DISSERTATION

MUON NEUTRINO RECONSTRUCTION WITH MACHINE-LEARNING TECHNIQUES AT THE ICARUS DETECTOR

Submitted by

Justin J. Mueller

Department of Physics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2024

Doctoral Committee:

Advisor: Michael Mooney

John Harton
Alexander Brandl
Samuel Brewer
Kazuhiro Terao

ABSTRACT

MUON NEUTRINO RECONSTRUCTION WITH MACHINE-LEARNING TECHNIQUES AT THE ICARUS DETECTOR

The ICARUS T600 LArTPC detector successfully ran for three years at the underground LNGS laboratories, providing a first sensitive search for LSND-like anomalous electron neutrino appearance in the CNGS beam. After a significant overhauling at CERN, the T600 detector has been placed in its experimental hall at Fermilab, fully commissioned, and the first events observed with full detector readout. Regular data-taking began in May 2021 with neutrinos from the Booster Neutrino Beam (BNB) and neutrinos six degrees off-axis from the Neutrinos at the Main Injector (NuMI). Modern developments in machine learning have allowed for the development of an end-to-end machine-learning-based event reconstruction for ICARUS data. This reconstruction folds in 3D voxel-level feature extraction using sparse convolutional neural networks and particle clustering using graph neural networks to produce outputs suitable for physics analyses. The analysis presented in this thesis demonstrates a high-purity and high-efficiency selection of muon neutrino interactions in the BNB suitable for the physics goals of the ICARUS experiment and the Short-Baseline Neutrino Program.

ACKNOWLEDGEMENTS

Throughout my PhD career I have been fortunate to have the support of many people. Early in my time at CSU, I became acquainted with my advisor Mike Mooney, who has been a constant source of guidance and support. It became clear to me early on that Mike was not only a great scientist, but also an excellent mentor with a genuine interest in the success of his students and the ability to pick out student projects that are both interesting and impactful. I attribute much of my success and growth as a scientist to Mike's continued mentorship and guidance. I would also like to thank my committee members, John Harton, Alex Brandl, Sam Brewer, and Kazuhiro Terao, for their time and feedback on my research. In particular, Kazu has been a great source of advice on this analysis and has provided a unique perspective on many aspects of the work.

My classmates at CSU, particularly Ryan, Derek, and David, have also been a great source of support and camaraderie. They were all instrumental in helping me navigate the early challenges of graduate school and have been great friends throughout my time at CSU. I will carry many fond memories of our interactions and shared experiences with me as I move forward in my career. After moving to Fermilab to work on the commissioning of the ICARUS detector, I was introduced to several people who made my transition to the lab smooth and enjoyable. Bruce Howard was one of the first people I met and worked with at the lab during the ICARUS commissioning and was always willing to help me with any questions I had or assist in hardware work when needed. Jacob Zettlemoyer became a close friend and colleague during my time at Fermilab and I attribute many of my successes to conversations with him. Outside of work, our weekly bowling nights have been a great source of fun and relaxation.

Many ICARUS collaborators have been instrumental in my success. During my hardware work, I have had the pleasure of working with and learning from many experts in the field, including Claudio Montanari, Angela Fava, Alberto Guglielmi, and Gabrielle Rampazzo. Donatella Torretta, Geoff Savage, and Bill Badgett provided significant guidance and operational support during many of my early hardware commissioning projects on ICARUS. Undoubtedly, the com-

puting support I received by the ICARUS production team and Vito Di Benedetto was instrumental in getting the files necessary for my analysis. Many other ICARUS collaborators, who are too numerous to list here, have provided support either directly or indirectly, and I am grateful for their contributions to the success of the ICARUS experiment.

Within the Mooney group at CSU, I have been fortunate to work with many talented and dedicated students, including Ivan Caro Terrazas, Lane Kashur, Dan Carber, Matt Siden, Ryan LaZur, and others. Lane in particular has paved the way for much of the work presented in this thesis while working on a complementary analysis. I have learned a great deal my fellow students and have enjoyed the camaraderie and shared experiences that come with working in a group. The SLAC group, Kazuhiro Terao, Francois Drielsma, Laura Domine, Dae Heun Koh, and Yeon-jae Jwa, have been valuable in the development of the reconstruction used in this analysis. Many of the ideas and techniques used in this analysis were developed in collaboration with the SLAC group, and they have all put in many hours of work to make this analysis possible. Their dedication and passion for seeing this analysis through to completion has been inspiring and I am grateful for the example they have set.

Without my family, I would not have been able to achieve the success I have had in my academic career. My parents, Jamie and Shannon, and my siblings Kaylyn, Nathan, and Mckenna, have never wavered in their support of my academic pursuits and ambitions even when it meant moving across the country to pursue my PhD. Moreover, many other family members have been a source of support and encouragement throughout my academic career, including my grandparents, aunts, uncles, and cousins. Truly, I would not be where I am today without the supportive environment that my family has provided me. Finally, I would like to thank my significant other, Cyrina, for her unwavering support and encouragement throughout the entirety of my PhD even while pursuing her own PhD.

I look forward to taking the skills and knowledge I have gained during my time at CSU and Fermilab and applying them to new challenges in the future. Many of the people who have been

instrumental in my success will remain close collaborators and friends as I move forward. I am grateful for the opportunities I have had and look forward to this next step in my career at Fermilab.

DEDICATION

To my family and friends who have supported me throughout my academic career.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Overview of Neutrino Physics

This chapter provides a historical introduction to the neutrino and its role in the Standard Model. Section 1.1 discusses the history leading to the discovery of the neutrino. An overview of the neutrino and its place in the Standard Model is provide in Section 1.2. Section 1.3 then presents an overview of neutrino interactions with matter including coherent elastic neutrino-nucleus scattering, quasielastic scattering, resonant interactions, and deep inelastic scattering. Section 1.4 introduces the phenomenon of neutrino oscillations and the implications for neutrino physics. Finally, Section 1.5 discusses the modification to neutrino oscillations in the minimal extension of the three-neutrino model to include a single sterile neutrino.

1.1 History of the Neutrino

The history of the neutrino begins in the early twentieth century and touches upon some of the most profound mysteries of the universe. As physicists delved deeper into the structure of matter, they encountered perplexing phenomena that existing theories struggled to explain.

The study of beta decay, a process by which a neutron is converted to a proton and an electron is emitted from the nucleus, provided the first evidence for the existence of the neutrino. At the time, beta decay was expected to behave as a two-body decay in which the outgoing electron carried a distinct energy equal to the difference between the initial and final nuclear states. The study of alpha and gamma decays, involving the emission of a helium nuclei and gamma ray, respectively, each revealed a narrow energy spectrum for the outgoing particle that is consistent with a two-body decay. In the early 1900s, however, several experiments studying beta decay found spectra that were not monoenergetic. In 1914, James Chadwick demonstrated that the beta decay spectrum was continuous [17], an apparent contradiction to the law of conservation of energy. Moreover, experiments showed that beta decay changed the nuclear spin of an atom by an integer amount

rather than the spin $\frac{1}{2}$ of the electron. This observation conflicted with the then-established law of conservation of angular momentum.

In 1930, the physicist Wolfgang Pauli proposed the existence of a new particle to resolve this open problem [18]. In a letter addressed to his ”radioactive ladies and gentlemen," Pauli posited a neutral, nearly massless particle that would carry away the missing energy and momentum in beta decays. This hypothetical particle would thus preserve the principles of conservation of energy and momentum. He referred to this particle as the "neutron," a name that was later adopted for the massive, chargeless particle discovered by James Chadwick in 1932. The name "neutrino," meaning "little neutral one" in Italian, was later coined by Enrico Fermi to distinguish Pauli's particle from Chadwick's neutron.

The existence of the neutrino remained purely theoretical until the mid-twentieth century when experimentalists began to seek evidence of its presence. In the 1950s, a series of experiments led by Frederick Reines and Clyde Cowan successfully detected neutrinos emitted from a nuclear reactor, providing the first experimental confirmation of Pauli's elusive particle [19]. These experiments earned Reines the 1995 Nobel Prize in Physics and secured the neutrino's place amongst the other fundamental particles. Subsequent experiments further explored the properties of neutrinos, including their interactions with matter and the phenomenon of neutrino oscillations. The study of neutrinos has since become a cornerstone of modern particle physics, providing a window into the fundamental forces and particles that govern the universe.

1.2 Neutrinos in the Standard Model

In the present day, we understand the neutrino as belonging to a class particles considered fundamental, meaning that they are not known to be composed of smaller constituents. The theoretical framework that describes the neutrino, the other fundamental particles, and the forces that govern their interactions is known as the Standard Model of particle physics. The Standard Model is a quantum field theory that describes the electromagnetic, weak, and strong forces as mediated by gauge bosons. A summary of the particles and forces in the Standard Model is shown in Figure

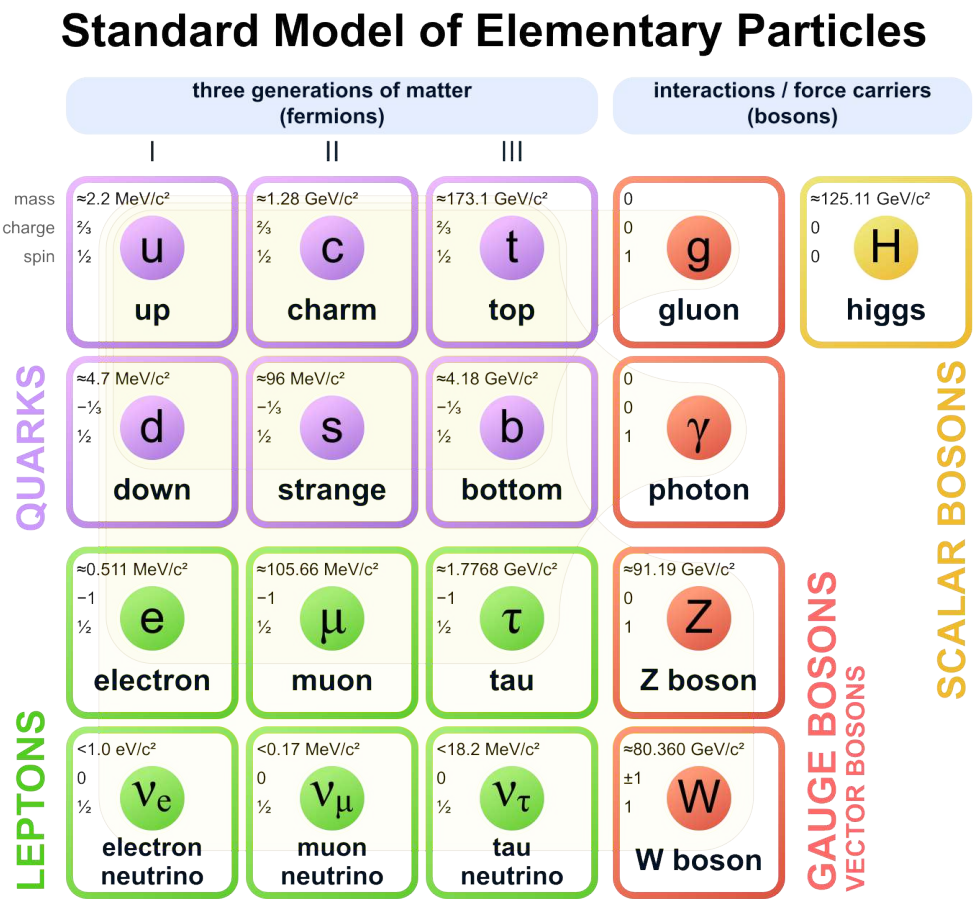


Figure 1.1: The particles and forces of the Standard Model of particle physics. Image from Wikimedia Commons.

1.1. Bosons are defined as particles with integer spin. The photon and gluon, which mediate the electromagnetic force and strong nuclear force respectively, are massless bosons with spin 1. The Z and $W^{\pm}$ bosons, which mediate the weak force and therefore neutrino interactions with matter, are massive bosons with spin 1. Though not a force carrier, the Higgs boson, which is responsible for the mass of the other particles via the Higgs Mechanism [20], is a massive boson with spin 0. Although a quantum theory of gravity is not included in the Standard Model, a hypothetical gauge boson with spin 2, known as the graviton, is predicted to mediate the gravitational force.

Matter is composed of the fundamental particles from the fermion sector of the Standard Model. Fermions are characterized by half-integer spin and are divided into two categories: quarks and leptons. These are further subdivided into three generations, each containing two quarks and

two leptons. The quarks are the up (u), down (d), charm (c), strange (s), top (t), and bottom (b) quarks. Pairwise, these represent the first, second, and third generations of quarks respectively. The lepton sector consists of the electron (e), muon (μ), tau (τ), electron neutrino (ν_e), muon neutrino (ν_μ), and tau neutrino (ν_τ). These represent the first, second, and third generations of charged leptons and their associated neutrino respectively. Each of these particles has an associated antiparticle with the same mass and opposite charge, typically denoted by a bar over the particle symbol. The antiparticle of the electron neutrino, for example, is denoted by $\bar{\nu}_e$.

In the context of the Standard Model, the neutrino is a massless, elusive particle that only interacts via the weak nuclear force, which as the name suggests has significantly weaker couplings to matter than the electromagnetic and strong nuclear forces. The example oft-repeated to set the scale of the weak force is that a neutrino can pass through a light-year of lead with roughly equal chance of interacting as not interacting. This weak interaction with matter makes neutrinos difficult to detect, but physicists have made great efforts to advance detector technology and experimental techniques to study neutrinos since the detection of the first (anti)neutrino in 1956 by Cowan and Reines and the subsequent first observation of a neutrino interaction in 1970 at Argonne National Laboratory (Figure 1.2).

The massless description of neutrinos in the Standard Model was initially motivated by the fact that neutrinos have only been observed with left-handed helicity [21]. The helicity of a particle is determined by the direction of a particle's spin relative to its momentum. The dependence of helicity on the particle's momentum means that a change of reference frame can cause the helicity of a particle to change; that is, helicity is not a Lorentz invariant quantity. In contrast, the chirality of a particle is a Lorentz invariant quantity that describes the "handedness" of a particle and is equal to the helicity for massless particles. The left-handed chirality of neutrinos is a consequence of the weak force's coupling to only left-handed particles and right-handed antiparticles. The non-observation of neutrinos with right-handed helicity in weak interactions led to the assumption that neutrinos are massless as only massless particles can be strictly left-handed.

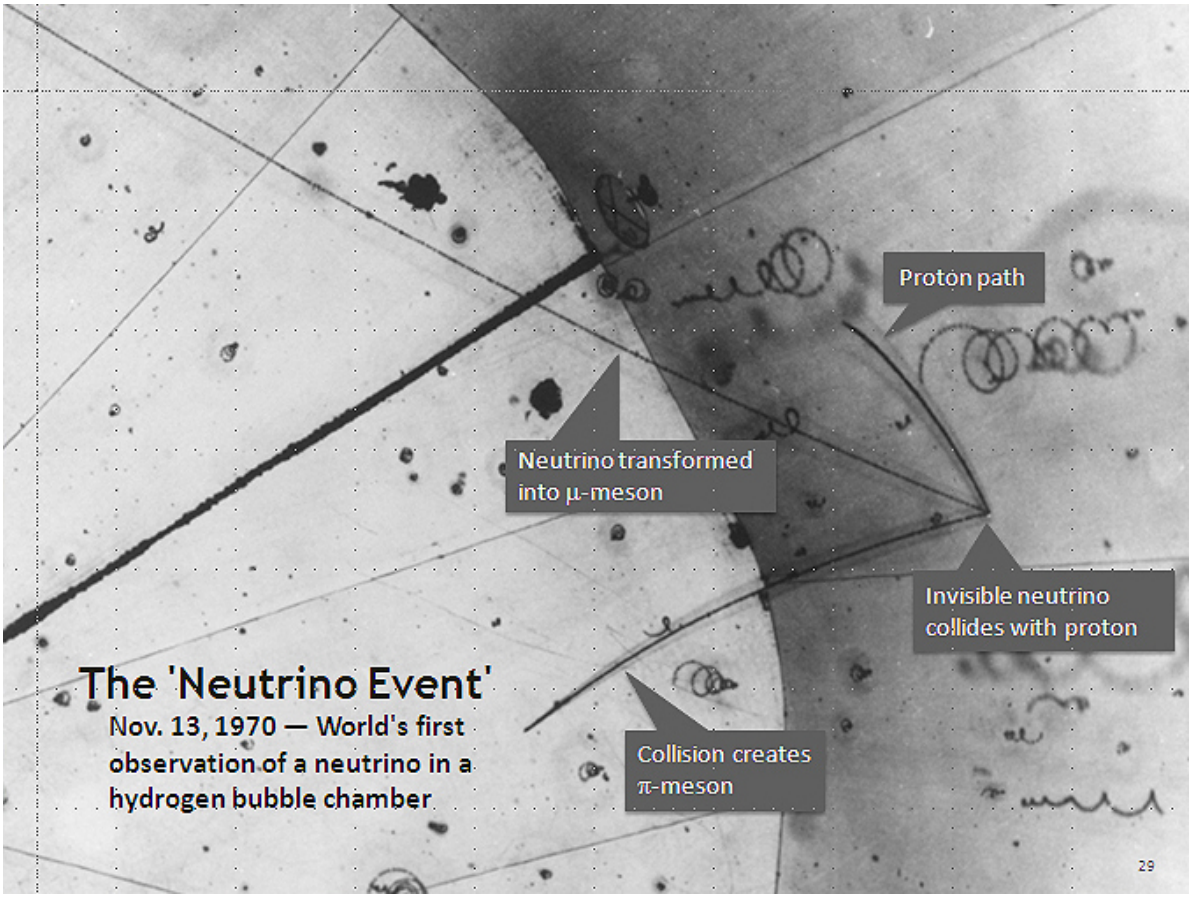


Figure 1.2: The first observation of a neutrino interaction in a 12-foot hydrogen bubble chamber at Argonne National Laboratory in 1970. The neutrino enters from the right colliding with a proton and producing a muon, pion, and recoiling proton. Each outgoing particle is visible as a track in the image coalescing at the interaction vertex.

Despite these assumptions inherent in the Standard Model, recent experimental evidence has confirmed that neutrinos do in fact have mass, albeit sufficiently small to make helicity and chirality approximately equal. The discovery of neutrino oscillations, a phenomenon in which neutrinos change flavor as they propagate through space, has provided the first direct evidence of neutrino mass. Direct measurements of the neutrino mass have also been made and have most recently set a upper limit on the neutrino mass of $0.8\ \mathrm{eV/c^2}$ [22]. The discovery of neutrino mass has profound implications for the Standard Model and has motivated the development of new theories to explain the origin of neutrino mass and the nature of neutrino oscillations.

1.3 Neutrino Interactions

As discussed in the previous section, neutrinos interact with matter via the weak nuclear force, which is mediated by the exchange of the charged $W^{\pm}$ and the neutral Z bosons. These give rise to the corresponding Charged Current (CC):

$$\begin{aligned} \nu_l + A &\rightarrow l^- + X \\ \bar{\nu_l} + A &\rightarrow l^+ + X \end{aligned} \tag{1.1}$$

and Neutral Current (NC) interactions:

$$\begin{aligned} \nu_l + A &\rightarrow \nu_l + X \\ \bar{\nu_l} + A &\rightarrow \bar{\nu_l} + X \end{aligned} \tag{1.2}$$

where $l \in \{e, \mu, \tau\}$ is used to represent the lepton flavor, A represents the nuclear target, such as a quark, nucleon, or atom, and X represents any number of final-state particles resulting from the interaction. It is worth noting that CC interactions provide a visible indication of the flavor of the parent neutrino in the form of the outgoing charged lepton. This makes CC interaction signals of particular interest to neutrino experiments.

Neutrino interactions can be further subdivided by the size of the nuclear target and the type of interaction produced. A diagram from [2] showing the total neutrino and antineutrino production rates for different interaction types is shown in Figure 1.3. Below 50 MeV neutrinos coherently scatter off the entire nucleus in an elastic collision. This process, known as coherent elastic neutrino-nucleus scattering (CEνNS), produces a faint interaction signature of a recoiling nucleus and was only recently detected in 2017 by COHERENT [23]. At higher energies, neutrinos can interact with individual nucleons in the nucleus. The most common interaction below about 1 GeV is quasielastic (QE) scattering, in which the neutrino interacts with a single nucleon producing a charged lepton and ejecting a nucleon. Though quite simple in nature, the challenge in identifying QE interactions lies in the fact that other interaction types may produce similar visible final-state particles.

QE interactions remain dominant up to about 1 GeV; however, at around 0.5 GeV the energy of the interaction becomes sufficient to produce resonant baryonic states, e.g. a delta baryon Δ. The excited resonant state then decays quickly into some final state, typically a nucleon and a single pion, but occasionally the decay may emit multiple pions, a kaon, some other meson, or a photon [2]. This process is dominant from about 1 GeV to 4 GeV. These energies are also sufficient for a process called coherent inelastic neutrino-nucleus scattering, which produces a pion with a particularly forward-scattered distribution [24].

Beginning around 2 GeV and becoming dominant above 4 GeV, the neutrino can resolve and interact with individual quarks directly. Known as deep inelastic scattering (DIS), this process involves high momentum transfer and can break apart the nucleon, often resulting in a large number of hadronic particles. The degree of complexity in the final state makes DIS interactions difficult to identify and reconstruct.

The landscape of neutrino interactions presents a complex challenge for experimentalists. Often, it is desirable to separate interaction types and characterize them separately or leverage the known final states for more precise energy reconstruction. The reality is that a single exclusive final state of a neutrino interaction often reflects a complex mixture of processes by which that

final state may be produced. For this reason an inclusive selection, consisting of all the different interaction channels as shown in Figure 1.3, is often chosen for analyses which are agnostic to the underlying interaction channel. This has the benefit of reducing the impact of systematic uncertainties associated with the modeling of neutrino-nucleus interactions and of increasing the statistics of the sample and usage fraction of the data. The inclusive channel is therefore often the most sensitive channel for oscillation measurements.

1.4 Neutrino Oscillations

Neutrino oscillations were first predicted by Bruno Pontecorvo in 1959 [25], though this prediction was restricted to ν_e and ν_μ oscillations as the known neutrinos at the time. Neutrino oscillations were first observed by Ray Davis in the 1960s in the Homestake experiment [26] as a discrepancy between the predicted and observed solar neutrino flux. Conclusive proof that this discrepancy was due to neutrino oscillations would not come until the Super-Kamiokande [27] and Sudbury Neutrino Observatory [28] collaborations reported their results in the late 1990s and early 2000s.

Although the exact mechanism responsible for generating the neutrinos mass is not yet known, the phenomena of neutrino flavor oscillations generated by the neutrino masses is well-modeled and understood. Neutrino flavor oscillations are a quantum mechanical effect that arise from the non-orthogonality of the three flavor states of the weak interaction $\nu_\alpha \in \{\nu_e, \nu_\mu, \nu_\tau\}$ and the three mass eigenstates of the Hamiltonian $\nu_j \in \{\nu_1, \nu_2, \nu_3\}$. The consequence of this non-orthogonality is that a neutrino produced in a flavor eigenstate is equivalently a superposition of all possible mass eigenstates. The relation between the flavor and mass eigenstates is given by the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix, U [29]:

$$|\nu_\alpha\rangle = \sum_j U_{\alpha j} |\nu_j\rangle \tag{1.3}$$

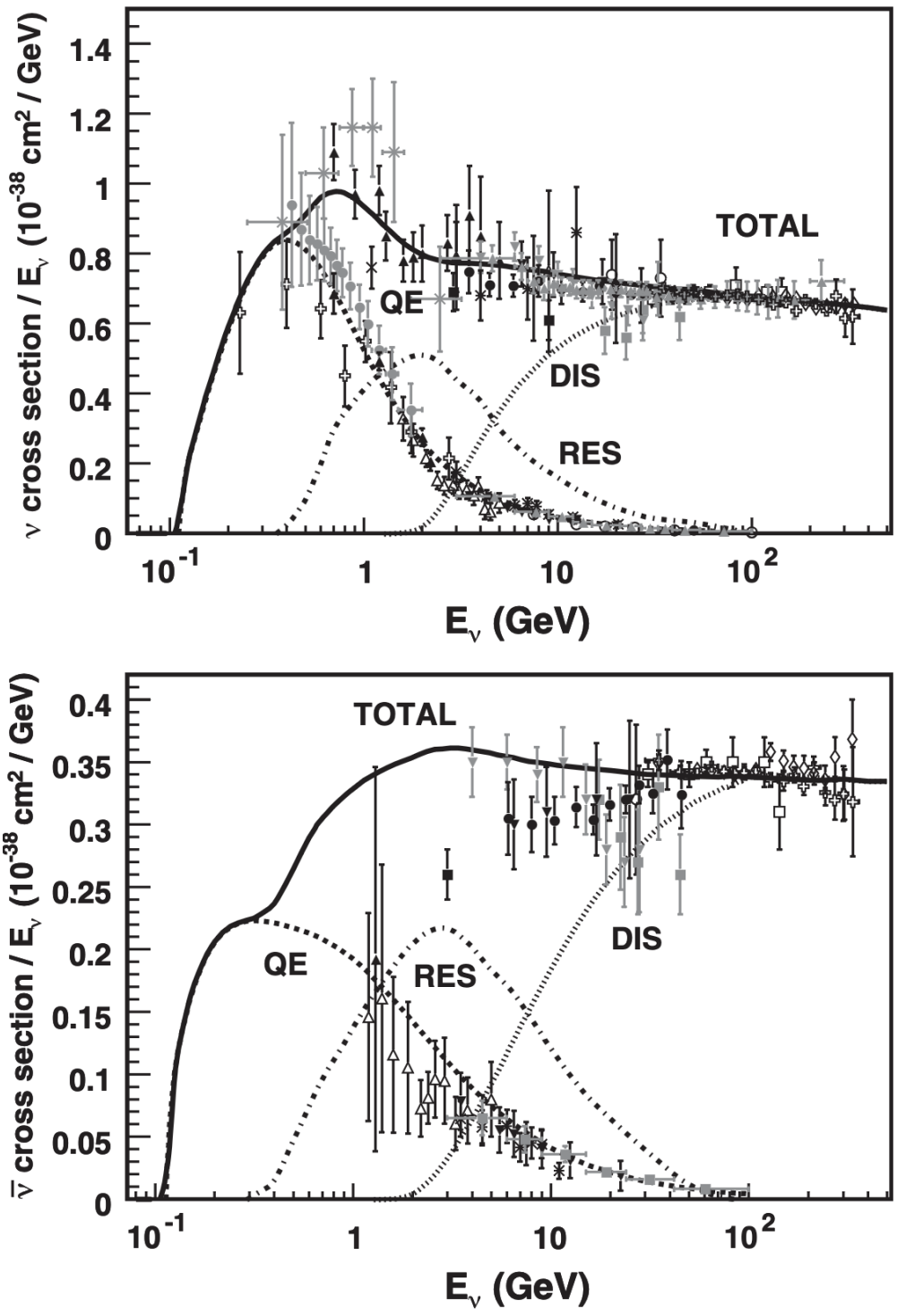


Figure 1.3: The total neutrino and antineutrino per nucleon CC cross sections divided by neutrino energy for different interaction types as a function of neutrino energy. Example predictions for each are provided by the NUANCE generator. The figure is taken from [2].

$$|\nu_j\rangle = \sum_\alpha U^*_{\alpha j}|\nu_\alpha\rangle \tag{1.4}$$

For three neutrino flavors, the PMNS matrix is a 3×3 unitary matrix describing a rotation from the flavor basis to the mass basis and parameterized by nine degrees of freedom. Five of these degrees of freedom can be absorbed into the phases of the particle fields, leaving a degree of freedom for the overall phase and three mixing angles describing the rotation [30]. The mixing angles are traditionally denoted by θ_{12}, θ_{13}, and θ_{23}, and the overall phase is denoted by δ_{CP}. The parameter δ_{CP} controls the degree of "charge-parity" symmetry violation and will be discussed later. The PMNS matrix can be written in terms of these parameters as:

$$\begin{aligned} U &= \begin{bmatrix} U_{11} & U_{12} & U_{13} \\ U_{21} & U_{22} & U_{23} \\ U_{31} & U_{32} & U_{33} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{CP}} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{CP}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{CP}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{CP}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{CP}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{CP}} & c_{23}c_{13} \end{bmatrix} \end{aligned} \tag{1.5}$$

