} \cdot M_{ij}}{\delta M^2_{ij}} && (12)\\
\chi^2_{p1} &= \frac{(\sum_k a_{ik} - 1)^2}{0.01^2} && (13)\\
\chi^2_{p2} &= \frac{b^2_{ij}}{0.025^2} && (14)\\
\chi^2_{p3} &= \lambda \cdot \sum_i \sum_k |a_{ik}| && (15)
\end{aligned}
$$

The index i spans all flashes associated with the remaining bundles. The index j spans all PMTs. The index k spans all TPC clusters associated with the remaining bundles that are associated with the flash i. The M_{ij} and δM_{ij} represent the measured PE and its uncertainties, respectively. The P_{ikj} represents the predicted PE for a bundle associated with TPC cluster k and PMT flash i at PMT j. The a_{ik} are unknown parameters to be solved for during the fit. If $a_{ik} = 1$, the bundle associated with TPC cluster k and PMT flash i is a good match. If $a_{ik} = 0$, the bundle is a bad match. χ^2_{p1} applies the constraints that the total strength of each TPC cluster should be equal to unity. In reality, each TPC cluster should only be used once. The b_{ij} are also unknowns to be solved for in the fit. The b_{ij} term takes into account the possibility that some of the PMT flashes may not be associated with any visible TPC clusters. The PMTs can view activity outside the TPC active volume. The χ^2_{p2} gives the constraint that the b term should be close to zero. The final term in the chi-square, χ^2_{p3}, represents the application of the compressed sensing techniques, which prefers a best-fit solution where most of the a_{ik} are zero. The regularization strength λ is chosen to be 0.1. After this first round of fitting, the bundles with their associated a_{ik} less than 0.05 are deleted.

A second round fit is performed. The chi-square function is slightly altered:

$$\chi^2 = \sum_i \sum_j \chi^2_{ij} + \chi^2_{p1} + \chi^2_{p3} \tag{16}$$

$$\chi^2_{ij} = \frac{(M_{ij} - \sum_k a_{ik} \cdot P_{ikj})^2}{\delta M^2_{ij}} \tag{17}$$

$$\chi^2_{p1} = \frac{(\sum_k a_{ik} - 1)^2}{0.01^2} \tag{18}$$

$$\chi^2_{p3} = \lambda \cdot \sum_i \sum_k |a_{ik}| \tag{19}$$

The b term is removed in this fit. Once the second round fit is concluded, the bundles corresponding to the same TPC cluster are examined. The bundle that corresponds to the largest coefficient a_{ik} is retained; all others discarded. This bundle is treated as the final matched pair for this TPC cluster. The rest of bundles are removed.

Figure 27 shows an example of match light and charge signals. In this case, one PMT flash is matched to two sizable TPC clusters. The larger cluster is responsible for most of the observed PEs. The event topology of the larger cluster is identified to be a stopped muon with a Michel electron decay. This stopped muon appears to be fully contained when compared to the expected active volume (red box). This is due to the existence of space charge in the TPC bulk. For this particular event, there are 31 matched pairs. For the completeness, we show every matched pair in Fig. 28, Fig. 29, Fig. 30, Fig. 31, and Fig. 32. The excellent performance of the matching algorithm is self-evident from these event displays.

3.3 TPC Filter (Single-Click Hand-Scan)

The last step, hand-scan, aims to select ν_μ CC events. Backgrounds are categorized into four types: through-going muon, stopped muon, light-mismatch, and other. The primary aim of the hand-scan is to remove cosmic backgrounds. Dirt and/or ν NC events are expected to contaminate the ν_μ CC selection.

In the hand-scan step, the 3D Bee event display [20] was utilized. Instruction for how we use the Bee display to perform the hand-scan is provided in Appendix A. Several features are checked in the hand-scan step as criteria for ν_μ CC designation:

- **Light mismatch** Compare agreement between channel-wise predicted and measured PMT PE map.
- **Muon track** One or more tracks are expected for ν_μ CC selection.

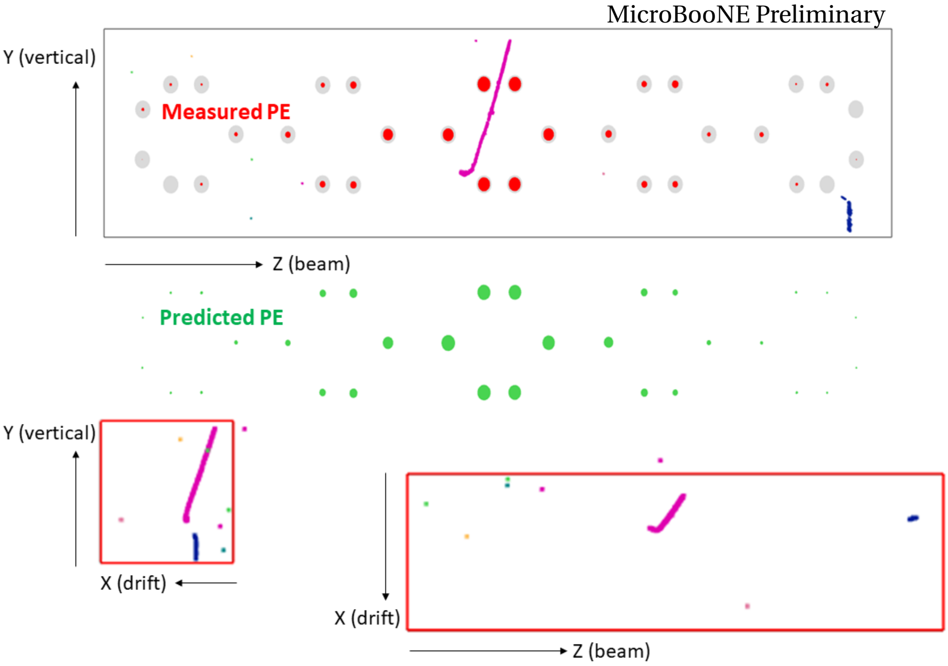


Figure 27: An example matched pair, in which the PMT flash is in coincidence with the BNB beam trigger. There are two relatively large TPC clusters that are matched to this flash. The larger cluster is a stopped muon with a Michel electron. The red solid circles in the top panel represent the measured PMT signals for this 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 black box represents the MicroBooNE detector active volume relative to the trigger time. The red box in the bottom panel represents the MicroBooNE detector active volume relative to the flash time. The matched TPC clusters should be fully contained inside the red box. For this event, since the flash time is in coincidence with the BNB beam trigger, which is close to the trigger time, the red box is effectively overlapped with the black box.

- **TPC boundary** The track will be checked if it is inferred to pierce any face of the TPC active volume. A space charge boundary is implemented in this step by eye. An uncertainty of several centimeters for this TPC boundary cut is anticipated. *No TPC fiducial volume cut was implemented, however, several cm uncertainty in hand-scan is expected.*
- **Directionality** If the tracks are not fully contained in the TPC active volume, directionality information is needed. The significant wiggling of the track at one end is used to determine the directionality of the track, especially for a long track. *No calorimetry information was used.*

In addition to ν_μ CC selection, the through-going muon and stopped muon could also be categorized using the TPC boundary and directionality information. More information could be used if an obvious Michel electron topology is present. The events which are neither ν_μ

Figure 28: The first set of 6 matched pairs for the same event as Fig. 27 (there are a total of 31 matched pairs in this event). The red solid circles in the top panel represent the measured PMT signals for this 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 black box represents the MicroBooNE detector active volume relative to the trigger time. The red box in the bottom panel represents the MicroBooNE detector active volume relative to the flash time. The matched TPC clusters should be fully contained inside the red box.

