Selection of ν_μ charged–current induced interactions with N>0 protons and performance of events with N=2 protons in the final state in the MicroBooNE detector from the BNB

MICROBOONE-NOTE-1056-PUB

The MicroBooNE Collaboration

October 12, 2018

Abstract

Using the MicroBooNE liquid argon time projection chamber (LArTPC) at Fermilab's Booster Neutrino Beam, we examine samples of charged current events from neutrino scattering on argon with exactly two or with any number of protons in final states with no other hadrons except neutrons. We described the particle identification method employed for protons, and we compare measured kinematic distributions of muons and protons to predictions from several Monte Carlo event generators.

Contents

1 Introduction **3**

2 Motivation **3**
2.1 MicroBooNE simulation 7
2.2 Data and Monte Carlo data sets 8

3 The ν_μ Charged Current inclusive selection **9**

4 Particle identification in LArTPC **9**

5 The ν_μ Charged Current N Proton Selection **13**
5.1 Signal Definition and Background Categories 13
5.2 Event Selection Scheme 14
5.3 Proton Momentum Threshold Determination 15
5.4 Proton Multiplicity Study 17
5.5 Signal and background analysis 17
5.5.1 Efficiency dependence on model 18
5.6 CCNProton Kinematical observables 19
5.7 Study of tracking resolution 22

6 The ν_μ charged current two proton events **23**
6.1 Kinematical distributions for the CC2Proton sample with a proton threshold momenta of 300 MeV/c 24
6.2 Discussion on CC2Proton analysis 29

7 Conclusions **30**

8 On-beam data Event Displays **31**

1 Introduction

This note describes the development and application of a fully-automated algorithm to identify final state protons in the MicroBooNE LAr TPC detector. The technique utilizes a χ^2 discriminator comparing the dE/dx along a candidate track to the theoretical expectation for a proton traversing liquid argon. Liquid Argon TPCs provide improved response to final state hadrons and greater sensitivity when comparing data to theoretical neutrino interaction models, final state interaction models, and nuclear models. Proton candidate reconstruction and identification is an important development on the path to improved cross section measurements and comparison with these models. The identification of final state particles is a key component to improved studies of neutrino interactions and comparison with theoretical models. The focus of the note is the presentation of a new proton identification method, efficiency and purity determination of this algorithm, and the application within the ν_μ–CC inclusive framework [8]. Finally, we present comparisons between data and two model choices of GENIE MC generated samples within this event selection.

2 Motivation

The study of charged–current neutrino interactions in Ar is an important component on the path to improving neutrino interaction models and reducing the uncertainties for current and future neutrino oscillation measurements. Past CCQE results came from MiniBooNE [10] without reconstruction of final-state nucleons. The neutrino beam energy spectrum is almost identical for MiniBooNE is almost identical to MicroBooNE. This allows results to improve upon MiniBooNE measurements. The most studied signal is now CC0π (sometimes called "CCQE-like") where the signal has one muon, no pions, and any number of nucleons in the final state. Many experiments have measured the CCQE–like interaction because it has a large cross-section in most accelerator based oscillation experiments, and is often the main signal channel for oscillation measurements. This note focuses on the study of ν_μ induced CC events containing a reconstructed muon candidate and at least one proton candidate. In future measurements, this channel will allow for the disentanglement of effects from neutrino interaction models, Final State Interaction (FSI) models and other theoretical concerns.

The MiniBooNE results were the primary indication for extending the 2p–2h mechanism (well known process in nuclear theory) into neutrino experiment interpretation; this is significant because the final state has an enhancement in nucleon multiplicity and can modified the final muon and hadron kinematics. Using the same CCQE-like definition mentioned earlier, MINERvA [9] and T2K [11] have used measured cross sections where both true CCQE and 2p2h events are included in signal while suppressing the dominant background (pion production). Pion production events where the pion is absorbed in the nucleus were also included.

Each detector experiment has a characteristic minimum proton energy based upon detector performance and beam properties, 102.3 MeV (450 MeV/c on momenta) for MINERvA and 500 MeV/c momenta for T2K. The identification of low momentum protons at MicroBooNE, along with an increased detector angular acceptance, gives greater sensitivity to nuclear structure and proton and/or pion FSI in studying these events. Although all Monte Carlo event generators now include 2p2h mechanisms, data samples gives only indirect evidence for it and the improved proton identification is important to have direct evidence. To study the 2p–2h mechanism with greater sensitivity, an explicit selection of ν_μ Charged Current two proton candidates is performed as a sub–set of the CCNProton sample (CC2Proton). We study the CC2Proton as a function of several reconstructed observables:

- Muon momenta and angle with respect to neutrino beam
- Protons momenta and angles with respect to neutrino beam
- Angle between the two protons

We present studies that, at the MicroBooNE detector with our current status of simulation and reconstruction, the proton must have a momentum above 300 MeV/c (kinetic energy of 46.8 MeV) to be detected and reconstructed.

Proton multiplicity is explored within this study due to its importance in order to achieve more accurate neutrino energy measurements, this is of strong importance for experiments with

unknown neutrino energy. The measured proton multiplicity in this note is compared with our default neutrino event generator, GENIE. However, in the distributions shown on Figure 1 we compare the predicted proton multiplicity with different generators. In this case, no reconstruction have been applied and they are only simulated ν_μ events in the MicroBooNE detector from the BNB beam with no other hadrons than protons in the final state (CCNp events, N=>0), and proton multiplicity accounts for protons above $300 MeV/c$ in momentum. As seen in Figure 1, when comparing different neutrino event generators, in this case two different model choices for GENIE (left plot GENIE Default, right plot GENIE Alternative, see Section 2.1) and NuWRO (bottom plot, see [13]), proton multiplicity can differ and even more noticeable is that the predicted composition at each multiplicity is as well different along the different generators.

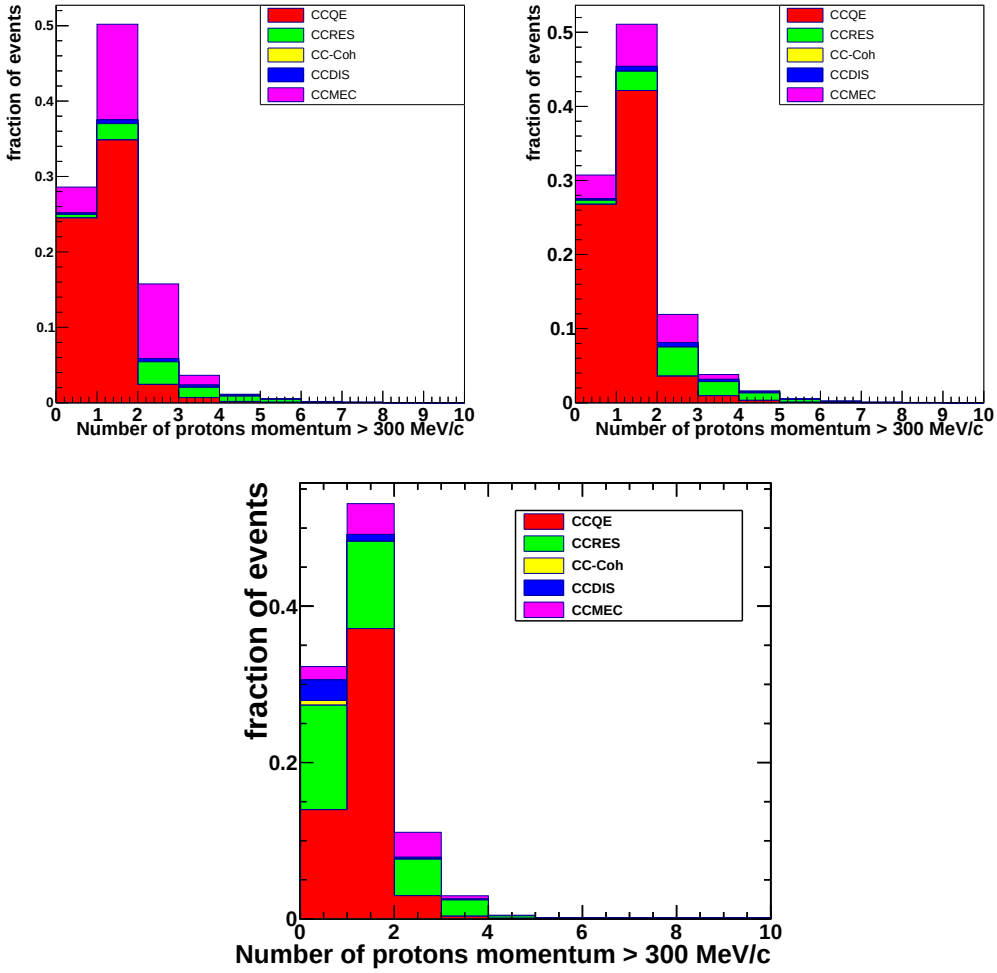

Figure 1: Comparison of the predicted proton multiplicity with different generators: GENIE Default (top left), GENIE Alternative (top right) and NuWRO (bottom). In this case, no reconstruction have been applied and they are only simulated ν_μ events in the MicroBooNE detector from the BNB beam with no other hadrons than protons in the final state (CCNp events, N=>0). Proton multiplicity accounts for protons above $300 MeV/c$ in momentum. Events are normalized to unity.

In the study described in this note we investigate events with at least one proton above $300 MeV/c$ in momentum. For these events we can infer the neutrino energy by using the same methodology applied by the ArgoNeuT experiment [1]. We calculate the neutrino energy as:

$$E_\nu = E_\mu + \sum T_p + E_{separ+excit} + T_{recoil} \quad (1)$$

$$T_{recoil} = \sqrt{(P^2_{T,miss} - M^2_{A-np})} - M_{A-np}$$

where, M_{A-np} is the mass of the nucleus, and depends on the number of protons in the final state. Excitation energy of argon, $E_{separ+exit}$, equals to 30.4 MeV and $P_{T,miss}$ is the missing transverse momentum. The accuracy of this approximation, when applied into reconstructed CCNp events in MicroBooNE, is shown on Figure 2. In this Figure we perform the distribution of the true neutrino energy minus calculated neutrino energy stacked by neutrino interaction type, for GENIE Deafult (left) and GENIE Alternative (right).

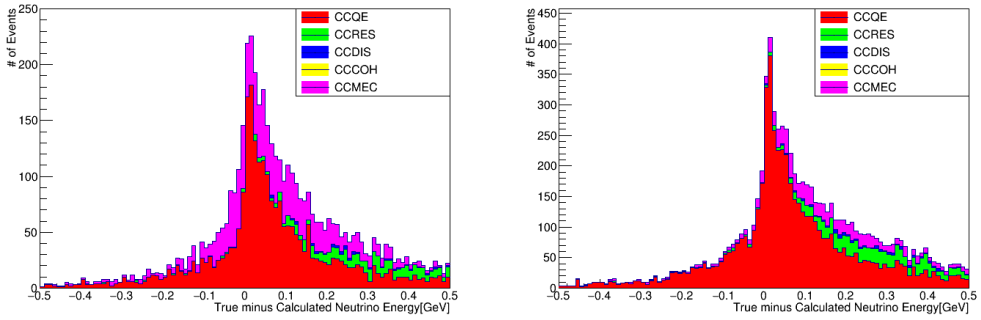

Figure 2: Distribution of true neutrino energy minus calculated neutrino energy for GENIE Default (Left) and GENIE Alternative (Right).

Additionally, the CC plus two proton exclusive channel (CC2p) is studied because of the significance of this sample with regards to meson exchange current (MEC) production in neutrino interaction models, the impact of pion absorption processes, and sensitivity to short range correlations (SRC) in nuclear models.

The ArgoNeuT collaboration investigated the sample of CC2p with lower proton momentum threshold ($200 MeV/c$) in order to investigate SRC effects within the Ar nuclei [1]. ArgoNeuT collected 30 events in this topology. The study for SRC was performed for two different observables, requiring the two protons to be above the Fermi momenta in Ar ($250 MeV/c$). The first observable investigated by ArgoNeut was the opening angle between the two emitted protons in the lab frame. In MicroBooNE, we are able to produce same study with higher statistics, see Section 6. The main differences to the ArgoNeuT study is the lower ν energy of the BNB with respect to the beam energy at ArgoNeuT and a slightly higher proton momentum threshold (being currently $300 MeV/c$ at MicroBooNE). In Figure 3 we show the cosine of the opening angle between the two protons in the lab frame according to GENIE Default (top left), GENIE Alternative (top right) and NuWRO (bottom, NuWRO with same configuration as [17]), no reconstruction has been applied and we account for protons above $300 MeV/c$ in momentum. We can observe big differences along the three simulations. In particular MEC and resonant pion production (we understand the pion has been absorbed and produce two protons in the final state) shapes are different along the simulations. In the two GENIE configurations there are big differences as well due to the different contribution per channel, mostly due to MEC impact. It ishttps://v2.overleaf.com/project/5b9ae2ea021c98798aee8d2c noticeable that we expect to have an important contribution from both resonant pion production and MEC in the back-to-back events in the lab frame distributions, the so-called hammer events, as discussed in [1] and [17]. However, GENIE Default is not predicting a comparable amount of resonant production with respect to MEC. As well, it is important to notice that NuWRO predicts a second peak at $cos\theta \sim 0.1$ which is not predicted by GENIE. In Section 6, we compare GENIE to the MicroBooNE data. We have not included yet in this analysis all the MicroBooNE open data, which will be included once systematical error performance is evaluated for this analysis. We expect then to be able to discriminate within these different generators and model choices.