Here $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$. As a consequence of the weak interaction's coupling to flavor eigenstates, neutrinos are produced and detected in flavor eigenstates. Their time evolution in a vacuum, or translation from the production to detection point, is governed by the free particle Hamiltonian and therefore depends on their mass eigenstates. The total energy of a mass eigenstate (E_j) can be approximated as:

$$E_j = \sqrt{p^2 + m_j^2} \approx p_j + \frac{m_j^2}{2p_j} \approx E + \frac{m_j^2}{2E} \tag{1.6}$$

where p_j is the momentum of the mass eigenstate, m_j is the mass of the mass eigenstate, and E is the kinetic energy of the neutrino. It is important to note that $c = 1$ in this notation. The approximation is valid for ultra-relativistic neutrinos, with their momentum p_j much larger than their mass m_j. Owing to their extremely small mass, this is valid for the three Standard Model neutrinos. Using this approximation, the time evolution of the jth mass eigenstate can be written as:

$$|\nu_j(t)\rangle = e^{-iHt} \approx e^{-i\left(E+\frac{m_j^2}{2E}\right)t}|\nu_j(0)\rangle \approx e^{-i\frac{m_j^2}{2E}t}|\nu_j(0)\rangle \approx e^{-i\frac{m_j^2}{2E}L}|\nu_j\rangle \tag{1.7}$$

where the relativistic assumption $L \approx t$ replaces the unobserved travel time in the neutrino frame with the measurable distance between the production and detection points. The overall common phase of e^{-iEL} is ignored, as it is a global phase that does not affect the probability of the neutrino oscillating between flavors. The probability of a neutrino produced in a flavor eigenstate $|\nu_\alpha\rangle$ being detected in a flavor eigenstate $|\nu_\beta\rangle$ is given by:

$$\begin{aligned}
P_{\nu_\alpha\to\nu_\beta}(L,E) &= |\langle\nu_\beta \,|\, \nu_\alpha(L)\rangle|^2 \\
&= \left|\langle\nu_\beta|\sum_j U_{\alpha j}|\nu_j(L)\rangle\right|^2 \\
&= \left|\langle\nu_\beta|\sum_j U_{\alpha j}e^{-i\frac{m_j^2 L}{2E}}|\nu_j(0)\rangle\right|^2 \\
&= \left|\langle\nu_\beta|\sum_{j,k} U_{kj}^* U_{\alpha j}e^{-i\frac{m_j^2 L}{2E}}|\nu_k\rangle\right|^2 \\
&= \left|\sum_{j,k} U_{\beta j}^* U_{\alpha j}e^{-i\frac{m_j^2 L}{2E}}\right|^2 \\
&= \sum_j\sum_k U_{\beta j}^* U_{\alpha j} U_{\alpha k}^* U_{\beta k}e^{-i\frac{\Delta m_{jk}^2 L}{2E}}
\end{aligned} \tag{1.8}$$

where $\Delta m_{jk}^2 = m_j^2 - m_k^2$ is the difference in mass squared between the jth and kth mass eigenstates. This expression already makes clear the dependence of the oscillation probability on the mass differences, and that the transition probability reduces to $\delta_{\alpha\beta}$ in the case of degenerate neu-

trino masses, therefore producing no oscillations in that case. The double sum can be split into a sum where $j = k$ and a sum where $j \neq k$, where the latter can be further simplified using the unitarity of U and recognizing that only the real portion of the sum survives due to $U_{ab} = U^*_{ba}$:

$$P_{\nu_\alpha \to \nu_\beta}(L, E) = \sum_j |U_{\alpha j}|^2 |U_{\beta j}|^2 + 2Re\left(\sum_j \sum_{k>j} U^*_{\beta j} U_{\alpha j} U^*_{\alpha k} U_{\beta k} e^{-i\frac{\Delta m^2_{jk} L}{2E}}\right) \tag{1.9}$$

If we consider the limiting case of $L = 0$, we obtain the useful relation:

$$\delta_{\alpha\beta} = \sum_j |U_{\alpha j}|^2 |U_{\beta j}|^2 + 2Re\left(\sum_j \sum_{k>j} U^*_{\beta j} U_{\alpha j} U^*_{\alpha k} U_{\beta k}\right) \tag{1.10}$$

We can then use this relation and begin to pick out the real terms which contribute to the oscillation probability and reduce the expression using some trigonometric identities. The final expression for the oscillation probability is then given by:

$$\begin{aligned}
P_{\nu_\alpha \to \nu_\beta}(L, E) &= \delta_{\alpha\beta} + 2Re\left(\sum_j \sum_{k>j} U^*_{\beta j} U_{\alpha j} U^*_{\alpha k} U_{\beta k}\left(e^{-i\frac{\Delta m^2_{jk} L}{2E}} - 1\right)\right) \\
&= \delta_{\alpha\beta} + 2\sum_j \sum_{k>j} Re\left(U^*_{\beta j} U_{\alpha j} U^*_{\alpha k} U_{\beta k}\right)\left(\cos\frac{\Delta m^2_{jk} L}{2E} - 1\right) \\
&\quad + 2\sum_j \sum_{k>j} Im\left(U^*_{\beta j} U_{\alpha j} U^*_{\alpha k} U_{\beta k}\right)\sin\frac{\Delta m^2_{jk} L}{2E} \\
&= \delta_{\alpha\beta} - 4\sum_j \sum_{k>j} Re\left(U^*_{\beta j} U_{\alpha j} U^*_{\alpha k} U_{\beta k}\right)\sin^2\frac{\Delta m^2_{jk} L}{4E} \\
&\quad + 2\sum_j \sum_{k>j} Im\left(U^*_{\beta j} U_{\alpha j} U^*_{\alpha k} U_{\beta k}\right)\sin\frac{\Delta m^2_{jk} L}{2E}
\end{aligned} \tag{1.11}$$

For antineutrinos, the oscillation probability $P_{\bar{\nu_\alpha} \to \bar{\nu_\beta}}$ can be found by replacing U with U^* in the above expression. From Equation 1.11, we can see that the difference in oscillation probability for neutrinos and antineutrinos is given by:

$$P_{\nu_\alpha \to \nu_\beta}(L, E) - P_{\bar{\nu_\alpha} \to \bar{\nu_\beta}}(L, E) = 4\sum_j \sum_{k>j} Im\left(U^*_{\beta j} U_{\alpha j} U^*_{\alpha k} U_{\beta k}\right)\sin\frac{\Delta m^2_{jk} L}{2E} \tag{1.12}$$

Referring back to Equation 1.5, we can see that the only parameter that can produce an imaginary component to U and therefore a nonzero difference in neutrino and antineutrino oscillation probability is δ_{CP}. This is known as the charge-parity (CP) violating phase, and from this result we can see that it alone controls the degree of CP violation in the lepton sector. The presence of CP violation in the lepton sector is a major open question in particle physics, and the measurement of the CP-violating phase is a major goal of current and future neutrino experiments.

There are several important features of neutrino oscillations that merit summarizing. First, much of this derivation can be generalized to further neutrino flavor and mass eigenstates with the addition of new parameters. Under the three-flavor description, neutrino mixing is modeled through the PMNS matrix U with four parameters: three mixing angles θ_{12}, θ_{13}, and θ_{23}, and the CP-violating phase δ_{CP}. The oscillation probability also contains an explicit dependence on the mass differences between eigenstates, which adds two more independent parameters to the model. Though $\Delta m^2_{jk} = m^2_j - m^2_k$ is a signed quantity, the dominant oscillation term $\sin^2 \frac{\Delta m^2_{jk} L}{4E}$ carries no sign information and is strictly positive. The last term in Equation 1.11 does allow for sign information to be carried by the oscillation probability, but is only nonzero in the presence of CP violation. This means that experimental sensitivity to the ordering of the mass eigenstates through neutrino oscillations is dependent on CP violation. It is worth noting, however, that matter effects can also generate a potential (due to the presence of electrons in matter and the charged-current elastic forward scattering process available to electron neutrinos) that shifts the value of the nu_e mass eigenstate, and therefore the oscillation probability, in a way that is dependent on the mass ordering [31].

From an experimental design perspective, the controllable parameters are the baseline L and the neutrino energy E. These parameters enter as a ratio in the oscillation probability, making it useful to describe oscillation experiments in terms of the ratio L/E. In accelerator-based neutrino experiments, the neutrino energy can be controlled to some extent by the design of the beamline, but otherwise varies quite dramatically between sources of reactor neutrinos, atmospheric neutrinos, and solar neutrinos. In all cases, precise reconstruction of the neutrino energy directly impacts

sensitivity to the oscillation parameters. The distance between the neutrino source and the detector is significantly easier to control for accelerator-based and reactor-based experiments, and is a key design consideration in an experiment. The degree to which the PMNS matrix has been measured has only been possible through careful and clever experiment designs that aim to isolate the parameters controlling neutrino oscillations.

1.5 Sterile Neutrinos

The previous section described neutrino oscillations in both the standard three-flavor scenario and the more general n-flavor scenario. The number of light active neutrino flavor states, that is, those which undergo electroweak interactions, has been determined to be three through precision measurements of the Z boson width at the Large Electron-Proton Collider (LEP) [3] as shown in Figure 1.4. This puts a strict limit on the number of light neutrino flavors, but does not preclude the existence of additional neutrino states that do not interact via the weak force. These states are known as sterile neutrinos.

Though the three-neutrino paradigm has been extremely successful in describing the results of neutrino oscillation experiments, there have been some recent anomalies which cannot be explained by this model. Discrepancies in the number of neutrinos observed in short-baseline experiments have revealed a tension between the three-neutrino model and the data, which may be explained by the addition of a new (sterile) neutrino with a mass around 1 eV modifying the oscillation probability at short baselines [32]. An overview of the short-baseline anomalies will be given in the next chapter.

A minimal extension to the three-neutrino model, called the 3+1 model, introduces a single sterile neutrino. This additional neutrino flavor state introduces a new mass splitting and extends the PMNS matrix by seven terms:

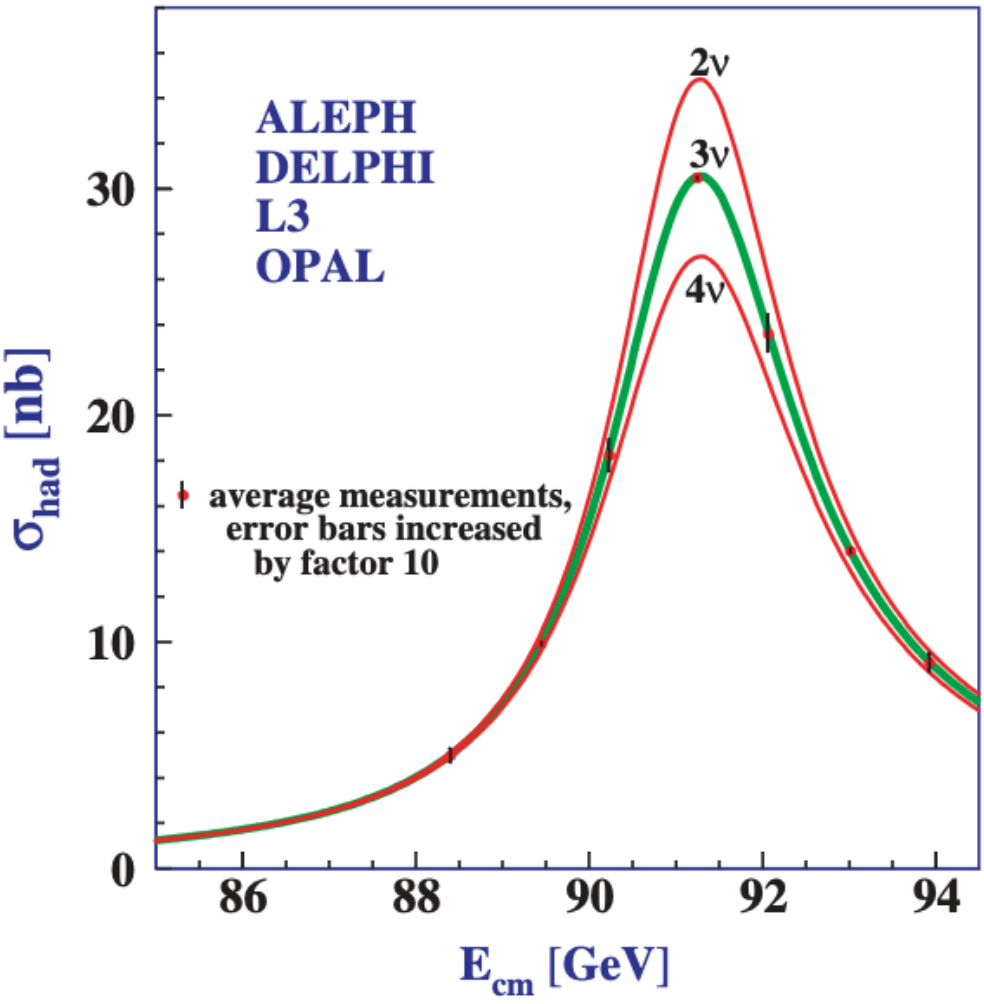


Figure 1.4: The measurement of the hadron cross section around the Z boson resonance as a function of the center-of-mass energy. Each line represents the prediction of the cross section given an expected number of neutrinos that the Z boson can decay into. The data (red points) agree well with the three-neutrino hypothesis. The figure is taken from [3].

$$U_{\alpha j} = \begin{bmatrix} U_{e1} & U_{e2} & U_{e3} & U_{e4} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} & U_{\mu 4} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} & U_{\tau 4} \\ U_{s1} & U_{s2} & U_{s3} & U_{s4} \end{bmatrix} \tag{1.13}$$

In this 3+1 extension of U, s corresponds to the sterile flavor eigenstate and a fourth neutrino mass eigenstate has been added. The matrix elements ($U_{s1}, ..., U_{s4}$) cannot be constrained through direct measurements due to the non-interacting nature of the sterile neutrino. This extension of the PMNS matrix also introduces three new neutrino mixing angles θ_{i4} and two new CP-violating phases. In the short-baseline regime, $E/L \ll \Delta m^2_{41}$, and with $\Delta m^2_{41} \gg \Delta m^2_{21}, \Delta m^2_{31}$, the oscillation probabilities are as follows [33]:

$$P_{\nu_\alpha \to \nu_\beta}(L, E) \approx \delta_{\alpha\beta} - 4 \left|U_{\alpha 4}\right|^2 \left(\delta_{\alpha\beta} - \left|U_{\beta 4}\right|^2\right) \sin^2 \frac{\Delta m^2_{41} L}{4E} \tag{1.14}$$

Accelerator and reactor neutrino sources are best suited for probing neutrino oscillations in this regime. These types of experiments operate at energies insufficient to produce tau particles due to their high rest mass of about 1.8 GeV, so we can consider purely ν_e and ν_μ sources and begin constraining the matrix elements relevant for these channels. From Equation 1.14, the oscillation probability for the available channels are given by:

$$P_{\nu_e \to \nu_e}(L, E) \approx 1 - 4 \left|U_{e4}\right|^2 \left(1 - \left|U_{e4}\right|^2\right) \sin^2 \frac{\Delta m^2_{41} L}{4E} \tag{1.15}$$

$$P_{\nu_\mu \to \nu_\mu}(L, E) \approx 1 - 4 \left|U_{\mu 4}\right|^2 \left(1 - \left|U_{\mu 4}\right|^2\right) \sin^2 \frac{\Delta m^2_{41} L}{4E} \tag{1.16}$$

$$P_{\nu_\mu \to \nu_e}(L, E) \approx 4 \left|U_{\mu 4}\right|^2 \left|U_{e4}\right|^2 \sin^2 \frac{\Delta m^2_{41} L}{4E} \tag{1.17}$$

The probabilities calculated allow for sensitivity to three of the parameters characterizing the 3+1 model: U_{e4}, $U_{\mu 4}$, and Δm^2_{41}. Equations 1.15 and 1.16 are disappearance channels, whereas

Equation 1.17 is an appearance channel. These channels will be discussed in more detail in the next chapter.

Chapter 2

Neutrino Experiments

Neutrinos are produced in a variety of processes in the universe, including nuclear reactions in stars, cascading particle showers from cosmic ray interactions in the atmosphere, beta decays during fission in nuclear reactors, and particle collisions in particle accelerators. Each of these sources has been leveraged by a variety of experiments to study the properties of neutrinos, including the parameters describing neutrino oscillations.

Sections 2.1 - 2.4 will briefly summarize the experiments carried out with each of these neutrino sources along with their contribution towards our understanding of neutrino oscillations as parameterized in Equation 1.5. Section 2.5 will then discuss the short-baseline anomalies observed in the LSND and MiniBooNE experiments, which have motivated the Short-Baseline Neutrino (SBN) Program.

2.1 Solar Neutrino Experiments

The Sun and other low-mass stars are predominately powered by the proton-proton (pp) chain, a process by which hydrogen is fused into helium. An alternative fusion process, the Carbon-Nitrogen-Oxygen (CNO) cycle, is also present in more massive stars with higher temperatures, though it represents only about 1% of the Sun's solar luminosity. Each of these processes produces electron neutrinos as a byproduct at several stages of the reaction chain, as shown in Figure 2.1. The primary solar neutrino flux component is the pp chain, which produces neutrinos with energies up to 400 keV. The CNO cycle produces neutrinos with energies up to 1.7 MeV. The later stages of the pp chain produce neutrinos with energies up to around 20 MeV. The breakdown of the solar neutrino flux as a function of neutrino energy is shown in Figure 2.2 [4].

Historically, solar neutrino experiments were the first to observe neutrino oscillations, beginning with the Homestake experiment in the 1960s. The Homestake experiment measured the number of $^{37}\mathrm{Ar}$ atoms produced by the interaction of solar neutrinos with chlorine in a large under-

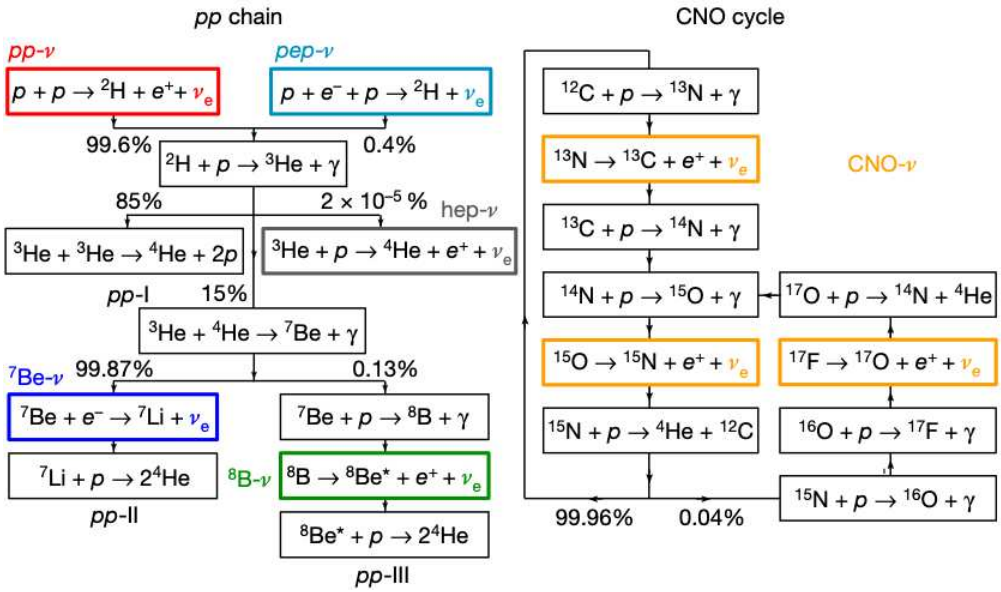


Figure 2.1: The solar neutrino production chain. The primary fusion reaction in the sun is the proton-proton chain (left), which produces electron neutrinos. The CNO cycle is a secondary fusion reaction that produces heavier elements (right) [4].

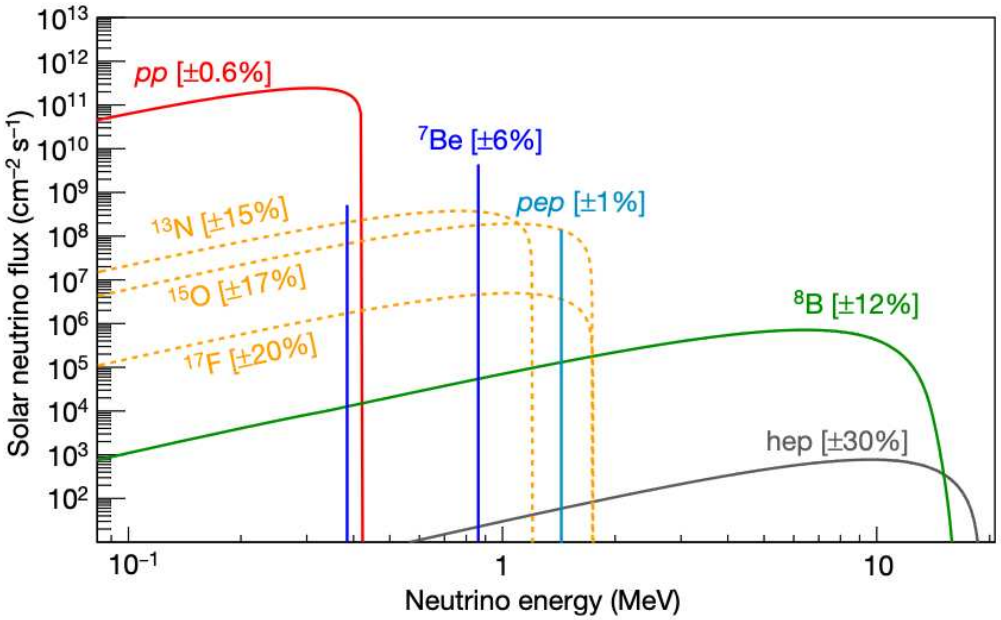


Figure 2.2: Breakdown of the solar neutrino flux as a function of neutrino energy for the different production modes (see Figure 2.1) [4].

ground tank through the reaction $\nu_e + {}^{37}\mathrm{Cl} \rightarrow e^- + {}^{37}\mathrm{Ar}$. The observed flux of solar neutrinos was found to be about one-third of the expected solar flux, a discrepancy that became known as the solar neutrino problem. Other experiments such as GALLEX, GNO, and SAGE also observed a deficit in the solar neutrino flux. The resolution of this problem would not come until the Sudbury Neutrino Observatory (SNO) experiment's measurement of the solar neutrino flux in 2001, which attributed this discrepancy to the phenomenon of neutrino oscillations [28].

Solar neutrinos are the best candidates to measure $\sin^2(\theta_{12})$ and Δm^2_{21}, as the oscillation length is well-matched to the solar radius. Several solar neutrino experiments have measured these parameters, including the Homestake experiment, GALLEX, SAGE, Borexino, Super-Kamiokande, and SNO. Recent global fit analyses have provided values of $\sin^2(\theta_{12}) = 0.303 \pm 0.012$ and $\Delta m^2_{21} = 7.41^{+0.21}_{-0.20}$ [1].

2.2 Atmospheric Neutrino Experiments

High-energy particles from space, mostly protons and atomic nuclei, are constantly bombarding Earth's atmosphere producing a cascade of secondary particles and neutrinos. The primary source of atmospheric neutrinos is the decay of charged pions and kaons produced in the hadronic showers initiated by cosmic rays, typically around 15 kilometers above the Earth's surface. The decay of these charged mesons produces muons and their corresponding neutrinos, where the muons subsequently decay to produce electrons, electron neutrinos, and muons neutrinos (Figure 2.3). The neutrinos produced in this cascade exhibit a wide range of energies from about 100 MeV to more than 10 TeV, though the flux decreases rapidly with increasing energy.

The Super-Kamiokande experiment in Japan was among the first to announce experimental evidence for the existence of neutrino oscillations [34]. The experiment measured the rate of atmospheric electron and muon neutrinos as a function of the zenith angle using a large water Cherenkov detector (Figure 2.4). A zenith angle of 0 degrees corresponds to neutrinos coming from directly overhead, while a zenith angle of 180 degrees corresponds to neutrinos coming from the opposite side of the Earth - effectively making the zenith angle a measurement of the neutrino

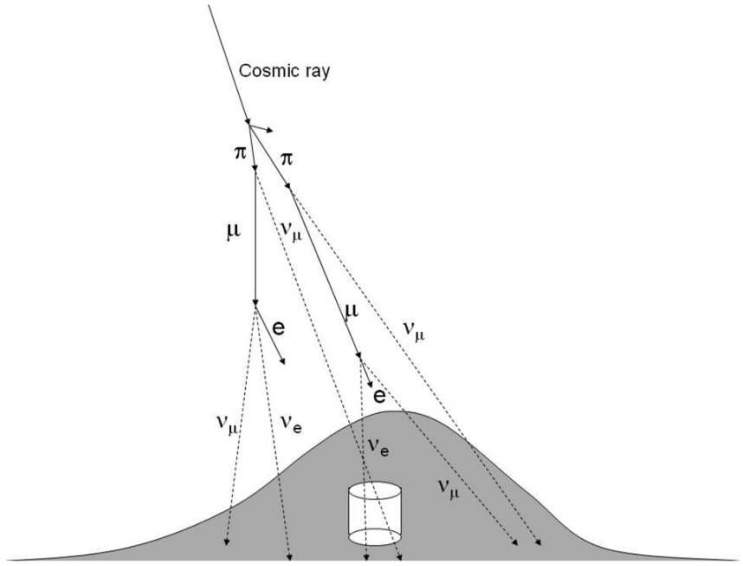


Figure 2.3: The atmospheric neutrino production chain. Cosmic rays produce charged pions and kaons, which subsequently decay to produce muons and neutrinos. The muons then decay to produce electrons, electron neutrinos, and muon neutrinos [5].

path length within the Earth. The Super-Kamiokande experiment observed a discrepancy between upward-going neutrinos and downward-going neutrinos, which was consistent with neutrino oscillations.

Atmospheric neutrinos provide excellent sensitivity to the parameter $\sin^2(\theta_{23})$. Global fits to atmospheric neutrino data from Super-Kamiokande, IceCube, and the upgraded IceCube DeepCore have provided values of $\sin^2(\theta_{23}) = 0.451^{+0.019}_{-0.016}$ ($0.569^{+0.016}_{-0.021}$) for the normal (inverted) neutrino mass ordering scenario [1].

2.3 Reactor-Based Neutrino Experiments

Nuclear reactors are an excellent source of electron antineutrinos, which are produced in the beta decay of fission products. The primary sources of reactor neutrinos in reactor neutrino experiments are the decays of $^{235}\mathrm{U}$, $^{238}\mathrm{U}$, $^{239}\mathrm{Pu}$, and $^{241}\mathrm{Pu}$ isotopes. The neutrinos produced in these decays have energies of a few MeV (Figure 2.5) and are detected through the inverse beta decay process $\bar{\nu}_e + p \rightarrow e^+ + n$, which produces electromagnetic showers via e^+e^- annihilation and neutron capture.

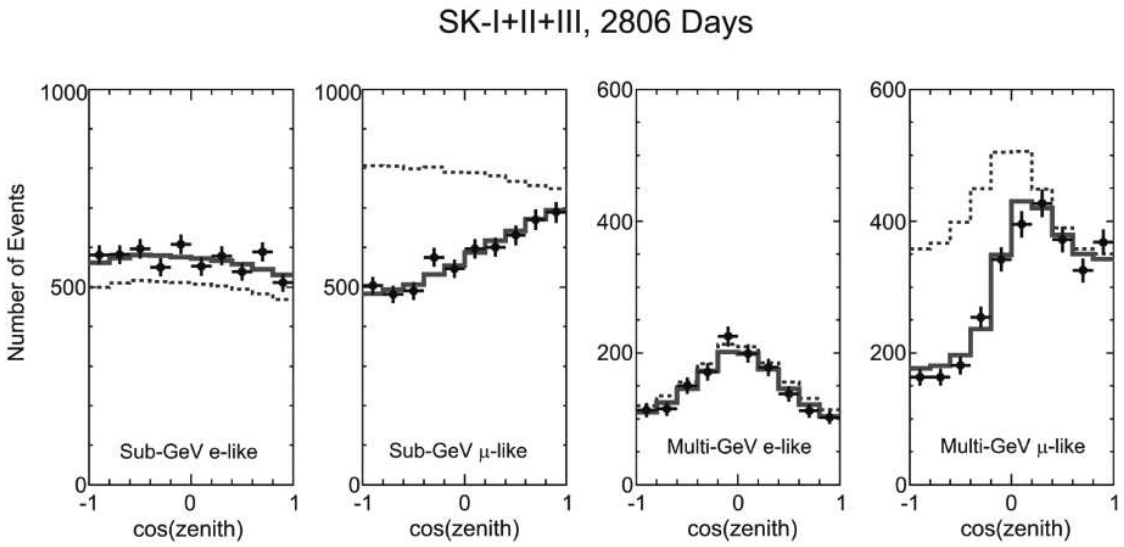


Figure 2.4: The Super-Kamiokande experiment measured the rate of atmospheric neutrinos as a function of the zenith angle. The data showed a discrepancy between upward-going neutrinos and downward-going neutrinos, consistent with neutrino oscillations [5].

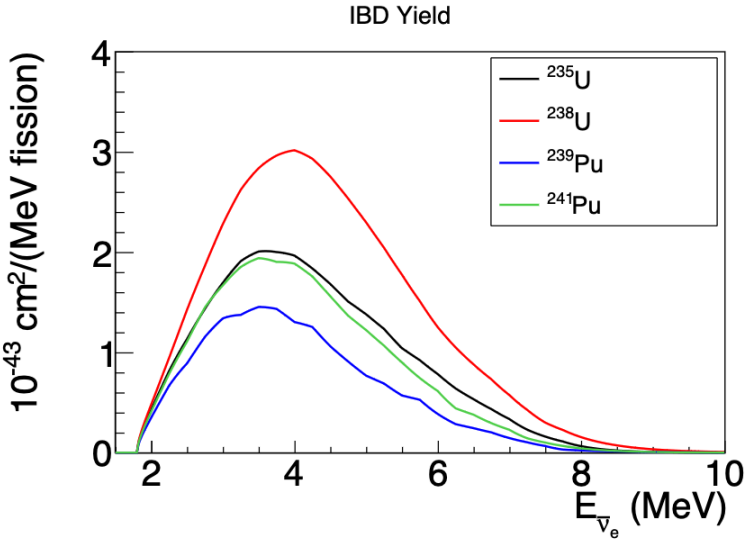


Figure 2.5: The reactor neutrino energy spectrum. The primary sources of reactor neutrinos are the decays of ^{235}U, ^{238}U, ^{239}Pu, and ^{241}U isotopes [6].

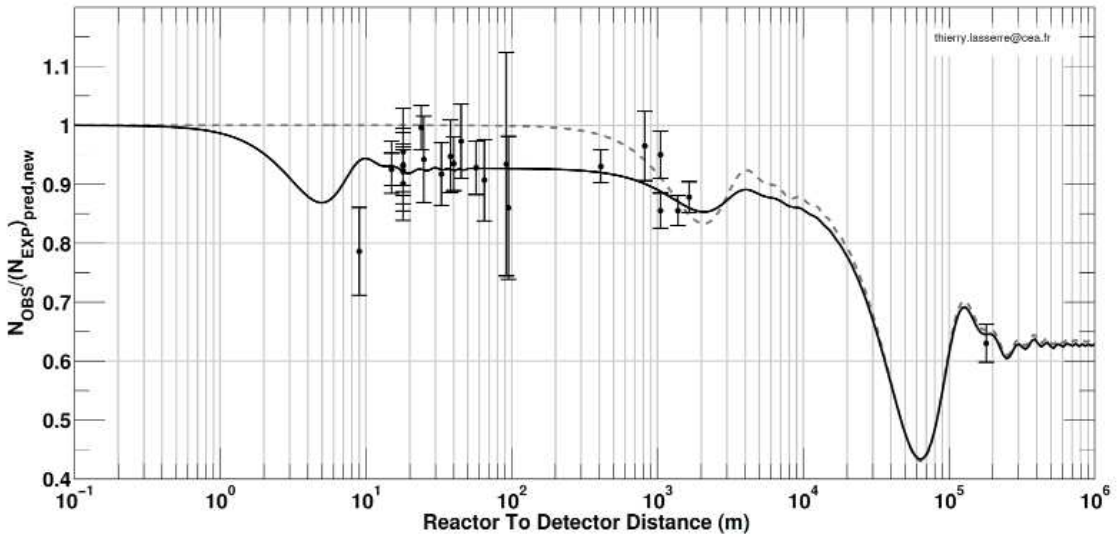


Figure 2.6: The reactor antineutrino anomaly. The graph shows the ratio of observed and expected events as a function of distance between detector reactor for all reactor experiments. The solid and dashed lines respectively represent the expectation under the three-neutrino model and the 3+1 model, which adds a single sterile neutrino [6].

