Tests of neutrino interaction models with the MicroBooNE detector

by

Aleena Rafique

B.Sc., Punjab University, 2009

M.Sc., Quaid-i-Azam University, 2011

M.Phil., Quaid-i-Azam University, 2013

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2018

Abstract

I measure a large set of observables in inclusive charged current muon neutrino scattering on argon with the MicroBooNE liquid argon time projection chamber operating at Fermilab. I evaluate three neutrino interaction models based on the widely used GENIE event generator using these observables. The measurement uses a data set consisting of neutrino interactions with a final state muon candidate fully contained within the MicroBooNE detector. These data were collected in 2016 with the Fermilab Booster Neutrino Beam, which has an average neutrino energy of 800 MeV, using an exposure corresponding to 5.0×10^{19} protons-on-target. The analysis employs fully automatic event selection and charged particle track reconstruction and uses a data-driven technique to separate neutrino interactions from cosmic ray background events. I find that GENIE models consistently describe the shapes of a large number of kinematic distributions for fixed observed multiplicity, but I show an indication that the observed multiplicity fractions deviate from GENIE expectations.

Tests of neutrino interaction models with the MicroBooNE detector

by

Aleena Rafique

B.Sc., Punjab University, 2009

M.Sc., Quaid-i-Azam University, 2011

M.Phil., Quaid-i-Azam University, 2013

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2018

Approved by:

Major Professor
Dr. Tim Bolton

Copyright

Abstract

I measure a large set of observables in inclusive charged current muon neutrino scattering on argon with the MicroBooNE liquid argon time projection chamber operating at Fermilab. I evaluate three neutrino interaction models based on the widely used GENIE event generator using these observables. The measurement uses a data set consisting of neutrino interactions with a final state muon candidate fully contained within the MicroBooNE detector. These data were collected in 2016 with the Fermilab Booster Neutrino Beam, which has an average neutrino energy of 800 MeV, using an exposure corresponding to 5.0×10^{19} protons-on-target. The analysis employs fully automatic event selection and charged particle track reconstruction and uses a data-driven technique to separate neutrino interactions from cosmic ray background events. I find that GENIE models consistently describe the shapes of a large number of kinematic distributions for fixed observed multiplicity, but I show an indication that the observed multiplicity fractions deviate from GENIE expectations.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I take this opportunity with pride and enormous gratification to express the feelings of gratefulness from the bottom of my heart to all the persons who backed me directly or indirectly throughout the materialization of this research work at this magnitude, without which this thesis might not have been written, and to whom I am greatly indebted.

First and the foremost, I deem it a rare opportunity to pronounce my deep sense of gratitude to my Ph.D. advisor, Dr. Tim Bolton. I am very thankful to him for his brilliant ideas, valuable and constructive guidance, and assistance throughout the planning, development, and accomplishment of this research work. His willingness to give his time so generously has been very much appreciated. I always could rely on his remarkable ability to sense and distill any problem so quickly. His kind behavior, encouraging nature, and inspiring attitude made it very easy for me to understand precisely and perform accurately this work throughout the period of research.

I can't thank my parents enough for their 24/7 availability for me despite being 12,000 km away. I am heartily thankful to you for the myriad of ways in which, throughout my life, you have actively supported me in my determination to find and realize my potential. A very special thank you for providing me moral support whenever I needed and showing your care and support in incredible ways.

I greatly acknowledge many scientists, professors, and postdocs from MicroBooNE experiment, who continuously provided insightful comments, encouragement, and support throughout my research work. Thanks to Sam Zeller, Bonnie Fleming, Matt Toups, Michele Weber, Andy Furmanski, Joel Mousseau, Wes Ketchum, Mike Mooney, Joseph Zennamo, Tingjun Yang, Georgia Karagiorgi, Steve Dytman, Anne Schukraft, Glenn Horton-Smith, and Sowjanya Gollapinni

for providing guidance on my research work.

Thanks to my friends at Fermilab, Raquel, Libo, Tia, Katherine, Lu, Athula, Simone, and David, who made a friendly environment for working at Fermilab.

Loving thanks to my friends from KSU, Elahe, Rachitha, Varuna, Peyman, Sajed, Farzaneh, Alex, Sawyer, and Huynh, who played such important roles along the journey, as we mutually engaged in making sense of the various challenges we faced and in providing encouragement to each other at those times when it seemed impossible to continue.

Despite these acknowledgements, the weaknesses and limitations of this thesis remain all mine.

Chapter 1

Introduction

A growing number of neutrino physics experiments use liquid argon as a neutrino interaction target nucleus in a time projection chamber [1]. Experiments that use or will use liquid argon time projection chamber (LArTPC) technology include those in the Short-Baseline Neutrino (SBN) program [2] at Fermilab, centered on searches for non-standard neutrino oscillations, and the long-baseline DUNE experiment [3]. The SBN program consists of the MicroBooNE experiment [4], an upgraded ICARUS experiment [5], and the new SBND experiment [6]. The DUNE experiment seeks to establish the mass ordering of the three standard model neutrinos and the charge parity violation parameter phase δ_{CP} in the PMNS neutrino mixing matrix [7].

All LArTPC neutrino oscillation-related measurements require a precise understanding of neutrino scattering physics and the measured response of the LArTPC detector to final state particles. These depend on: (a) the neutrino flux seen by the experiment, (b) the neutrino scattering cross sections, (c) the interaction physics of scattering final state particles with argon, (d) transport and instrumentation effects of charge and light in the LArTPCs, and (e) software reconstruction algorithms. In practice item (a) is determined by a combination of hadron production cross section measurements and precise descriptions of neutrino beamline components. Item (b) is most commonly provided by the GENIE [8] neutrino event generation model for neutrino-argon scattering. Items (c)-(e) are incorporated into a detailed suite of GEANT-based [9] simulation and event re-

construction products called LArSoft [10].

GENIE has been built up from models of the most important physical scattering neutrino-nucleon mechanisms for the SBN and DUNE energy regimes ($0.5-5$ GeV) : quasi-elastic (QE) scattering $\nu_\ell N \to \ell^- N', \nu_\ell N'$, resonance production (RES) $\nu_\ell N \to \ell^- R, \nu_\ell R'$, and non-resonant multi-hadron production referred to as deep inelastic scattering (DIS): $\nu_\ell N \to \ell^- X, \nu_\ell X'$ [11]. The underlying neutrino-nucleon scattering processes receive significant modification from the nuclear environment, including the effects of Fermi motion of target nucleons, many-nucleon effects, and final state interactions (FSI) [12]. While GENIE has received a fair amount of "tuning" (the process of finding a set of GENIE parameters chosen to optimize agreement with a particular data set) from previous electron and neutrino scattering measurements, considerable uncertainties remain in the modeling of both the underlying neutrino-nucleon scattering and the nuclear environment effects [13].

Relatively few neutrino scattering measurements on argon exist [14–19], especially for the recoil hadronic system. Most of these report low-statistics exclusive final states. Nearly all existing neutrino scattering constraints on GENIE models derive measurements on scattering from carbon, which has 30% of argon's atomic mass number and a 22% lower neutron-to-proton ratio. I take a step in improving the empirical understanding of neutrino scattering from argon here by performing a large set of comparisons of observed inclusive properties of charged current scattering, measured at the MicroBooNE experiment in the Fermilab Booster Neutrino Beam (BNB) [20], to predictions from several variants of GENIE. These comparisons are generated by applying fully automated event reconstruction and signal selection tools to a subset of MicroBooNE's first collected data. While this analysis must focus in large part on reducing cosmic ray backgrounds, sensitivity to GENIE model parameters remains.

In chapter 2, I introduce the physics of neutrinos. In chapter 3, I describe the Fermilab BNB. Chapter 4 describes MicroBooNE detector. Chapter 5 summarizes the event reconstruction procedure in MicroBooNE. Chapter 6 describes the neutrino interaction models. Chapter 7 presents the

comparison of observed charged particle multiplicity distributions, and of charged track kinematic distributions for each multiplicity, to predictions from GENIE, and chapter 8 gives an overall conclusion.

This work was performed while I was a member of MicroBooNE collaboration. I used many tools produced by others in the collaboration and by me. During the course of this work, I produced several technical notes, public notes, and a physics paper. The work described in this thesis is largely derived from [21] which I authored.

Chapter 2

Physics of Neutrinos

2.1 Introduction to Neutrinos

Neutrinos are fundamental and abundant particles of the universe. These properties make neutrinos very interesting to study. Despite their abundance, we know very little about neutrinos. This is due to the fact that neutrinos have very high penetration power. Neutrinos coming from sun mostly penetrate whole earth without making interaction with anything along its way. This makes the detection of neutrinos very difficult. Hence it is a challenge to catch and study these rarely interacting particles.

The standard model neutrinos lie in the category of "leptons". Neutrinos interact only via the weak subatomic force and gravity, and have spin "$1/2$". There are three types (or "flavors") of neutrinos in the standard model which consist of electron, muon, and tau neutrinos. Neutrino types are differentiated by their interaction properties.

2.2 Neutrino Prediction and Discovery

In 1920's, the process that led into prediction of neutrinos was the beta decay which is a radioactive process in which a nucleus of atomic number Z changes to one having atomic number Z+1, and an

electron is emitted.

$$X_Z \rightarrow X_{Z+1} + e^- \tag{2.1}$$

In order to conserve energy in this process, the outgoing electron must have a fixed value of kinetic energy. Contrary to this requirement, it was observed that the electron rather had a continuous energy spectrum [22]. At first, it was assumed that the energy conservation is violated in beta decay process but later on, in 1930, Wolfgang Pauli proposed the presence of another outgoing neutral particle in the beta decay [23], which he named "neutrino", that carries the rest of energy such that the total outgoing energy remains same as incoming total energy of the process.

Neutrinos were discovered by Clyde Cowan and Frederick Reines in 1956 [24]. They used a nuclear reactor located at Savannah River Plant in South Carolina as the neutrino source for their experiment. The experiment used two tanks of $CdCl_2$ dissolved in water. These tanks were sandwiched between three layers of scintillator which contained 110 photomultiplier tubes. This experiment was designed to observe inverse beta decay process in which an anti-neutrino interacts with a proton producing a neutrino and a positron.

$$\bar{\nu} + p \rightarrow n + e^+ \tag{2.2}$$

The positron quickly annihilates with an electron producing two gamma rays. The scintillator material gives off flashes of light in response to the gamma rays, and these light flashes are detected by photomultiplier tubes. The thermal neutrons having large cross sections, were detected by process "neutron capture" [25] by Cd nucleus. The coincidence of two events, electron-positron annihilation and neutron capture, was the signature of the anti-neutrino interaction. Frederick Reines was awarded Nobel Prize in 1995 for this discovery.

2.3 Neutrino Sources

Two main categories involve natural and artificial sources of neutrinos. Natural sources include atmospheric, solar, and supernovae neutrinos. Artificial or man-made sources include reactor and accelerator neutrinos.

Atmospheric neutrinos are the neutrinos that are produced by the interaction of cosmic rays in the earth atmosphere. These cosmic rays have a broad mass spectrum and are dominated by protons. When these protons interact with air molecules in the earth atmosphere, they produce mesons which further decay into leptons (antileptons) and corresponding neutrinos (antineutrinos). The energy spectrum for these neutrinos are mostly in GeVs. The earth sees symmetric flux of cosmic rays in all directions.

Solar neutrinos are generated as a result of the nuclear fusion process that is responsible for the tremendous energy radiated by sun. The neutrinos emitted have relatively low energy that are created by fusing hydrogen (having one proton) to helium (having two protons and two neutrons). This process is possible only with the conversion of protons to neutrons. Mostly electron neutrinos are emitted as a result of fusion processes that are happening inside the sun's core.

Supernovae neutrinos are created by the explosive deaths of stars. When a massive star collapses into a neutron star [26] near the end of its life, its binding energy gets radiated out in the form of neutrinos. These neutrinos can have any flavor and their emission timescale is several tens of seconds.

Nuclear rectors are the excellent source of electron neutrinos and antineutrinos from β decay. In fission reaction, energy is generated by breaking a heavy nucleus into light nuclei. In heavy nuclei, neutron to proton ratio is large. These unstable neutrons decay via beta decay producing neutrinos. The reactor neutrinos generally have energies in few MeV. This small amount of energy limits our scope of experiments so reactor neutrino source is not an ideal source for doing sophis-

ticated neutrino experiments.

The accelerator neutrinos work on the principle of accelerating a proton beam up to certain energy before they hit a target after which they produce mesons. These mesons are then allowed to decay into lepton and neutrinos. The neutrino energies from accelerator sources are usually in GeVs. This thesis describes an experiment which uses accelerator neutrinos to produce the neutrino beam.

Figure 2.1 shows a plot of neutrino energies versus cross-section (interaction probability) from different sources.

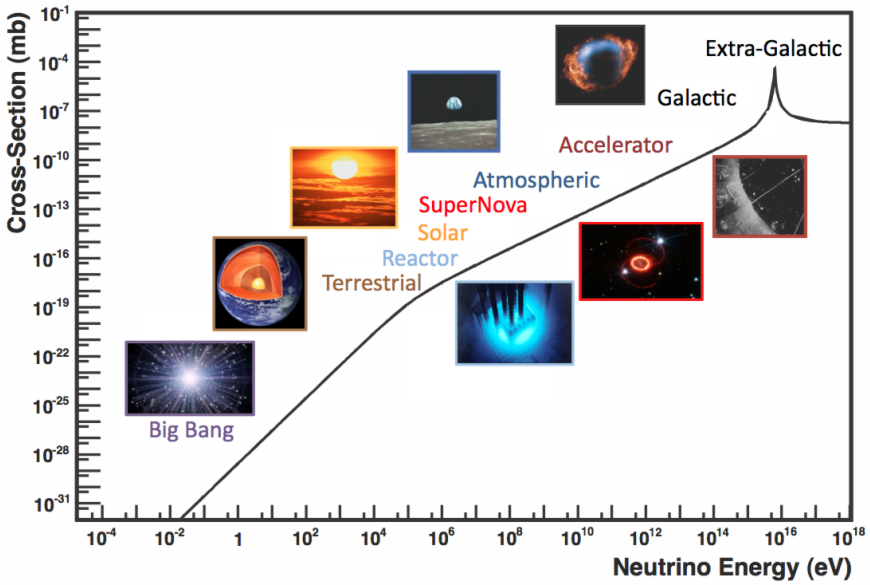


Figure 2.1: Neutrino energies versus cross section for various neutrino sources. Figure from Ref. [27].

2.4 Neutrino Oscillations

The phenomena of neutrino changing its flavor to a different flavor after traveling some distance is called neutrino oscillation. Neutrino oscillations were first observed in 1968 in Ray Davis's Homestake experiment [28]. In this experiment, solar neutrinos, which are mostly electron neutri-

nos, were being studied and a deficit in their flux with respect to standard model prediction was observed. That deficit was not fully understood until 2001. The presence of neutrino oscillations was confirmed in 2001 by combining oscillation results obtained from the Super-Kamiokande experiment [29] that used atmospheric neutrinos and at the same time from Sudbury Neutrino Observatory (SNO) experiment [30] that used solar neutrinos.

Neutrino oscillations lead towards non-zero neutrino mass. The phenomenon in which a neutrino with one flavor eigenstate, after propagating certain distance L, changes to another flavor eigenstate is called neutrino oscillation. As the flavor eigenstates can be written as a superposition of the mass eigenstates, this phenomenon is possible only if mass eigenstates are non-zero. The three mass eigenstates are represented as ν_1, ν_2, and ν_3 with masses m_1, m_2, and m_3 respectively. The experiments can only measure the mass-squared splitting ($m_{ij} = m_i^2 - m_j^2$) between the masses and hence the absolute masses of the mass eigenstates are unknown. This introduces the two possible ordering configurations of the masses called "normal mass hierarchy" and "inverted mass hierarchy", as described in detail in Sec. 2.5.2.

A neutrino flavor eigenstate can be written as a quantum linear superposition of the neutrino mass eigenstates by way of the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) 3×3 unitary mixing matrix [31], $U^*_{\alpha i}$.

$$|\nu_\alpha >= \sum_i U^*_{\alpha i} |\nu_i > \tag{2.3}$$

where ν_α is a flavor eigenstate with $\alpha = \mu, e, \tau$ and ν_i is a mass eigenstate with $i = 1, 2, 3$.

Generally, three neutrino mixing matrix can be written as follows:

$$U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{-i\delta} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.4}$$

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$ with θ_{ij} being the mixing angles and δ_{ij} are the CP violating

phases (described in Sec. 2.5.3).

In general, there are two basic choices of searches in the neutrino oscillation experiments: appearance or disappearance search. In appearance experiments, the type of neutrinos at the production point are known and then another neutrino type is observed after traveling a distance L. In disappearance experiments, the number of a certain type of neutrinos is known at the production point and then a deficit in the number of neutrinos of the same type is observed after traveling a distance L.

The two neutrino oscillation case represents the approximate solution in many experiments. The mixing matrix for two neutrino oscillation case can be written as rotation matrix form

$$U = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \tag{2.5}$$

where θ being the effective mixing angle [32].

A neutrino with initial flavor α propagates as a superposition of two mass eigenstates ν_1 and ν_2 as

$$|\nu_\alpha(t)\rangle = \cos\theta e^{-iE_1 t}|\nu_1\rangle + \sin\theta e^{-iE_2 t}|\nu_2\rangle \tag{2.6}$$

In relativistic approximation

$$E_i = p_i + \frac{m_i^2}{2p_i} \tag{2.7}$$

the oscillation probability for a neutrino of flavor α to be detected having flavor β can be written as follows,

$$P_{\alpha\to\beta} = |\langle\nu_\alpha|\nu_\beta\rangle = \sin^2 2\theta \sin^2\left(1.27\frac{\Delta m^2 L}{E}\right) \tag{2.8}$$

where L is the baseline distance (distance from neutrino production to detection point), E is the average neutrino energy, and θ is the mixing angle.

2.5 Open Questions in Neutrino Physics

With the neutrino oscillation discovery, we have started to understand complex properties of neutrinos. Since neutrino detection is very difficult, we believe that there is a lot more yet to be understood about them. Some of the major unsolved problems in neutrino physics are the following:

- How many neutrino species exist? Are there sterile neutrinos?
- What is the scale of neutrino mass and mass hierarchy?
- Is there a CP violation in neutrino sector?
- Are neutrinos Dirac or Majorana particles?

I will explain, in detail, some of these questions and the experiments that are being designed to address these questions in the following subsections.

2.5.1 Sterile Neutrino Hint and Search

One of the important question in neutrino physics is to look for the existence of a neutrino beyond the three known standard model neutrinos. Some hints have been observed for the existence of such neutrino from experiments including Liquid Scintillator Neutrino Detector (LSND) [33] and MiniBooNE [34] experiments.

In 2001, using a beam of $\overline{\nu}_\mu$ at Los Alamos Neutron Science Center, LSND experiment saw an excess of $\overline{\nu}_e$ events above the expected backgrounds (see Figure 2.2) [35]. Considering the LSND baseline ($L = 30$ m) and average neutrino energy ($E = 40$ MeV), the experimental value for $\Delta m^2 \approx 1eV^2$ disagreed with the previous neutrino oscillation measurements where Δm^2_{12} and Δm^2_{23} were an order of magnitude less than $1eV^2$. Since the Z-boson decay width constrains the number of weakly interacting neutrino states to three, one possible explanation for this different Δm^2 value is the possible existence of potential additional "sterile" neutrino states.

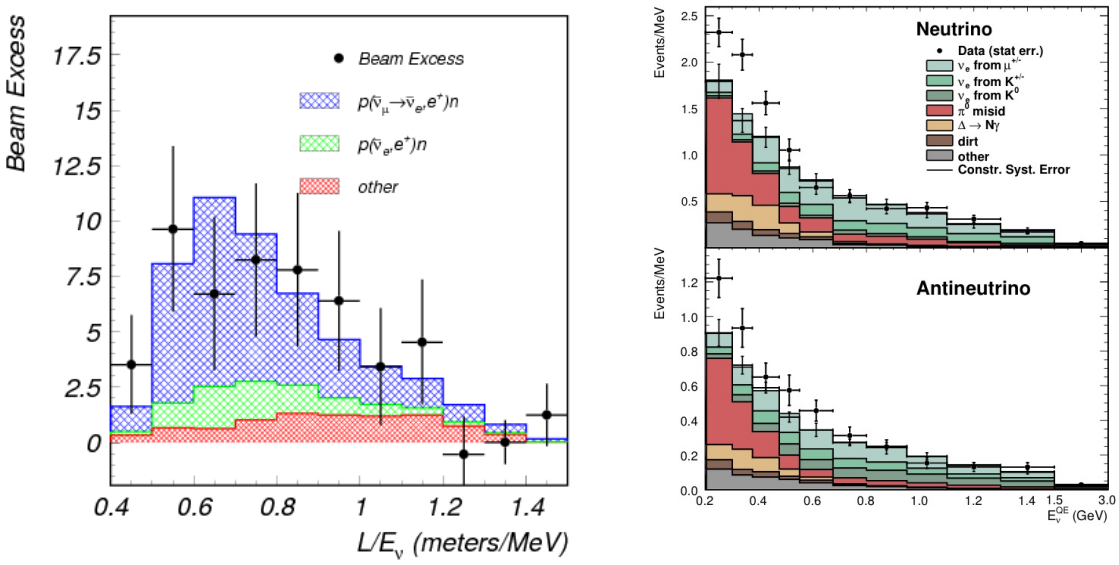


Figure 2.2: The LSND anomaly. Figure from Ref. [35] (left) and the low energy excess result from MiniBooNE experiment in neutrino and antineutrino mode. Figure from Ref. [36] (right).

The MiniBooNE experiment was designed to test the LSND result. The MiniBooNE experiment employs spherical detector located 541 m from the BNB neutrino source at Fermilab. Figure 2.3 shows the detector that consists of a sphere with a diameter of 12.2 meters filled with 818 tons of mineral oil located underneath more than 3 meters of earth overburden. MiniBooNE uses Cherenkov light for the neutrino detection. The detector was filled with mineral oil which was used as the neutrino target material. The particles exiting the target nuclei as a result of neutrino interactions produce Cherenkov light which was detected by the photomultiplier tubes (PMTs) lining the signal region of the detector. Figure 2.2 shows the result of the $\overline{\nu}_e$ appearance search in neutrino mode in MiniBooNE. In this figure, the dominant background is π_0 misidentification that arrises due to the fact that it is difficult to distinguish electrons from photons in MiniBooNE. We can clearly see the excess of data over prediction in the low energy region ($<$ 500 MeV). One possibility of the excess of data over prediction could be the underestimation of π_0 background. Since MiniBooNE saw similar excess result in the antineutrino mode as well, the other possibility leads towards the existence of sterile neutrinos.

The MicroBooNE experiment, which uses a technology that can distinguish electron from pho-

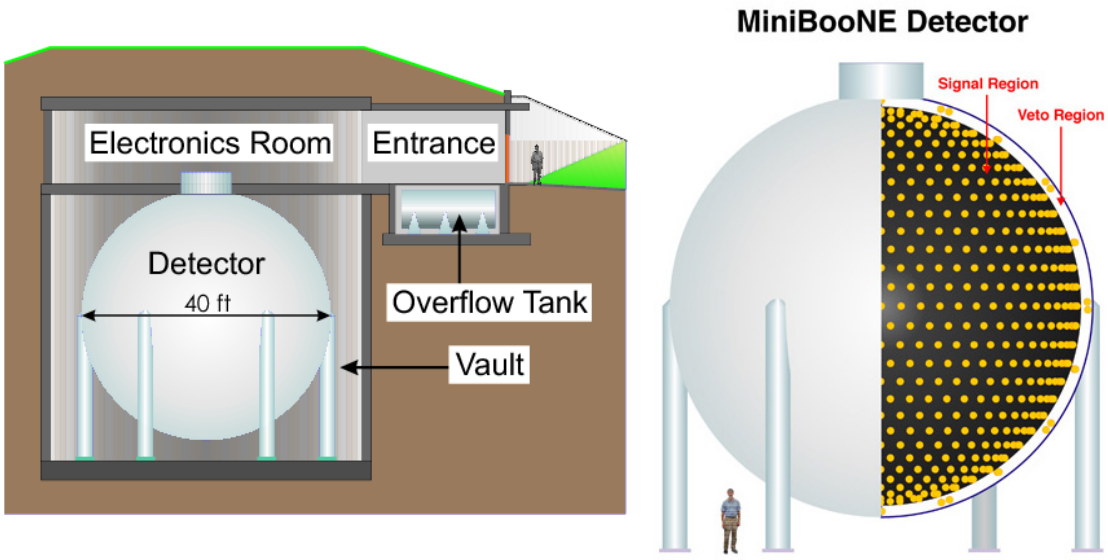


Figure 2.3: The MiniBooNE detector (left) and a drawing of the MiniBooNE detector showing the distribution of PMT's in the signal and veto regions (right). Figures from Ref. [34].

ton, is designed to test the MiniBooNE result and hence can confirm/nullify the data excess observed in the low energy region of MiniBooNE and can help with the sterile neutrino search.

2.5.2 Neutrino Mass Hierarchy

The recent discovery of neutrino oscillation proves the existence of non-vanishing neutrino masses. Now more experimental and theoretical efforts are needed so that the formulation of a new theory that includes the neutrino masses can be developed. The solar, atmospheric, accelerator, and reactor neutrino experiments have determined the necessary parameters describing the three-generation oscillations. The puzzle is yet to understand the neutrino mass hierarchy (or the neutrino mass ordering). In other words, it is still to be understood whether the ν_3 neutrino mass eigenstate (atmospheric neutrino) is heavier or lighter than the ν_1 and ν_2 (solar neutrino doublet). The case where ν_3 is heavier than ν_1 and ν_2 is called "normal" mass hierarchy, otherwise the hierarchy is called "inverted".

Figure 2.4 shows both possible hierarchies. As it is shown that ν_1 neutrino mass eigenstate have largest contribution coming from ν_e (lightest charged lepton), we expect "normal" hierarchy to be the one where mass of ν_1 is smaller than mass of ν_2 and ν_3. The inverted mass hierarchy shows

the opposite situation.

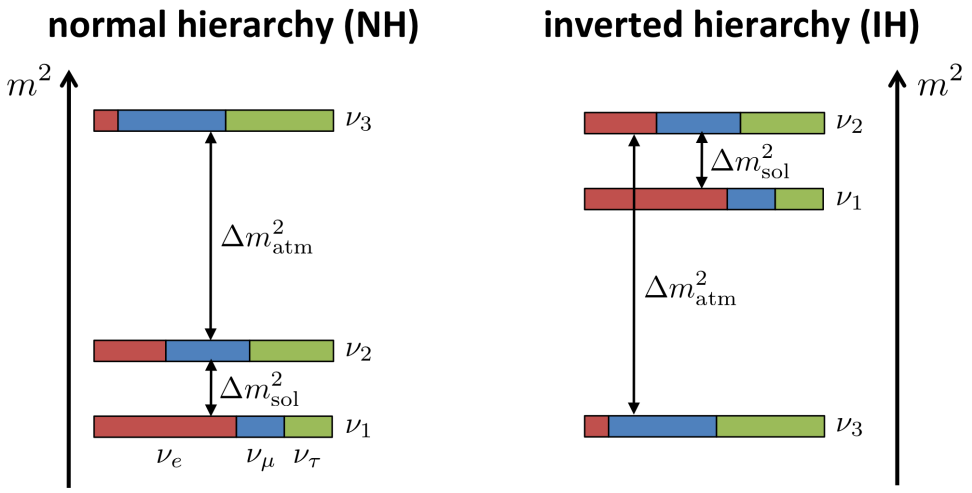


Figure 2.4: The normal and inverted neutrino mass hierarchies pattern is shown as mass squared.

Getting the knowledge of neutrino mass hierarchy is not trivial since in all experiments, the parameter that we determine is the mass-squared splitting, not the absolute mass of the eigenstates. Therefore, it is very challenging to determine the true mass hierarchy. Despite of its challenging nature, it is very important to determine the true neutrino mass hierarchy since the ultimate experimental and theoretical goal for all the neutrino experiments is to formulate a theory that can describe and explain the neutrino mixing and mass pattern which can be related to the charged lepton masses (and possibly to the quark mixing and masses).

In long baseline neutrino experiments such as DUNE, NOνA [37], HK [38], and T2K [39] have a potential to distinguishing between the normal ($m_3^2 - m_1^2 > 0$) and inverted ($m_1^2 - m_3^2 > 0$) neutrino mass hierarchies. This is due to the fact that earth matter effects, that play an important role in long baseline experiments, have opposite sign in the oscillation probability expression for the two hierarchies.

2.5.3 CP Violation

The CP-symmetry is a combination of charge and parity (mirror) symmetry. If we take a left-handed neutrino and apply both charge (C) and mirror (P) conjugation operators, we get a right-handed antineutrino which does exist.

In PMNS matrix, the value of θ_{13} that is obtained from the short baseline reactor and accelerator experiments is related to δ_{CP} which is known as the CP-violating phase. Therefore, only if we are sensitive to θ_{13}, we can measure the value of δ_{CP}. The value of δ_{CP} signifies how much CP is violated.

Since CP operator changes a particle into antiparticle, we can obtain the value of δ_{CP} by comparing results from both neutrino and antineutrino beams. Since antineutrino interaction cross section is lower than neutrino interaction cross section, it takes almost double amount of time in collecting antineutrino data compared to neutrino data hence this is not an ideal way to get δ_{CP}. The value of δ_{CP} modifies the oscillation pattern as a function of energy of neutrino for a long baseline experiment. In this way, an experiment that takes only neutrino data can, in principle, measure the δ_{CP}.

Some recent and future long baseline neutrino experiments such as NOνA, HK, and DUNE are designed to have good sensitivities to δ_{CP}. By combining the result from several experiments, the best sensitivity will be obtained.

2.6 Importance of Neutrino Interactions

To understand and answer the above-mentioned open questions about neutrinos (including neutrino oscillations), we need to better understand how neutrinos interact with the matter since the only way to detect neutrinos is to identify the products of their interactions. Understanding neutrino interactions with matter plays an important role in reducing the impact of systematic effects on

the neutrino oscillation measurement, especially when we are stepping towards the precision era and exploring CP violation, mass hierarchy, and sterile neutrino search. All of these analyses will require a precision understanding of neutrino interactions at a given neutrino energy range based on the particular detector that will be used for that particular analysis measurement. The only way to understand neutrino oscillations is by measuring neutrino interactions. Chapter 6 describes the neutrino interactions in detail. The thesis work described in chapter 7 is an effort to understand neutrino interactions at $\sim$1 GeV neutrino energy.

Chapter 3

The Booster Neutrino Beam

3.1 Introduction

This chapter describes the formation of neutrino beam in the Booster Neutrino Beam-line at the Fermi National Accelerator Laboratory. This beam has served the MiniBooNE experiment in the past, currently operating for MicroBooNE, and in future will serve SBND and ICARUS detectors in addition to MicroBooNE.

In order to properly interpret the charged particle multiplicity analysis results described in chapter 7, it is very important to understand how BNB is created and what is the BNB flux through MicroBooNE detector.

3.2 Booster Neutrino Beam Formation

In order to produce a neutrino beam, Booster first accelerates a beam of protons to 8 GeV kinetic energy (8.89 GeV/c momentum) and then sends those protons down to neutrino beam-line. The protons then hit a target and produce hadrons, primarily pions, which are further allowed to decay into neutrinos and leptons. A schematic diagram of the process of neutrino beam formation is shown in Figure 3.1. The subsequent sections describe this procedure in detail.