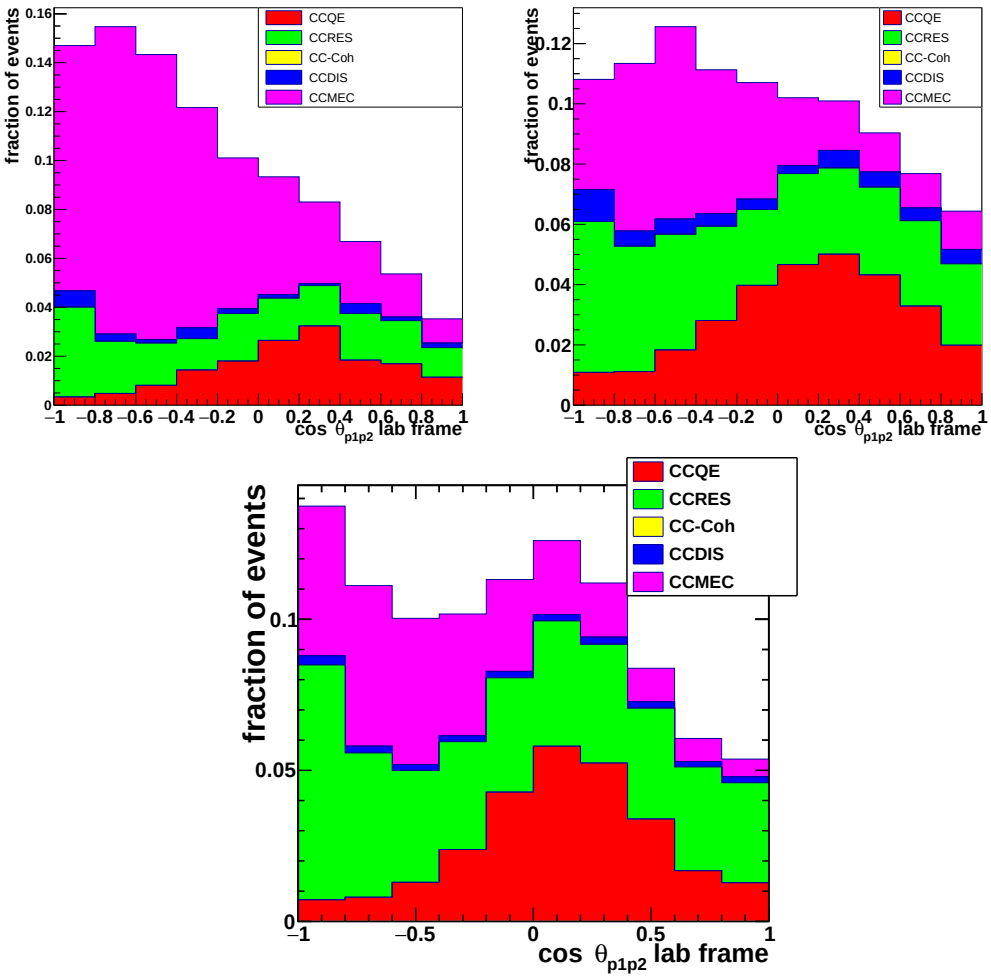

Figure 3: Opening angle between the two protons in the lab frame according to GENIE Default (top left), GENIE Alternative (top right) and NuWRO (bottom) in the lab frame for the CC2p topology. Events are normalized to unity.

The second observable studied by ArgoNeuT was the two body current events trying to reconstruct initial nucleon-nucleon configuration. In this case, it is performed the cosine of the angle between the least energetic proton and the struck nucleon. The struck nucleon 3-momenta, $\vec{p_n}$, is calculated using:

$$\vec{p_n} = \vec{p_1} - \vec{q_3}$$

Where, $\vec{p_1}$ is the 3-momenta of the most energetic proton and the 3-momentum transfer, $\vec{q_3}$, is calculated using the approach explained before in this Section. In Figure 4 we show the cosine of the opening angle between the least energetic proton (p_2) and the struck nucleon (p_n) in the center of mass (CM) frame according to GENIE Default (top left), GENIE Alternative (top right) and NuWRO (bottom), for the CC2p topology. No reconstruction is applied and we only account for protons with momenta above $300 MeV/c$.

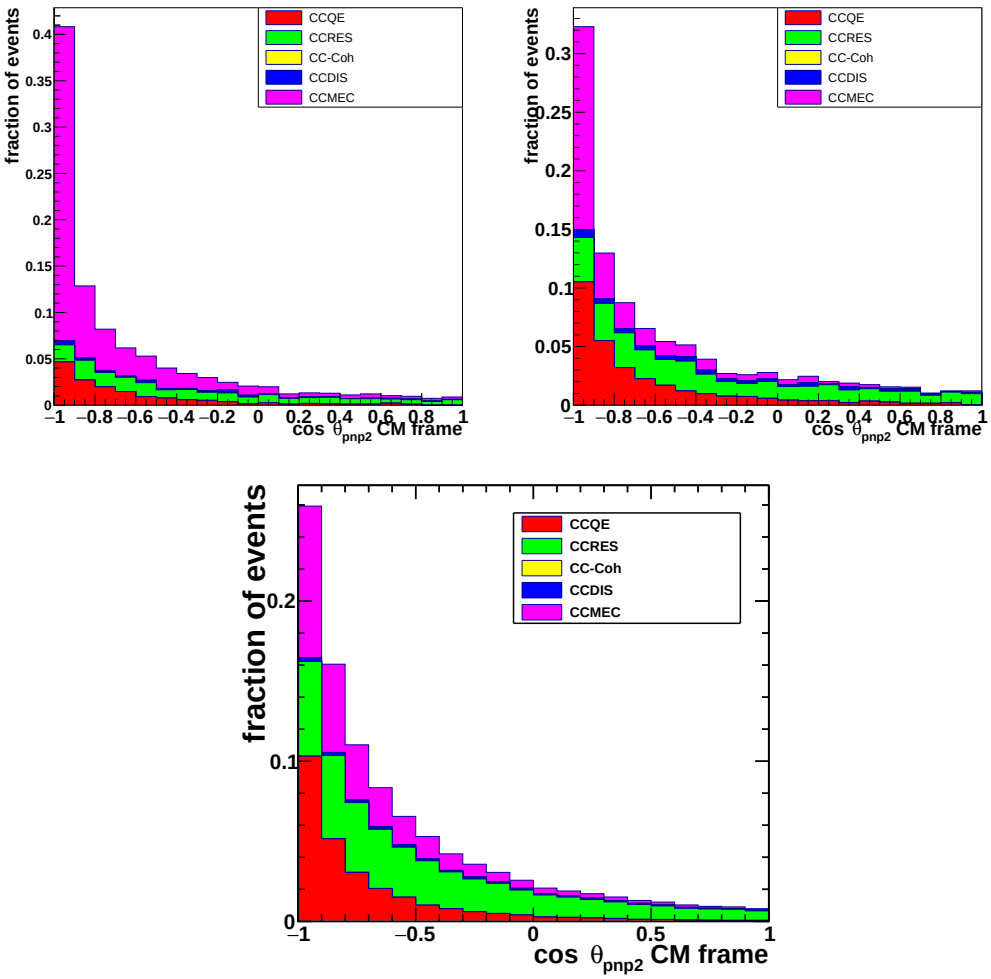

Figure 4: Cosine of the opening angle between the least energetic proton (p_2) and the struck nucleon (p_n) in the CM frame according to GENIE Default (top left), GENIE Alternative (top right) and NuWRO (bottom). No reconstruction is applied and we only account for protons with momenta above $300 MeV/c$. Events are normalized to unity.

For Figure 4 we actually observe that distributions along the different generators are not as different in shape, except for the most back-to-back bin. For that bin the increased contribution in both GENIE choices with respect to NuWRO is strong, and comparisons with our data will be of high importance to understand this effect. In this analysis, we have not applied a cut on $P_{T,miss}$, which will be investigated before to apply. The back-to-back events int he CM are expected to come from CCQE (mostly SRC CCQE events, but as well some CCQE without SRC) and some MEC. It is surprising the hug amount of MEc events produced in the back-to-back configuration by GENIE. We expect that our data comparisons for the two angles introduced here, which were first studied by ArgoNeuT, will help to understand the different simulations.

2.1 MicroBooNE simulation

MicroBooNE performs a simulation chain as described in [8], for the flux, neutrino interaction, overlay of cosmic background and detector simulation. We perform all distributions comparing data to GENIE $v2-12-2$. There are two different GENIE model choices when comparing with of our data with simulated BNB neutrino interactions, "GENIE Default" and "GENIE Alternative". We refer to our baseline GENIE simulation as "Default Genie + Empirical MEC", and is based on the default Genie model set, including MEC events [16]. The alternative set of models we are comparing to in this note is referred to as "Genie Alternative". Details of both versions are

listed in Table 1. The GENIE default model set is the basis for all MicroBooNE analyses to date. Nevertheless, the models there are most applicable to the energies of the MINOS detector (3-7 GeV) rather than MicroBooNE (0.2-2 GeV). The addition of multi–nucleon interaction effects ($2p2h$) through the GENIE Empirical-MEC model is tuned to get good agreement with the MiniBooNE data. The main differences between the GENIE Default and GENIE Alternative are described in Table 1. The main difference is a complete replacement of the nuclear model and quasielastic-like models. The local Fermi gas (LFG) [14] [12] model has a more realistic momentum distribution and replaces the Relativistic Fermi Gas model. The combination of the Nieves quasielastic model [15] and the Valencia $2p2h$ model [14] gives a better description of the MiniBooNE quasielastic-like data and replaces the previous CCQE (Llewellyn-Smith) LS Model and Empirical-MEC model. At neutrino energies less than 1 GeV, added effects of RPA (long range nucleon-nucleon correlations) and Coulomb distortion on the outgoing muon have been shown to be important. Each model set describes MiniBooNE data [10] reasonably well but that doesn't have to hold for MicroBooNE data because of greater sensitivity to nuclear structure and FSI in a heavier nucleus.

Model element	GENIE Default	GENIE Alternative
Nuclear Model	Bodek-Ritchie Fermi Gas	Local Fermi Gas
Quasi-elastic	Llewellyn-Smith	Nieves
Meson-Exchange Current	Empirical	Nieves
Resonant	Rein-Seghal	Berger-Seghal
Coherent	Rein-Seghal	Berger-Seghal
FSI	hA	hA2014

Table 1: GENIE Deafult and GENIE Alternative model choices.

MicroBooNE utilizes the *Corsika* simulation package to simulate cosmogenic events that overlap in time with the GENIE-generated neutrino interaction and are then included in the same simulation of energy deposition and detector response. To reproduce events in the trigger time window that may not produce any neutrino interaction event, MicroBooNE uses off-beam data instead of simulation, see Section 2.2. Then, comparisons of on-beam data are done with respect to MC+off-beam events being normalized such that the total yields are equal (area normalized) and maintaining the *a priori* mixture between MC and off-beam data.

2.2 Data and Monte Carlo data sets

MicroBooNE started taking neutrino interaction data in October of 2015. The data set used in this analysis ranges from February to July 2016. Two different data streams are utilized in this note:

on-beam: Taken only when the arrival of the neutrino beam spill from the BNB is expected.

off-beam: Taken during periods when no beam was received and with the same detector conditions as the on-beam data. It is used for a data-driven measurement of cosmic backgrounds.

The trigger summary, POT (Protons On Target), and total number of events for the samples used in these studies are summarized in Table 2.

Sample	P.O.T	Trigger Count
BNB on-beam	$4.411e19$	9775610 BNB Spills
BNB off-beam	N/A	73710197 EXT Triggers
BNB+cosmic MC (GENIE Default)	$2.010e20$	N/A
BNB+cosmic MC (GENIE Alternative)	$4.087e20$	N/A

Table 2: The detector datasets used for this analysis along with the simulated BNB samples used to predict the event yield. The factors used for normalization are listed. GENIE Deafult and GENIE Alternative model choices are described in the previous section. EXT Triggers is the number of triggers recorded orthogonal in time to the arrival of a BNB spill.

We use the POT and trigger information to normalize both the MC and off-beam data to our on-beam data. However, for the purposes of the analysis presented here, we applied an overall are normalization once the MC and off-beam event yields are established. We will present POT

normalized distributions once full uncertainty studies have been completed. For now, shape-only comparisons with the two GENIE choices are the main focus of this note.

3 The ν_μ Charged Current inclusive selection

The analyses presented here use as their pre-selection the ν_μ CC inclusive identification as defined for the Neutrino 2018 analysis, which is described in MicroBooNE public note [8]. This pre-selection focuses on isolating events with a single muon that was the result of a ν_μ interaction, while rejecting muons resulting from cosmic interactions. The pre-selection uses a mixture of optical and geometric information to enhance the neutrino signal over the cosmic background, and accepts contained and uncontained tracks. The ν_μ CC inclusive selection does not contain explicit cuts on track direction or final state particle or track multiplicity, and so does not bias subsequent selection of different topologies. The selected muon candidate in the event is the track with the longest reconstructed length and can be contained or uncontained in the detector[1]. The final efficiency, defined as the number of CC-ν interactions within the fiducial volume after selection divided by the number of CC-ν interactions generated within the fiducial volume is $\approx 55\%$. About 30 % of the events which pass selection do not contain a neutrino event, as measured in the off-beam data.

4 Particle identification in LArTPC

After the ν_μ CC inclusive sample is selected, we study the discrimination between MIP and non-MIP particles to identify protons in the final state of the ν_μ interaction. We perform this study for all reconstructed tracks (excluding the longest track within the event which is most commonly a muon), and do not include reconstructed showers in candidate sample. Once these tracks are identified from Pandora reconstruction [5], particle ID will be performed using the energy loss as a function of residual range and identify the Bragg peak signal that will allow us to use the full power of the LArTPC granularity. Since protons from a ν interaction are unlikely to exit the detector at the MicroBooNE energies (200 - 1500 MeV in neutrino energy) and exiting tracks would have a biased residual range, we will consider as proton candidates those tracks that are fully contained in the containment volume ('CV'). Studies were done separating fully contained tracks from those exiting the detector (using the CV definition) and, as expected, a strong bias was seen in the uncontained sample.

The samples used in this analysis have been calibrated for dQ/dx [2] using off-beam data (dominated by muons). The calibration was done to get a flat detector response, after space charge effects [3] applied in simulation, and angular distortions [6].

After applying the dQ/dx calibration, the dE/dx has been estimated based upon the Box model [3] [4]. A full study with a high-purity proton sample in data for the recombination model characterization has been done and an *effective parametrization* has been applied in data. This *effective parametrization* corrects for effects observed in data indicating lower recombination than the expected one. It was observed than the previous parameterization was implying a higher recombination effect than that observed in data when calculating the dE/dx using a theoretical approximation (using the residual range [4] information to obtain the dE/dx). As the calibration and E-field mapping are convoluted in the recombination formula, a decision was made to correct for this convoluted effect by modifying the Box model parameters in the data. An improved approach is in development in order to discriminate calibration, E-field mapping, and recombination parameters for a final measurement of recombination in the MicroBooNE detector. These *effective recombination* parameters were not applied to the Monte Carlo simulated samples. However, the MC is simulated with the same parameters that are used in the MC reconstruction. It should be

[1] It should be noted that we use two different volume requirements in event selection. The "fiducial volume" is a smaller volume in the active detector and the candidate event vertex must be within this fiducial volume. The "containment volume" is a less restrictive volume and all non–muon–candidate tracks must be contained within the containment volume. Section 5.2 gives full details of the two volumes.

