Finding first neutrino events with MicroBooNE

Ben Carls[1], Jessica Esquivel[2], Andy Furmanski[3], Ryan Grosso[4], Ariana Hackenburg[5], Wesley Ketchum[1], Anne Schukraft[1], Kazu Terao[6], Tracy Usher[7], and Katherine Woodruff[8]

[1]FNAL
[2]Syracuse University
[3]University of Manchester
[4]University of Cincinnati
[5]Yale University
[6]Nevis Laboratories, Columbia University
[7]SLAC
[8]New Mexico State University

September 25, 2015

Abstract

We present an analysis path for identifying a highly-purified selection of neutrino interactions from our large cosmic ray backgrounds using the first hours of BNB data from MicroBooNE. This selection is aimed at providing events that can be shown in our approved event displays and released publicly in a timeline manner. Since event rates for 5 Hz running of the BNB are large and cosmic background events will outnumber neutrino interactions by far, an automated procedure is necessary to select neutrino candidate events. We show this can be achieved based on the timing information of flashes observed by the PMT system and topology of events reconstructed in 2D and 3D from TPC information. Based on these studies, we believe an automated selection that will provide us with a rate of O(100) events per day (assuming nominal rate and POT per pulse from the BNB), where the cosmic background contamination is on the order of 20%. We will use the procedures developed in this analysis to further study expected signal and background rates from data and simulation, and apply the filters described here on our first data.

*University of Manchester
†University of Chicago

1 Introduction

1.1 Goal

The goal of this analysis is to positively identify Booster Neutrino Beamline (BNB) neutrino interaction events in the MicroBooNE detector in data collected during the first day(s) of running. Events will need to be (1) identified as likely neutrino interaction events, and (2) separated from the large "empty-beam-spill" and "cosmic-coincident-beam-spill" backgrounds. The former will be done by identifying events with a detected flash of scintillation light during the 1.6 μs beam-spill length of the BNB, and by identifying reconstructed objects from the TPC that are consistent with being from a neutrino interaction. The latter will be done by rejecting events that do not have flashes of scintillation light during the beam-spill period, and/or that lack clearly identifiable neutrino-interaction-like topologies in the TPC reconstruction. After selection, events will be cross-checked in 2D and 3D event displays for verification of the selection performance, and to determine which event displays will be part of a near-term public release.

The selections defined here are targeted to reduce the ratio of readout events containing a neutrino interaction to readout events that contain only cosmic backgrounds from an initial ratio of 1 neutrino to 675 backgrounds to a ratio of 1 to 0.5 or better (rates are explained in the following subsection). This is equivalent to reducing the background of cosmic events by a factor 1000 or more. At this level, the total number of events passing selection will be in the few hundreds per day. From that point, if further separation of neturino events from cosmic-only background events is necessary, events from the first days of data-taking that pass these automated filters can be confirmed and distinguished from cosmic events by studying event displays.

The selection of events from this analysis are intended to be used for MicroBooNE's first public displays of neutrino interactions. Therefore, we want to select events which are clearly identifiable as a neutrino interaction by not only MicroBooNE collaborators, but also colleagues from other experiments and the public. These events should show a clear interaction vertex with at least two tracks originating from it, and a flash of scintillation light that is coincident with the BNB beam-spill period. This analysis is not optimized for obtaining a high-efficiency or ultra-high-purity sample of neutrinos, which may be determined in later studies.

1.2 Event rates

At the start of this report, we are introducing the following definitions that should help the understanding of the studies presented below.

- **Readout event**: an event of 4.8 ms length that gets read out whenever we receive an accelerator timing signal. The trigger for the readout occurs at 1.6 ms (3200 2-MHz samples) into the readout event time. The readout event may or may not contain a neutrino interaction.
- **Neutrino event**: a readout event that contains a neutrino interaction.
- **Cosmic event**: a readout event that does not contain a neutrino interaction but only cosmic background tracks and showers.

Drift HV (kV)	Drift time (ms)	CRY sim (muons)	CORSIKA sim (muons)	Data (handscan)	Data (reco tracks)
−58	2.2	8.3	12.1		
−70	2.0	7.6	11.0	8.8	7.3

Table 1: Expected cosmic rates per readout window. These numbers come from studies presented in DocDB 4742. The data "handscan" results (DocDB 4742) find an average of 8.8 track-like objects in 50 data events after extrapolating down from the full readout window (4.8 ms) to a 2.0 ms drift time. The data "reco tracks" results come from looking at *trackkalmanhit* 3D-reconstructed tracks in cosmics data. The numbers, roughly, agree.

- **Accidental**: A flash which was not induced by the beam, but occurred during the 1.6 μs beam spill. Includes all backgrounds such as noise, cosmic particles (other than muons), and possible radioactivity.

This paragraph shows a very simple calculation on how many neutrino interactions we can expect in MicroBooNE.

- A booster spill has $4 \cdot 10^{12}$ POT.
- If we are running with 5 Hz, we will therefore get $1.7 \cdot 10^{18}$ POT per day.
- Based on MCC6.1 simulation, we know that 20,000 neutrino interactions in the full active volume of the MicroBooNE TPC (87 tons) corresponds to $5.3 \cdot 10^{19}$ POT[1].
- This results in roughly 650 neutrino interactions per day in the TPC active volume.
- In other words: we will have about 1 neutrino interaction per 660 beam spills.

Note that these calculations include all interaction types and all neutrino flavors, and does not assume a fiducial volume cut.

Another important number to take into account in the following is how many cosmic muon background tracks are observed in a drift window. The expected number of cosmic ray induced tracks per readout window, from studies by Matt Bass (DocDB 4742) and Sowjanya Gollapinni (DocDB 4751), are listed below. The rate changes with the applied cathode voltage because the drift time changes with the electric field strength. The summary of expected rates from simulation and a hand-scan of data are shown in Table 1.

A short note on background from neutrino interactions in the dirt outside of LArTF: Studies by Katherine et al. have estimated that the rate of events that do have traces from dirt interactions inside the TPC active volume is 81 per 1.7e16 POT (corresponding to one hour of running at 5 Hz). However, only 2 of these 81 events are expected to have muon track fragments inside the TPC active volume. This is a small background that has not been further studied in this note. We expect that, since the topology cuts defined below are looking for a neutrino vertex inside the active volume, the dirt event contamination after

[1]taken from CC incl. study, see DocDB 4633.

TPC-based selection should be small, as incoming muons should not appear contained. Meanwhile, dirt events will help pinpoint flash-timing studies by increasing the number of flashes seen in-time with the beam spill.

1.3 General strategy

The general strategy for the analysis is as follows:

1. Identify flashes from scintillation light produced in-time with the beam spill. Reject events that do not have such a flash.

2. On remaining events, perform 2D and 3D reconstruction, utilitizing clustering, tracking, and vertex-finding algorithms to identify neutrino-interaction-like topologies. Reject events that don't contain these neutrino candidates. For events that do, distinguish them from other cosmic rays in the event.

The optical- and TPC-based reconstruction and selection filters are separate, allowing for individual evaluation of them. They are described in Sections 2 and 3, respectively. Section 4 contains a more detailed description of the exact data-processing strategy, with an overview of the expected number of events after each step of the filtering.

The results presented in this note are currently derived from a variety of different sources: cosmics data, and Monte Carlo simulation with cathode voltage at -128 kV and -70 kV. In particular, analysis of the flash-finding has produced expected filtering rates based on data; meanwhile, the TPC reconstruction has based most of its expected rates on simulation with cathode voltage at -128 kV. We will update expected cosmic backgrounds and neutrino ID efficiencies as samples become available and they are processed.

Additionally, as a collaboration we are still moving towards stable and well-performing reconstruction algorithms that properly handle the complicated noise behavior seen in the data from the detector. Thus, we expect there may be some changes, especially in TPC reconstruction, that may affect exact purities and efficiencies. We do not expect changes in the reconstruction software to alter the end results very much.

2 Flash Finding

This section describes the first event filter to be applied based on reconstructed optical information, namely optical flash. Technical details of data samples used and analysis scripts as well as LArSoft event filters to be used are described in the last part of the section.