Reactor-based neutrino experiments are sensitive to the parameter $\sin^2(\theta_{13})$ and Δm^2_{13}. The Daya Bay, Double Chooz, RENO, and KamLAND experiments have contributed to the measurement of these parameters. Recent global fits to reactor neutrino data have provided values of $\sin^2(\theta_{13}) = 0.02225^{+0.00056}_{-0.00059}$ and $\Delta m^2_{13} = 2.507^{+0.026}_{-0.027} \times 10^{-3}$ eV^2 for the normal neutrino mass ordering scenario and $\sin^2(\theta_{13}) = 0.02223^{+0.00058}_{-0.00058}$ and $\Delta m^2_{13} = 2.486^{+0.025}_{-0.028} \times 10^{-3}$ eV^2 for the inverted neutrino mass ordering scenario [1].

It is worth noting that several reactor neutrino experiments have also observed a deficit in the measured reactor neutrino flux compared to the expected flux. This deficit (Figure 2.6) is known as the Reactor Antineutrino Anomaly (RAA), and it has been interpreted as a possible indication of the existence of a fourth sterile neutrino state. Recent results, however, have suggested that the anomaly may in fact be due to a bias in the nuclear fission models used to predict the reactor neutrino flux [35].

2.4 Accelerator-Based Neutrino Experiments

The final mechanism of neutrino production discussed in this section, and the primary focus of the rest of this thesis, is neutrino production via particle accelerators. The idea of using particle accelerators to produce neutrinos came about in the late 1950s and early 1960s, on the same timeline as Pontecorvo's proposal to use a proton beam to produce neutrinos from pion decays [25] and Schwartz's calculations of the neutrino flux from a proton beam. The Alternating Gradient Synchrotron (AGS) at Brookhaven National Laboratory was commissioned in 1962 and subsequently became the first beam neutrino experiment. The AGS produced a beam of neutrinos by directing 15 GeV protons onto a beryllium target, which produced pions and kaons that subsequently decayed to produce neutrinos. This experiment demonstrated the existence of two types of neutrinos, ν_e and ν_μ [36], and earned Leon Lederman, Melvin Schwartz, and Jack Steinberger the 1988 Nobel Prize in Physics *"for the neutrino beam method and the demonstration of the doublet structure of the leptons through the discovery of the muon neutrino."*

The method of producing neutrinos with particle accelerators has since been refined. The modern accelerator-based neutrino beam typically begins by stripping protons from a gas (such as hydrogen) and accelerating them to high energies to form a proton beam. The proton beams used in neutrino experiments typically have energies up to $\mathcal{O}(100\ \mathrm{GeV})$. The proton beam is then directed onto a high-density target, creating a shower of secondary particles, including pions and kaons. These pions and kaons are the primary production mode of neutrinos in accelerator-based experiments, and their charged nature allows them to be focused via a magnetic focusing horn onto a long decay region. The collimation of the charged secondaries is important because the resulting decay-in-flight of the pions and kaons produces neutrinos with a well-defined direction. Moreover, the direction of the horn's current can be used to select for strictly positively-charged or negatively-charged pions and kaons, which allows for the creation of a predominately neutrino or antineutrino beam. At the end of the decay region there is typically a beam dump to absorb the remaining protons and secondary particles. The result of this process is a beam of neutrinos

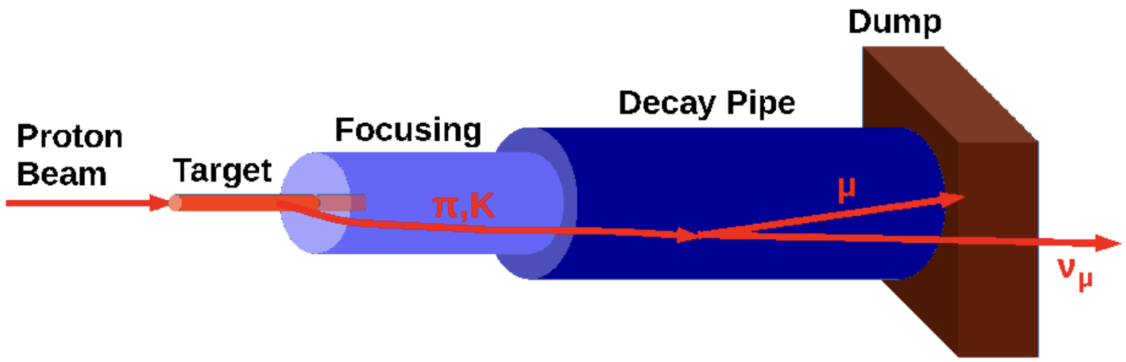


Figure 2.7: A diagram of a typical neutrino beamline. Protons are accelerated to high energies and directed onto a target to produce pions and kaons, which subsequently are focused by a magnetic horn and decay to produce a collimated beam of neutrinos incident on a detector [7].

(or antineutrinos) with energies ranging from $\mathcal{O}(100\ \mathrm{MeV})$ to $\mathcal{O}(100\ \mathrm{GeV})$. A diagram showing a typical neutrino beamline is illustrated in Figure 2.7.

As discussed in the previous chapter and represented in Equation 1.11, the tunable parameters in neutrino oscillation experiments are the baseline L and the neutrino energy E. The physical location of the detector relative to the beam source sets the baseline, whereas the energy is set by the beam energy and subsequent neutrino production modes. Accelerator-based neutrino experiments are typically subdivided into long-baseline (LBL) and short-baseline (SBL) experiments based on the relative ratio L/E. The goal of SBL experiments is to measure the properties of neutrinos in the presence of minimal neutrino oscillations, meaning that the L/E is so small that the oscillation probability is negligible under the standard three-neutrino oscillation framework. These experiments are well-suited for cross section measurements, searches for non-standard neutrino physics, and searches for eV-scale sterile neutrinos. LBL experiments, on the other hand, typically are physically placed at a L/E that maximizes neutrino oscillations and therefore increases sensitivity to the mixing angles and mass splittings. In either SBL or LBL experiments, multiple detectors are typically used to minimize systematic uncertainties associated with the neutrino flux and neutrino-nucleus interactions. The two-detector configuration allows for precise measurements of neutrino flavor appearance and disappearance.

2.4.1 Long-Baseline Neutrino Experiments

LBL experiments are primarily sensitive to Δm^2_{31}, $\sin^2(\theta_{23})$, $\sin^2(\theta_{13})$, the CP-violating phase δ_{CP}, and the neutrino mass ordering. Several LBL experiments have been carried out to measure these parameters and have offered world-leading complementary results on neutrino oscillation physics.

The Main Injector Neutrino Oscillation Search (MINOS) experiment was comprised of two steel scintillator sampling calorimeters: the near detector located at the Fermi National Accelerator Laboratory (Fermilab) and the far detector located 735 kilometers away in the Soudan Underground Laboratory in northern Minnesota. The MINOS detectors were exposed to the Neutrinos at the Main Injector (NuMI) beam, which originates at Fermilab. The most recent combined analysis of ν_μ disappearance and $\nu_\mu \rightarrow \nu_e$ appearance at MINOS provided a measurement of $|\Delta m^2_{32}| = |2.32 - 2.53| \times 10^{-3}$ eV^2 (68% C.L.) and $\sin^2(\theta_{23}) = 0.35 - 0.65$ (90% C.L.) in the normal neutrino mass ordering scenario and $|\Delta m^2_{32}| = |2.32 - 2.53| \times 10^{-3}$ eV^2 (68% C.L.) and $\sin^2(\theta_{23}) = 0.34 - 0.67$ (90% C.L.) in the inverted neutrino mass ordering scenario [37].

The two current generation neutrino oscillation experiments are the T2K (Tokai-to-Kamioka) experiment in Japan and the NOvA (NuMI Off-Axis ν_e Appearance) experiment in the United States. Both experiments have a primary goal of measuring θ_{13} using the $\nu_\mu \rightarrow \nu_e$ appearance channel and a predominately ν_μ beam. The T2K experiment uses the J-PARC facility in Tokai as the source of the neutrino beam and the location of the near detector. The far detector is the Super-Kamiokande detector located 295 kilometers away in Kamioka. The NOvA experiment utilizes the same NuMI beamline that MINOS used with a near detector at Fermilab and a far detector located 810 kilometers away in Ash River, Minnesota. The two collaborations have been pursuing a joint analysis of their data to provide a combined measurement of the neutrino oscillation parameters.

Two other LBL experiments have been proposed and are in various stages of construction. The upgrade to the T2K experiment, Hyper-Kamiokande, is a proposed water Cherenkov detector with a fiducial mass of 260 kilotons - about 25 times the size of its predecessor. The Deep Underground Neutrino Experiment (DUNE) will use massive liquid argon time projection chambers (LArTPCs)

located at the Sanford Underground Research Facility in South Dakota. The DUNE experiment will consist of four 17-kiloton LArTPCs and will be exposed to the Long-Baseline Neutrino Facility (LBNF) beam originating at Fermilab. The LArTPC technology will be discussed in the next chapter.

2.5 Short-Baseline Anomalies

Table 2.1: Global measurements of three-flavor oscillation parameters up to the 1σ Best-Fit (BF) and the 3σ range. Note that $\Delta m^2_{3l} \equiv \Delta m^2_{31} > 0$ for the normal ordering scenario and $\Delta m^2_{3l} \equiv \Delta m^2_{32} < 0$ for the inverted ordering scenario. Values taken from [1].

Parameter	Normal Hierarchy		Inverted Hierarchy	
	1σ BF	3σ Range	1σ BF	3σ Range
$\sin^2\theta_{12}$	$0.303^{+0.012}_{-0.012}$	$0.270 \rightarrow 0.341$	$0.303^{+0.012}_{-0.011}$	$0.270 \rightarrow 0.341$
$\sin^2\theta_{23}$	$0.451^{+0.019}_{-0.016}$	$0.408 \rightarrow 0.603$	$0.569^{+0.016}_{-0.021}$	$0.412 \rightarrow 0.613$
$\sin^2\theta_{13}$	$0.02225^{+0.00056}_{-0.00059}$	$0.02052 \rightarrow 0.02398$	$0.02223^{+0.00058}_{-0.00058}$	$0.02048 \rightarrow 0.02416$
δ_{CP} (deg)	232^{+36}_{-26}	$144 \rightarrow 350$	276^{+22}_{-29}	$194 \rightarrow 344$
Δm^2_{21} (10^{-5} eV^2)	$7.41^{+0.21}_{-0.20}$	$6.82 \rightarrow 8.03$	$7.41^{+0.21}_{-0.20}$	$6.82 \rightarrow 8.03$
Δm^2_{3l} (10^{-3} eV^2)	$2.507^{+0.26}_{-0.27}$	$2.427 \rightarrow 2.590$	$-2.486^{+0.025}_{-0.028}$	$-2.570 \rightarrow -2.406$

The current three-neutrino oscillation framework has been successful in describing the results of solar, atmospheric, reactor, and LBL neutrino experiments. However, there have been results at SBL experiments which appear to be in tension with the three-flavor neutrino model. Such anomalies may be the result of some mis-modeled background or missing systematic uncertainty in measurements, but they may also be indicative of new physics beyond the Standard Model. This section will cover the LSND and MiniBooNE experiments, which have each observed an excess of electron-like events that are not consistent with the three-neutrino oscillation framework. The results of these experiments have motivated the Short-Baseline Neutrino (SBN) Program.

2.5.1 The LSND Experiment

The Liquid Scintillator Neutrino Detector (LSND) was a short-baseline neutrino oscillation experiment that ran from 1993 to 1998 at the Los Alamos National Laboratory. LSND was situated

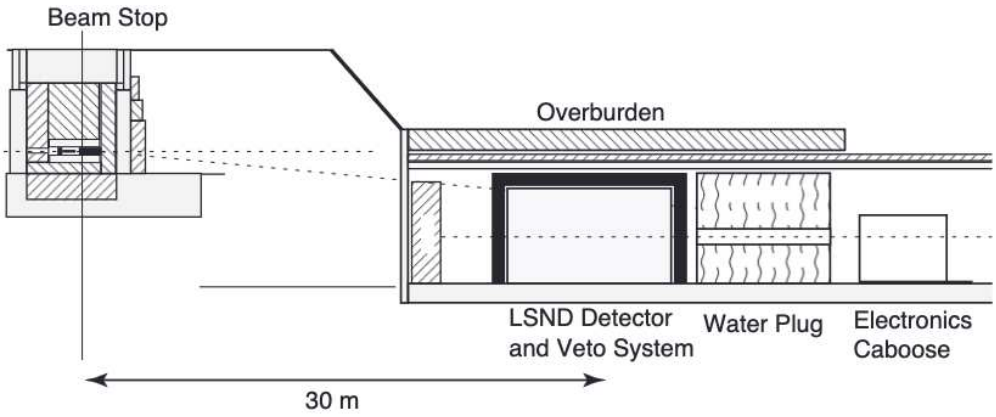


Figure 2.8: A diagram of the LSND detector and beamline. The detector was situated approximately 30 meters from the neutrino production target. The figure is taken from [8].

in the 0.8 GeV proton beam at the Los Alamos Neutron Science Center (LANSCE). The beam primarily produced pions, which then produced neutrinos predominately through the $\mu^+ \to e^+\nu_e\bar{\nu}_\mu$ decay-at-rest production mode. The LSND experiment was designed to search for $\bar{\nu}_\mu \to \bar{\nu}_e$ oscillations through the process:

$$\bar{\nu}_\mu \xrightarrow{oscillation} \bar{\nu}_e + p \to e^+ + n \tag{2.1}$$

The LSND detector was filled with 167 metric tons of mineral oil and liquid scintillator and instrumented with 1220 photomultiplier tubes (PMTs). It was situated approximately 30 meters from the neutrino production target (Figure 2.8). The design of the detector allowed for the detection of both the Cherenkov light and scintillation light produced by charged particles in the interaction final state, where the combination of this information provided particle identification, track angle reconstruction, and energy estimation. Cherenkov light is produced by charged particles traveling faster than the speed of light in the medium and produces a ring-like signature observable with the PMTs install on the walls of the detector. The threshold for the Cherenkov radiation process is dependent on the mass of the particle, effectively making the LSND detector insensitive to more massive particles like protons and neutrons.

The neutrino source had a small amount of intrinsic $\bar{\nu}_e$ contamination, so the measurement of this process, which has a large and well-known cross section, provided a sensitive way to search

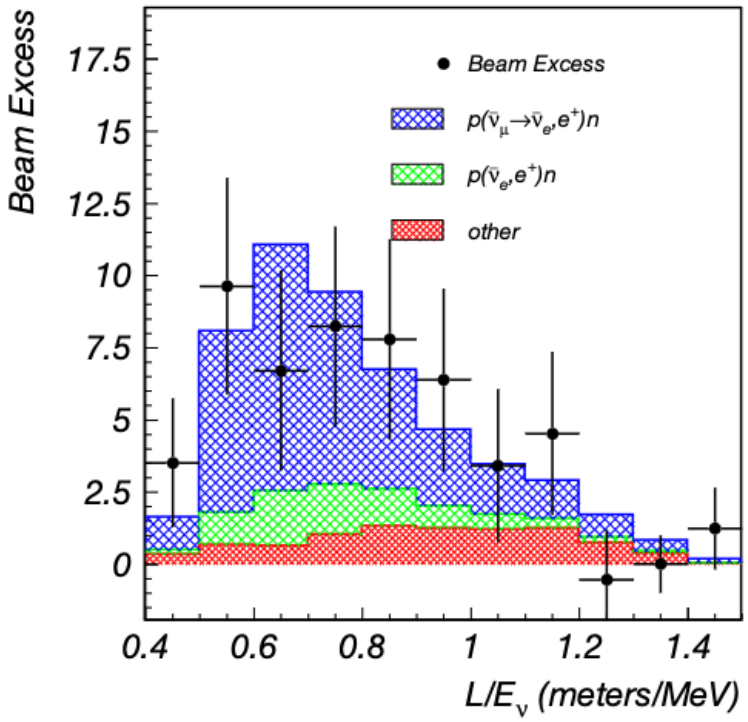


Figure 2.9: The LSND $\bar{\nu}_e$ event spectrum. The measured events are shown as black points, while the background from other sources, including $\bar{\nu}_e$'s intrinsically present in the beam, are shown as the green and red regions of the histogram. The blue histogram shows the hypothesized $\bar{\nu}_e$ events from $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ oscillations at the measured value of Δm^2. The figure is taken from [8].

for $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ oscillations. This inverse beta decay process was detected by observing the positron and neutron produced in the final state: the positron produced a prompt signal in the detector, while the neutron was captured on a proton in the detector and produced a delayed 2.2 MeV γ signal. The LSND collaboration observed an excess of $\bar{\nu}_e$ events over the expected background from other sources, including the intrinsic $\bar{\nu}_e$ contamination in the beam, as shown in Figure 2.9. The excess was observed at a significance of 3.8σ with a best-fit Δm^2 in the region of $0.2 - 10$ eV^2 [8]. The LSND result was surprising because it was not consistent with the three-neutrino oscillation framework that was beginning to be established at the time.

2.5.2 The MiniBooNE Experiment

Following the surprising results of LSND at Los Alamos, the Mini Booster Neutrino Experiment (MiniBooNE) was constructed at Fermilab to further test the LSND signal. MiniBooNE

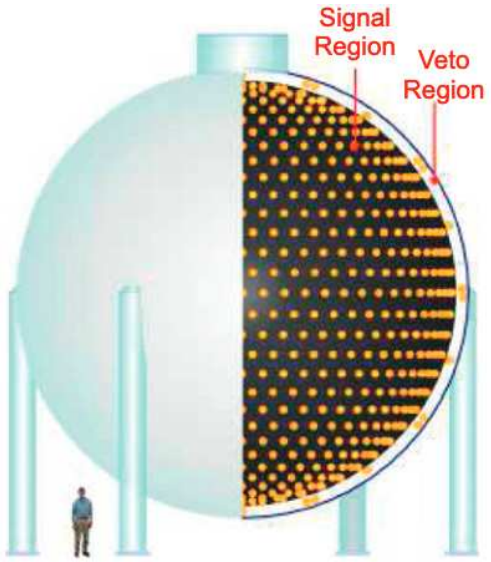


Figure 2.10: A diagram of the MiniBooNE detector. The detector was situated approximately 500 meters from the neutrino production target.

consisted of a large Cherenkov detector filled with 800 tons of mineral oil and instrumented with 1520 PMTs, most of which were reused from the LSND experiment. The design of the detector was a 12.2 meter diameter sphere with the PMTs arranged primarily facing inwards, but with some facing outwards to provide a veto for cosmic rays. The spherical shape was chosen to maximize the ratio of volume to instrumented surface area, which was important for event reconstruction. This design choice also resulted in a fully active volume with no dead regions - an important feature to prevent neutral current $\pi^0 \rightarrow \gamma\gamma$ background from mimicking single electron events. A diagram of the MiniBooNE detector is shown in Figure 2.10.

MiniBooNE was situated along Booster Neutrino Beam (BNB), which extracted 8 GeV protons from the Fermilab Booster accelerator and directed them onto a beryllium target and a magnetic focusing horn. The focusing horn was used to focus and preferentially select the sign of charged pions produced in the target, which then decayed to produce neutrinos. The BNB produced a predominantly ν_μ beam with less than a percent of ν_e contamination. Though situated at a larger distance from the neutrino source than LSND, the higher energy of the neutrinos placed MiniBooNE at an L/E that was nearly the same as LSND. The interaction channels studied at MiniBooNE were similar to those studied at LSND, namely [13]:

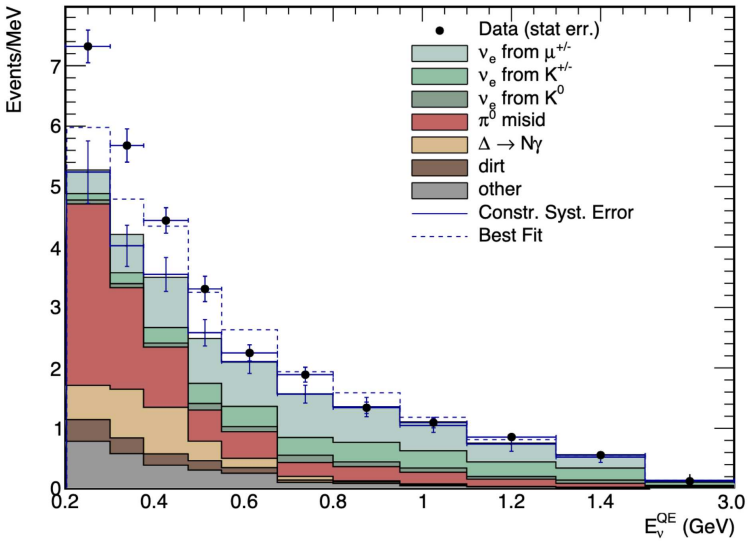


Figure 2.11: The MiniBooNE electron-like event spectrum. The measured events are shown as black points, while the background from other sources, including ν_e's intrinsically present in the beam, are shown as the shaded regions of the histogram. An excess of 4.8σ above the expected event rate (without anomalous neutrino oscillations) is observed [9].

$$\nu_\mu \xrightarrow{oscillation} \nu_e + n \rightarrow e^- + p \tag{2.2}$$

$$\bar{\nu}_\mu \xrightarrow{oscillation} \bar{\nu}_e + p \rightarrow e^+ + n \tag{2.3}$$

The MiniBooNE experiment collected data from 2002 to 2019 and observed an excess of electron-like events in the neutrino mode data, as shown in Figure 2.11. The shaded region shows the expected background from other sources, including ν_e's intrinsically present in the beam. The data points reflect a 4.8σ excess above the expected event rate in the region below about 800 MeV. The measurement is limited by systematic uncertainties, as the significance of the excess using statistical uncertainties only is 12.2σ. The excess is also higher than the best-fit prediction under the 3+1 sterile neutrino hypothesis. A joint analysis between LSND and MiniBooNE was also reported in the latest updated analysis and showed consistency in both energy and magnitude of the events with a combined significance of 6.0σ [9].

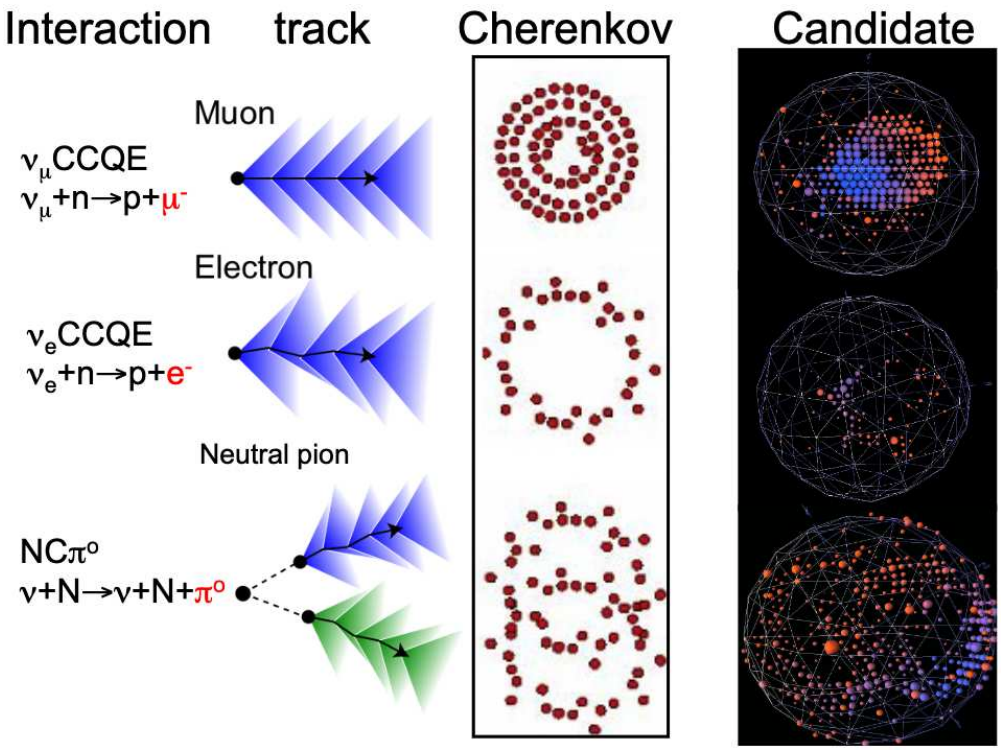


Figure 2.12: Examples of event topologies in the MiniBooNE detector. The top row shows ν_μ CCQE events, the middle row shows ν_e CCQE events, and the bottom row shows NC $\pi^0 \to \gamma\gamma$ events. The image is taken from [10].

The flavor of the final-state lepton is the key to resolving the flavor of the parent neutrino. Muon production is indicative of a muon neutrino interaction, while electromagnetic activity is indicative of electron neutrino or π^0 events. Muons travel in long, straight paths and produce a well-defined Cherenkov ring signature, while electrons produce a more diffuse Cherenkov ring owing to the cascading electromagnetic activity in the interaction. Events that produce a γ, which subsequently produces an e^+e^- pair, will create two diffuse rings that are not always separable. Moreover, π^0 decays ($\pi^0 \to \gamma\gamma$) may result in only a single ring if one of the γ's escapes the detector. Examples of these event topologies are shown in Figure 2.12.

The challenges associated with electron/photon discrimination at MiniBooNE are a significant source of background events and systematic uncertainty in the ν_e appearance search. While the appearance of ν_e events is consistent with the introduction of at least one sterile neutrino with a Δm^2 high enough to drive oscillations at short baselines, misidentified $\pi^0 \to \gamma\gamma$ and resonant

decays such as $\Delta \rightarrow N + \gamma$ must also be considered as explanations. The largest of these two, the π^0 background, was measured by MiniBooNE by reconstructing the two-gamma invariant mass to select events consistent with the π^0 mass [9]. Additionally, the $\Delta \rightarrow N + \gamma$ background was constrained by a NC π^0 measurement, which has contributions from Δ resonant production [9].

These results from LSND and MiniBooNE, both involving excesses of electron-like events with difficult-to-constrain photon backgrounds, have motivated the SBN Program. The detectors of the SBN Program utilize the LArTPC technology to provide excellent particle identification and electron/photon separation capabilities. The LArTPC technology and the SBN Program will be discussed in the next chapter.

Chapter 3

The Short-Baseline Neutrino Program

The Short-Baseline Neutrino (SBN) Program is a suite of three detectors utilizing the liquid argon time projection chamber (LArTPC) technology and dedicated to the resolution of neutrino oscillation anomalies at short baselines. The LArTPC technology is the principal technology of choice for modern neutrino detectors and offers a unique combination of high-resolution tracking, calorimetry, and particle identification capabilities, which are essential for the study of neutrino interactions. This chapter provided an overview of the LArTPC technology (Section 3.1) and the Short-Baseline Neutrino Program (Section 3.2).

3.1 The LArTPC Technology

The energy of an energetic particle traveling through argon is transferred to the atoms primarily in two ways: ionization and atomic excitation. The ionization process occurs when the energy transfer is sufficient to liberate an electron from the parent argon atom. Atomic excitation occurs when the particle transfers energy to the atom, but not enough to remove an electron. An excited argon atom quickly pairs with another argon atom to form an excited dimer, which then decays to the dissociative ground state by emission of a 128 nm ultraviolet photon, as represented in Equation 3.1. In addition to light production via atomic excitation, the process of ionization and subsequent recombination of liberated electrons and argon ions also produces light as shown in Equation 3.2. At zero electric field, the number of photons released by minimum-ionizing particles (MIPs), such as muons or pions, is approximately 40,000 photons per MeV and decreases by about a factor of two at an electric field of $\mathcal{O}(100\ \mathrm{V/cm})$ [38]. In each process, the emission of scintillation light proceeds through the radiative decay of the excited dimer. The dimer may exist in either a singlet or triplet state, referring to how the spin of the electron and the argon dimer couple. The single state decays promptly with a lifetime of $\mathcal{O}(6\ \mathrm{ns})$ and the triplet state decays with a lifetime of

$\mathcal{O}(1.5\ \mu\text{s})$. Collectively, these make up the "prompt" and "late" components of the scintillation signal available for detection in a LArTPC.

$$\text{Ar}^* + \text{Ar} \rightarrow \text{Ar}_2^* \rightarrow 2\text{Ar} + \gamma \tag{3.1}$$

$$\text{Ar}^+ + \text{Ar} + e^- \rightarrow \text{Ar}_2^+ + e^- \rightarrow \text{Ar}_2^* \rightarrow 2\text{Ar} + \gamma \tag{3.2}$$

The application of a strong electric field across the argon volume reduces the rate of recombination, therefore preserving an ionization signal which traces the path of the charged particle. The freed electrons are drifted along this applied electric field to the readout electronics, where they are collected and converted into a signal. Electronegative impurities in the argon, such as oxygen and water, can capture the drifting electrons, thereby reducing the signal. This is also the reason that noble elements are an attractive choice for such detectors - they are chemically inert with low electronegativity.