[2] dQ/dx is the deposited charge in a given hit in a wire.

[3] The space charge effect is the build-up of slow-moving positive ions in a detector due to, for instance, ionization from cosmic rays, leading to a distortion of the electric field within the detector. This effect leads to a displacement in the reconstructed position of signal ionization electrons in LArTPC detectors, as well as variations in the amount of charge quenching experienced by ionization throughout the volume of the TPC.

[4] The residual range is defined as the distance from a given hit to the last hit in the track trajectory.

noted that only the collection plane hits [5] are used for the particle identification due to a better understanding and modeling of the calorimetry and detector response, and due to the status of our reconstruction and simulation at this stage.

The new particle identification (PID) method is based on a χ^2 calculation (using dE/dx and residual range) between the measured response and the predicted proton hypothesis from Geant4 simulation. Geant4 uses a Bethe-Bloch formula for predicting proton dE/dx versus residual range, and applies smearing on the calculated values to try to mimic the detector response in data. The dE/dx *vs* residual range distributions in argon for the different particle hypotheses are shown in Figure 5. They show the theoretical distributions using Geant4 simulation. Each data point corresponds to the smeared dE/dx value at its corresponding residual range.

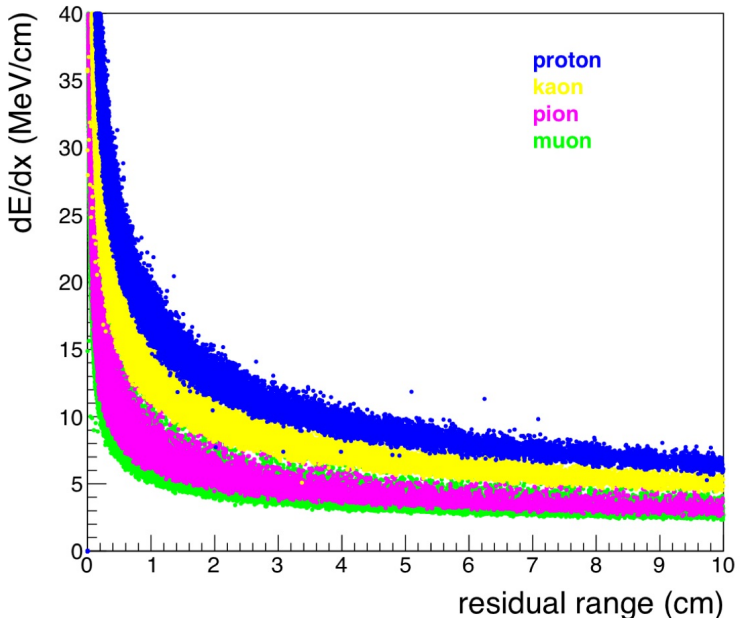

Figure 5: Theory distributions of the dE/dx with respect to the residual ranges of the different particle types in argon, using a Geant4 simulation. Each point corresponds to a dE/dx value at its corresponding residual range value.

We extract for each bin in residual range the mean value of the dE/dx from the Geant4 curves, Figure 5. Within the Geant4 simulation, the residual range is samples with a bin size of 0.08 cm.

For each selected track (either in MC or data) we obtain the reconstructed dE/dx per hit and calculate the χ^2 between the track and the Geant4 determined mean dE/dx per residual range bin (from the simulated values, see Figure 5). These χ^2 values are then summed for all hits on the selected track, excluding first and last hit from the reconstructed tracks. To avoid mis-measurement of the range, we exclude the first and last hits of the track from this calculation. This is due to the fact that the residual range calculation for the first and last hit may be wrong since the exact position of the hit between the wires is unknown. It also avoids complications from particle activity effects (*i.e.* scattering) and energy deposition overlap at the interaction vertex. The χ^2 value is then normalized by the number of degrees of freedom (ndof), which correspond to the number of hits in collection plane.

The χ^2 value can be determined for several particle assumptions, but in this analysis we exclusively use the proton hypothesis (thus the *proton* subscript).

$$PID = \chi^2_{proton}/ndof = \sum_{hit}(\frac{(dE/dx_{measured} - dE/dx_{theory})}{\sigma_{dE/dx}})^2/ndof$$

Where $\sigma_{dE/dx}$ is the estimated resolution of the dE/dx. Values used for error resolution in dE/dx come from studies performed by the ArgoNeuT Collaboration[2]. This does not take into

[5] MicroBooNE has three wire planes with different orientation with respect to each other. The collection plane wires are vertical with respect to the neutrino beam direction.

account that MicroBooNE has $3mm$ wire pitch (instead of the $4mm$ in ArgoNeuT), which should overestimate the error in resolution. We prefer to use this conservative approach due to the large smearing in the dEd/dx observed in the MicroBooNE detector.

In Figure 6, we show the distribution of the calculated PID ($\chi^2_{proton}/ndof$) values for all tracks within the selected ν_μ CC inclusive sample. The left plot is for contained tracks in the event being reconstructed as part of the ν_μ CC interaction; right plot is for uncontained tracks. The predicted yield from simulated data plus off-beam data is normalized to the number of selected events in on-beam data. While there are some data/MC discrepancies, both MC and off-beam data (being dominated by muons) show the separation power of the new proton identification method between MIP/non–MIP particles using the PID value.

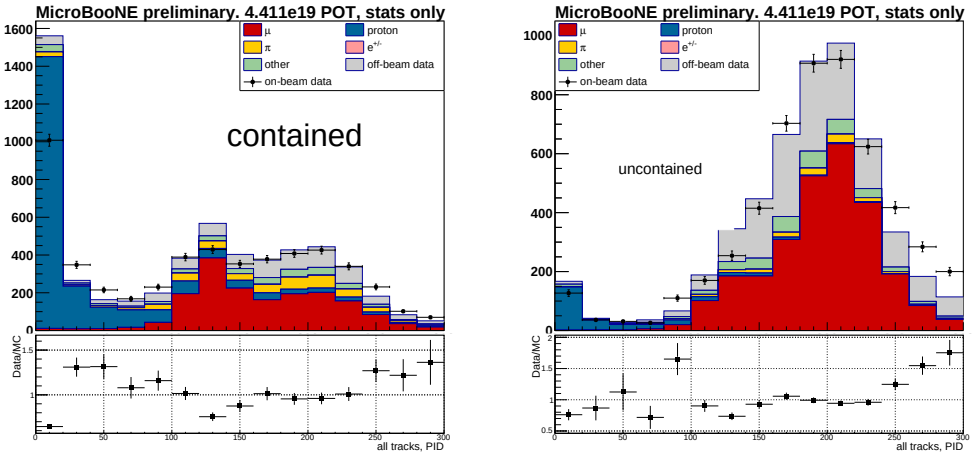

Figure 6: $\chi^2_{proton}/ndof$ values for the tracks within the selected ν_μ CC inclusive sample. Left plot shows contained tracks ($\chi^2/ndof_{data/MC} = 441.3/14$, includes statistics only) and the right plot shows uncontained tracks ($\chi^2/ndof_{data/MC} = 493.2/14$, includes statistics only). The total predicted yield is area normalized to the number of selected events in on-beam data.

The separation of tracks into contained and uncontained samples allows us to undertsand the importance of the presence of a Bragg peak in particle identification. From Figure 6, we observe that uncontained tracks are shifted to higher values of $\chi^2_{proton}/ndof$. This is reasonable since uncontained tracks will not have Bragg peak energy deposits. Therefore the detector response is less likely to appear proton-like regardless of the actual particle type. While contained muons will have lower χ^2 values closer to proton–like than uncontained muons due to the Bragg peak at the stopping point of the muon. We can also interpret the sample of protons (observed in MC) outside the proton–like region as protons which interact with the detector material (argon), then there is no Bragg peak energy deposition. Improvements in the ability to identify interacting protons after scatters will be a focus in the next generation of analysis in MicroBooNE. For this first stage of proton selection, they will be part of our inefficiency and not selected. The effect is quantified in Section 5.2 discussing proton identification efficiency determinations.

Looking at the distributions in data and MC, we are confident that the proton identification has improved purity compared to previous identification methods and can be used for topological selection.

Taking into account both the particle power discrimination (muon *versus* proton) and the data/MC shape agreements, we decide to apply for the muon and proton candidates a PID requirement:

- muon candidate has a PID ($\chi^2_{proton}/ndof$)>88.
- proton candidates have a PID ($\chi^2_{proton}/ndof$)<88.

When considering this choice of PID value, we considered the data/MC comparisons but also the on-beam *vs* off-beam yields. Since we expect the off-beam data to be dominated by muons, the sample allows us to determine the MIP passing rate for the proton ID discriminant. The chosen cut value used for this PID method have been determined by studying both efficiency and purity

of the proton and muon candidates. The chosen PID cut value is the one that optimizes both efficiency and purity for the proton candidates. In Tables 3 and 4, the purity, both before and after the PID cut, is listed along with the relative efficiency (after the PID cut with respect to before the PID cut) for the selected sample of one muon and at least one proton sample. The only conditions applied (before and after the PID cut) is for the muon candidate to be the longest track and proton candidates to be contained in the FV. All tracks have at least five hits in the collection plane. In Table 4, all the protons candidates are considered in the calculations. For the leading proton candidate, longest candidate track in the event, purity and efficiency have been estimated to be higher than for the shorter proton candidates. However, both efficiency and purity of the method are excellent when considering all the tracks in the event.

Muon enhanced sample	Before PID requirement	After PID	relative efficiency after PID requirement
true μ	93.0%	94.0%	99.14%
true proton	1.8%	1.0%	–
other	5.2%	5.0%	–

Table 3: Composition of the selected muon candidate sample before and after the PID requirement. Relative efficiency is included.

Proton enhanced sample	Before PID requirement	After PID	relative efficiency after PID requirement
true μ	18.7%	2.0%	
true proton	64.5%	92.6%	85.2%
other	16.8%	5.4%	

Table 4: Composition of the selected proton candidates before and after the PID requirement. Relative efficiency is included.

Several studies has been performed in order to understand better the data/MC discrepancies in our PID discriminant. Figure 7 shows the dE/dx versus residual range before any PID requirement (top plots) and after the PID requirement for the proton candidates, bottom plots. By looking into the dE/dx *vs* residual range for all contained tracks when requiring χ^2_{proton}<88 (proton requirement) on on-beam data (left) we observe the main reason of the disagreement at low χ^2_{proton} values: dE/dx values in data are more smeared than in our MC simulation. This larger smearing in data biases the χ^2_{proton} values in data to higher values and the over expectation in MC at lower χ^2_{proton} values. In Figure 7, data are compared to the theoretical predictions according the Bethe-Bloch equation and agreement between the selected candidates and the central Bethe-Bloch prediction can be seen.

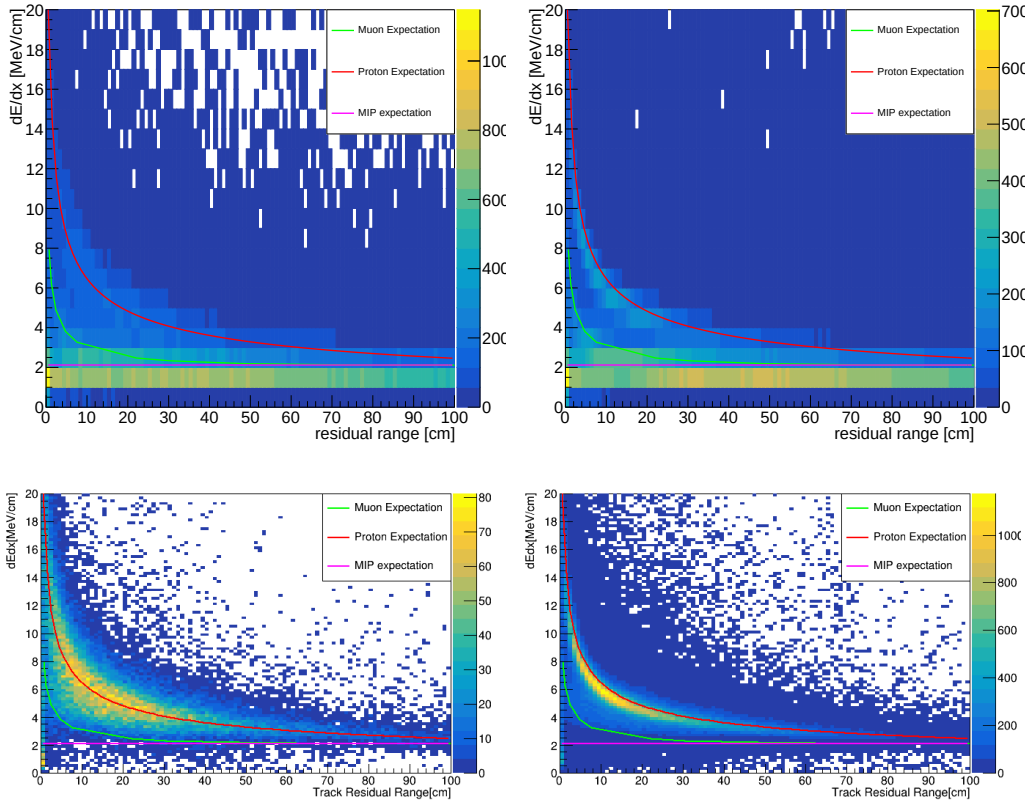

Figure 7: dE/dx *vs* residual range for all contained tracks within the selected ν_μ CC inclusive sample. Top plots for all the tracks before PID requirement. Bottom plots with PID requirement, χ^2_{proton}<88 (proton requirement). on-beam data, left, MC in right plots. Theoretical predictions according the Bethe-Bloch equation for the different particle types are included.

We are currently working on perform another PID technique, on the top of this one, for those tracks that have not been selected as a protons. We know that scattered protons don't produce Bragg peak and for this reason we loos identification power for the most energetic protons. However, these protons can be recovered and investigations on performing efficiently and with high purity is ongoing.