2.1 Flash Reconstruction

Flash reconstruction is described in detail in Ben Jones's thesis (DocDB 4736). A flash is defined as the collection of light observed at the same time across the detector. Flashes are reconstructed by first identifying optical hits: signals on individual optical detectors above a specified photoelectron (PE) threshold. Optical hits from all the PMTs are accumulated in 1μs bins of time, and if a bin rises above a set PE threshold, the collection of hits in that bin that overlap

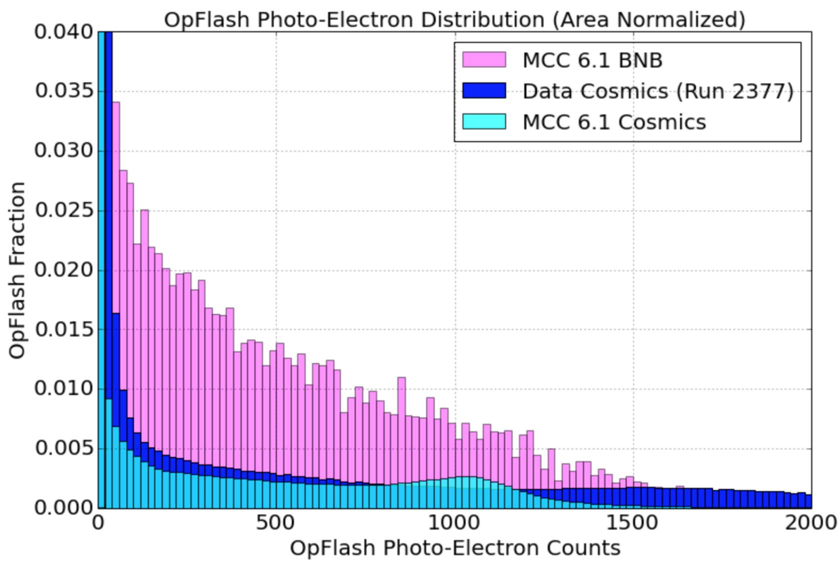


Figure 1: The distribution of number of PE in each flash. MCC 6.1 BNB, Cosmic, and Data are overlayed where data is taken from run 2377. For MC BNB distribution the flashes are taken from the 1.6 μs beam spill window, and cosmics are taken from 3 ms window. Distributions are then normalized within the shown range of Flash PEs. The majority of flashes from cosmogenic sources are small compared to the typical neutrino flash amplitude, and could be cut out with a moderate PE cut.

in time are defined as the ones that make up a flash. Properties of the flash—average time, average z and y positions and widths, etc.—are determined based on the set of hits.

The total size of a flash is described by the number of photoelectrons in the flash in total (summed over all PMTs in the flash). Neutrino interactions and cosmic muons are expected to produce a significant amount of light, however accidentals such as cosmogenic secondaries, noise, and other low energy backgrounds (many of which are not modeled in MC) will produce flashes with very low PE. For this reason cutting on the number of PEs in a flash is a good way of removing many backgrounds.

The number of PE distribution over all cosmic MC flashes is shown in figure 1. MCC 6.1 was used.

2.2 Flash Finding Efficiency

Using BNB Monte Carlo simulation, the flash finding efficiency for neutrino interactions was estimated. Figure 3 shows the efficiency for finding flashes as a function of the minimium PE in a flash, and figure 4 shows the same but zoomed in to the most interesting region. As is expected, with no minimum PE requirement the efficiency is close to 100% (the inefficiency coming from attenuation, PMT intrinsic efficiency and other effects). Requiring a minimum of 50 photoelectrons in a flash still retains 81.9% of beam interactions. This

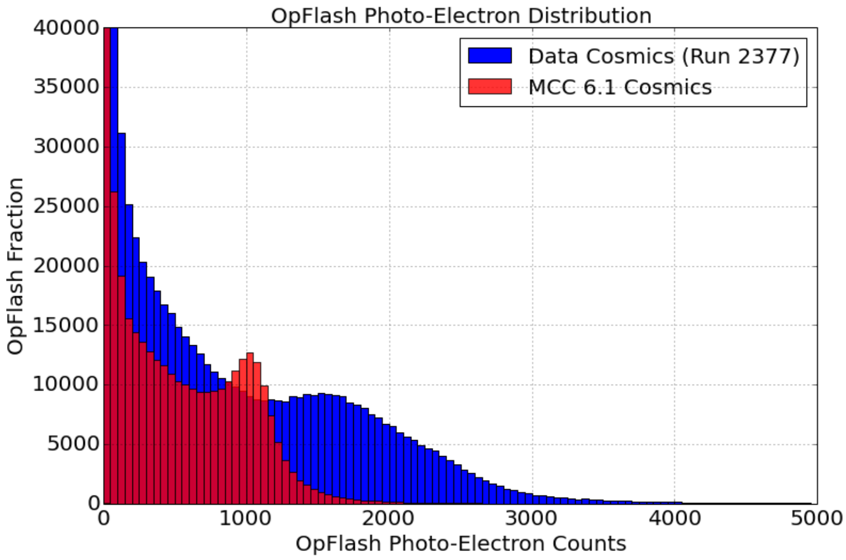


Figure 2: The distribution of number of PE in each flash. MCC 6.1 Cosmic, and Data are overlayed where data is taken from run 2377. MC is scaled to the same number of events that were recorded in this data run.

efficiency is expected to be sufficient for this analysis, though more sophisticated cuts may be developed in the future.

Figure 5 shows the efficiency for finding flashes from beam events, as a function of the beam interaction point in the x-direction (corresponding to the distance from the wall of PMTs). The efficiency looks rather flat for $x < 100$ cm, with some slight reduction in efficiency as the events get close to the far edge of the TPC.

2.3 Cosmic Mis-ID

A cosmic "mis-ID" will occur when a flash from a cosmic happens to overlap with a beam spill search window. It is also in principle possible to have dark noise hits in coincidence, however this is expected to be a negligible contribution. It is clear from figure 1 that there is a very large accidental contribution at low PE values which is removed by a very moderate PE cut.

Unfortunately, as figure 2 shows, there are some substantial differences between the MCC6 6.1 cosmic MC and recently collected data. The largest difference is that the peak coming from cosmic events is higher in data than in MC. There are a number of differences between the MC and data here, for example the drift field in data is lower and the purity is higher, which lead to more light being collected by the PMTs in data.

Also, there is some evidence that the liquid argon filters release small amounts of radon into the cryostat [2, 3], in which case some of the flashes in data will come from radon decays, however these flashes are also expected to be small, and a moderate PE cut will remove these events.

However, after all these considerations, the conclusion remains the same – placing a cut at a moderate PE value will remove a large fraction of accidental

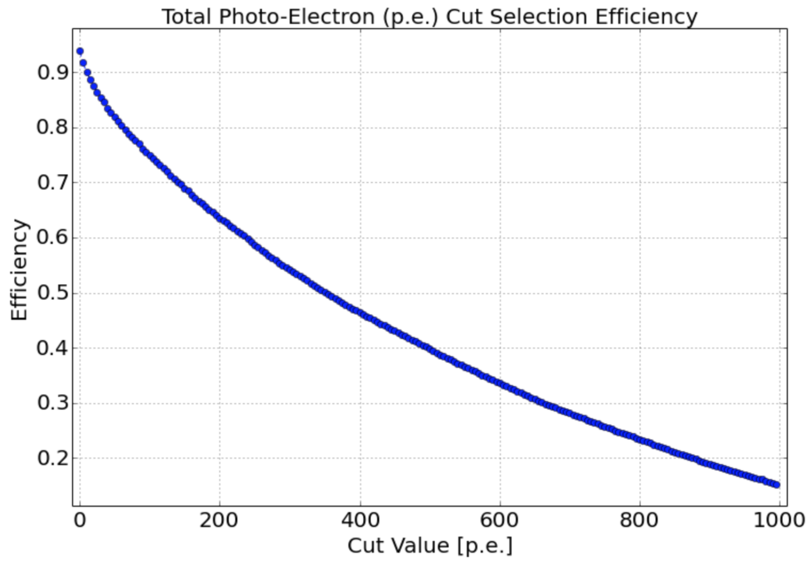


Figure 3: Efficiency for selecting beam events as a function of minimum PE cut.

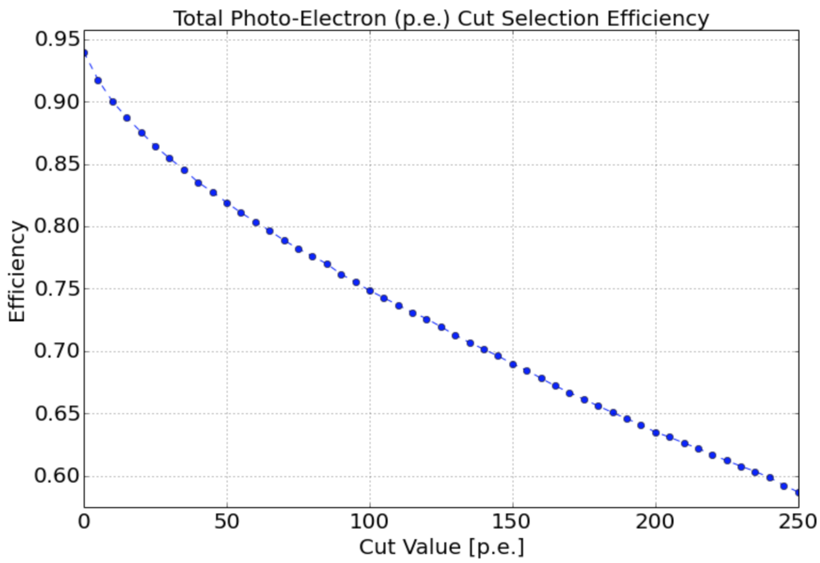


Figure 4: Efficiency for selecting beam events as a function of minimum PE cut. Zoomed into most interesting region. Note the efficiency does not hit 100% with as the cut PE drops to zero due to the cut on flash timing, requiring an observed flash to be in the 1.6 μs beam-spill period.