In 1974, William Willis and Veljko Radeka demonstrated the use of liquid argon ionization chambers as a total absorption calorimeter. They showed that the ionization signal produced by charged particles traversing the liquid argon volume could be used to measure the total charge deposited by particles traversing through the detector [39]. Two years later, Nygren proposed the concept of a TPC using a gaseous argon-methane mixture as a tracking detector in the Positron-Electron Project (PEP) at SLAC. The design used an electric field to drift ionization electrons generated by charged particles in the gas towards each end of the cylindrical chamber, where a single layer of proportional wire readout planes amplified and detected the arriving electrons. This design allowed for the three-dimensional reconstruction of charged particle activity in the detector volume [40]. In 1977, Carlo Rubbia proposed the use of LArTPCs for neutrino detection [41], which has since become a principal technology in modern neutrino detectors.

In the modern day, a typical LArTPC-based neutrino detector consists of several important components: an anode plane, a cathode plane, a field cage, and a photon detection system (PDS),

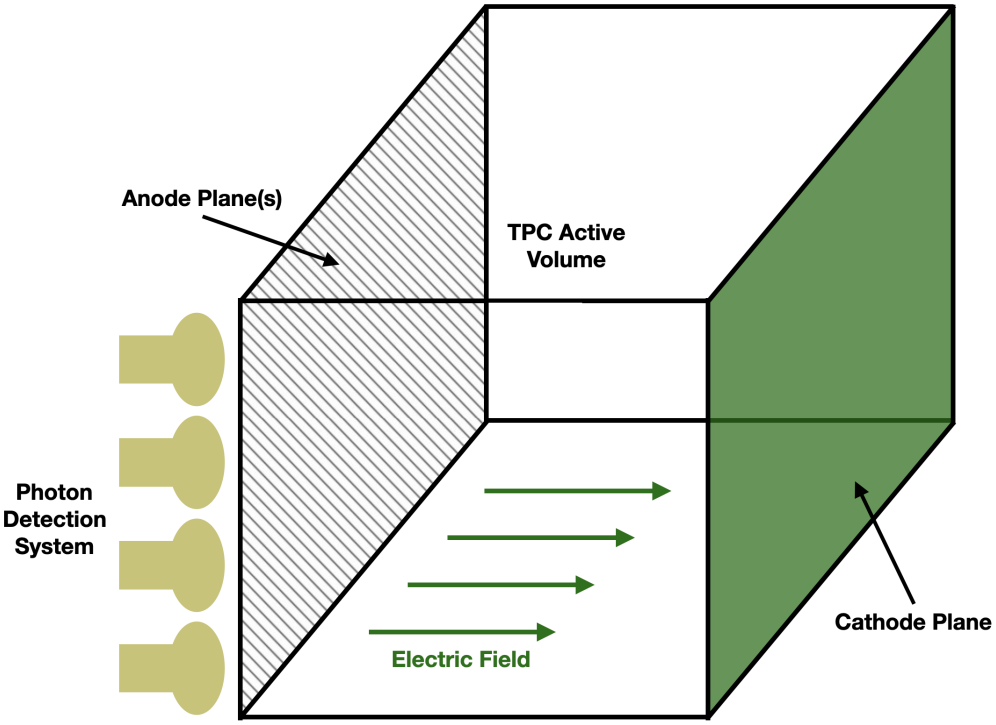
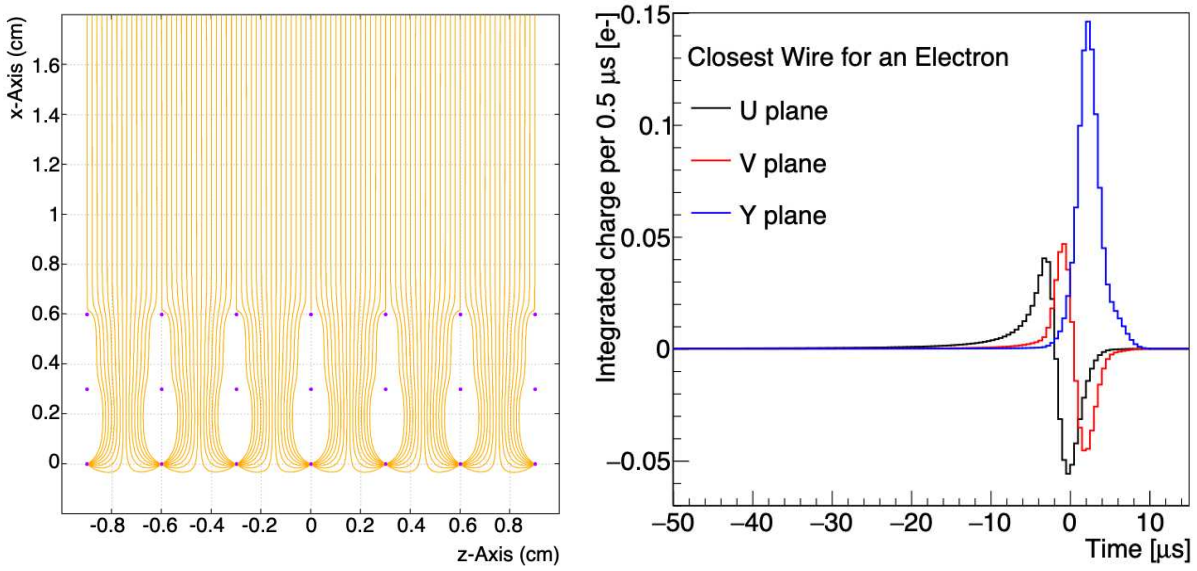


Figure 3.1: Schematic diagram of a typical LArTPC-based neutrino detector. The LArTPC consists of an anode plane, a cathode plane, a field cage, and a photon detection system (PDS).

as shown schematically in Figure 3.1. The cathode plane is a large, electrically conductive surface which is held at a negative voltage with respect to the anode plane. The field cage is a series of electrically conductive rings which are held at intermediate voltages between the anode and cathode planes. A resistor divider chain is typically used to step down the voltage magnitude from the cathode to the anode at a constant gradient. The field cage is used to shape the electric field in the detector volume, ensuring that the ionization electrons experience a uniform electric field across the entire drift distance.

The anode plane consists of a series of wire planes, each with wires oriented at an angle different to those of the other planes and biased with a voltage such that the drifting ionization electrons are deflected around all but the last wire plane. The final wire plane is biased such that the drifting ionization electrons are collected instead. An example of simulated ionization electron trajectories from MicroBooNE, which has two induction planes and a collection plane, is shown in Figure 3.2a. Each wire near the path of an ionization electron will experience a particular induced current



(a) **(b)**

Figure 3.2: **(a)** The simulated path traversed by an ionization electron in the MicroBooNE LArTPC. The wire planes are each represented by a row of dots indicating their orientation into the page. The image represents a 2D simulation done with GARFIELD. In reality, the wire planes are not parallel with each other. **(b)** The simulated average signal shapes in the MicroBooNE LArTPC for the first induction plane (U plane), the middle induction plane (V plane), and the collection plane (Y plane). Images taken from [11].

signal as the electron drifts through the field. The signal on induction-type planes, meaning planes that are not collecting the drifting ionization electrons, will exhibit a rising signal as the ionization electron drifts towards it followed by a rapid switch in polarity as the ionization electron drifts past the wire towards the subsequent plane. The first plane will also experience a long-range induction effect, which tends to stretch the signal towards earlier time, though this effect can be mitigated by the addition of a shielding (grid) plane in between the first induction plane and the cathode. The signal observed on the collection plane will exhibit a unipolar shape as the drifting ionization electrons are collected by the wires. The simulated average signal shapes for the three wire planes in MicroBooNE are shown in Figure 3.2b as an example.

Each wire plane produces a two-dimensional projection of the three-dimensional ionization signal corresponding to the charged particle traversing the detector volume. The spacing of the wires, typically a few millimeters, ultimately determines the spatial resolution of the detector. An example event display of a ν_μ CC candidate from the collection plane of the ICARUS detector

is shown in Figure 3.3. The vertical axis is the time axis with an extent of about a millisecond, whereas the horizontal axis represents the wire axis with a spatial extent of about two meters. The event display clearly shows the high spatial resolution of the detector, as evidenced by the clarity with which one can reconstruct fine details of the tracks such as the delta rays from the muon track. The color represents the local charge deposition density; it is possible to visually and algorithmically separate the proton, which has a higher energy deposition rate and tends to stop in a shorter distance, from the muon. The combination of two-dimensional projections from each wire plane makes it possible to reconstruct a three-dimensional image of the same activity. The process of reconstructing events and extracting high-level information for analyses will be discussed in Chapter 5.

Fundamentally, two different types of particle topologies exist in interactions as seen with the wire planes: tracks and electromagnetic showers. Tracks represent the path of a charged particle through the detector volume and are generally straight and continuous. Bends in the track may be caused by Coulomb scattering of the particle on the argon atoms or distortions in the electric field. Electromagnetic showers are cascades of particles produced by a high-energy photon or electron. Showers are characterized by a series of electromagnetic processes such as pair production, Compton scattering, and bremsstrahlung. The shower is visible in the detector as a collection of spatially separate "pieces" of charge deposition with gaps indicating the path taken by a photon.

One of the challenging backgrounds associated with the LSND and MiniBooNE experiments was interactions which produced one or more photons, as they can be difficult to distinguish from the electron-containing signal. The LArTPC technology provides two ways to address this background: the spatial gap between the photon creation point and the shower start and the energy profile at the start of the electromagnetic shower. When a photon is emitted at the interaction vertex, it produces no visible signal until it pair produces into an e^+/e^- pair, thus beginning the visible portion of the electromagnetic shower. This gap is not present for showers induced by electrons. Similarly, the deposited energy per unit length, or dE/dx, at the start of the shower will either match

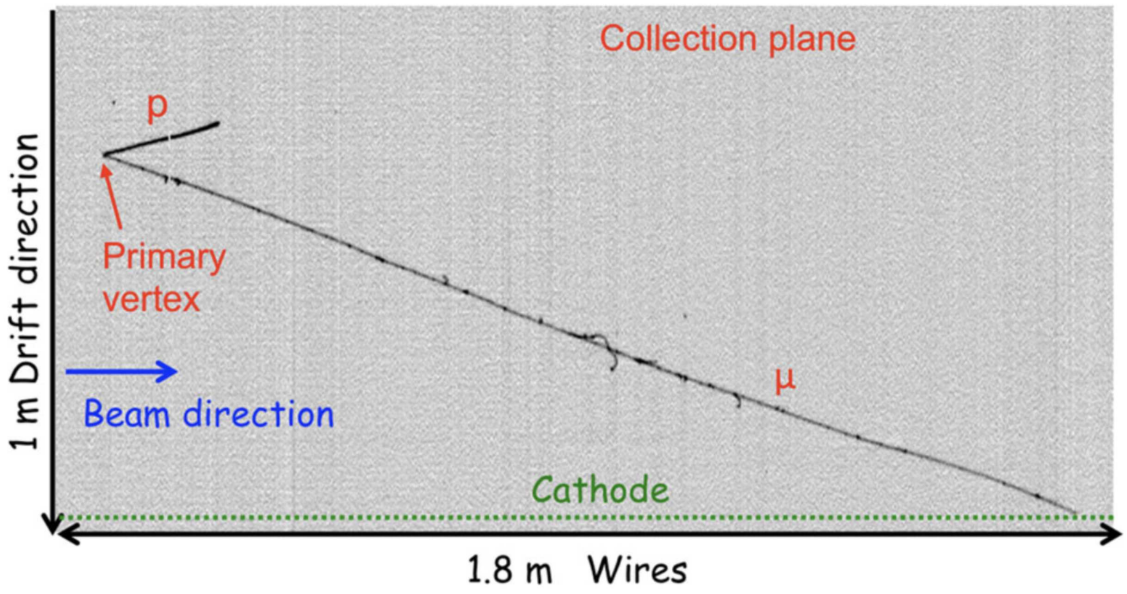


Figure 3.3: An example event display of a ν_μ CC candidate from the ICARUS detector. Image taken from [12].

the profile of a single electron or that of a stacked e^+e^- pair. In the latter case, the dE/dx will be twice as high as that of a single electron.

The electric field in the detector volume of a typical LArTPC is $100 - 500$ V/cm, which is sufficient to drift the ionization electrons to the anode plane(s) in a few milliseconds. Neutrinos from accelerator-based sources typically are delivered in $\mathcal{O}(\mu\text{s})$ windows, which makes the timing information available from the TPC subsystem alone relatively imprecise. This is one of the motivations for leveraging the $\mathcal{O}(\text{ns})$ timing information from the prompt scintillation light produced by the interaction. This is accomplished through the use of the PDS, shown schematically behind the anode wire plane in Figure 3.1, which collects scintillation photons and produces a digitized signal of the "light flash" produced by the interaction. The association of this optical signal with ionization charge signal information provides timing information with a resolution matching the PDS and can be used to reject non-beam backgrounds such as cosmic rays.

To summarize, the LArTPC technology is characterized by $\mathcal{O}(\text{mm})$ resolution of interactions in the detector volume, precise calorimetric information for energy reconstruction and particle identification, and $\mathcal{O}(\text{ns})$ resolution of event timing for background rejection. Liquid argon is

an attractive medium for neutrino detection due to its high density, low electronegativity, and high scintillation light yield. The design of a LArTPC is also modular, and when combined with argon's relatively high abundance in the atmosphere it allows for the cost-effective construction of large detectors.

3.2 The Short-Baseline Neutrino Program

In 2014, the Particle Physics Projects Prioritization Panel (P5) recommended a near-term, world-leading, short-baseline experimental neutrino program with strong international and domestic participation. Specifically [42]:

- **P5 Recommendation #12:** In collaboration with international partners, develop a coherent short- and long-baseline neutrino program hosted at Fermilab.
- **P5 Recommendation #15:** Select and perform in the short term a set of small-scale short-baseline experiments that can conclusively address experimental hints of physics beyond the three-neutrino paradigm. Some of these experiments should use liquid argon to advance the technology and build the international community for LBNF at Fermilab.

In the modern day, the Long-Baseline Neutrino Facility (LBNF) is the infrastructure project at Fermilab which will host the Deep Underground Neutrino Experiment (DUNE). These recommendations are the genesis of the SBN Program at Fermilab, as proposed in [14], which is dedicated to resolving the short-baseline neutrino anomalies observed in the LSND and MiniBooNE experiments while furthering the development of the LArTPC technology for DUNE. Section 3.2.1 will provide a brief overview of the physics goals of the SBN Program, followed by a description of the Booster Neutrino Beam in Section 3.2.2. The three detectors of the SBN Program, the Short-Baseline Near Detector (SBND), MicroBooNE, and ICARUS, will be respectively described in Sections 3.2.3, 3.2.4, and Chapter 4.

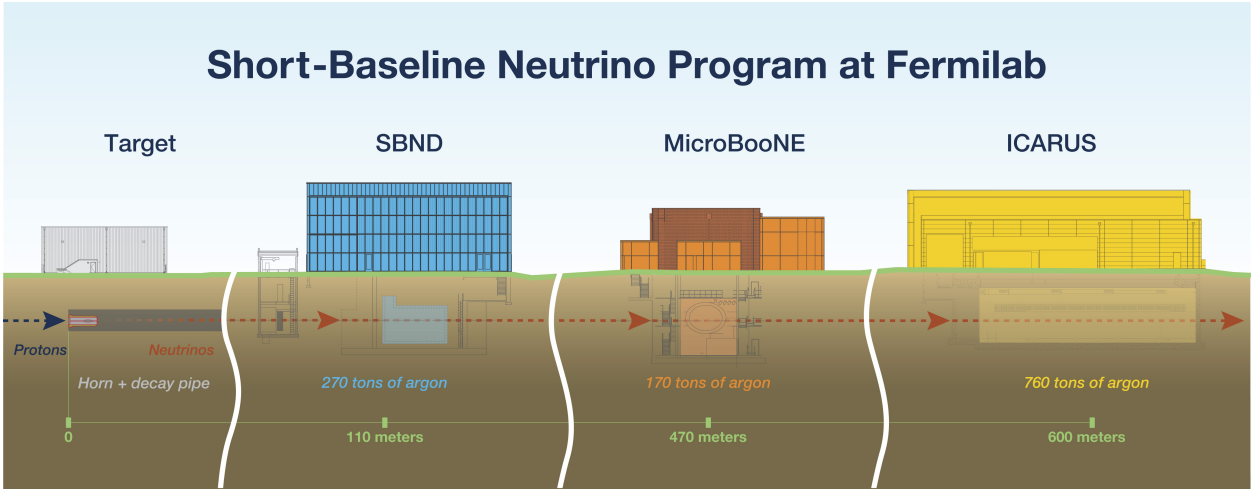


Figure 3.4: Schematic diagram of the Short-Baseline Neutrino Program at Fermilab. The three detectors, SBND, MicroBooNE, and ICARUS T600, are located on-axis in the Booster Neutrino Beam. Image from Fermilab Visual Media Services.

3.2.1 Physics Goals

The SBN Program brings together three detectors utilizing the LArTPC technology and located on-axis in the Booster Neutrino Beam. Figure 3.4 schematically summarizes the locations of the three detectors relative to the BNB target and their sizes. The detector locations were chosen to optimize the sensitivity to neutrino oscillations and to minimize the impact of flux systematic uncertainties. SBND is located closest to the BNB target and is designed to characterize the neutrino flux. The two far detectors, MicroBooNE and ICARUS, are positioned to have complementary sensitivities to short-baseline neutrino oscillations. The shared technology and beamline of the three detectors means that common systematic uncertainties cancel in the comparison of the results from the three detectors to first order. As an example, if the neutrino flux or ν-Ar cross sections are mis-modeled, the effect will be the same in all three detectors, and the comparison of the results from the three detectors will be unaffected. This is a significant advantage of the SBN Program over previous experiments which have observed short-baseline neutrino anomalies.

The SBN Program, following the recommendations of P5, has four primary physics goals:

- Conclusively resolve anomalies observed at short-baselines
- Characterize $\nu - Ar$ interactions

- Search for physics beyond the Standard Model
- Develop the LArTPC technology, software, and analysis tools for DUNE

These four primary physics goals will each be described in detail below.

Conclusively resolve anomalies observed at short-baselines The primary physics goal of the SBN Program is to conclusively resolve the short-baseline neutrino anomalies observed in the LSND and MiniBooNE experiments, as described in Section 2.5.1 and 2.5.2. As discussed in Section 1.5, the 3+1 minimal extension to the three-neutrino model provides effective oscillation probabilities for ν_μ disappearance (Equation 3.3) and ν_e appearance (Equation 3.4) with which the SBN Program can explore the parameter space of the sterile neutrino hypothesis:

$$\begin{aligned} P^{3+1}_{\nu_\mu \to \nu_\mu} &= 1 - 4\left|U_{\mu 4}\right|^2 \left(1 - \left|U_{\mu 4}\right|^2\right) \sin^2 \frac{\Delta m^2_{41} L}{4E} \\ &\equiv 1 - \sin^2 2\theta_{\mu\mu} \sin^2 \frac{\Delta m^2_{41} L}{4E} \end{aligned} \tag{3.3}$$

$$\begin{aligned} P^{3+1}_{\nu_\mu \to \nu_e} &= 4\left|U_{\mu 4}\right|^2 \left|U_{e 4}\right|^2 \sin^2 \frac{\Delta m^2_{41} L}{4E} \\ &\equiv \sin^2 2\theta_{\mu e} \sin^2 \frac{\Delta m^2_{41} L}{4E} \end{aligned} \tag{3.4}$$

These two channels are the primary targets with the highest sensitivity due to the fact that the BNB is a predominately ν_μ beam.

Electron neutrino candidates include intrinsic ν_e CC interactions as well as other beam-related mis-identified backgrounds. The SBN proposal [14] chose a signal definition which requires the presence of a single electron shower with $E_e > 200$ MeV and with an assumed identification efficiency of 80% after restricting the candidate interaction vertices to the fiducial volume. The fiducial volume cut requires that the vertex of the interaction to be at least 25 cm from the faces parallel to the beam, 30 cm from the detector face upstream of the beam, and 50 cm from the detector face downstream of the beam in order to maintain reconstruction fidelity. The 80% efficiency target was estimated from hand-scanning simulated events and requires verification with automated reconstruction algorithms.

In addition to the intrinsic ν_e CC interactions, there are several sources of background which can mimic the signal. Neutrinos interacting through NC channels may produce γ showers above the 200 MeV shower cut, which can be mis-identified as electrons. For example, NC interactions which produce any number of π^0 in the final state or radiative resonant decays are sources of γ's which can mimic the signal. Cuts on the number of reconstructed showers in an interaction along with conversion gap and dE/dx cuts, which leverage the excellent spatial resolution and calorimetry of the LArTPC technology, can be used to suppress these backgrounds.

Due to the relatively plentiful ν_μ CC interactions, mis-reconstructed ν_μ CC interactions with an electromagnetic shower can also fake a ν_e CC interaction. If the muon is not reconstructed or if the muon is mis-identified as a pion, the interaction may meet the signal definition. Tracks that are sufficiently long are almost entirely muons, so a cut on the track length can be used to suppress these backgrounds. Cosmogenic photons, either from a cosmic muon or produced in the atmospheric shower itself, may also create electromagnetic showers which can mimic the signal. Additional shielding in the form of a concrete overburden can reduce the rate of primary photons significantly, but does not eliminate the background imposed by primary cosmic muons. Instead, reconstruction cuts utilizing the interaction topology and timing information can be used to suppress these backgrounds, or robust estimation of the cosmic background through simulation or data can be used to characterize the background.

ν_μ CC candidates are similarly selected assuming an 80% reconstruction and identification efficiency after a fiducial volume cut has been applied (the same as described previously). This signal definition requires a muon in the interaction's final state. Beam-related backgrounds include NC charged pion production, as charged pions have a similar dE/dx profile to muons and may therefore be mis-identified as muons. Charged pion tracks are typically quite short with most traveling less than half a meter in liquid argon, so a cut on the track length can be used to suppress these backgrounds. A minimum length of 50 cm for the candidate muon track was chosen in the SBN proposal to minimize these backgrounds.

Cosmogenic muons are a significant background for the ν_μ disappearance search. Cosmogenic muons are most often out-of-time with respect to the beam window, so a cut on the timing of the interaction can be used as a first step to suppress these backgrounds. Furthermore, muons entering the detector can be tagged in a variety of ways. A muon which stops in the detector will have one end point at the detector boundary and exhibit a clear stopping signature in the form of a high rate of charge deposition at the opposite end of the track. The cosmic ray tagging system, which is a subsystem surrounding each of the three SBN detectors dedicated to tagging particles as they enter or exit the detector, can also be used to provide further cosmic background rejection. The combination of interaction timing and the requirement of an exiting or fully-contained track topology are sufficient to suppress the cosmic background to a manageable level.

Precise measurement of the energy of the neutrino interaction is essential for optimal sensitivity to neutrino oscillations and is contingent upon the calorimetric capabilities of the LArTPC technology. The reconstruction of the electron neutrino energy relies on the precise reconstruction of the resulting electromagnetic shower's energy. This is done by summing the charge deposited in the shower and converting it to an energy by accounting for the electronics gain, electron lifetime, and recombination effects.

The reconstruction of the muon neutrino energy is dependant on whether the resulting muon track is fully contained or exiting. If the muon is fully contained, the energy can be estimated using the length of the muon track and the theoretically well-known rate of energy loss for muons in liquid argon. If the muon exits the volume, its energy must instead be estimated using the degree of scattering from Coulomb interactions present along the track. In order to attain sufficient energy resolution through multiple Coulomb scattering measurements, the SBN proposal places a 100 centimeter minimum length cut on exiting muon tracks.

Characterize $\nu - \mathrm{Ar}$ interactions Precise neutrino-nucleus cross section measurements are a fundamental prerequisite for every neutrino oscillation experiment, especially for the future long-baseline experiment DUNE. The MeV-to-GeV neutrino energy range exposes the detectors of the SBN Program to a rich landscape of neutrino interactions on argon, as discussed in Section 1.3,

ranging from the emission of a single nucleon to more complex final states with multiple pions or other hadrons. The LArTPC technology is particularly well suited to studying complex final states due to its high-resolution tracking and calorimetry capabilities. The slightly different detector geometries and locations of each detector in the SBN Program also enables important cross-checks of the neutrino-nucleus cross section measurements.

Search for physics beyond the Standard Model The SBN Program is also well-positioned to search for physics beyond the Standard Model. Sterile neutrino decay is a mechanism which may explain the LSND and MiniBooNE anomalies [43–45]. In this process, an active neutrino interacts via a neutral-current process inside the detector and produces a heavy neutrino which subsequently decays into a light neutrino and a photon. The experimental signature is an interaction vertex in the upstream portion of the detector followed by a photon signal in the downstream region of the detector. The near detector has the advantage of higher statistics due to its location, while MicroBooNE and ICARUS have the advantage of a larger volumes for observing such a decay.

SBN will also be able to search for sub-GeV dark matter produced by the proton beam. Several recent studies have highlighted the sensitivity of the SBN experiments to light exotic new particles such as dark matter tridents [46], Higgs portal scalars [47], and elastically [48] and inelastically [49] scattering dark matter. This sensitivity to a wide range of new physics is enhanced by the location of off-axis location of MicroBooNE and ICARUS with respect to the higher-energy NuMI beam.

Develop the LArTPC technology, software, and analysis tools for DUNE The scope and complexity of DUNE requires the development of new technologies and software tools that rise to the challenge of its physics goals. An important component of the viability of the LArTPC technology for DUNE is the scalability to the large detector volume required for the experiment while maintaining the high-resolution tracking and calorimetry capabilities. The SBN Program provides an opportunity to test similar cryostat, TPC, and PDS designs as those that will be used in DUNE.

The SBN Program requires an advanced suite of algorithms for signal processing, event reconstruction, and simulation, and analysis. These algorithms must account for detector effects such as the diffusion of the drifting ionization electrons, electron-ion recombination, impact of argon impurities on ionization charge loss, and electronics noise. Many of these effects can be characterized in the SBN detectors and will have immediate relevance for DUNE. The scalability of these algorithms to larger data volumes is also an important aspect of the development of the LArTPC technology for DUNE. Factorizing these algorithms into a common software framework will enable the sharing of tools and efficient development of new algorithms. The Liquid Argon Software (LArSoft) package is a realization of this detector-independent organization of common software tools for the simulation, reconstruction, and analysis of LArTPC data [50]. This package is shared between the detectors of the SBN Program and DUNE, and much of the development of LArSoft is currently driven by SBN Program.

3.2.2 The Booster Neutrino Beam

The detectors of the SBN Program make use of the existing Booster Neutrino Beam at Fermilab. The BNB is a conventional horn-focused neutrino beam fed by with 8 GeV protons from Fermilab's Booster accelerator. The BNB was originally designed and optimized for the MiniBooNE detector, and provided neutrinos for the experiment from 2002 to 2019. One of the design considerations for the beamline was the maximization of neutrino flux at roughly 500 MeV and the minimization of the flux at higher energies to reduce π^0 production from NC interactions. These choices were made to optimize the sensitivity of MiniBooNE to the LSND anomaly [13].

The Fermilab Booster is a 474-meter-circumference synchrotron operating at 15 Hz. The Booster accelerates 400 MeV protons from the Fermilab LINAC to 8 GeV, which are then extracted and directed to the BNB target. The harmonic number of the Booster is 84, which means that the protons are extracted in 84 buckets of which 81 are filled with protons. The resulting beam structure is a series of 81 proton bunches with a 2 nanosecond width and a 19 nanosecond spacing. The proton beam is directed by a switch magnet to either to the Main Injector or the BNB.

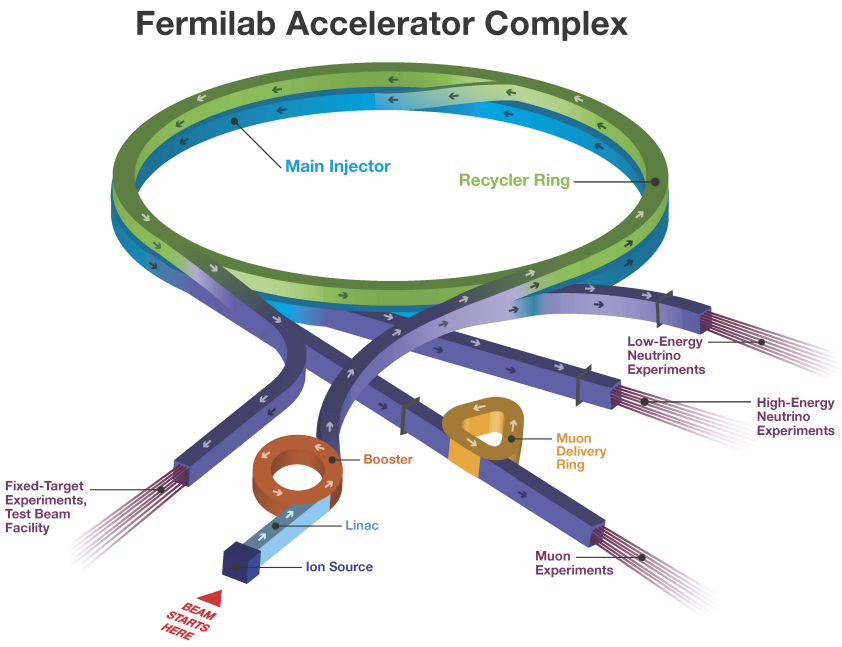


Figure 3.5: Schematic diagram of the Fermilab accelerator complex. The Booster Neutrino Beam is fed by 8 GeV protons from the Booster accelerator. Image taken from the Fermilab Visual Media Services

Figure 3.5 shows these components of the Fermilab accelerator complex. Protons sent to the BNB are focused onto a beryllium target, producing primarily charged pions and kaons. The charged particles produced by the proton beam are then focused by a pulsed magnetic horn system, shown schematically in Figure 3.6. The direction of the horn current determines the sign of the focused particles, and therefore the dominant sign of the neutrinos produced. The nominal running mode of the BNB is "neutrino mode" where the horn focuses positively-charged particles, producing a predominantly neutrino beam.