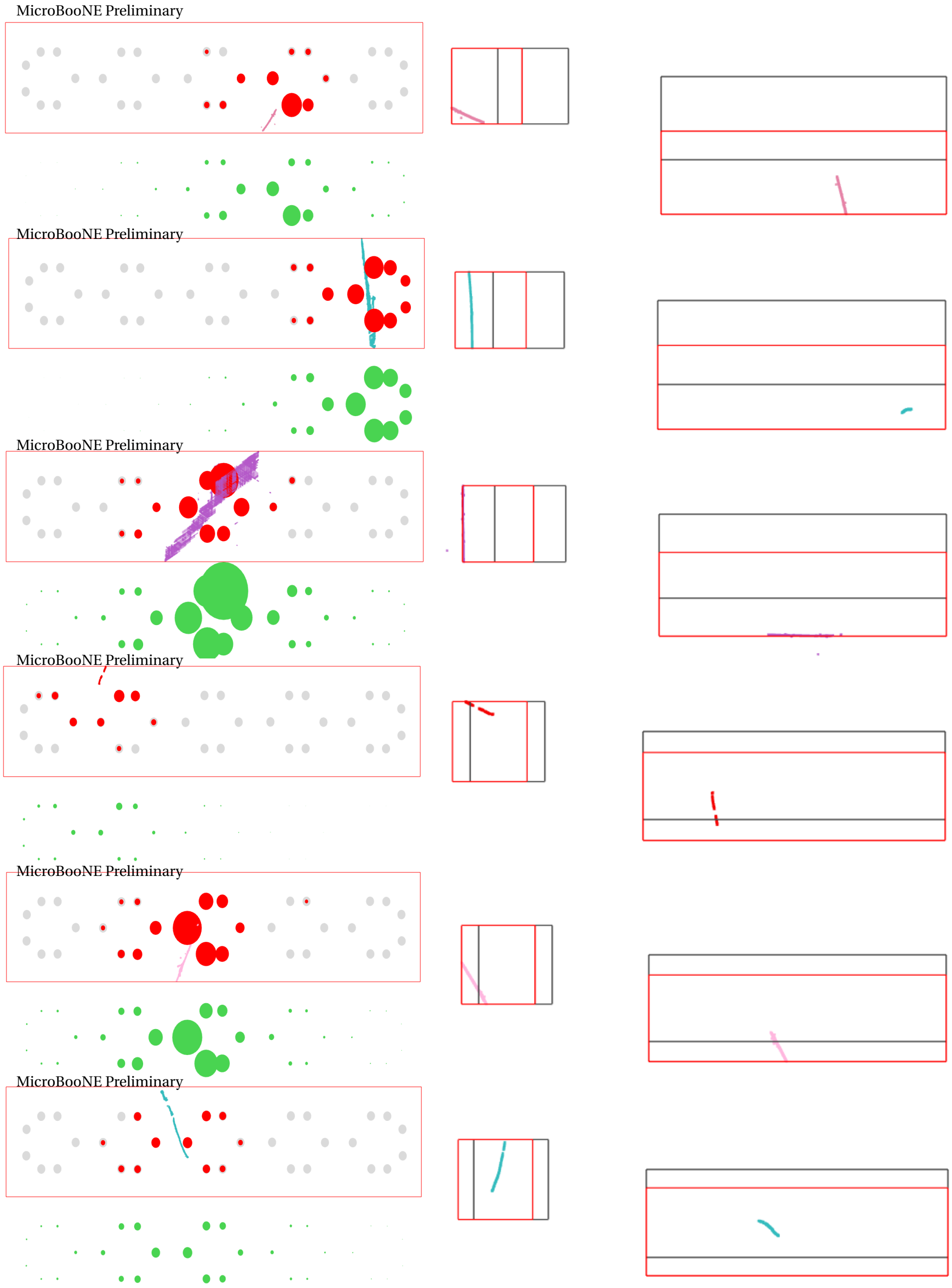


Figure 29: The second set of 6 matched pairs for the same event as Figs. 27 and 28 (there are a total of 31 matched pairs in this event). The red solid circles in the top panel represent the measured PMT signals for this flash. The area of the solid red circle is proportional to the measured PE. The green solid circles in the top panel represent the predicted PMT signals based on the two TPC clusters. The black box represents the MicroBooNE detector active volume relative to the trigger time. The red box in the bottom panel represents the MicroBooNE detector active volume relative to the flash time. The matched TPC clusters should be fully contained inside the red box.

CC candidates nor through-going muon, stopped muon, or light mismatch are categorized as

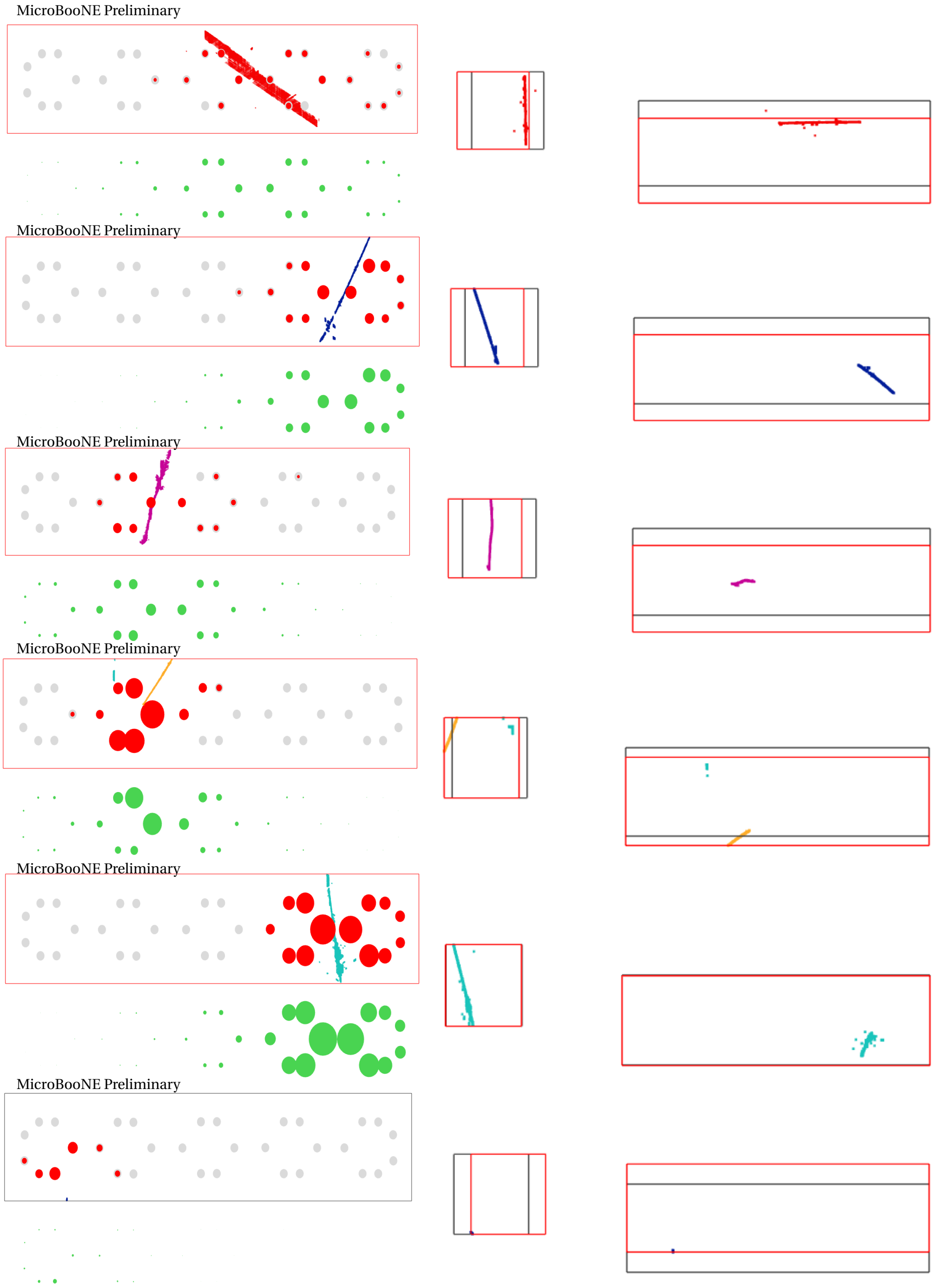


Figure 30: The third set of 6 matched pairs for the same event as Figs. 27, 28, and 29 (there are a total of 31 matched pairs in this event). The red solid circles in the top panel represent the measured PMT signals for this flash. The area of the solid red circle is proportional to the measured PE. The green solid circles in the top panel represent the predicted PMT signals based on the two TPC clusters. The black box represents the MicroBooNE detector active volume relative to the trigger time. The red box in the bottom panel represents the MicroBooNE detector active volume relative to the flash time. The matched TPC clusters should be fully contained inside the red box.

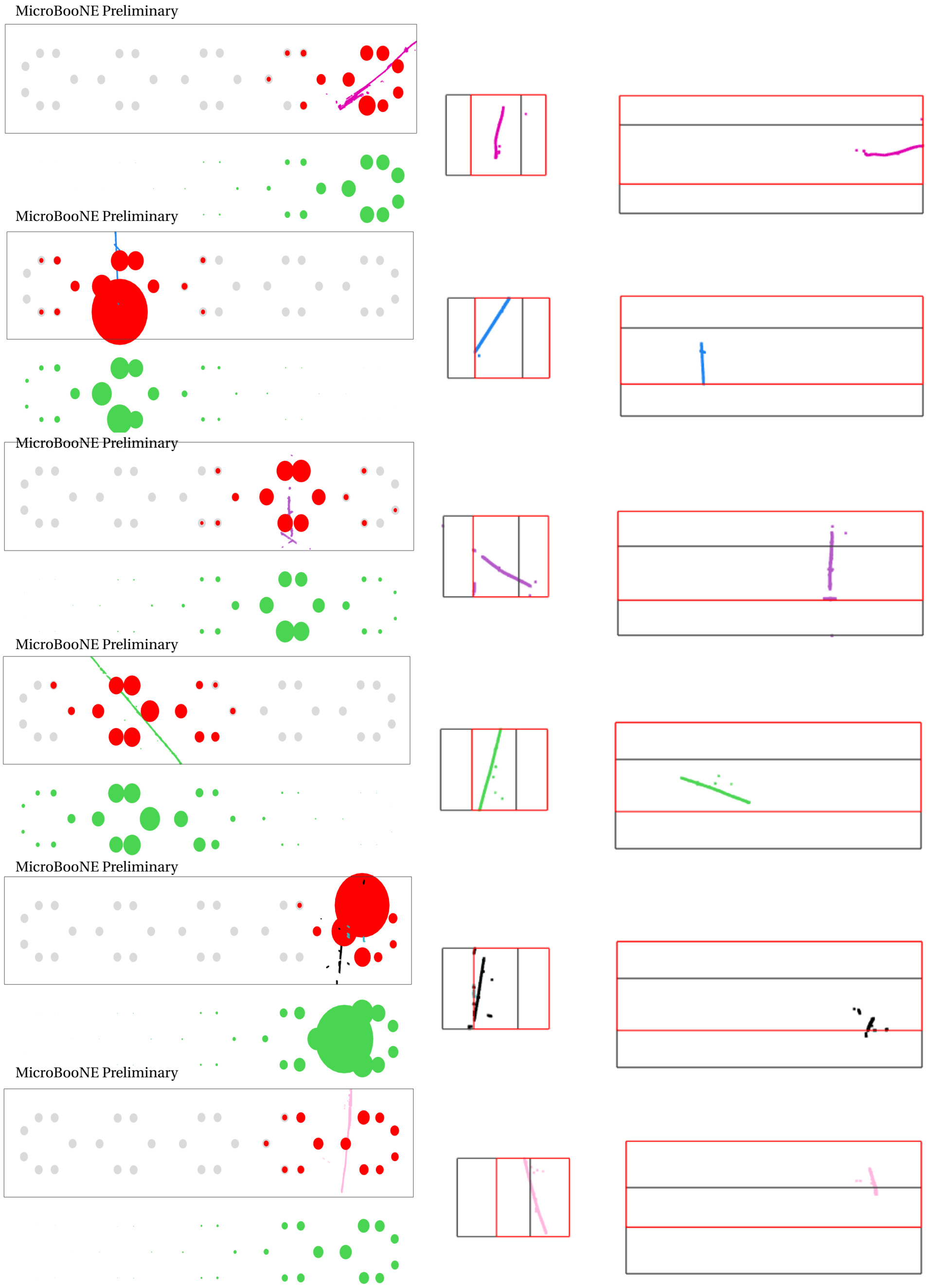


Figure 31: The fourth set of 6 matched pairs for the same event as Figs. 27, 28, 29, and 30 (there are a total of 31 matched pairs in this event). The red solid circles in the top panel represent the measured PMT signals for this flash. The area of the solid red circle is proportional to the measured PE. The green solid circles in the top panel represent the predicted PMT signals based on the two TPC clusters. The black box represents the MicroBooNE detector active volume relative to the trigger time. The red box in the bottom panel represents the MicroBooNE detector active volume relative to the flash time. The matched TPC clusters should be fully contained inside the red box.