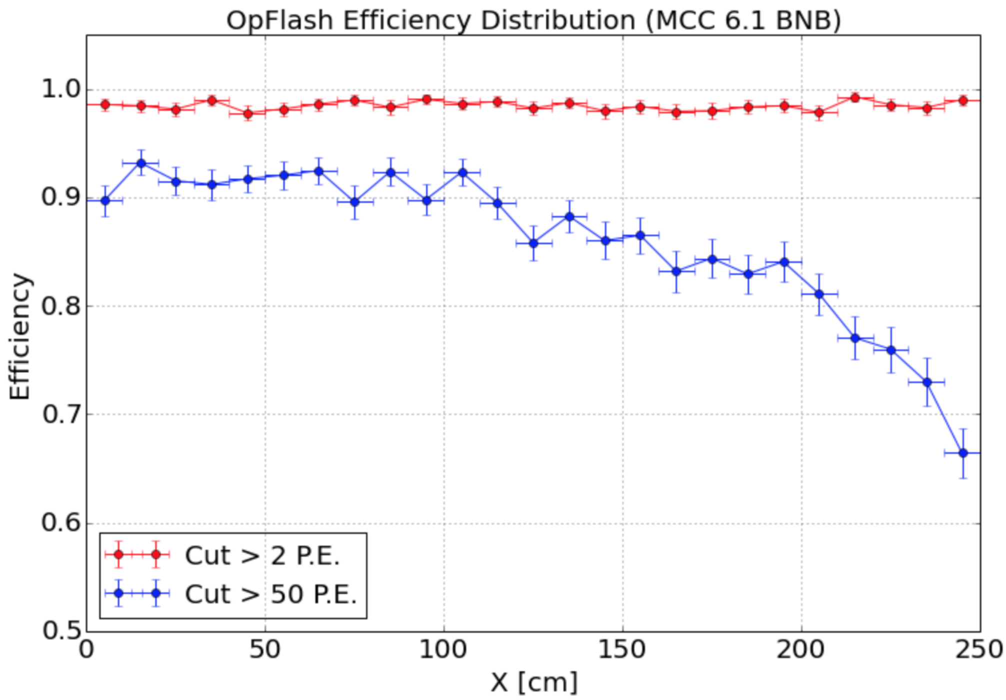


Figure 5: Efficiency for selecting beam events as a function of x-position (in cm) of the flash, for a 2 PE, and a 50 PE threshold cut.

and cosmic events, without significantly reducing the number of neutrino events.

Figure 6 shows the measured mis-ID rates as a function of PE threshold for cosmic data and MC. Clearly most flashes from non-beam backgrounds are very low energy, so for the same 50 PE cut which achieves an overall 80% neutrino selection efficiency, almost all of the accidental backgrounds are removed, leaving 0.8% of events with a flash due to non-beam background. This number agrees with the prediction one would get using the cosmic ray-induced tracks observed in data (see Table 1). For 8 observed tracks in a 2.0 ms drift window, we would expect $8 \times \frac{1.6}{2000} = 0.0064$ cosmics during the 1.6 μs beam spill. It is not surprising that long cosmic muon tracks would leave in excess of 50 PE in the reconstructed flash.

The distribution of number of flashes (above 50 PE) seen in a 2 ms drift window (2 ms is the drift time at 70 kV) for data and MC is shown in figure 7. This follows a poisson distribution, as expected, and the data distribution peaks at a value corresponding to 5 kHz.

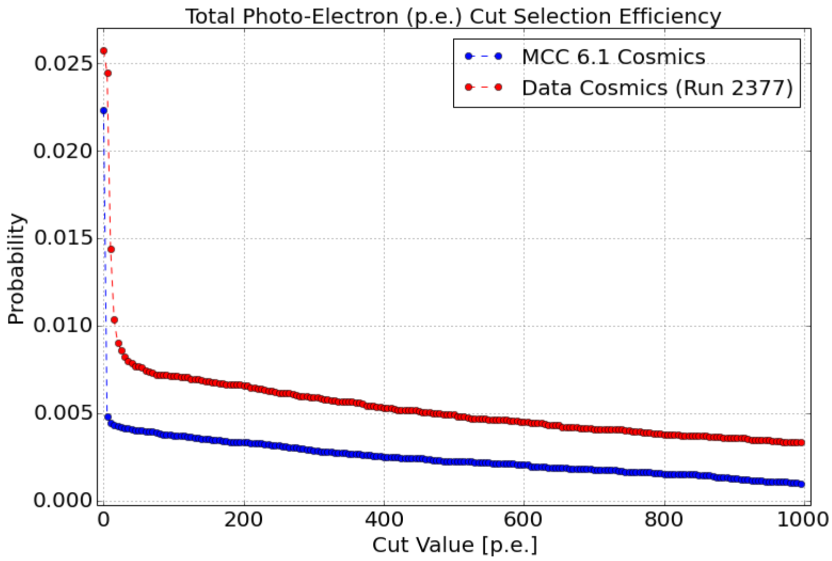
2.4 Beam Timing

An essential component for a filter using flashes is to get the appropriate time from the flash, and make a cut restricting flashes to be consistent with the neutrino beam spill period. As mentioned previously in Section 2.3, requiring a flash with $N_{\mathrm{PE}} > 50$ that is in time with the 1.6 μs BNB spill period will reduce the number of events by a factor of more than 150. However, before we can make a cut on flash timing, we must understand the timing of the trigger and PMT readout relative to the arrival of neutrinos from the BNB. While efforts have been made to understand this timing during commissioning, the best way to verify this timing is to identify a 1.6 μs window near the expected beam-time (which is the trigger time) where the number of flashes is significantly above the cosmic-ray background[2]. The following section calculated the necessary statistics for this measurement to be made.

First: a quick note on the PMT readout, which is explained in more detail in [4]. Unlike the case for the TPC wires, we do not get unbiased readout over an entire drift period (or number of drift periods) from the PMT readout system. Instead, we receive short windows of readout in one of two ways: (1) from individual PMT channels going above a set "discriminator" threshold; and, (2) from all PMT channels when an input signal, like a beam gate or fake beam gate, is sent into the PMT readout boards. This latter form provides unbiased readout of the PMT channels, and we expect to obtain this unbiased readout for a period of 23.4 μs around the trigger time. This unbiased readout period—the *'beam readout window'*—is much larger than the spill period, and is very likely to contain the beam spill period based on our current understanding of the detector.

Figures, 8, 9 and 10 show the expected distribution of flash times with respect to trigger time, for 8 hours, 12 hours, and one day (24 hours). The POT and expected event rates are summarised in table 2.4 All of these plots were produced using cosmic data and BNB Monte Carlo. In each plot every bin is 1.6μs wide (one full beam spill).

[2] We also record a well understood signal, known as RWM, which will tell us instantly if there is a dramatic failure which would make us miss the neutrinos.



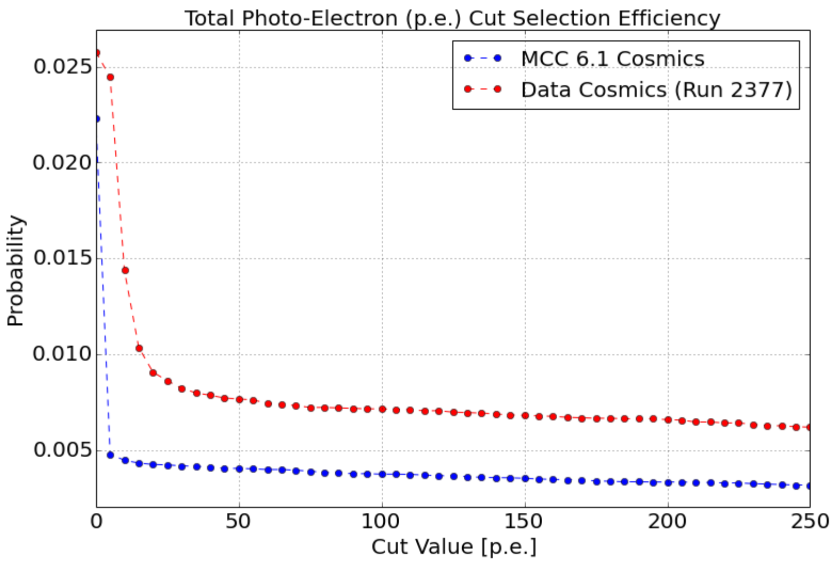


Figure 6: Fraction of cosmic events with a flash in the beam spill period as a function of the minimum PE of that flash, for (blue) MC and (red) data.