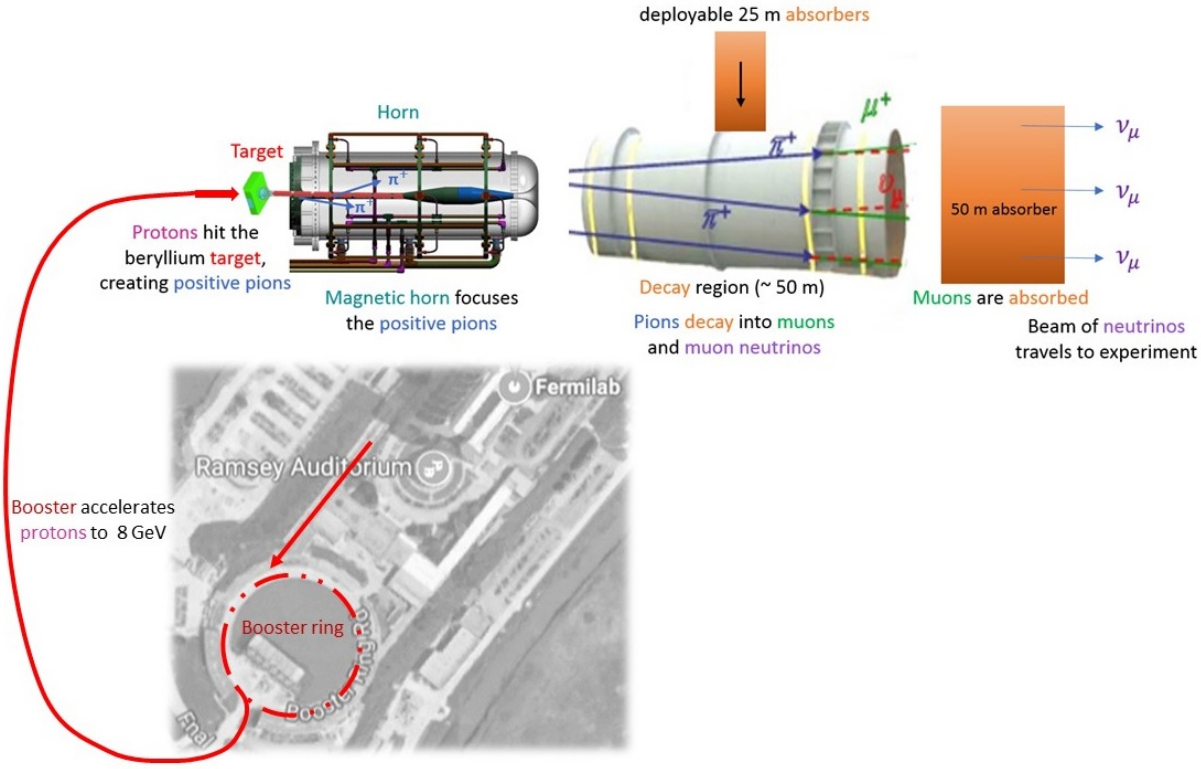


Figure 3.1: Schematic diagram for the BNB. The primary proton beam from the Booster enters the target hall, hits the target, produces mesons that decay into neutrinos in the decay region. Figure from Ref. [40].

3.2.1 Primary Proton Beam

The primary proton beam is created by converting hydrogen gas molecules (H_2) to H^- ions via Cockroft-Walton generator [41]. These ions are initially accelerated to 1 MeV kinetic energy and then are sent to the linear accelerator where they are accelerated up to 400 MeV kinetic energy using alternating electromagnetic fields. Electrons are removed when they pass through a carbon foil and the proton beam is sent to the Booster synchrotron where it acquires about 8 GeV kinetic energy by varying the magnetic field strength in time.

It is also important to note that the proton beam comes in bunches of protons called "beam spills", where each beam spill contains about 4×10^{12} protons and is separated by 1.6 μs time window from the next spill. The proton beam leaves the Booster synchrotron ring once it acquire 8 GeV kinetic energy and then is sent to target hall where it is impinged on a thick beryllium (Be) target.

3.2.2 Secondary Neutrino Beam

In the target hall, seven cylindrical slugs of beryllium are served as the proton beam target. The target is divided into short segments so that any forces on the assembly due to off-axis and asymmetrical heat loads from the proton beam can be minimized. The target consists of seven slugs, each one is 10.16 cm long and 1 cm in diameter. The Be is chosen as the target because it minimizes the beam power load on the target cooling system, has a high pion production yield, and minimizes the radioactivity due to proton exposure. Two toroids upstream of the target measure the number of protons-on-target (POT) within 2% of error.

The Be target is located inside a single toroidal aluminum alloy focusing electromagnet, called "horn" which is used to focus the charged mesons that are produced as a result of beam-target collisions. A schematic diagram and picture of the BNB horn are shown in Figure 3.2. Magnetic fields are generated between two coaxial conductors of the horn which is used to focus mesons. The current flows back and forth between inner and outer conductors generating a toroidal magnetic field. The maximum magnetic field strength of the horn is 1.5 T and the pulsed current has a peak at 170 kA. The horn polarity can be changed by changing the direction of current to focus either positively or negatively charged mesons, which in turn leads to switch between the neutrino (neutrino-mode) or antineutrino (antineutrino-mode) beam production.

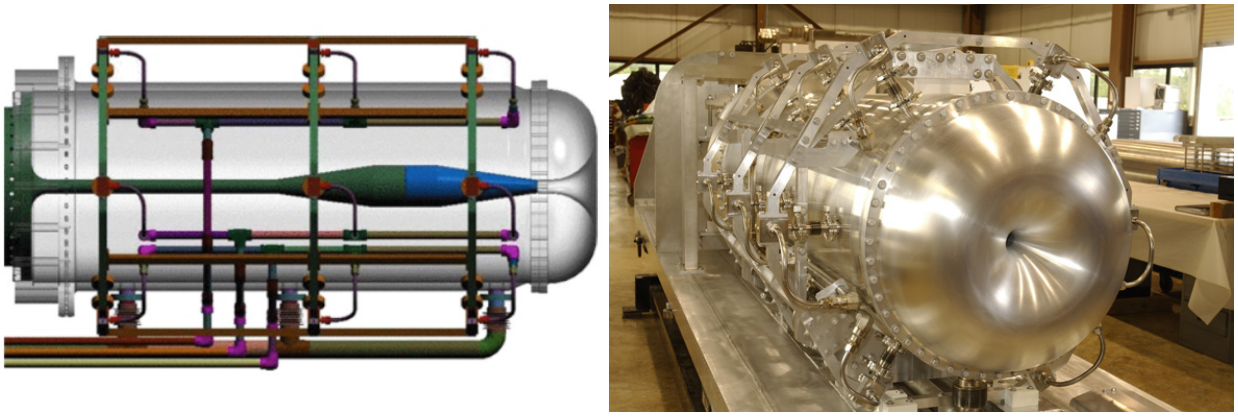


Figure 3.2: A schematic side view of the horn (left) and the BNB horn front view (right). Figures from Ref. [42].

Focused mesons are then allowed to pass through a 50 m long air-filled tunnel, called the decay

region, where the majority of mesons decay into muons and muon neutrinos. All charged particles are then absorbed into a steel absorber placed at the end of the decay region, thus allowing only neutrinos to pass. In this way a neutrino beam is produced in the Booster neutrino beam-line.

3.3 Neutrino Flux Predictions at MicroBooNE

The neutrino flux through MicroBooNE is determined using a GEANT4-based Monte Carlo (MC) simulation. This procedure is done in several steps. First the beam line geometry including its location, shape and BNB composition components is defined. Then according to the expected beam optics properties, the primary proton beam is generated upstream of the beryllium target. Then as the result of p-Be interactions, the particle production is simulated. Custom tables are developed for the production of neutrons, protons, $\pi^{\pm}$, $K^{\pm}$, and K^{0} created from production models based on external data. The details on the neutrino flux predictions in MiniBooNE (located right behind MicroBooNE) are given in [43]. The neutrino flux prediction at MicroBooNE is shown in Figure 3.3 which shows the BNB flux as a function of neutrino energy for different neutrino types.

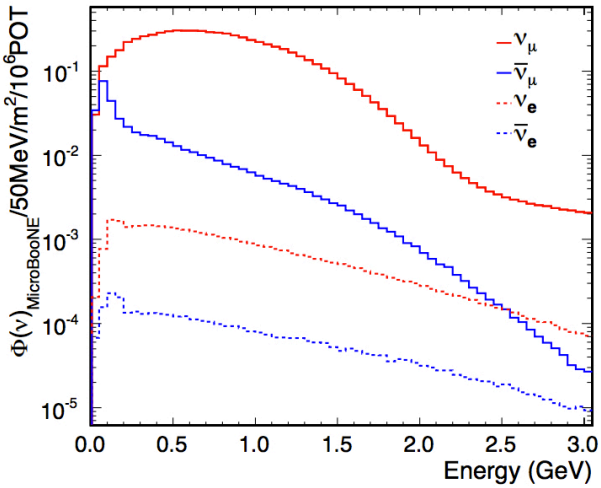


Figure 3.3: Booster Neutrino Beam flux as a function of neutrino energy at the MicroBooNE location. Figure from Ref. [44].

Chapter 4

The MicroBooNE Experiment

4.1 Introduction

The MicroBooNE detector (Fig. 4.1) is a Liquid Argon Time Projection Chamber (LArTPC) installed on the BNB at Fermilab. It has active mass of about 85 tonnes. MicroBooNE is currently the largest LArTPC that is operational in the United States. MicroBooNE is a high-resolution detector designed to be able to accurately identify neutrino interactions. It began collecting neutrino beam data in October of 2015.

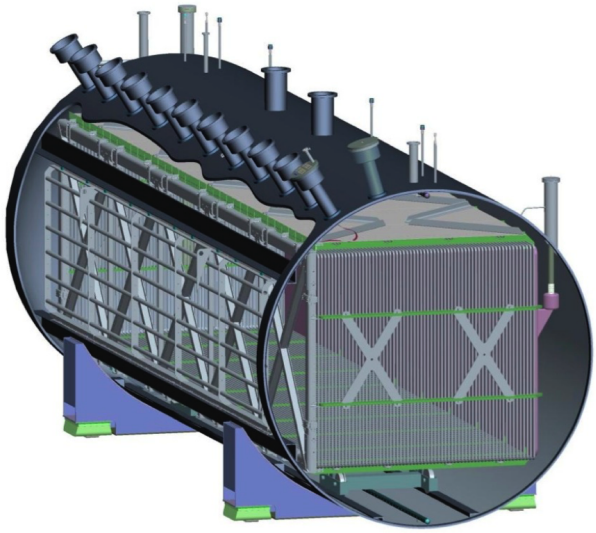


Figure 4.1: TheMicroBooNE detector. Figure from Ref. [4].

4.2 Goals of MicroBooNE

One of the primary goal of MicroBooNE experiment is to address the short baseline neutrino oscillations, primarily the excess of electron-like events at low neutrino energies that is observed by the MiniBooNE experiment as described in Sec. 2.5.1. Having placed on the same beamline and baseline as of the MiniBooNE, MicroBooNE uses a different approach by using a different technology to address the anomaly observed by MiniBooNE.

The other goal of MicroBooNE is to understand the neutrino interaction cross sections with argon at $\sim$ 1 GeV neutrino energies. These precision cross section measurements are critical for the neutrino oscillation analyses in addition to understanding the neutrino interactions in general. Also, understanding the neutrino-argon interactions is very important for the future short and long baseline LArTPC experiments. At BNB energies, there are several neutrino-argon interaction processes as well as nuclear effects that lead to the final state particles, as described in chapter 6.

MicroBooNE's additional goals involve the search for proton decay and supernova. The detector is capable of recording neutrinos coming from a galactic supernova and is developing algorithms for measuring the proton decay that would be beneficial for the future large scale LArTPCs.

4.3 Liquid Argon Technology

The choice of liquid argon as a target for neutrinos is attractive due to several reasons.

- Liquid argon, being a dense medium, results in a high neutrino-argon interaction probability which leads to a high event rate.
- The noble liquids produce large amount of ionization and scintillation light when a charged particle traverses through them.
- Argon is an affordable and abundant (1% of atmosphere) gas compared to the rest of noble gases.

Although liquid argon technology brings many desirable features, here are some points one needs to also consider about this target choice. Since argon atom has a complicated nuclear structure (relative to helium and hydrogen for example), it will introduce nuclear effects, which must be considered in data analyses. Also, to keep the target cool to very low temperatures up to $-300°$F, additional design considerations must be taken into account to ensure stable operations.

4.4 LArTPC

The rectangular shaped LArTPC (Fig. 4.2) is the core of the MicroBooNE detector that drifts and collects the charge produced due to ionization that is liberated by the charged particles traversing through the liquid argon. The MicroBooNE TPC has an active mass of about 85 tonnes of liquid argon. It is 10.4 meters long in the beam direction, 2.3 meters tall, and 2.6 meters in the electron drift direction. It has three major components: the cathode, the field cage, and the anode. A high negative voltage (-70 kV) is applied to the cathode of the detector via a feedthrough. The field cage, consisting of a series of field rings connected to each other by a voltage divider chain, starts from the cathode and ends on the anode. The anode consists of three wire planes: two induction planes (referred to as "U" and "V" planes) and a collection plane (CP) (referred to as "Y" plane). The U and V planes are oriented at $\pm 60°$ relative to vertical whereas Y plane wires are oriented vertically. The U and V planes have 2400 wires each and are kept at -110 V and 0 V respectively whereas the Y plane has 3256 wires and is kept at $+230$ V. The distance between two adjacent wires as well as spacing between two consecutive wire planes (also called as "wire pitch") is 3 mm.

When the incident neutrino beam hits a target argon nucleus inside TPC, the charged particles get emitted. These charged particle traverse through medium where they ionize many argon atoms along their way. Due to the external applied electric field, the electrons start to drift towards cathode. The drifting electrons require 2.3 ms to drift across the full width of the TPC from cathode to anode. These electrons pass through the first two (induction) wire planes producing bipolar induction signals. The third (collection) wire plane collects the charge of the drifting

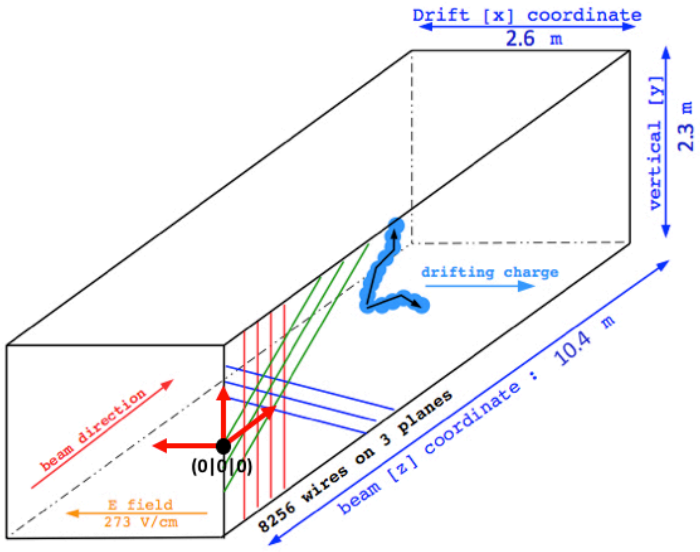


Figure 4.2: A schematic of the MicroBooNE TPC showing the coordinate system and wire planes. Figure from Ref. [45].

electrons in the form of a unipolar signal.

4.5 Light Collection System

The light collection system of MicrBooNE consisting of 32 8-inch photomultiplier tubes (PMTs) is located behind the anode plane wires immersed in the liquid argon contained within a cryostat having 170 tonne capacity. Since liquid argon is a very good scintillator, the light information coming from the neutrino-argon interactions in the TPC is stored and utilized by the light collection system for proper functioning of the detector. Figure 4.3 presents a diagram of an optical unit (left) and a picture of an installed light collection system in the MicroBooNE (right).

One of the critical capabilities of the light collection system having a nanosecond timing resolution is to enable the precise determination of the time of neutrino interaction, which crucially aids in the reduction of cosmic ray backgrounds. The other very important function is to aid in triggering of the detector when a light pulse is observed in coincidence with the beam spill. This reduces the rate of cosmic backgrounds by $1/600$, which subsequently helps reducing the data storage.

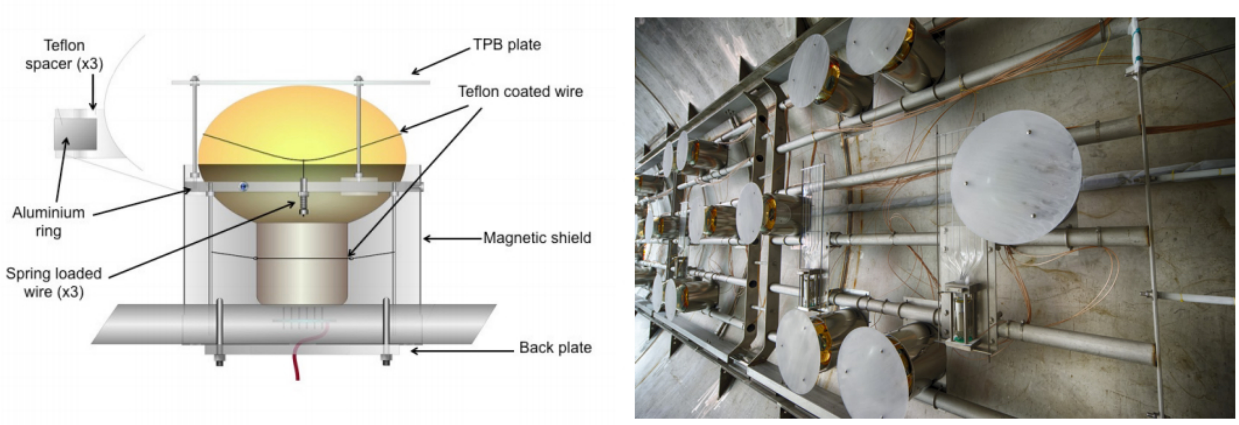


Figure 4.3: Diagram of the optical unit (left) and units mounted in MicroBooNE cryostat (right). Figures from Ref. [4].

4.6 Signal Processing

The raw signal developed on the TPC wires and PMTs must be processed before using in analyses [46]. The MicroBooNE readout electronics allow for measurement of both the time and charge created by drifting electrons on each wire. First, the analog signals/waveforms that are developed on LArTPC 8256 wires from induction and collection planes are amplified and shaped. The data from shaped waveforms and PMTs is then sent to readout system that is digitized at 2 MHz using 12-bit ADCs. A data acquisition (DAQ) system readout window consisting of 9600 recorded samples (4.8 msec) for all wires then noise-filters [47] and deconvolves the waveforms utilizing the offline software algorithms. Reconstruction algorithms are then used on these output waveforms to reconstruct the times and amplitudes of charge depositions (hits) on the wires from particle-induced ionization in the TPC bulk. The DAQ system then stores the data on disk. Figure 4.4 shows a schematic diagram for the MicroBooNE signal processing and readout stages.

4.7 Cosmic rays in MicroBooNE

The MicroBooNE detector lacks appreciable shielding from cosmic rays (CR) since the detector is at the earth's surface and has little overburden. Most events that are recorded and processed through an online software trigger which requires that the total light recorded with the PMT sys-

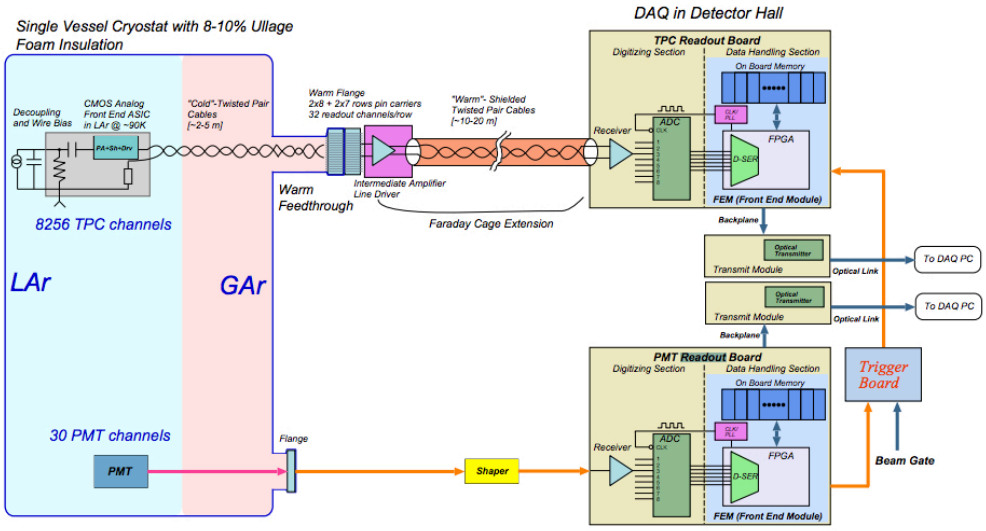


Figure 4.4: MicroBooNE signal processing and readout stages. Signals fist pass through cold electronics (at liquid Ar temperatures on the left), then signal coming from TPC and PMT pass through feedthroughs into warm electronics readout boards. In DAQ, signals are digitized and combined with external triggers coming form the accelerator and then are sent to disk for storage. Figure from Ref. [4].

tem exceeds 6.5 photoelectrons (PE) during neutrino beam operations ("on-beam data") contain no neutrino interactions. Triggered events with a neutrino interaction typically have the products of up to 20 cosmic rays coincident with the beam spill in the event readout window (4.8 msec) contributing to a recorded event along with the products of the neutrino collision. Figure 4.5 shows a typical 3D event display from MicroBooNE that shows the average number of cosmic rays in an event. A large sample of events recorded under identical conditions as the on-beam data, minus the coincidence requirement with the beam, ("off-beam-data") has been recorded for use in characterizing cosmic ray backgrounds. A straightforward on-beam minus off-beam background subtraction is difficult, as the off-beam data does not reproduce all correlated detector effects associated with on-beam events that contain a neutrino interaction with several overlaid cosmic rays. The situation is particularly complicated with events containing neutrino interactions with only one observed track, which share the same topology with the most common single-muon CR configuration. Monte Carlo simulations of the CR flux using the CORSIKA package [48] provide useful guidance; however, the ability of these simulations to describe the very rare CR topologies that closely match neutrino interactions is not well known.

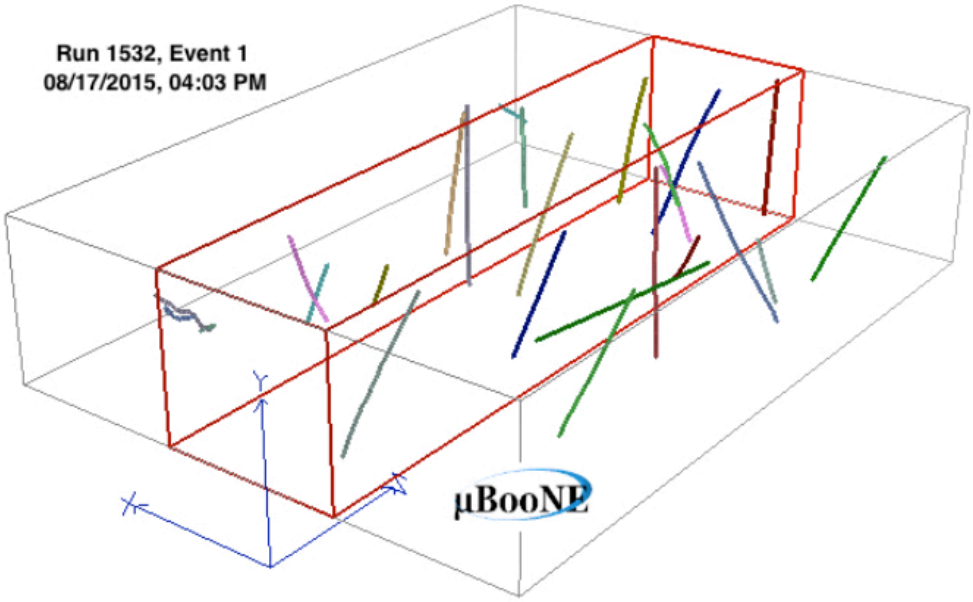


Figure 4.5: A 3D event display from MicroBooNE showing average number of observed cosmic rays in an event. Figure from Ref. [49].

For these reasons this analysis employs a method to separate neutrino interaction candidates from CR backgrounds that is driven by the data itself. Even though CR tracks should always appear to at least enter the detector, they can satisfy the experimental condition of being fully contained if a segment of the CR track falls outside the data acquisition readout time window, or if a segment of the track fails to be identified due to instrumentation- or algorithm-related inefficiencies. The separation of neutrino interaction candidates from CR backgrounds rests on the observation that a neutrino ν_μ CC interaction produces a final state μ^- that slows down as it moves away from its production point at the neutrino interaction vertex due to ionization energy loss in the liquid argon. As it slows down, its rate of restricted energy loss, dE/dx_R, increases, and deviations from a linear trajectory due to multiple Coulomb scattering (MCS) become more pronounced. A CR muon track can produce an apparent neutrino interaction vertex if it comes to rest in the detector or it is not fully reconstructed to the edge of the TPC, but the CR track will exhibit large dE/dx_R and MCS effects in the vicinity of this vertex. Furthermore, the vast majority of ν_μ CC muons travel in the neutrino beam direction ("upstream" to "downstream"), whereas CR muons move upstream or downstream with equal probability.

Chapter 5

Event Reconstruction in MicroBooNE

In order to fully utilize the imaging capabilities offered by LArTPC technology, new approaches to pattern recognition are required. For this purpose, several automated event reconstruction techniques are being developed in MicroBooNE. The reconstruction package I used in this analysis is known as *"Pandora"* [50–52]. This chapter briefly summarizes the Pandora algorithm. The text in this chapter follows closely the Pandora paper [52].

In MicroBooNE, the Pandora algorithm has two stages. The first reconstruction stage, PandoraCosmic, is used to reconstruct cosmic and delta rays in an event whereas the second stage, PandoraNu, is used to reconstruct neutrino interaction products. Figure 5.1 shows the Pandora reconstruction chain in MicroBooNE. In PandoraCosmic, all anode plane hits are used to identify cosmic-ray particles. Then the hits associated with these cosmic-ray particles are removed from the input hit collection and the remaining set of hits is sent as an input to the second stage, PandoraNu, where the neutrino candidate vertices and particles are created.

5.1 Hit Reconstruction

Hit reconstruction makes use of the anode plane waveforms from the TPC. Raw signals from the TPC, recorded on the wires, pass through a noise filter, deconvolution, and calibration. A peak

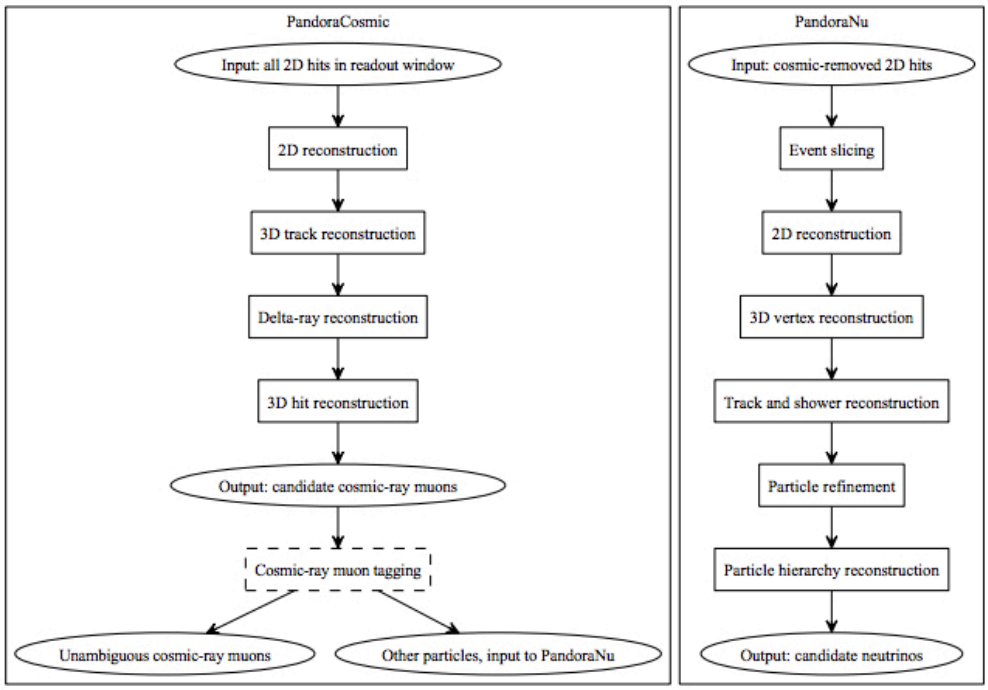


Figure 5.1: Pandora reconstruction chain in MicroBooNE. Figure from Ref. [52].

finder looks for bumps (energy deposition) in the resultant waveforms and fits them to gaussian curves. These resultant reconstructed hits are called "gaushits". The quality of the fit as well as additional parameters such as peak time and area under gaussian fit (charge) is also stored.

5.2 Track Reconstruction

Track reconstruction algorithms in Pandora are tuned to reconstruct cosmic rays and neutrino interaction track separately. The reconstruction of both type of tracks begin in a similar way. First, the hits connected in space and time are combined into two-dimensional (2D) clusters for each of the three anode planes. This initial clustering provides clusters with high purity than completeness. In other words, those clusters are preferred in which the energy deposited on the hits come from one true particle only, over those where a large fraction of the total hits are associated with the true particle. Then several cluster-merging algorithms are applied to group multiple 2D clusters to improve their completeness. The pairs of clusters in a close proximity or pointing towards each others are identified. Based on the overall event topology, the chains of associated clusters are

identified. Track reconstruction algorithms are also used to merge the clusters across unresponsive wire regions. Figure 5.2 shows an example of two cosmic crossing tracks from MicroBooNE simulation where the cluster merging algorithms work across the unresponsive wires and merge the different clusters into one, after the application of the full algorithm chain. .