5 The ν_μ Charged Current N Proton Selection

In this section, we describe the signal and background definitions along with the final event selection. After determining the χ^2_{proton} threshold for optimal proton selection, the proton identification is used to select ν_μ CC candidate events with at least one proton.

5.1 Signal Definition and Background Categories

The signal is defined as ν_μ charged current interactions with no pions in the final state and at least one proton. A momentum threshold on protons to be greater than 300 MeV/c has been included in order to account for our current efficiency detection in MicroBooNE. The determination of this momentum threshold is discussed in the next subsection.

Backgrounds are separated into several categories:

- Cosmic Ray particles (off-beam): After requiring an online BNB trigger, around 90% of recorded events still do not contain a neutrino interaction in the TPC. Any cosmic ray particles mimicking neutrino interactions are placed in this category.
- ν_μ CC0π0p: This category encompasses any ν_μ CC interaction with no pions or protons above the energy threshold.

- ν_μCC other: This category consists of events where at least a pion was produced in the final state of the interaction.
- Other: All other event classes are considered to be "Other". This is primarily NC events and other interactions from outside of the fiducial volume, and also includes a small number of ν_e events.

5.2 Event Selection Scheme

To select signal candidate events, we implemented the CC inclusive selection as pre-cuts, as described in section 3. The CC inclusive selection drastically reduces most of the background from cosmics and NC events. To further separate the signal from background after the CC inclusive selection, the following cuts were performed:

- **Number of reconstructed tracks greater than 1**: There must be at least two tracks associated to the neutrino vertex candidate. The longest track is selected as the muon candidate, all other tracks associated to the same neutrino vertex are considered as proton candidates.
- **All event vertices in Fiducial Volume**: From the CC inclusive selection, all the event vertices are required to be inside Fiducial Volume (FV) defined in Figure 8:

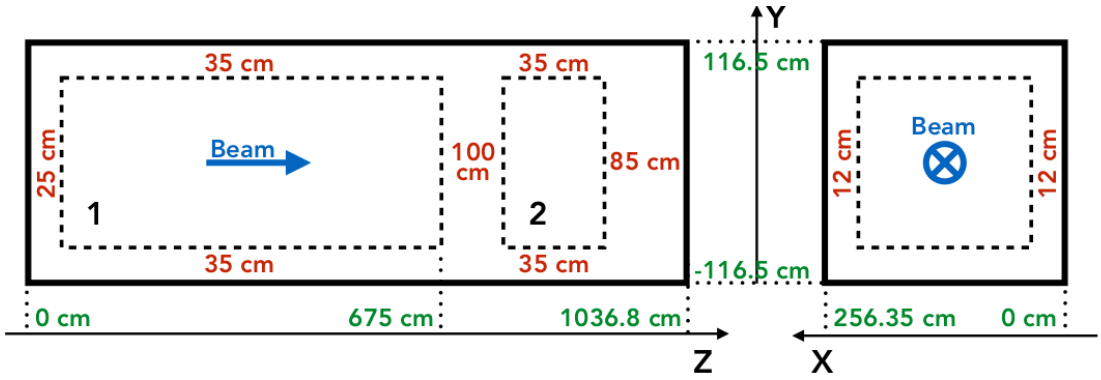

Figure 8: Fiducial Volume used in CC inclusive selection for the reconstructed neutrino vertex cut [8].

Note that the 100 cm gap in z direction is due to the dead wires in the TPC collection plane.

- **Proton candidates tracks contained**: All proton candidate tracks must be contained where the large Containment Volume (CV) is defined as:
  - 10 cm in x-direction from any TPC edge (drift direction)
  - 20 cm in y-direction from any TPC edge (vertical to the drift direction)
  - 10 cm in z-direction (beam-direction) from any TPC edge

  This CV volume requirement is applied to the end point of each proton-candidate track, since the start of the track follows the FV condition defined before. The CV allows for increased efficiency when selecting proton candidate tracks while still rejecting cosmics and dead detector regions during vertex selection.
- **Minimum number of hits**: all proton-candidate tracks are required to have at least 5 hits in the collection plane. This is a quality cut to ensure high quality reconstruction but also due to the PID method exclusion of the first and last hits on the track. The effective minimum number of hits used later in the PID χ^2 method then becomes three.
- **PID cut**: Based on the study in Section 4, we used $\chi^2_{proton} < 88$ as the cut value of the proton identification.

In constructing the event topology, the longest track is considered to be the muon candidate and can exit the detector. If it exits the detector, the momentum is estimated from the scattering of the track along its path (referred to as Multiple Coulomb Scattering, or MCS [7]). If it is contained in the containment volume, the momentum is calculated by the track range. Since all proton candidate tracks are contained, the momentum is based on the track length.

Table 5 shows the event composition of all the ν MC events using only MC truth information, no selection applied, it includes events outside the FV. CC Incl is the total number of true CC inclusive events (CC inclusive defined as all the CC events coming from ν_μ CC events with the neutrino vertex within the FV shown in the Figure 8). CC0πNp is the total number of charged current events with zero pions and at least one proton within the same FV. CC0πNp w/ p_{thresh} is the total number of events with at least one proton momentum greater than 300 MeV/c within the FV. These samples are used to determine the signal efficiency and proton ID efficiency.

Total MC Evts	CC Incl (in FV)	CC0πNp (in FV)	CC0πNp w/ p_{thresh} (in FV)
100%	16.8%	11.9%	8.7%

Table 5: The event composition in the MC sample from MC truth information without any requirement on reconstruction. No selection applied.

5.3 Proton Momentum Threshold Determination

The proton selection cuts, described in the previous Section 5.2, are used to select proton candidate tracks within the CC inclusive sample. But in order to classify an event into a given signal topology, a proton momentum threshold which takes into account the detector acceptance is applied to divide events between 0-proton, 1-proton, 2-proton, etc. To decide upon a proton momentum threshold, the proton identification efficiency was studied. Figure 9 shows the selection efficiency with respect to the leading proton momentum per each event (GENIE Default and GENIE Alternative are shown). Based upon the efficiency distribution, a proton momentum threshold of 300 MeV/c was chosen to separate the candidate events into the different topologies. This momentum threshold translates into approximately 47 MeV in kinetic energy and a path length of approximately 1.5 cm in Ar for a proton. This agrees well with the requirement that each track has at least 5 collection hits considering the $3mm$ wire pitch in MicroBooNE. Both the Default and Alternative GENIE MC give results that are equivalent within statistical uncertainties and gives confidence that the efficiency is neutrino-interaction model independent, at least for these two modeling choices.

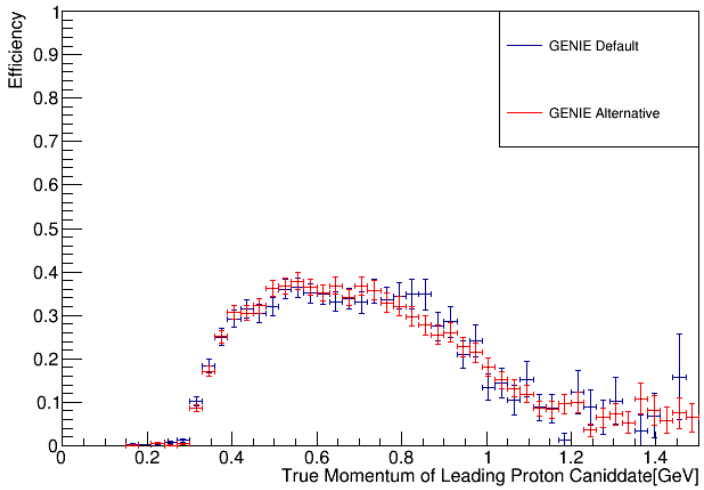

Figure 9: Selection efficiency for the signal sample with respect to the leading proton momentum.

An interesting effect is the decreased identification efficiency for protons with momentum above 800 MeV/c. High momentum proton inefficiency was suspected to be due to the hadronic re-

interactions and therefore the absence of a Bragg peak. There was some concern that the containment cut (CV) could cause the drop in efficiency, but it was found that after the vertex FV cut, only 2% of protons generated out of the nucleus escape the CV. Within the simulated sample, GEANT4 classifies the last process to affect a final state particle at the end of its trajectory, and we can differentiate between particles whose trajectory ends with a Bragg peak from those that have an inelastic process. Figure 10 shows the probability of an inelastic scatter for the protons in the liquid argon as a function of the proton momentum. With increasing momentum, the interaction probability goes to almost 100%. Then there will be no Bragg peak at the track end. These inelastic interactions are the origin of the decreased proton identification efficiency at higher proton momentum. At low momentum, the efficiency drops to almost 0 at 300 MeV/c. We therefore use this value (300 MeV/c) as the momentum threshold for defining the signal topology in terms of number of protons.

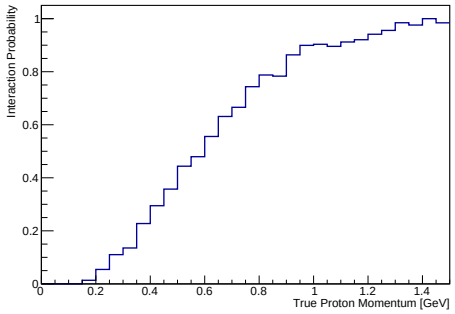

Figure 10: Interaction probability of the leading proton momentum.

Using the full selection and the 300 MeV/c proton momentum threshold, Table 6 shows the purity and efficiency in the CCNProton selection, for selected CCNProton events in MC (GENIE Default).

Cuts	N-selected	Purity	Efficiency
pthresh	6356	0.762	0.286

Table 6: Number of CCNProton selected events in MC, purity and efficiency after requiring at least one proton with momentum greater than 300 MeV/c. According to GENIE Default.

Sample Normalization

- Normalization of Simulated BNB events to on-beam data exposure: to normalize the MC to the data, one must weight by the ratio of the generated POT exposure to the on-beam data sample exposure, which are given in Table 2.
- Normalization off-beam data to on-beam data: The off-beam data sample is used to measure the non-beam related backgrounds. To normalize properly this background to the on-beam sample, one needs the total number of external triggers before the software trigger and scale this exposure to the number of total BNB spills for the on-beam data sample. This is done using the information provided in Table 2.
- Normalization of combined predicted yield (area normalization): After adding the Simulated BNB events to the off-beam data, the combined prediction is then normalized to the number of selected events in on-beam data. Only the MC sample scaling is modified during this procedure with the contribution from off-beam data fixed.

Table 7 shows the number of data events after all the cuts including the proton momentum threshold cut. The decision was made to normalize the predicted event yield to the number of selected data events to emphasize shape differences between theory and data along with showing variations between the various theoretical models presented. An absolute normalization will be provided as soon as systematic uncertainty studies are finished.

Cuts	on-beam	off-beam	Scaled off-beam
pthresh	1133	502	66.6

Table 7: Event selection of on-beam and off-beam data after requiring there are at least one proton with momentum greater than 300 MeV/c.

5.4 Proton Multiplicity Study

We study the proton multiplicity in simulated data to understand signal topology migration. We show the number of true protons versus the number reconstructed protons in the MC sample. Table 8 shows the number of true protons and the number of reconstructed and identified proton candidates. These studies have been done with the GENIE Default sample and includes a proton threshold of 300 MeV/c.

	Nreco=1	Nreco=2	Nreco=3	Nreco=4	Nreco=5
Ntrue=1	4404	128	4	0	0
Ntrue=2	663	621	9	0	0
Ntrue=3	113	115	49	1	1
Ntrue=4	18	36	23	5	0
Ntrue=5	7	17	10	2	0

Table 8: Number of true protons and reconstructed protons after the proton momentum cut.

From this table we can see that most of the single proton events that have been selected in the CCNProton sample are reconstructed correctly (only one proton reconstructed). Single proton events that have no reconstructed proton are not included in this table and are part of our inefficiency. In some of the events with more than one proton, there are fewer protons that were reconstructed compared to the true number of protons. The inefficiency naturally increases with the proton multiplicity. Note that the threshold requirement in this table is only applied to the leading proton for this analysis and there are many additional protons below that threshold.

5.5 Signal and background analysis

Table 9 shows the relative fraction of CC (charged current) QE (quasi-elastic), MEC, RES (resonant pion production), COH (coherent pion production) and other events ν interactions in the selected sample in MC. These studies have been done with GENIE Default. Table 10 shows the composition of the signal before and after the momentum threshold cut with all the signals divided into 3 categories by number of protons.

	Total_sig	QE	RES	DIS	COH	MEC
without pthresh cut	4995	2778	524	75	0	1618
with pthresh cut	4861	2729	502	72	0	1558

Table 9: Composition of the signal without momentum threshold cut. GENIE Default.

	Total_sig	CC0π1P	CC0π2P	CC0πnP(n>2)
without pthresh cut	4995	3615	1049	331
with pthresh cut	4861	3954	800	107

Table 10: Composition of the signal without p threshold cut.

Some key points to note are:

- There is a bias towards forward-going particles.This is mostly due to the number of collection hits when the tracks are forward-going (the track being perpendicular to the collection plane wires). For backwards-going particles, they have lower momentum spectrum in the lab frame and therefore a lower identification efficiency.

- There are drops in efficiency when the muon or proton is aligned parallel or towards to detection wires. This is a limitation of the current reconstruction algorithm and is a known deficiency.

To examine the BNB-induced and cosmic backgrounds in more detail, it is useful to separate them into additional categories.

- CC0π0P: No pions were produced, also no protons above threshold.
- CC1πNP: One pion was produced and emitted in the final state of the interaction.
- CCNπNP: Multiple pions were produced and emitted in the final state of the interaction.
- CCν_e: The candidate tracks are from an electron neutrino interaction with any other final state particles.
- NC: The candidate tracks are from an NC interaction. This is usually an NC π, with a proton and a charged pion in the final state.
- OOFV: True interaction vertex is outside the Fiducial Volume, but tracks coming from the neutrino interaction are selected.
- Cosmic: Only cosmic-induced tracks are selected.
- Mixed: Of the two or more tracks selected, at least one is from true neutrino interaction, but another is from cosmic origin. This kind of backgrounds are combined with background from Cosmic and are shown in the column of the cosmic background in Table 11.