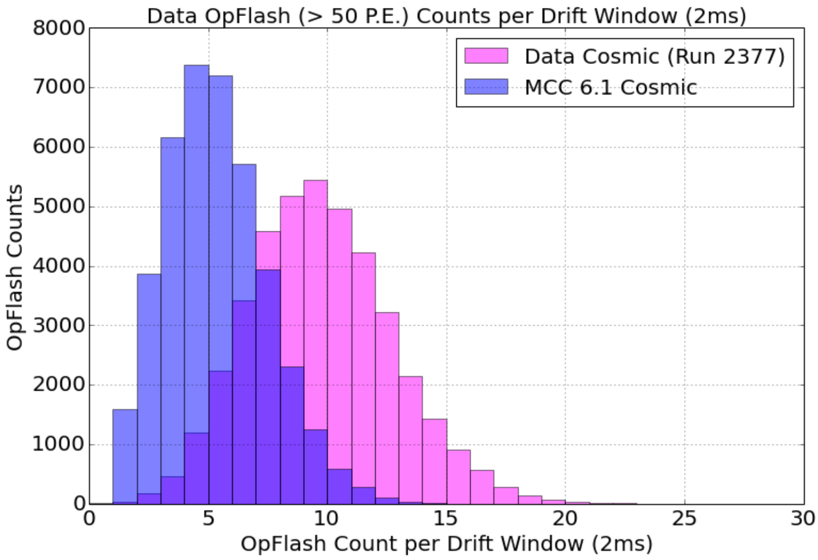


Figure 7: The distribution of number of flashes above 50 PE in 2.0 ms drift window for data and MCC 6.1 where data is taken from run 2377.

BNB integrated time (hr)	Expected POT (10^{17}protons)	N_{flashes}, Cosmics in 23.4 μs / in 1.6 μ s	N_{flashes}, ν in 1.6 μ s	$N_\nu / \sqrt{N_\nu + N_{\text{cosmic}}}$
2	1.4	3911 / 274	43	2.4
8	5.8	15645 / 1096	175	4.9
12	5.8	23349 / 1580	262	6.1
24	17	46544 / 3215	524	8.6

Table 2: Expected number of cosmic and neutrino events with flash in a 23.4 μs and 1.6 μs window around the trigger arrival time. The exposure period estimates assume a 5 Hz BNB pulse rate with 4×10^{12} protons on target per BNB pulse.

The plots show that it will take roughly one full 24 hour period of data taking to determine the exact time of the beam spill with respect to the trigger. Once the location of the 1.6μs-beam spill period relative to the trigger time is determined, we will have accurate start and end times for the beam spill, and use that for filtering out events that are inconsistent with having scintillation light from a BNB neutrino interaction.

2.5 Event Rates After Flash-Finding Filter

We expect about 12%[3] of cosmic-ray events to survive a filter that requires a flash with $N_{\text{PE}} > 50$ to be inside the 23.4μs-long unbiased beam readout window of the PMT readout. Further reducing the time-selection to 1.6μs reduces the cosmic-ray event survival probability to 0.8%, as noted in Section 2.3. Assuming a 5 Hz beam rate, this will leave 2100 and 135 cosmic-ray events per hour passing

[3]Obtained by scaling up the 0.8% for 1.6 μs to 23.4 μs

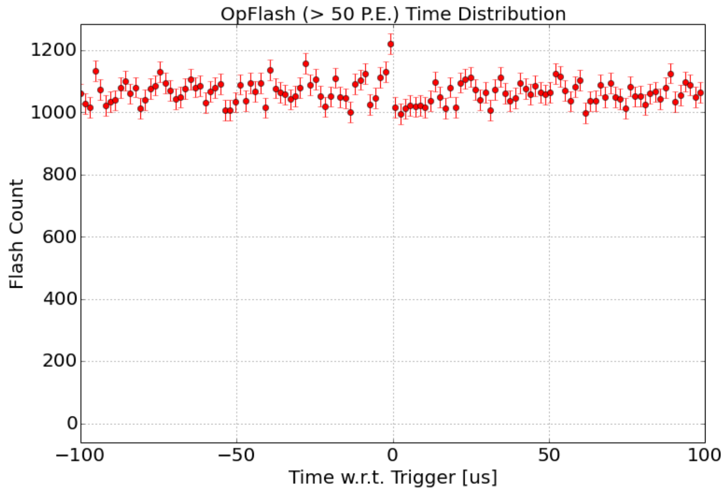


Figure 8: Predicted distribution of flash times, with respect to the trigger time, for 8 hours of data taking.

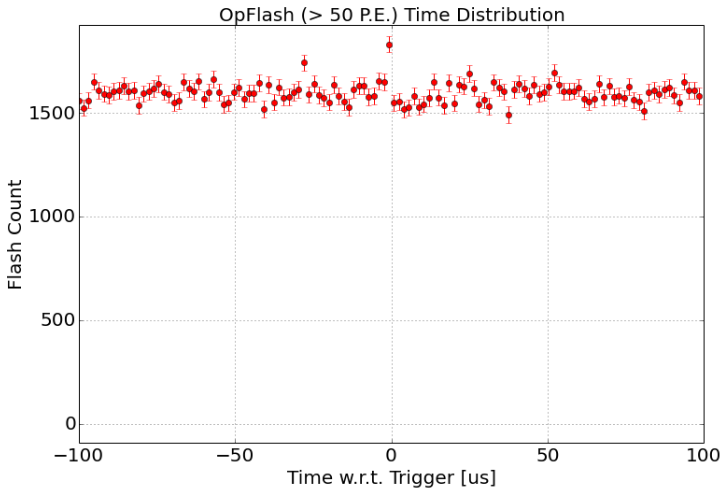
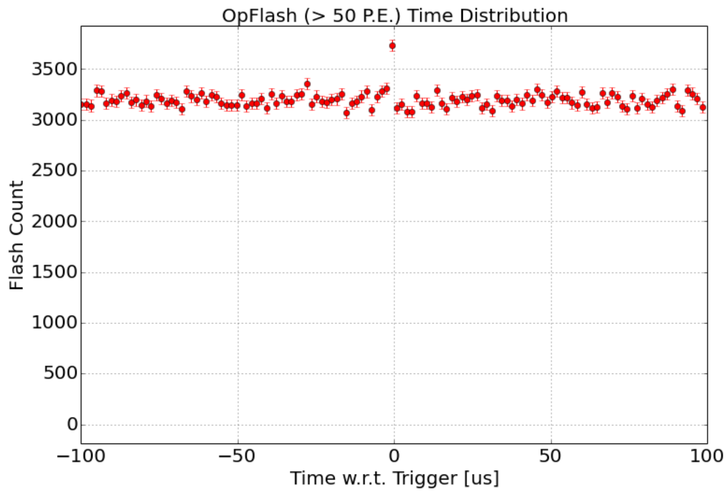


Figure 9: Predicted distribution of flash times, with respect to the trigger time, for 12 hours of data taking.



u

Figure 10: Predicted distribution of flash times, with respect to the trigger time, for 1 day (24 hours) of data taking.

the flash-finding filter, respectively. The efficiency for neutrino events to pass the flash amplitude and timing filter is, as noted in Section 2.2, 82%. Using the rough numbers offered in Section 1.2, we expect about 650 total neutrino events per day, and thus would expect about 22 neutrino events per hour to pass this filter.

2.6 Technical Remark

In this section three kinds of event samples were used.

1. MCC 6.1 BNB Monte Carlo simulation
2. MCC 6.1 Cosmic Monte Carlo simulation
3. Cosmic data sample from run 2377

Data files for above samples can be found under the persistent dCache area:

```
/pnfs/uboone/persistent/users/uboonepro/NeutrinoID_OpFlashData/
```

The actual list of events, identified by a combination of run, subrun, and event id, are generated from data file themselves and stored in the git repository. Data files are output of an actual LArSoft filter module that can be found under MicroBooNE software repository, uboonecode:

```
uboone/OpticalDetectorAna/FlashTrigger_module.cc
```

The event list text files can be found under the sub-directory:

```
uboone/OpticalDetectorAna/opflash_ana_neutrino_id
```

under which following analysis scripts, used to generate plots in this section, can be found as well.

EventListPrintOut.ipynb ... Generate event list text files.

FlashFilterEfficiencyAnalysis.ipynb ... Fig.3,4,6.

FlashRateAnalysis.ipynb ... Fig.7.

FlashTimeAnalysis.ipynb ... Fig.8,9,10.

FlashXDependency.ipynb ... Fig.5

3 TPC topology selection

In order to further reduce the background of cosmics events to our neutrino event selection, and to select neutrino events that are suitable for first public displays, we need to make use of reconstruction of the data from the TPC wires. We have developed two independent selection streams that use information from the TPC: one based on reconstructed 2D clusters, and the other based on reconstructed 3D tracks.

The main idea is the same in 2D and 3D. It is to look for a neutrino interaction inside the TPC active volume, which can be identified by two or more tracks originating from the same vertex. An example event display is shown in figure 15. Additional requirements enhance the purity and are described below.