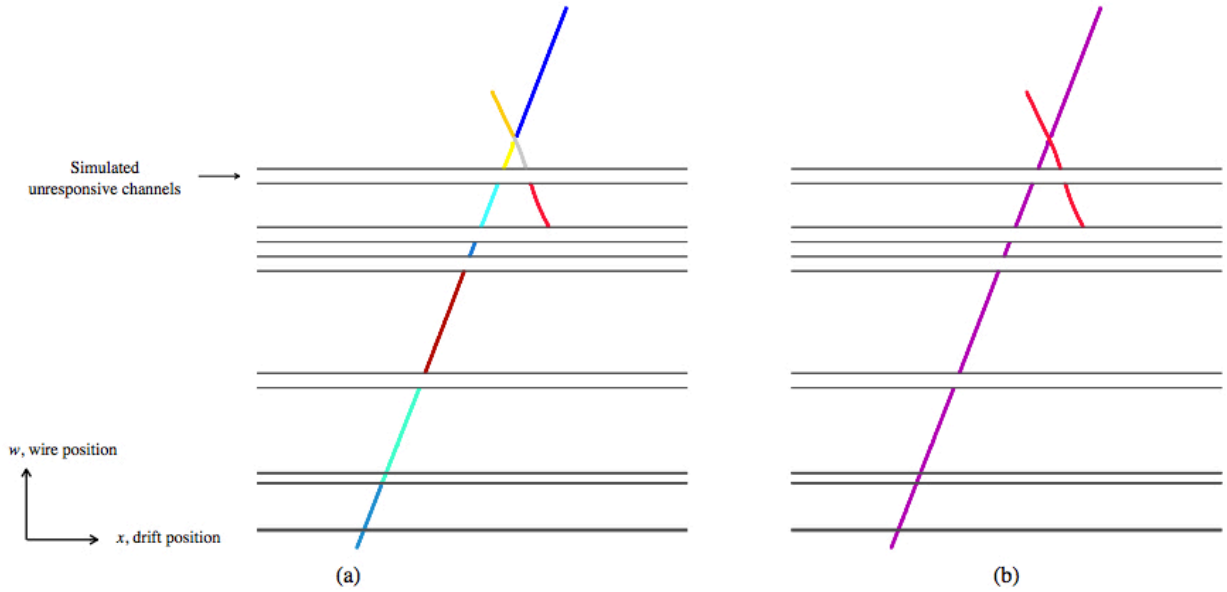


Figure 5.2: Clusters produced by two crossing cosmic ray muon tracks. Different colors represent different clusters before applying the full chain of cluster merging algorithms (a); after applying the full chain of cluster merging algorithms (b). Figures from Ref. [52].

2D clusters across different anode planes are grouped together to form three dimensional (3D) tracks. The challenge is to identify consistent groups of clusters across different planes. This cluster matching is done according to the following steps: The algorithm first determines a pair of clusters in different plane views that overlap in time. Then the position of cluster in the third plane view is predicted. This position is then compared with the sliding-fit position of the candidate cluster in third plane view. A χ^2 representing the quality of the fit is calculated and stored.

5.3 Neutrino Vertex Reconstruction

In this very important step of reconstruction, first a list of all possible vertex positions in an event is created. Then a pair of clusters that overlap in time in two different plane views is chosen and

their end positions are compared. Each pair of cluster, from two plane views, thus provides four possible candidate vertex positions. The decision of selecting the best candidate neutrino vertex depends on an algorithm that assigns scores to each candidate vertex and the candidate with the highest score is selected to be the most likely neutrino vertex. The three components of the score are the following:

- **Energy-kick score:** This makes use of the fact that the primary particles from neutrino interaction all point towards the vertex and are energetic compared to the secondary particles.
- **Asymmetry score:** This removes candidates having symmetric number of hits in upstream and downstream of the candidate vertex position. This is based on the fact that the true neutrino vertex should have asymmetric number of hits on either end.
- **Beam deweighting score:** Here, the knowledge of the beam direction is used to preferentially select vertex candidates with low z coordinates.

5.4 Optical Reconstruction

Light collected on the 32 PMTs in MicroBooNE is used to reconstruct optical hits. To reduce sensitivity to possible fluctuations in the signal baseline, a threshold-based hit-reconstruction algorithm requires PMT pulses of a minimum charge for a hit to be reconstructed. A weighted sum of PMT hits (optical flashes) are reconstructed by requiring a time coincidence of $\sim$1 μs between hits on multiple PMTs. The relative timing and charge collection of optical hits, along with the spatial locations of the PMT, within an optical flash is then used to associate the flash with reconstructed tracks in the TPC, a process known as flash-matching.

5.5 Reconstruction Performance

To check the performance of Pandora reconstruction package in MicroBooNE, we define a metric "efficiency" that represents the fraction of true particles of a type ($\mu^{\pm}$, $\pi^{\pm}$, p, K etc) that match

with at least one reconstructed particle. Figure 5.3 represents the number of hits versus efficiency plot for muons and protons from all simulated BNB interactions. We see that the efficiencies for reconstructing muon and proton tracks are approximately 70% and 45%, respectively, at the analysis thresholds (80 and 15 collection plane hits for muons and protons, respectively).

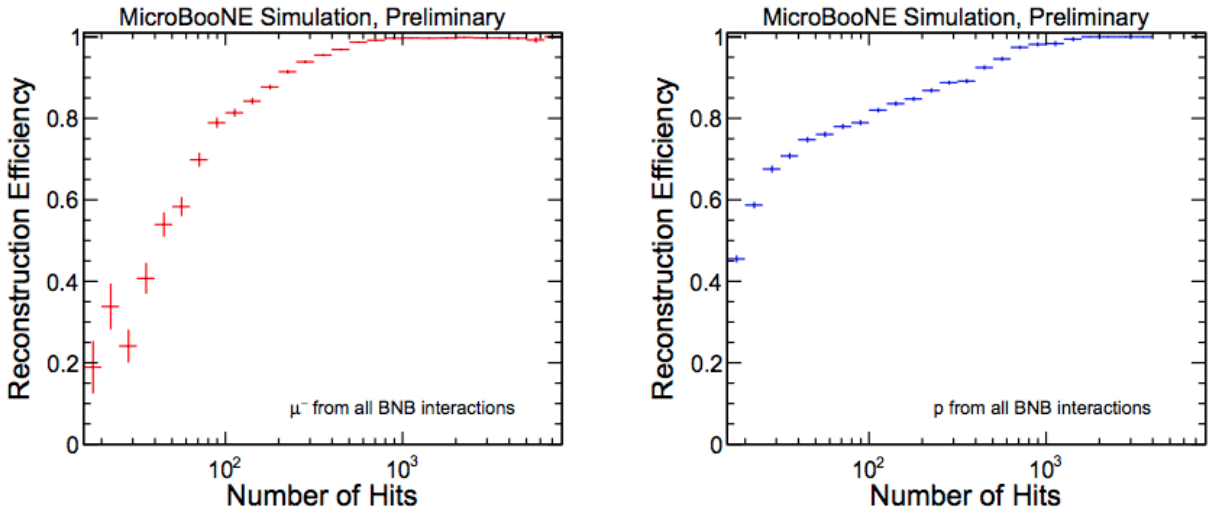


Figure 5.3: Number of hits versus track reconstruction efficiency in all simulated BNB interaction for muons (left) for protons (right). Figures from Ref. [51].

Chapter 6

Neutrino Interactions

In order to do an exciting neutrino physics measurement using a neutrino experiment, it is very important to estimate the number of neutrino interactions we expect to observe in the detector and to understand the nature of these interactions. The expected number of neutrino interactions in a detector is directly proportional to the neutrino flux at the detector location, the number of targets in the detector, and the neutrino interaction cross section.

$$N_\nu(E) \sim \Phi_\nu(E) \times N_{target} \times \sigma_\nu(E) \tag{6.1}$$

The understanding of neutrino interactions is very important for several analyses beyond cross sections that include oscillation measurements, CP violation search, and mass hierarchy studies.

Neutrino-nucleus interactions are separated into three energy regions: high, intermediate, and low energy regions. At high neutrino energies (on the order of 100 GeV), the de Broglie wavelength is very small (~ 0.1 fm) thus the nuclear effects are suppressed. At intermediate energies (on the order of 1 GeV, used by many accelerator based neutrino experiments), the de Broglie wavelength becomes 1 fm and thus nuclear effects become important. At low energies (on the order of MeV, used by reactor and solar neutrino experiments), de Broglie wavelength becomes greater than the nuclear diameter and thus the initial and final states are specific nuclear levels.

Neutrinos interact only via the weak force. They interact rarely and their cross sections are very small, typically of the order of 10^{-38} cm^2 at around 1 GeV neutrino energies.

There are two types of neutrino interactions: charged current (CC) and neutral current (NC) interactions. CC interactions are mediated via $W^{\pm}$ whereas NC interactions are mediated via Z bosons. In CC interactions, the outgoing lepton flavor (μ, e, τ) is always the same as the incoming neutrino flavor. On the other hand, in the NC interactions, there are no outgoing leptons and hence we can not get any information about the incoming neutrino flavor. Figure 6.1 presents one example each of CC and NC neutrino interaction types and the corresponding candidate events from MicroBooNE data.

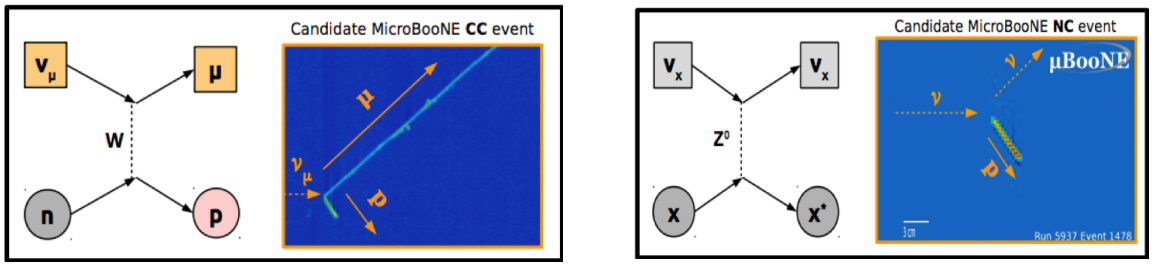


Figure 6.1: Charged current (CC) neutrino interaction topology and and a candidate CC event from the MicroBooNE data (left); Neutral current (NC) neutrino interaction topology and and a candidate NC event from the MicroBooNE data (right). Figure from Ref. [53].

This analysis makes use of the CC interactions only since the event selection requires to have a candidate lepton in the final state of interaction (see chapter 7 for details). There are several scattering processes in CC interactions. At MicroBooNE average neutrino energies ($\sim$ 1 GeV), the dominant processes are quasi-elastic (CCQE), resonance (CCRES), and deep inelastic (CCDIS) interactions. The details of each of them are described below.

6.0.1 CCQE

The CCQE interactions are the most dominant interactions at MicroBooNE neutrino energies and require relatively lower neutrino energies. In these interactions, the incoming neutrino scatters off of a free nucleon inside the nucleus. A $W^{\pm}$ gets exchanged at the interaction point and the final state particles include a lepton and a nucleon. Since BNB has a high flux of ν_{μ}, most of CCQE events in MicroBooNE can be observed by identifying a muon (which is usually a long, minimum

ionizing track) and a proton (which is usually a short, highly ionizing track) as final state particles in the detector.

$$\nu_\mu + n \rightarrow \mu^- + p \tag{6.2}$$

Figure 6.2 shows the most dominant CCQE event topology observed in MicroBooNE. These interactions occur at relatively low energies and the cross section of these interactions at MicroBooNE energies is largest among all interaction types.

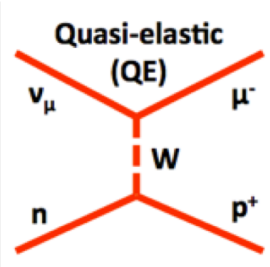


Figure 6.2: CCQE event topology.

6.0.2 CCRES

In CCRES interactions, the nucleus excites into a baryonic resonance state and then de-excites and creates different particles. There are three interaction modes for CCRES interactions depending on the final state particles.

$$\nu_\mu + p \rightarrow \Delta^{++} \rightarrow \mu^- + p + \pi^+ \tag{6.3}$$

$$\nu_\mu + n \rightarrow \Delta^{+} \rightarrow \mu^- + p + \pi^0 \tag{6.4}$$

$$\nu_\mu + n \rightarrow \Delta^{+} \rightarrow \mu^- + n + \pi^+ \tag{6.5}$$

Figure 6.3 shows an example of a CCRES event observed in MicroBooNE. Identification of pion is bit challenging since it is difficult to differentiate between muon and pions. Several techniques are being developed in MicroBooNE to identify different types of particles but the study

performed in this analysis is an inclusive study therefore I do not need to identify and distinguish different types of particles.

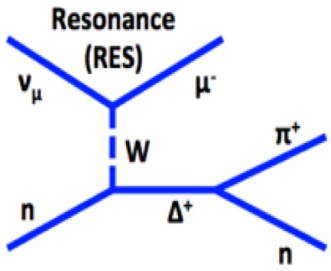


Figure 6.3: CCRES event topology.

6.0.3 CCDIS

The CCDIS interactions require relatively higher neutrino energies. In these interactions, the nucleus breaks up and multiple hadrons are produced in addition to the lepton.

$$\nu_\mu + n \rightarrow \mu^- + X \tag{6.6}$$

Figure 6.4 shows a CCDIS event topology. These interactions occur at relatively high energies and the cross section of these interactions at MicroBooNE energies is smaller than the other two interaction types described above.

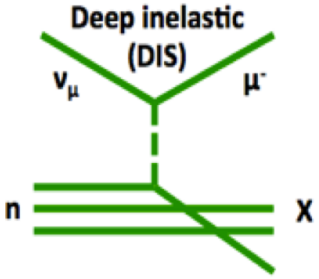


Figure 6.4: CCDIS event topology.

6.1 GENIE: Neutrino Event Generator

The development of a neutrino physics generator is a challenging task due to various reasons. One of them is because of the lack of theoretical and phenomenological understanding of neutrino interactions at few GeV energy range. At this energy range, a variety of nuclear scattering mechanisms become important. It is challenging to piece together different models with different ranges of validity to generate events that cover full phase space. Additional problem is the lack of availability of experimental data in this energy range that can be used to tune the generator. Another challenge is to incorporate the rapidly evolving nature of neutrino physics experiments and theory. This is due to the fact that neutrinos are less understood particles compared to many other particles from standard model. The variety of topics in neutrino physics that are currently being studied include understanding of neutrino oscillations, sterile neutrinos search, CP violation search, and mass hierarchy problem. This wide variety of topics of interest makes it a difficult task to model and generate neutrino interactions.

GENIE [8] (Generates Events for Neutrino Interaction Experiments) is a generator used to model neutrino interactions. It is not based on a particular theory instead it is a combination of different neutrino interaction models added in different proportions based on the data obtained from past and present neutrino experiments. It is a generator that has been extensively used, validated, and tuned with the external data coming from different neutrino experiments. It has not yet been tested in detail on the experiments that use argon as a target atom. The analysis described here provides the most comprehensive test of GENIE ever performed on the argon data.

In this analysis, I test three models: GENIE default, GENIE+MEC, and GENIE+TEM. Details of each of these models are described below.

6.1.1 GENIE Default

The GENIE default model used in this analysis includes all the CC and NC processes and corresponding interaction channels (QE, RES, and DIS) described above. There are few more additional channels including coherent pion production [54] which has a very small cross section at Micro-

BooNE neutrino energies. In GENIE, the models used for QE, RES, and DIS interactions are Llewellyn-Smith model [55], Rein-Sehgal model [56], and an effective leading order model using the modifications suggested by Bodek and Yang [57], respectively. In addition to these models, the Fermi gas model [58] is used to calculate the effects such as Fermi motion and Pauli blocking for bound nucleon targets.

For large target nuclei, the particles produce in primary neutrino interactions often ($\sim$ 30% in iron at few GeV neutrino energy) have final state interactions (FSI). These FSI bring complexities in neutrino interactions but these effects must be considered for atoms having large nuclei such as argon. Around 20% of the pions produced in RES events in carbon get absorbed as the result of FSI and can look like QE events in the detector. A proper understanding of FSI is critical for a high quality Monte Carlo (MC). Figure 6.5 shows an example of an RES interaction where the primary particles get absorbed within the same nucleus and other particles are produced in FSI.

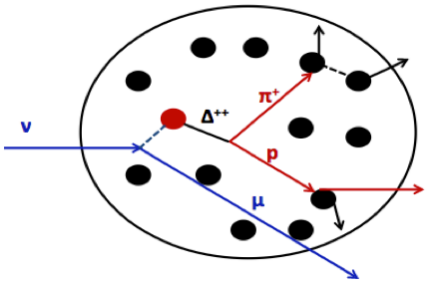


Figure 6.5: A resonance event where primary particles get absorbed within the same nucleus as the result of FSI.

6.1.2 GENIE+MEC

GENIE+MEC has all the processes and models that are present in GENIE default with an addition of Meson Exchange Current [59] model. The model referred to as MEC in the following plots is the empirical model for MEC by Dytman [60]. The MEC is often referred to as two-particle-two hole (2p2h) [61] final states in neutrino-nucleus scattering. In this interaction, neutrino scatters off of a bound state of a pair of neutron and proton. As the result of this interaction, two protons are produced. Figure 6.6 shows a picture of MEC interaction.

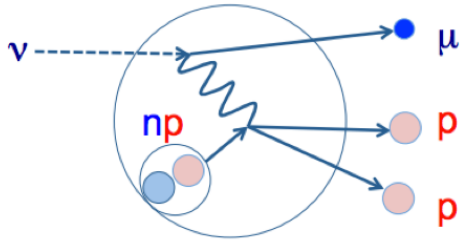


Figure 6.6: MEC interaction picture.

6.1.3 GENIE+TEM

GENIE+TEM has all the processes and models that are present in GENIE default with an addition of Transverse Enhancement Model [62]. The TEM, the empirical superscaling function is modeled with the effective spectral function (ESF) [63]. This model enhances the transverse component of the QE cross section in the low neutrino energy regime and it has been used successfully in the electron scattering data.

Chapter 7

Tests of Neutrino Interaction Models in MicroBooNE

7.1 Introduction to Observed Charged Particle Multiplicity and Kinematic Distributions

Neutrino interactions in the MicroBooNE detector produce charged particles that can be reconstructed as tracks in the liquid argon medium of the MicroBooNE LArTPC. These interactions can be characterized by a number of inclusive properties. The charged particle multiplicity, or number of primary charged particles, n, is a simple observable characterizing final states in high-energy-collision processes, including neutrino interactions. I note that in MicroBooNE the observable charged particle multiplicity corresponds to that of charged particles exiting the target nucleus participating in the neutrino interaction.

The charged particle multiplicity distributions (CPMD) comprise the set of probabilities, P_n, associated with producing n charged particles in an event, either in full phase space or in restricted phase space domains. In addition to the observed CPMD, kinematic properties of all charged particle tracks for each multiplicity can be examined. Determination of inclusive event properties such as the CPMD and of individual track kinematic properties at Fermilab BNB neutrino energies

naturally fits into the modern strategy [11] of presenting neutrino interaction measurements in the form of directly observable quantities.

Inclusive measurements expand the empirical knowledge of neutrino-argon scattering that will be required by the DUNE experiment and the Fermilab SBN program. As physical observables, the CPMD and other distributions can also be used to test models, or particular tunes of models such as GENIE. These models are typically constructed from a set of exclusive cross section channels, and tests of inclusive distributions can provide independent checks.

I describe here an evaluation of several variants of GENIE against *observed* charged particle distributions, including the observed CPMD in MicroBooNE data collected in 2016 in the Fermilab BNB. For the observed CPMD, I mean the conditional probability, after application of certain detector selection requirements, of observing a neutrino interaction with n charged tracks relative to the probability of observing a neutrino interaction with at least one charged track:

$$O_n = \frac{N_{\text{obs},n}}{\sum_{m=1}^{\infty} N_{\text{obs},m}}, \tag{7.1}$$

where $N_{\text{obs},n}$ is the number of neutrino interaction events with n observed tracks.

This analysis requires at least one of the charged tracks to be consistent with a muon; hence the O_n are effectively observed CPMD for ν_μ charged current (ν_μ CC) interactions. The ν_μ NC, ν_e, $\bar{\nu}_e$, and $\bar{\nu}_\mu$ backgrounds, in total, are expected to be less than 10% of the final sample. The muon candidate is included in the charged particle multiplicity, and all events thus have $n \geq 1$. For each multiplicity, I have available the kinematic properties of charged tracks. These can in principle be related to the 4-vector components of each track; however, I choose distributions of directly observable quantities in the detector: visible track length and track angles.

Values for O_n depend on cross sections for producing a multiplicity n, $\sigma_{CC,n}$, as well as the BNB neutrino flux and detector acceptance and efficiency:

$$\begin{aligned} N_{\text{obs},n} = &\sum_{\nu}\sum_{n'}\int dE_\nu \Phi_\nu\left(E_\nu\right) \\ &\cdot\int d\Pi_{n'} \frac{d\sigma_{CC,n'}\left(E_\nu,\Pi_{n'}\right)}{d\Pi_{n'}} \varepsilon_{n,n'}\left(E_\nu,\Pi_{n'}\right), \end{aligned} \tag{7.2}$$

where E_ν is the neutrino energy, $\Phi_\nu(E_\nu)$ is the neutrino flux summed over ν_μ, $\bar{\nu}_\mu$, ν_e, and $\bar{\nu}_e$ species, $d\Pi_{n'}$ represents the n'-particle final state phase space, $\varepsilon_{n,n'}(E_\nu, \Pi_{n'})$ is an acceptance and efficiency matrix that gives the probability that an n' charged particle final state produced in phase space element $d\Pi_{n'}$ is observed as an n'-particle final state in the detector, and $d\sigma_{CC,n'}(E_\nu, \Pi_{n'})/d\Pi_{n'}$ are the differential cross sections for producing a multiplicity n'. One can likewise express the distribution of any observed kinematic distribution X_n corresponding to an observed multiplicity n as

$$\begin{aligned} dN_{\mathrm{obs},n} = &\sum_\nu \sum_{n'} \int dE_\nu \Phi_\nu(E_\nu) \\ &\cdot \int d\Pi_{n'} \frac{d\sigma_{CC,n'}(E_\nu, \Pi_{n'})}{d\Pi_{n'}} \hat{\varepsilon}_{n,n'}(E_\nu, \Pi_{n'} \to X_n), \end{aligned} \tag{7.3}$$

where $\hat{\varepsilon}_{n,n'}(E_\nu, \Pi_{n'} \to X_n)$ is the probability that an n' charged particle final state produced in phase space element $d\Pi_{n'}$ produces the observed value X_n of the kinematic variable in the detector. In practice I obtain the O_n and distributions of X_n directly from data and compare these to values derived from evaluating Eqs. 7.2 and 7.3 using a Monte Carlo simulation that includes GENIE neutrino interaction event generators coupled to detailed GEANT-based models of the Fermilab BNB and the detector.

The observed CPMD and inclusive observed kinematic distributions have several desirable attributes. The $\sigma_{CC,n}$ are all large up to $n \lesssim 4$ at these neutrino energies; therefore only modest event statistics are required. Only minimal kinematic properties of the final state are imposed (the track definition implies an effective minimum kinetic energy), and complexities associated with particle identification and photon reconstruction are avoided. At the same time, the observed quantities reveal much of the power of the LArTPC in identifying and characterizing complex neutrino interactions. The observed CPMD and associated kinematic distribution ratios will have reduced sensitivity to systematic normalization uncertainties associated with flux and efficiency compared to absolute cross section measurements. Figure 7.1 presents four candidate neutrino events having different observed multiplicities from MicroBooNE data. Red dots correspond to neutrino interaction vertices.

A disadvantage of the use of observed CPMD and other kinematic quantities is their lack of

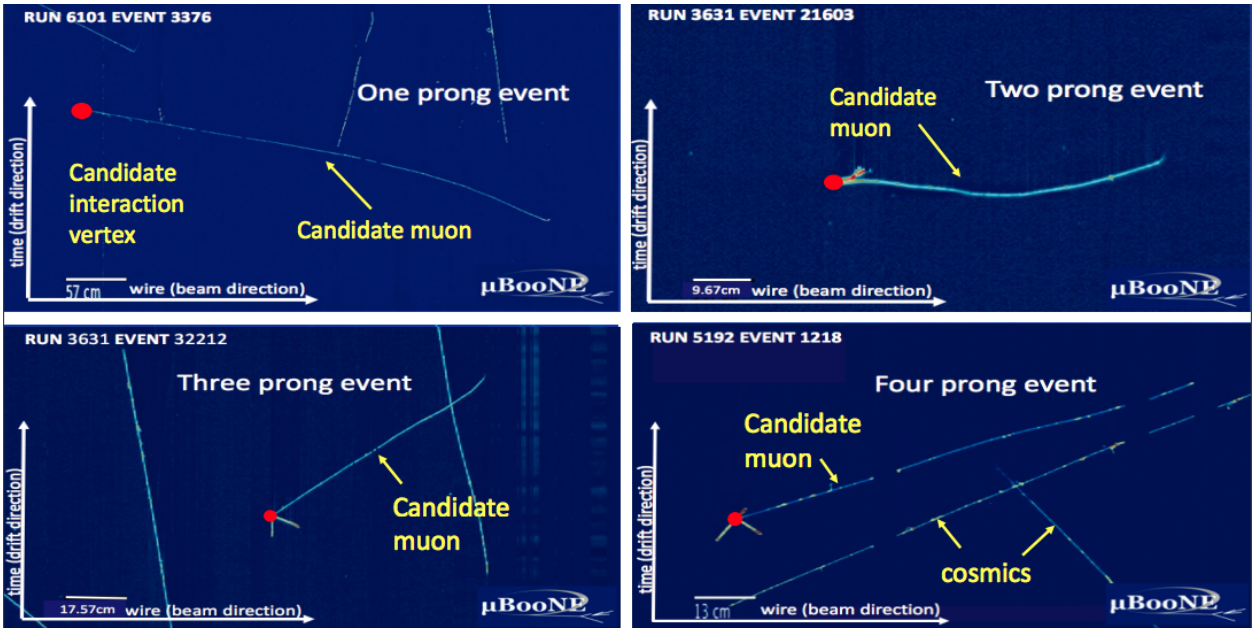


Figure 7.1: Four different multiplicity (prong) candidate neutrino events from MicroBooNE data. Red dots represent neutrino interaction vertices in each event. Figures from Ref. [53].

portability. One must have access to the full MicroBooNE simulation suite to use the O_n to test models.

7.2 Event Selection and Classification

7.2.1 Data

This analysis uses two data samples:

- "On-beam data", taken only during periods when a beam spill from the BNB is actually sent. The on-beam data used in this analysis were recorded from February to April 2016 using data taken in runs in which the BNB and all detector systems functioned well [64]. This sample comprises about 15% of the total neutrino data collected by MicroBooNE in its first running period (October 2015 to summer 2016),

- "Off-beam data" taken with the same software trigger settings as the on-beam data, but during periods when no beam was received. The off-beam data were collected from February

to October 2016.

7.2.2 Simulation

The LArSoft software framework is used for processing data events and Monte Carlo simulation (MC) events in the same way. Three simulation samples are used in this analysis:

- Neutrino interactions simulated with a default GENIE model overlaid with CORSIKA CR events ("MC default"),
- MC default augmented by the GENIE implementation of the Meson Exchange Current model overlaid with CORSIKA CR events ("MC with MEC"),
- MC default augmented by the GENIE implementation of the Transverse Enhancement Model overlaid with CORSIKA CR events ("MC with TEM").

The generator stage (production of a set of final state four-vectors for particles originating from the argon nucleus as a result of the ν_μ-Ar interaction in GENIE) employs GENIE (version $v2.8.6d$ for the MC default and $v2.10.6$ for the MC with MEC and TEM) with overlaid simulated CR backgrounds using CORSIKA version $v7.4003$ with a constant mass composition model (CMC) [65] at 226 m above sea-level elevation. Simulated secondary particle propagation utilizes GEANT version $v4.9.6.p04d$ using a physics particle list *QGSP_BIC* with custom physics list for optical photons, and the detector simulation employs LArSoft version $v4.36.00$. All GENIE samples were processed with the same GEANT and LArSoft versions for detector simulation and reconstruction. These samples thus allow for relative comparison of different GENIE models to the data.

7.2.3 Event Selection

Event selection starts by requiring an optical flash within the 1.6 μs duration beam spill window and the summed light collected by the PMT to exceed 50 PE. Reconstructed vertices must be contained in the fiducial volume of the detector, defined as 10 cm from the border of the active volume in x, 10 cm from the border of the active volume in z, and 20 cm from the border of the active

volume in y (see Fig. 4.2 for detector coordinates). At each candidate neutrino interaction vertex, a candidate muon track is identified as the longest of all tracks starting within 5 cm of the vertex. The candidate muon track is further required to be fully contained within the detector, where containment requires both ends of the track to lie within the same fiducial volume required for an event vertex, to have at least 75 cm 3D track length, and to have an event vertex located within 80 cm in z of the PMT-reconstructed position of an optical flash. Considerable CR backgrounds remain after these pre-selection procedures, with signal/background $\approx 1/1$.

Pre-selected events then pass through a second stage filter that imposes further quality conditions on track candidates. Start and end points of the candidate muon must lie in detector regions with functional collection plane wires. The candidate muon track must start 46 cm below the top surface of the TPC in order to suppress CR backgrounds, must start within 3 cm (reduced from 5 cm) of the selected vertex position, must have at least 80 hits in the collection wire plane, and must not have significant wire gaps in the start and end 20 collection plane-hit segments used in the pulse-height (PH) test (Sec. 7.2.4) and the multiple Coulomb scattering test (Sec. 7.2.4).

Events satisfying all of the above criteria comprise the final data sample. Table 7.1 lists the event passing rates for the on-beam data, off-beam data, and the MC default samples at different steps of the event selection. The passing rates in on-beam data are consistent with expectations for a mix of CR-only events, as provided by the off-beam data, and events containing a neutrino interaction in addition to cosmic rays, as provided by the MC.

The observed multiplicity of a selected event is defined to be the number of particles starting within 3 cm of the selected vertex that have at least 15 collection plane hits where the Pandora MicroBooNE track reconstruction algorithms perform optimally. There is no containment requirement for tracks other than the candidate muon track. Table 7.2 lists the number of selected events in each multiplicity bin with relative event rates for on-beam data, off-beam data, and MC default samples.

The minimum collection plane hit condition corresponds to a minimum range in liquid argon

of 4.5 cm, and the requirement thus excludes charged particles below a particle-type-dependent kinetic energy threshold from entering the sample that ranges from 31 MeV for a $\pi^{\pm}$ to 69 MeV for a proton. No acceptance exists for particles with kinetic energies below these thresholds, which roughly increases as the secant of the track angle with respect to the neutrino beam direction.

The average current Pandora-based track reconstruction efficiency in MicroBooNE is $\langle\varepsilon\rangle \approx 45\%$ (see Sec. 5.5) at the 15 collection plane hits threshold. This relatively low value, with implicit kinetic energy thresholds, creates a common occurrence called "feed-down" wherein events produced with n tracks at the argon nucleus exit position are reconstructed with an observed multiplicity $n' < n$. For example, $n = 1$ is commonly observed because one of the two tracks in a quasi-elastic-like event fails to be reconstructed due to low acceptance or tracking efficiency.

The candidate muon containment requirement limits its energy to be $\lesssim 1.2$ GeV depending on the muon scattering angle. This results in a sample biased towards relatively higher inelasticity, E_H/E_ν, with E_H being the energy transferred from the neutrino to the hadronic system in the collision.

Figure 7.2 shows the GENIE expectations for muons, protons, and pions in true kinetic energy. The kinetic energy thresholds associated with the 15 collection plane hit requirement for short tracks and the 75 cm 3D track length requirement for the long track are evident.

Table 7.1: Passing rates for event selection criteria applied to on-beam data, off-beam data, and MC default samples. Numbers are absolute event counts. Quantities in parentheses give the relative passing rate with respect to the step before (first percentage) and the absolute passing rate with respect to the starting sample (second percentage). Table from Ref. [21].