Cuts	CC0π0P	CC1πNP	CCNπNP	CCν_e	NC	OOFV	Cosmics
CCincl	26%	28%	3.1%	0.3%	3.2%	33.3%	6.2%
ntrks	12.3%	42.5%	5.2%	0.2%	3.3%	28.7%	6.8%
contained	13.9%	45.0%	5.1%	0.2%	3.8%	28.0%	4.0%
minColl	13.2%	45.6%	5.2%	0.3%	4.0%	27.8%	3.9%
PID	4.2%	45.9%	4.0%	0.4%	7.2%	35.6%	2.7%
pthresh	3.6%	46.3%	4.0%	0.5%	7.3%	35.5%	2.8%

Table 11: Composition of the beam-induced background and different selection cuts quoted as percentages of the total beam-induced background. In this table, all the backgrounds are categorized into 7 categories instead of 3 as mentioned in section 5.1. Cuts are defined as: CCincl (CC inclusive sample cut), ntrks (number of reconstructed tracks greater than 1), contained (proton candidates tracks contained), minColl (minimum number of hits), PID (PID cut), pthresh (proton energy threshold definition)

From Table 11 we can see that most of the backgrounds are from CC one pion events, and events from outside Fiducial Volume. Work is ongoing to reject events where a pion in emitted from the ν interaction and interacts to produce a proton(s), as well as to tag neutral pions through observing photons near the vertex. Some charged pions will reinteract too close to the neutrino-argon vertex for the detector or reconstruction algorithms to resolve the separation into the primary protons. These backgrounds will have to be constrained by direct measurements of charged pion production on argon in MicroBooNE and measurements of pion-argon scattering from test beam experiments (LArIAT).

The out-of-fiducial volume (OOFV) backgrounds can be reduced with an improved calibration for the space-charge effect, which is actively being developed.

5.5.1 Efficiency dependence on model

While we have identified some of the cuts made in this analysis that are known to lead to large changes in the efficiency across some kinematic variables it is important to verify that variations in the neutrino interaction models do not bias the proton identification efficiency.

Figure 11 and Figure 12 show the efficiency of the CCNProton sample as a function of 4 kinematic variables. Many of these variables are coupled through the interaction physics, as well

as being impacted by the detector geometry and selection cuts. This can make their interpretation complicated.

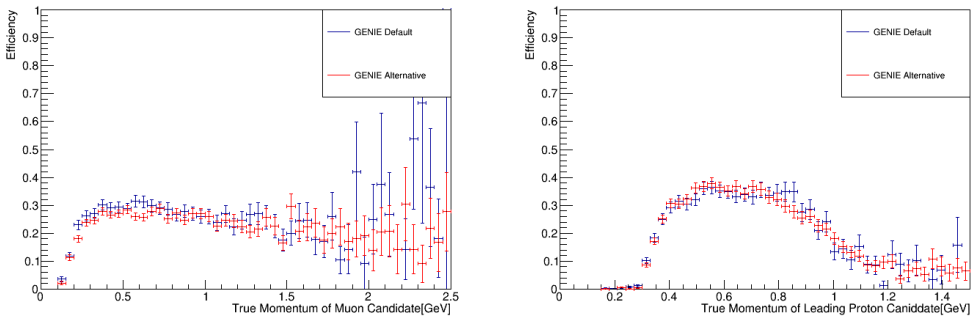

Figure 11: Signal reconstructed efficiency with respect to the muon candidate momentum and leading proton momentum before the momentum threshold cut.

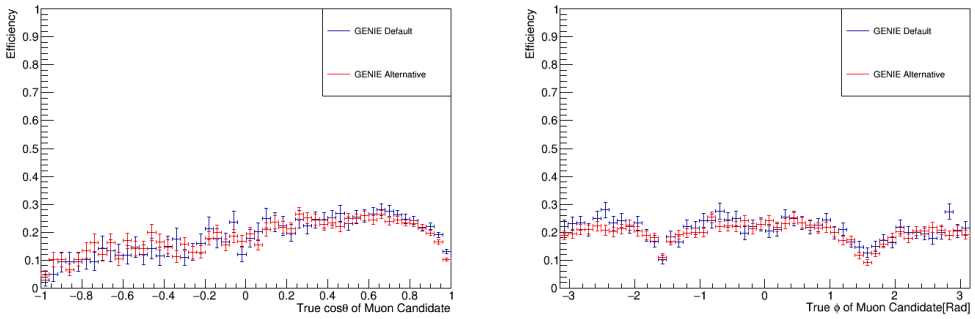

Figure 12: Signal reconstructed efficiency with respect to $\cos\theta_\mu$ and ϕ_μ before the momentum threshold cut.

When considering a variable such as muon momentum, the efficiency must be integrated over all other variables, and large variations in the efficiency over the angular and proton variables can translate to model dependence in the muon momentum efficiency. Given the large number of features we expected in certain regions of kinematic phase space, this is a concern. To test that we haven't introduced any large model bias, we consider the two sets of models within GENIE (both shown in Figure 11). Importantly, although the efficiencies do have several strong features, both GENIE model sets predict the same shapes, showing that there are no obvious or strong model dependencies within the GENIE context. The largest systematic deviations between the two model sets are of the order a few percent and not beyond the statistical fluctuations of the generated samples. More efficiency studies will follow using different generators than just GENIE to ensure model independence.

5.6 CCNProton Kinematical observables

In this section, we show the comparision between observed data and the predicted yield using the simulated BNB samples and the off-beam sample. All the distributions are made after the proton threshold cut performed to the leading proton candidates. The total predicted yield is area normalized to the number of selected events in on-beam data in such a way as to maintain the *a priori* mixture between simulated data and off-beam cosmic background.

Figure 13 to Figure 18 show the comparison of the muon candidate's momentum and angles between GENIE Default and GENIE Alternative. Figure 15 to Figure 19 show the comparison of the leading proton candidate's momentum and angles between GENIE Default and Alternative. Generally speaking, the GENIE Alternative agrees with the data better than GENIE Default, and

with a lower χ^2. The deficit in the cosθ_μ distribution in the forward going region is also reduced in GENIE Alternative. But we emphasize that we have not yet performed a full systematic uncertainty estimate and these plots are only intended to give an impression of current agreement and potential sensitivity. Figure 17 shows the distribution of the number of proton candidates in GENIE Default and Alternative compared to data.

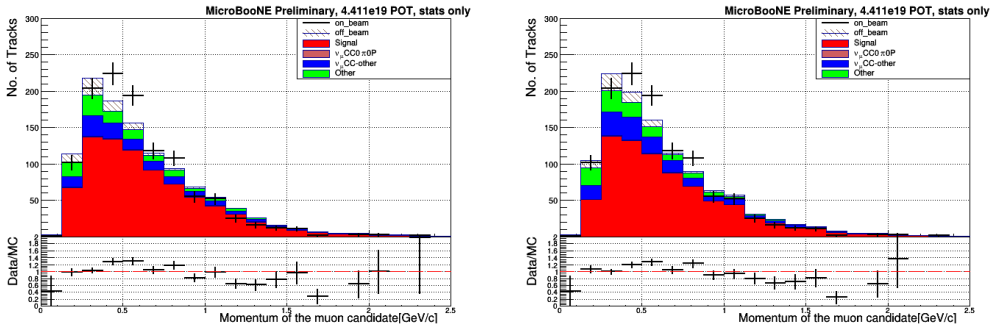

Figure 13: Distribution of the muon candidate momentum (GENIE Default:Left, $\chi^2/ndof$ = 48.9/20 vs GENIE Alternative: Right, $\chi^2/ndof = 39.8/20$, includes statics only).

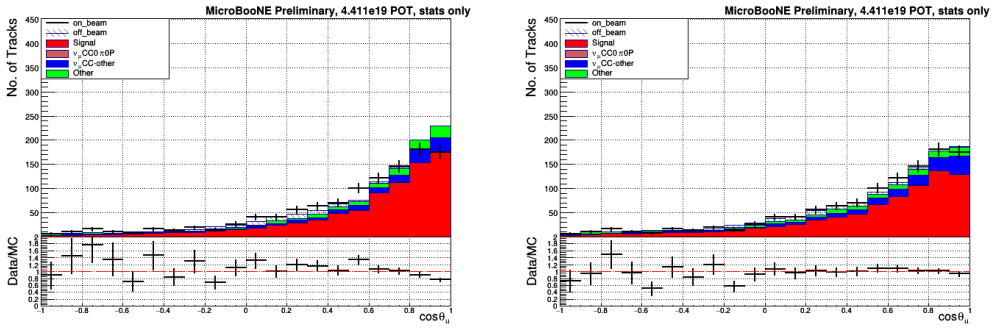

Figure 14: Distribution of cosθ of muon candidate (GENIE Default: Left, $\chi^2/ndof$ = 31.6/2 vs GENIE Alternative: Right, $\chi^2/ndof = 14.4/2$, includes statics only).

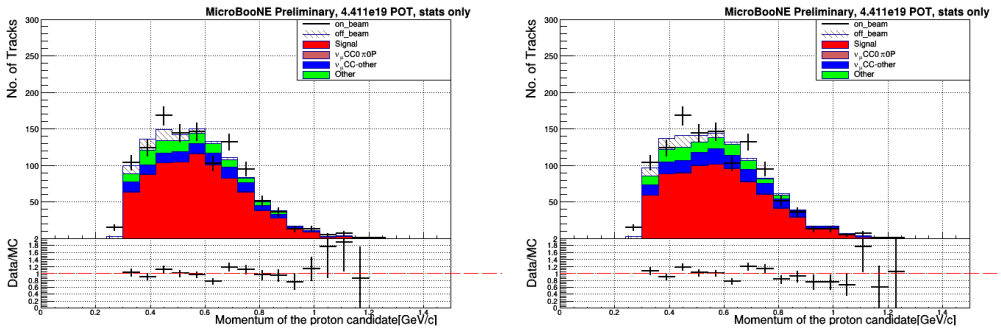

Figure 15: Distribution of leading proton candidate momentum (GENIE Default: Left, $\chi^2/ndof$ = 24.7/20 vs GENIE Alternative: Right, $\chi^2/ndof = 30.0/20$, includes statistics only).

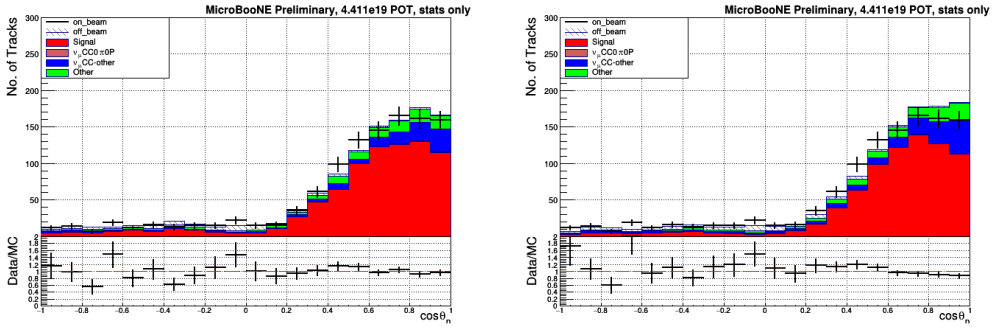

Figure 16: Distribution of cosθ of leading proton candidate (GENIE Default:Left, $\chi^2/ndof$ = 19.1/2 vs GENIE Alternative: Right, $\chi^2/ndof = 29.9/2$, includes statistics only).

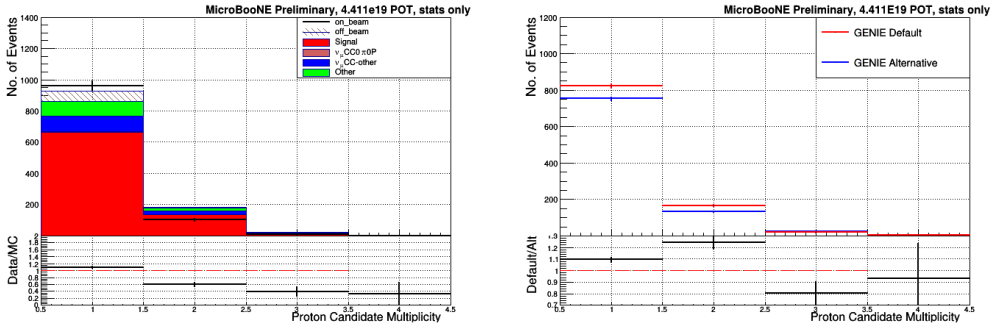

Figure 17: Area normalized distribution of proton multiplicity for the GENIE Default Model on the left, $\chi^2/ndof = 42.7/4$ includes statistical uncertainty only. Right plot show a comparison of GENIE Default and GENIE Alternative models unit normalized with statistical uncertainty only.

The neutrino interactions in the MicroBooNE detector should be independent of ϕ, the polar angle around the beam axis. As the detector is located on the surface without significant overburden, there is a ϕ dependence on the contribution of cosmic background for the CC1muNp selection. The plot in Figure 18 shows the ϕ distribution for muon candidates and Figure 19 show the ϕ distribution for proton candidates in the CC1muNp selection. It should be noted that the off-beam cosmic background (open, hashed histogram) appear at $\pi/2$ in the muon candidate distribution, and at $-\pi/2$ in the proton candidate distribution. This may be related to broken cosmic track entering in the single proton event selection, most probably a cosmic track oriented to the two oposite corners of the detector. While a small contribution to the sample, the back-to-back nature of the cosmic background gives a hint on how to further reduce the background. Notably, there is no variation between the GENIE Default and GENIE Alternative predictions. Note that the off-beam cosmic background in the CC1mu2p selection is small enough that this effect is not visible in the plots in Section 6.

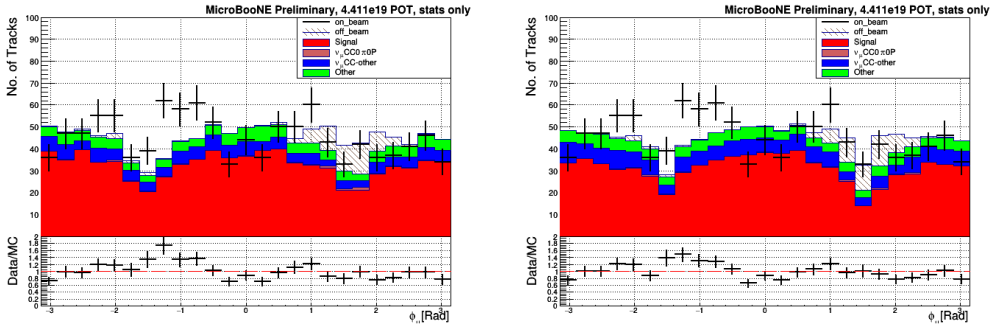

Figure 18: Distribution of ϕ of muon candidate (GENIE Default:Left, $\chi^2/ndof = 60.0/25$ vs GENIE Alternative:Right, $\chi^2/ndof = 47.3/25$, includes statistics only).