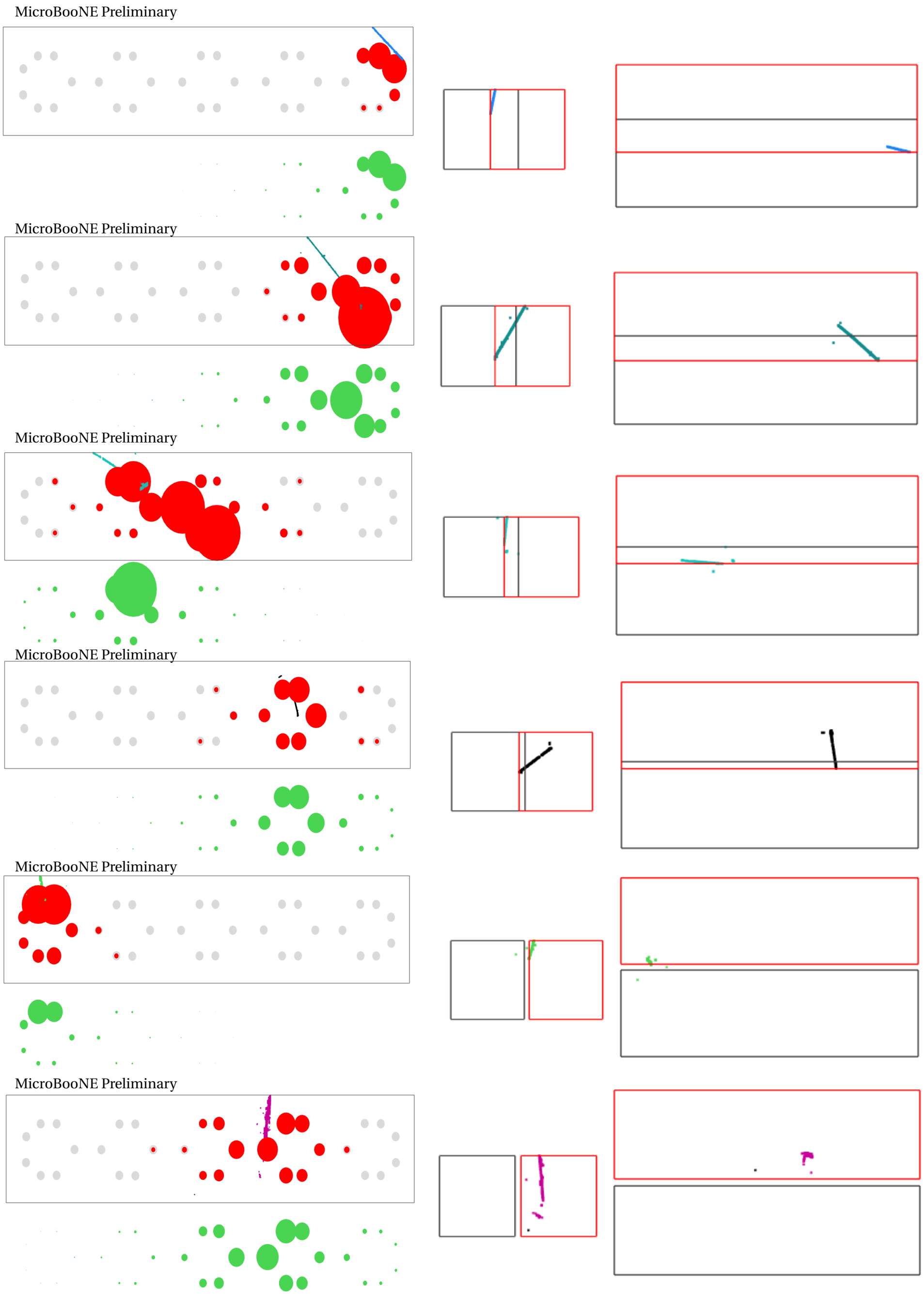


Figure 32: The fifth set of 6 matched pairs for the same event as Figs. 27, 28, 29, 30, and 31 (there are a total of 31 matched pairs in this event). The red solid circles in the top panel represent the measured PMT signals for this flash. The area of the solid red circle is proportional to the measured PE. The green solid circles in the top panel represent the predicted PMT signals based on the two TPC clusters. The black box represents the MicroBooNE detector active volume relative to the trigger time. The red box in the bottom panel represents the MicroBooNE detector active volume relative to the flash time. The matched TPC clusters should be fully contained inside the red box.

"other backgrounds", e.g. neutrino NC interactions, part of the ν_μ CC interactions very close to the boundary, and any other low visible energy events with very short tracks or dots.

Examples of fully contained ν_μ CC (Fig. 33) and partially contained ν_μ CC (Fig. 34), through-going muon (Fig. 35 in combination with the expected effective boundary due to the space charge), stopped muon (with Michel electron in Fig. 36, or wiggling at one end Fig. 37), and "light mismatch" Fig. 38. are shown below. The list of 1254 BNB events for hand-scan can be found at `https://www.phy.bnl.gov/~chao/uboone/scan-1/`.

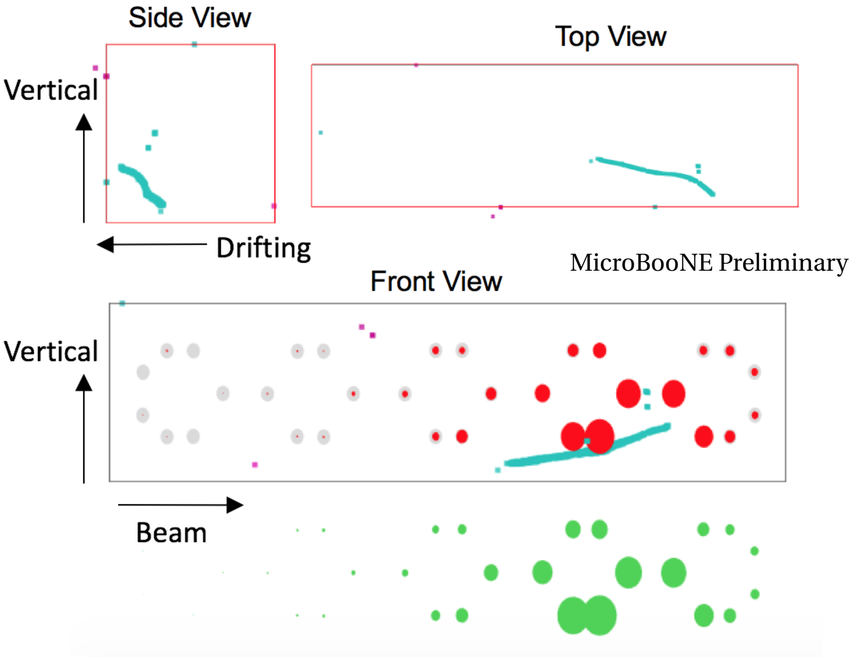


Figure 33: Fully contained ν_μ CC event. `https://www.phy.bnl.gov/twister/bee/set/7ab3a69b-09cf-473e-a220-33615b6f1e55/event/12/`

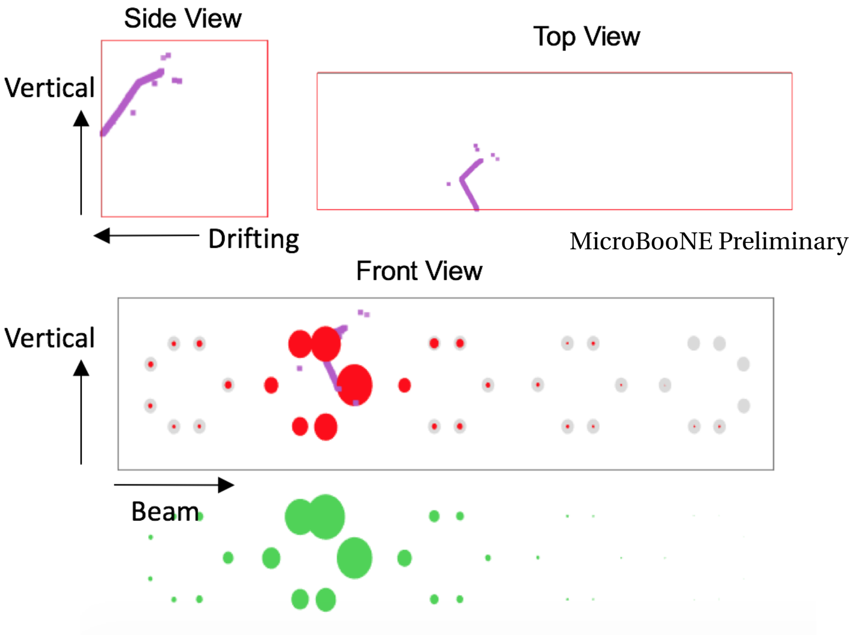


Figure 34: Partially contained ν_μ CC event. `https://www.phy.bnl.gov/twister/bee/set/e06d5db6-3e16-4e05-8787-8fc3fcd3927f/event/42/`, this is the second in-beam matched cluster within [3, 5] μs for this event.