	On-beam data		**Off-beam data**		**MC default**	
Selection cuts	**Events**	**Passing rates**	**Events**	**Passing rates**	**Events**	**Passing rates**
Total events	547616		2954586		188880	
ν_μ events passing pre-cuts	4049	(0.74%/0.74%)	14213	(0.48%/0.48%)	7106	(3.8%/3.8%)
Events passing dead region cut	3080	(76%/0.56%)	10507	(74%/0.36%)	5632	(79%/2.9%)
Events with $y > 70$ cm	2438	(79%/0.44%)	7883	(75%/0.27%)	4795	(85%/2.6%)
long track to vertex distance < 3 cm	2435	(99%/0.44%)	7862	(99%/0.27%)	4781	(99%/2.5%)
Events with ≥ 80 CP hits	1930	(79%/0.35%)	5279	(67%/0.17%)	4387	(92%/2.3%)
Events passing wire gap cuts	1795	(93%/0.33%)	4954	(94%/0.16%)	4016	(92%/2.1%)

Table 7.2: Selected number of events from the on-beam data, off-beam data, and MC default samples and their corresponding acceptance rates on the multiplicity basis. Table from Ref. [21].

Multiplicities	**On-beam data**		**Off-beam data**		**MC default**	
	Events	**Event rate**	**Events**	**Event rate**	**Events**	**Event rate**
Total events	1795		4954		4016	
mult = 1	1379	77%	4113	83%	2599	65%
mult = 2	389	22%	828	17%	1186	30%
mult = 3	26	1.4%	12	0.2%	210	5%
mult = 4	1	0.06%	1	0.2%	18	0.4%
mult = 5	0	0%	0	0%	3	0.07%

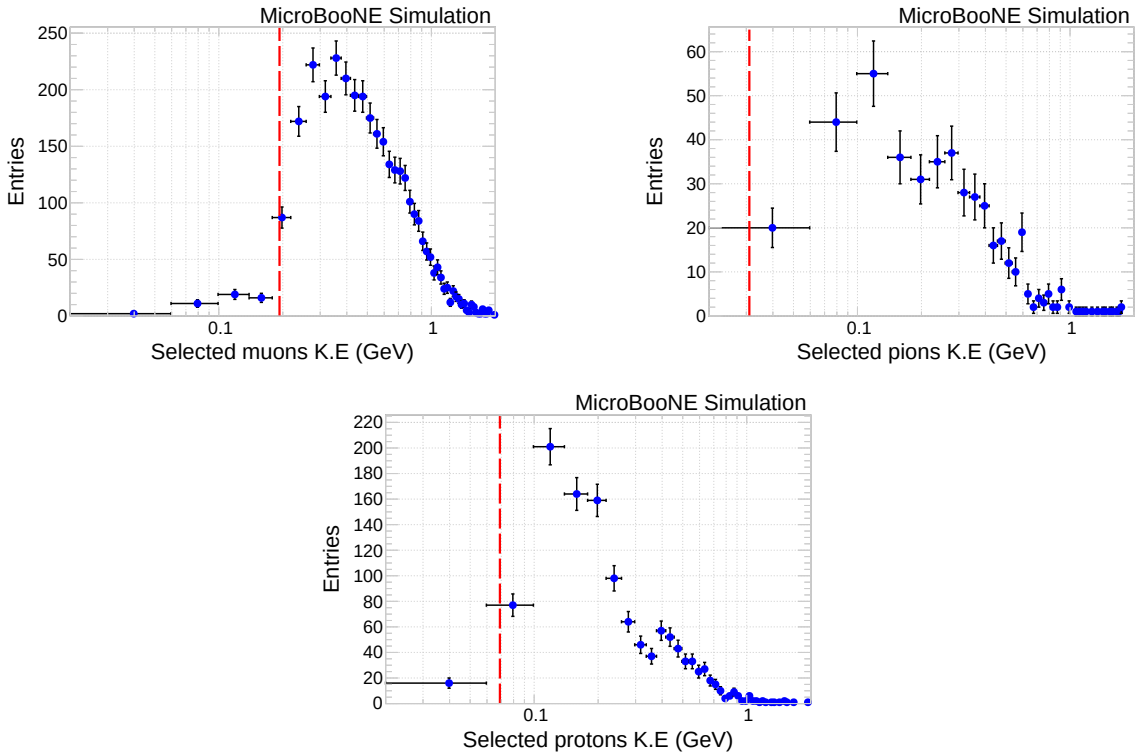


Figure 7.2: GENIE expectations for true kinetic energy distributions for selected muons (top left), pions (top right), and protons (bottom). The kinetic energy thresholds associated with the 15 collection plane hit requirement for short tracks and the 75 cm length requirement for the long track are represented by dashed red lines. Figures from Ref. [21].

7.2.4 Event Classification

Selected events are next classified into four categories based on whether they pass or fail the PH test and the MCS test described in the following sections. These are the candidate muon track direction-based tests which are used to separate neutrino signal and CR background contributions

in the sample. Table 7.3 lists the event selection rates for the on-beam data, off-beam data, and the MC default samples in each category. The final samples are called *neutrino-enriched*, *mixed*, or *background-enriched* sub-samples depending on whether events pass both tests, pass either one of the two tests, or fail both tests, respectively.

Table 7.3: Final categories from the on-beam data, off-beam data, and MC default samples. Numbers are absolute event counts. The percentages correspond to the fraction of events in each category. Table from Ref. [21].

Categories	**On-beam data**		**Off-beam data**		**MC default**	
PH, MCS	**Events**	**Event rate**	**Events**	**Event rate**	**Events**	**Event rate**
PASS, PASS	802	44%	1252	25%	2464	61%
PASS, FAIL	334	19%	1013	20%	704	18%
FAIL, PASS	304	17%	1049	21%	442	11%
FAIL, FAIL	355	20%	1640	33%	406	10%

Pulse Height Test

A neutrino-induced muon from a CC event will exhibit an increasing rate of energy loss as one moves downstream along its track. A visual diagram for the PH test is shown in Fig. 7.3. I take into account the expected behavior of the rate of restricted energy loss [66], dE/dx_R, with the following procedure:

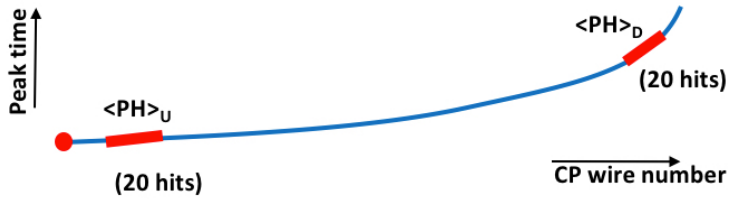


Figure 7.3: Diagram showing PH test for a candidate muon track. Figure from Ref. [21].

- Compute the truncated mean of the pulse heights deposited in 20 consecutive collection plane hits, $\langle PH \rangle_U$, starting 10 hits away from the upstream end of the muon track that is taken as a proxy for the upstream restricted energy loss. The truncated mean is formed by taking the average of the 20 PH after removing individual PH that do not lie within the range

of $20\% - 200\%$ of the average [67]:

$$\langle PH \rangle_U = \frac{\sum_{n=11}^{n=30} PH_n \left(0.2 \langle PH \rangle < PH_n < 2.0 \langle PH \rangle\right)}{\sum_{n=11}^{n=30} \left(0.2 \langle PH \rangle < PH_n < 2.0 \langle PH \rangle\right)}, \tag{7.4}$$

which can be determined iteratively with an initial approximation that $\langle PH \rangle$ is the arithmetic average. Use of the truncated mean PH rather than the average PH minimizes effects of large energy loss fluctuations,

- Form a similar quantity from 20 consecutive collection plane hits that end 10 collection plane hits away from the downstream end of the track, $\langle PH \rangle_D$,

- Form the test $p = \langle PH \rangle_U < \langle PH \rangle_D$. Muons from ν_μ CC interactions will pass this test with a probability $P(PH)$. Muons from CR background can be characterized by the probability that they pass the negation of the test $\bar{p} = \langle PH \rangle_U > \langle PH \rangle_D$, denoted as $Q(PH)$.

Figure 7.4 presents the PH downstream to upstream ratio distribution for neutrino events only from MC default (signal MC) and off-beam data (cosmic data) samples. The expected signal is considerably enriched relative to the background for PH ratios greater than 1 and I use this value to define the PH test used in the analysis.

Multiple Coulomb Scattering Test

A neutrino-induced muon from a CC event will generally exhibit an increasing degree of multiple Coulomb scattering (MCS) about a nominal straight line trajectory as one moves from upstream to downstream along the track. A visual diagram for the MCS test is shown in Fig. 7.5. I take into account the expected MCS behavior by an independent test with the following procedure:

- Form three 20 collection plane-hit long track segments at the upstream, downstream, and geometric center of the track. The upstream and downstream segments are displaced by 10 collection plane hits from the upstream and downstream ends of the track, respectively,

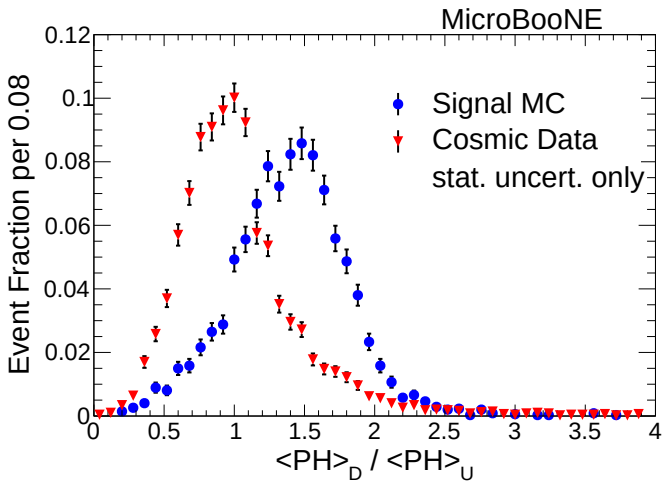


Figure 7.4: Pulse height (PH) downstream to upstream ratio. Events with $\frac{\langle PH \rangle_D}{\langle PH \rangle_U} > 1$ pass the PH test. Figure from Ref. [21].

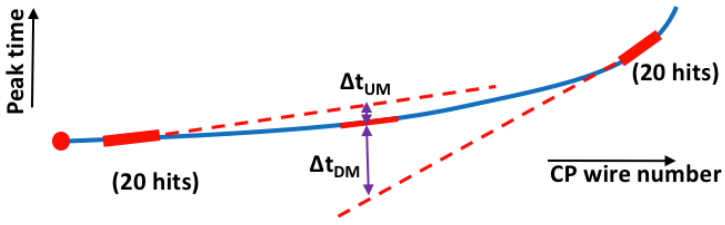


Figure 7.5: Diagram of the MCS test for a candidate muon track. Figure from Ref. [21].

- Perform a simple linear least squares fit of hit time vs. (wire) position using the 20 contiguous collection plane hits at the upstream end of the track. Denote the determined line as L_U. Perform a similar fit using the 20 collection plane hits at the downstream end of the track. Denote the determined line as L_D. Finally perform one more similar fit from the 20 collection plane hits located about the geometric center of the track. Denote this line as L_M,

- Compare the hit time predicted at the geometric center of the track, t_C, by L_M, which uses hits about the geometric center, to the time predicted at the geometric center of the track by the projection of L_U from the beginning of the track:

$$\Delta t_{UM} = |t_C(L_U) - t_C(L_M)|. \tag{7.5}$$

- Repeat the process except compare t_C from L_M to the time predicted at the geometric center of the track by the projection of L_D from the end of the track:

$$\Delta t_{DM} = |t_C(L_D) - t_C(L_M)|. \tag{7.6}$$

- Form the test $q = \Delta t_{UM} < \Delta t_{DM}$. Since MCS should become, on average, more pronounced along the downstream end of the track as the momentum decreases, this provides a second directional test on the muon track candidate. Muons from ν_μ CC interactions will pass this test with a probability $P(MCS)$. Muons from CR background tests can be characterized by the probability that they pass the negation of the test $\bar{q} = \Delta t_{UM} > \Delta t_{DM}$, denoted as $Q(MCS)$.

Figure 7.6 presents the MCS downstream to upstream ratio $\Delta t_{DM}/\Delta t_{UM}$ distribution for neutrino events only from MC default (signal MC) and off-beam data (cosmic data) samples. We observe that MCS ratio for the signal dominates over the background for values greater than 1 and I use this value to define the MCS test used in this analysis.

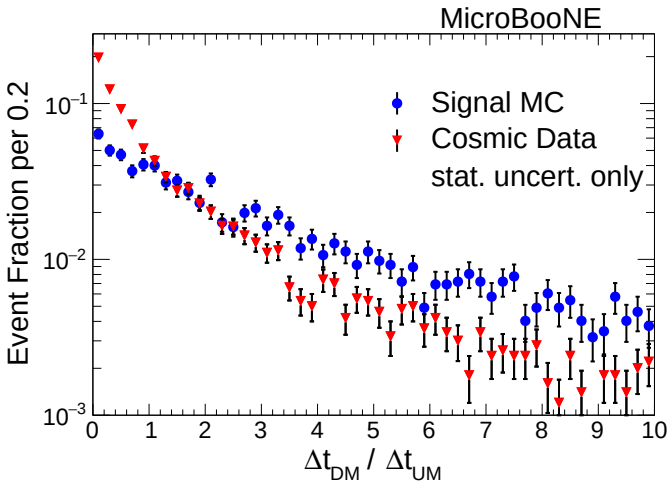


Figure 7.6: Multiple Coulomb scattering (MCS) downstream to upstream ratio. Events with $\frac{\Delta t_{DM}}{\Delta t_{UM}} > 1$ pass the MCS test. Figure from Ref. [21].

7.3 Signal Extraction

7.3.1 Data-Driven Signal+Background Model

On-beam data consists of a mixture of neutrino interaction and CR background events. I designate a passing of the PH or MCS test by the symbol ν, and a failure of either of the two tests by the symbol *CR*, thus creating the categories "$\nu\nu$", "νCR", "$CR\nu$", "$CRCR$", which contain a corresponding number of events $N_{\nu\nu}$, $N_{\nu CR}$, $N_{CR\nu}$, and N_{CRCR}. Here the first symbol corresponds to the failure/passing of the MCS test and the second corresponds to the failure/passing of the PH test. The "$\nu\nu$" and "$CRCR$" categories are expected to have relatively high neutrino or CR purity, respectively; while the "νCR" and "$CR\nu$" have mixed purity.

I next build a model for the number of events in each category as follows:

$$\hat{N}_{\nu\nu} = P(MCS|PH)\,P(PH)\,\hat{N}_{\nu} \tag{7.7}$$
$$+ (1 - Q(PH) - Q(MCS) + Q(MCS|PH)\,Q(PH))\,\hat{N}_{CR},$$
$$\hat{N}_{CR\nu} = (1 - P(MCS|PH))\,P(PH)\,\hat{N}_{\nu} \tag{7.8}$$
$$+ (Q(MCS) - Q(MCS|PH)\,Q(PH))\,\hat{N}_{CR},$$
$$\hat{N}_{\nu CR} = (P(MCS) - P(MCS|PH)\,P(PH))\,\hat{N}_{\nu} \tag{7.9}$$
$$+ (1 - Q(MCS|PH))\,Q(PH)\,\hat{N}_{CR},$$
$$\hat{N}_{CRCR} = Q(MCS|PH)\,Q(PH)\,\hat{N}_{CR} \tag{7.10}$$
$$+ (1 - P(PH) - P(MCS) + P(MCS|PH)\,P(PH))\,\hat{N}_{\nu}.$$
$$\tag{7.11}$$

The quantities $\hat{N}_{\nu\nu}$, $\hat{N}_{CR\nu}$, $\hat{N}_{\nu CR}$, and $\hat{N}_{CRCR}$ are model parameters corresponding to the observed number of events $N_{\nu\nu}$, $N_{\nu CR}$, $N_{CR\nu}$, and N_{CRCR}, respectively. $\hat{N}_{\nu}$ and $\hat{N}_{CR}$ are the estimated number of neutrino and CR events, in the sample, to be determined by a fit described below. The quantities $P(PH)$ and $P(MCS)$ represent the average probabilities that a neutrino interaction muon passes the *PH* or *MCS* test condition, while $Q(PH)$ and $Q(MCS)$ denote the mean probabilities that a cosmic ray muon *fails* one of these tests. The conditional probability $P(MCS|PH)$ denotes the fraction of time that a neutrino interaction muon event that passes the *MCS* condition after it has passed the *PH* condition, and the conditional probability $Q(MCS|PH)$ denotes the fraction of time that a cosmic ray event muon fails the *MCS* test after failing the *PH* test.

As the *MCS* and *PH* conditions result from different physical processes (muon-nucleus and muon-electron scattering, respectively), and the *MCS* and *PH* test are formed from different measurements (time and charge, respectively), the *PH* and *MCS* tests are nearly independent with $P(MCS|PH) \approx P(MCS)$ and $Q(MCS|PH) \approx Q(MCS)$. In the analysis I find evidence for weak, but non-negligible, correlations between the tests, and use the conditional probabilities to take these into account.

I collect data and construct a similar model for off-beam data, which contains no neutrino content, dividing the events into the same categories as above, and fitting the observed number of events in each category, $N'_{\nu\nu}$, $N'_{\nu CR}$, $N'_{CR\nu}$, and N'_{CRCR} to the parameterizations:

$$\hat{N}'_{\nu\nu} = (1 - Q(PH) - Q(MCS) + Q(MCS|PH)\,Q(PH))\,\hat{N}'_{CR}, \tag{7.12}$$

$$\hat{N}'_{CR\nu} = (Q(MCS) - Q(MCS|PH)\,Q(PH))\,\hat{N}'_{CR}, \tag{7.13}$$

$$\hat{N}'_{\nu CR} = (1 - Q(MCS|PH))\,Q(PH)\,\hat{N}'_{CR}, \tag{7.14}$$

$$\hat{N}'_{CRCR} = Q(MCS|PH)\,Q(PH)\,\hat{N}'_{CR}. \tag{7.15}$$

In this case the $\nu\nu$ and $CRCR$ categories are expected to be enriched samples containing muons characteristic of neutrino interactions and cosmic rays, respectively, while the $CR\nu$ and νCR samples have a mixed composition. $\hat{N}'_{CR}$ is the estimated CR content of the sample (in practice the number of events in the sample).

This algorithm uses the eight categories of events in on-beam and off-beam data to estimate the neutrino content in each multiplicity bin. To calculate the MC distributions, I replace the on-beam data with the MC samples and perform the fit again. The same off-beam data sample was used in both fits. In the absence of correlations, the quantities $\hat{N}_\nu$, $\hat{N}_{CR}$, $\hat{N}'_{CR}$, $P(PH)$, $P(MCS)$, $Q(PH)$, and $Q(MCS)$ can be directly determined from the data with no model inputs. The addition of the two conditional probabilities $P(MCS|PH)$ and $Q(MCS|PH)$ requires use of a model to determine the correlation between the PH and MCS tests. These correlations are implemented through the parameterizations

$$P(MCS|PH) = \frac{\alpha_\nu P(MCS)}{1 + (\alpha_\nu - 1)\,P(MCS)} \text{ and} \tag{7.16}$$

$$Q(MCS|PH) = \frac{\alpha_{\mathrm{CR}} Q(MCS)}{1 + (\alpha_{\mathrm{CR}} - 1)\,Q(MCS)}. \tag{7.17}$$

The two new parameters α_ν and α_{CR} are obtained from Monte Carlo simulation of neutrino data and from the off-beam data, respectively. In order to have minimum MC dependence, we chose to fix α_ν and α_{CR} (ratios of probabilities) from the pure BNB events and the off-beam data samples

that remain fixed throughout the whole fitting process while the other probabilities change with different data samples. Maximum likelihood fit was first performed on pure BNB MC events by keeping the cosmic parameters ($Q(PH)$, $Q(MCS)$, $Q(MCS|PH)$, and α_{CR}) fixed for all categories. The purpose was to extract the value of α_ν and keep it fixed later on. Similarly, the maximum likelihood fit was then performed on off-beam data sample by keeping the neutrino parameters ($P(PH)$, $P(MCS)$, $P(MCS|PH)$, and α_ν) fixed for all categories to extract the value of α_{CR} which was then kept fixed in the subsequent fits. If $\alpha_\nu = 1$, no correlation would exist between the tests, whereas a large α_ν would imply near total correlation, with similar conditions applied to α_{CR}.

7.3.2 Fitting Procedure

I construct a likelihood function based on the probability distribution for partitioning events into one of four categories of a multinomial distribution, for both on-beam and off-beam data. The multinomial probability of observing n_i events in bin i, with $i = 1,2,3,4$, with the probability of a single event landing in bin i equal to r_i is

$$M(n_1,n_2,n_3,n_4;r_1,r_2,r_3,r_4) = \frac{(n_1+n_2+n_3+n_4)!}{n_1!n_2!n_3!n_4!} r_1^{n_1} r_2^{n_2} r_3^{n_3} r_4^{n_4}. \tag{7.18}$$

The n_i are the observed number of events in each bin, and the r_i are functions of the model parameters.

The likelihood also incorporates the Poisson statistics of observing $n_1+n_2+n_3+n_4$ in both the on-beam and off-beam data:

$$P_{on\text{-}beam} = \frac{\hat{N}^N}{N!} e^{-\hat{N}}, \tag{7.19}$$

$$P_{off\text{-}beam} = \frac{\hat{N}'^{N'}}{N'!} e^{-\hat{N}'}, \tag{7.20}$$

with

$$\hat{N} = \hat{N}_{\nu\nu} + \hat{N}_{CR\nu} + \hat{N}_{\nu CR} + \hat{N}_{CRCR}, \tag{7.21}$$

$$\hat{N}' = \hat{N}'_{\nu\nu} + \hat{N}'_{CR\nu} + \hat{N}'_{\nu CR} + \hat{N}'_{CRCR}, \tag{7.22}$$

$$N = N_{\nu\nu} + N_{CR\nu} + N_{\nu CR} + N_{CRCR}, \tag{7.23}$$

$$N' = N'_{\nu\nu} + N'_{CR\nu} + N'_{\nu CR} + N'_{CRCR}. \tag{7.24}$$

The final likelihood function is

$$\begin{aligned} L_{TOT} = {} & M_{on\text{-}beam}\left(N_{\nu\nu}, N_{CR\nu}, N_{\nu CR}, N_{CRCR}; \frac{\hat{N}_{\nu\nu}}{\hat{N}}, \frac{\hat{N}_{CR\nu}}{\hat{N}}, \frac{\hat{N}_{\nu CR}}{\hat{N}}, \frac{\hat{N}_{CRCR}}{\hat{N}}\right) \\ & \times M_{off\text{-}beam}\left(N'_{\nu\nu}, N'_{CR\nu}, N'_{\nu CR}, N'_{CRCR}; \frac{\hat{N}'_{\nu\nu}}{\hat{N}'}, \frac{\hat{N}'_{CR\nu}}{\hat{N}'}, \frac{\hat{N}'_{\nu CR}}{\hat{N}'}, \frac{N'_{CRCR}}{\hat{N}'}\right) \\ & \times \frac{\hat{N}^N}{N!} e^{-\hat{N}} \times \frac{\hat{N}'^{N'}}{N'!} e^{-\hat{N}'}. \end{aligned} \tag{7.25}$$

The model parameters $\hat{N}_\nu$, $\hat{N}_{CR}$, $\hat{N}'_{CR}$, $P(PH)$, $P(MCS)$, $Q(PH)$, and $Q(MCS)$ and their statistical uncertainties are estimated via the maximum likelihood method, implemented by minimizing the negative-log-likelihood

$$\mathscr{L}_{TOT} = -\ln L_{TOT}, \tag{7.26}$$

using the MIGRAD minimization in the standard MINUIT [68] package in ROOT [69].

The fitting procedure can be used to obtain estimates for $\hat{N}_\nu$, $\hat{N}_{CR}$, $\hat{N}'_{CR}$, $P(PH)$, $P(MCS)$, $Q(PH)$, and $Q(MCS)$ for each multiplicity. When the probability parameters $P(PH)$, $P(MCS)$, $Q(PH)$, and $Q(MCS)$ are consistent between multiplicities, I use all multiplicities together in their determination for improved statistical precision and vary only the three parameters $\hat{N}_\nu$, $\hat{N}_{CR}$, and $\hat{N}'_{CR}$ for each individual multiplicity.

7.3.3 Results with Simulated Events

Maximum likelihood fits were performed on all three GENIE simulation samples to extract the values of seven parameters $\hat{N}_\nu$, $\hat{N}_{CR}$, $\hat{N}'_{CR}$, $P(PH)$, $Q(PH)$, $P(MCS)$, and $Q(MCS)$). Parameters α_ν and α_{CR} were extracted from MC and off-beam data samples and kept fixed for the subsequent fits. As expected, the *PH* and *MCS* probabilities show no statistically significant difference between the three GENIE models considered. Table 7.4 lists the values obtained from the fit for the above-mentioned parameters in the default MC and the MicroBooNE data

The number of neutrino events in the simulated data samples were extracted for each observed multiplicity and compared to the known number from the event generation. Table 7.5 and Fig. 7.7 summarize this comparison. I find that the fit results agree within statistics with the known inputs, indicating a lack of bias in this signal estimation technique. I have also verified that this method is insensitive to the signal-to-background ratio of the sample over a range corresponding to 0.2 – 5.0 times that estimated in the data.

Table 7.4: Fit parameter results and corresponding uncertainties for the default MC and data samples. The same off-beam data sample was used in both fits. All uncertainties are from the fit and are purely statistical. Table from Ref. [21].

	Fit results	
Parameters	**Default MC**	**Data**
$\hat{N}_\nu$	3405±159	1023±170
$\hat{N}_{CR}$	611±150	782±169
$\hat{N}'_{CR}$	5002±71	5002±71
$P(PH)$	0.848±0.018	0.766±0.050
$P(MCS)$	0.770±0.0123	0.730± 0.039
$Q(PH)$	0.542±0.007	0.552±0.007
$Q(MCS)$	0.537±0.007	0.534± 0.007
α_ν	1.32±0.05	1.32±0.05
α_{CR}	1.36±0.04	1.36±0.04

Table 7.5: Fitted and true number of neutrino events for the MC default sample for different multiplicity bins. The last column shows good agreement between the fit results and true content for different bins. Table from Ref. [21].

Multiplicities	**Fit N_ν**	**True N_ν**	**True-Fit χ^2/ndf**
1	2070±63	2152	1.7
2	1112±44	1092	0.2
3	210±14	208	0.0
4	18±4	18	0.0
5	3±2	3	0.0

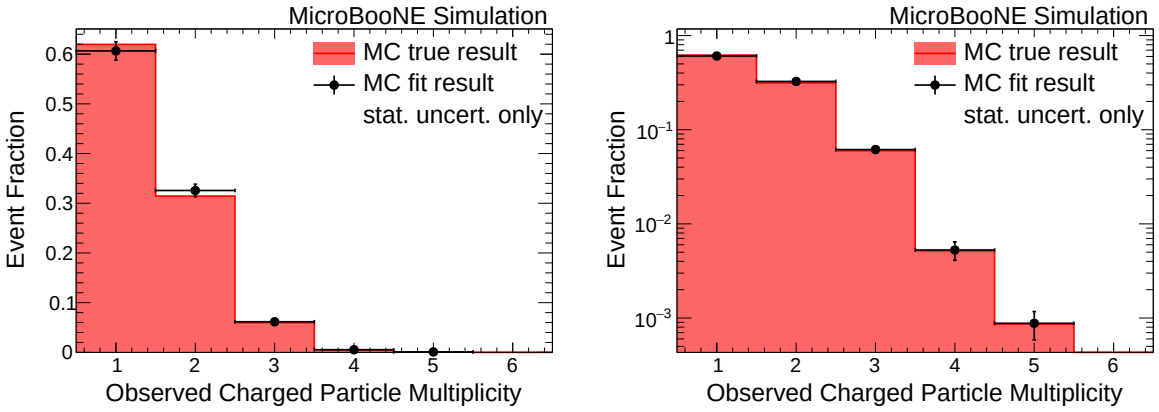


Figure 7.7: Overlaid true and fitted observed neutrino multiplicity distributions from the MC default sample in linear scale (left) and in log y scale (right). Figures from Ref. [21].

7.4 Statistical and Systematic Uncertainty Estimates

Table 7.6 presents the percentage estimates for statistical and systematic uncertainties from different sources. Figure 7.8 presents a plot of each uncertainty source as a function of observed multiplicity.

7.4.1 Statistical Uncertainties

Statistical uncertainties are returned from the MINUIT package used in our fitting for both data and MC samples. These uncertainties include contributions from the CR background in this fitting procedure, and this procedure includes the CR background systematic uncertainty in a contribution to the total statistical uncertainty. Both data and MC statistics contribute substantially to the overall

Table 7.6: Statistical and systematic uncertainties estimates from data and MC. Table from Ref. [21].

	Uncertainty Estimates			
Uncertainty Sources	**mult=1**	**mult=2**	**mult=3**	**mult=4**
Data statistics	4%	10%	20%	99%
MC statistics	2%	3%	7%	22%
Short track efficiency	7%	11%	25%	33%
Long track efficiency	1%	2%	4%	7%
Background model systematics	2%	2%	0%	0%
Flux shape systematics	0%	0.4%	0.2%	0.5%
Electron lifetime systematics	0.5%	0.1%	6%	5%

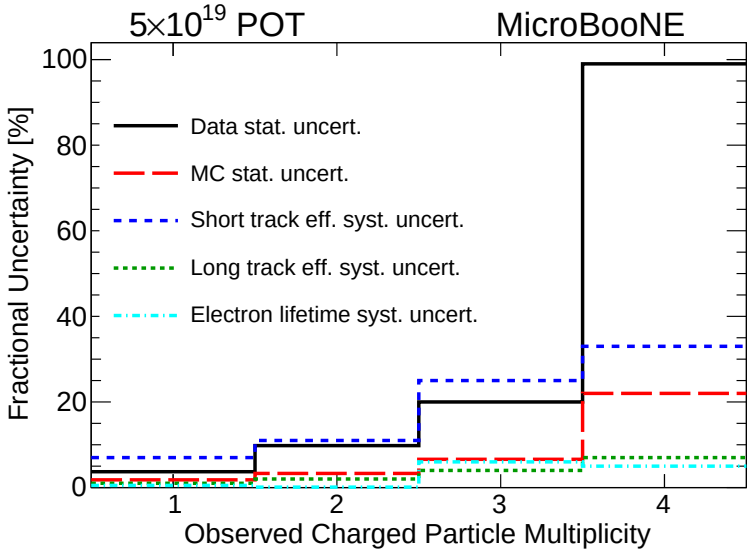


Figure 7.8: Percentage uncertainty distributions from different systematic and statistical sources as a function of observed charged particle multiplicity. Figure from Ref. [21].

uncertainties in the data, as shown in Fig. 7.8.

7.4.2 Short Track Efficiency Uncertainties

The dominant systematic uncertainty originates from the differences in the efficiency between data and simulation for reconstructing short-length hadron tracks. The overall efficiencies of the Pandora reconstruction algorithms [51] are a strong function of the number of hits of the tracks, with a plateau not being reached until of order of several hundred hits. The inclusive efficiencies

for reconstructing protons or pions at the 15 collection plane hit threshold is estimated to be $\langle \varepsilon \rangle = 0.45 \pm 0.05$. The absolute efficiency value is not used in this analysis, but I use this estimate to conservatively assign a mean efficiency uncertainty of $\delta = 15\%$.