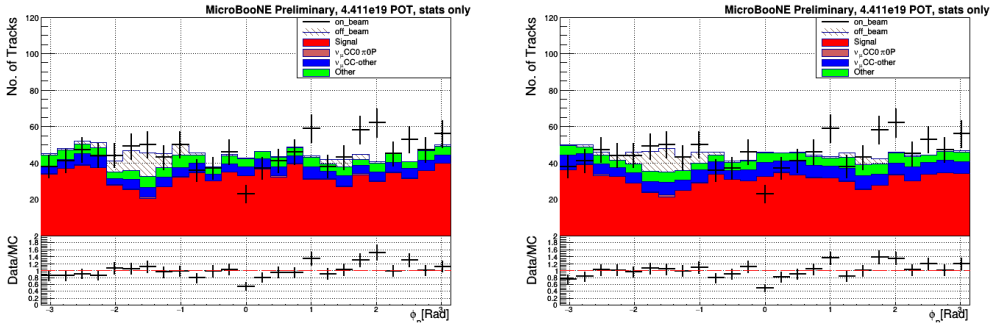

Figure 19: Distribution of ϕ of leading proton candidate (GENIE Default:Left, $\chi^2/ndof = 42.5/25$ vs GENIE Alternative: Right, $\chi^2/ndof = 41.7/25$, includes statistics only).

5.7 Study of tracking resolution

The next distributions show tracking resolution, with the measured value versus true value in MC. Figure 20 and Figure 21 shows the resolution of momentum and cosθ, respectively, of muon candidate and proton candidate. The momentum of exiting muons is calculated from MCS; while the momentum of the contained muons is calculated from the track range. Therefore, the resolution of the contained muon is better than the exiting muons. Resolution performance studies have been done using GENIE Default sample.

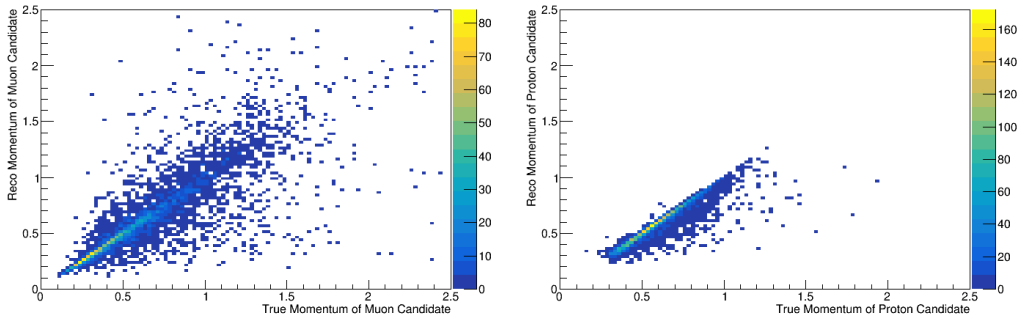

Figure 20: Reconstructed momentum vs true momentum distributions of muon candidate (left) and proton candidates (right).

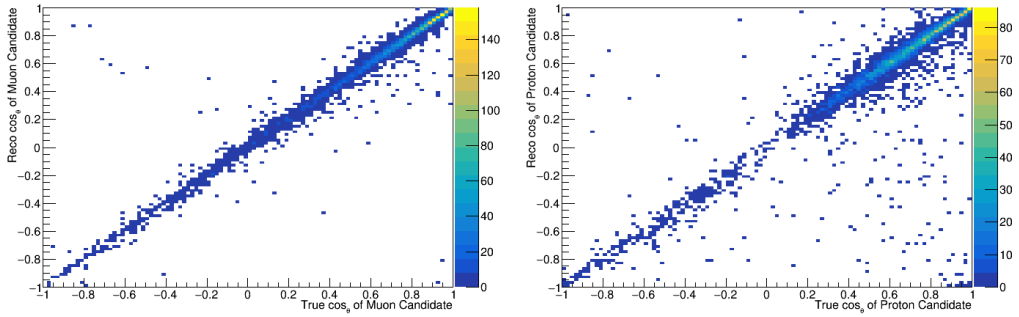

Figure 21: Reconstructed cosθ vs true cosθ distributions of muon candidate (left) and proton candidates (right).

6 The ν_μ charged current two proton events

To enhance the sensitivity to specific neutrino interaction models (e.g. MEC) in future cross sections measurements, an additional signal sample is defined by requiring exactly 2 protons above the momentum threshold in the event. The CC2Proton event selection uses the same particle identification cuts as the CCNProton selection but adds a requirement on only 2 proton candidates in the event. In the signal definition both protons are required to be above the momentum threshold of 300 MeV/c, this means that both protons are above the Fermi Momentum in Ar (250 MeV/c). The longest proton candidate is called 'proton 1' while the shorter proton candidate is called 'proton 2'. In this sample selection, events where the number of proton candidates is greater than 2 are classified as background.

Composition in the plots are categorized as:

- ν_μ CC0π0p: No pions and no protons in the event (events with all protons below the momentum threshold will enter in this category).
- ν_μ CC0π1p: No pions and one proton in the event (events with all protons below the threshold and one proton above threshold will enter in this category).
- ν_μ CC0π2p: No pions and two protons in the event (events with two protons above threshold will enter in this category). This is the signal for this analysis.
- ν_μ CC0πNp: No pions and N>2 protons in the event (events with N>2 protons above threshold will enter in this category).
- ν_μ CC1πNp: One pion (charged or neutral) and any number of protons.
- ν_μ CCNπNp: More than one pion (charged or neutral) and any number of protons.
- ν_e CC events.
- NC events.
- OOFV: neutrino events with vertex out-of-fiducial volume events.
- Cosmic: cosmic event according to MC. If any of the three tracks selected per event (muon, proton 1 or proton 2 candidate) belongs to a cosmic event, the event is cathegorized as cosmic.
- off-beam data: off-beam data that has been included as part of the background and normalized to those that should correspond to on–time cosmics.

The use of this simulated event classification has been done in order to isolate backgrounds in more detail and in particular to distinguish background due to non-detected particles as opposed to the mis-identification of final state particles. Table 12 shows purity and efficiency (with respect to CC2Protons) according to the final state proton multiplicity. We compare the selected sample composition between the GENIE Default and GENIE Alternative samples. Our main contamination comes from undetected protons at very low momenta and pions.

Topology	Composition (GENIE Default)	Composition (GENIE Alternative)
$CC0\pi 0Proton$	0.1%	0.3%
$CC0\pi 1Proton$	6.3%	6.9%
$CC0\pi 2Proton$	83.%	77.9%
$CC0\pi NProton$	0.%	0%
$CC1\pi NProton$	5.5%	9.3 %
$CCN\pi NProton$	1.%	0.65%
$CC\nu_e$	0.%	0%
NC	3.35%	4.3%
$OOFV$	0.2%	0.3%
cosmic	0.7%	0.4%
Efficiency $CC0\pi 2Proton$	14.7%	15.2%

Table 12: Composition and efficiency of the selected sample after applying all the CC2Proton cuts, with proton threshold at 300 MeV/c in momenta. Efficiencies are also shown. Using both GENIE Default and Alternative.

6.1 Kinematical distributions for the CC2Proton sample with a proton threshold momenta of 300 MeV/c

We show the kinematical distributions for muons and protons, from the CC2Proton selection. Figure 22 shows the momenta distributions for the selected muons and protons when comparing to GENIE Default. Figure 23 shows the same distributions when comparing to GENIE Alternative.

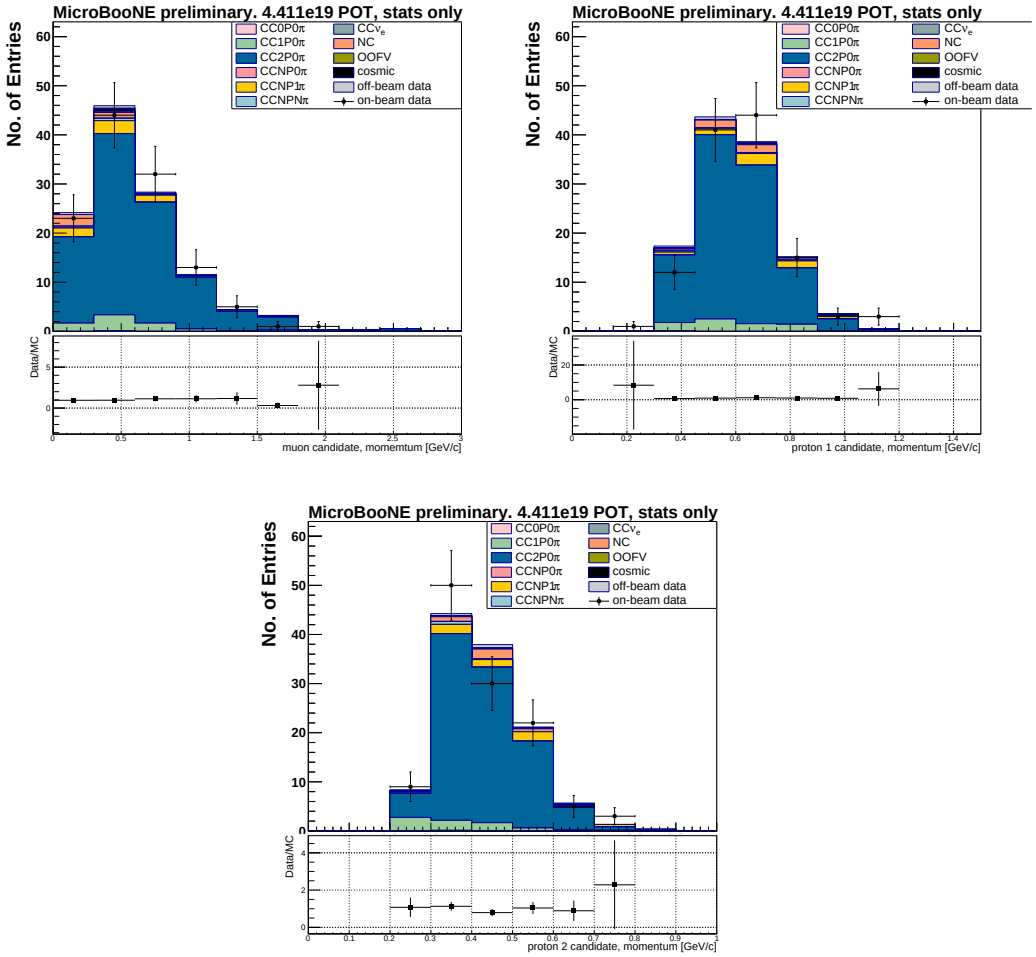

Figure 22: Momenta distributions for each candidate particle according to the final topology of the event. Left top for the muon ($\chi^2/ndof_{data/MC} = 3.8/7$, includes statistics only), right top for the longest proton ($\chi^2/ndof_{data/MC} = 22.7/7$, includes statistics only) and bottom for the shorter proton ($\chi^2/ndof_{data/MC} = 4.8/6$, includes statistics only). Using GENIE Default.

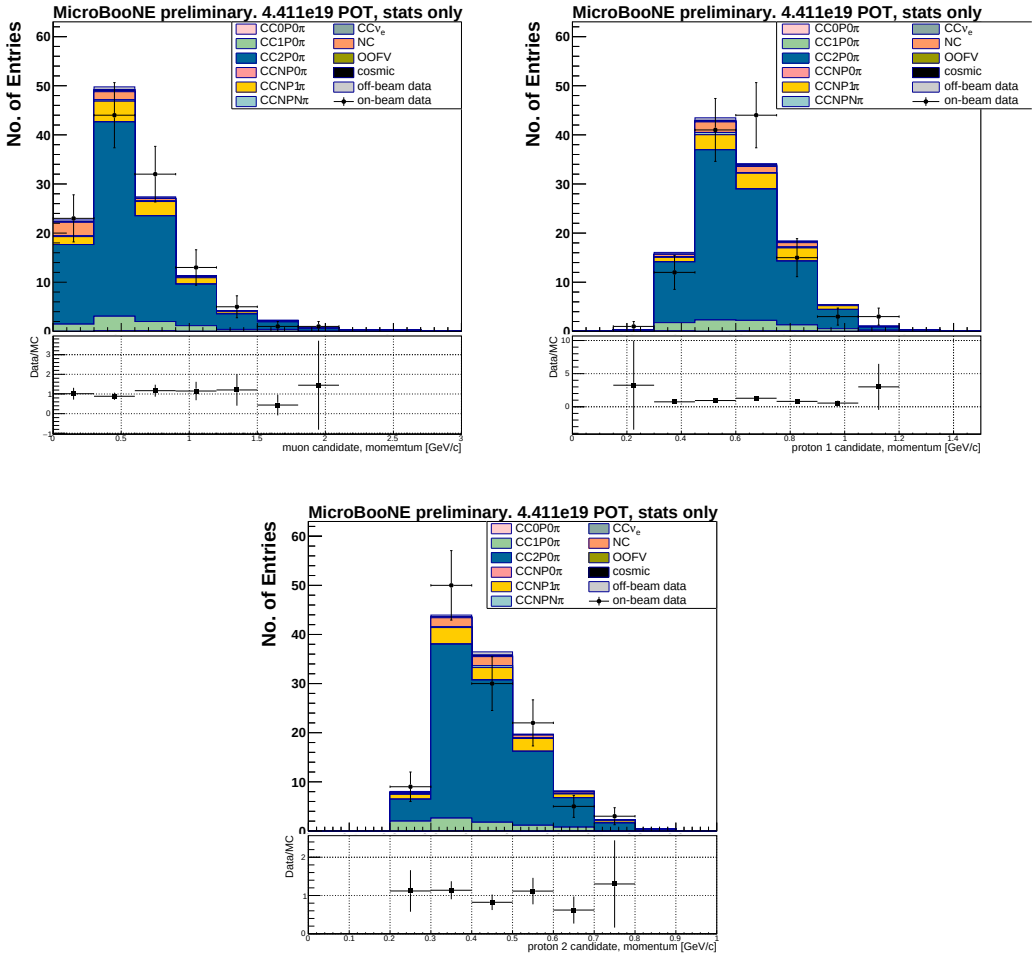

Figure 23: Momenta distributions for each candidate particle according to the final topology of the event. Left top for the muon ($\chi^2/ndof_{data/MC} = 2.9/7$, includes statistics only), right top for the longest proton ($\chi^2/ndof_{data/MC} = 11.4/7$, includes statistics only) and bottom for the shorter proton ($\chi^2/ndof_{data/MC} = 3.8/6$, includes statistics only). Using GENIE Alternative.