The focused charged pions and kaons enter an air-filled cylindrical decay region after passing through the horn system. This region provides about 50 meters of length for the charged particles to decay before reaching the steel and concrete absorber. The charged pions and kaons decay predominately into muons and muon neutrinos, where the muons subsequently decay to produce both muon and electron neutrinos. In the rest frame of each of the decay particles the neutrinos

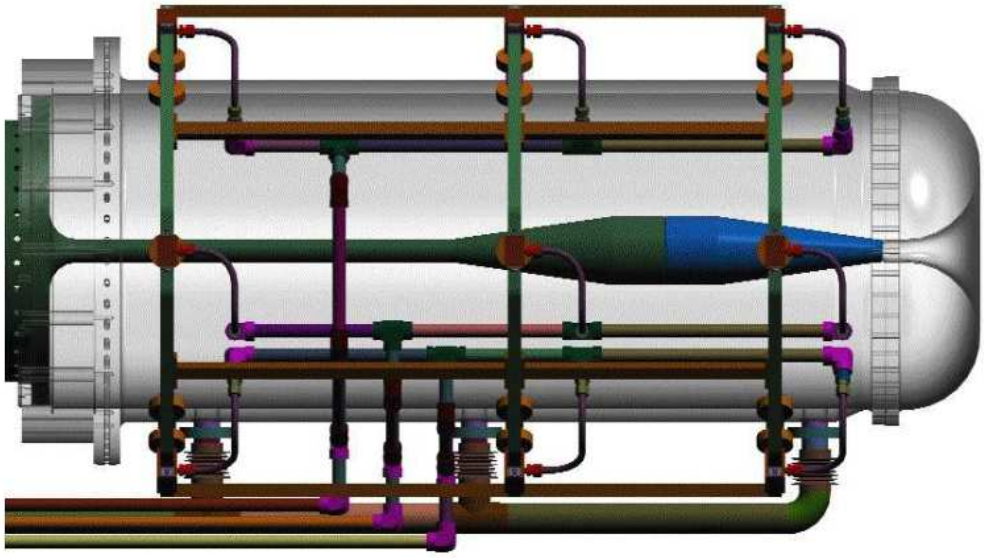


Figure 3.6: Schematic diagram of the BNB horn system. The horn system is used to focus the charged pions and kaons produced in the target and to direct them towards the decay pipe. Image taken from [13].

are produced isotropically, but due to the boost of the parent particle the neutrinos are produced in a forward cone in the lab frame oriented towards the detectors of the SBN program. The BNB neutrino production channels and their branching ratios are shown in Table 3.1.

The flux of neutrinos at each detector location is shown in Figure 3.7. Note that the rate at SBND (left) is 20-30 times higher than at the MicroBooNE and ICARUS detectors. The far detectors also have a broader ν_μ spectrum as a result of the narrower solid angle viewed by each detector relative to the near detector, though this is not expected to have a significant impact on the joint analysis systematics. The ν_e/$\bar{\nu}_e$ fluxes have a more similar shape in all three detectors. In neutrino mode, the flux is dominated by ν_μ ($\sim$ 93.6%), followed by $\bar{\nu}_\mu$ ($\sim$ 5.9%), and with an intrinsic ν_e / $\bar{\nu}_e$ contamination at the level of 0.5% below 1.5 GeV. The primary source of the ν_μ flux is pion decay-in-flight except above $\sim$ 2 GeV, where kaon decay becomes the dominant source. The pion-to-muon decay chain is also the primary contributor to the intrinsic ν_e flux, with the remainder coming from K^+ and K^0_L decays.

Table 3.1: Table showing the BNB neutrino production channels and associated branching ratio.

Particle	Lifetime [ns]	Decay Mode	Branching Ratio [%]
π^+	26.03	$\mu^+\nu_\mu$	99.9877
		$e^+\nu_e\nu_\mu$	0.0123
K^+	12.38	$\mu^+\nu_\mu$	63.56
		$\pi^0 e^+\nu_e$	5.07
		$\pi^0\mu^+\nu_\mu$	3.35
K^0_L	51.16	$\pi^- e^+\nu_e$	20.333
		$\pi^+ e^-\bar{\nu}_e$	20.197
		$\pi^-\mu^+\nu_\mu$	13.551
		$\pi^+\mu^-\bar{\nu}_\mu$	13.469
μ^+	2196.98	$e^+\nu_e\bar{\nu}_\mu$	100.0

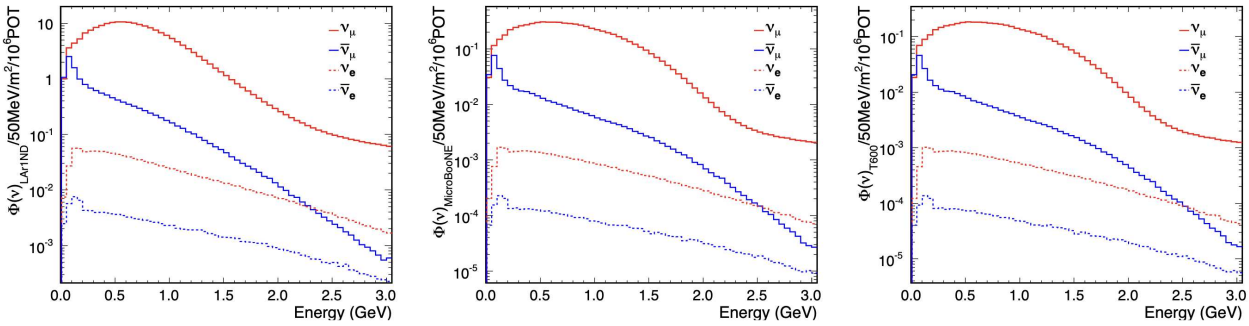


Figure 3.7: The neutrino flux at each detector location along the BNB beamline. The flux is shown for ν_μ and ν_e at the SBND (left), MicroBooNE (center), and ICARUS (right) detectors. Image from [14].

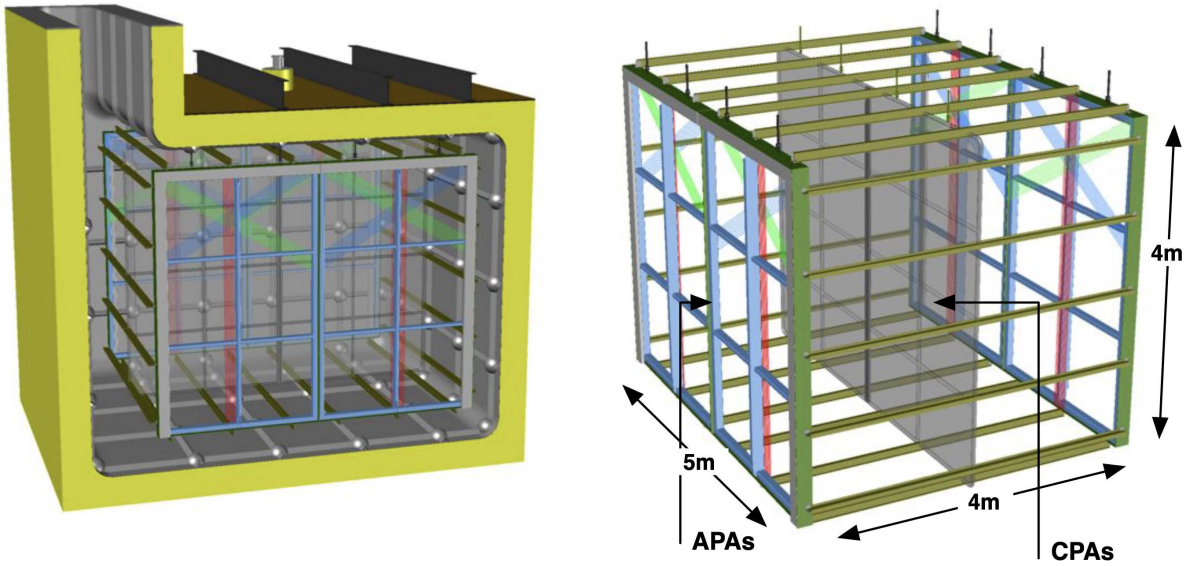


Figure 3.8: Schematic diagram of the SBND detector. The SBND detector consists of two independent TPCs separated by a central cathode plane. Each TPC is instrumented with two wired APAs. The detector is equipped with a PDS system and a CRT system. Image from [14].

3.2.3 Short-Baseline Near Detector

The design of SBND builds upon the many years of LArTPC detector development from the ArgoNeuT, MicroBooNE, and ICARUS experiments. A diagram of SBND is shown in Figure 3.8. Overall, the design philosophy of SBND is to serve as a prototype for DUNE while also functioning as a physics experiment. The detector is constructed as a 4 meter wide, 4 meter high, and 5 meter long (in the beam direction) volume which is subdivided into two independent TPCs by an opaque central cathode plane. Instrumenting both of the walls opposite the central cathode are two sets of wired Anode Plane Assemblies (APA). Each APA consists of two wired induction planes at $\pm 60^{\circ}$ from the vertical and a collection plane with vertical wires, with all wires spaced at 3 millimeters. The cathode is held at -100 kV yielding a drift electric field of 500 V/cm corresponding to the 2 meter drift length of each TPC and a maximum drift time of 1.28 milliseconds. Each of the 11,264 wires is read out by a custom-designed cold electronics system immersed in the argon and attached to the APA frame.

SBND has a PDS system consisting of a total of 120 photomultiplier tubes (PMTs) and 192 X-ARAPUCAs spread across each wall behind the APAs. Tetraphenyl butadiene (TPB), which

is used to shift the wavelength of the scintillation light from 128 nm to the 420 nm that can be detected by the PDS, is used to coat reflective foils on the cathode. A portion of the PMTs and X-ARAPUCAs are coated with TPB, while the rest remain uncoated. This mixed design makes it possible to separate the reflected and direct components of the scintillation light and can be used to improve resolution of the interaction along the drift direction. The PDS system is used to provide a trigger for the TPC readout and to provide a precise timing reference for the event reconstruction.

Due to the location of the detector on the surface, SBND is equipped with a full-coverage Cosmic Ray Tagger (CRT) system. Each CRT module consists of extruded scintillator strips read out by silicon photomultipliers (SiPMs). The design of SBND includes seven CRT planes: six to cover each side plus an additional one placed 3 meters above the first top CRT plane to form a "CRT telescope" that provides better angular resolution for cosmic muons entering the top of the detector. The full coverage of the CRT system will provide tagging of cosmic muons entering, exiting, or crossing the detector. The CRT system provides important input for cosmic background rejection and for selecting cosmic muon samples for calibration purposes.

3.2.4 MicroBooNE

The MicroBooNE experiment was proposed as a follow-up experiment to the MiniBooNE experiment primarily to address the observation of an excess of events consistent with ν_e at low energies. A diagram of the MicroBooNE detector is shown in Figure 3.9. The MicroBooNE detector is constructed as a 2.6 meter wide, 2.3 meter high, and 10.4 meter long (in the beam direction) volume with a single TPC drift volume. The cathode plane is located on one wall of the detector and held at a voltage of -70 kV, yielding a drift electric field of 273 V/cm corresponding to the 2.6 meter drift and a maximum drift time of 2.3 milliseconds. The original target electric field was 500 V/cm, but instabilities in the cathode high voltage led to the detector running at the lower drift electric field. Despite this, the 273 V/cm drift electric field was sufficient for physics of interest. The TPC is instrumented with three wired anode planes opposite the cathode which are

oriented at $\pm 60^\circ$ and 0° from the vertical. Each of the 8,256 wires is read out by a custom-designed cold electronics system immersed in the argon [51].

Located behind the anode plane are 32 PMTs which are used to detect the scintillation light produced by interactions. TPB-coated acrylic plates are placed in front of the PMTs to shift the wavelength of the scintillation light to wavelengths that the PMTs can detect. The PMTs are used to provide a trigger for the TPC readout and to provide a precise timing reference for the event reconstruction. The MicroBooNE detector is also equipped with a CRT system consisting of extruded plastic scintillator strips. There are six CRT planes corresponding to each side of the detector, which provides tagging of cosmic muons entering, exiting, or crossing the detector with nearly full coverage.

MicroBooNE ran successfully from 2015 to 2020 and accumulated five full datasets corresponding to more than 1.3×10^{21} protons on target. The experiment has published results on a variety of topics including numerous detector physics measurements [52–55], differential cross sections for ν_e [56] and ν_μ [57] interactions on argon, and searches for physics beyond the Standard Model. The experiment has also published results on the search for a low-energy excess showing inconclusive evidence for a sterile neutrino or other beyond the Standard Model physics [58–62]. The full SBN Program will be needed to conclusively resolve the short-baseline neutrino anomalies.

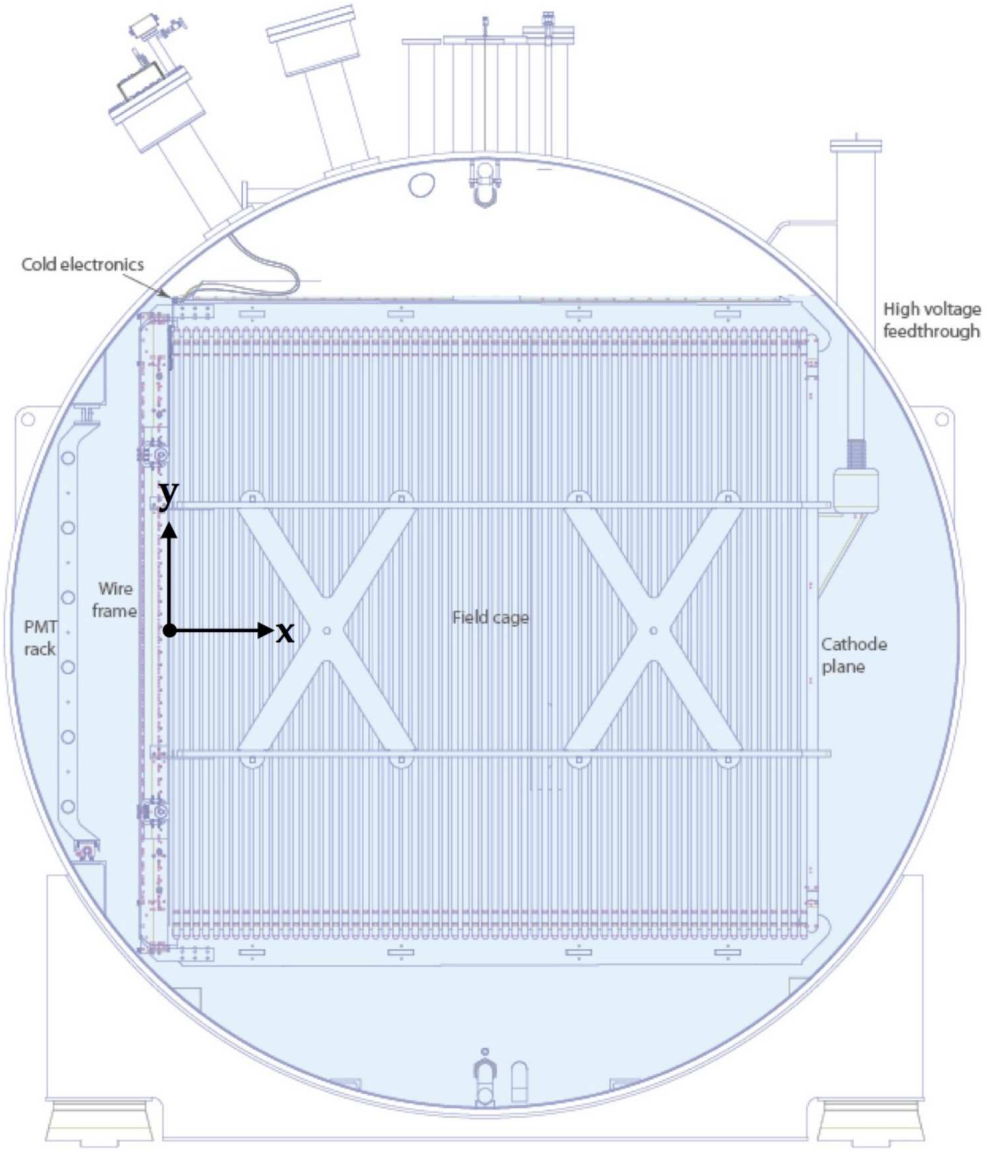


Figure 3.9: Schematic diagram of the MicroBooNE detector. The MicroBooNE detector consists of a single TPC drift volume with three wired anode planes opposite the cathode. The detector is equipped with a PDS system and a CRT system. Image from [15].

Chapter 4

The ICARUS T600 Detector

The ICARUS T600 detector is the world's first large-scale LArTPC detector, and at the time of construction the largest detector of its kind ever built. Its construction realized many years of research and development by the ICARUS collaboration with successively larger prototypes. This chapter will provide an overview of the history of the ICARUS T600 detector (Section 4.1) and the design of each of the three subsystems that make up the detector: the TPC system (Section 4.2), the photon detection system (Section 4.3), and the cosmic ray tagger (Section 4.4). A summary of the trigger system and data collection is provided in Section 4.5. The TPC noise characteristics of the detector will be discussed in Section 4.6, including the measurement and simulation of the noise. The performance of the noise model is evaluated by comparing the noise characteristics in data to Monte Carlo simulation in Section 4.6.3.

4.1 History of the ICARUS T600 Detector

After the original proposal of the LArTPC technology for imaging of particle interactions, an extensive research and development program was undertaken to demonstrate the feasibility of the technology. This program included ten years of small-scale studies of LAr volumes as a proof of concept and to serve as test beds for LAr purification methods, read-out schemes, and electronics, as well as five years of studies with prototypes of increasing mass and complexity. These prototypes demonstrated the stability of the detector technology and its ability to image particle interactions. The ICARUS T600 detector was the culmination of this program and was designed to be a large-scale LArTPC with a liquid argon mass of 760 tons.

The assembly of the ICARUS T600 detector began in 1999 in Pavia, Italy. The T600 detector is constructed as two separate cryostats, each with internal dimensions of 3.6 meters in width, 3.9 meters in height, and 19.6 meters in length. Each half-module, called a "T300," houses an inner detector made of two TPCs (sharing a common cathode), a field-shaping cage, and a PDS. The

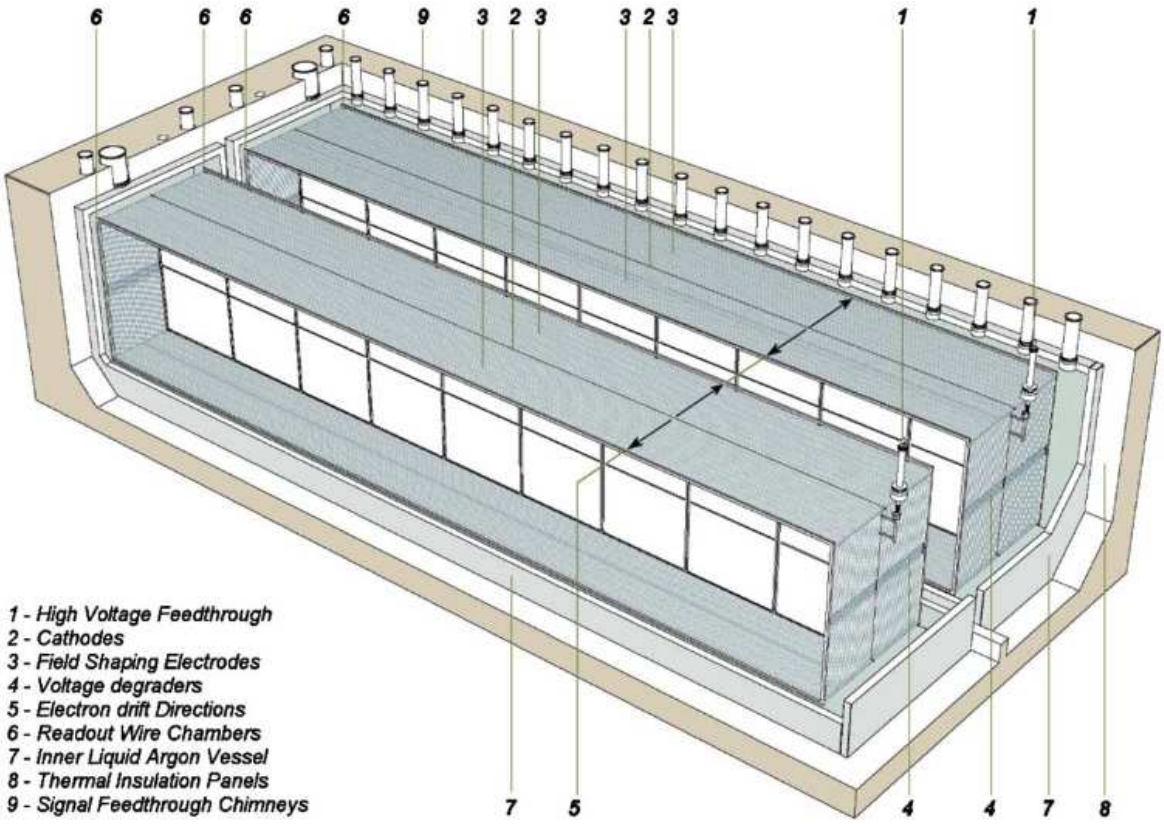


Figure 4.1: An artist cutaway view of the ICARUS T600 detector. The two T300s are shown side-by-side with the cryostat and the TPCs visible. Image from [16].

dimensions and shape of the two T300s were defined by the requirements that the containers be transportable through the Italian highways to the underground Gran Sasso National Laboratory (LNGS). The design of the T600 detector is shown in Figure 4.1.

The first of the two T300s was brought into operation in 2001 with an initial three-month test run with cosmic rays on the surface, allowing for the first time an extensive study of the main detector features [16]. After this successful test, the detector was de-commissioned and, in 2004, the full T600 was transported to Hall B of the underground LNGS. Numerous commissioning activities were necessary to bring the detector to full operation, which occurred in the first months of 2010. This marked the start of a three-year run of the T600 detector in the CERN to Gran Sasso (CNGS) neutrino beam [63]. The physics program included setting limits on ν_e appearance [64, 65], refuting the claim of super-luminal neutrinos from the OPERA experiment [66, 67], and a demonstration of electron lifetime exceeding 15 ms [68], in addition to the general success at

Figure 4.2: The ICARUS T600 detector after installation in the SBN Far Detector building at Fermilab. Image from Fermilab Creative Services.

operating the large detector and reconstructing neutrino interactions with unprecedented detail [63].

The T600 detector was decommissioned starting in June 2013. The detector was then transported to CERN at the end of 2014 for a significant overhaul in view of its future as part of the SBN Program at Fermilab. The T600 was moved to Fermilab in 2017 where it was installed in the SBN Far Detector building in August 2018. Cold commissioning and subsequent filling with liquid argon began in February 2020 and was completed by around May 2020. The detector activation took place in August 2020 with the raising of the cathode and anode planes to their nominal voltages, and the first cosmic events were subsequently recorded. An image showing the T600 detector after installation at Fermilab is shown in Figure 4.2.

4.2 TPC System

Each of the T300 submodules contains a pair of TPCs with drift volumes bifurcated by a central cathode. The anode for each TPC is made of three parallel wire planes separated by 3 mm, with the stainless-steel 100 µm wires of each plane oriented at different angles to the vertical axis. The first induction plane has wires oriented horizontally, whereas the middle induction plane and the collection plane have wires oriented at ± 30 degrees from the vertical axis. The wires of the first induction plane are split vertically at the center of the TPCs forming two separate wire planes with a length of 9 m each. The other two planes are continuous and have wires with a length of up to 3.8 m.

The wires are connected to the readout electronics through a series of feed-through flanges that are distributed along the length of the TPCs. A schematic illustrating the end of one of the TPCs is shown in Figure 4.3. Wires are connected in groups of 32 to cables, which serve to transmit the signal on the wires up to the feed-through flanges and to the readout crates. A readout crate holds nine readout boards, each having the capability to digitize signals from two cables for a total of 64 channels. The TPC electronics are described in more detail in [69] and [70].

Due to the angle of each wire plane, there are necessarily some wires with no attachment point at the top edge of the detector. These wires are instead accessed through a single feed-through flange on the corner of each TPC that hosts three readout crates. Two of these crates digitize signals from the front induction plane, whereas the other digitizes signals from either the middle induction plane or the collection plane depending on which plane on that end of the detector has wires inaccessible from the top. The cables are assigned to the readout boards in such a way that each of the two cables of a given board are connected to spatially separate regions of the wire plane.

The remaining cables from the middle induction plane and the collection plane pass through eighteen flanges along the length of the detector. Each readout board of these crates is connected to exactly 32 middle induction plane wires and 32 collection plane wires. This means that not only are the two cables of each board attached to wires from spatially separate regions, but the

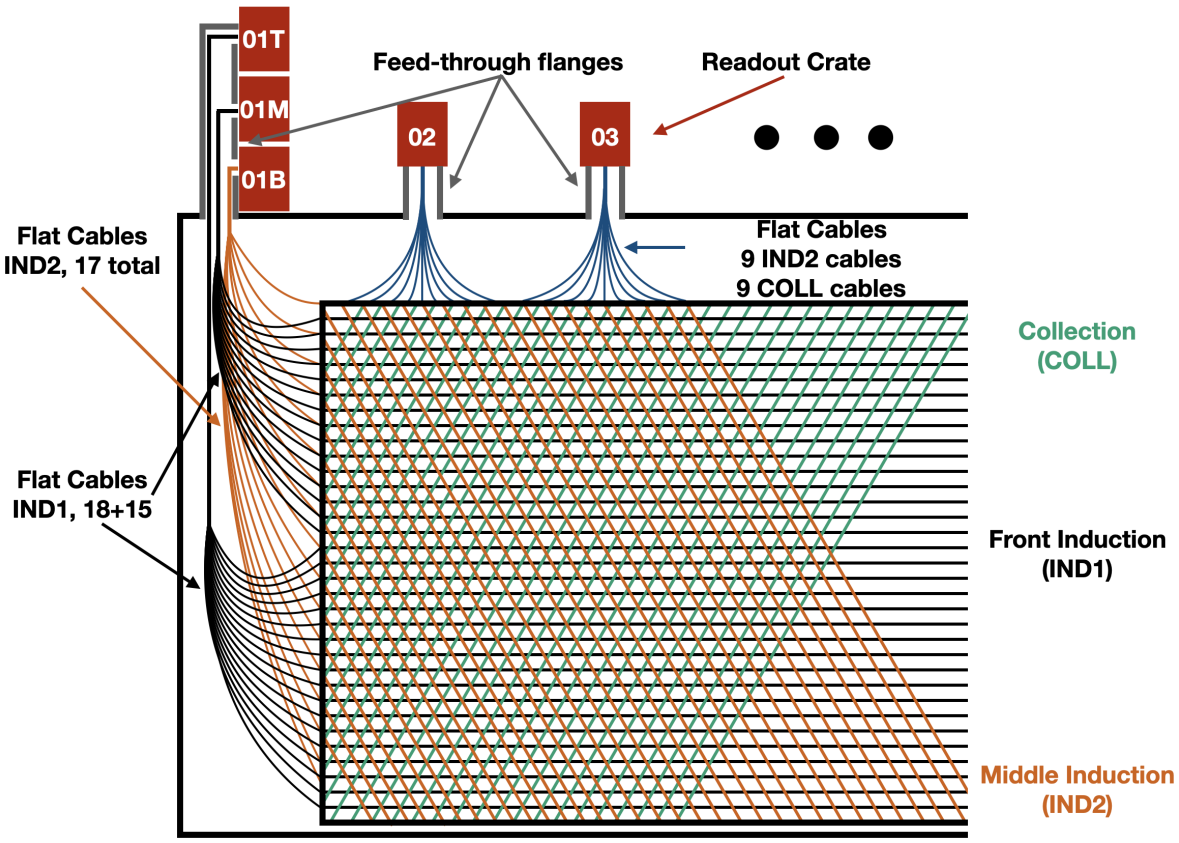


Figure 4.3: A diagram showing the geometry of the ICARUS TPC readout. The wires are connected in groups of 32 to cables and routed up to the feed-through flanges. A readout crate hosts nine readout boards, each digitizing the signal from two cables for a total of 64 channels per board. Each line on the anode plane represents 32 wires spaced by 3 mm.

wires are also from totally independent planes. The fact that there is no single board with all 64 channels from a spatially connected region is relevant for the filtering of coherent noise, which will be discussed in a later section.

The cathode is set at a voltage of −75 kV corresponding to a drift field of 500 V/cm across the 1.5 m drift of each TPC. With a measured electron drift velocity of 0.1572 cm/µs, the maximum drift time for ionization in the detector is 0.95 ms [12]. The wire planes are biased to voltages of −250 V, −30 V, and 250 V for front induction plane, middle induction plane, and collection plane respectively. These wire bias voltages were chosen to maximize transparency of the induction planes to drifting ionization charge.

The electron lifetime has generally been sufficient for high-quality data-taking throughout the operation of the detector. The electron lifetime is measured by fitting the exponential decay of the charge collected on the collection plane as a function of drift time. Figure 4.4 shows the electron lifetime as a function of the calendar date. The steady-state electron lifetime is a function of the gaseous/liquid argon filter efficiencies, the rate of impurity contamination, and the rate of argon recirculation. Each cryostat has a separate system of cryogenic pumps for recirculating argon gas through a purification system, which results in slightly different electron lifetimes for each cryostat due to the above factors. Over time, the filter efficiency drops as the filters become saturated with impurities, thus necessitating the regeneration of the filters. The filters for the West cryostat were regenerated at the end of 2022 and resulted in a significant increase in electron lifetime. A similar increase was observed in October of 2023 for the East cryostat after its filters were regenerated. Most importantly, the electron lifetime has been stable and at sufficient levels for high-quality data-taking throughout each of the physics runs.

4.3 Photon Detection System

The ICARUS T600 photon detection system consists of 360 8" Hamamatsu PMTs deployed behind the three anode wire planes of the detector, for an even 90 PMTs per TPC. This reflects an increase of roughly a factor of four over the number of PMTs installed during the CNGS run.