I then estimate the effect of an efficiency uncertainty on multiplicity by the following procedure: Consider a track in an event in a multiplicity bin N. If one lowers the tracking efficiency by the factor $1-\delta$, then there is a $1-\delta$ probability that the track reconstructed and the event stayed in that multiplicity bin, and a probability δ that the track would not have been reconstructed and that the event would thus have a *lower* multiplicity. If the overall multiplicity is N, with $N-1$ short tracks and one long track, and each track's reconstruction probability is reduced by a factor $1-\delta$, then an overall fraction of events $(1-\delta)^{N-1}$ will remain in the bin, and a fraction $1-(1-\delta)^{N-1}$ will migrate to lower multiplicity bins. The fraction of tracks that migrate to multiplicity $N' < N$ from bin N, $f(N';N,\delta)$, is given by binomial statistics:

$$f\left(N';N,\delta\right) = \frac{(N-1)!}{(N'-1)!\,(N-N')!}\,(1-\delta)^{N'-1}\,\delta^{N-N'}. \tag{7.27}$$

I use this result to generate the expected observed CPMD in simulation that would emerge from lowering the tracking efficiency by the factor $1-\delta$ compared to the default simulated CPMD. The difference between the two distributions is then taken as the systematic uncertainty assigned to short track efficiency, with the assumption that the effect of increasing the default efficiency by a factor $1+\delta$ would produce a symmetric change. Table 7.7 summarizes this study for the three GENIE models used. The observed multiplicity $= 1$ probability increases because of "feed down" of events from higher multiplicity, due to the lowered efficiency, mainly from observed multiplicity $= 2$. The other observed multiplicity probabilities decrease accordingly. The largest effects are in high multiplicity bins because the loss of events from lowering the efficiency by the factor $(1-\delta)$ varies as $(1-\delta)^{N-1}$ for multiplicity bin N. Monte Carlo simulations show that "fake tracks" that could move events to higher multiplicity are rare.

Table 7.7: Relative change in observed multiplicity probabilities corresponding to a -15% uniform reduction in short charged particle tracking efficiencies for three GENIE models: default, MEC, and TEM. The missing entry for observed multiplicity 5 in TEM is due to no event being generated with that observed multiplicity. Table from Ref. [21].

Observed multiplicity	$\frac{\Delta P_n}{P_n}$ **Default**	$\frac{\Delta P_n}{P_n}$ **MEC**	$\frac{\Delta P_n}{P_n}$ **TEM**
1	$+7\%$	$+7\%$	$+8\%$
2	-11%	-12%	-12%
3	-25%	-25%	-25%
4	-33%	-36%	-39%
5	-44%	-48%	–

7.4.3 Long Track Efficiency Uncertainties

To first order, the efficiency for reconstructing tracks with length > 75 cm is not expected to affect the observed multiplicity distribution, as it is common to all multiplicities and cancels in the ratio when forming observed multiplicity probabilities. At second order, however, a multiplicity dependence that changes the distribution of observed multiplicity without affecting the overall number of events is possible. A plausible model for this is that higher multiplicity in an event helps Pandora better define a vertex, and thus increases the chance that the event passes the ν_μ CC selection filter.

I estimate the size of this effect by comparing the efficiencies obtained with the Pandora package for simulated quasi-elastic final states in which both the proton and muon are reconstructed, to charged pion resonance final states in which the proton, pion, and muon are all reconstructed. From this study I conclude that the efficiency for finding the muon in final states where all charged particles are reconstructed could be up to 3% higher for charged pion resonance events (observed multiplicity 3) than quasi-elastic events (observed multiplicity 2). I then assume, for the purpose of uncertainty estimation, that this relative enhancement seen for higher observed multiplicity events in the MC is absent in the data.

Table 7.8 summarizes this study. Effects are generally small compared to those seen in Table 7.7. No dependence on GENIE variant is found.

Table 7.8: Relative change in observed multiplicity probabilities corresponding to increasing the conditional probability for reconstructing the long track by 3% for each additional track found in the event, as suggested by Pandora studies of QE and charged pion resonance production for three GENIE models: default, MEC, and TEM. The missing entry for observed multiplicity 5 in TEM is due to no event being generated with that observed multiplicity. Table from Ref. [21].

Observed multiplicity	$\frac{\Delta P_n}{P_n}$ **Default**	$\frac{\Delta P_n}{P_n}$ **MEC**	$\frac{\Delta P_n}{P_n}$ **TEM**
1	-1%	-1%	-1%
2	$+2\%$	$+2\%$	$+2\%$
3	$+4\%$	$+4\%$	$+2\%$
4	$+7\%$	$+7\%$	$+7\%$
5	$+9\%$	$+9\%$	–

7.4.4 Background Model Uncertainties

In the signal extraction fitting procedure, two conditional parameters (α_ν and α_{CR}) were extracted from the Monte Carlo simulation and off-beam data. To calculate the systematic uncertainties on these parameters, their values were varied by $\pm 1\sigma$ of their statistical uncertainty. Those values were propagated in the observed charged particle multiplicity distribution. We also extracted the α_ν and α_{CR} values separately from the GENIE default, GENIE+TEM, and GENIE+MEC models. The effect from this systematic variation were found to be very small.

7.4.5 Flux Shape Uncertainties

Variations in flux can be parameterized by

$$\Phi(E_\nu) \rightarrow (1+\Delta(E_\nu))\,\Phi(E_\nu)\,, \tag{7.28}$$

where $\Phi(E_\nu)$ is the neutrino flux at neutrino energy E_ν and $\Delta(E_\nu)$ is the fractional uncertainty in the flux at that energy. An energy-independent $\Delta(E_\nu)$ has no effect on the observed multiplicity distributions as this measurement is independent of absolute normalization. On the other hand, raising the high energy flux relative to the low energy flux could enhance the contributions of higher multiplicity resonance and DIS processes. I consider only highly correlated energy-dependent shifts, denoted as $\Delta_i(E_\nu)$ for $i = 1-6$ via an approximate procedure that should be conservative. These shifts, shown in Fig. 7.9, are allowed to modify the BNB flux within uncertainties deter-

mined by the MiniBooNE collaboration [20]. The first two variations simply shift all flux values up ($\Delta_1(E_\nu)$) or down ($\Delta_2(E_\nu)$) according to the flux uncertainty envelope. The next two enhance the high energy flux ($\Delta_3(E_\nu)$) or low energy flux ($\Delta_4(E_\nu)$) linearly with neutrino energy, with the variation taken to be zero at the average neutrino energy. The final two variations enhance high energy flux ($\Delta_5(E_\nu)$) or low energy flux ($\Delta_6(E_\nu)$) logarithmically with neutrino energy, with the variation taken to be zero at the average energy. As expected, shifts that are positively correlated across all energies produce negligible differences, but even shifts that produce sizable distortions between high and low energies contribute a systematic uncertainty contribution that are small.

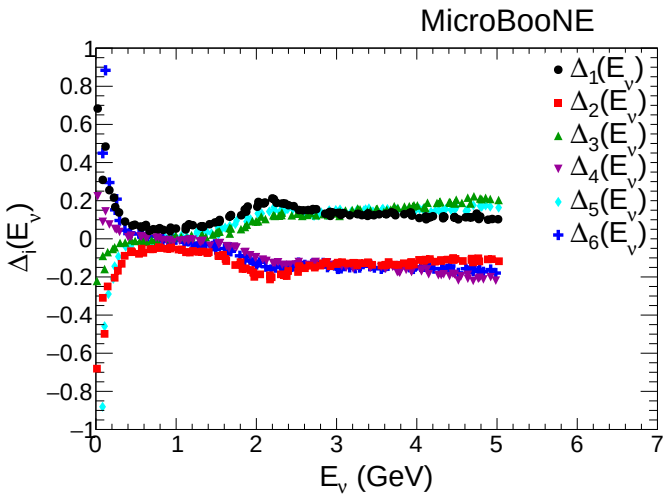


Figure 7.9: Beam flux shifts for the parameterizations $\Delta_i(E_\nu)$, $i = 1-6$. The variations $\Delta_1(E_\nu)$ and $\Delta_2(E_\nu)$ define the envelope of flux uncertainties for the BNB. Figure from Ref. [21].

7.4.6 Electron Lifetime Uncertainties

The measured charge from muon-induced ionization can vary within the detector volume due to the finite probability for drifting electrons to be captured by electronegative contaminants in the liquid argon. This capture probability can be parameterized by an electron lifetime τ. I perform this analysis on simulation with two lifetimes that safely bound those measured during detector operating conditions, $\tau = 6$ msec and $\tau = \infty$ msec. The resulting distribution of percentage uncertainty as a function of multiplicity in Fig. 7.8 shows that the electron lifetime uncertainties minimally affect the multiplicity.

7.4.7 Other Sources of Uncertainty Considered

A systematic comparison was performed on all kinematic quantities entering this analysis between off-beam CR data and the CR events simulated with CORSIKA. No statistically significant discrepancies were observed between event selection pass rates applied to off-beam data and MC simulation.

A check of possible time-dependent detector response systematics was also performed by dividing the data into two samples and performing the analysis separately for each sample. Differences between the two samples are consistent within statistical fluctuations.

The data are not corrected for ν_μ NC, ν_e, $\bar{\nu}_e$, or $\bar{\nu}_\mu$ backgrounds. An assumption is made that the Monte Carlo simulation adequately describes these non ν_μ CC backgrounds. Section 7.6.1 shows that these backgrounds, in total, are expected to be less than 10% of the final sample; their impact on the final distributions is generally small.

7.5 Results

7.5.1 Observed Charged Particle Multiplicity Distribution

Following the implementation of the signal extraction procedure and verification through closure test on MC events, I execute the same maximum likelihood fit on data. Table 7.4 lists the values of the fit parameters obtained for the data; and Table 7.9 lists the number of neutrino events in different multiplicity bins for the data. While this method does not require this to be the case, I note that the fitted PH and MCS test probabilities $P(PH)$, $Q(PH)$, $P(MCS)$, and $Q(MCS)$ agree in data and simulation within statistical uncertainties. This provides evidence that the simulation correctly describes the muon PH and MCS tests used in the analysis.

Area normalized, bin-by-bin fitted multiplicity distributions from three different GENIE predictions overlaid on data are presented in Fig. 7.10 where data error bars include statistical uncertainties obtained from the fit and the MC error bands include MC statistical and systematic uncertainties that are listed in Table 7.6 added in quadrature.

In general the three GENIE models agree within uncertainties with one another, and agree

qualitatively with the data. There are indications that GENIE overestimates the mean charged particle multiplicity relative to the data.

Table 7.9: Fitted number of neutrino events for the data sample in different multiplicity bins. The uncertainties correspond to the statistical uncertainty estimates obtained from the fit. The percentages correspond to the fraction of events in each category. Table from Ref. [21].

Multiplicities	**Fitted N_ν**
1	732±53
2	260±29
3	26±5
4	1±1
5	0±0

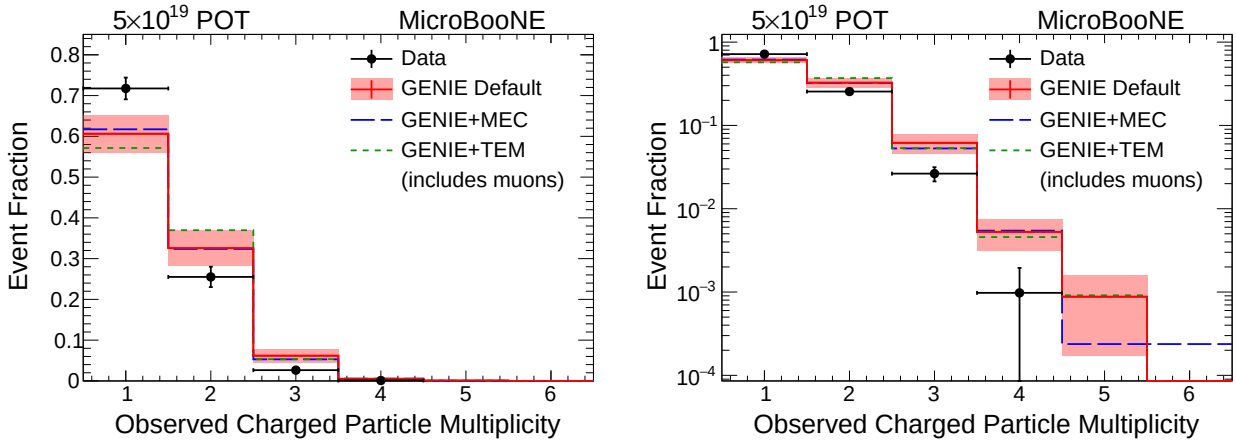


Figure 7.10: Bin-by-bin normalized multiplicity distributions using 5×10^{19} POT MicroBooNE data compared with three GENIE predictions (left) in linear scale, (right) in log scale. The data are CR background subtracted. Data error bars include statistical uncertainties obtained from the fit. Monte Carlo error bands include MC statistical uncertainties from the fit and systematic uncertainty contributions added in quadrature. Figures from Ref. [21].

7.5.2 Observed Kinematic Distributions

A key technical feature of this analysis entails performing tests on the pulse height and multiple Coulomb scattering behavior of hits on the long contained track in each event. This allows a categorization of events in each multiplicity into four categories according to whether the

long track passes or fails the PH and MCS tests: $(PASS, PASS)$, $(PASS, FAIL)$, $(FAIL, PASS)$, and $(FAIL, FAIL)$. I have shown that the $(PASS, PASS)$ category is "neutrino-enriched" and the $(FAIL, FAIL)$ category is cosmic-ray-dominated. The mixed cases $(PASS, FAIL)$ and $(FAIL, PASS)$ provide samples with intermediate signal-to-background ratios.

This fit to the distribution of the eight event categories in on-beam and off-beam data allows me to estimate the number of neutrino events $M_{\nu,i}(PH, MCS)$ and the number of corresponding background CR events $M_{CR,i}(PH, MCS)$ for each observed multiplicity i. Once $M_{\nu,n}(PH, MCS)$ and $M_{CR,n}(PH, MCS)$ are established, I can obtain a prediction for the content of any bin k of any kinematic quantity X_{ij} associated with track j in an observed multiplicity i event in any (PH, MCS) test combination:

$$\text{model}\left(X_{ij}, PH, MCS\right)_k = M_{\nu,i}(PH, MCS)\,\hat{x}_{\nu,ij}(PH, MCS)_k + M_{CR,i}(PH, MCS)\,\hat{x}_{CR,ij}(PH, MCS)_k . \tag{7.29}$$

Here $\hat{x}_{\nu,ij}(PH, MCS)_k$ is an area-normalized histogram of X_{ij} for "true neutrino events" in a given category obtained from a "MC" sample, and $\hat{x}_{CR,ij}(PH, MCS)_k$ is an area-normalized histogram of X_{ij} for CR events obtained from off-beam data. This distribution can be compared to the corresponding one for data in each category, $\text{data}\left(X_{ij}, PH, MCS\right)_k$.

In short, I assume that the observed distribution of events consists of a mix of neutrino events plus CR events. The proportions of the mix in each category are fixed by the output of this fit, which, by construction, constrains the normalization of the model to equal that of the data. We emphasize that only the PH and MCS tests have been used to extract the neutrino interaction signal sample; no information from any quantity X_{ij} is used.

7.5.3 Checks on Distributions lacking Dynamical Significance

Several kinematic properties of neutrino interactions depend only weakly on the neutrino interaction model; these include the reconstructed vertex positions, the initial and final coordinates of the long track, and the azimuthal angles of individual tracks. These distributions provide checks on the overall signal-to-background separation provided by the test-category fits and flux and detector

modeling. They also test for differences between the modeling of neutrino events, which depend on the GEANT detector simulation, and CR events, which use the off-beam data and thus do not depend on detector simulation.

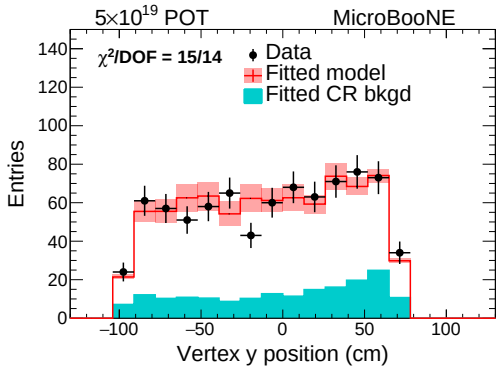
As an example, I show the observed distributions for the selected vertex y position for the candidate muon track from the full selected sample in Fig. 7.11. For this and all subsequent distributions, the on-beam data events are indicated by plotted points with statistical error bars. The model prediction is shown by a colored band (red for GENIE default, green for GENIE+TEM, and blue for GENIE+MEC) with the width of the band indicating the correlated statistical plus efficiency systematic uncertainty from using common $N_{\nu,n}, N_{CR,n}$ values for all bins of all distributions of a given multiplicity bin n. The CR contribution to a distribution in a given category is shown by the shaded cyan region. For example, Fig. 7.11 compares the on-beam data to GENIE default MC sample and also shows the CR background.

The signal-enriched $(PASS, PASS)$ category for vertex y has the nearly flat distribution expected for a neutrino event sample with a small CR background. Note that in this selection, I only allow candidate muon tracks initial y position < 70 cm. This cut rejects many cosmic rays that produce a downward trajectory in the final selected sample, thus the remaining background is dominated by cosmic rays with an apparent upward trajectory. This can be seen in the background-enriched sample $(FAIL, FAIL)$ in the vertex y distribution where a peak at negative y values corresponds to "upwards-going"CR.

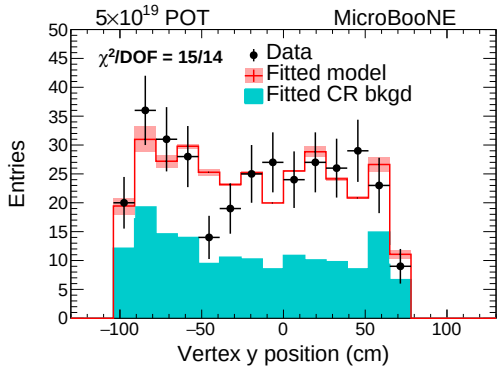
Figure 7.12 shows the distribution of azimuthal angle ϕ, defined in the plane perpendicular to the beam direction, of the muon candidate track for the full selected sample. The CR-dominated $(FAIL, FAIL)$ category shows the expected peaking at $\phi = \pm\pi/2$ from the mainly vertically-oriented CR. The asymmetry in the peak's structure is due to the requirement on vertex y position described previously in Sec. 7.2.3. By contrast the signal-enriched $(PASS, PASS)$ category has the nearly flat distribution expected for a neutrino event sample with a small CR background.

Similar levels of agreement exist between data and simulation for distributions of the event vertex x and z positions, for the (x, y, z) position of the end point of the muon track candidate, and for the azimuthal angles of individual tracks in multiplicity 2 and 3 topologies. I thus conclude that the simulation and reconstruction chain augmented by this method for estimated CR backgrounds

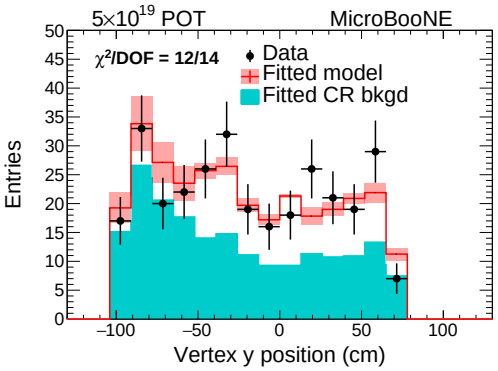
satisfactorily describes features of the data that have no dependence on the neutrino interaction model.



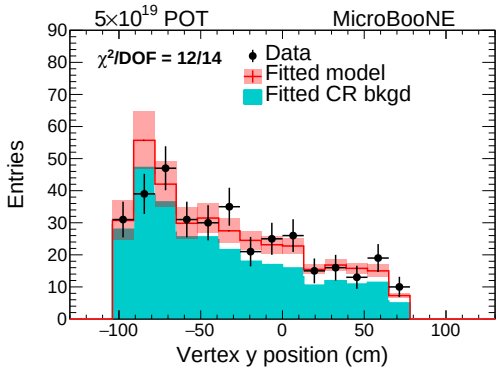
(a) Neutrino-enriched sample (PH pass, MCS pass)



(b) Mixed sample (PH pass, MCS fail)



(c) Mixed sample (PH fail, MCS pass)

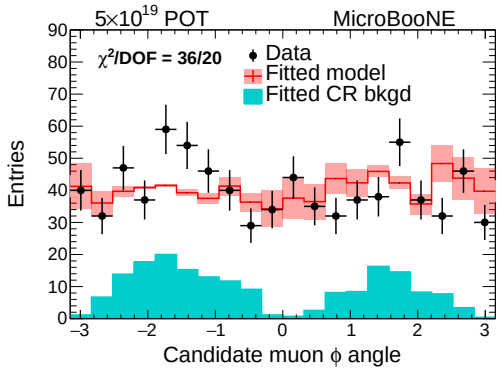


(d) Background-enriched sample (PH fail, MCS fail)

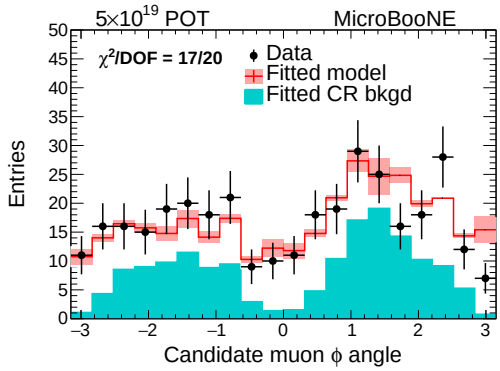
Figure 7.11: Neutrino interaction reconstructed vertex position along y-axis for data and GENIE default MC. Neutrino-enriched sample is nearly flat as expected. The asymmetry in the CR-background-enriched category corresponds to "upwards-going cosmics" which is a known feature of the selection. Figures from Ref. [21].

7.5.4 Dynamically Significant Distributions

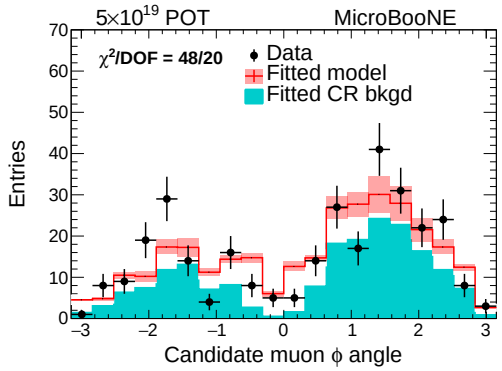
Events with N reconstructed tracks have potentially $4N$ dynamically significant variables—the components of each particle 4-vector—which will have distributions that depend on the neutrino interaction model. Azimuthal symmetry of the beam eliminates one of these, leaving $3, 7, 11$, and 15 dynamically significant variables for multiplicities $1-4$, respectively. In the following, I use the notation X_{ij} to label a dynamical variable x associated with track j in an observed multiplicity



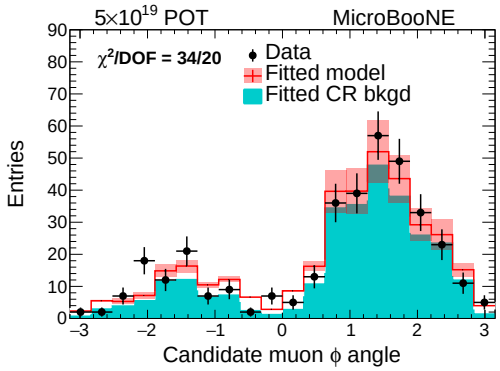
(a) Neutrino-enriched sample (PH pass, MCS pass)



(b) Mixed sample (PH pass, MCS fail)



(c) Mixed sample (PH fail, MCS pass)



(d) Background-enriched sample (PH fail, MCS fail)

Figure 7.12: Candidate muon azimuthal angle distribution from the full selected sample for data and GENIE default MC. The neutrino-enriched sample is nearly flat as expected. The CR-background-enriched sample has expected peaks at $\pm\pi/2$. Figures from Ref. [21].

i event. For example, $\cos\theta_{11}$ describes the cosine of polar angle distribution of the only track in multiplicity 1 events, while L_{22} would describe the length of the second (short) track in multiplicity 2 events. The notation with three subscripts, X_{ijk}, represents a distribution of the difference in variable x associated with tracks j and k in an observed multiplicity i event.

For one-track events, three variables exist. I use the observed length L_{11} as a proxy for kinetic energy, and the cosine of the scattering angle with respect to the neutrino beam direction $\cos\theta_{11}$. The azimuthal angle ϕ_{11} has no dynamical significance and must be uniformly distributed due to the cylindrical symmetry of the neutrino beam.

Since the particle mass is not determined in this analysis, I are free to introduce a third dynamically significant quantity that is sensitive to particle mass, which I take to be

$$\sin\Theta_{11} = |\hat{s}_{11} \times \hat{t}_{11}|\,, \tag{7.30}$$

where $\hat{s}_{11}$ is a unit vector parallel to the track direction at the event vertex, and $\hat{t}_{11}$ is a unit vector that points from the start of the track to the end of the track in the detector. The variable Θ_{ij} measures the angular deflection of a track over its length due to multiple Coulomb scattering. Its dependence on track momentum and energy differs from that of track length. For most of the MicroBooNE kinematic range, I expect light particles (π and μ) to scatter more, and thus produce a broader $\sin\Theta_{ij}$ distribution, than protons over the same track length.

Figure 7.13 shows the distributions of L_{11}, $\cos\theta_{11}$, and $\sin\Theta_{11}$, from the neutrino-enriched sample compared to the GENIE default model. Fig. 7.14 presents the L_{11} distribution for the GENIE+MEC and GENIE+TEM models. This is the distribution where the agreement between data and GENIE+TEM model, compared to the agreement between data and the other two models, is largest. Figure 7.15 presents the $\cos\theta_{11}$ distribution for GENIE+MEC and GENIE+TEM models. This is the distribution where the agreement between data and the GENIE default compares least favorably than to the GENIE+MEC and GENIE+TEM models.

For brevity in the following, except where noted, I only show comparisons of data to predictions from the GENIE default model. Comparisons to GENIE+TEM and GENIE+MEC show qualitatively similar levels of agreement. Differences for specific distributions can be examined in

terms of the χ^2 test statistic values in Table 7.10.

For two track events, seven dynamic variables exist. These include properties of the long track that parallel the choices for one-track events, L_{21}, $\cos\theta_{21}$, and $\sin\Theta_{21}$, similar quantities for the second track, L_{22}, $\cos\theta_{22}$, and $\sin\Theta_{22}$, plus a quantity that describes the correlation between the two tracks in the event, which I take to be the difference in azimuthal variables $\phi_{221} = \phi_{22} - \phi_{21}$. Since track 2 can exit the detector, the meaning of L_{22} and $\sin\Theta_{22}$ differ somewhat from L_{21} and $\sin\Theta_{21}$. Two other two-track correlated variables of interest, which are not independent, are the cosine of the opening angle,

$$\cos\Omega_{221} = \cos\theta_{21}\cos\theta_{22} + \sin\theta_{21}\sin\theta_{22}\cos\left(\phi_{22} - \phi_{21}\right), \tag{7.31}$$

and the cosine of the acoplanarity angle

$$\cos\theta_A = \frac{\hat{s}_{21}\cdot\left(\hat{z}\times\hat{s}_{21}\right)}{\left|\hat{z}\times\hat{s}_{21}\right|} \tag{7.32}$$

$$= \sin\theta_{21}\sin\left(\phi_{22} - \phi_{21}\right), \tag{7.33}$$

with $\hat{z}$ a unit vector in the neutrino beam direction and $\hat{s}_{21}$ is a unit vector parallel to the first track direction at the event vertex. For the scattering of two initial state particles into two final state particles $(2 \to 2)$, one expects from momentum conservation $\phi_{221} = \pm\pi$ and $\cos\theta_A = 0$. Deviations of ϕ_{221} from $\pm\pi$ or of $\cos\theta_A$ from 0 could be caused by undetected tracks in the final state, from NC events in the sample, or from effects of final state interactions in CC events.

The opening angle serves a useful role in identifying spurious two-track events that result from the tracking algorithm "breaking" a single track into two tracks, most commonly in cosmic ray events. Broken tracks produce values of $\cos\Omega_{221}$ very close to -1. Figures 7.16 and 7.17 show the distributions of (L_{21} and L_{22}) and ($\cos\theta_{21}$ and $\cos\theta_{22}$) from the neutrino-enriched sample, compared to the GENIE default model. Figure 7.18 presents the distributions of $\cos\theta_{21}$ using GENIE+MEC and GENIE+TEM models. This is the distribution where the agreement between data and GENIE default model, compared to the agreement between data and the other two models, is largest. Figures 7.19 and 7.20 show the distributions of ($\sin\Theta_{21}$ and $\sin\Theta_{22}$) and ($\phi_{22} - \phi_{21}$ and

$\cos\Omega_{221}$) from the neutrino-enriched sample, compared to the GENIE default model.

For three-track events, eleven dynamic variables exist. A straightforward continuation of the previous choices leads to the choice of L_{31}, $\cos\theta_{31}$, $\sin\Theta_{31}$, L_{32}, $\cos\theta_{32}$, $\sin\Theta_{32}$, $\phi_{32}-\phi_{31}$, L_{33}, $\cos\theta_{33}$, $\sin\Theta_{33}$, and $\phi_{33}-\phi_{31}$ as the eleven variables. Other azimuthal angle difference such as

$$\phi_{32}-\phi_{33}=(\phi_{32}-\phi_{31})-(\phi_{33}-\phi_{31}) \tag{7.34}$$

are not independent. Figures 7.21, 7.22, 7.23, 7.24, 7.25, and 7.26 show the distributions of (L_{31}, L_{32}, and L_{33}), ($\cos\theta_{31}$, $\cos\theta_{32}$, and $\cos\theta_{33}$), ($\sin\Theta_{31}$, $\sin\Theta_{32}$, and $\sin\Theta_{33}$), ($\phi_{32}-\phi_{31}$ and $\cos\Omega_{321}$), ($\phi_{33}-\phi_{31}$ and $\cos\Omega_{331}$), ($\phi_{32}-\phi_{33}$ and $\cos\Omega_{323}$) from the neutrino-enriched sample, compared to the GENIE default model.