The advantage of having 2D and 3D channels running in parallel is that if the performance of a particular algorithm that we optimized for is bad on actual experimental neutrino data, we still have event candidates coming through the other channel. Therefore, we are planning to have events pass the topology filter that either pass the 2D selection or pass the 3D selection.

The topology cuts presented below are currently optimized on MCC6.1 simulation, which was using 128 kV cathode voltage[4]. Since then, changes have been applied to both the simulation and the reconstruction in order to include what we learned from first experimental data taking of cosmics. Numbers in this section will be updated once a new MC dataset is available.

Note, that we are calculating passing rates for the TPC topology selection only. In addition, the number of background events will reduce by using flash finding as described above. An approximate efficiency factor for cosmic events to pass the flash finding is the factor of 0.008 described in Section 2.5, which we are applying to the passing rates to give a feeling for final event numbers. Note that flash matching is not yet used here. This means that the track(s) in a cosmic background that fake the neutrino interaction can be different from the track that has its flash during the beam spill. If we add flash matching to the selection and only run the topology selection on events that are flash matched, we will further increase the efficiency. This will be explored when flash matching performance studies on real data are available.

3.0.1 Cosmic tagging

The first step is based on the geometry cosmic tagging of events. Sarah's geometry tagger [1] runs on reconstructed 3D tags and assigns scores to each reconstructed track in the event that tag tracks that are most likely caused by cosmics.

[4]The MCC6.1 files were generated with larsoft version v04_09_00 and are located in /pnfs/uboone/scratch/users/bcarls/cc_inc_mc/v04_09_00. All data sets (BNB only, Cosmic only, BNB + Cosmic) contain 20,000 events. If not stated otherwise, all events are used.

Tracks	Tagged as neutrino (score = 0)	Tagged as "uncontained" (score > 0)
True neutrino	80.3%	19.7%
True cosmic	6.3%	93.7%

Clusters	Tagged as neutrino (score = 0)	Tagged as "uncontained" (score > 0)
True neutrino	100%	0%
True cosmic	31.1%	68.9

Figure 11: Matrix for tagging and mis-id'ing of neutrino and cosmic induced tracks in MCC6.1 with trackkalmanhit and cccluster (y-plane only).

Cosmic tagging scores for tracks:

- 1: the track is tagged as entering and exiting
- 0.95: the track is a delta ray associated with a tagged track
- 0.5: the track is either exiting or entering, but not both
- 0.4: the track is entering or exiting through the z boundary
- 0: the track is not tagged.

Clusters get assigned a cosmic score of either 0 or 1 accordingly.

Figure 11 shows that over 90% of cosmic induced tracks are tagged as cosmic. These tracks are not further considered when looking for neutrino topologies in an event. We are losing about 20% of neutrino induced tracks by requiring that the events are contained.

In order to tag tracks as entering or exiting, the algorithm checks if the tracks go through a boundary region. The boundary region for this study was set to 5 cm distance from all sides of the TPC. Based on handscan studies with recent experimental data, it has been decided to increase this boundary to 10 cm for all sides of the TPC, and we are going to use a 10 cm boundary for this selection. The 10 cm boundary has already been used for processing recent experimental data (e.g. the data shown in table 7), but not yet in any of the simulation used in this note.

Cosmic tagging is less efficient for clusters than for tracks, since full 3D-position information of the entering and exiting point of the cluster is not available. The rejection of clusters is mostly based on the timing information of the cluster.

Since cosmic tagging is using the timing information of events to exclude every cluster or track that happens outside the drift window, it will depend on the drift time and therefore on the HV setting. The drift time at 128 kV is 1.6ms, at 70 kV it is 2 ms, so we are expecting about $2/1.6 = 1.25$ times more cosmic induced tracks or clusters in the drift window. Figure 12 does show a comparison of geometry cosmic scores for 128 kV simulation and data

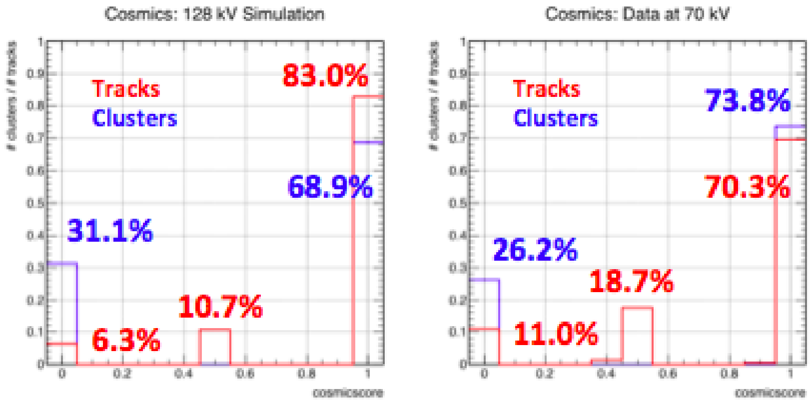


Figure 12: Cosmic scores for 128 kV simulation and data taken at 70 kV. Note that both have been processed with different versions. This is the reason why there are scores with 0.4 and 0.95 for the experimental data, this was not yet added when processing simulation. Histograms are area normalized to 1. Simulation uses 9200 events, data uses 180 events.

taken at 70 kV[5]. The cosmic tagging efficiency as seen from Figure 12 varies since several changes in the reconstruction between the two processing versions change also influence the result. From the comparison it can be said that there is no indication of any problem with cosmic tagging for lower drift fields.

3.1 Selection based on 2D clusters

In first observations with experimental cosmics data, it has been shown that often 2D clustering performs reliably well, while the performance of 3D track reconstruction can be more sensitive to the performance of lower-level quantities, and thus performance can fluctuate strongly depending on, for example, different noise filters. This motivates having a neutrino selection filter running only on 2D objects, and in particular on clusters in the collection (Y) plane, which has the best information.

3.1.1 Primary Cuts

The first step is to determine which clusters to consider. We require that the clusters are in the Y-plane, the cluster cosmics tagger score was < 0.4 and the cluster had at least ten hits. We then select only readout events which contain at least two clusters that meet these requirements.

After seeing that we were still being overwhelmed by cosmic events per day, we stepped back and thought about what neutrino events we were looking to see

[5]This data as been processed with v04_22_00 using RawDigitFilter instead of noise filter. MC at 128kV is from MCC6.1 and has been processed with v04_09_00. Files are located in `/uboone/data/users/aschu/CosmicTagging/CosmicTaggingComparison/`. The script to evaluate the tagging efficiency is in uboonecode feature branch `feature/Anne_NeutrinoIDFilter` in `uboonecode/uboone/TPCNeutrinoIDFilter/AnalysisScripts/ClusterVsTrackTagging.C`.

and made cuts based on this. The following primary cuts are made to remove as many cosmics as possible while leaving "nice" looking neutrino events.

The next cut is made to remove long, vertical clusters. This cut was applied after going through cosmic events that are passing and seeing that most clusters are long and at high projected angles while the neutrino events are forward going clusters. We require a good cluster to have either a projected start angle less than 30 degrees from the z-axis or be less than 200 wires long. We add the length condition so that we don't cut any short high-angle clusters that may correspond to a proton, or other highly ionizing particle, accompanying a long muon cluster. The 200 wire length cut roughly equates to 0.6 m in the z-direction, assuming a 3-mm wire spacing. The projected cluster start angle, α, is defined by $\tan\alpha = \Delta T/\Delta W$, where T is time ticks and W is wires.

Lastly, we require all clusters to be at least 15 wire ticks or 15 time ticks long. This was a general cut to remove cosmics that were being matched with small deltas without removing short proton clusters that could be matched to a long muon cluster. We make the length cut in either the time or z direction because we don't want to remove more time-going short tracks. The reason that we don't cut out short tracks because our ideal neutrino events have both a long MIP cluster and a short, more highly ionizing cluster starting at the same point. This is explained in more detail in the next section.

3.1.2 Secondary Cuts

The goal of the secondary cuts is to match a long, low-angle cluster with a short, high-charge cluster. We only look at clusters that have passed all of the primary cuts. We first select a cluster with a length greater than 500 wires, which corresponds to approximately 1.5 m in the Z-direction. This long cluster is guaranteed to have a start angle of less than 30 degrees because of our preliminary cuts.

We then search for any cluster which begins within approximately 3 cm (10 wires and 30 time ticks) from the low-Z end of the long cluster. We also require this cluster to be shorter than the initial cluster. Since start and end points of the clusters may not be in the right direction, we compare both the start and end points of the shorter cluster to the point of the long cluster which has a lower wire value. This is to cut out backward-going interactions.