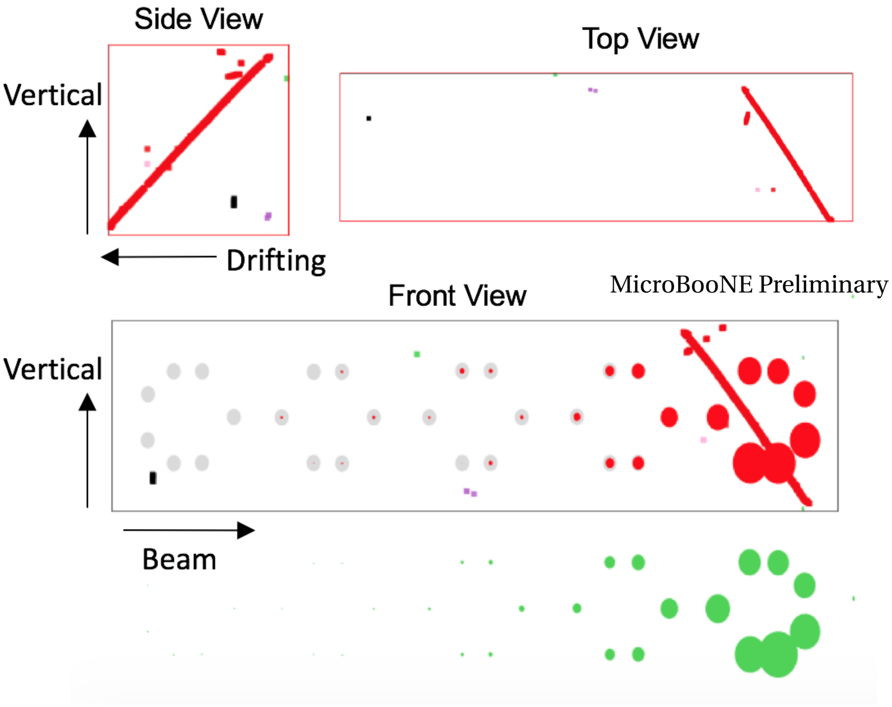


Figure 35: Through-going muon. The upper end in the side view is on the space charge boundary. `https://www.phy.bnl.gov/twister/bee/set/7ab3a69b-09cf-473e-a220-33615b6f1e55/event/67/`

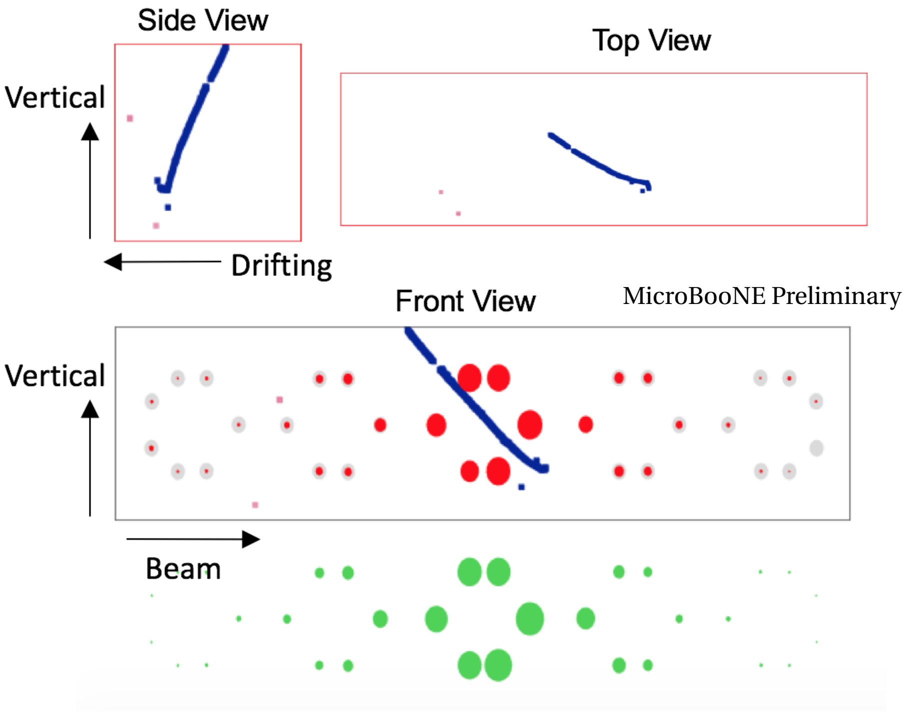


Figure 36: Stopped muon with a Michel electron. `https://www.phy.bnl.gov/twister/bee/set/7ab3a69b-09cf-473e-a220-33615b6f1e55/event/83/`

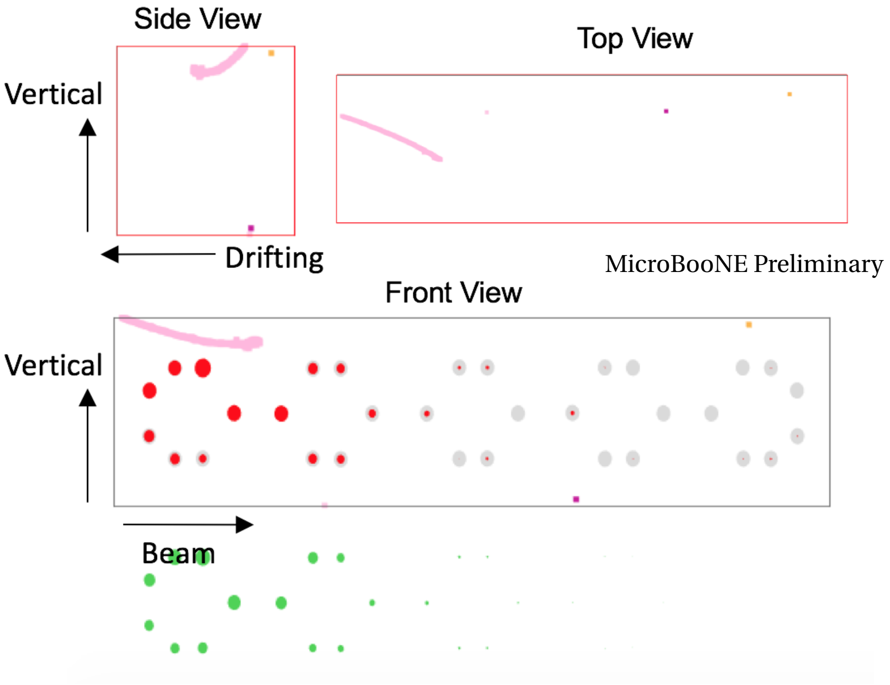


Figure 37: Stopped muon without Michel electron with a wiggling at the end inside the TPC. `https://www.phy.bnl.gov/twister/bee/set/7ab3a69b-09cf-473e-a220-33615b6f1e55/event/38/`

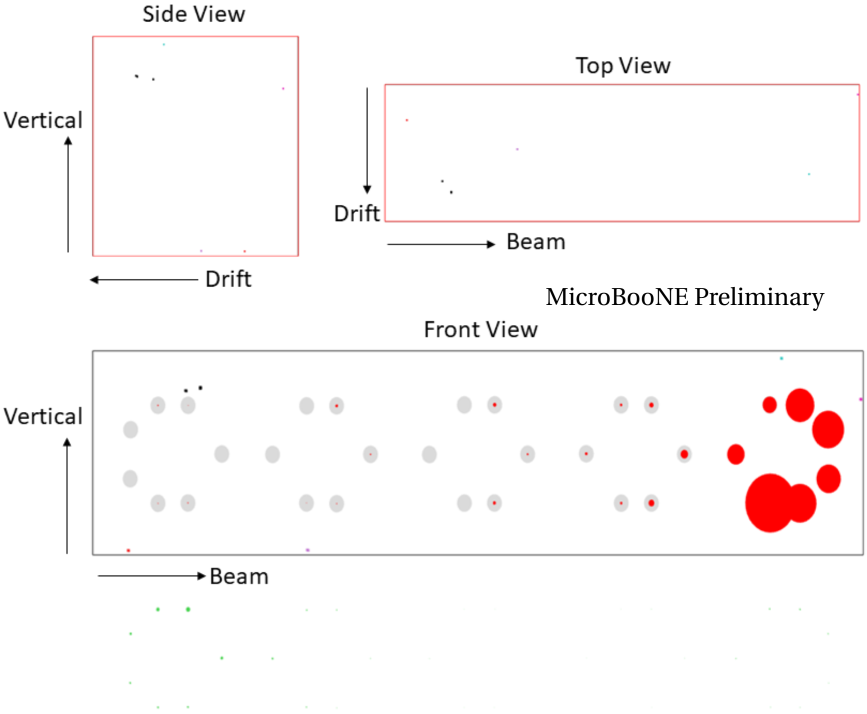


Figure 38: Example of an event categorized as "light mismatch". The many-to-many matching algorithm aggressively finds the best TPC cluster to explain the flash. The LAr visible by the PMTs is vastly larger than the TPC active volume. The instance of a "light mismatch" event is most likely indicative of a flash trigger corresponding to activity external to the TPC active volume. `https://www.phy.bnl.gov/twister/bee/set/7ab3a69b-09cf-473e-a220-33615b6f1e55/event/14/`

3.4 Performance

This section describes the ν_μ CC semi-automated scan result on the 3.96 k BNB and 3.15 k external BNB events.

The 3.96 k BNB events (beam on) are sampled from the MicroBooNE data (software trigger) collected from February 23, 2016, to April 17, 2016 (run numbers 5121 to 5946). The 3.15 k external BNB events are sampled from the data collected from February 23, 2016, to May 19, 2016 (run numbers 5121 to 6340). After many-to-many cluster-light matching, 1254 and 737 events survive in the BNB and external BNB samples, respectively.