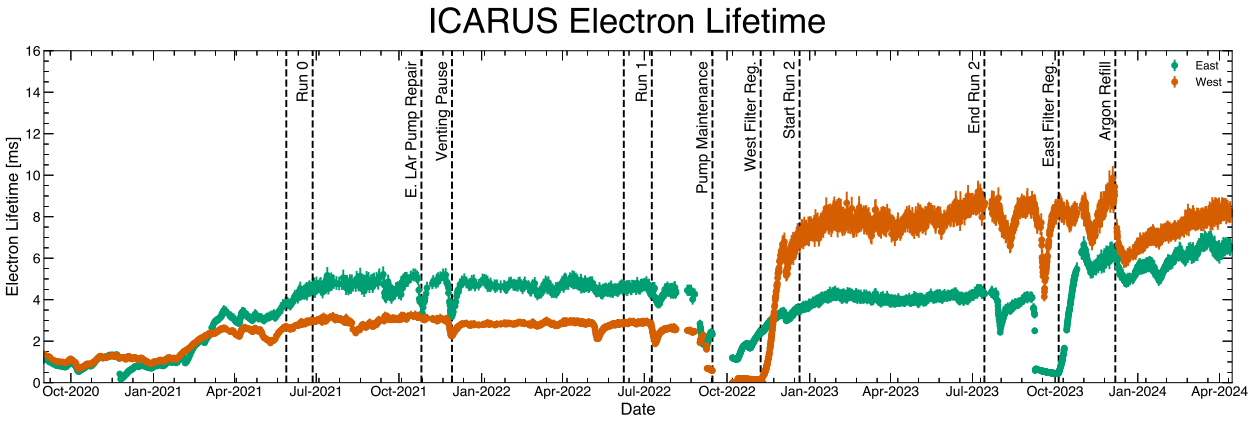


Figure 4.4: The electron lifetime as a function of the calendar date.

Since the glass domes of the PMTs are not inherently transparent to the 128 nm scintillation light produced in the LAr, they are each coated with a thin layer of TPB to shift the light to a visible wavelength. The PMTs are mounted on the wire chamber mechanical frames and are positioned about 5 mm from the collection plane. A stainless steel cage is mounted around each PMT to mitigate the induction of fake signals on the nearby wire planes by the much larger PMT signals. An image of the inside of one of the T300 submodules illustrating the PMTs is shown in Figure 4.5.

The PMTs are accompanied by a dedicated laser calibration system for gain equalization, timing, and monitoring of the PMTs. Laser pulses generated by a laser diode are transmitted via optical fibers to each PMT where it delivers a precisely timed and controlled pulse of light. The timing calibration principally accounts for relative offsets between the PMTs due to hardware differences, and can be further refined to the sub-nanosecond level using cosmic muons. Minimally this allows for analyses to select interactions occurring within the beam window, but it can also be used to resolve the bunch structure of the beam for further cosmogenic background rejection or for searches looking for heavy beyond the Standard Model particles.

4.4 Cosmic Ray Tagger

The location of the T600 detector on the surface at Fermilab exposes the detector to 11 kHz of cosmic rays, presenting an additional challenge to neutrino reconstruction and selection. Addi-

Figure 4.5: An image of the inside of one of the T300 submodules. The PMTs are visible mounted on the walls of the chamber behind the anode wire plane. The anode wire planes are only visible from the reflections glancing off the wires.

tional shielding from cosmic rays is provided by a 3 m concrete overburden situated above the detector. This overburden is designed to reduce the background contributions from charged hadrons and high-energy photons, which are attenuated significantly by the concrete. The overall rate of cosmic activity is reduced by about half with the overburden in place, but it has relatively little effect on the cosmic muon flux.

In order to efficiently identify neutrino interactions among this large background of cosmic rays, the ICARUS T600 detector is equipped with a cosmic ray tagger (CRT) system which provides a nearly 4π solid angle coverage of the detector. The CRT system consists of three distinct subsystems, each using slightly different technology: the top CRT, side CRT, and bottom CRT. The top CRT sits above the detector and beneath the concrete overburden and consists of 123 detector modules covering the full top and top rim of the detector. It is designed to intercept more than 80% of the cosmic muon flux. Each module is composed of two orthogonal layers of scintillator bars read out by wavelength-shifting fibers and SiPMs. The orthogonal arrangement of the scintillator bars in the module leads to better spatial resolution on each CRT hit. The top CRT is a brand new system designed specifically for the ICARUS T600 detector.

The side CRT subsystem deploys the salvaged scintillator modules that were formerly used by the MINOS experiment. Each module consists of adjacent $800\,\text{cm} \times 4\,\text{cm} \times 1\,\text{cm}$ strips of polystyrene scintillator with wavelength-shifting optical fibers embedded in the center and read out by SiPMs. The side CRT subsystem is double-layered consisting of an inner and outer wall. On the South side of the detector these layers are oriented orthogonally to each other, whereas the layers on the other sides are parallel and horizontal.

The bottom CRT subsystem is composed of two daisy chains of seveb modules each that are positioned underneath the warm vessel. These modules were repurposed from the veto modules of the Double Chooz reactor neutrino experiment [71]. Each of the modules consists of two layers of polystyrene scintillator strips oriented parallel to each other. Unlike the top and side CRT subsystems, the scintillator strips of the bottom CRT subsystem are not read out by SiPMs, but rather by PMTs. Due to hardware failures and difficulty integrating the bottom CRT subsystem

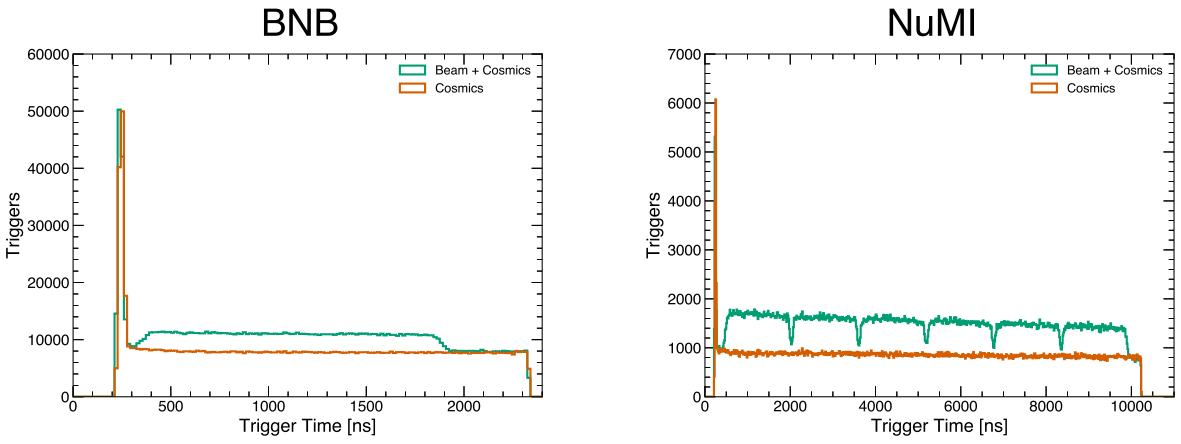
into the DAQ, the system was not included in the first two physics runs of the ICARUS T600 detector.

4.5 Data Collection

Data collection at ICARUS begins with the issuance of a trigger from the ICARUS trigger system corresponding to a series of conditions that must be met. At the highest level, there are two types of trigger requirements employed at ICARUS: light-based and fixed. The first type, called the "Majority" requirement, implements a simple logic on the number of PMT pairs that have simultaneously seen an optical signal above a certain threshold. The latter type, called the "Minimum Bias" or "MinBias" requirement, simply issues a trigger every N beam gates irrespective of the light activity in the detector. The Majority trigger is used for the collection of neutrino events, whereas the MinBias trigger is used for trigger efficiency studies.

In the case of the Majority trigger requirement, the PMTs on each wall of the TPC are divided into adjacent pairs, then each wall is divided into consecutive longitudinal slices of length 6 m (30 PMTs). At the boundaries of these divisions the trigger efficiency is reduced due to the light signal being split between adjacent slices, but this is mitigated by introducing further 6 m slices centered at each of these boundaries ("overlapping windows").

The Majority trigger condition has three controllable parameters: the per-PMT threshold, the number of PMT pairs required in the beam window, and the number of PMT pairs required outside of the beam window. Only one of the latter two conditions needs to be met, in combination with one of the two beam windows, in order to issue a trigger. In the most recent physics run (Run 3), the Majority trigger was supplemented with the addition of "adders," which additively stack the signal from the PMTs in certain regions of the detector. This was done to mitigate the scenario in which an interaction produces a large, but localized, light signal that is not seen by enough PMTs to issue a trigger. This is particularly relevant for interactions near the detector boundaries or close to the PMTs.



(a) **(b)**

Figure 4.6: The number of triggers for the BNB (left) and NuMI (right) beams. The green line shows the number of triggers that meet the Majority requirement within the beam window, whereas the orange line shows the number of triggers meeting the same requirement, but outside of the beam window. Triggers which are related to the beam (i.e. neutrinos) are present as an excess in the green distribution with respect to the orange distribution.

Due to the off-axis location of the detector with respect to the NuMI beamline, it is exposed to both the BNB and NuMI beams, with beam windows that are 1.6 µs and 9.6 µs long, respectively. The trigger system receives the "Early Warning" signal from the Fermilab Accelerator Division for each of these beams a fixed time before the protons reach the target, thus setting the start of the beam window. A fixed time after each beam window, a second "off-beam" window is opened that is wholly separate from the neutrino beam. Triggering on this window provides an accurate estimate of cosmogenic backgrounds which meet the Majority requirement ("in-time cosmics") and can be used for background subtraction in analyses. Figure 4.6 shows the distribution of triggers for the BNB and NuMI beams for the on-beam and off-beam Majority data streams. The excess of triggers in the on-beam window with respect to the off-beam window is indicative of neutrino interactions. The time resolution of the trigger time is not sufficient to see the individual buckets of each beam, but the 1.6 µs length of the BNB beam window is clearly visible. In NuMI, the three unfilled buckets are visible as the gap between subsequent 1.6 µs bunches. The higher intensity of the NuMI beam results in a larger on-beam excess relative to the off-beam window.

Table 4.1: A summary of the physics data-taking runs for ICARUS.

	Run 1	Run 2	Run 3
Start date	June 9, 2022	December 20, 2022	March 14, 2024
End date	July 10, 2022	July 16, 2023	–
BNB POT [$\times 10^{20}$]	0.411	2.049	–
NuMI POT [$\times 10^{20}$]	0.679	2.742	–
BNB collection efficiency [%]	93.2	95.0	–
NuMI collection efficiency [%]	92.9	95.6	–
Has bottom CRT	No	No	Yes
Has overlapped trigger window	No	Yes	Yes
Has adders	No	No	Yes
PMT threshold [ADC]	400	390	390
Majority condition (on-beam, off-beam)	5, 10	5, 9	4, 7

A summary of the three physics runs at ICARUS, including the configurable parameters of the trigger system and the exposure as measured in protons on target (POT), is shown in Table 4.1. Run 3 is still ongoing at the time of writing, so its ultimate POT exposure is not yet known. Beam efficiency quantities are calculated as the ratio between the POT delivered by the Accelerator Division and the POT for which the detector was taking data. Run 2 is currently the largest dataset and reflects a stable detector configuration, and is therefore chosen as the dataset for the analysis presented in this thesis.

The ICARUS data acquisition (DAQ) system utilizes the ArtDAQ framework, which is a toolkit for building high-performance, real-time data acquisition systems [72]. The ArtDAQ framework is built around the conceptual workflow of reading data from detector elements (BoardReaders) and sending it to a central location (EventBuilders) for assembling the detector data into events. The BoardReader processes act as an interface between the lower-level hardware of each component of the detector and the EventBuilder processes.

When a trigger condition is met, the trigger BoardReader pushes its data fragment to an EventBuilder process. Upon receipt of the trigger fragment, the EventBuilder requests data from all other BoardReaders in the system. The BoardReaders of the TPC, PMT, and CRT systems each send their data fragments to the EventBuilder, representing 1.6 ms, 30 µs, and 6 ms windows of

data respectively, thus completing the event. The EventBuilder assigns a unique event number to each event and writes the event to disk. The data is split into different streams based on the trigger condition (Majority vs. MinBias), whether it was on-beam or off-beam, and by the beam type (BNB vs. NuMI). An average event after lossless compression is about 160 MB in size, most of which is TPC data. The DAQ system is capable of stably supporting rates in excess of 5 Hz, though the typical operational rates are closer to 1 Hz or below [12].

4.6 TPC Electronics Noise Characterization and Simulation

The characterization of TPC electronics noise is important for understanding its impact on ionization signals and the reconstruction of particle interactions in the detector volume. The ICARUS TPC electronics noise is characterized principally through measurements of the absolute noise scale, frequency characteristics, and channel-to-channel correlations, as is detailed below (Section 4.6.1). These measurements of the noise are used as input to a data-driven model (Section 4.6.2) implemented in Monte Carlo simulation. The accuracy of this noise model is presented in Section 4.6.3.

4.6.1 Noise Measurement

For measurements of noise, it is important to use waveforms that are free of signal to avoid biases that may result from signal contamination. Data taken with the cathode voltage turned off was chosen for these measurements due to the lack of signal in the waveforms from drifting ionization electrons. A set of waveforms from a single readout crate before any noise filtering is applied is shown in Figure 4.7. The visible horizontal bands represent noise common to groups of 64 channels on the same readout board. The exact source of this coherent noise is not known, but due to the channel grouping it is believed to be generated on the readout board. A similar coherent noise feature was observed at MicroBooNE despite the different front-end electronics used for their TPC readout [73]. There is an additional, smaller coherent noise structure that can be seen as a weak anti-correlation between channels on adjacent boards. This is believed to be due

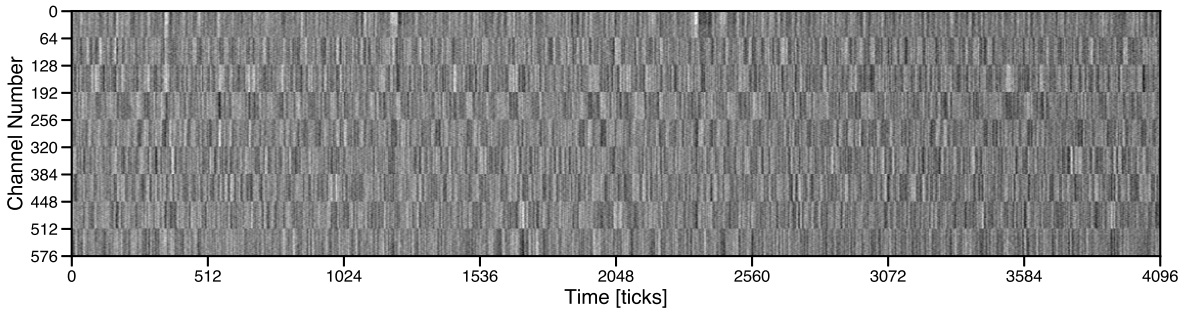


Figure 4.7: An example set of waveforms for a single readout crate. A readout crate contains nine readout boards, each with 64 channels. Data was collected with no electric field applied.

to capacitive or inductive coupling between the cables in the close confines of the feed-through flange.

The channel-to-channel correlation, ρ, can be defined as

$$\rho_{ij} = \frac{\vec{w_i} \cdot \vec{w_j}}{\sigma_i \sigma_j}$$

where $\vec{w_i}$ is the waveform for channel i and σ_i is the root-mean-square (RMS) of its waveform. This can be calculated pairwise for each channel within a readout crate and averaged across many events. Channel-to-channel correlations between channels not in the same readout crate are not significant and are correspondingly not discussed here.

The geometry of the readout motivates two distinct classifications for readout crates: crates which connect only to front induction plane wires and crates which connect to a mix of middle induction plane and collection plane wires. The additional wire and cable length for channels in the front induction plane results in higher overall noise. The cables connecting the wires to the front-end are also in significantly closer proximity due to the path down to the wires and the presence of three sets of cables in a single feed-through flange.

Figure 4.8 shows the channel-to-channel correlations for channels within the same readout crate. The left plot shows only channels belonging to readout crates serving front induction wires, whereas the right plot shows only channels belonging to readout crates which serve a mix of middle induction and collection wires. The main block diagonal structure of highly-correlated channels

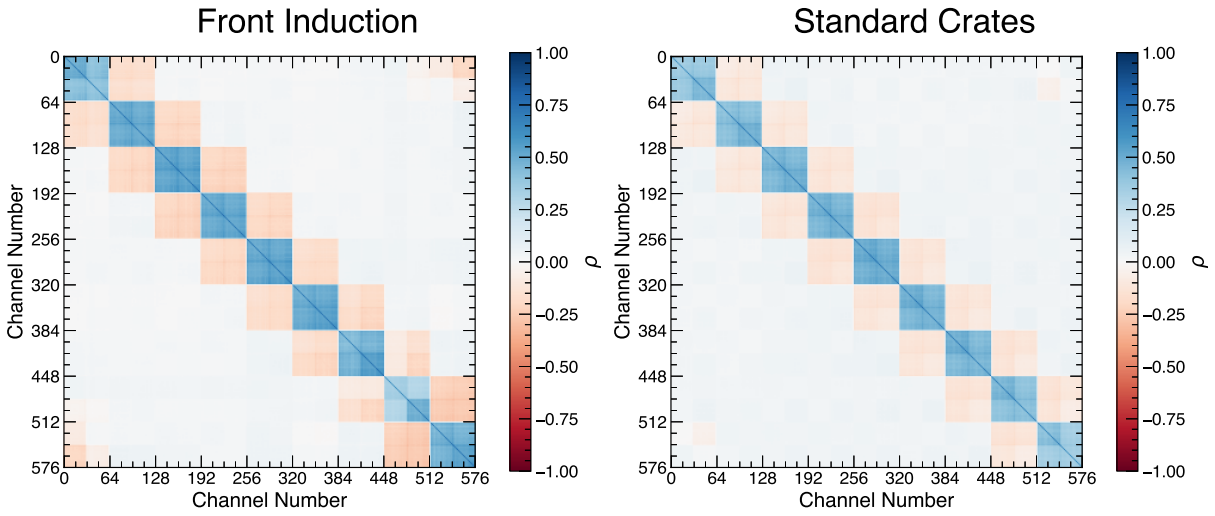


Figure 4.8: Correlation matrix of noise across each readout crate. Correlation coefficients are calculated pairwise for channels in the readout crate, then averaged across all selected readout crates. The left plot shows only crates which connect to front induction plane wires, whereas the right plot shows the crates which connect to a mix of middle induction plane and collection plane wires.

reflects the presence of noise that is coherent across channels of the same readout board. The off-diagonal structure of anti-correlation is believed to be due to capacitive or inductive coupling between the cables of adjacent boards. This effect is observed to be stronger for the front induction plane, which is consistent with the closer proximity and higher path overlap of cables for these wires. Though each cable represents 32 wires, the broader correlated component for channels on the same readout board introduces anti-correlations are nearly uniform across the full group of 64 channels.

The presence of these significant correlations necessitates some degree of noise filtering to maintain reconstruction fidelity. ICARUS employs a coherent noise removal algorithm similar to the one used by MicroBooNE [73]. The coherent noise removal algorithm operates on a group of channels and first defines the coherent noise component for that group using the median value of the waveforms for each time tick. This produces a waveform which is expected to represent noise fluctuations that are common to the entire group, which are visually identifiable in images of the raw waveforms. The resulting waveform is then subtracted from each waveform in the group to produce a set of corrected waveforms.

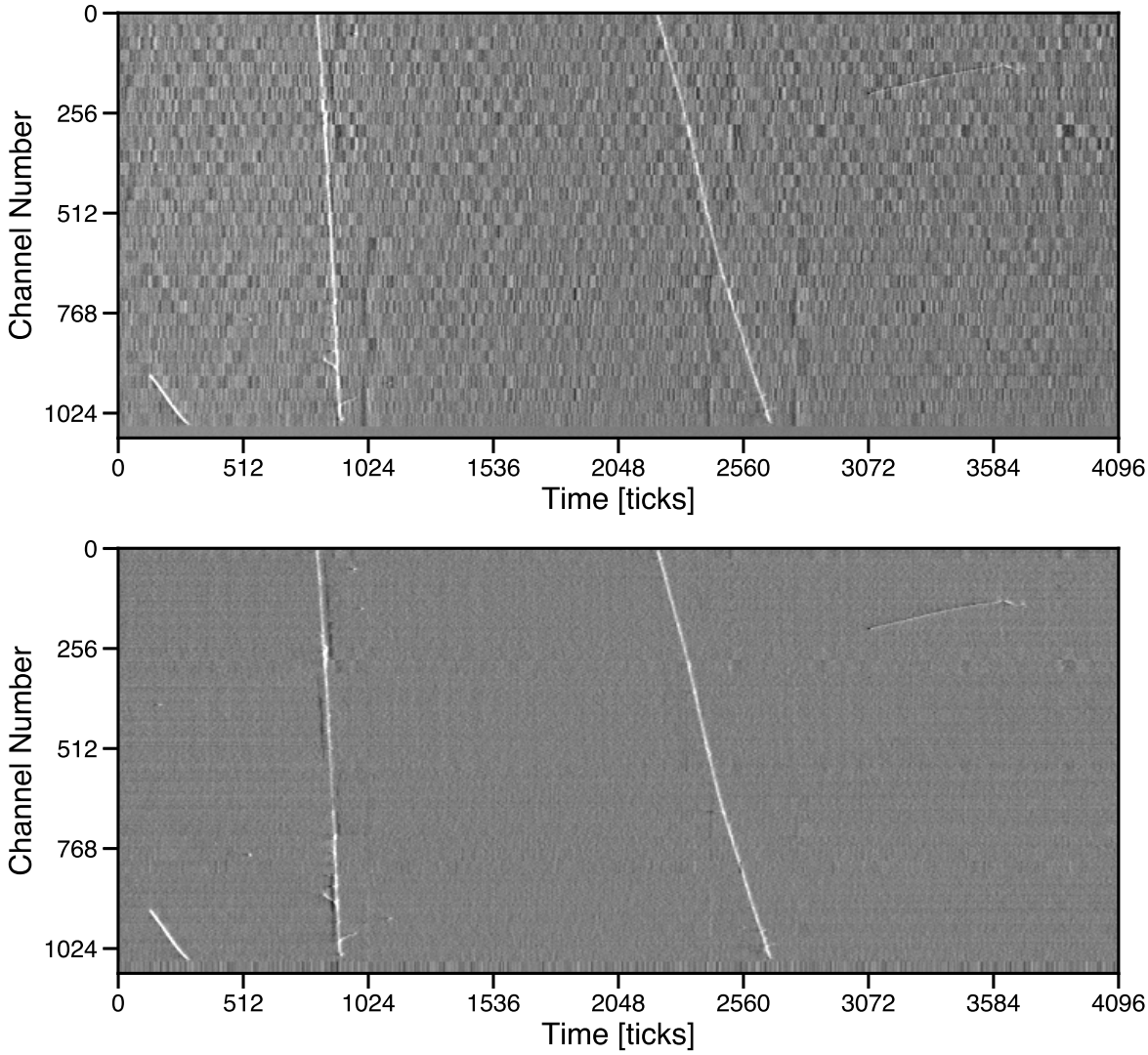


Figure 4.9: An example set of waveforms containing signal from cosmic muons before and after noise filtering. The image spans the South half of the front induction plane from one of the TPCs.

At ICARUS, channels belonging to the same readout board are used to define the groups for this noise filtering as motivated by the noise correlation matrices. A downside of this algorithm is that signal from tracks which are isochronous across the group of channels may be impacted as the track itself is coherent in the same manner as the targeted noise. This is partially mitigated by the fact that no group of 64 channels contains signal from a single spatially connected region - each sub-group of 32 channels always sees a distinct region of the detector. Figure 4.9 shows an event display of some cosmic muon tracks before and after the coherent noise removal algorithm is applied. The coherent noise is visible as the vertical bands running through adjacent channels.

The portion of the waveform that remains after the removal of the coherent noise component is representative of the noise naturally present on the channel due to intrinsic noise sources. The separation between coherent and intrinsic components of the noise allows for a more detailed characterization of the noise. The absolute noise levels are measured by the RMS of each waveform before and after the removal of coherent noise using data taken with the cathode voltage turned off. The distributions per plane are shown in Figure 4.10 in both units of ADC and Equivalent Noise Charge (ENC) using the conversion factor of 550 e^-/ADC found in [69]. The front induction plane exhibits significantly higher noise due to the longer flat cables and wires. The middle induction plane and the collection plane have similar noise levels owing to the fact that they have similar wire and cable lengths. The substantial variation in cable length for front induction plane channels drives most of the additional width of front induction plane noise distributions.

The frequency characteristics of each component of the noise can be measured using the discrete Fast Fourier Transform (FFT). Figure 4.11 shows the FFT spectra per plane before ("full") and after ("intrinsic") the coherent noise removal. The underlying intrinsic noise populates the expected Rayleigh distribution, and is similar for all three planes. The coherent noise is present as an additional, less smooth distribution on top of the intrinsic noise. The coherent noise also exhibits two broad peaks at specific frequencies that are not yet attributed to a specific source. At the lowest frequency bins, there is a sharp increase due to low-frequency oscillations in the waveforms. These oscillations are not coherent across groups of channels as evidenced by the full noise and intrinsic

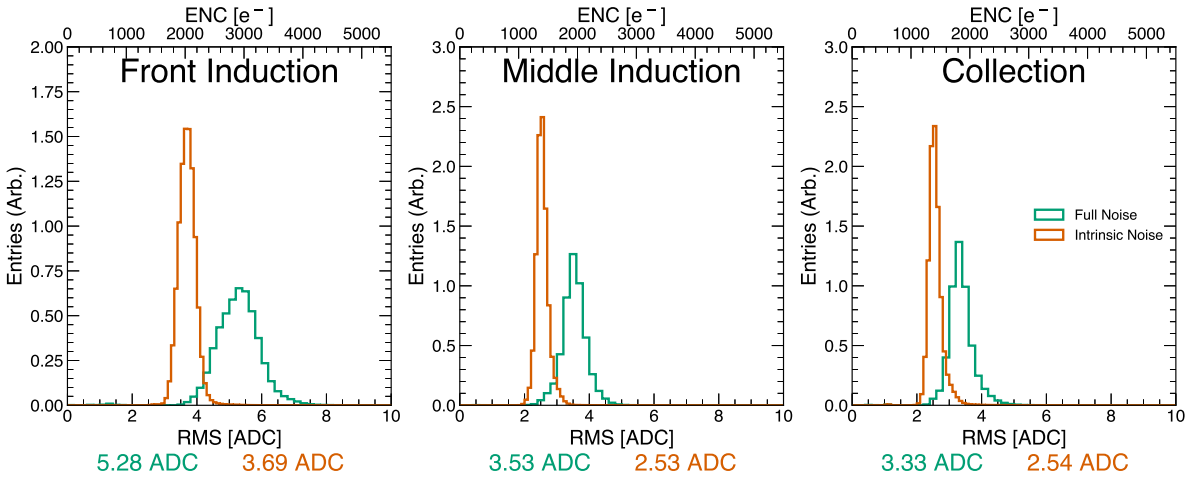


Figure 4.10: Noise distributions per plane as characterized by the RMS of each waveform before (top) and after (bottom) coherent noise removal. Shown below each plot are the medians associated with each distribution.

noise spectra exhibiting the same low-frequency trend. Low-frequency oscillations have also been visually observed on single channels. The spectra for the middle induction plane and the collection plane are nearly identical, but differ greatly from those of the front induction plane. This is can be explained by the different readout geometry and total channel capacitance, the latter of which generally has the effect of scaling the noise.

Outside of hardware interventions, the detector noise has been historically very stable, as shown in Figure 4.12. The biggest source of noise reduction was the disconnection of the Resistive Temperature Devices (RTDs) that read the temperature at various locations within the argon, after which the noise was both significantly reduced and considerably less variable. Due to the location of the RTDs on the inner walls of the cryostat, the collection plane was the most directly exposed plane, though the effect was highly localized and on average larger on front induction plane. There is also a reduction in the coherent noise due to the installation of fifth-order Chebyshev low-pass filters on one of power supply lines of the readout boards. These filters were installed in stages, as annotated in the plot. The general trend of the middle induction plane presenting higher noise than the collection plane is observable in the history of the full noise. This appears to be related to the coherent noise component as the difference disappears after coherent noise removal. While the

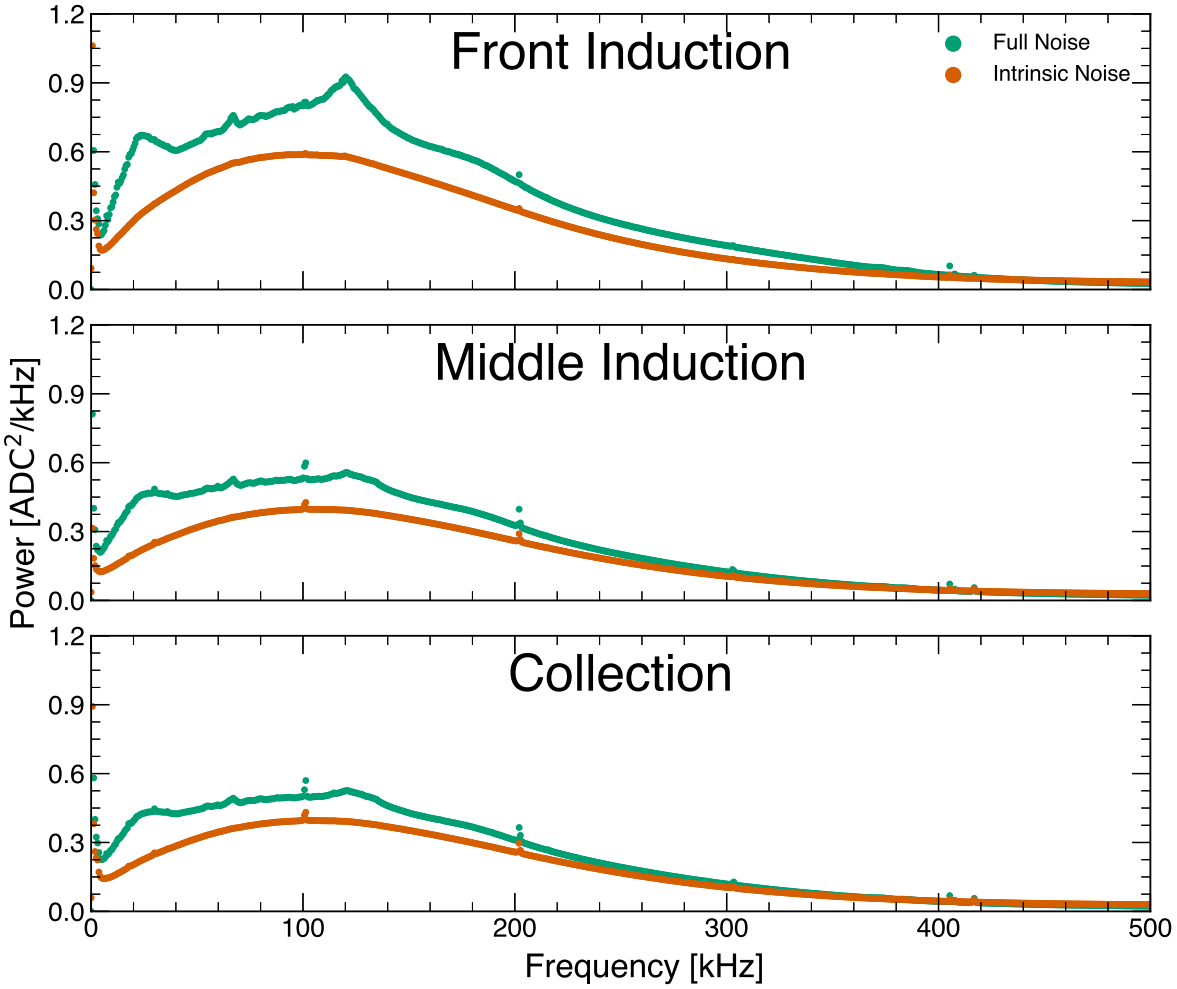


Figure 4.11: Frequency characteristics of the noise per plane as represented by the FFT spectra before and after coherent noise removal.