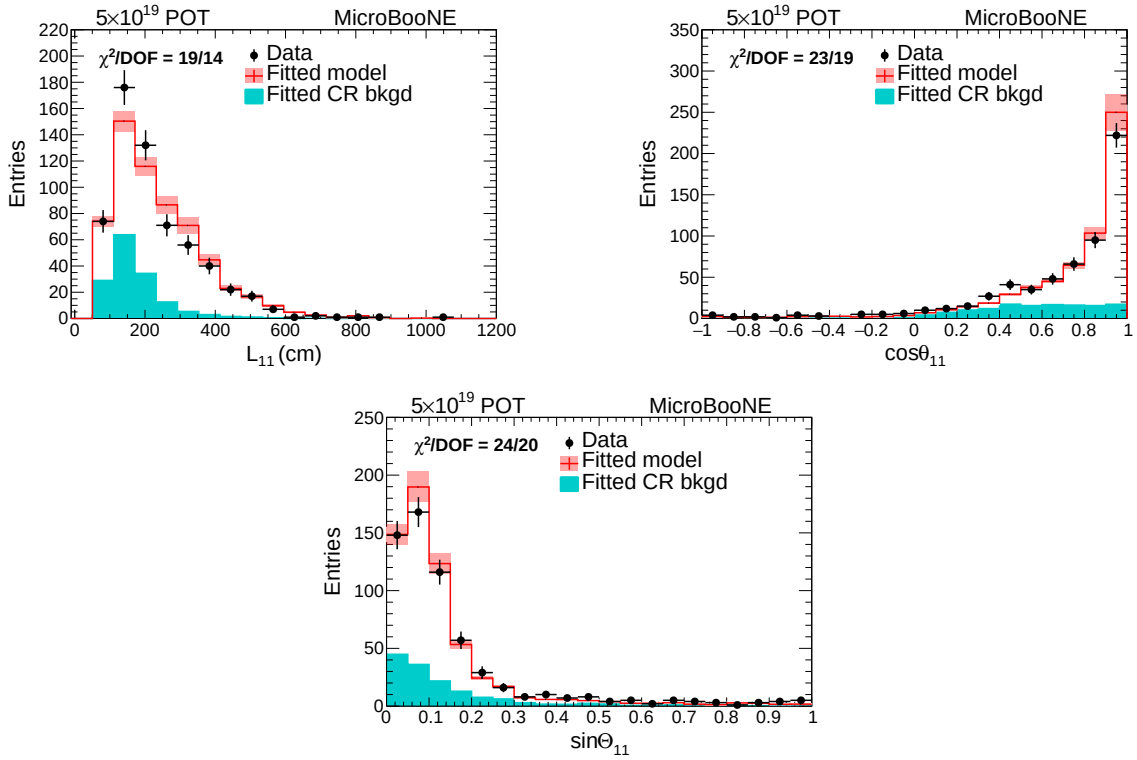


Figure 7.13: **Multiplicity = 1** candidate muon track length (top left), $\cos\theta$ (top right), and $\sin\Theta$ (bottom) distributions from neutrino-enriched sample for data and GENIE default MC. Figures from Ref. [21].

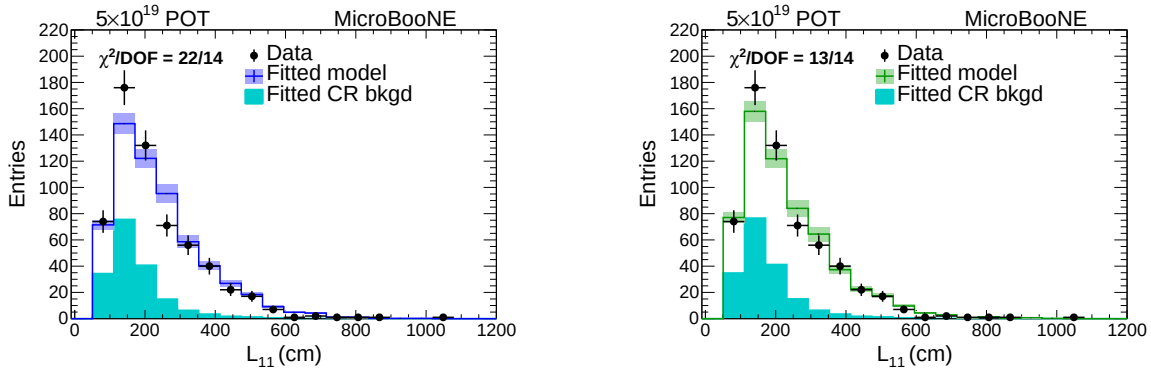


Figure 7.14: **Multiplicity = 1** candidate muon track length distribution using GENIE+MEC model (left); using GENIE+TEM model (right) from neutrino-enriched sample. Figures from Ref. [21].

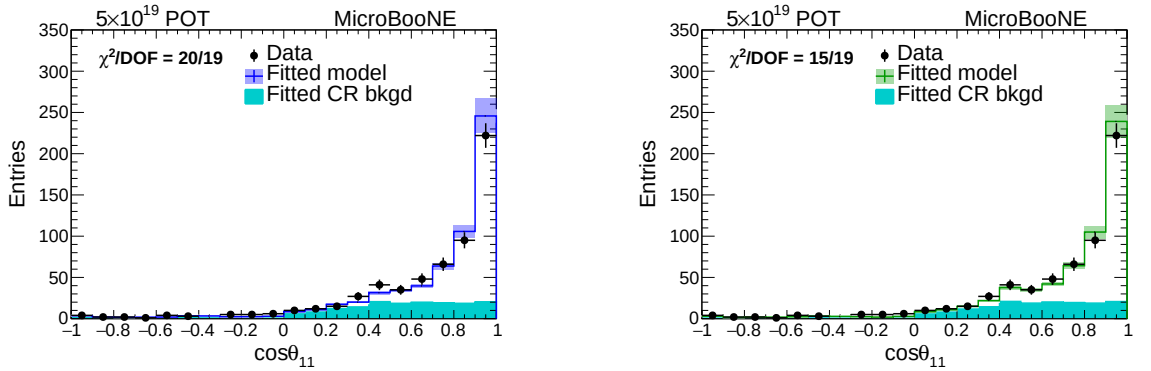


Figure 7.15: **Multiplicity = 1** candidate muon $\cos\theta$ distribution using GENIE+MEC model (left); using GENIE+TEM model (right) from neutrino-enriched sample. Figures from Ref. [21].

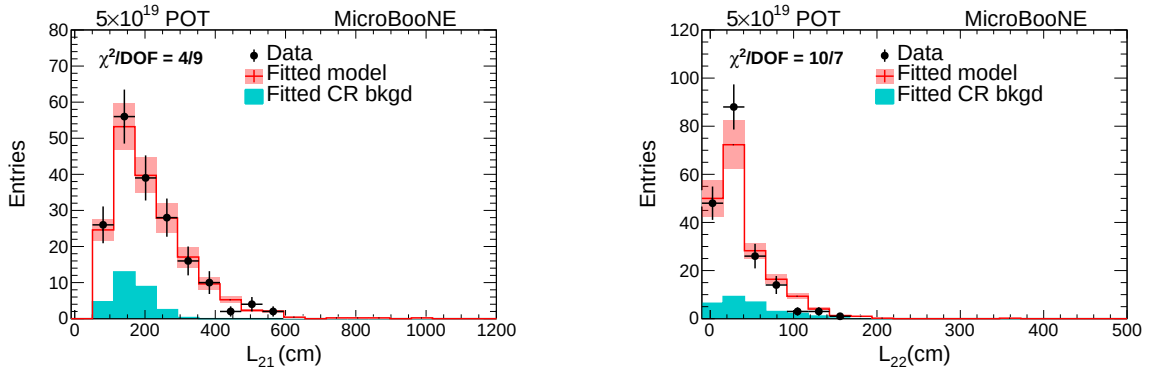


Figure 7.16: **Multiplicity = 2** Track length distribution for candidate muon (left); for second track of the event (right) from neutrino-enriched sample for data and GENIE default MC. Figures from Ref. [21].

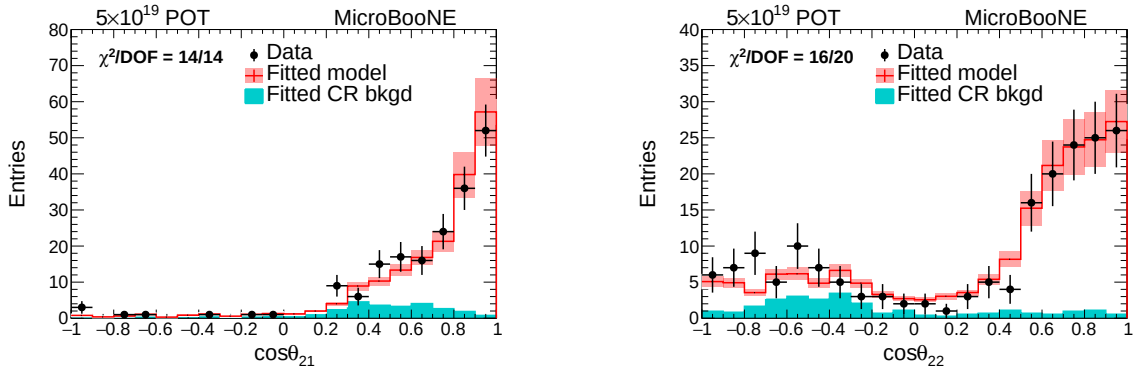


Figure 7.17: **Multiplicity = 2** $\cos\theta$ distribution for candidate muon (left); for second track of the event (right) using GENIE default model from neutrino-enriched sample for data and GENIE default MC. Figures from Ref. [21].

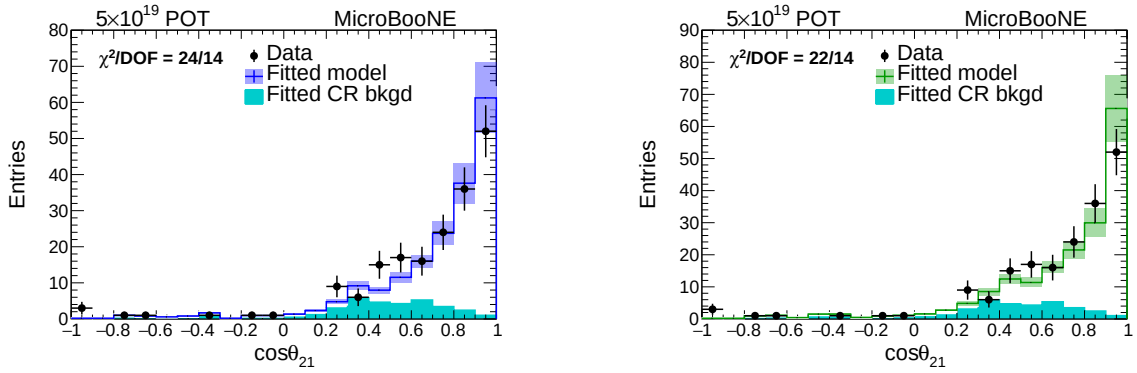


Figure 7.18: **Multiplicity = 2** $\cos\theta$ distribution for candidate muon using GENIE+MEC model (left); using GENIE+TEM model (right) from neutrino-enriched sample. Figures from Ref. [21].

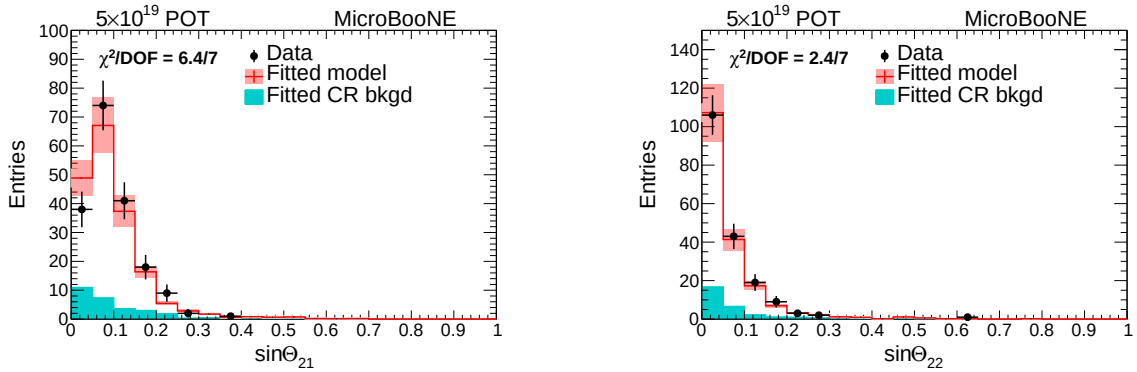


Figure 7.19: **Multiplicity = 2** $\sin\Theta$ distribution for candidate muon (left); for second track of the event (right) from neutrino-enriched sample for data and GENIE default MC. Figures from Ref. [21].

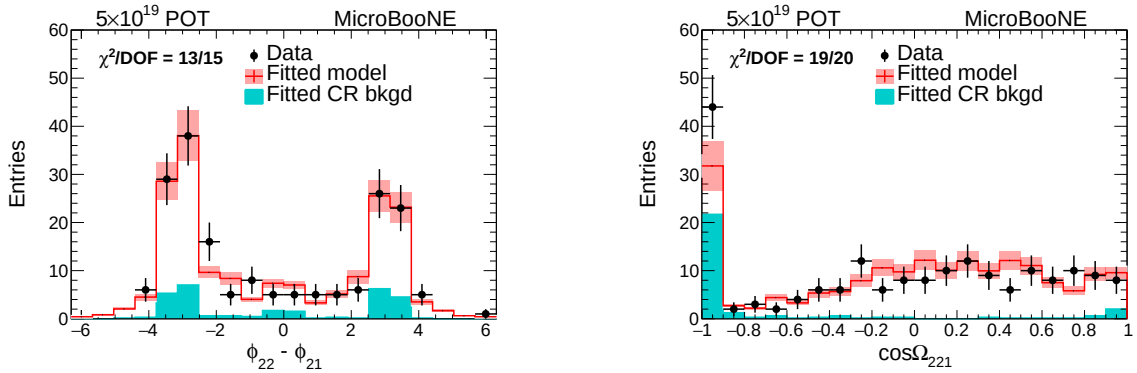


Figure 7.20: **Multiplicity = 2** $\phi_2 - \phi_1$ distribution (left); $\cos\Omega_{21}$ distribution (right) from neutrino-enriched sample for data and GENIE default MC. Figures from Ref. [21].

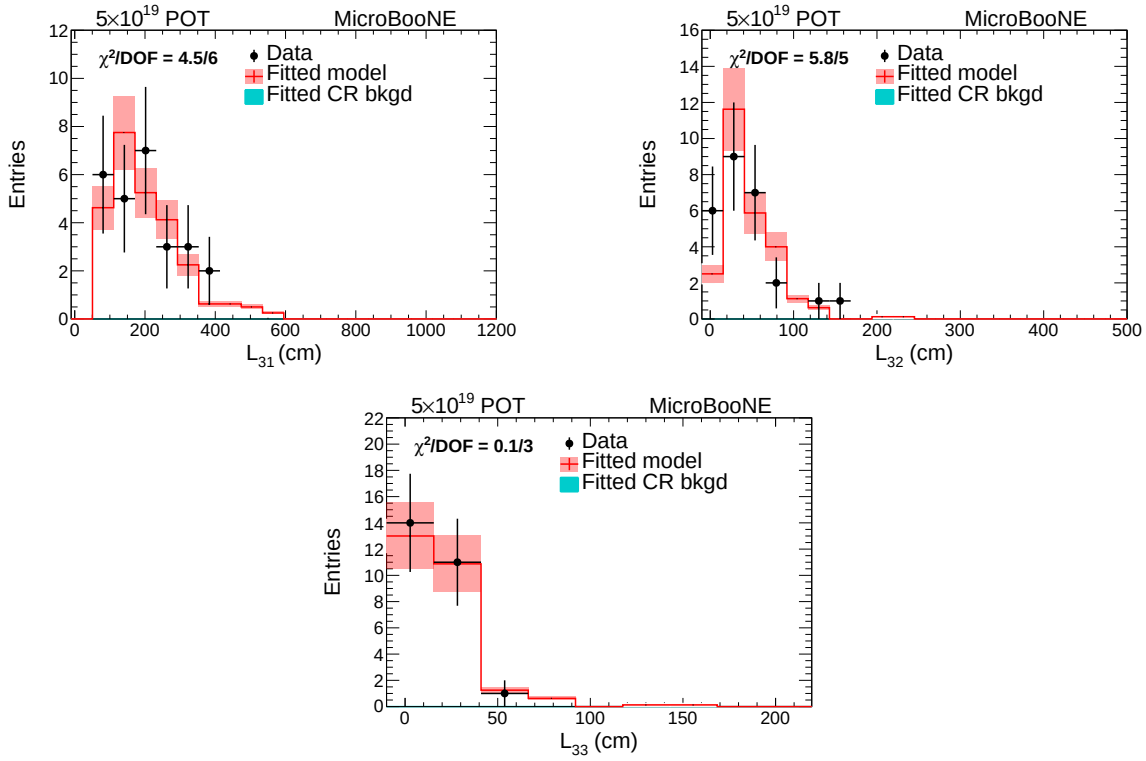


Figure 7.21: **Multiplicity = 3** Track length distribution for candidate muon (top left); for second longest track (top right); for shortest track of the event (bottom) from neutrino-enriched sample for data and GENIE default MC. Figures from Ref. [21].

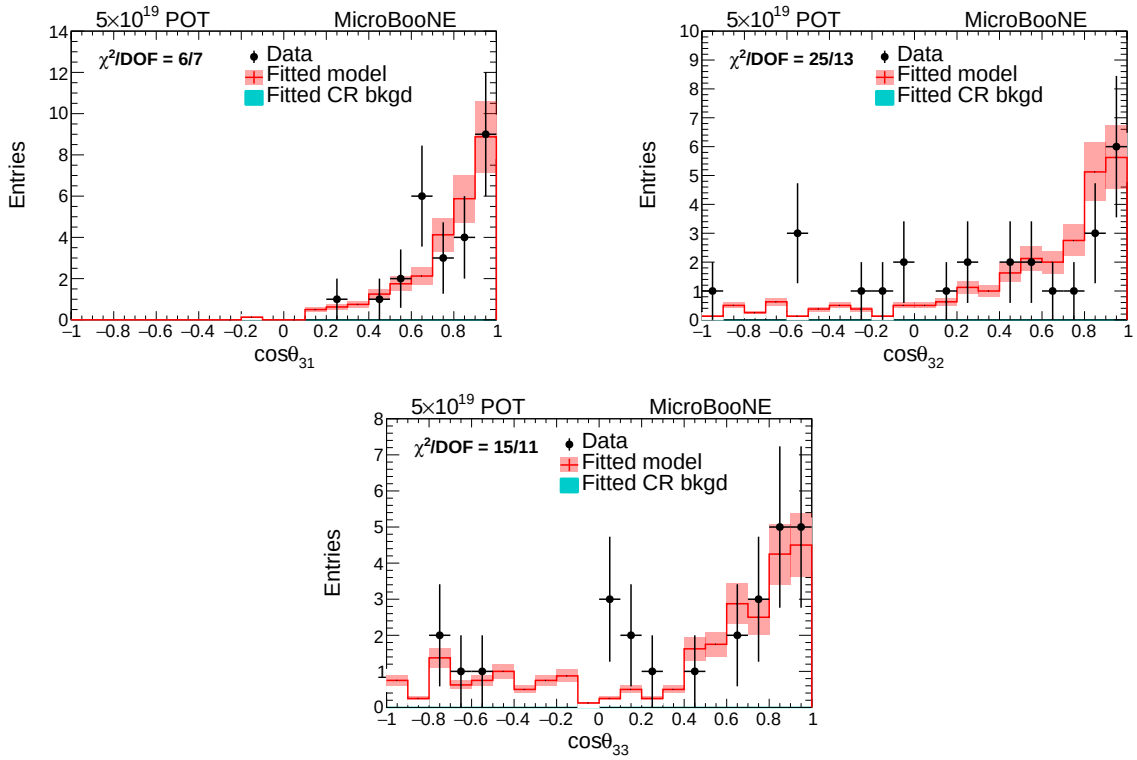


Figure 7.22: **Multiplicity = 3** $\cos\theta$ distribution for candidate muon (top left); for second longest track (top right); for shortest track of the event (bottom) from neutrino-enriched sample for data and GENIE default MC. Figures from Ref. [21].

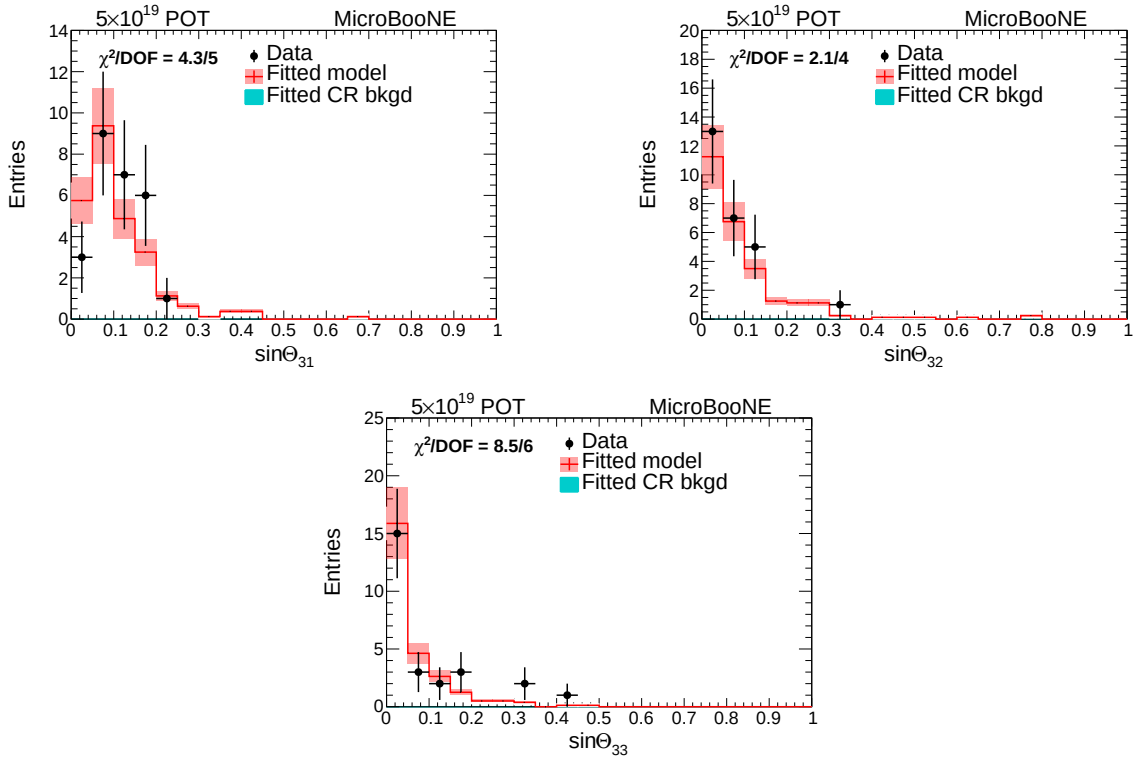


Figure 7.23: **Multiplicity = 3** $\sin\Theta$ distribution for candidate muon (top left); for second longest track (top right); for shortest track of the event (bottom) from neutrino-enriched sample for data and GENIE default MC. Figures from Ref. [21].

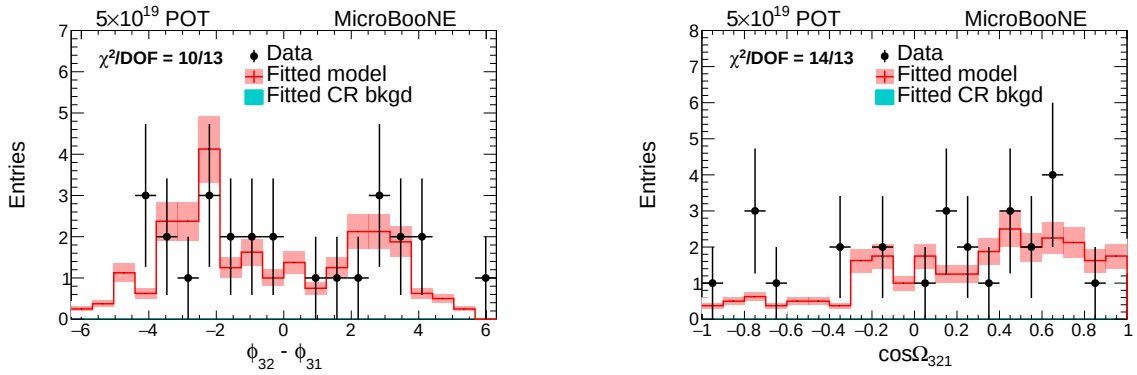


Figure 7.24: **Multiplicity = 3** $\phi_2 - \phi_1$ distribution (left); $\cos\Omega_{21}$ distribution between first and second track (right) from neutrino-enriched sample for data and GENIE default MC. Figures from Ref. [21].

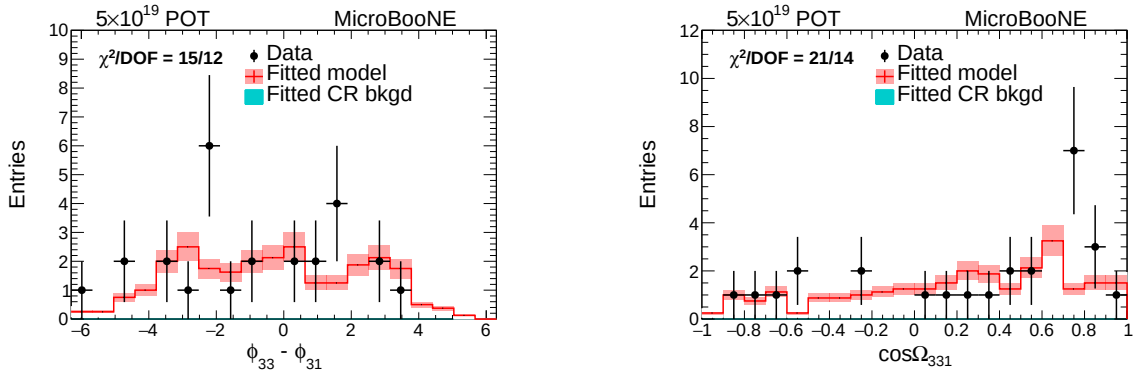


Figure 7.25: **Multiplicity = 3** $\phi_3 - \phi_1$ distribution (left); $\cos\Omega_{31}$ distribution between first and third track (right) from neutrino-enriched sample for data and GENIE default MC. Figures from Ref. [21].

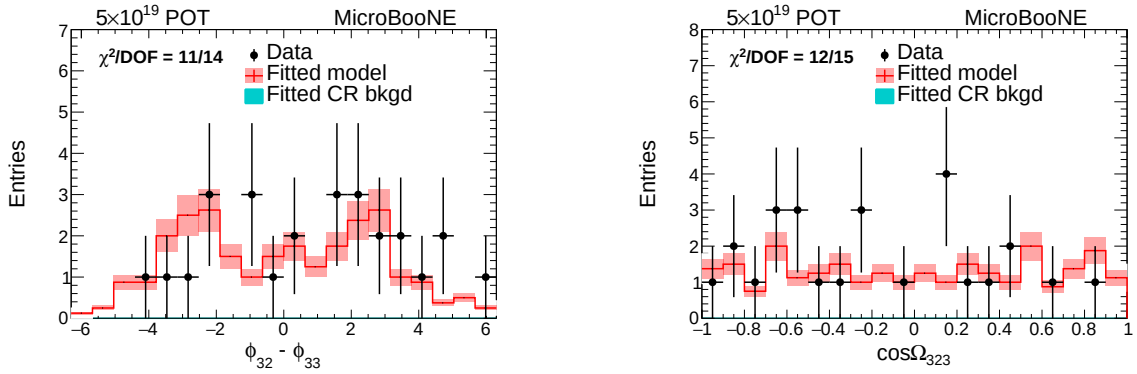


Figure 7.26: **Multiplicity = 3** $\phi_2 - \phi_3$ distribution (left); $\cos\Omega_{23}$ distribution between second and third track (right) from neutrino-enriched sample for data and GENIE default MC. Figures from Ref. [21].

7.5.5 χ^2 Tests for Kinematic Distributions

I quantify agreement between model and observation through use of χ^2 tests on the kinematic distributions described in Sec. 7.5.4. I use only the "neutrino-enriched" sample of events in which the candidate muon passes both the *PH* and *MCS* tests. Data are binned into histograms, with a bin k for a variable x_{ij}, d_{ijk}, and compared to model predictions constructed by assuming that the number of events in a bin k of a variable X_{ij}, m_{ijk}, consists of contributions from neutrino and CR background contributions. I shorten the notation in Eq. 7.29 to

$$m_{ijk} = M_{\nu,i}\hat{x}_{\nu,ijk} + M_{CR,i}\hat{x}_{CR,ijk}, \tag{7.35}$$

where $M_{\nu,i}$ and $M_{CR,i}$ are the number of neutrino and CR events, respectively, predicted to be in the neutrino-enriched category for multiplicity i (as described in Sec. 7.5.2); and $\hat{x}_{\nu,ijk}$ and $\hat{x}_{CR,ijk}$ the fraction of neutrino and CR events, respectively, falling in the k bin for variable x_{ij} as predicted by the GENIE model and the off-beam CR sample, respectively. The $\hat{x}_{\nu,ijk}$ and $\hat{x}_{CR,ijk}$ are shape distributions normalized to one:

$$\sum_{k=1}^{bins} \hat{x}_{\nu,ijk} = \sum_{k=1}^{bins} \hat{x}_{CR,ijk} = 1. \tag{7.36}$$

I then construct a χ^2 for x_{ij} using a Poisson form appropriate for the low statistics in many bins:

$$\chi^2_{ij} = 2\sum_{k=1}^{bins} \left(m_{ijk} - d_{ijk} - d_{ijk}\ln m_{ijk} + d_{ijk}\ln d_{ijk}\right). \tag{7.37}$$

Table 7.10 summarizes the results of these χ^2 comparison tests for 21 independent kinematic variables to the three GENIE models. Only bins with at least one data event and one model event were used in the calculation of χ^2. The number of degrees of freedom associated with the χ^2 test was set equal to the number of bins used for that histogram. I note here that these tests for consistency are defined at the level of statistical uncertainties only; systematic uncertainties are not incorporated into the χ^2 terms.

I summarize here salient features of Table 7.10 as follows: All three models consistently

Table 7.10: χ^2 test results for dynamically significant variables for all three GENIE models. Only the uncorrelated statistical uncertainties from data are used in forming the χ^2. Contributions from systematic uncertainties are not included. The last five listed distributions are not included in the total χ^2/DOF since these quantities can be expressed in terms of others. Table from Ref. [21].

	χ^2/DOF		
Distributions	**GENIE default**	**GENIE+MEC**	**GENIE+TEM**
L_{11}	19/14	22/14	13/14
L_{21}	4.0/9	4.6/9	7.3/9
L_{22}	10/7	8.4/7	16/7
L_{31}	4.5/6	3.4/6	5.5/6
L_{32}	5.8/5	3.9/6	6.5/6
L_{33}	0.1/3	0.7/3	0.5/3
$\cos\theta_{11}$	23/19	20/19	15/19
$\cos\theta_{21}$	14/14	24/14	22/14
$\cos\theta_{22}$	16/20	15/20	16/20
$\cos\theta_{31}$	6.0/7	4.2/7	9.2/7
$\cos\theta_{32}$	25/13	20/13	15/13
$\cos\theta_{33}$	15/11	13/11	17/11
$\sin\Theta_{11}$	24/20	21/20	25/20
$\sin\Theta_{21}$	6.4/7	3.6/7	6.3/7
$\sin\Theta_{22}$	2.4/7	3.4/7	2.4/6
$\sin\Theta_{31}$	4.3/5	6.0/5	9.1/5
$\sin\Theta_{32}$	2.1/4	2.5/4	1.6/4
$\sin\Theta_{33}$	8.5/6	7.0/5	9.5/6
$\phi_{22}-\phi_{21}$	13/15	12/15	14/15
$\phi_{32}-\phi_{31}$	10/13	9.2/13	10/14
$\phi_{33}-\phi_{31}$	15/12	13/12	8.7/11
$\phi_{32}-\phi_{33}$	11/14	11/14	11/14
$\cos\Omega_{221}$	19/20	13/20	13/20
$\cos\Omega_{321}$	14/13	13/13	17/13
$\cos\Omega_{331}$	21/14	16/14	12/14
$\cos\Omega_{323}$	12/15	18/15	19/15
Total χ^2/DOF	**228.1/216**	**216.9/216**	**229.6/216**

describe the data, with summed χ^2 per degree-of-freedom (χ^2/DOF) of 228.1/216, 216.9/216, and 229.6/216, respectively, and corresponding p-values of $P_{\chi^2} = 27\%$, 47%, and 25% for GENIE default, MEC, and TEM, respectively. No tune of GENIE is superior to any other with any meaningful statistical significance for the distributions I have considered. The acceptable values of χ^2 are consistent with the hypothesis that the combination of a GENIE event generator, the

MicroBooNE BNB flux model, and the MicroBooNE GEANT-based detector simulation satisfactorily describe the properties of neutrino events examined in this analysis. All elements of the MicroBooNE analysis chain thus appear to be performing satisfactorily; and no evidence exists for missing systematic effects that would produce data-model discrepancies outside the present level of statistics.