Figure 24 shows the *cosθ* distributions for each particle with respect to the beam direction. Left top for the muon, right top for the longest proton and bottom for the shorter proton, comparing to GENIE Default. Figure 25 shows the same distributions when comparing to GENIE Alternative.

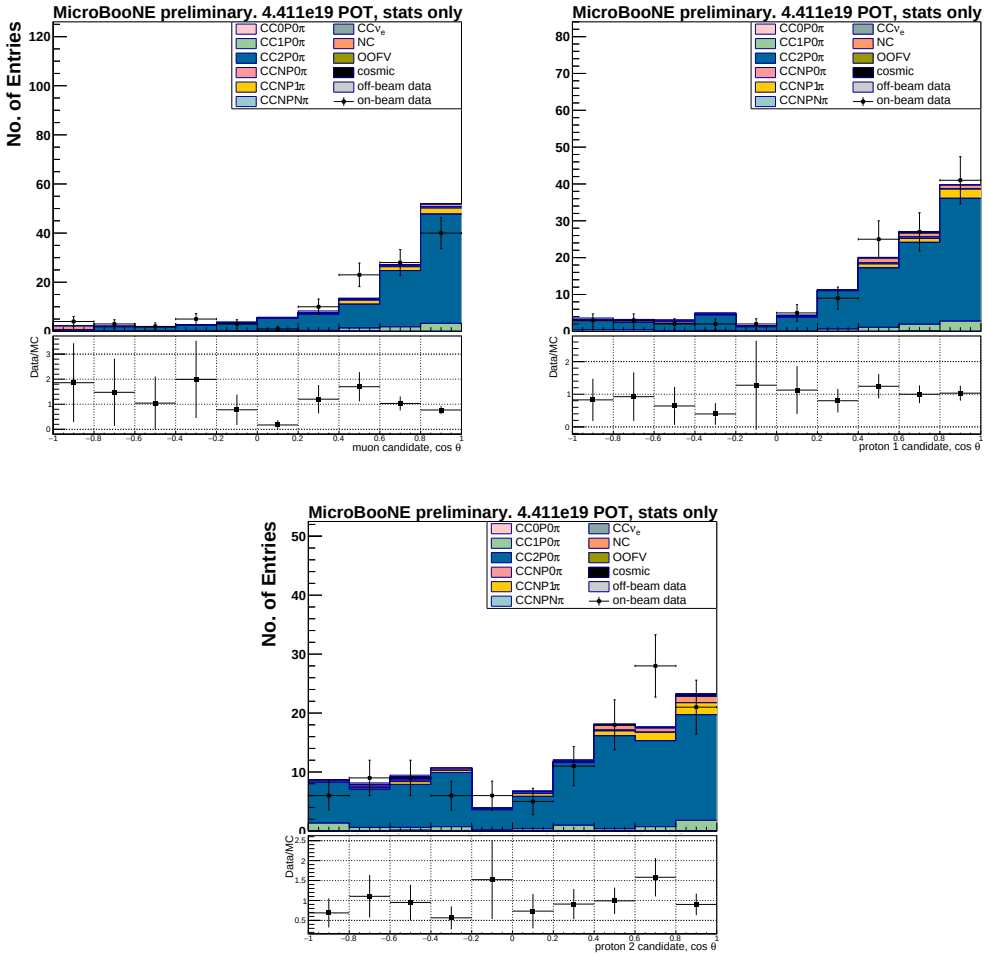

Figure 24: *cosθ* distributions for each candidate particle according to the final topology of the event. Left top for the muon ($\chi^2/ndof_{data/MC} = 15.2/9$, includes statistics only), right top for the longest proton ($\chi^2/ndof_{data/MC} = 4.5/9$, includes statistics only) and bottom for the shorter proton ($\chi^2/ndof_{data/MC} = 10.5/9$, includes statistics only). These are the distributions after applying all the cuts explained in this public note. Using GENIE Default.

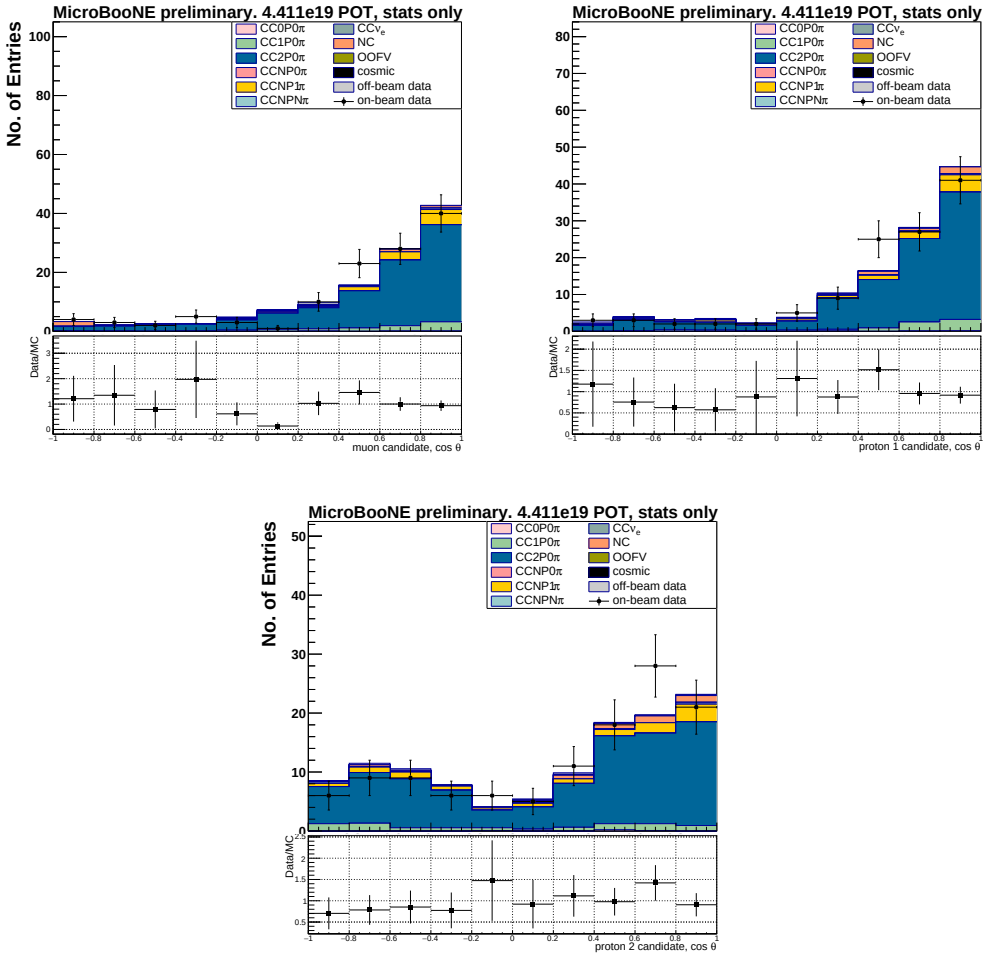

Figure 25: $cos\theta$ distributions for each candidate particle according to the final topology of the event. Left top for the muon ($\chi^2/ndof_{data/MC} = 12.5/9$, includes statistics only), right top for the longest proton ($\chi^2/ndof_{data/MC} = 6.4/9$, includes statistics only) and bottom for the shorter proton ($\chi^2/ndof_{data/MC} = 6.5/9$, includes statistics only). These are the distributions after applying all the cuts explained in this public note. Comparing to GENIE Alternative.

As an example with a reconstructed variable sensitive to changes in neutrino interaction model, we show the opening angle between the two protons in the sample. Figure 26 shows the $cos\theta_{p_1p_2}$ distributions, $cos\theta_{p_1p_2}$ the opening angle between the two protons. It is interesting to note that GENIE Alternative changes in shape with respect to GENIE Default and its agreement with the data seems to be a bit better. Still the ratio plots shows significant differences in data with both of the MC. The main difference in GENIE Alternative seems to be due to fewer predicted MEC events than the default, since MEC events are more likely to be generated with back-to-back nucleons we expect to see most of the discrepancies in the back-to-back region. For the limited statistics we present here and the need of systematic studies we need to be carefully before extract a physics conclusion.

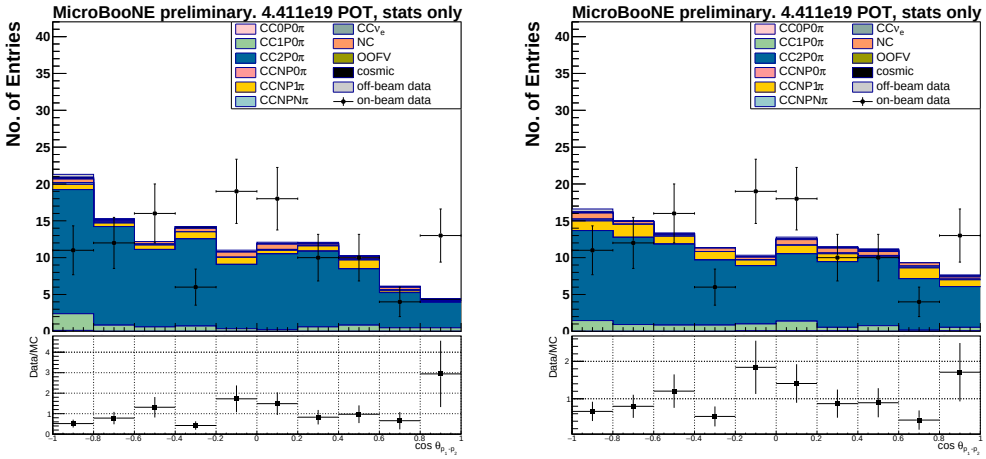

Figure 26: $cos\theta_{p_1p_2}$ distributions, being $\theta_{p_1p_2}$ the opening angle between the two protons. These are the distributions after applying all the cuts explained in this public note. Left for GENIE Default ($\chi^2/ndof_{data/MC} = 21.0/9$, includes statistics only), right for GENIE Alternative ($\chi^2/ndof_{data/MC} = 17.8/9$, includes statistics only).

6.2 Discussion on CC2Proton analysis

We observe that in general the shape comparisons of the different distributions are very similar in most of the cases to both GENIE choices. Both of the GENIE choices provide a reasonably good description of the data for some of the distributions, while in most of the cases this description is not accurate. In particular, for the proton opening angle in the lab frame, GENIE predicts more back-to-back events and the shapes are different than the data. However, as already mentioned, due to the presented low statistics and the need of systematic studies we prefer to extract physics conclusions when these studies are finished. The total number of events selected in data, on-beam and off-beam, is summarized in Table 13. To have an accurate measurement of the disagreement between data and predicted yield, full systematic studies will be performed in the future.

Sample	Events passing CC2Proton selection
on-beam data	119
off-beam data (normalized by number of triggers)	1.32

Table 13: Selected events for on-beam data and off-beam data passing the CC2Proton selection. off-beam data is normalized using number of triggers as explained within the normalization discussion. The total predicted yield is area normalized to the number of selected events in on-beam data.

Differences in data/MC are being investigated, in particular to disentangle detector effects. Possible reasons of disagreements at this stage are, taking into account detector simulation and reconstruction only:

- Particle multiplicity dependence.
- Other detector effects, such as DIC (dynamic induced charge), or more in general, E–field mapping.

One should notice that the CC2Proton selection has a very negligible cosmic contamination, this allow us to perform more accurate measurements on the different kinematic observables.

For completion, we show in Table 14 the neutrino interaction type in the selected CC2Protons events according to the two GENIE models.

Reaction type	GENIE Default	GENIE Alternative
CCQE	13.3%	23.4 %
Resonant	32.7%	47.3%
DIS	5.9%	6.8%
Coherent	0.1%	0%
Meson Ex-change Current	45.4%	19.6%

Table 14: Topology compositions for selected events in MC GENIE Default and Alternative passing the CC2Proton selection.

There is a clear difference in MEC production according to the different GENIE model choices. While in Default, MEC is dominant, it is not the case for the Alternative mode, in which resonant pion production is the dominant (mostly going into pion absorption within the nucleus).

7 Conclusions

We have developed a fully automated selection of ν induced protons which has been applied to ν_μ induced proton events in MicroBooNE. Studies of particle identification and anomalies in the data and MC agreement have been shown and discussed, while we claim that at this stage we are presenting limited statistics with respect to the available data (see [8]) and systematics studies will follow . We achieve a purity for protons using the selected PID of 92.5% and an efficiency of the method of 85.2%. The ν_μ CCNProton selection uses the ν_μ CC inclusive pre–selection before applying additional requirements for identifying final state protons creating a high-purity sample of CCNProton events and establishing a proton threshold of $300 MeV/c$ in the signal. This is the lowest threshold achieved in LArTPC using a fully automated reconstruction.

The proton identification method uses the calorimetry information from the MicroBooNE detector, demonstrating the power of the LArTPC technology to perform PID efficiently even for low energy particles (momenta $<$ 400 MeV/c). We achieve lower proton energy thresholds than other neutrino experiments as T2K and MINERvA but as well a higher angular acceptance (4π) which will allow to better sensitivity to the different neutrino scattering models.

The calculated efficiency of 29% and purity of 77%, on the CCNProton selection, are adequate for this study and will be improved before continuing on to cross section measurements. Some of the kinematic distributions demonstrate good agreement between our data and the simulation. To understand all kinematic discrepancies, we need to further quantify detector effects and disentangle then from neutrino scattering simulation.