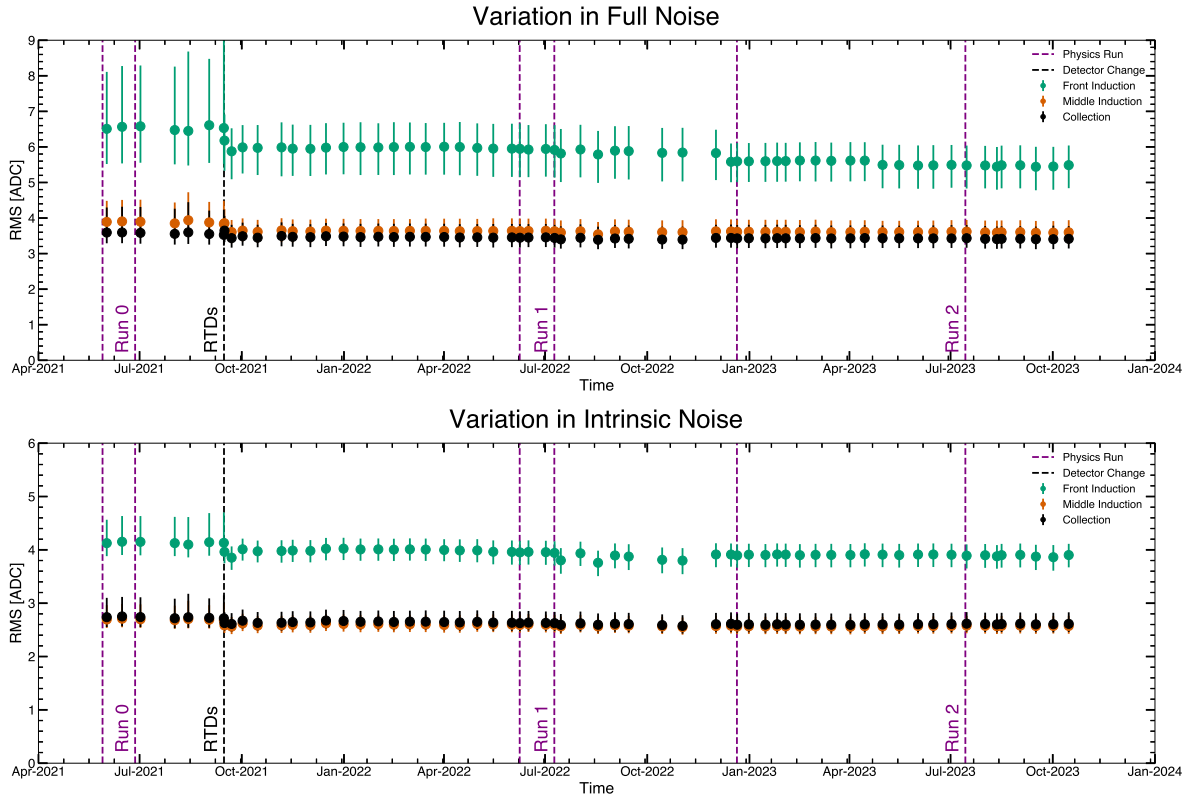


Figure 4.12: Full noise (top) and intrinsic noise (bottom) plotted as a function of time. The uncertainty bars represent the inner 68.27% of the distribution centered at median value.

full noise shows noticeable changes at distinct points in time, the intrinsic noise is exceptionally stable, suggesting that it is only the coherent noise component which has evolved over time. The noise conditions within each physics run are expected to be relatively constant.

4.6.2 Noise Simulation

A data-driven noise model is implemented in simulation using an algorithm provided by the Wire-Cell toolkit [11]. The algorithm allows for the freedom to configure a noise component by defining groups of channels, the noise spectrum associated with each group, and whether the noise component is coherent across the group or intrinsic to each channel. The noise component is simulated for each channel by drawing the amplitudes from the associated spectrum with a randomly chosen phase, then applying the inverse FFT. Coherent noise components have the additional requirement that the phases are shared for all channels within the same coherent grouping. The total

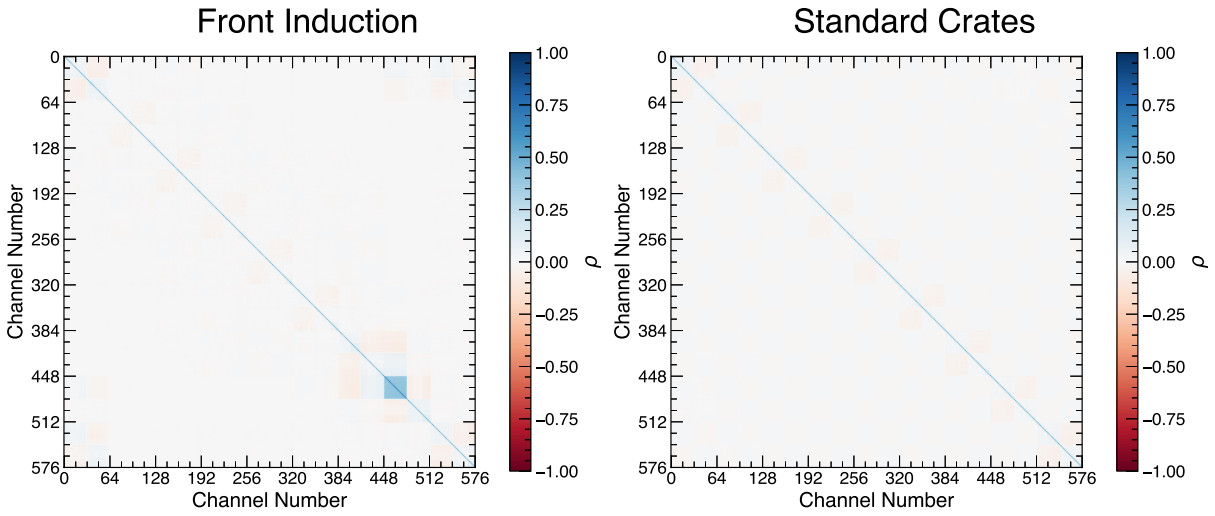


Figure 4.13: Channel-to-channel correlation matrix averaged across all crates serving front induction wires (left) and crates serving a mix of middle induction plane and collection plane wires (right) calculated after removal of the coherent noise common to channels on the same readout board.

waveform for each channel is calculated as the sum of the signal waveform and all noise waveforms from all components.

As discussed in the previous section, the dominant coherent noise component is observed across channels within the same readout board. After removal of this coherent noise, the correlation matrix can be recalculated to determine if further coherent noise sources are present. Figure 4.13 shows the resulting correlation matrices for crates broken into the same two categories described in the previous section. There is a single population of 32 channels in the correlation matrix for front induction with a remaining high-level of correlation. This group corresponds to a set of channels with no physical wires or cable, which necessarily occurs since there are an odd number of cables in front induction. These wires become highly self-correlated after the subtraction of the coherent noise waveform. The remaining portions of the correlation matrices do not show large residual correlations, and therefore no further correlated component is modeled.

The noise model is configured with two noise components: an intrinsic one that is uncorrelated and a coherent one that is correlated across the full group of 64 channels. In both cases, the input spectra reflect the average of the group of 64 wires.

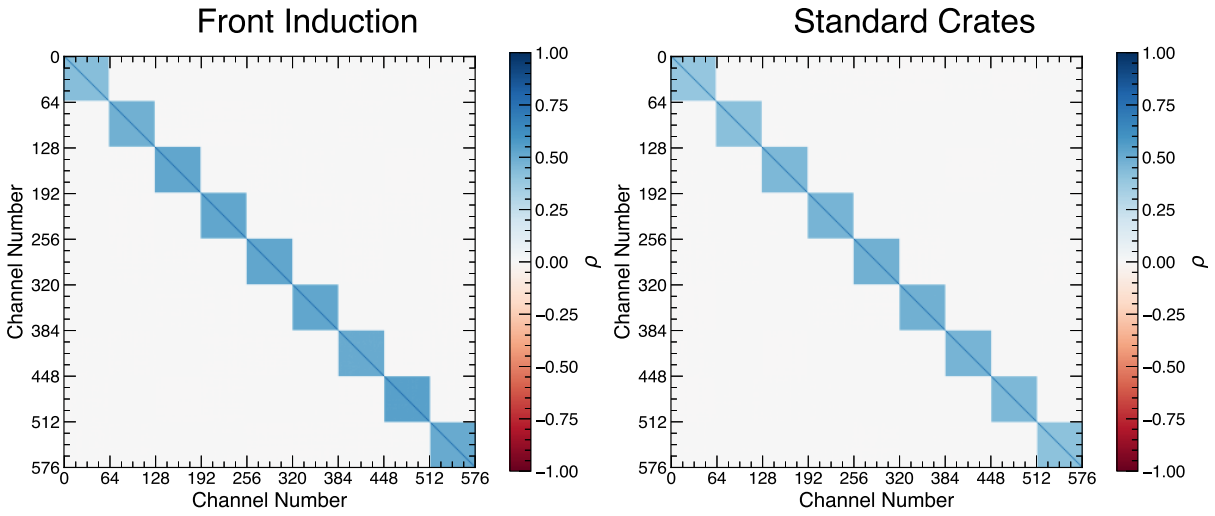


Figure 4.14: Channel-to-channel correlation matrix from Monte Carlo simulation averaged across all crates serving front induction plane wires (left) and crates serving a mix of middle induction plane and collection plane wires (right).

4.6.3 Comparison of Data and Monte Carlo Simulation

After configuring the noise model, it is used to generate a sample of Monte Carlo simulation events for a comparison with data. The data sample used for this comparison is the same as was used to generate the input spectra. Both the Monte Carlo simulation sample and data sample were analyzed in exactly the same way, so the only differences in the results are expected to be from any inaccuracies in the noise model.

The correlation matrices for front induction plane crates and standard crates is shown in Figure 4.14. The noise model accurately reproduces the coherent component of the noise common to channels within the same readout board. The anti-correlated component between channels of adjacent readout boards is not modeled and therefore not reproduced. This anti-correlated component observed in data is itself coherent across the readout board and is removed by the coherent noise removal process, so this is not expected to have a noticeable impact on the overall noise levels.

The RMS calculated per channel and per event can be used to characterize the accuracy of the noise model in modeling the absolute noise scale. Figure 4.15 shows these distributions per plane before (top) and after (bottom) coherent noise removal. The full noise shows good agreement for front induction, but there is a systematic bias in the middle induction plane and the collection

plane. Each readout board in a standard crate contains equal amounts of middle induction plane and collection plane wires, so the noise model effectively models the average of the two. This, along with the fact that the middle induction plane exhibits slightly higher noise, results in a bias in opposite directions for both planes. After coherent noise removal, the agreement between data and Monte Carlo simulation improves significantly, and the bias shown by middle induction and collection is reduced as demonstrated in Figure 4.16. It is worth noting that the waveforms after coherent noise removal are used downstream in the reconstruction, so inefficiencies in the noise model that appear in the full noise and not in the intrinsic noise are expected to have only second-order effects.

The noise spectra in data and Monte Carlo simulation can be compared to verify that the same spectra that were used as input to the simulation are reflected in the produced waveforms. Figure 4.17 shows a comparison of the average FFT spectra per plane for data and Monte Carlo simulation. The shapes are reproduced nearly identically, but there are some minor discrepancies in the overall magnitude. The front and middle induction planes are consistently slightly under-predicted by Monte Carlo simulation, whereas the collection plane is slightly under-predicted after coherent noise removal. These observations are consistent with the results from the RMS noise comparisons shown earlier.

Further characterization of the noise model accuracy can be done by examining two different metrics: the event-to-event variations and the channel-to-channel variations in the noise levels. The event-to-event variations are calculated as the difference of the measured RMS value from the median for the corresponding channel. The width of this distribution is driven by statistical fluctuations from the combination of frequency components of random phases and by short-scale temporal variations in the noise itself. The overall width of the data distribution can be represented as the quadrature sum of the Gaussian widths associated with the two underlying processes. The Monte Carlo simulation, which uses a static noise distribution, is only broadened by the statistical fluctuations. By comparing data and Monte Carlo simulation, the additional Gaussian broadening necessary to match Monte Carlo simulation to data, parameterized by σ_t, can be extracted and used

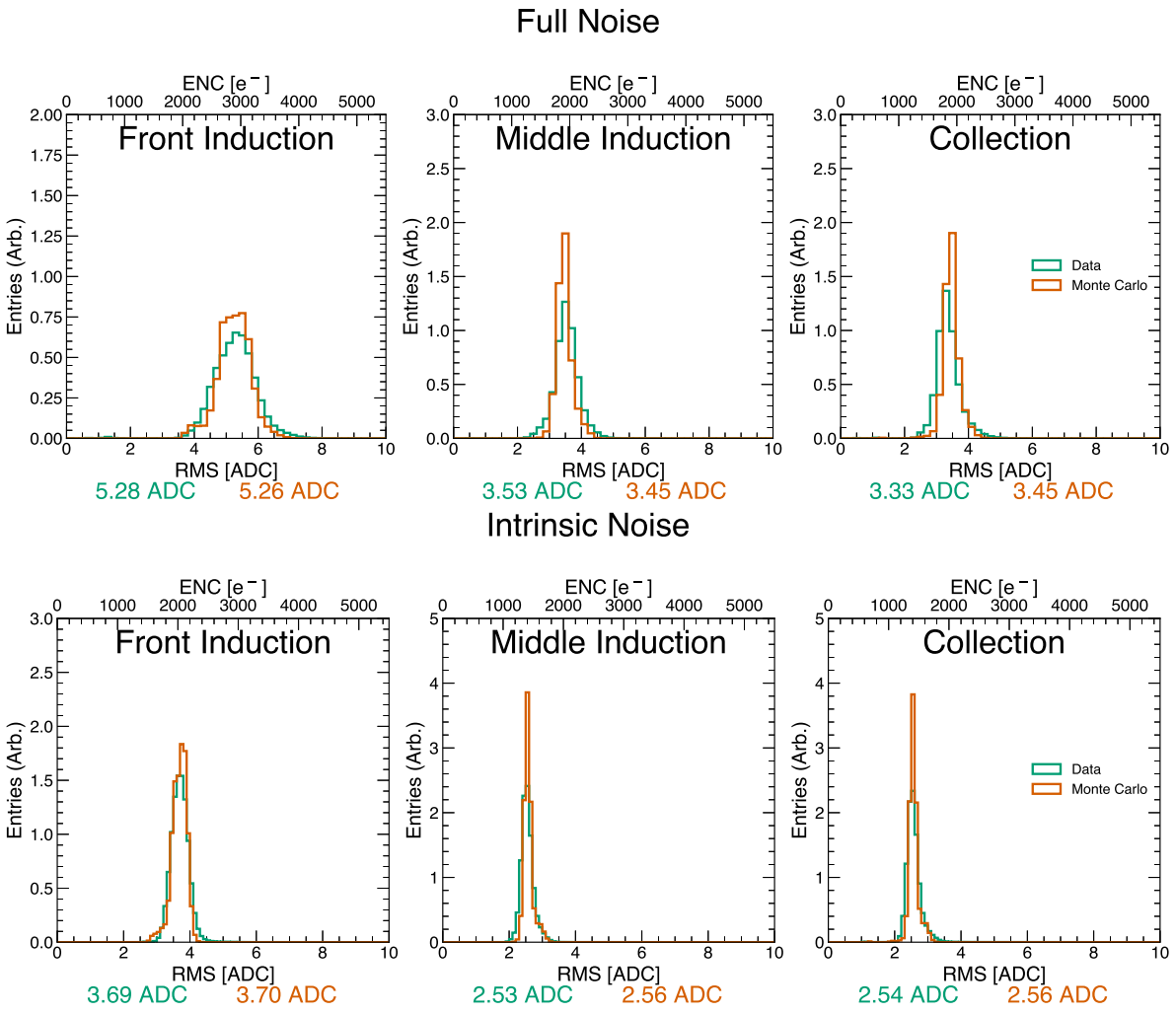


Figure 4.15: Noise distributions in data and Monte Carlo simulation per plane as characterized by the RMS of each waveform. The top plot shows the full noise and the bottom plot shows the noise after coherent noise removal. The reported values are the medians associated with each distribution.

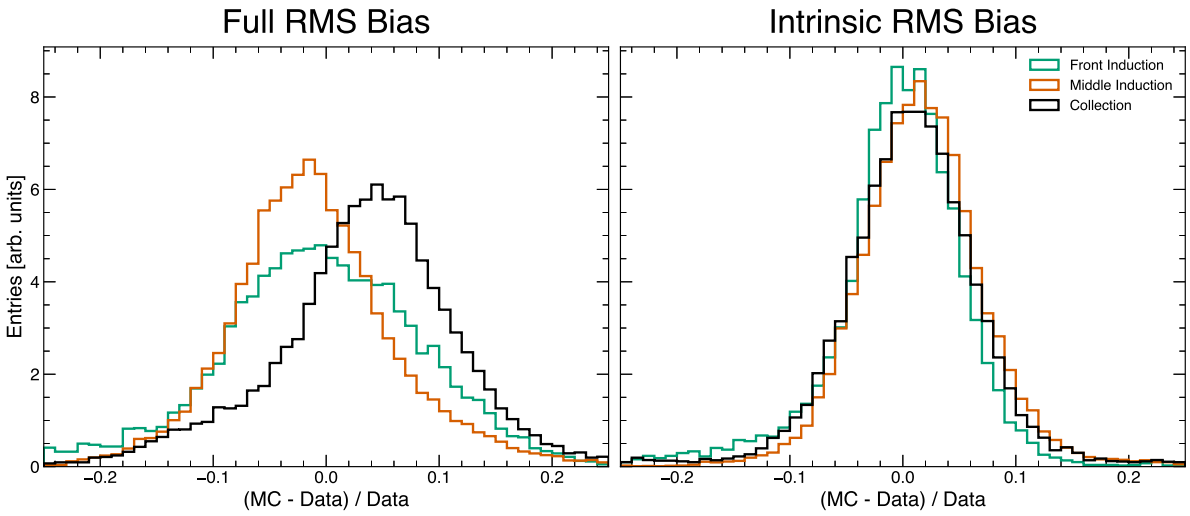


Figure 4.16: The bias between data and Monte Carlo simulation calculated as the same-channel difference in RMS normalized by RMS in data. The left plot shows the bias in the full noise and the right plot shows the bias in the intrinsic noise.

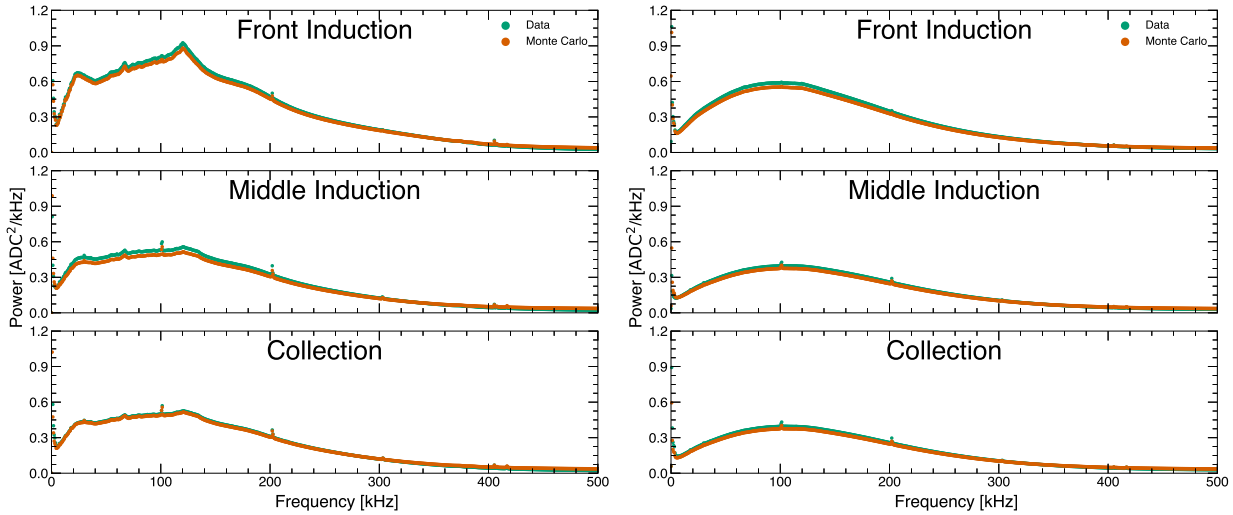


Figure 4.17: Frequency characteristics of the noise in data and Monte Carlo simulation per plane as represented by the FFT spectra. The left plot shows the spectra for the full noise and the right plot shows the spectra after coherent noise removal.

to directly characterize the temporal variations in the noise. Figure 4.18 shows these distributions and the associated results for the full noise and the intrinsic noise for all three planes.

The channel-to-channel variations are calculated as the difference of the measured RMS value from the median for the group of 64 channels. In addition to the statistical and temporal variations discussed above, this distribution experiences broadening from the spatial variation in noise levels across channels within the same readout board. As before, the quadrature difference of the widths of the data and Monte Carlo simulation distributions, now parameterized as σ_s, characterizes the inaccuracy of the noise model at modeling spatial variations smaller than the 64 channel grouping used to configure each noise component. These distributions are shown in Figure 4.19. The channel-to-channel variations highlight the bias in the full noise observed previously in the middle induction and collection planes that results from modeling the average of two planes with slightly different noise levels. This disappears after coherent noise removal, which suggests the discrepancy between the two planes is a coherent noise source.

The relative sizes of σ_s and σ_t suggest that spatial variations in the noise are the dominant contribution to mis-modeling of the noise ($\sigma_s > \sigma_t$). The discrepancies between data and Monte Carlo simulation shown in Figure 4.15 are primarily drive by these unmodeled spatial variations. Correspondingly, the temporal variations in the noise over short time scales are negligible and do not need to be modeled. Further improvements to the noise model should target spatial variations in the noise by decreasing the group size for the intrinsic noise or by adding additional coherent components for smaller group sizes.

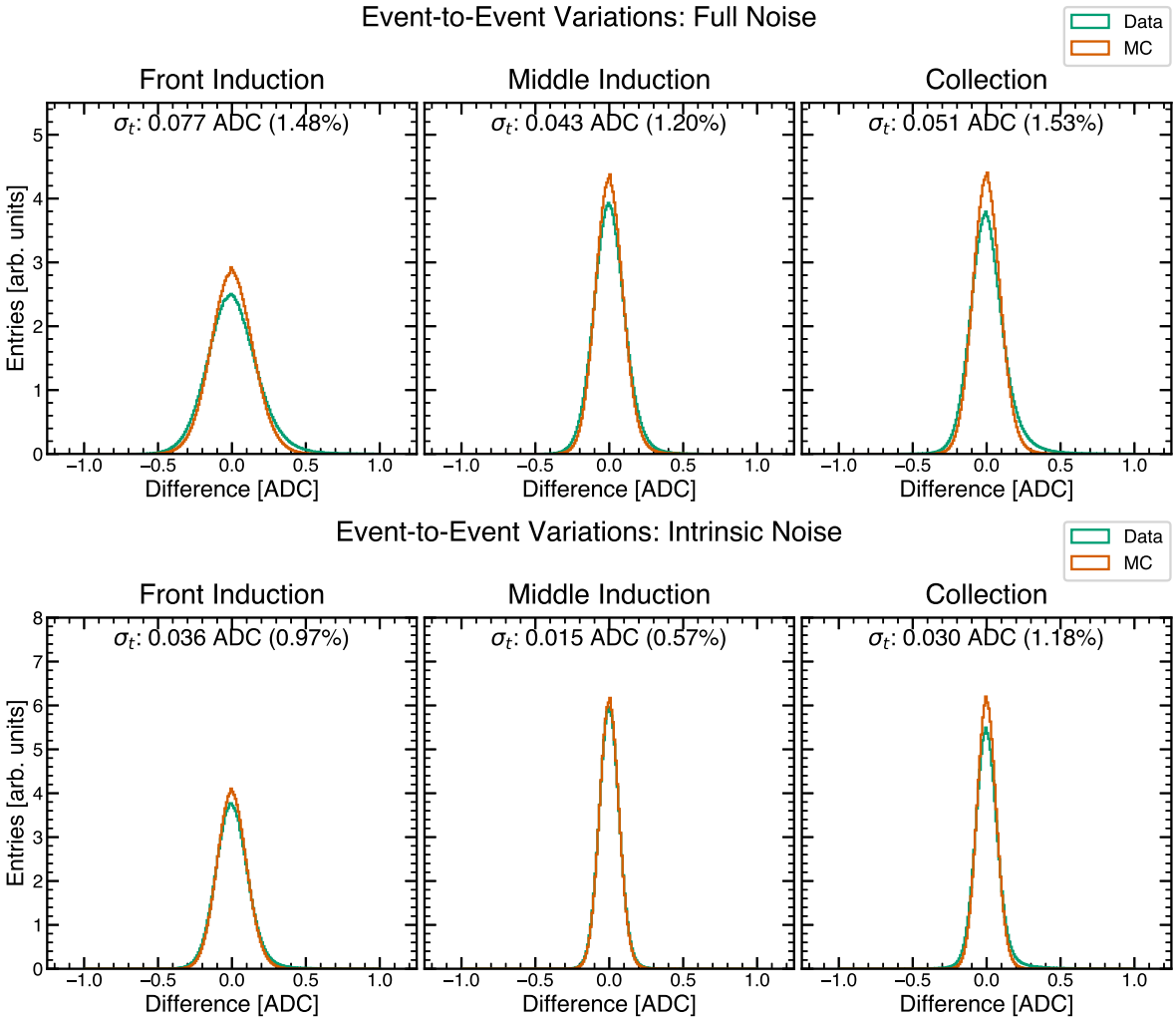


Figure 4.18: Noise variations in data and Monte Carlo simulation. Event-to-event variations are calculated per-event as the difference between the measured RMS and the median RMS for the channel. The reported values represent the Gaussian smearing necessary to match Monte Carlo simulation to data and the corresponding percentage is the result after normalizing by the median noise on the plane. The top plot shows the variations in the full noise and the bottom plot shows the variations in the intrinsic noise.

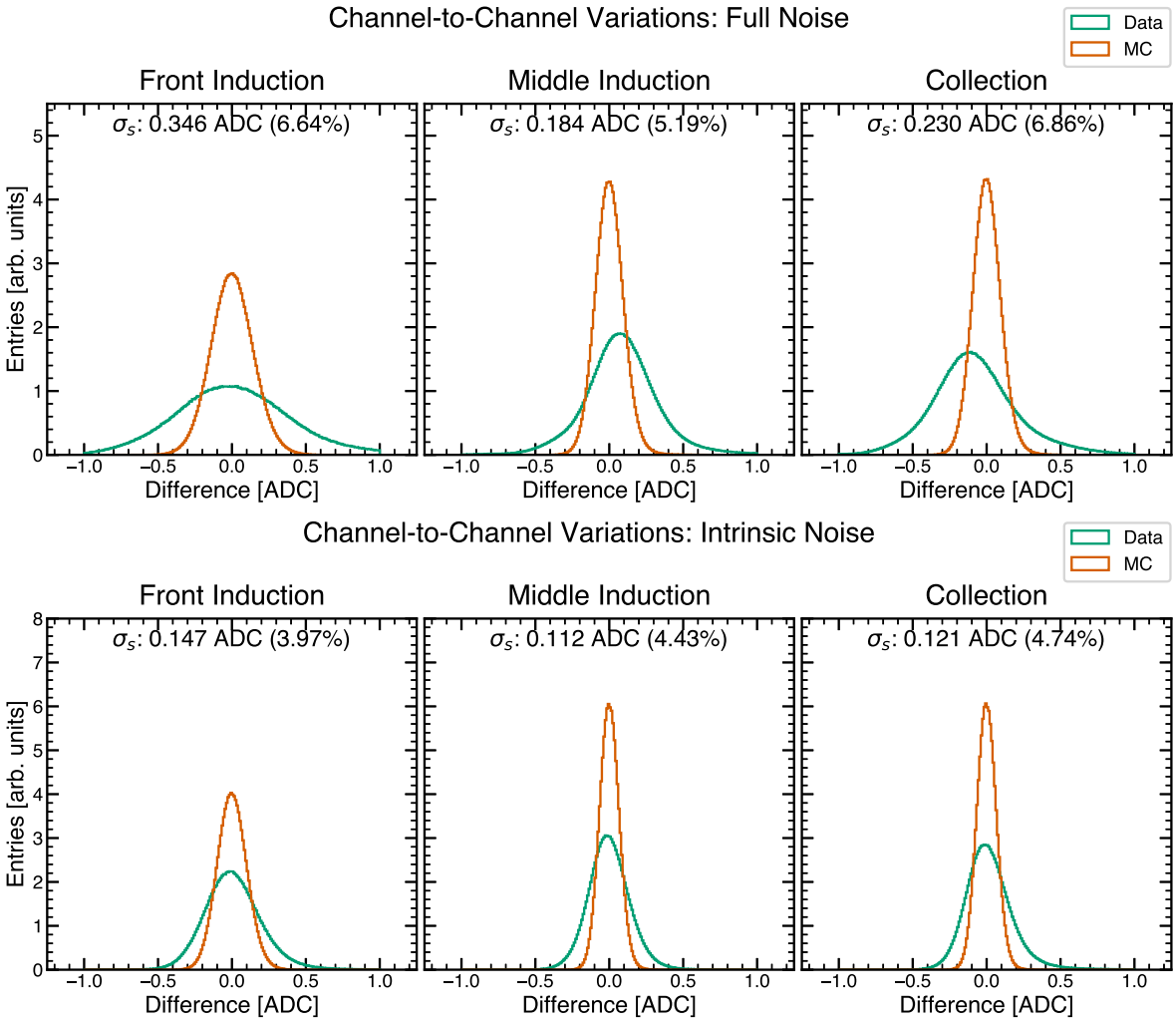


Figure 4.19: Noise variations in data and Monte Carlo simulation. Channel-to-channel variations are calculated per-event and per-channel as the difference of the measured RMS from the median RMS for the entire group of 64 channels. The reported values represent the Gaussian smearing necessary to match Monte Carlo simulation to data and the corresponding percentage is the result after normalizing by the median noise on the plane. The top plot shows the variations in the full noise and the bottom plot shows the variations in the intrinsic noise.