Aggregating Table 7.10 different ways uncovers no significant discrepancies. The χ^2 tests on leading track $\cos\theta$ and $\sin\Theta$ yield satisfactory results for all multiplicities. Combined χ^2/DOF for all distributions associated with a particular multiplicity likewise exhibit adequate agreement. The most poorly described single distribution is that for the length of the muon candidate in multiplicity 1 events. The P_{χ^2}, while acceptable, are 16% and 8% for the GENIE default and GENIE+MEC, respectively. The GENIE+TEM model has $P_{\chi^2} = 53\%$.

The χ^2 values for different distributions in a given multiplicity are calculated using the same events, which gives rise to concerns about correlations between different distributions. I have performed studies that verify that the χ^2 values would be highly correlated if the model and data disagreed by an overall normalization, but that otherwise the χ^2 tests on different distributions exhibit independent behavior, even when the same events are used. The P_{χ^2} values for different distributions do not cluster near 0 or 1, which is consistent with the view that the projections display approximately independent statistical behavior.

In summary, all GENIE models successfully describe, through χ^2 tests, the shapes of a complete set of dynamically significant kinematic variables for observed charged particle multiplicity distributions 1, 2, and 3. The statistical power of these tests from the overall data statistics available corresponds to approximately 4%, 7%, and 20% for multiplicity 1, 2, and 3, respectively.

7.5.6 χ^2 Tests for Multiplicity Distribution

While I find satisfactory agreement between GENIE models and kinematic distribution shapes using χ^2 tests that incorporate only statistical uncertainties, the situation differs for the overall multiplicity distribution. Here, I find statistical $\chi^2_M/\text{DOF} = 30/4$, $22/4$, and $28/4$ for the default, MEC, and TEM GENIE models, respectively. However, in the case of multiplicity, a significant

systematic uncertainty exists for tracking efficiency that must be taken into account before any conclusion can be drawn.

I incorporate a tracking efficiency contribution to the χ^2 test by defining

$$\chi^2_M = \sum_{i=1}^{2} \frac{\left(D_i - K\hat{M}(\delta)_i\right)^2}{\sigma_i^2} + \sum_{i=3}^{5} 2\left[K\hat{M}(\delta)_i - D_i \ln\left(K\hat{M}(\delta)_i\right) - D_i + D_i \ln(D_i)\right] \tag{7.38}$$
$$+ \sum_{i=1}^{5} 2\left[\hat{M}(0)_i - M_i \ln\left(\hat{M}(0)_i\right) - M_i + M_i \ln(M_i)\right] + \left(\frac{\delta}{0.15}\right)^2.$$

Here D_i is the number of neutrino events estimated by the signal extraction procedure (Sec. 7.5.2), and σ_i is the estimated uncertainty on D_i using the signal extraction procedure. For multiplicity 3 and higher, the uncertainty on D_i is purely statistical as the CR background becomes negligible. The quantities M_i are the number of events in the MC sample with multiplicity i. Finite statistics in the MC sample are incorporated by interpreting the M_i as Poisson fluctuations about their true values $\hat{M}(0)_i$ in the third term of Eq. 7.38. This analysis does not absolutely normalize MC to data, hence the relative normalization of data to MC is allowed to float via the parameter K in the first term of Eq. 7.38. The normalization constant K, while not used directly in the model test, is consistent with the predicted value from the default GENIE model.

As discussed in Sec. 7.4.2, changing the per-track efficiency by a constant fraction δ in the model would shift events between multiplicities according to

$$\hat{M}(\delta)_4 = \left[\hat{M}(0)_4\right](1-\delta)^3, \tag{7.39}$$

$$\hat{M}(\delta)_3 = \left[\hat{M}(0)_3 + 3\hat{M}(0)_4\,\delta\right](1-\delta)^2, \tag{7.40}$$

$$\hat{M}(\delta)_2 = \left[\hat{M}(0)_2 + 2\hat{M}(0)_3\,\delta + 3\hat{M}(0)_4\,\delta^2\right](1-\delta), \tag{7.41}$$

$$\hat{M}(\delta)_1 = \left[\hat{M}(0)_1 + \hat{M}(0)_2\,\delta + \hat{M}(0)_3\,\delta^2 + \hat{M}(0)_4\,\delta^3\right]. \tag{7.42}$$

For the nominal model used in the MC simulation $\delta = 0$. As discussed in Sec. 7.4.2 I estimate the uncertainty on δ to be 15%, and I introduce this into χ^2_M through the "pull term" $(\delta/0.15)^2$.

I minimize χ^2_M with respect to the tracking efficiency pull parameter δ, the MC-to-data nor-

malization K, and the five MC statistical quantities $\hat{M}(0)_i$, $i = 1-5$. This procedure yields

$$\chi^2_M/\text{DOF} = 6.4/3 \text{ (default)}, 4.3/3 \text{ (MEC)}, 5.8/3 \text{ (TEM)}, \tag{7.43}$$

$$\delta = 0.32 \text{ (default)}, 0.27 \text{ (MEC)}, 0.32 \text{ (TEM)}. \tag{7.44}$$

I find that a satisfactory χ^2 value can be obtained for the multiplicity distribution itself, albeit at the cost of a $\approx 2\sigma$ pull in the parameter δ.

7.6 Discussion

7.6.1 GENIE Predictions for Observed Multiplicity

At BNB energies, the nominal GENIE expectations for charged particle multiplicities at the neutrino interaction point are $\approx (80\%)$ $n = 2$ (from quasi-elastic scattering, $\nu_\mu n \to \mu^- p$, neutral pion resonant production $\nu_\mu n \to \mu^- R^+ \to \mu^- p\pi^0$, and coherent pion production $\nu_\mu \text{Ar} \to \mu^-\pi^+\text{Ar}$); $\approx (20\%)$ $n = 3$ (resonant charged pion production $\nu_\mu p \to \mu^- R^{++} \to \mu^- p\pi^+$); and $\approx (1\%)$ $n \geq 4$ (from multi-particle production processes referred to as DIS). However, final state interactions (FSI) of hadrons produced in neutrino scattering with the argon nucleus can subtract or add charged particles that emerge from within the nucleus. These multiplicities are further modified by the selection criteria.

The following list summarizes qualitative expectations for components of observed multiplicities from particular processes. These components can include contributions from the primary neutrino-nucleon scatter within the nucleus and secondary interactions of primary hadrons with the remnant nucleus. Secondary charged particles are usually protons, which are expected to be produced with kinetic energies that are usually too low for track reconstruction in this analysis. However, more energetic forward-produced protons from the upper "tail" of this secondary kinetic energy distribution may contribute. Note that the particle-type-dependent kinetic energy thresholds for charged particles entering the sample range from 31 MeV for a $\pi^\pm$ to 69 MeV for a proton.

- Multiplicity > 3, mainly predicted to be "DIS events" in which at least three short tracks are reconstructed. "DIS" is the usual term for multi-particle final states not identified with any particular resonance formation.

- Multiplicity $= 3$, mainly predicted to be $\mu^- p \pi^+$ events from Δ resonance production in which all three tracks are reconstructed. "Feed down" from higher multiplicity would be small due to the relatively small DIS cross section at MicroBooNE energies.

- Multiplicity $= 2$, mainly predicted to be QE $\mu^- p$ events and resonant $\mu^- p \pi^0$ events in which the proton is reconstructed, with a sub-leading contribution from "feed down" of resonant charged pion production events where one track fails to be reconstructed.

- Multiplicity $= 1$, mainly predicted to be "feed down" from QE $\mu^- p$ and $\mu^- p \pi^0$ events in which the proton is not reconstructed, with contributions from other higher multiplicity topologies in which more than one track fails to be reconstructed.

Figure 7.27 illustrates these expectations from GENIE. I note that, as expected, the three-track topology is dominated by resonant pion production in the default GENIE model, while the two-track and one-track topologies are QE-dominated with non-negligible resonance feed-down.

My observation of discrepancy of data compared to simulation in three-track compared to two-track topologies, shown in Fig. 7.10, is consistent with the low ν_μ CC pion cross sections compared to GENIE reported by MINERvA [70] using hydrocarbon targets at the somewhat higher neutrino energy from the Fermilab Neutrinos at the Main Injector (NuMI) beam. The T2K experiment [71] also reports a low pion production cross section relative to GENIE expectations using water targets in a neutrino beam with comparable energy to the BNB. However, MiniBooNE measured a charged pion production rate more in agreement with GENIE using mineral oil as a target in the same Fermilab BNB as used by MicroBooNE [72].

MicroBooNE also observes more one-track events than GENIE predicts, as shown in Fig. 7.10. This corroborates ArgoNeuT's observation that approximately 35% of neutrino interactions on argon targets with no pions detected in the final state also contained no observable proton [73].

While my observed multiplicity distribution disagrees with GENIE expectations and shows consistency with a number of other experiments, we cannot, as noted in Sec. 7.5.6 definitively exclude an alternate explanation of the discrepancy in terms of a tracking efficiency error at this time.

The kinetic energy thresholds limit acceptance in such a way that protons produced in FSI may not significantly contribute to the observed CPMD. Furthermore, this analysis requires a forward-going long contained track as a muon candidate, which restricts the final state phase space. Also, this analysis makes use of fully automated reconstruction. Therefore, results of this analysis should not be directly compared to the low energy proton multiplicity measurement reported by ArgoNeuT [17].

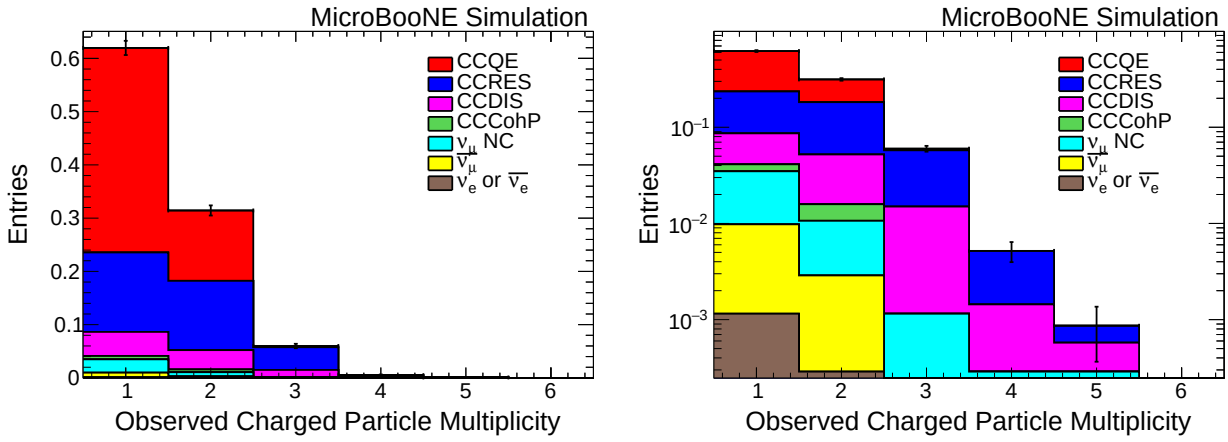


Figure 7.27: Observed (stacked) multiplicity distributions for different neutrino interaction types from BNB-only default MC simulation in linear scale (left); and in log y scale (right). Black error bars represent MC statistical uncertainties. Figures from Ref. [21].

7.6.2 GENIE Predictions for Kinematic Distributions

Kinematic distributions for fixed multiplicity suffer much less from tracking-related systematic uncertainties than the multiplicity probabilities; hence GENIE expectations for the shapes of kinematic distributions can be compared directly to data. The MEC and TEM tunes of GENIE primarily change the normalization of QE-like event topologies relative to resonance type topologies, and secondarily modify properties of low energy final state protons that would usually not satisfy

the acceptance criteria. Shape comparisons would thus not be expected to differentiate MEC and TEM from the GENIE default, and I have verified this expectation with my χ^2 tests. Accordingly I confine the following discussion to the default GENIE tune.

Figure 7.28 shows the predictions for reconstructed L_{11}, $\cos\theta_{11}$, and $\sin\Theta_{11}$ from the neutrino-enriched sample, using the GENIE default model. The muon track candidate is only mildly affected by the details of the recoiling hadronic system, and thus QE, RES, and DIS production produce similar shape contributions to L_{11}, $\cos\theta_{11}$, and $\sin\Theta_{11}$.

Figures 7.29, 7.30, 7.31, and 7.32 present the distributions of reconstructed (L_{21} and L_{22}), ($\cos\theta_{21}$ and $\cos\theta_{22}$), ($\sin\Theta_{21}$ and $\sin\Theta_{22}$), and ($\phi_{22}-\phi_{21}$ and $\cos\Omega_{221}$) respectively from the neutrino-enriched sample, using the GENIE default model. There is again minimal difference between QE, resonance, and DIS channels in track length, $\cos\theta$, or $\sin\Theta$ for the leading track. However, the QE channel produces contributions to the distributions in $\cos\theta_{21}$ and $\cos\theta_{22}$ that are considerably less forward-peaked than the resonance channel contributions. Distributions of these quantities in the data appear to be consistent with this picture. I also note that the $\sin\Theta_{22}$ distribution receives a contribution from QE scattering peaked at small values, consistent with expectations for a proton, and a broader distribution more similar to that of the leading muon track candidate that is consistent with the hypothesis that a charge pion can reconstruct as the second track in resonance contributions.

Striking differences between QE and RES contributions exist in the $\phi_{22}-\phi_{21}$ distribution between QE and resonance contributions in the Fig. 7.32. The QE contributions demonstrate the clear $\phi_{22}-\phi_{21}=\pm\pi$ peak expected for $2\rightarrow 2$ scattering. The gap between the $\pm\pi$ peaks is dominated by contributions from resonance feed-down.

Figures 7.33, 7.34, 7.35, 7.36, 7.37, and 7.38 present the reconstructed distributions of (L_{31}, L_{32}, and L_{33}), ($\cos\theta_{31}$, $\cos\theta_{32}$, and $\cos\theta_{33}$), ($\sin\Theta_{31}$, $\sin\Theta_{32}$, and $\sin\Theta_{33}$), ($\phi_{32}-\phi_{31}$ and $\cos\Omega_{321}$), ($\phi_{33}-\phi_{31}$ and $\cos\Omega_{331}$), ($\phi_{32}-\phi_{33}$ and $\cos\Omega_{323}$), respectively from the neutrino-enriched sample, using the GENIE default model. The three-track sample in GENIE is dominated by resonance contributions, and the data sample, although of limited statistics, has a CR background consistent with zero. I can thus compare in detail GENIE predictions for kinematic shape distributions to data. GENIE's predictions agree with observations.

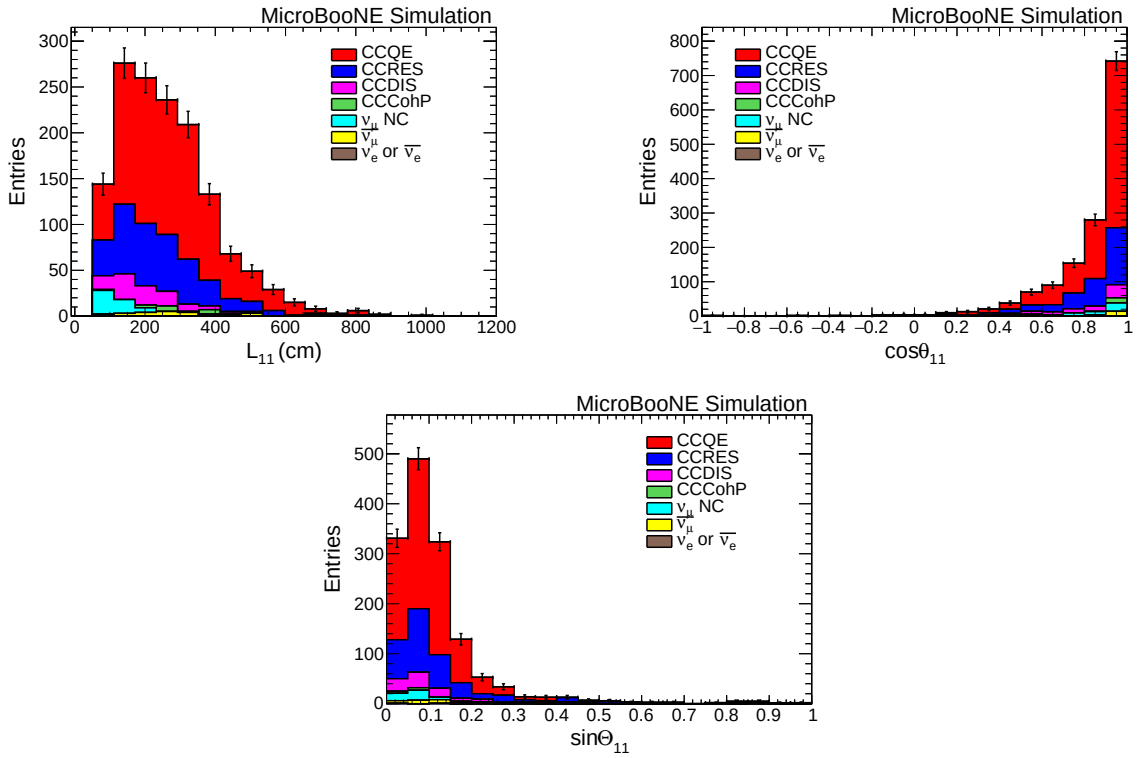


Figure 7.28: **Multiplicity = 1** GENIE default predictions for candidate muon track length (top left), $\cos\theta$ (top right), and $\sin\Theta$ (bottom) distributions. Black error bars represent MC statistical uncertainties. Figures from Ref. [21].

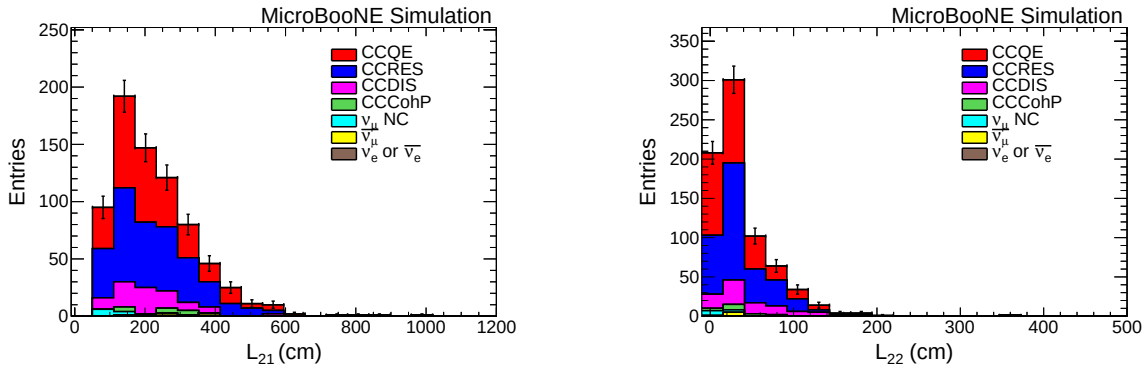


Figure 7.29: **Multiplicity = 2** Track length distribution for candidate muon (left); for second track (right) from GENIE default MC. Black error bars represent MC statistical uncertainties. Figures from Ref. [21].

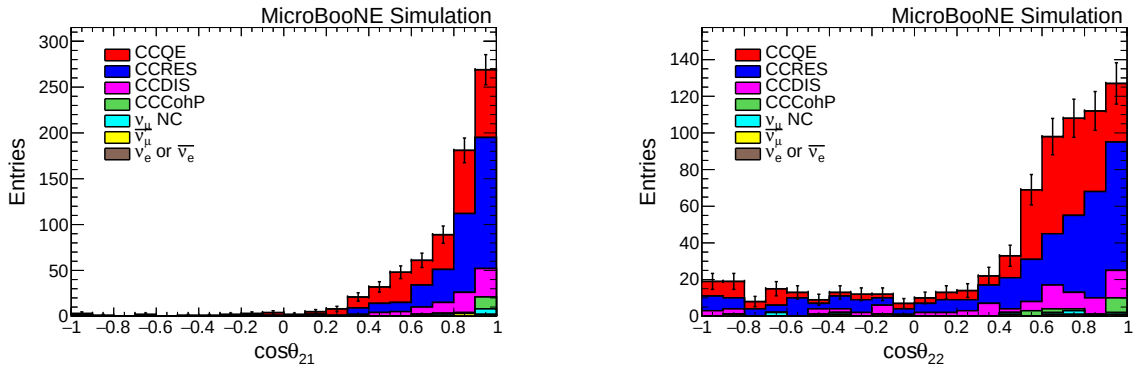


Figure 7.30: **Multiplicity = 2** Cosine of polar angle distribution for candidate muon (left); for second track (right) from GENIE default MC. Black error bars represent MC statistical uncertainties. Figures from Ref. [21].

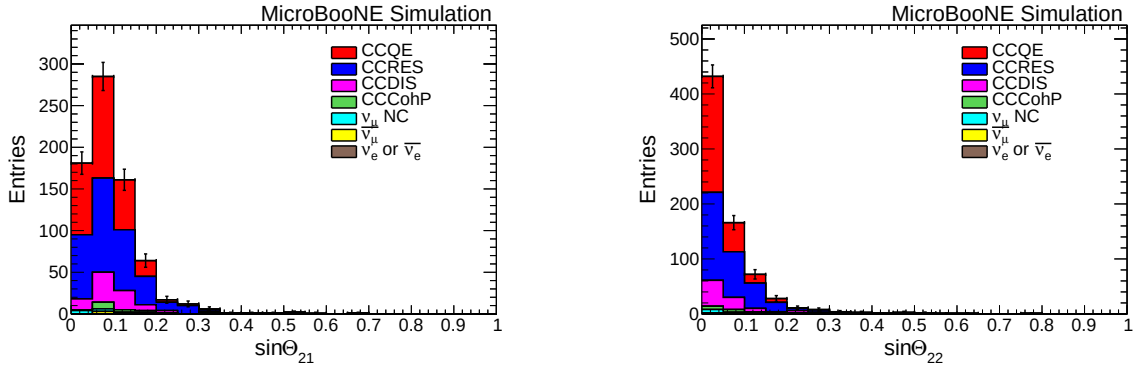


Figure 7.31: **Multiplicity = 2** SinΘ for candidate muon (left); for second track (right) from GENIE default MC. Black error bars represent MC statistical uncertainties. Figures from Ref. [21].

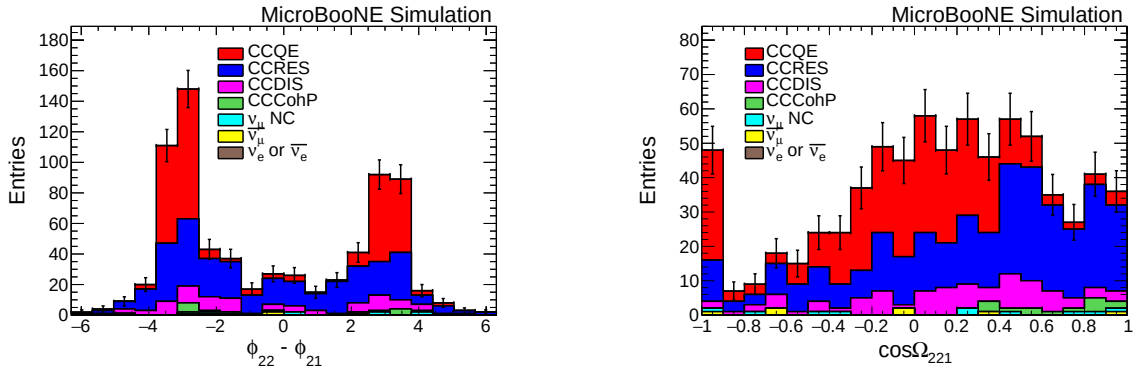


Figure 7.32: **Multiplicity = 2** $\phi_2 - \phi_1$ distribution (left); Cosine of opening angle distribution (right) from GENIE default MC. Black error bars represent MC statistical uncertainties. Figures from Ref. [21].

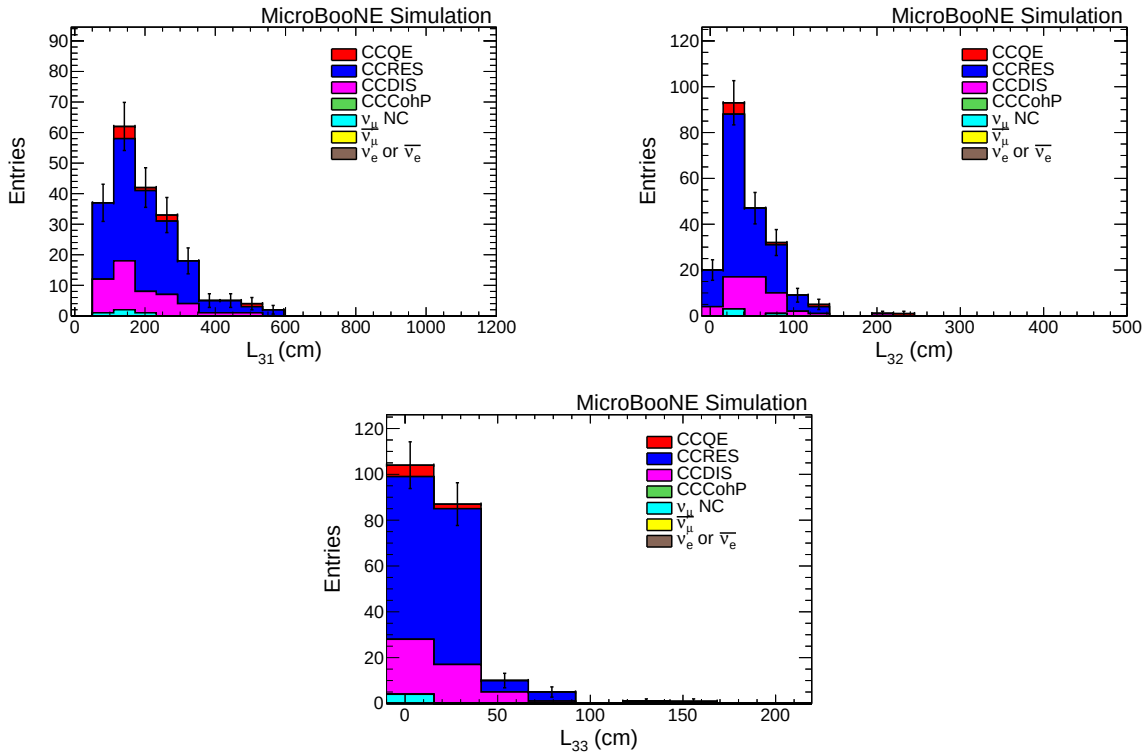


Figure 7.33: **Multiplicity = 3** Track length distribution for candidate muon (top left); for second longest track (top right); for shortest track (bottom) from GENIE default MC. Black error bars represent MC statistical uncertainties. Figures from Ref. [21].

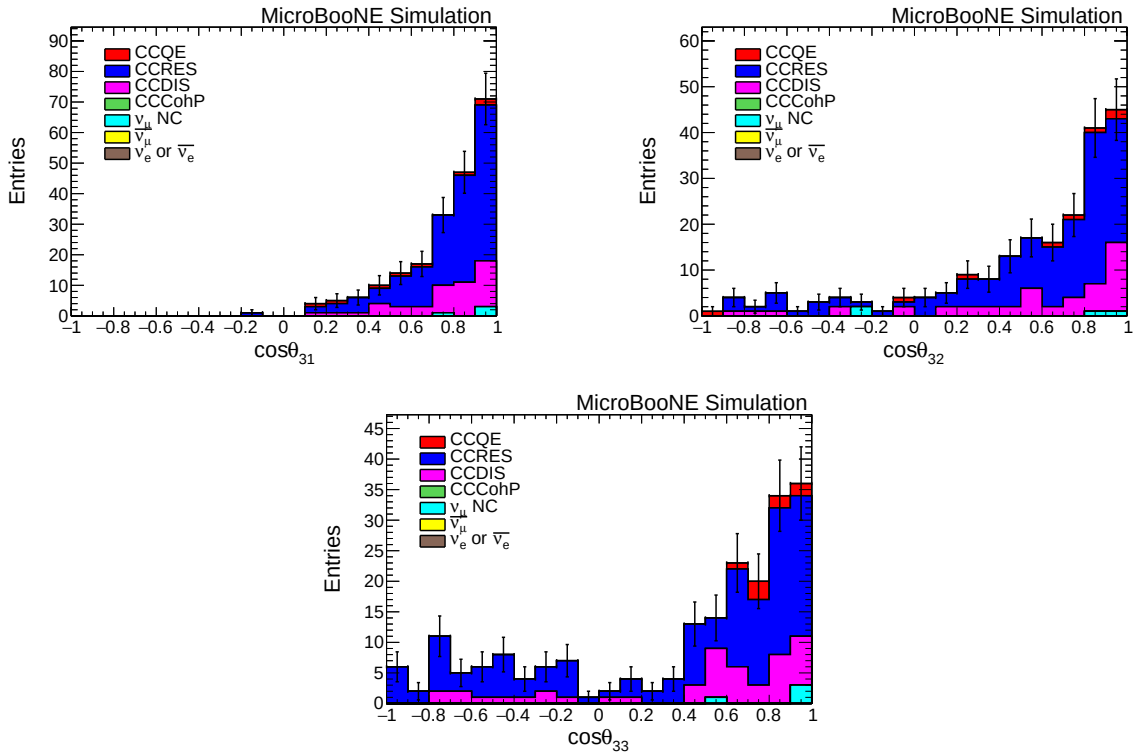


Figure 7.34: **Multiplicity = 3** Cosine of polar angle distribution for candidate muon (top left); for second longest track (top right); for shortest track (bottom) from GENIE default MC. Black error bars represent MC statistical uncertainties. Figures from Ref. [21].