Six scanners categorized the 1254 BNB and 737 external BNB events. The scanned results were cross-compared among scanners resulting in a convergence on event-wise type designation. A blind test (mixing 107 BNB events and 110 external BNB events in addition to the data samples used for neutrino selection) was preformed for each scanner to calibrate the false rate of misidentifying comics as neutrino events.

3.4.1 Semi-automated scan result

The categorization of events after matching for each trigger stream is provided in Table 5. Beside the candidates of charge-current ν_μ, which contains an energetic muon in the final states, there are four categories of backgrounds. More details regarding the description of each background can be found in Sec. 3.3.

Trigger stream	$\nu_\mu CC$	Light mismatch	Through going μ	Stopping μ	Other	Total
External BNB	0	187	415	95	40	737
BNB	113	356	560	171	54	1254
R^{Ext}	0	0.25	0.56	0.13	0.05	100%
R^{BNB}	0.09	0.28	0.45	0.14	0.04	100%

Table 5: Results of the semi-automated scan after event classification. For BNB data stream, the overall passing rate (the ratio of the selected candidates over initial software triggers) of the semi-automated ν_μ CC scan is $\frac{113}{3960} = 2.85\%$.

Cosmic background contamination The external BNB sample offered hand-scanners training to identify and characterize common cosmic backgrounds. The blind test (on a mixed subset of beam on and beam off events) showed that the false-positive rate of misidentification of cosmic as neutrino events is zero. However, given the finite statistics of this blind test sample (a 217 event sample of which 110 events were from the external BNB (beam off) sample), an upper limit of 1 event was given to estimate the false-positve rate which is 1%.

Dirt and/or ν NC background contamination Though the cosmic background is significantly reduced, dirt or ν NC interactions (both related to the neutrino beam) are expected to contaminate the $\nu_\mu CC$ candidates. Special topologies similar to $\nu_\mu CC$ interactions, e.g. π plus p and balanced transverse momentum or interactions with low visible energy, are likely ν NC backgrounds. A muon from a dirt $\nu_\mu CC$ event can enter and stop inside the detector, mimicking an outgoing muon from a $\nu_\mu CC$ interaction in the active TPC volume. For this circumstance, the directionality of the track is important to discriminate between signal and backgrounds. Therefore, for all the ν_μ CC neutrino candidates from the scanning, a second round examination was performed. For any suspicious neutrino candidates that were not fully contained or partially contained with obvious wiggling (directionality), additional plots[7] of 2D projected charge on each wire plane view were checked in order to search for a Bragg peak in any view to further determine the directionality. An example of the diagnostic plot is shown in Fig. 39 corresponding to the event in Fig. 36. As a result of the second round examination, we have identified 14 events out of the 113 selected ν_μ CC events to most likely be backgrounds. These events were not removed from the hand-scan results.

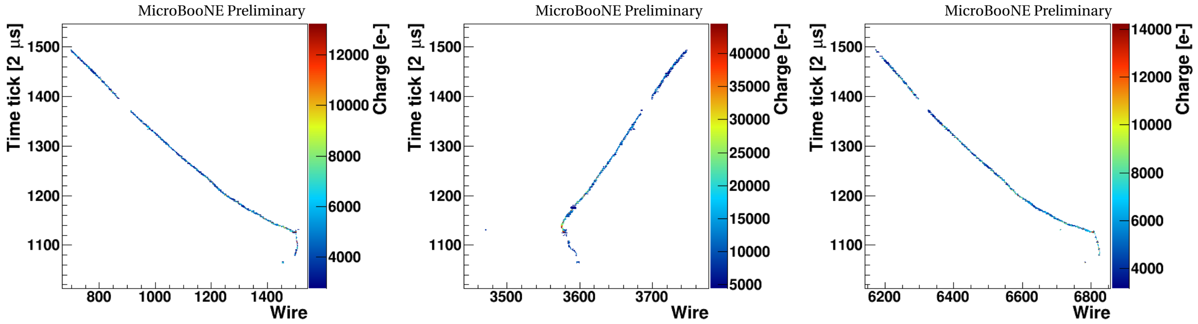


Figure 39: An example of the 2D charge plots for U, V, and Y plane view, respectively, from left to right. This is used to check potential false-positive neutrino candidates and estimate the background contamination (mainly from stopped muons). If we zoom in to these plots, the Bragg peak is obvious at the end of the muon track where the Michel electron is produced.

3.4.2 Cosmic rejection power

The cosmic rejection power is shown in Table 6. All the values were derived exclusively using the external BNB (beam off) sample. As mentioned above, 1% of the cosmics is expected to

---

[7] These plots are linked to the Bee display of each event and you can find them by clicking "Plots->Diagnostic plots" in the Bee display menu.

contaminate the neutrino selection after matching.

The number of cosmics per event is estimated to be ~26, corresponding to a cosmic ray rate of 5.5 kHz [21]. The "many-to-many matching" and "TPC filter" act on the TPC object level instead of the event level.

Procedure	Absolute Cosmic Rejection Power (Relative)
Hardware Trigger	1
Software Trigger	0.041
PMT filter*	9.6×10^{-3} (0.234)
Many-To-Many Matching	3.7×10^{-4} (0.038)
TPC Filter (Hand Scan)	$<3.7\times10^{-6}$ (<0.01)

Table 6: Cosmic rejection power at each step with our selected ν_μ CC candidates. The cosmic ray rate in the MicroBooNE detector is estimated to be 5.5 kHz [21], corresponding to ~26 cosmics per readout. The PMT filter* here includes a cut on events based on flash time and a cut on events after matching that do not have a matched TPC cluster. The many-to-many matching provides a one flash to one TPC activity relation, reducing the cosmic TPC activities by a factor of 26 ($\frac{1}{26}$ = 0.038). The hand scan cosmic rejection power, 0.01, comes from the blind test. Procedures for the last two rows of the table act on the TPC object level; preceding rows are procedures acting on the event level.

3.5 Automated cosmic tagger

We have developed a couple of simple automated tools to tag the light mismatch and through-going muon categorized events from out initial hand scan. The purpose of these automated tools were to (1) validate the numbers in the semi-automated hand scan and (2) as a first step towards a fully-automated scan. These tools were developed following the initial scan.

A *low-visible-energy tagger* was developed to cut on matched events with very small, dot-like clusters. These events comprise a portion of the "light mismatch" category from Table 5. No PMT light information was employed in the low-visible-energy tagger; however, there is a *strong* correlation with the light measurement far exceeding the predicted PE for these events. An example of a "light mistmatch" categorized event is provided in Fig. 38.

A second simple automated tool, the *through-going muon tagger*, was developed to identify track-like clusters and check if the two ends of identified clusters are at the TPC active volume boundary. Similar tools to those used in the clustering stage (Sec. 2.2) were employed in application of this tool. A cut is made on identified track-like clusters if cluster endpoints are within 3 cm of the space charge boundary. Evaluation of the effective detector

Trigger stream	Light mismatch	Through going μ	Non-excluded
R_{auto}^{Ext}	0.21	0.44	0.35
R_{auto}^{BNB}	0.22	0.31	0.47

Table 7: Automated tagging of light mismatch and through-going muon. “External BNB normalized” is the external BNB normalized to per BNB event passing matching. The ’non-excluded’ values are used to calculate the cosmic rejection power.

boundary due to space charge is provided in Appendix B.

Note that the simple automated tools described are preliminary and known to have inefficiencies. No dQ/dx information, multiple Coulomb scattering, or other advanced pattern recognition were used. Further development is ongoing for these tools as well as additional automated tools (not described here) for a *fully-automated* neutrino selection.

We applied the present tools (low-visible-energy tagger and through-going muon tagger) on the 3.96 k BNB (beam on) and the 3.15 k external BNB (beam off) data samples. The categorization of events with these simple automated tools is shown in Table 7. These automated tools have negligible effect on signal efficiency. Of the 113 selected ν_μ CC events from the hand scan, the simple automated tools tagged four previously identified background events (from the second round examination) and incorrectly tagged a bonafide ν_μ CC event as a background.

The current automated cosmic rejection power is 1.3×10^{-4} (derived from external BNB "Non-excluded" value of 0.35 as shown in Table 7). As noted above, these automated tools are preliminary implementations. For example, neither of the automated tools were designed to address the stopped muon background. Additionally, the existing automated tools have some inefficiencies, which can be seen by comparing individual categories of R^X in Table 5 to values in Table 7. Improving these automated tools is an area of active development. In particular, employing handles like calorimetry and multiple Coulomb scattering among other advanced pattern recognition tools are planned to further increase the purity of the neutrino candidate sample.

4 SUMMARY AND OUTLOOK

This note summarizes the principles underlying *Wire-Cell tomographic imaging* and describes application of this technique to MicroBooNE data. Described are the inherent benefits in reconstructing the TPC signal directly in 3D. The 3D image is faithfully consistent with each 2D projective measurement. Also, the 3D image is not casualty to the ambiguity of overlapping tracks and showers as in the projected image. These considerations vastly simplify downstream reconstruction (e.g. clustering) and neutrino selection mechanisms (e.g. many-to-many matching of TPC clusters to TPC flashes).

The ν_μ CC semi-automated scan result is an important step towards realizing the projected capability of high efficiency and high purity neutrino selection with single-phase LArTPCs. It lends credence to the efficacy of the new tools and techniques to reconstruct and analyze LArTPC data described in this note. Currently, the single-click hand-scan serves as a preliminary TPC filter, neither calorimetry nor advanced pattern recognition tools were utilized. The fully-automated TPC filter is under active development.

REFERENCES

[1] X. Qian, C. Zhang, B. Viren, and M. Diwan. Three-dimensional Imaging for Large LArTPCs. arXiv:1803.04850.

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

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

[4] C. Adams et al. Ionization Electron Signal Processing in Single Phase LArTPCs II. Data/Simulation Comparison and Performance in MicroBooNE. arXiv:1804.02583.