Now that we have a vertex match between a long and a short cluster, we make cuts on the charge and the projected opening angle of the pair. We require that the shorter cluster has a higher start charge than the long cluster. Start charge is defined as the charge on the first wire in ADC counts. We also require that the projected opening angle (the difference in projected start angles of the two clusters) be between 11 and 90 degrees. This cut is mainly to remove matched clusters which are either entirely overlapping or are actually part of the same long particle track. The neutrino/cosmic events per day are shown in table 3. Figure 13 shows in detail the percent of clusters that pass each primary cut. These numbers will be different than the number of events remaining because the cosmics have, on average, many more clusters per event than neutrinos. Figure 14 shows in detail the percent of matched cluster pairs remaining per secondary cut from the total number of events containing clusters that passed all primary cuts.

[7] Analysis code can be found in uboonecode feature branch fea-

Cluster Set	No cuts	Primary cuts	Secondary cuts
Neutrino only	570	303	32
Cosmics only (no flash)	308,016	291,879	602
Cosmics only (with flash)	2464	2335	5
Neutrino/Cosmics	0.23	0.13	6.4

Table 3: Passing rates for 2D cluster cuts for a neutrino only MC set and a cosmics only MC set. Column one shows raw event rates with no cuts applied. Column two shows event rates after preliminary cuts are applied. Column three shows event rates after secondary cuts are applied. Line three shows the second line scaled with a factor of 0.008, which is a first guess efficiency estimate for the flash finding. All event numbers are normalized to events per day assuming we are running at 5 Hz.[7]

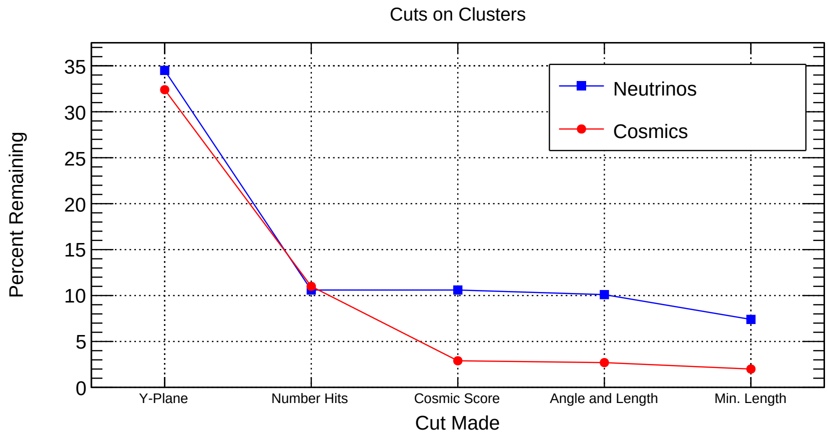


Figure 13: Percent of good clusters remaining for neutrinos and cosmics after primary cuts. The is relative to the total number of initial clusters.

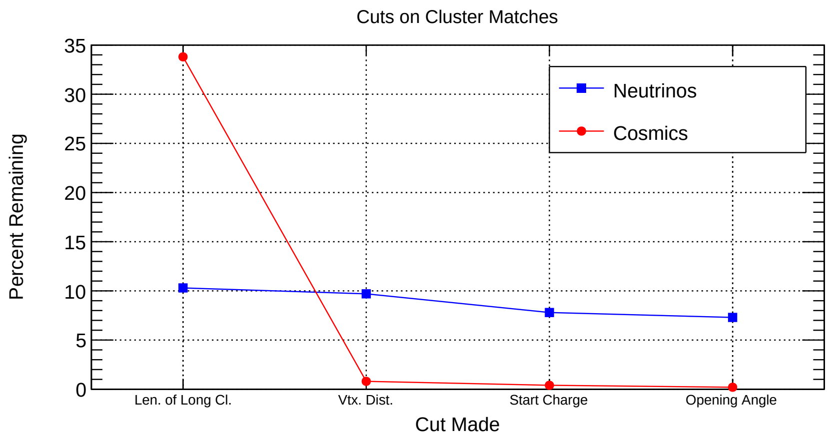


Figure 14: Percent of matched cluster pairs remaining for neutrinos and cosmics after secondary cuts. This is relative to the total number of events that contain clusters which pass the primary cuts.

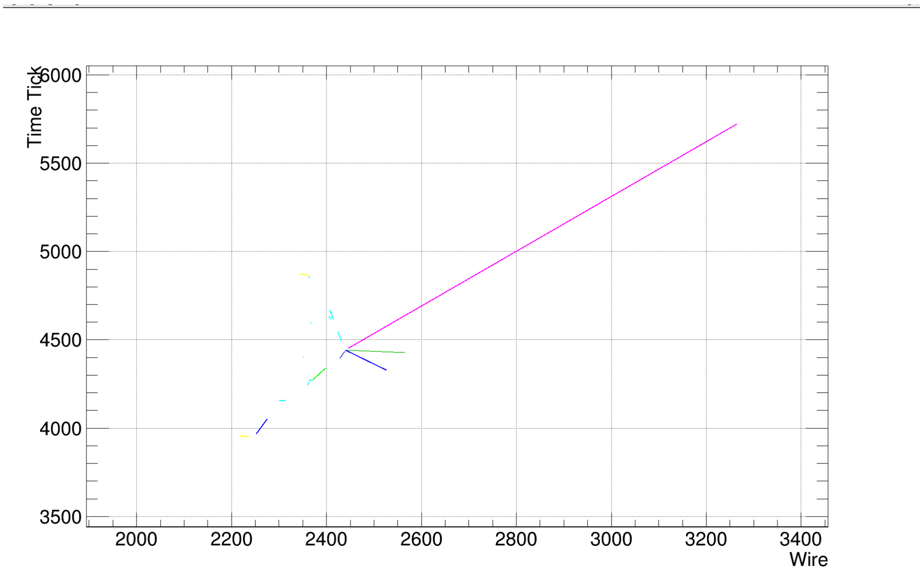


Figure 15: Y-plane view of neutrino Event that passed all cuts. The straight lines connect start and end points of 2D clusters. Zoomed in to interesting section.

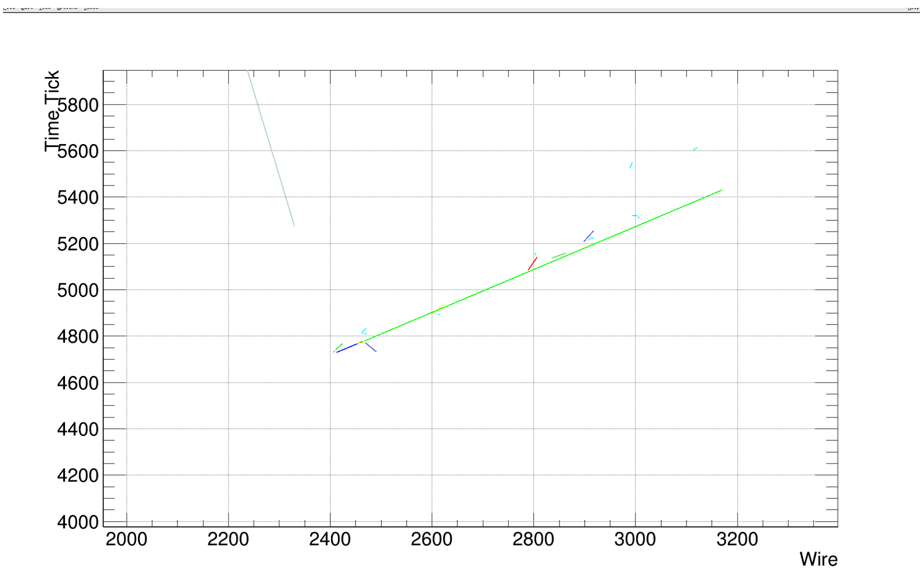


Figure 16: Y-plane view of cosmic Event that passed all cuts. The straight lines connect start and end points of 2D clusters. Zoomed in to interesting section.

It is important to note that there is no 2D vertexing algorithm or flash matching used to calculate these numbers. Using a vertex algorithm rather than matching by hand as well as implementing flash matching may increase our signal to background ratio. Figures 15 and 16 are neutrino and cosmic events that passed the primary and secondary cuts. By hand-scanning though, these can be easily distinguished.

3.1.3 Final Steps

As described above, the neutrino selection on 2D clusters provides a basic and robust approach to select neutrino candidate events. Based on MC studies, the background of cosmic events can be reduced to 25% of the neutrino event yield. We are therefore going to run this selection chain in addition to the 3D tracking approach that is described below. The efficiency of the 2D selection is lower than the efficiency of the 3D selection, but nevertheless, this is an alternate approach that is a good backup in case 3D track reconstruction does not perform well on BNB events for some reason.

3.2 Selection based on 3D tracks and vertices

3.2.1 Vertex with two tracks

The selection of neutrino candidate events is based on finding trios of a reconstructed vertex plus two tracks[8]. We loop over all reconstructed vertices and all reconstructed tracks (with a geometry cosmic tagger score < 0.4) and for each trio combination that can be formed, we calculate the following distances (illustrated in Figure 17):

- d: distance between the start points of the two tracks
- d_1: distance between vertex and start of track 1
- d_2: distance between vertex and start of track 2

Please note that the direction of the track gets flipped such that the point closest to the vertex is the start point. Note that this also means that if the same track is part of trio with a different reco vertex, its direction might be the other way. We do this flipping because track directionality (based on calorimetry or delta ray direction) is not yet implemented in e.g. trackkalmanhit.