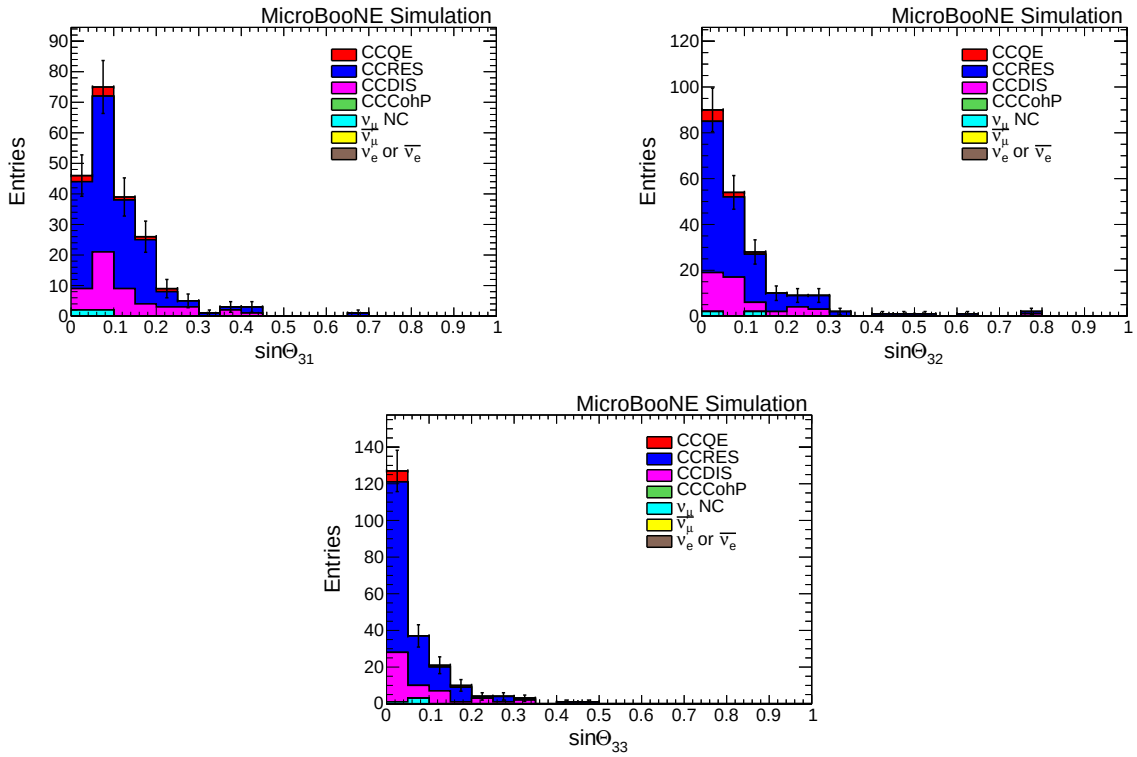


Figure 7.35: **Multiplicity = 3** SinΘ distribution for candidate muon (top left); for second longest track (top right); for shortest track (bottom) from GENIE default MC. Black error bars represent MC statistical uncertainties. Figures from Ref. [21].

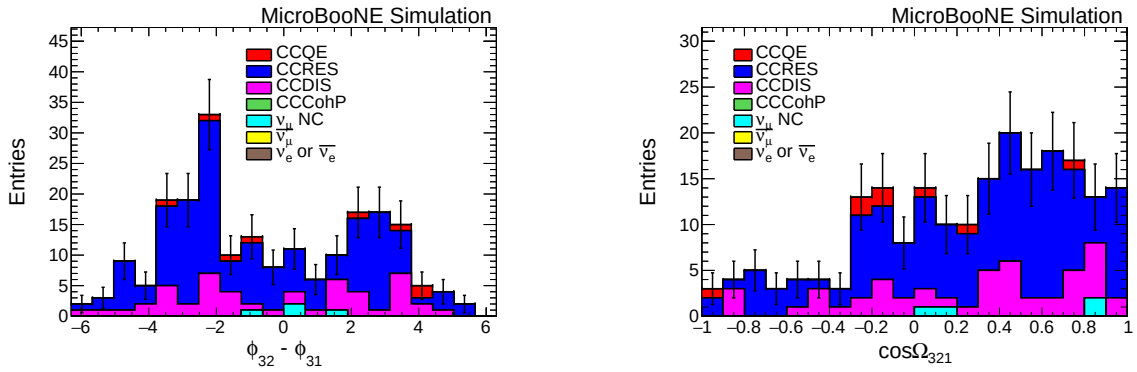


Figure 7.36: **Multiplicity = 3** $\phi_2 - \phi_1$ distribution (left); Cosine of opening angle distribution between first and second track (right) from GENIE default MC. Black error bars represent MC statistical uncertainties. Figures from Ref. [21].

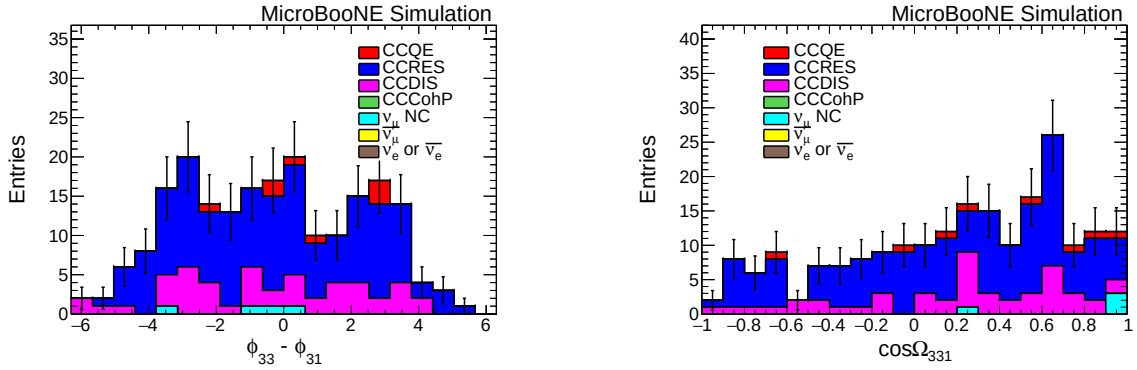


Figure 7.37: **Multiplicity = 3** $\phi_3 - \phi_1$ distribution (left); Cosine of opening angle distribution between first and third track (right) from GENIE default MC. Black error bars represent MC statistical uncertainties. Figures from Ref. [21].

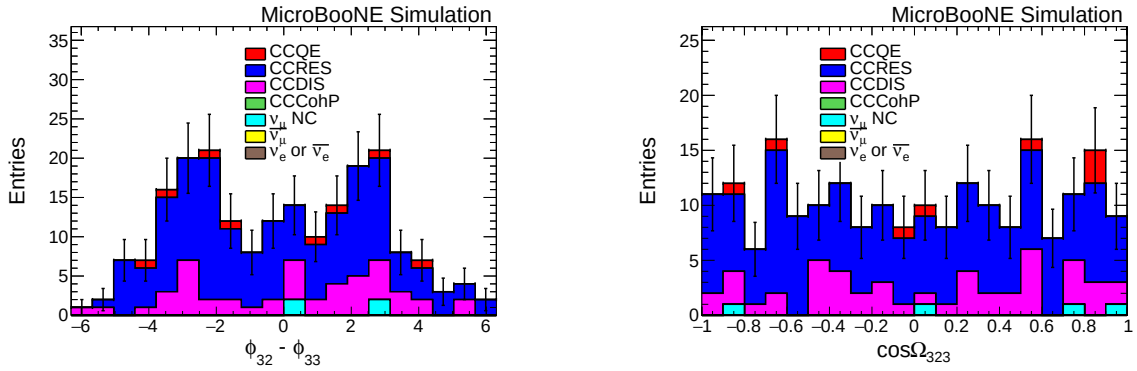


Figure 7.38: **Multiplicity = 3** $\phi_2 - \phi_3$ distribution (left); Cosine of opening angle distribution between second and third track (right) from GENIE default MC. Black error bars represent MC statistical uncertainties. Figures from Ref. [21].

Chapter 8

Summary

I have completed an analysis that compares observed charged-particle multiplicities and observed kinematic distributions of charged particles for fixed multiplicities in a restricted final state phase space for neutrino scattering events in argon to predictions from three GENIE tunes processed through the MicroBooNE simulation and reconstruction chain. This analysis takes into account statistical uncertainties in a rigorous manner, and estimates the impact of the largest expected systematic uncertainties. I observe that all elements of the MicroBooNE measurement chain−detector performance, data acquisition, event reconstruction, Monte Carlo event generator, detector simulation, and flux modeling−perform well.

With particle-type-dependent kinetic energy thresholds of 31 MeV for $\pi^{\pm}$ and 69 MeV for protons, I find all three GENIE tunes consistently describe data in the shapes of 26 different kinematic distributions at fixed multiplicities. GENIE appears to over-predict the number of three-track events in data that would be expected from resonant pion production, and to under-predict the number of one-track events; however, I cannot rule out a higher than expected tracking efficiency uncertainty as an alternative explanation for these observations. This study thus empirically validates the use of GENIE in describing single-process (quasi-elastic, resonance) neutrino scattering on argon, but not the predictions for the relative contributions of different processes to the overall cross section. I find no significant differences at this stage in the experiment between the default GENIE tune or tunes that add MEC or TEM. Use of any of the three GENIE tunes for future

MicroBooNE analyses, or for physics studies of inclusive final states performed for the SBN and DUNE experiments, receives empirical validation from this work.

As part of this analysis, I have developed a data-driven cosmic ray background estimation method based on the energy loss profile and multiple Coulomb scattering behavior of muons. Within the available Monte Carlo statistics, I have shown that this method provides an unbiased estimate of the number of neutrino events in a pre-filtered sample, and, given current statistical precision, it is independent of the signal-to-background level, final state charged particle multiplicity, and other kinematic properties of the final state particles. This method can be applied to a broad range of charged current process measurements.

Bibliography

[1] C. Rubbia, CERN-EP-INT-77-08 (1977).

[2] M. Antonello *et al.* (MicroBooNE, LAr1-ND, and ICARUS-WA104 Collaborations), [arXiv:1503.01520 [physics.ins-det]] (2015).

[3] R. Acciarri *et al.* (DUNE Collaboration), [arXiv:1512.06148 [physics.ins-det]] (2015).

[4] R. Acciarri *et al.* (MicroBooNE Collaboration), JINST **12**, P02017 (2017).

[5] S. Amerio *et al.* (ICARUS Collaboration), Nucl. Instrum. Meth. **A527**, 329 (2004).

[6] Nicola McConkey for the SBND collaboration, J. Phys.: Conf. Ser. **888**, 012148 (2017).

[7] J. Beringer *et al.* (Particle Data Group), Phys. Rev. D **86**, 010001 (2012).

[8] C. Andreopoulos, *et al.*, Nucl. Instrum. Meth. **A614**, (2010) 87-104.

[9] S. Agostinelli *et al.* (GEANT4 Collaboration), Nucl.Instrum.Meth. **A506**, (2003).

[10] E. D. Church, [arXiv:1311.6774 [physics.ins-det]].
URL: http://larsoft.org/

[11] J. A. Formaggio and G. P. Zeller, Rev. Mod. Phys. **84**, 1307 (2013); T. Katori, Neutrino. Cross. Section. Newsletter (12/13/2014),
URL: http://pprc.qmul.ac.uk/∼katori/nu-xsec.html.

[12] R. A. Smith and E. J. Moniz, Nuclear Physics **B43**, (1972) 605-622.

[13] L. Alvarez-Ruso *et al.*, FERMILAB-PUB-17-195-ND-T, [arXiv:1706.03621 [hep-ph]].

[14] F. Arneodo *et al.* (ICARUS-Milano Collaboration), Phys. Rev. D **74**, 112001 (2006).

[15] C. Anderson *et al.* (ArgoNeuT Collaboration), Phys. Rev. Lett. **108**, 161802 (2012).

[16] R. Acciarri *et al.* (ArgoNeuT Collaboration), Phys. Rev. D **89**, 112003 (2014).

[17] R. Acciarri *et al.* (ArgoNeuT Collaboration), Phys. Rev. D **90**, 012008 (2014).

[18] R. Acciarri *et al.* (ArgoNeuT Collaboration), Phys. Rev. Lett. **113**, 261801 (2014). [erratum]: Phys. Rev. Lett. **114**, 039901 (2015).

[19] R. Acciarri *et al.* (ArgoNeuT Collaboration), Phys. Rev. D **95**, 072005 (2017).

[20] A. A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), Phys. Rev. D **79**, 072002 (2009).

[21] C. Adams *et al.* (MicroBooNE Collaboration), Phys. Rev. D [arXiv:1805.06887 [hep-ex]] (submitted).

[22] J. Chadwick, Verhandl. Dtsc. Phys. Ges. **16**, 383 (1914).

[23] W. Pauli, Open letter sent to a group of physicists meeting in T?bingen (1930).

[24] Frederick Reines and Clyde L. Cowan, Nature **178**, (1956).

[25] Niels Bohr, Nature **137**, (1936) 344−348.

[26] J. M. Lattimer and M. Prakash, The Astrophysical Journal Letters **550**, (2001) 426−442.

[27] J. A. Formaggio, G. P. Zeller, Rev. Mod. Phys. **84**, (2012) 1307−1341.

[28] Raymond Davis, Jr., Don S. Harmer, and Kenneth C. Hoffman, Phys. Rev. Letters **20**, (1968).

[29] Y. Fukuda *et al.* (Super-Kamiokande Collaboration), Phys. Rev. Letters **81**, (1998).

[30] John N. Bahcall, M. Concha Gonzalez-Garcia, and Carlos Pea-Garay, Journal of High Energy Physics **2001**, (2001).

[31] Ziro Maki, Masami Nakagawa, and Shoichi Sakata, Progress of Theoretical Physics **28**, (1962).

[32] William J. Marciano, Phys. Rev. D **20**, (1979).

[33] C. Athanassopoulos *et al.* (LSND), Nucl. Instrum. Meth. **A388**, (1997) 149−172.

[34] A.A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), Nucl. Instrum. Meth. **A599**, (2009) 28−46.

[35] A. Aguilar *et al.* (LSND Collaboration), Phys. Rev. D **64**, (2001) 112007.

[36] A. A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), Phys. Rev. Lett. **110**, (2013) 161801

[37] S. Mufson *et al.* (NOvA Collaboration), Nucl. Instrum. Meth. **A799**, (2015) 1−9.

[38] K. Abe *et al.* (Hyper-Kamiokande working group), [arXiv:1109.3262 [hep-ex]] (2011).

[39] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **93**, 012006 (2016).

[40] Fermilab Target systems, URL: http://targets.fnal.gov/BNB_neutrino_beam.html.

[41] A. Aguilar *et al.* (LSND Collaboration), IEEE Transactions on Nuclear Science **18**, (1971).

[42] Fermilab BNB horn,
https://web.fnal.gov/project/TargetSystems/BNB_target_hall/SitePages/BNB%20Horn.aspx

[43] A. A. Aguilar-Arevalo *et al.* ((MiniBooNE Collaboration)), Phys. Rev. D **72**, (2002).

[44] MicroBooNE approved plots,
URL: https://microboone-exp.fnal.gov/public/approved_plots/plots/4269/BNB_flux_uboone.pdf

[45] The MicroBooNE Collaboration, MICROBOONE-NOTE-1024-PUB (2017).

[46] Jyoti Joshi, Xin Qian, [arXiv:1511.00317 [physics.ins-det]] (2015).

[47] R. Acciarri *et al.* (MicrobooNE Collaboration), JINST **12**, P08003 (2017).

[48] D. Heck *et al.*, Forschungszentrum Karlsruhe Report FZKA 6019, 1998.
URL: https://www.ikp.kit.edu/corsika/70.php.

[49] MicroBooNE approved plots, Event displays
URL: https://microboone-exp.fnal.gov/public/approved_plots/Event_Displays.html

[50] J. S. Marshall and M.A. Thomson, Eur. Phys. J. **C75**, (2015).

[51] The MicroBooNE Collaboration, MICROBOONE-NOTE-1015-PUB (2016).

[52] R. Acciarri *et al.* (MicroBooNE Collaboration), Eur. Phys. J. C **78**, 82 (2018).

[53] Aleena Rafique, DPF 2017,
https://indico.fnal.gov/event/11999/session/9/material/slides/0?contribId=359

[54] L. Alvarez-Ruso, L. S. Geng, S. Hirenzaki, and M. J. Vicente Vacas, Phys. Rev. C **75**, 055501 (2007).
[erratum]: Phys. Rev. C **80**, 019906 (2009).

[55] C. H. Llewellyn Smith, Phys. Rept. **3**, 261 (1972).

[56] D. Rein and L. M. Sehgal, Ann. Phys. **133**, 79 (1981).

[57] A. Bodek and U. K. Yang, J. Phys. G **29**, (2003) 1899−1906.

[58] Alberto Mengoni and Yutaka Nakajima, Journal of Nuclear Science and Technology, **34**, (1994) 151−162.

[59] J. E. Amaro *et al.*, Phys. Lett. B **696**, (2011) 151−155.

[60] T. Katori, AIP Conf. Proc. **1663**, 030001 (2015).

[61] Omar Benhar, Alessandro Lovato, and Noemi Rocco, [arXiv:1312.1210 [nucl-th]] (2013).

[62] A. Bodek, H. S. Budd, and M. E. Christy, Eur. Phys. J. C **71**, 1726 (2011).

[63] A. Bodek, M. E. Christy, and B. Coopersmith, Eur. Phys. J. C **74**, 3091 (2014).

[64] The MicroBooNE Collaboration, MICROBOONE-NOTE-1013-PUB (2016).

[65] C. Forti *et al.*, Phys. Rev. D **42**, 3668 (1990).

[66] C. Patrignani *et al.* (Particle Data Group), Chin. Phys. C **40**, 100001 (2016).

[67] W.K. Sakumoto *et al.*, Nucl. Instrum. Meth. **A294**, (1990) 179−192.

[68] F. James and M.Roos, Comput. Phys. Commun. **10**, (1975) 343−367.

[69] Rene Burn and Fons Rademakers, Nucl. Instrum. Meth. Phys. Research **A389**, (1997) 81−86.

[70] C. L. McGivern *et al.* (MINERvA Collaboration), Phys. Rev. D **94**, 052005 (2016).

[71] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **95**, 012010 (2017).

[72] P. A. Rodrigues, AIP Conference Proceedings **1663)**, 030006 (2015).

[73] K. Partyka (ArgoNeuT Collaboration), AIP Conf. Proc. **1663**, 080007 (2015).

[74] The MicroBooNE Collaboration, MICROBOONE-NOTE-1010-PUB (2016).

Appendix A

MicroBooNE Detector Stability

A.1 Introduction

These MicroBooNE detector stability studies described in this chapter are derived from a MicroBooNE public note [64], which I authored. Data quality monitoring plays a crucial role in detector and analysis reliability because it ensures that the detector performance is optimal to maximize the precision of the physics results. Reconstruction stability is an important requirement for the ν_μ charged-current inclusive analysis in MicroBooNE. Stable run selection criterion is used in conjunction with other requirements, including good operating conditions for the detector and, in the case of data collected during BNB operation, beam-quality requirements, to ensure data of sufficient quality for MicroBooNE's ν_μ charged-current inclusive analysis. This chapter presents detector stability plots of the selected MicroBooNE reconstruction variables used by ν_μ charged-current inclusive analysis.

A.2 Detector Reconstruction Algorithms and Objects

I have chosen to use three products of the reconstruction to monitor the detector performance: 3D charged-particle tracks, "flashes" (described below), and interaction vertices. These quantities are chosen since these are the most relevant to the event selection of the ν_μ charged-current (CC)

inclusive analysis [74].

A "reconstructed track" is a result of a pattern recognition and tracking algorithm, where a 3D charged-particle trajectory is formed by matching the TPC 'hits' and clusters (formed by grouping hits that are connected in space and time within a wire plane) across the three wire planes of the TPC. I present here the number of reconstructed tracks from the output of one of tracking algorithms—the "KalmanHit" tracking algorithm—where all tracks are treated independently without the use of any neutrino-interaction vertex association. The KalmanHit algorithm is one of the primary tracking algorithms used to reconstruct muon tracks in the CC inclusive analysis.

A "PandoraNu vertex" is the reconstructed vertex created using the Pandora pattern-recognition algorithm suite [51]. It is created by searching for pairs of 2D clusters from different wire-plane views to create plausible candidate 3D vertex positions, and then it selects the best candidate from each match by examining the distribution of 2D hits around projections of the candidates into each view. By only requiring two planes for the reconstruction it reduces the impact of non-responsive or noisy wires.

An "optical flash" is reconstructed by grouping 'optical hits' (characterized by their amplitude, width, time relative to the trigger, and the photo multiplier tube (PMT) on which they occurred) that occur coincident within a window of roughly 1 μs in time across the detector. Therefore, an optical flash is a reconstructed detection of prompt scintillation light produced by charged particles in the detector volume. The ν_μ CC inclusive analysis requires a flash with amplitude greater than 50 photo-electrons (PE) over all PMTs to reduce noise and other background events, and so I check the average number of reconstructed flashes above this threshold per event to help ensure a stable selection.

A.3 Reconstruction Stability over Time

Detector stability plots provide a high level check of data and simulation by monitoring any change in detector hardware and reconstruction software run by run, highlighting the impact of such changes on various reconstructed parameters. The following is a list of key variables for the ν_μ CC inclusive analysis that were monitored for each run.

- Average number of tracks (from the KalmanHit algorithm) per event and its spread.
- Average number of vertices (from the PandoraNu algorithm) per event and its spread
- Average number of $>$ 50 PE flashes per event

These quantities are a reflection of the overall stability and quality of the collected data, and are used as a component selecting good data-taking periods. In addition to above listed quantities, y and z center of flashes distributions were also used for selecting good data-taking periods but are not presented in this note.

In this section I present the stability of these quantities in data collected from the MicroBooNE detector since late February, 2016. I analyze data both with and without exposure to the BNB neutrino beam, called "on-beam" and "off-beam" data, respectively. Due to the long readout time of the detector (4.8 ms), all events (where by 'event' I mean data acquired during the 4.8-ms readout time) are dominated by the presence of cosmic ray interactions.

The spread of these variables was calculated for all events over a run by taking the square root of the variance of the distribution. The distribution of each of these quantities over all runs was fit to a Gaussian and runs lying within 3-σ bounds of the fitted Gaussian distribution in both data streams were selected. In addition 3-σ bound cuts were also applied to y and z center of flashes plots for stable run selection. Table A.1 and A.2 present average and spread (RMS) values from the Gaussian fits of the selected quantities for on-beam and off-beam data respectively. The average values match rather well, and the spreads are reasonably close, but slightly larger in off-beam data due to less statistics in off-beam data stream. A fraction of 91.7% of the delivered protons on target (POT) from the BNB pass this criterion. As both on-beam and off-beam data streams pass through the same event-triggering algorithms, and, as previously mentioned, the events are dominated by cosmic rays, the properties of reconstructed objects are largely comparable in the two data streams. Figure A.1, A.2, and A.3 present detector stability plots for Kalmanhit tracks, PandoraNu vertices, and $>$ 50 PE flashes per event respectively. Figure A.4, A.5, and A.6 present corresponding normalized y-projections of detector stability plots for both off-beam and on-beam data streams.

Table A.1: Average and spread (RMS) values with errors obtained from Gaussian fits of selected quantities for on-beam data. Table from Ref. [64].

Variables & Algorithms	**On-beam data**	
	mean	**spread (σ)**
Avg. number of KalmanHit tracks (per event)	15.85 ± 0.01	0.24 ± 0.01
Spread of number of KalmanHit tracks	4.62 ± 0.01	0.13 ± 0.01
Avg. number of PandoraNu vertices (per event)	10.54 ± 0.01	0.21 ± 0.01
Spread of number of PandoraNu vertices	5.04 ± 0.01	0.17 ± 0.01
Avg. number of $>$ 50 PE flashes (per event)	31.23 ± 0.03	0.78 ± 0.03

Table A.2: Average and spread (RMS) values with errors obtained from Gaussian fits of selected quantities for off-beam data. Table from Ref. [64].

Variables & Algorithms	**Off-beam data**	
	mean	**spread (σ)**
Avg. number of KalmanHit tracks (per event)	15.81 ± 0.01	0.30 ± 0.01
Spread of number of KalmanHit tracks	4.61 ± 0.01	0.20 ± 0.01
Avg. number of PandoraNu vertices (per event)	10.46 ± 0.01	0.30 ± 0.01
Spread of number of PandoraNu vertices	5.01 ± 0.01	0.23 ± 0.01
Avg. number of $>$ 50 PE flashes (per event)	31.30 ± 0.04	0.83 ± 0.03

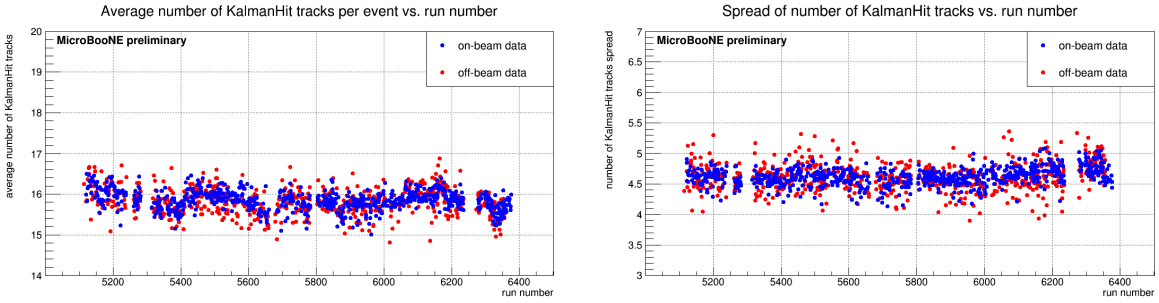


Figure A.1: Average number of KalmanHit tracks per event vs. run number with selected stable runs (left) and spread of number of KalmanHit tracks vs. run number with selected stable runs (right). Figures from Ref. [64].

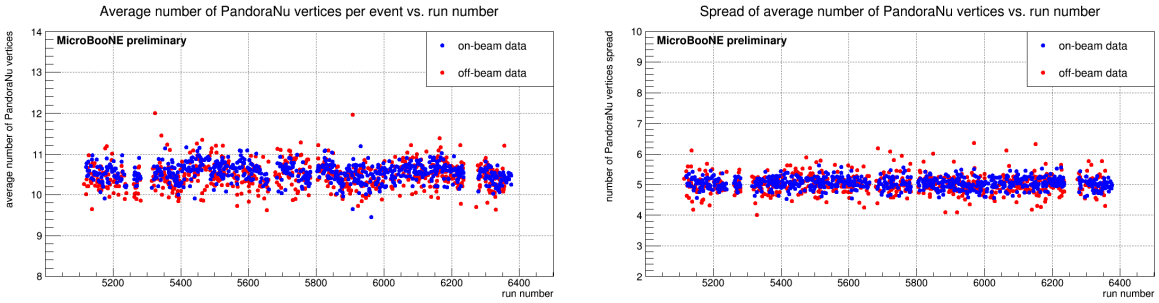


Figure A.2: Average number of PandoraNu vertices per event vs. run number with selected stable runs (left) and spread of number of PandoraNu vertices vs. run number with selected stable runs (right). Figures from Ref. [64].

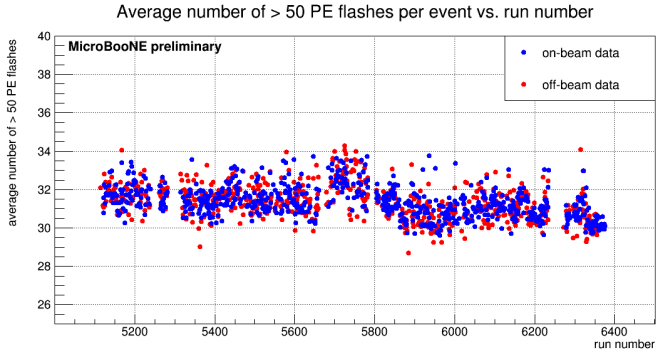


Figure A.3: Average number of $>$ 50 PE flashes per event vs. run number with selected stable runs. Figure from Ref. [64].

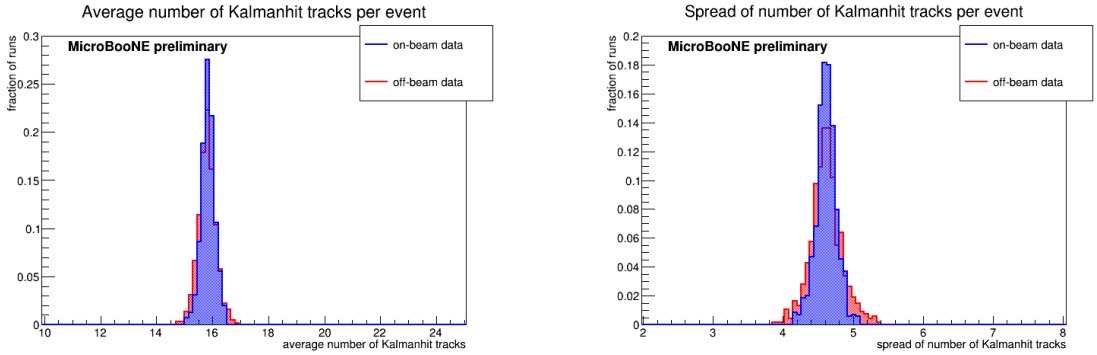


Figure A.4: Average number of KalmanHit tracks per event with selected stable runs (left) and spread of number of KalmanHit tracks with selected stable runs (right). Figures from Ref. [64].

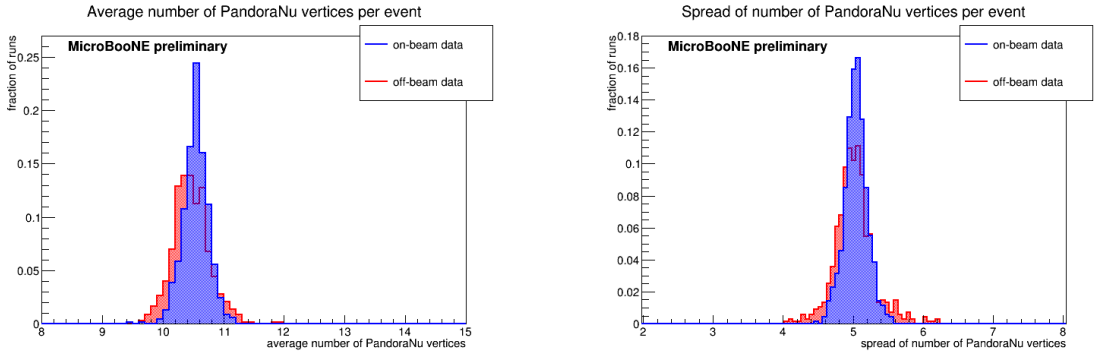


Figure A.5: Average number of PandoraNu vertices per event with selected stable runs (left) and spread of number of PandoraNu vertices with selected stable runs (right). Figures from Ref. [64].

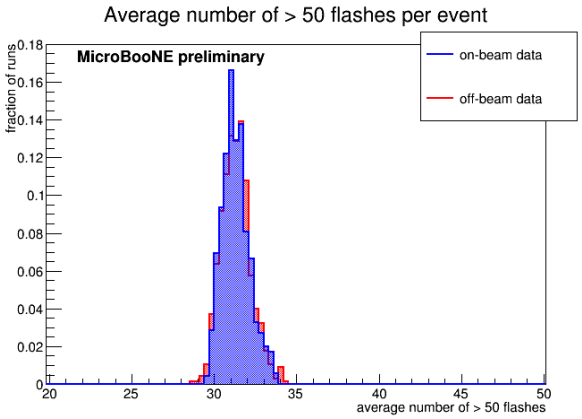


Figure A.6: Average number of $>$ 50 PE flashes per event with selected stable runs. Figure from Ref. [64].