[5] Bruce Baller. Liquid argon TPC signal formation, signal processing and reconstruction techniques. *JINST*, 12(07):P07010, 2017, 1703.04024.

[6] W. Tang, X. Li, X. Qian, H. Wei, and C. Zhang. Data Unfolding with Wiener-SVD Method. *JINST*, 12(10):P10002, 2017, 1705.03568.

[7] H. Nyquist. Thermal Agitation of Electric Charge in Conductors. *Phys. Rev.*, 32:110–113, Jul 1928.

[8] B. K. Natarajan. *SIAM J. Comput.*, **24**:227, 1995.

[9] E. J. Candès, J. K. Romberg, and T. Tao. *Comm. Pure Appl. Math.*, 59:1207, 2006.

[10] D. L. Donoho. *IEEE Trans. Info. Theory*, **52**:1289, 2006.

[11] H. Yu and G. Wang. *Physics in Medicine and Biology*, **54**:2791, 2009.

[12] M. Yuan and Y. Lin. *J. Roy. Stat. Scocienty. Series B (Statistical Methodology)*, **68**:49, 2006.

[13] R. Tibshirani. *J. Roy. Stat. Society. Series B (Statistical Methodological)*, **58**:267, 1996.

[14] S.S. Chen, D.L. Donoho, and M.A. Saunders. *SIAM J. Sci. Comput.*, **20**:33, 1998.

[15] Hastie T Friedman J and Tibshirani R. *Journal of statistical software*, 33:1, 2010.

[16] Jose Luis Blanco and Pranjal Kumar Rai. nanoflann: a C++ header-only fork of FLANN, a library for nearest neighbor (NN) wih kd-trees. `https://github.com/jlblancoc/nanoflann`, 2014.

[17] Wire-cell proto type code. `https://github.com/BNLIF/wire-cell-2dtoy`, 2018.

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

[19] Boost c++ libraries. `http://www.boost.org/`.

[20] `https://www.phy.bnl.gov/wire-cell/bee/`.

[21] R. Acciari et al. Measurement of cosmic-ray reconstruction efficiencies in the MicroBooNE LArTPC using a small external cosmic-ray counter. 2017, 1707.09903.

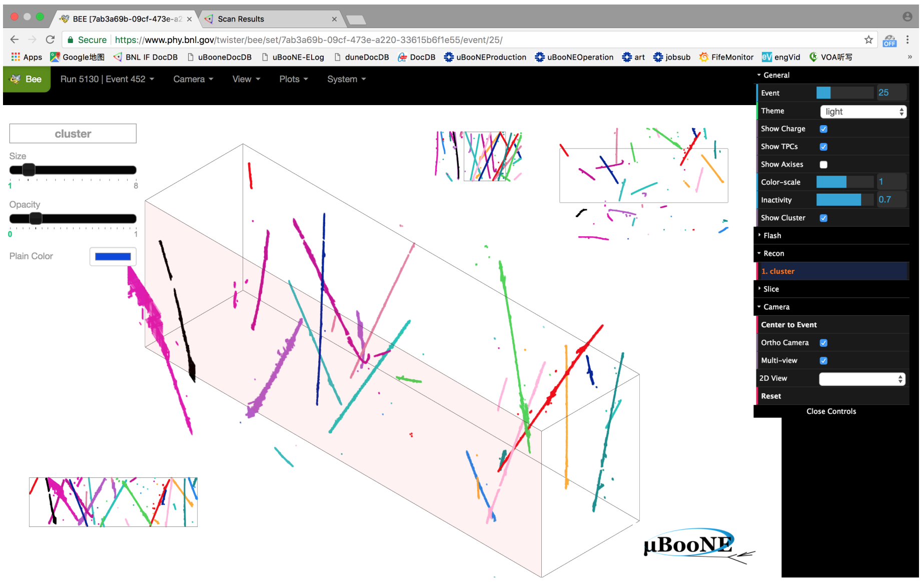
[22] Michael Mooney. The MicroBooNE Experiment and the Impact of Space Charge Effects. In *Proceedings, Meeting of the APS Division of Particles and Fields (DPF 2015): Ann Arbor, Michigan, USA, 4-8 Aug 2015*, 2015, 1511.01563.

[23] `https://lar.bnl.gov/properties/spacecharge.html`.

A INSTRUCTION OF HAND SCAN AND GALLERY OF SELECTED ν_μ CC CANDIDATES

All the 1254 BNB events after matching filter are the candidates of hand-scan. The 3D bee display for these events can be found at at `https://www.phy.bnl.gov/~chao/uboone/scan-1/`. The categorization is indicated. The hand scan can be performed with a single key-stroke in Bee. One can do these scans very efficiently and accurately with some training.

An instruction of the one-key hand-scan will be described step by step in the following. **Step 1:** Click on the bee link, and one will see the web interface as shown in Fig. 40. The 3D



Secure https://www.phy.bnl.gov/twister/bee/set/7ab3a69b-09cf-473e-a220-33615b6f1e55/event/25/

Event 25
Theme light
Show Charge ☑
Show TPCs ☑
Show Axises ☐
Color-scale 1
Inactivity 0.7
Show Cluster ☑
Ortho Camera ☑
Multi-view ☑
2D View

Figure 40: Bee display interface for hand-scan.

view of the MicroBooNE TPC is shown as well as three projective views (front, side, and top). One can move around the 3D view by the mouse.

Step 2: use '/' to find the in-beam matched cluster(s) and flashes as shown in Fig. 41. On the anode plane of the 3D view, 32 circles represent the 32 PMTs for light collection. The red circle areas represent the measured PE (photon electron) and the green circle areas represent the predicted PE based on the matched cluster. There maybe more than one in-time flash, press '/' a few times will cycle through all of them. The bottom right corner shows the time and the total PE for the matched flash.

Step 3: check from different angle for the matching cluster by 'z', 'x', 'c', or with the mouse. This is important to categorize the events. Keep in mind the space charge effect (best seen in the x-y view). For example, use 'x' to get the interactive side view (drifting-vertical view) as shown in Fig. 42. The criteria of neutrino selection are described in Sec. 3.4, please have a try. Enjoy!

Note: if one wants to see other matched flashes, use ',' or '.' to navigate forward or backward. If one wants to see all the other flashes, use 'shift'+',' or 'shift'+'.' to navigate. To see everything, on the right panel, click Flash, un-check "Matching cluster". One can go to next event ('shift+n') or previous event ('shift+p'). Don't forget to press 'Esc' if any key doesn't seem to work.

Diagnostic plots: to make an estimation of the hand-scan purity, the cleaned-up (get rid of all the other clusters) 2D projected charge is plotted for each wire plane view if the cluster is matched with any in-beam flash. Although this step is not involved in our current

Event 25
Theme light
Show Charge ☑
Show TPCs ☑
Show Axises ☐
Color-scale 1
Inactivity 0.7
Show Cluster ☑
Ortho Camera ☑
Multi-view ☑
2D View

#36: (4.02 us, 2015.75 pe)
matching: 80,7

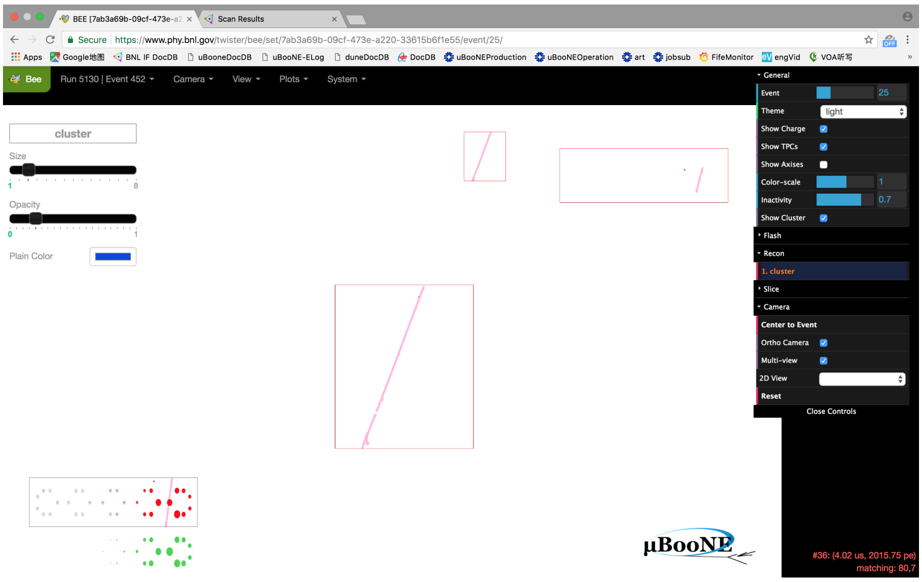
Figure 41: Bee display of matched cluster(s) with a in-beam flash.

hand-scan, one can find these plots by clicking 'Plots->Diagnostic plots' in the menu on the top.

B EFFECTIVE DETECTOR BOUNDARY

For a surface running LArTPC like MicroBooNE, cosmic muons are one of the major backgrounds. Due to the relatively slow drifting velocity of ionization electrons, there could be 20-30 cosmic muons within each triggered readout. Some of them can go through the active volume, some of them can stop inside the detector. The major difference between a neutrino interaction which occurred inside the active volume and a cosmic muon background is that the track of a cosmic muon backgrounds is initiated outside and travel into the active volume. On the other hand, the tracks produced by neutrino interactions are initiated inside the active volume and travel towards outside the detector. While some of the tracks travel out of the active volume (also referred to as partially contained), some tracks can stop inside (also referred to as fully contained). Therefore, the relative position between the reconstructed track trajectories and the detector boundary is a crucial piece of information in separating cosmic backgrounds and neutrino interactions. In other words, the precise knowledge of the boundary of the active volume is important.