We then select $\max_d = \max(d, d_1, d_2)$ as the relevant characteristic for each trio found in the event. The distribution of $\max_d$ is shown in Figure 18.

The next step is to select the best trio of each event. The decision to keep or reject an event is made on the best trio of an event. The best trio is the one with the smallest maximum distance within it, i.e. $min(\max_{d,i})$, where i loops over all trios in the event. The distribution is shown in figure 19 for BNB neutrinos and in figure 20 for cosmics. It is obvious, that $min(\max_{d,i})$ is much smaller for neutrino events than for cosmic events.

ture/Anne_NeutrinoIDFilter in the script uboonecode/uboone/TPCNeutrinoIDFilter/-AnalysisScripts/2DClusterNuCuts.C

[8]The code to produce tables 4 to 6 and 7 and figure 18, 19, 20 can be found in uboonecode feature branch feature/Anne_NeutrinoIDFilter in the script uboonecode/uboone/TPCNeutrinoIDFilter/AnalysisScripts/StatisticsForPairPlusVertex.C

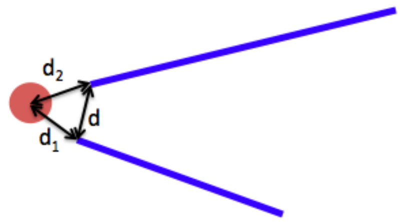


Figure 17: Sketch of the neutrino candidate topology (1 reco vertex + 2 reco tracks) and definitions of relevant distances.

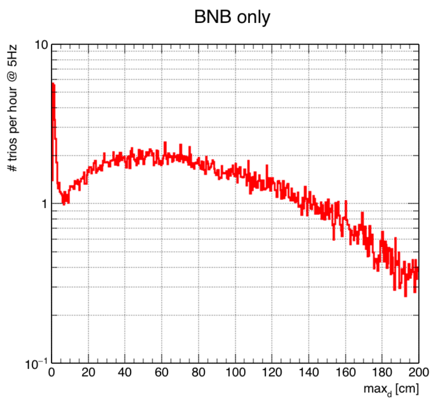


Figure 18: Distribution of $\max_d$ for all trios found in all BNB events. There are clearly two populations: the peak at small values are the trios that really belong together. The broad bump at higher values are random combinations of tracks and vertices.

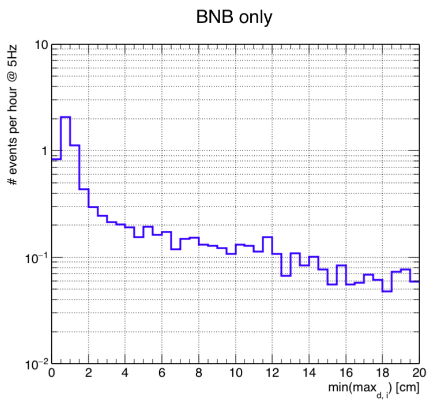


Figure 19: Distribution of $\min(\max_{d,i})$ for all BNB events.

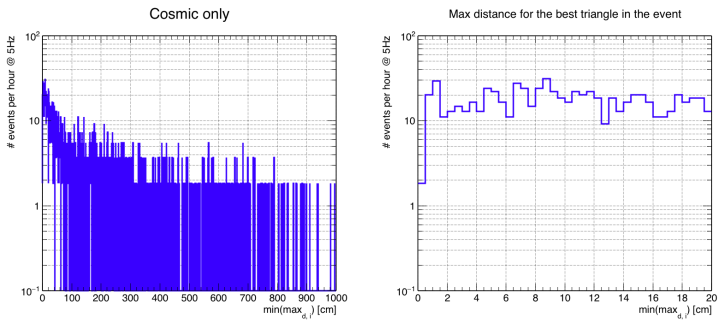


Figure 20: Distribution of $\min(\max_{d,i})$ for all Cosmic events. The plot on the right shows a zoom in to the first 20 cm.

Cut value in cm	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Neutrinos	0.6	2.5	3.4	3.8	4.0	4.2	4.3	4.4	4.6	4.7
Cosmics (no flash)	0	22.1	46.0	64.5	75.5	93.9	110.5	123.4	139.9	160.2
Cosmics (with flash)	0	0.17	0.37	0.52	0.6	0.75	0.88	0.99	1.12	1.28
Neutrino/Cosmics	-	14.2	9.2	7.3	6.6	5.5	4.9	4.5	4.1	3.7

Table 4: Passing rates for the combination of **trackkalmanhit and cccluster vtx** as a function of the cut value chosen for $\min(\max_{d,i})$. The second line shows neutrino numbers, the third line cosmic numbers not applying the flash finding. Line four shows the third line scaled with a factor of 0.008, which is a first guess efficiency estimate for the flash finding. All event numbers are normalized to events per hour assuming we are running at 5 Hz.

Note that this topology does not prevent us from finding neutrino topologies with more than two tracks. In a case with multiple tracks, the trio of the two tracks closest to the vertex will be used to make the decision. Also note that this selection does not prevent us from selecting events with a shower (or multiple showers) as long as they have two tracks (e.g. muon plus proton), which are reconstructed.

Different combinations of vertex and track reconstruction algorithms have been tested and the following combinations have been chosen:

- trackkalmanhit with cccluster, $\min(\max_{d,i}) < 3$ cm.
- trackkalmanhit with pandoraNu, $\min(\max_{d,i}) < 4.5$ cm.
- pandoraNu with cccluster, $\min(\max_{d,i}) < 5$ cm.

The cut value was chosen such that the cosmic background contamination, after the expected effects of the flash-finding filter, is less than 20%. Tables 4 to 6 show the passing rates for neutrinos and backgrounds based on MCC6.1. The location of the cut may need to be modified in data, based on the observed cosmic yield.

Table 7 shows the passing rates for experimental data for different combinations of tracking and vertexing algorithms. These can be compared to the

Cut value in cm	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Neutrinos	0.8	2.9	4.0	4.5	4.8	5.0	5.2	5.4	5.6	5.8
Cosmics (no flash)	1.8	22.1	51.6	62.6	75.5	90.2	103.1	119.7	132.6	156.5
Cosmics (with flash)	0.01	0.18	0.41	0.50	0.60	0.72	0.82	0.96	1.06	1.25
Neutrino/Cosmics	56.2	16.4	9.8	8.9	7.9	6.9	6.3	5.7	5.3	4.6

Table 5: Passing rates for the combination of **trackkalmanhit and pandoraNu vtx** as a function of the cut value chosen for $\min(\max_{d,i})$. The second line shows neutrino numbers, the third line cosmic numbers not applying the flash finding. Line four shows the third line scaled with a factor of 0.008, which is a first guess efficiency estimate for the flash finding. All event numbers are normalized to events per hour assuming we are running at 5 Hz.

Cut value in cm	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Neutrinos	0.2	1.4	2.5	3.0	3.3	3.5	3.6	3.8	3.9	4.0
Cosmics (no flash)	1.8	12.9	33.1	49.7	51.6	55.2	57.1	58.9	70.0	73.7
Cosmics (with flash)	0.01	0.10	0.27	0.40	0.41	0.44	0.46	0.47	0.56	0.59
Neutrino/Cosmics	12.6	13.5	9.4	7.5	8.0	7.9	8.0	8.0	6.9	6.8

Table 6: Passing rates for the combination of **pandoraNuKHit track and cccluster vtx** as a function of the cut value chosen for $\min(\max_{d,i})$. The second line shows neutrino numbers, the third line cosmic numbers not applying the flash finding. Line four shows the third line scaled with a factor of 0.008, which is a first guess efficiency estimate for the flash finding. All event numbers are normalized to events per hour assuming we are running at 5 Hz.

tables above. Note, that this is based on very low statistics, yet: the data sample contains only 720 events, which corresponds to a time period of 2.4 min for running at 5 Hz.

3.3 TPC selection conclusion

For the inital selection of neutrino candidate events, we want to run the following filters in parallel:

- 2D cluster selection as described in subsection 3.1
- 3D track and vertex selection for three different combinations of track reconstructions and vertex algorithms as described in subsection 3.2.1: trackkalmanhit + ccluster/linecluster vtx, pandoraNuKHit + ccluster/-linecluster vtx, trackkalmanhit + pandoraNu vtx. Cut values for minimum distances are listed above.

This gives four streams that output neutrino candidate events. We do expect a large overlap between these.

If it turns out that the background of cosmic events is overwhelming when running with real BNB events, the purity of the sample can easily be enhanced

[10] The data used here was from runs 1712, 1713, 1715, 1716, 1717. The total number of events is 700. The data was processed with `v04_22_00` using RawDigitFilter as a default. Files are located in `/pnfs/uboone/scratch/users/bcarls/neutrino_id/v04_22_00`.