The CC2Proton sample is a sub-sample of the CCNProton selection requiring a two proton–only selection. Performance under the different kinematic observables have been presented with respect to two different GENIE model choices. We are able to select 119 on-beam data events, with an off-beam cosmic contamination of 1.32 events. This corresponds to the selected data events in the $4.4 \cdot 10^{19}$ POT sample used, while we expect to select $\sim 430 events in the full open data sample$(1.6 $\cdot 10^{20}$ POT).

In this public note, we have demonstrated that MicroBooNE can select protons coming from neutrino interactions with high purity. We are also able to make measurements on the total number of protons in the final state in the MicroBooNE experiment. While the current measured efficiencies will be improved with upcoming reconstruction improvements, this analysis makes our current benchmark in proton identification in LArTPC with automatic reconstruction.

8 On-beam data Event Displays

Here are presented some of the selected on-beam data events within the CC0π2P selection scheme and last event shown corresponds to the selected four proton candidates event. Figures 27 and 28 are CC2Proton candidates, while Figure 29 shows an event that passes selection but upon visual inspection is suspected to be background. Figure 30 shows an event passing the CCNProton selection and containing 4 proton-candidate tracks. All figures corresponds to on-beam data. All figures show collection plane only.

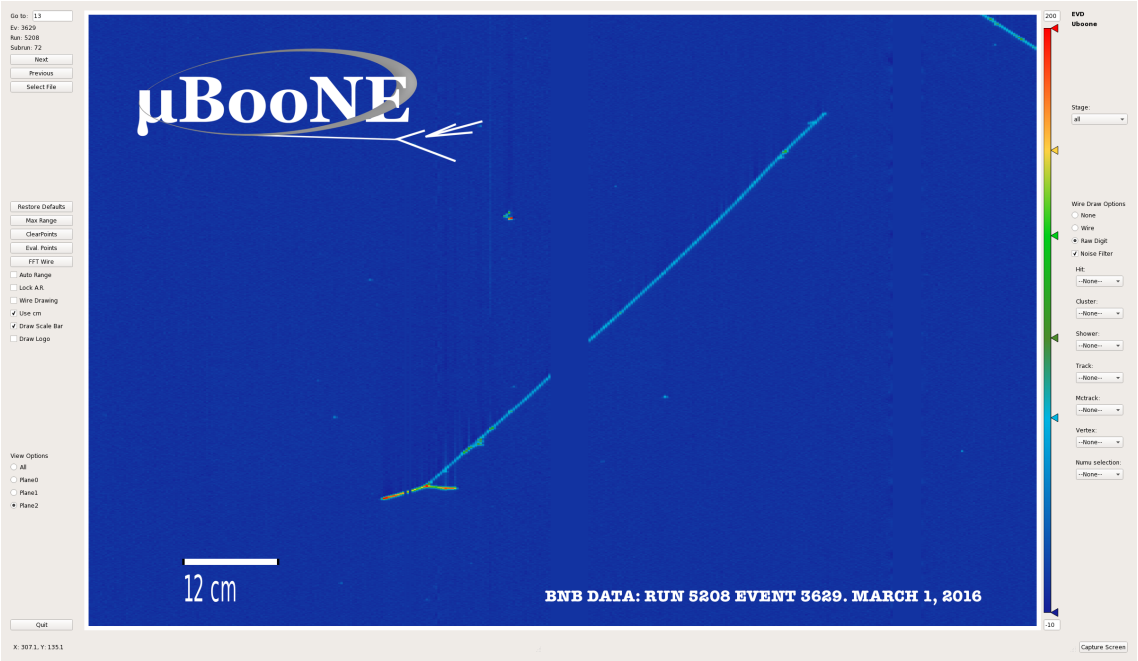

Figure 27: on-beam BNB data, ν_μ CC with 2 proton candidates in the final state.

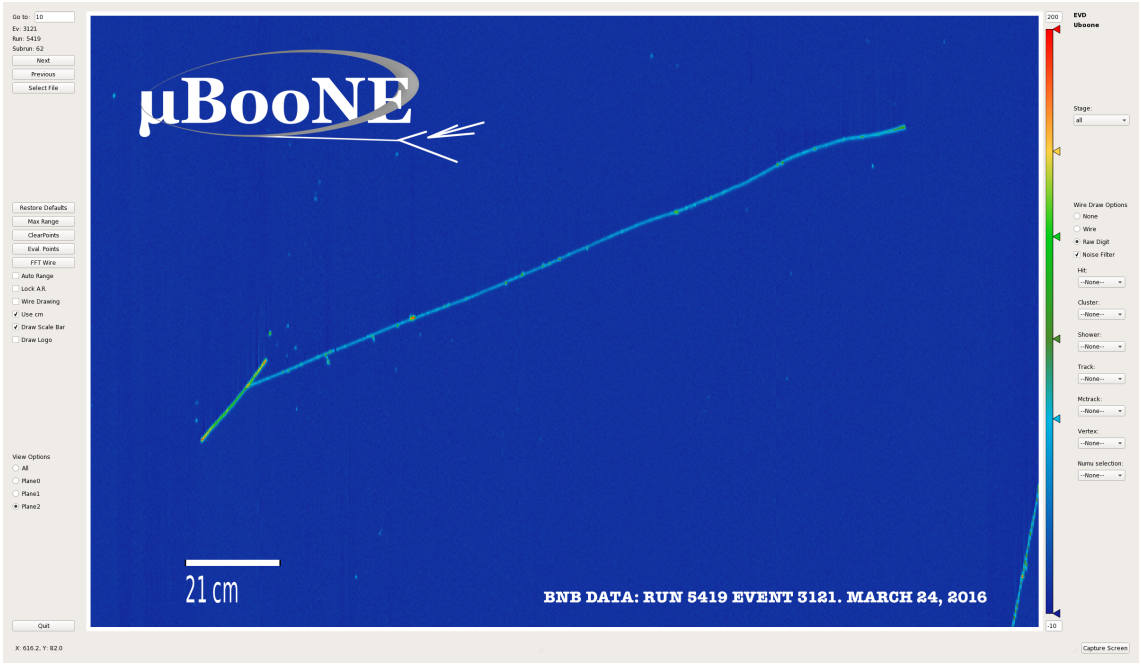

Figure 28: On-beam BNB data, ν_μ CC with 2 proton candidates in the final state.

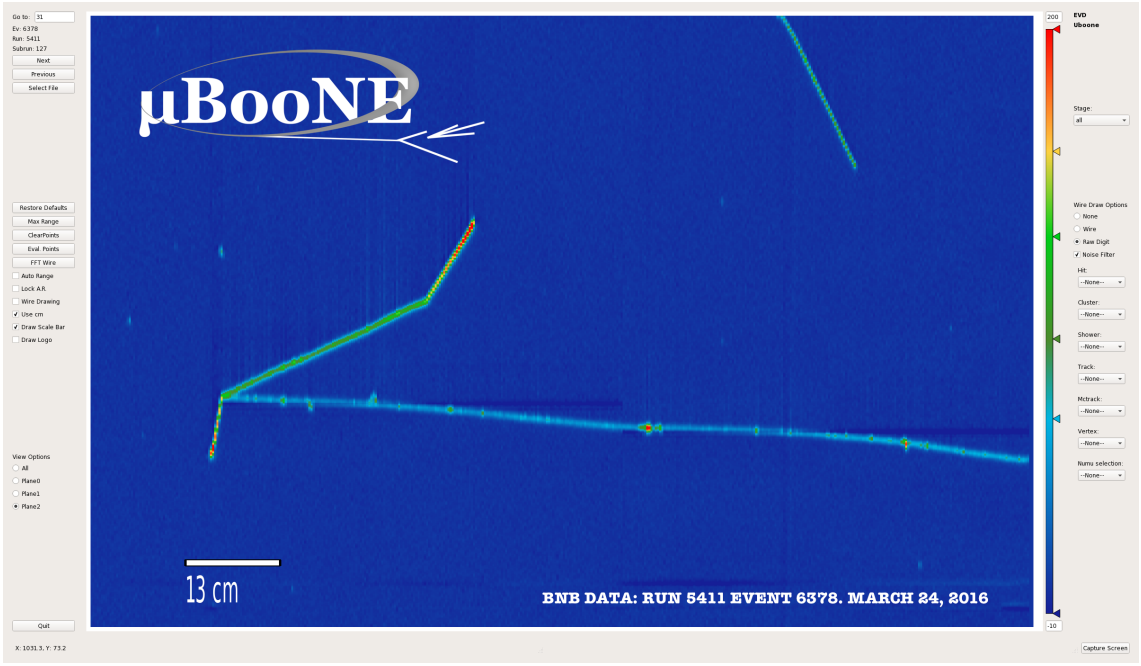

Figure 29: On-beam BNB data, event passing the CC2Proton selection.

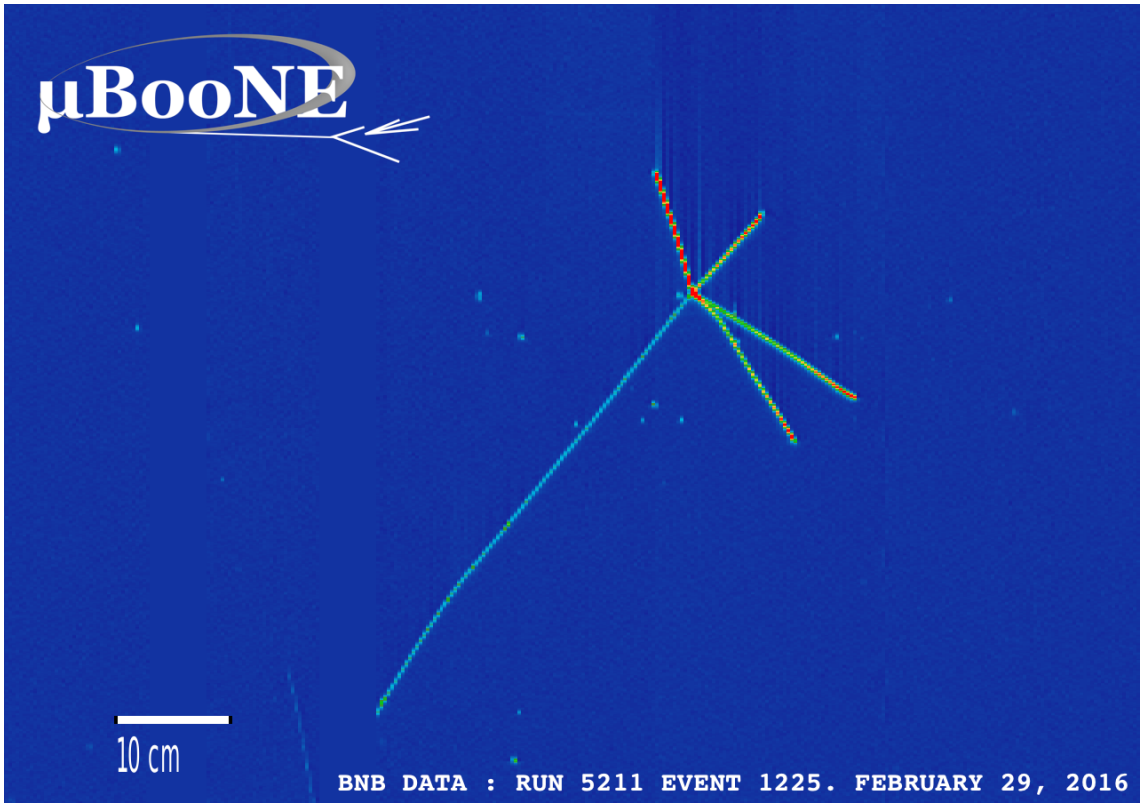

Figure 30: On-beam BNB data event passing the CCNProton selection with the longest track from the vertex being a muon track and 4 proton candidate tracks. This is the highest proton-multiplicity event in the current data sample.

References

[1] ArgoNeuT Collaboration. The detection of back-to-back proton pairs in charged-current neutrino interactions with the argoneut detector in the numi low energy beam line. *arXiv:1405.4261v1.*

[2] ArgoNeuT Collaboration. First measurement of the cross section for ν_μ and $\bar{\nu_\mu}$ induced single charged pion production on argon using argoneut. *PHYSICAL REVIEW D 98, 052002 (2018).*

[3] The ArgoNEUT Collaboration. *arXiv:1306.1712.*

[4] The ICARUS Collaboration. Study of electron recombination in liquid argon with the icarus tpc. *Nuclear Instruments and Methods in Physics Research A 523 (2004) 275–286.*

[5] The MicroBooNE Collaboration. *arXiv:1708.03135, Eur. Phys. J. C78, 1, 82 (2018).*

[6] The MicroBooNE Collaboration. Detector calibration using through going and stopping muons in the microboone lartpc. *MICROBOONE-NOTE-1048-PUB.*

[7] The MicroBooNE Collaboration. Determination of muon momentum in the microboone lar tpc using an improved model of multiple coulomb scattering. *JINST 12 P10010 (2017).*

[8] The MicroBooNE Collaboration. First muon-neutrino charged-current inclusive differential cross section measurement for microboone run 1 data. *MICROBOONE-NOTE-1045-PUB.*

[9] The MINERvA Collaboration. *arXiv:1305.2243v4.*

[10] The MiniBooNE collaboration. *arXiv:1002:2680 [hep-ex], Phys. Rev. D81, 092005 (2010).*

[11] The T2K Collaboration. *arXiv:1602.03652v2 [hep-ex].*

[12] R. Gran et al. Neutrino-nucleus quasi-elastic and 2p2h interactions up to 10 gev. *Phys. Rev. D.88 (2018), p. 113007.*

[13] The Wroclaw Neutrino Group. Nuwro - wrocław neutrino events generator.

[14] I. Ruiz Simo J.Nieves and M. J. Vicente Vacas. *Phys. Rev. C*, 83, 2011.

[15] I. Ruuiz Simo J.Nieves and M. J. Vicente Vacas. *Phys. Rev. C*, 83, 2004.

[16] T. Katori. Meson exchange current (mec) models in neutrino interaction generators. *AIP Conf. Proc. 1663 (2015), p. 030001.*

[17] K. Niewczas and J. Sobczyk. Search for nucleon-nucleon correlations in neutrino-argon scattering. *arXiv:1511.02502v2 [hex-exp] (2015).*