There are two effects that can lead to distortions in the effective boundary of the active volume. The first one is the space charge effect. While the ionization electrons are drifting


https://www.phy.bnl.gov/twister/bee/set/7ab3a69b-09cf-473e-a220-33615b6f1e55/event/25/

Run 5130 | Event 452

Event 25

Theme light

Show Charge ☑

Show TPCs ☑

Show Axises ☐

Color-scale 1

Inactivity 0.7

Show Cluster ☑

Ortho Camera ☑

Multi-view ☑

#36: (4.02 us, 2015.75 pe)
matching: 80,7

Figure 42: Bee display from different views to do the hand-scan.

towards the anode plane, the resulting positive charged ions are drifting towards the cathode plane. Since the mass of the ion is much larger than that of the electron, the drifting velocity of the ion is about five orders of magnitude slower than that of the electron. It could take several minutes for an ion to travel a full drift distance. For surface running single-phase LArTPCs like MicroBooNE, a constant source of positive charged ions is built up in the bulk of the TPC due to the traversal of cosmic muons. Therefore, a large amount positive charge is accumulated inside the active volume. The amount of charge can be large enough to distort the drifting electric fields. A more detailed discussion of the space charge effect can be found at Ref. [22, 23].

Intuitively, the impact of the space charge on the effective detector boundary can be easily understood. As ionization electrons drift towards the anode plane, they are going to be pulled by the positive charged ions present inside the active volume. Therefore, as they drift through the active volume, they are going to be pulled towards the center of the detector. As a result, the reconstructed position in the Y-Z plane, which represents the position where the ionization electrons reach the anode plane, appears to be squeezed to the detector center compared to its original location. The longer the ionization electron drifts inside the active volume, the more pronounced this effect becomes. Therefore, the effective detector boundary in terms of reconstructed position appears to be smaller than the naive expectation (rectangular box) of the active volume.

In order to understand the location of the effective detector boundary, we took the TPC cluster/PMT flash matched results after implementation of the many-to-many TPC cluster/PMT flash matching algorithm (discussed in Sec. 3.2). The assigned flash time is used to correct the TPC cluster drift distance. In order to enhance the purity of the sample (correctly matched pairs) and reduce the distortions in the distribution due to readout effects and other reconstruction inefficiencies, the following cuts are applied:

- Small clusters associated with larger clusters are removed.
- Clusters that bookend the TPC readout (as in close in proximity to the beginning and ending portions of the readout window are removed).
- A rigorous Kolmogorov-Smirnov distance test and a chi-square are evaluated to remove matched pairs that are not high-confidence assured.

For this study, since we mainly care about the detector boundary, the absolute efficiency of the selection is not required to be high. Figure 43 shows the results for one example event after applying these cuts. More events can be found at `https://www.phy.bnl.gov/twister/bee/set/ee3d70e4-186f-4523-b945-c0d755fc9049/event/list/`.

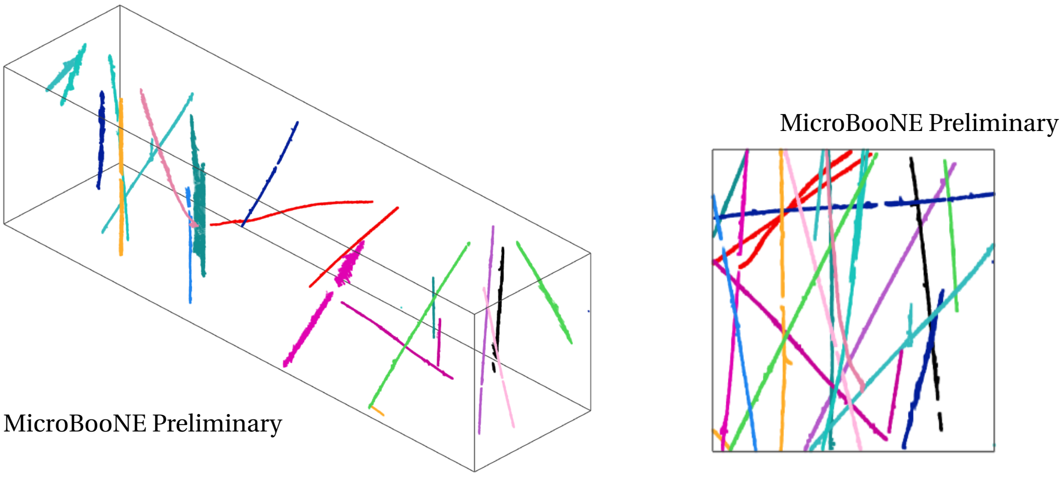


Figure 43: Taken from `https://www.phy.bnl.gov/twister/bee/set/ee3d70e4-186f-4523-b945-c0d755fc9049/event/4/`. Left and right panel shows different view angles. In particular, the right panel shows the projection in the Y-X plane with Y and X being vertical and drift direction. At large drift distance (right side along the X), gaps can be clearly observed at both top and bottom corners. These gaps are the natural results of the existence of space charge in the detector.

In order to map out the effective detector boundaries, about 1700 events are processed. The space points of the position-corrected tracks are projected into the X-Y and X-Z planes. Figure 44 shows the zoomed-in view of the four corners at large drift distance. A deficit of the space points in large X and high Y corresponding to the top corner is clearly visible. Similar deficits can be seen at large X and low Y (bottom corner), large X and high Z (downstream corner), and large X and low Z (upstream corner). The red lines in Fig. 44 represents a proposed cut position. We should also note that the effective detector boundary is in principle a 2D surface in a 3D space. It is not necessarily factorized into two 1D cuts as proposed in Sec. 44. Figure 45 shows the locations of the effective boundary in the X-Y plane as a function of different Z locations. There are hints of Z position-dependence in the X-Y effective boundary. The proposed position cuts (red lines) are also shown in the plot. For the purpose of rejecting cosmics, the principal concern is that the proposed cut is within the actual detector boundary. For this point of view, the proposed cut can satisfy the requirement. However, a mismatch between the cut and the actual detector boundary would lead to an uncertainty in the estimation of target mass. A more careful study of the 2D effective detector boundary is needed. Development on this is on-going.

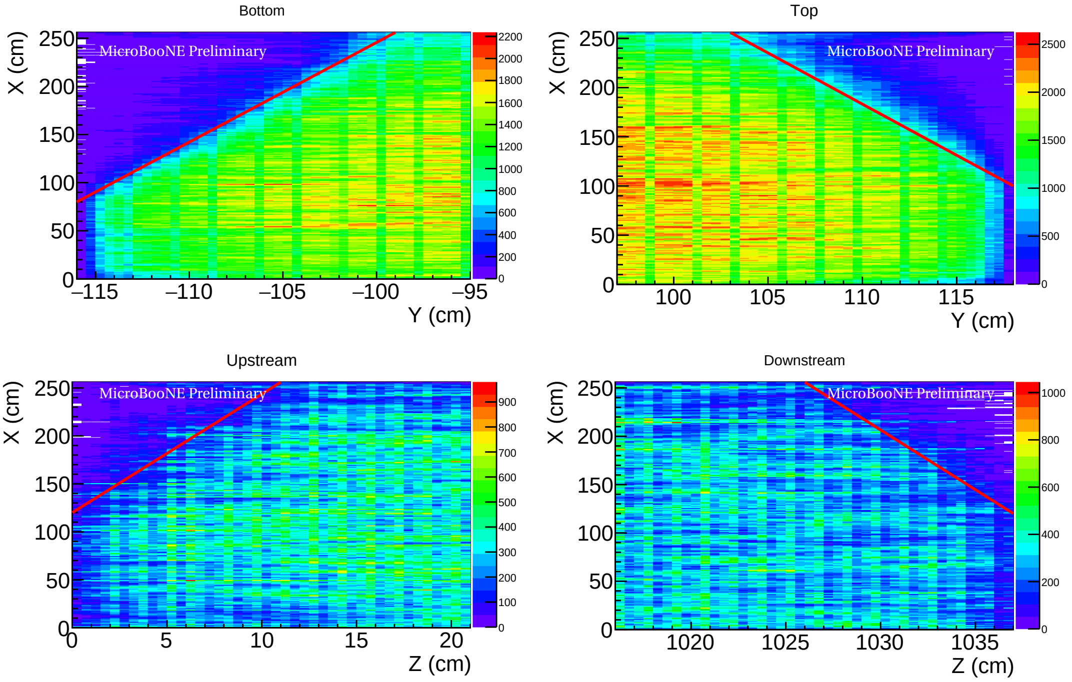


Figure 44: Illustration of the locations of effective boundaries at four corners. The location of the red lines illustrate the proposed cut position. See text for more discussions.