Cut value in cm	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
trackkalmanhit + pandoraNu	0	0.4	0.4	0.6	0.8	0.8	0.8	0.8	0.8	1.2
trackkalmanhit + cctrack	0	0	0	0	0	0	0	0	0	0
trackkalmanhit + linecluster	0	0	0	0	0	0	0	0	0	0
pandoraNuKHit + pandoraNu	0.4	0.8	1.6	1.6	1.8	1.8	1.8	1.8	2.2	2.4
pandoraNuKHit + cctrack	0	0	0	0	0	0	0	0	0	0
pandoraNuKHit + linecluster	0	0	0	0	0	0	0	0	0	0

Table 7: Passing rates for experimental cosmics data[10]as a function of the cut value chosen for $\min(\max_{d,i})$. Numbers have been scaled with a factor of 0.008, which is a first guess efficiency estimate for the flash finding. All event numbers are normalized to events per hour assuming we are running at 5 Hz. The reason for the zeros for cctrack and linecluster is, that these algorithms reconstruct a very low number of vertices in general compared to PandoraNu.

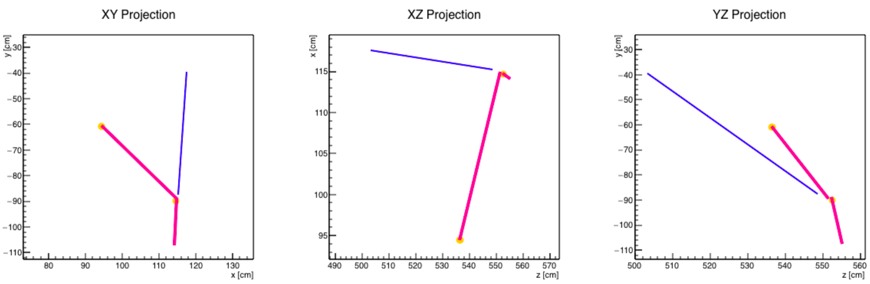


Figure 21: Projections of a BNB simulated event that does pass the 3D selection criteria. The two pink tracks are the two tracks that form the trio together with the vertex they are connected to (orange triangle).

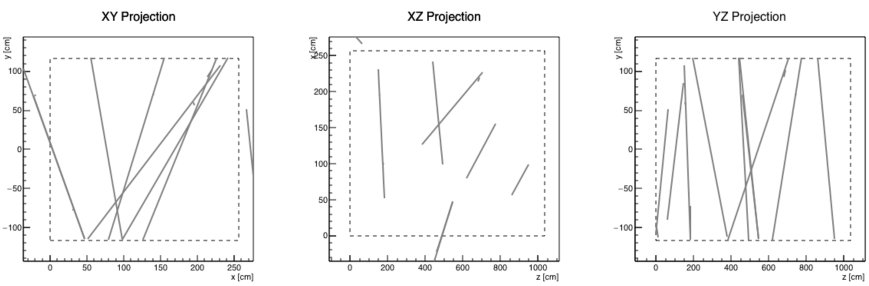


Figure 22: Projections of a cosmic only simulated event that does not pass the 3D selection criteria. The grey tracks are all tagged by the geometry cosmic tagger. There is only a single tiny contained track that can't fulfill the selection.

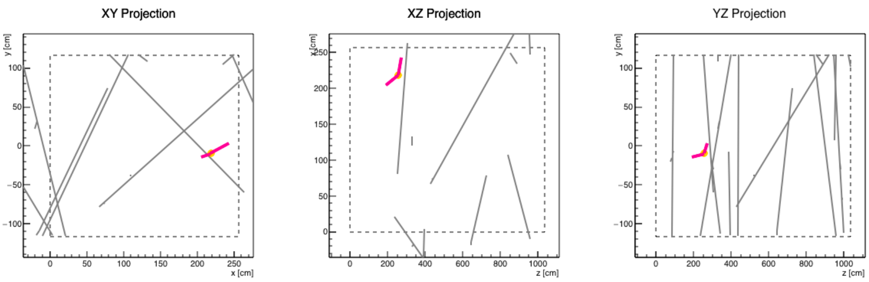


Figure 23: Projections of a cosmic only event that does pass the 3D selection criteria and creates a fake signal. The two pink tracks are the two tracks that form the trio together with the vertex they are connected to (orange triangle).

by requiring that events pass several of the selection streams, i.e. either the 2D and a 3D selection or two or more 3D selections. Since all necessary output for this additional selection is available, this can very quickly be added.

4 Plans for data processing

We will process the data using a series of LArSoft filter modules, to gradually reduce the number of candidate neutrino interaction events.

4.1 Filter 1: Loose Flash Filter

We will pass events through a filter that looks at flashes reconstructed during the unbiased PMT readout window that includes the beam spill. This window is 23.4 μs long, and will very likely contain the true beam spill window.

Based on the work shown in Section 2, we expect about 12% of events containing only cosmic rays to pass this requirement for a large flash in this 23.4 μs window. Assuming a 5 Hz BNB rate, this would yield 2100 events per hour.

From the calculation in Section 1.2, we expect around 27 neutrino interactions per hour. We expect 100% to pass the flash-timing cut, while 82% (from Section 2.2) of these events will pass the 50 PE threshold, leaving about 22 neutrino interactions per hour.

At this stage, the sample is roughly 1% neutrino interaction, and 99% cosmics.

4.2 Filter 2: TPC Reco Filter

Events that pass the loose flash filter may then be passed through the TPC reconstruction chain in order to apply the selection on neutrino interaction topologies described in Section 3. One of the best performing selections is the 3D selection using the TrackKalmanHit tracking algorithm with PandoraNu vertexing. Performance is listed in Table 5. If we take the 3 cm distance cut, then roughly 5 neutrino interactions survive per hour. Based on the reported number of 90.2 cosmic events per hour passing those cuts, without flash-filtering, that is an efficiency for cosmics of $90.2\times\frac{1}{(5\text{Hz})(3600\text{s})} = 0.5\%$. Thus, if 2100 cosmic events pass the first level of flash filtering each hour, then about 11 cosmic ray events per hour survive the addition of this TPC reconstruction filter. With a total selection of approximately 16 events per hour, this filter can provide a scannable set of data.

Events that pass this filter may be used to cross-check that the algorithms are working properly on data. Looking in the sideband region where the distance cut is less stringent would yield a higher number of cosmic rays relative to neutrino interactions. As mentioned in Section 3.3, more complicated cuts could increase the overall efficiency without sacrificing purity.

After running through a TPC-reconstruction-based filter, we expect the event sample to be about 30% neutrino interactions, and 70% cosmic-only events.

4.3 Filter 3: Tight Flash Filter

With data collected over the first 24 hours, the beam timing can be measured more precisely following the method described in Section 2.4. Assuming that after 8-24 hours of flash data (passing the loose flash timing filter) we can measure the beam spill to within 1.6 μs total width, then the cosmics events that have passed the filters up to this point will be reduced by a factor of ten. The neutrino events should be unaffected. Thus, the final selection will yield about 5 neutrino events per hour, but about only 1 cosmic event per hour passing all requirements. This will be a high-purity neutrino sample: roughly 80% neutrino interactions, compared to 20% cosmics events. From the first 24 hours of data then, we expect about 136 total neutrino interaction candidate events, which we expect to contain 120 neutrino interaction events, and 16 cosmics events.

5 Additional methods considered / Future improvements

5.1 Flash Matching

Flashes that are reconstructed will have light on multiple PMTs, but not all PMTs. Using information about the location of light deposits, and the size of each light deposit, it is possible to estimate where in the tank the flash originated. Of course, the flash will come from an extended source, but the centre-of-charge should correspond roughly to the reconstructed flash position. Flash matching is not currently used to select neutrino interactions, however in the future it is expected to be a valuable resource for increasing the purity and efficiency of neutrino event selections.

5.2 π^0 Topological Filter

Using 2D cluster and 3D shower reconstruction, it is possible to devise a selection for events that contain two electromagnetics showers consistent with coming from the decay of a neutral pion. This selection could be powerful in opening up more neutral current neutrino interactions into our first event selection, as well as provide a different, interesting topology to display in our first event set. Howeber, the π^0 stream is not yet ready for implementation and we are not planning to have it running in the first couple of days. A dedicated π^0 filter will be further developed and added to the proposal later on. Some first ideas for this can be found in DocDB 4770.

References

[1] Sarah Lockwitz, *DocDB3413*, April 25, 2014.

[2] H. Simgen, G. Zuzel, *Radon Emanation Measurements*, http://www.mpi-hd.mpg.de/gerda/lngs07/lngs07_slides/zuzel_gerda_nov07.pdf.

[3] F. Cavanna et al., *Radon emanation measurements at LNGS*, internal communication.

[4] D. Kaleko, *PMT Triggering and Readout for the MicroBooNE Experiment*, Proceedings from the LIDINE conference at Fermi National Accelerator Laboratory, June 2013.