The second detector effective boundary effect originates from the existence of many unresponsive channels at MicroBooNE. As shown in Fig. 2 and Fig. 12, the known unresponsive channels lead to unresponsive regions in the Y-Z plane, which are extended into the active volume along the X (drift) direction. When a cosmic track is entering into the active volume

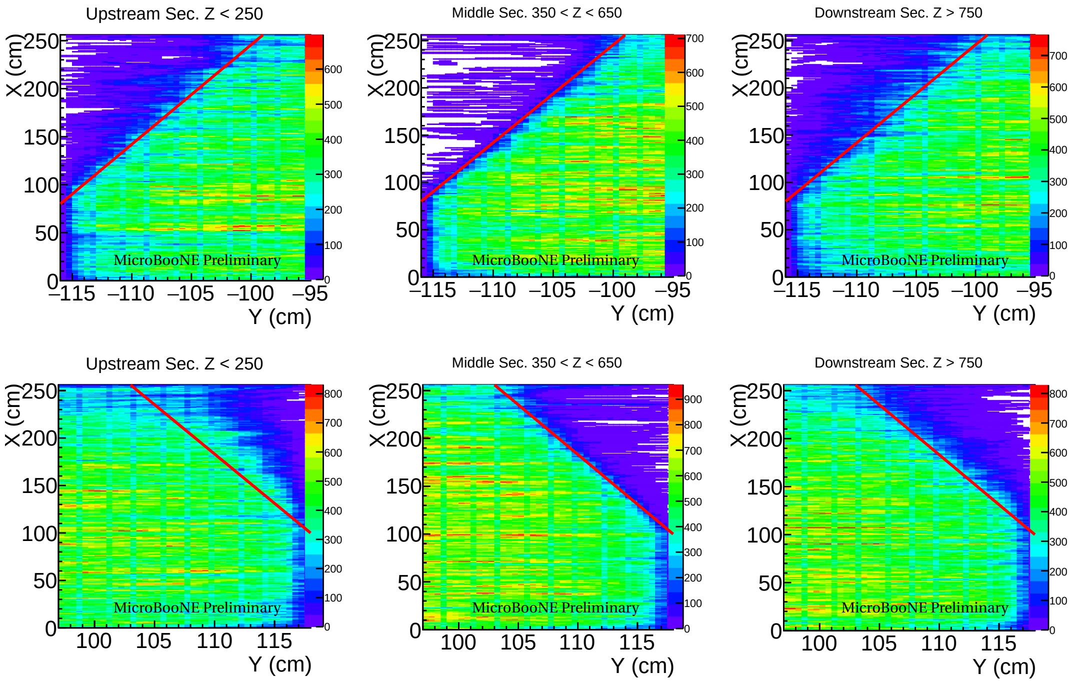


Figure 45: Illustration of the locations of effective boundary in the X-Y plane as a function of different Z locations. The location of the red lines illustrate the proposed cut position.

from one of these unresponsive regions. The initial part of the track inside the active volume would be shadowed by the existence of unresponsive region and not visible. In this case, the reconstructed cosmic muon track trajectory *could mimic* a contained neutrino interaction track. Fortunately, since the unresponsive channels are precisely known [2], the locations of the unresponsive regions are also precisely known. Therefore, a cut can be developed to deal with this effect. The detailed development of this cut is ongoing.

C NEUTRINO INDUCED SHOWER EVENTS

In Sec. 3.4, we described a semi-automatic event selection for charge-current (CC) ν_μ interaction. This exercise significantly improved our understanding on various background sources for CC ν_μ interactions. Since the primary physics goal of MicroBooNE is to search for anomalous CC ν_e interactions, a similar exercise should be conducted to study the backgrounds for neutrino-induced shower events. The details of this exercise is beyond the scope of the current technote and will be reported in a future technote. In this appendix, we display some potential candidates.

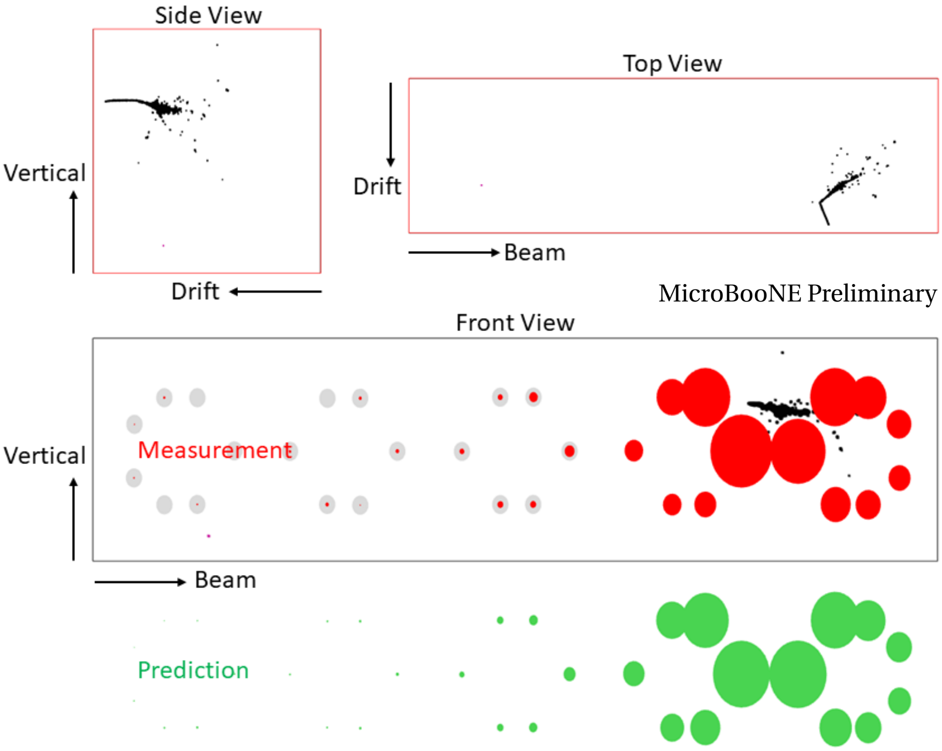


Figure 46: ν_e candidate `https://www.phy.bnl.gov/wire-cell/bee/set/8ce87fc2-ecd0-49c2-b227-2deec9a6a566/event/122/`

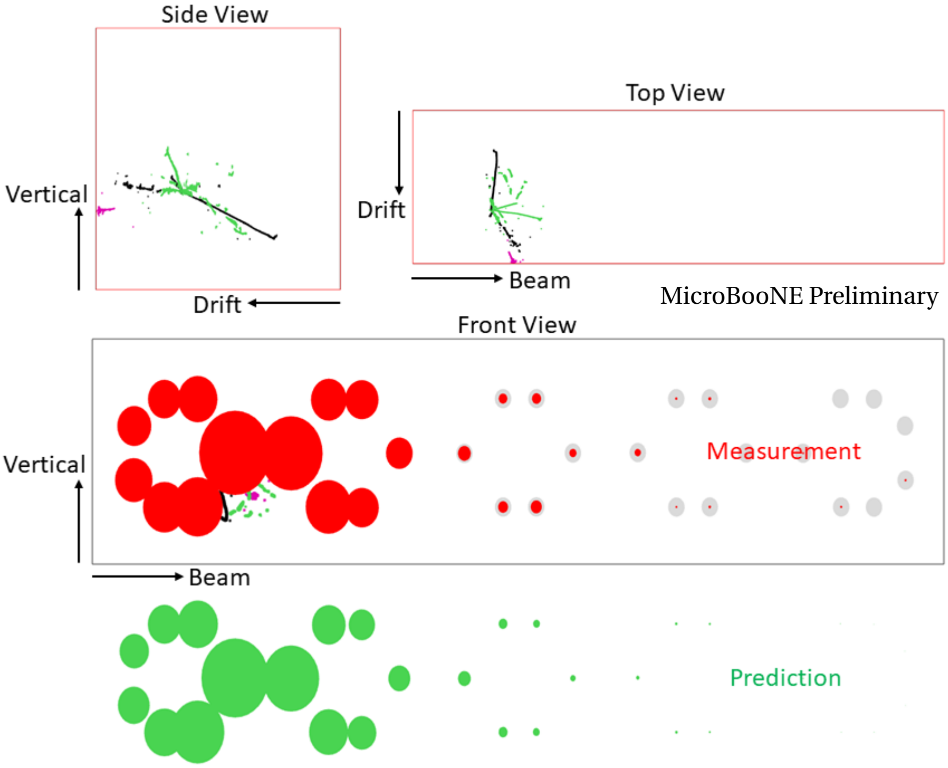


Figure 47: Multiple π^0 candidate `https://www.phy.bnl.gov/wire-cell/bee/set/8ce87fc2-ecd0-49c2-b227-2deec9a6a566/event/58/`

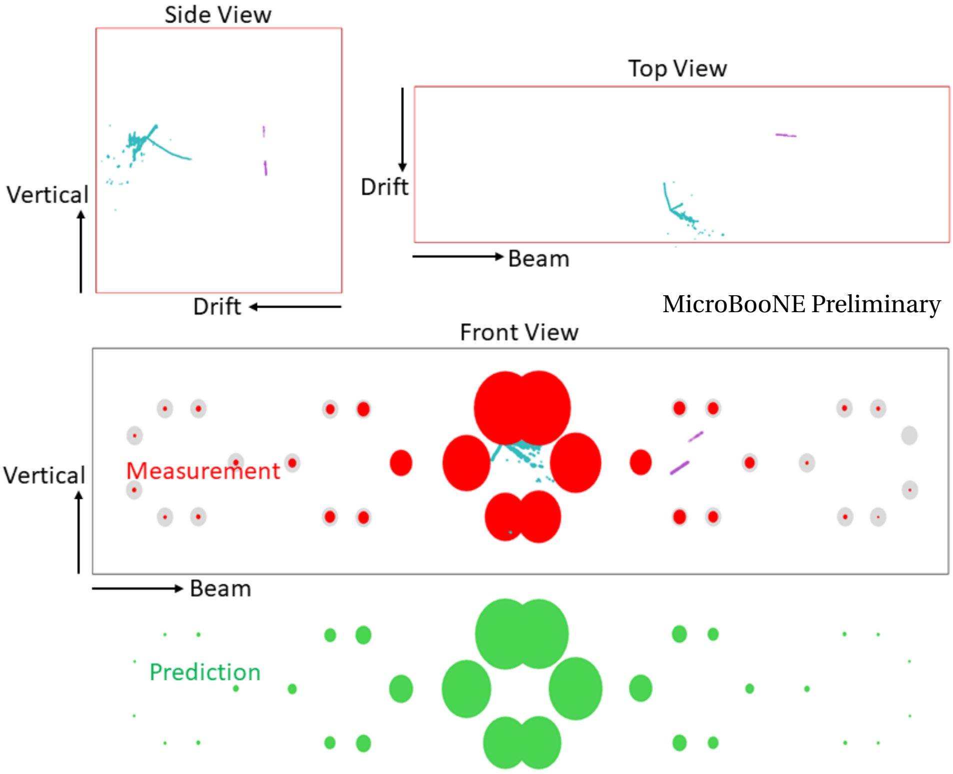


Figure 48: e^+e^- candidate `https://www.phy.bnl.gov/wire-cell/bee/set/8122be11-3423-4b7e-95fd-08c6e9791079/event/194/`