Chapter 5

Simulation and Reconstruction

This chapter describes the full simulation and signal processing chain used in ICARUS. Precise simulation and reconstruction of neutrino interactions is crucial for the overall sensitivity of analyses in neutrino experiments. Section 5.1 provides an overview of the full simulation path used in ICARUS, including the beam flux simulation, the neutrino interaction simulation, and the cosmic ray simulation. Section 5.2 describes the TPC signal processing chain used in ICARUS, including the deconvolution of the raw waveforms, the hit finding algorithm; the optical flash reconstruction is also discussed in this section.

5.1 The ICARUS Simulation

The ICARUS simulation is a Monte Carlo simulation designed to model neutrinos, neutrino-induced particles, and their interactions in the detector with high fidelity. The simulation serves to predict the expected signal and background rates observed in data in the detector along with their associated kinematic distributions. The simulation of events is broken into several stages. First, the beam flux simulation produces the expected number, flavor, and energy of neutrinos from the beam source (Section 5.1.1). Next, the neutrino-argon interaction simulation and the cosmic ray simulation produce the expected final-state particles and their kinematics (Section 5.1.2). These particles are then propagated through the liquid argon medium and any secondary interactions are simulated. Finally, the detector response to the simulated charge and light signals is modeled and used to produce the expected signals in the TPC, PMT, and CRT subsystems (Section 5.1.3).

5.1.1 Beam Flux Simulation

The beam flux simulation used at ICARUS is based principally on the MiniBooNE flux simulation [13]. The neutrino flux prediction is computed using a simulation based on the GEANT4 framework [74] that models the production of particles in proton-target interactions and the result-

ing decay of secondary particles in the decay pipe. A set of histograms is produced that describe the energy, momentum, and spatial position of neutrinos of each flavor. These flux histograms are then sampled in the next stage to generate neutrinos for the neutrino interaction simulation.

5.1.2 Event Simulation

This stage of the simulation handles the generation of final-state particles from neutrino interactions and from cosmic ray interactions in the detector. The GENIE neutrino event generator [75] is used to simulate neutrino interactions in the detector using the neutrino flux computed in the previous stage as input. The GENIE event generator factorizes the neutrino interaction process into a series of models: the nuclear model that describes the ground state of the struck nucleon, the primary interaction model that describes the interaction of the neutrino with the struck nucleon, the hadronization process that describes the formation of the resulting hadrons, and the final-state interactions that occur during the propagation of the hadrons through the nucleus.

The specific version of GENIE adopted for ICARUS is version 3.4 with the AR23_20i_00_00 model configuration. At the time of the writing of this thesis, the technical documentation for this tune does not exist and is being developed. This GENIE tune was designed for common use within SBN and DUNE with a primary goal of providing an extensive phase space for flexible re-weighting of the models. The models used in the ICARUS simulation are summarized in Figure 5.1. The nuclear model employed is the Correlated Local Fermi Gas model, which adds short-range correlations to the local Fermi Gas model describing the initial state of the nucleons, as inspired by [76]. The binding energy of the nucleon is also drawn from a distribution as opposed to a fixed value.

The Valencia model is used to describe quasielastic scattering and accounts for long-range nucleon-nucleon correlations within the nucleus using the random phase approximation (RPA). The "RPA effect" is used to suppress the cross section for CCQE processes at low Q^2 values [77]. The SuSAv2 model is used to describe the CC "two particle, two hole" (2p2h) processes and is currently the only fully relativistic model that can be extended without approximations to the

Table 5.1: Summary of the GENIE models used in the ICARUS simulation.

Interaction	Model
Nuclear Model	Correlated Local Fermi Gas
Quasielastic Scattering	Valencia
2p2h	SuSAv2
Resonance	Berger-Sehgal
Coherent Pion Production	Berger-Sehgal
Deep Inelastic Scattering	Bodek-Yang
Hadronization	AGKY
Final-state Interactions	INTRANUKE hA

full-energy range of interest for present and future neutrino experiments [78–80]. The Berger and Sehgal model [81] is used to model resonant production and coherent pion production, while deep inelastic scattering is described by the Bodek-Yang model [82]. The hadronization process is modeled using an effective low-mass hadronization model known as AGKY whose validity spans the full invariant mass range [83, 84]. Modeling of final-state interactions is handled by an intranuclear cascade model known as INTRANUKE hA [85]. The resulting final-state particles and their kinematics are then passed on to the GEANT4 simulation for propagation through the detector.

In addition to neutrino interactions, cosmic-ray interactions in the detector must also be simulated. The generation of cosmic-ray muons is handled by the CORSIKA cosmic ray simulation package [86]. The cosmic-ray muons are generated based on the expected flux of cosmic rays at the detector site and are passed to the GEANT4 simulation for propagation through the concrete overburden and the detector. Cosmic activity triggers the detector readout more often than neutrino interactions, though this background can be estimated in a data-driven way using the dedicated off-beam data stream originating from only cosmic activity.

5.1.3 Detector Simulation

The simulated particles generated by GENIE and CORSIKA are used as input to a GEANT4-based simulation of the geometry of the ICARUS detector. The simulation of the detector geom-

etry is based on the ICARUS T600 detector design and is implemented in the LArSoft software framework [50]. The GEANT4 simulation takes each primary particle generated by GENIE or CORSIKA and simulates its propagation through the detector in a step-wise fashion. At each step, the path length for each possible physics process, such as scattering or decay, is calculated and the process with the shortest path length is selected. These steps are also forcibly terminated on any physical or logical boundaries in the detector geometry. At each step, the particle deposits some of its energy in the medium. This process continues iteratively until all particles have come to rest or have left the detector volume. The deposited energy is aggregated into a set of 3D cubic voxels of length 0.3 mm for each dimension.

The deposited energy is translated into discrete electron depositions that form the ionization cloud. The initial amount of electrons liberated from the argon is calculated using the Modified Box model, which accounts for the recombination of ionization electrons in the liquid argon medium [87]. The ionization electrons are drifted to the anode plane using the electric field in the detector and accounting for multiple detector effects which can impact the drift of the ionization electrons. These include space-charge effects (SCE), which are the distortion of the electric field due to the buildup of slow-moving positive argon ions [88], the spatial diffusion of the ionization electrons [52], and attenuation due to argon impurities.

After accounting for these effects, the electron cloud is drifted to the anode wire planes and a simulated signal is produced by the Wire-Cell simulation [11]. Though in reality one must account for induced charge signals associated with drifting ionization charge that are collected by neighboring wires ("2D convolution"), the discussion here follows the "1D convolution" simplification. Mathematically, the signal produced in the wire planes is given by the convolution of the time profile of the original charge signal $S(t)$ with the detector response function $R(t' - t)$. The detector response function represents the instantaneous portion of the measured signal at time t' ($M(t')$) due to a component of the original signal at time t. The convolution is given by:

$$M(t') = \int_{-\infty}^{\infty} S(t)R(t' - t)dt \tag{5.1}$$

The detector response function carries within it both the wire field response, which describes the wire's response to the drifting ionization charge signal as dictated by the Shockley-Ramo theorem [89, 90], and the electronics response, which describes the response of the electronics to the wire signal. The response functions may be simulated or estimated from data. At ICARUS, the detector response function was tuned using cosmic muon data, the process of which will be published in an upcoming paper. This convolution process produces a simulated waveform corresponding to the drifted ionization electrons. As a final step, the simulated waveforms are overlayed with simulated noise waveforms to produce the final simulated signal waveforms in the TPC subsystem, as discussed in Section 4.6.2.

The optical signal simulation in the detector uses the energy deposited by the particles to produce scintillation photons. The amount of photoelectrons produced in each PMT is calculated using a photon visibility lookup table that maps the 3D position of the scintillation photons to the expected number of photoelectrons produced in each PMT. The photoelectrons are simulated in the PMT waveforms using a model of the PMT electronics response in addition to the simulated noise waveforms. The CRT subsystem is implemented as a set of individual sub-detectors in the GEANT4 simulation. Particles passing through these regions are used to create the expected signals ("hits") in the CRT subsystem.

5.2 Reconstruction

Event processing at ICARUS begins by taking the raw detector output, whether simulated or real, and transforming it into information that can be used in later stages of the reconstruction chain. For the TPC subsystem, this begins with the raw waveforms and the removal of the coherent noise as described in Section 4.6.1. This produces a set of noise-filtered waveforms that are passed to the signal processing stage. The signal processing stage is realized mathematically as a deconvolution of the noise-filtered waveforms with the detector response function, essentially an inversion of the process modeled by Equation 5.1. One can take advantage of the properties of the Fourier transform to separate the signal from the noise in the frequency domain, recasting

the deconvolution as a division of the Fourier-transformed signal by the Fourier-transformed response function. However, the solution must be extended to account for additional contributions of the electronics noise using a filter function $F(\omega)$ to attenuate frequencies attributed to noise. The signal function in the frequency domain is then given by:

$$S(\omega) = \frac{M}{(}\omega)R(\omega)F(\omega) \tag{5.2}$$

The original signal can then be obtained from Equation 5.2 by taking the inverse Fourier transform. It is worth noting that the deconvolution process specified above assumes that the induced signal in each of the wires is wholly independent. In reality, this is not the case as each wire is sensitive to the ionization charge drifting past neighboring wires. A more realistic model of the signal processing would involve a matrix inversion of the signal in each wire to account for this effect. At the time of this thesis being written, the ICARUS signal processing chain does not apply this 2D deconvolution and instead treats each wire signal as independent, though this improvement is entering its final stages of validation in the experiment. This process produces a set of deconvolved waveforms that represent the original time structure of the current induced by the ionization electrons in the TPC. Further information about this method can be found in the MicroBooNE paper discussing this topic [11].

Segments of waveforms corresponding to physical signals, or hits, are searched for in the deconvolved waveform with a threshold-based hit-finding algorithm. Each hit is then fit with a Gaussian, the area of which is proportional to the number of drift electrons generating the signal. The gain factor provides the conversion between the area of the Gaussian and the number of electrons and has been measured using cosmic muon data separately for each TPC. The efficiency of finding hits corresponding to physical signals is dependent on the noise level in the detector, the angle of the track with respect to the wire planes, and the number of ionization electrons produced by the particle. Isochronous tracks, or tracks which produce hits at nearly the same time across the wires, are most impacted by the coherent-noise-filtering algorithm. The front induction plane has the lowest hit-finding efficiency due to the lower signal-to-noise ratio, whereas the middle induc-

tion plane and collection plane have higher hit-finding efficiencies. Preliminary performance of the hit-finding algorithm at ICARUS has been presented in a paper [12]. The result of this process is a set of hits for each plane that represent the ionization charge deposited by the particle in the TPC. Calibrations are applied to these ionization charge depositions to recover the original energy depositions made by the particle.

The reconstruction of the scintillation light in the PMT subsystem is performed using a similar process. A threshold-based algorithm is applied to each recorded PMT signal waveform in the event to identify optical activity. These regions of optical activity, called "OpHits," are characterized by a start time, a duration, a maximal amplitude, and the integral of the signal. At the second stage of optical reconstruction, all OpHits from any PMT in coincidence within 100 ns are clustered together into an "OpFlash." The OpFlash is expanded to include all OpHits within 1 us of the first OpHit time. The OpFlash can be thought of as the total detected light associated with a single interaction in the detector. The geometric distribution of the OpHits within the OpFlash (across multiple PMTs) can be to associate the scintillation light signal with the ionization charge signal in the TPC.

At this stage of the reconstruction, a set of low-level reconstructed objects has been produced that represents the ionization charge as viewed by all three TPC wire planes and the scintillation light as viewed by the PMT subsystem. These objects are used as input for the higher-level reconstruction algorithms that aim to reconstruct interactions and the kinematics of their associated particles within the detector volume. ICARUS employs two distinct event reconstruction chains: the Pandora reconstruction package based on more traiditional multi-algorithm pattern recognition [91] and a machine-learning analysis chain based entirely on neural networks. The analysis presented in this thesis uses the machine-learning analysis chain, which will be described in the next chapter.

Chapter 6

The ICARUS Machine-Learning Analysis Chain

Traditionally, events in neutrino experiments based on the LArTPC technology are reconstructed using a set of algorithms that are tuned independently and rely on hand-engineered features. In recent years, machine learning (ML) has become a powerful tool in computer science, consistently breaking performance records in computer vision and natural language processing. The highly-visual nature of LArTPC data makes it a natural candidate for machine learning applications: the images of particles are easy to interpret by eye, but it is often hard to programmatically design algorithms to extract high-level features suitable for physics analyses. Moreover, the physics goals of the SBN Program assume high-purity and high-efficiency automatic selections that have so far remained out of reach of traditional pattern recognition techniques. This chapter describes the event reconstruction based on neural networks employed in the ICARUS experiment. Section 6.1 provides an overview of the different types of neural networks used in the ICARUS ML event reconstruction. An overview of the generators used to produce the training sample is provided in Section 6.2. Section 6.3 describes the full ML chain used in the ICARUS ML event reconstruction, from the input data to the final output. Finally, Section 6.4 describes the post-processing steps that are applied to the output of the neural networks to extract the final physics quantities.

6.1 Neural Networks

Artificial neural networks can be described as a family of parameterized functional approximations that are inspired by the structure and function of the biological neural networks in the brain. The basic building block of a neural network is the artificial neuron, which is a simple computational unit that receives a signal from connected neurons and produces an output signal. The output signal is computed by some non-linear function of the weighted sum of the input signals. Typically, neurons are aggregated into layers and the layers are stacked to form a network. Signals

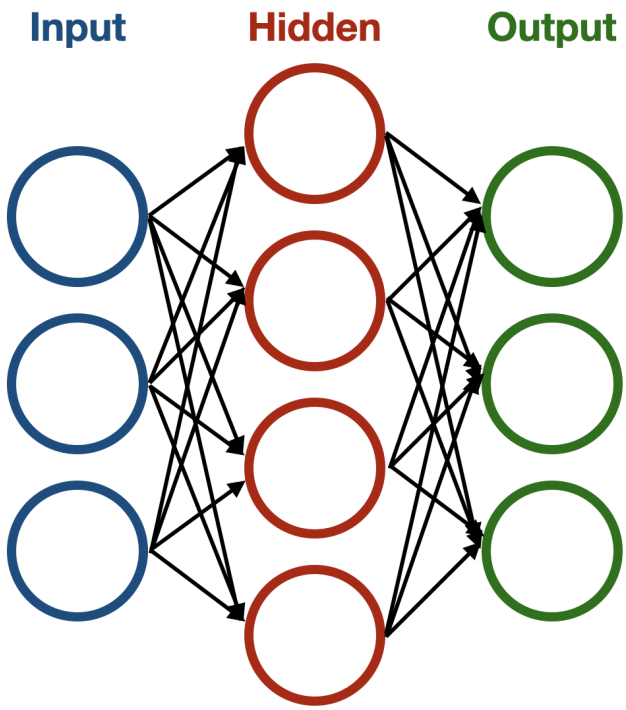


Figure 6.1: Schematic representation of a simple neural network with one hidden layer.

travel from the input layer to the output layer, possibly traversing multiple intermediate "hidden" layers, with the output of one layer serving as the input to the next layer. A neural network is said to be "deep" if it has more than one hidden layer. Figure 6.1 shows a schematic representation of a simple neural network with one hidden layer.

The parameters of a neural network are typically optimized through gradient-based methods that serve to minimize a loss function. The loss function is a measure of the difference between the predicted output of the network and the target value. In a supervised learning setting, the target value is known and the network is trained to predict this value. This necessitates the use of labeled data, where the target value/category is known. For applications in neutrino physics, the role of a training sample is played by simulated data, where the truth information is known from the simulation. The trained network can then be applied to real data, where the target value is not known, to make predictions.

6.1.1 Convolutional Neural Networks

The convolutional neural network (CNN) is a type of neural network that is particularly well-suited for image recognition tasks. Compared to basic neural networks, CNNs make the additional assumption that most of the relevant information exists in a local neighborhood around the signal of interest. This is a natural description of LArTPC data, where the images of particles are highly structured and the local features are important for classification of particles/events. This requirement of sensitivity to local features motivates the use of convolutional layers in CNNs. A convolutional layer computes the dot product between its weights and a small region of the input data. The spatial extent of this region is called the receptive field of the neuron. This convolutional operation is iteratively moved across the entire input image producing an activation map, which may be stacked with other activation maps to obtain the output. Pooling layers are often used in conjunction with convolutional layers to reduce the spatial dimensions of the input data. This is done by aggregating the values of neighboring pixels in the input data with an operation such as max pooling or average pooling. A pooling layer also has the effect of expanding the receptive field as the aggregation method pulls in information from a larger region of the input data.

The restricted spatial extent of the convolutional layers allows CNNs to have fewer parameters than fully-connected networks, which would be computationally prohibitive for images of the size of LArTPC data. The use of shared weights in the convolutional layers also allows CNNs to learn translation-invariant features; that is, features that are detected regardless of their position in the image. This is particular desirable for LArTPC data where the position of the particle and the underlying space points in the detector is not relevant for the classification task.

6.1.2 Graph Neural Networks

Graph neural networks (GNN) are a class of neural networks that are well-suited for datasets with a natural relational structure. The input data in a GNN is represented as a set of nodes and edges, where the nodes represent the entities of interest and the edges represent the relationships between the entities. The key design principle in GNNs is the use of message passing, a process by

which the features of each node and edge are updated by aggregating the features of its neighbors. This is an iterative process performed across the entire graph, thus allowing the network to learn complex relationships between the nodes. Among all graph learning tasks, two are of particular relevance in this thesis:

- **Node-learning tasks**: tasks that involve predicting the classification of a node. In the context of LArTPC data, an example is the classification of a particle as a muon or a primary particle coming out of the interaction vertex.
- **Edge-learning tasks**: tasks that involve predicting the classification of an edge. To relate to LArTPC data, an example is the classification of two particles being part of the same interaction, or of two shower fragments belonging to the same parent shower.

6.2 Training Datasets

The training of neural networks requires a labeled dataset, where the true output is known. In the context of LArTPC data, the labeled dataset is provided by the simulation of the detector response to a set of known particles. In order to avoid potential biases due to assumptions of particular physics models governing cosmic ray and neutrino generation, the reconstruction chain is trained using a set of simulated events produced by two physics-agnostic generators:

- **Multi-Particle Vertex (MPV)**: this generator produces a set of up to N particles originating from a common vertex. These particles include electrons, positrons, photons, muons, anti-muons, both negatively-charged and positively-charged pions, and protons.
- **Multi-Particle Rain (MPR)**: this generator produces a set of up to M isolated particles. These particles include electrons, positrons, muons, anti-muons, and protons.

The MPV generator is meant to emulate a pseudo-neutrino interaction with multiple particles emitted from a common vertex, whereas the MPR generator produces interactions that are similar to those produced by cosmic rays. For each particle, the generators are configured with a range of

Table 6.1: Summary of the neural networks used in the full ML event reconstruction chain in the ICARUS experiment.

Stage	Type	Description
UResNet Deghost	CNN	Classification of space points as reconstruction artifacts or real charge depositions
UResNet	CNN	Semantic segmentation (voxel-level classification of activity)
PPN	CNN	Prediction of start/end points of showers/tracks
Graph-SPICE	CNN	Coarse clustering of space points into particle fragments
GrapPA-Shower	GNN	Clustering of shower fragments into complete showers
GrapPA-Track	GNN	Clustering of track fragments into complete tracks
GrapPA-Interaction	GNN	Clustering of particles into complete interactions with PID and primary designation

kinetic energies than can be assigned as well as a maximum particle multiplicity. The kinetic energy is drawn from a uniform distribution within the configurable range. The maximum particle multiplicity for a given vertex is enforced by only drawing particles of types that are not already at their preconfigured maximum.

The range of all parameters controlling the generation of training events is ideally a superset of the phase space observed in ICARUS data and Monte Carlo simulation. This is to avoid regions of phase space where performance of the reconstruction is worse due to relatively few encounters of events in this region of phase space in the training sample. For example, the reconstruction may struggle to correctly identify low-energy protons if the training sample does not have sufficiently high representation of low-energy protons. The particles generated by these generators are fed into the full detector simulation as described in Chapter 5. This allows for the network to be trained on events that represent detector effects to the extent that they are modeled in the simulation.

6.3 The ICARUS Machine Learning Analysis Chain

The Neutrino ML group at SLAC has developed an end-to-end, ML-based reconstruction chain for the ICARUS experiment. The chain consists of a hierarchical set of neural networks that are optimizable end-to-end. The two types of neural networks described in the preceding section are employed: CNNs and GNNs. The full chain is shown schematically in Figure 6.2 and each stage is summarized in Table 6.1. This section reviews each stage of the chain in order, starting from the voxel-level tasks and building up to the high-level particle- and interaction-level classifications.

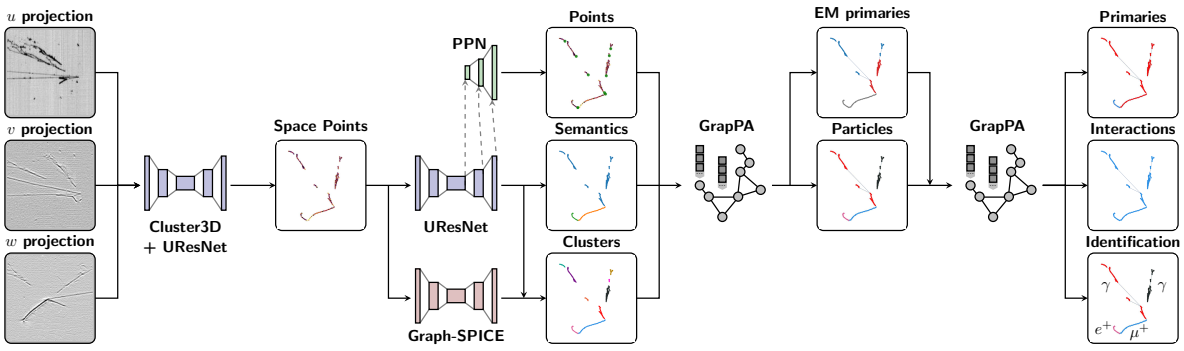


Figure 6.2: Schematic architecture of the end-to-end ML-based reconstruction chain used in the ICARUS experiment.

6.3.1 Tomographic Reconstruction

The reconstruction chain operates on a 3D set of points representing the charge depositions in the detector. Due to the detector's readout being intrinsically 2D, some additional processing is required to reconstruct the input 3D image. The process of reconstructing a higher-dimensional image from a set of lower-dimensional projections is generally known as tomographic reconstruction, and is a concept that is widely used in medical imaging. In the context of LArTPC data, the 3D space points are reconstructed from the hits found in each wire plane of the detector.

The algorithm that performs this task, called Cluster3D, is a part of the traditional reconstruction pipeline. Cluster3D considers all possible combinations of 2D hits and selects the ones with a consistent geometry that plausibly could have originated from the same 3D deposition. At a basic level, this means that the set of hits must originate from wires that intersect and they must have consistent hit times. Each match is assigned a score that factors in the distance between the hits in the time domain and the extent of each hit in time.

This algorithm considers both 3-hit combinations (triplets) and 2-hit combinations (doublets) to reconstruct the 3D space points. Inherent to the tomographic reconstruction process are ambiguities in the matching, resulting in the possibility of incorrect combinations. Space points formed by triplets are less likely to be an incorrect combination, though the inclusion of doublets increases the efficiency of reconstructing true 3D charge depositions. For the purposes of the ML reconstruction

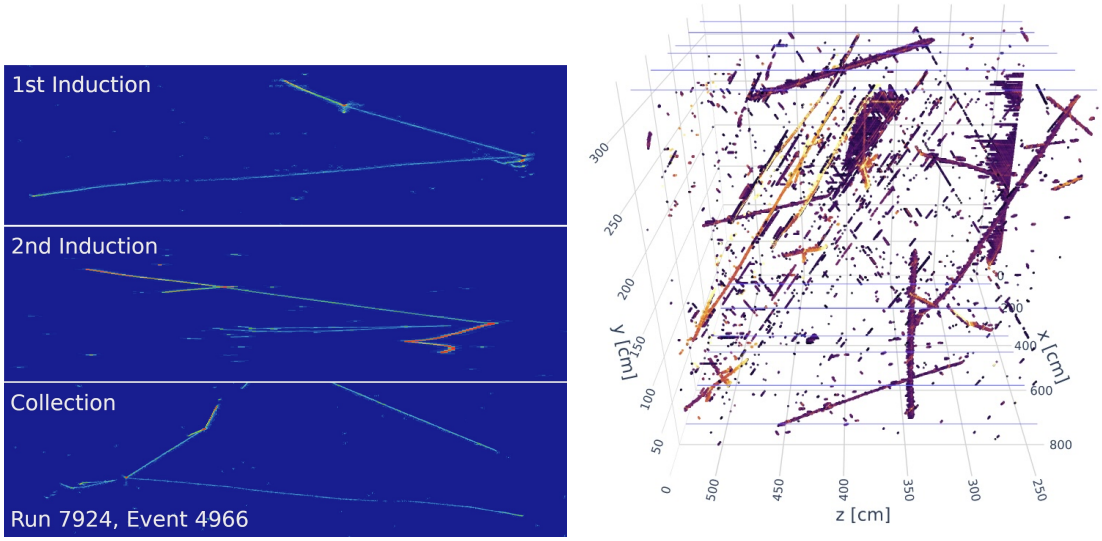


Figure 6.3: Example of a 2D event display of the same neutrino interaction in the ICARUS detector (left) and the corresponding 3D space points reconstructed by the Cluster3D algorithm (right). The image shown is a data event from Run 1.

chain described here, the Cluster3D algorithm has been tuned to be highly efficient at the cost of a higher number of incorrect combinations.

These incorrect combinations are known as tomographic artifacts, or "ghost points," and they are especially numerous for tracks parallel to the wire planes where the number of valid combinations of 2D hits is large. Figure 6.3 shows an example of a 2D event display of a neutrino interaction in the ICARUS detector and the corresponding 3D space points reconstructed by the Cluster3D algorithm. Of particular note is the presence of the ghost points in the 3D space points, which pose a challenge to visually identifying the true activity in the image. The first stage of the point classification task is to identify and remove these ghost points.

6.3.2 Point Classification

Ghost point removal As a first step, the reconstruction is tasked with the removal of the tomographic artifacts, or ghost points, from Cluster3D ("deghosting"). The neural network used for this implements the U-ResNet architecture, which was first popularized due to its success in biomedical imaging [92], to perform a binary classification of the space points, or semantic segmentation.

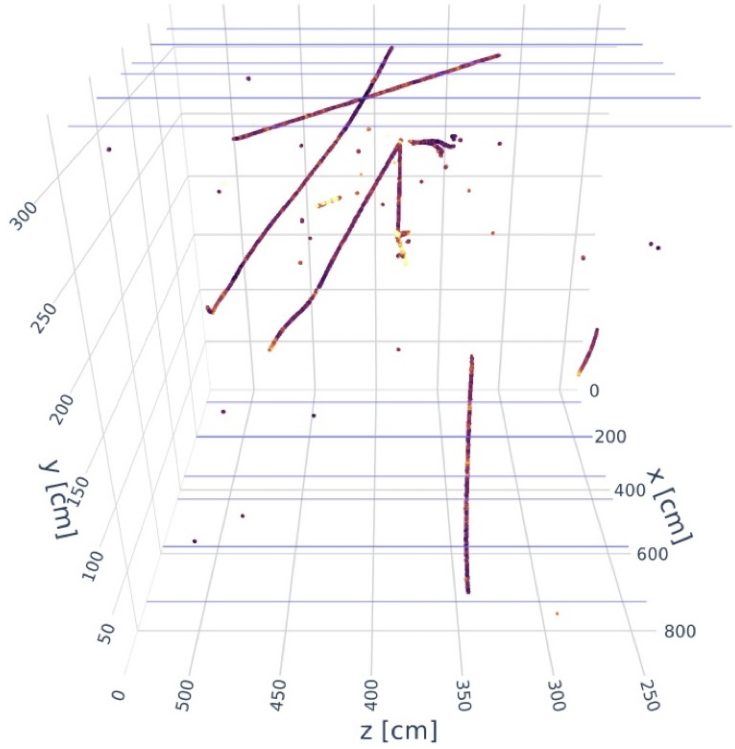


Figure 6.4: Example of the 3D space points after the deghosting step. The image shown is a data event from Run 1 and the same as in Figure 6.3.

Modifications were made to adapt U-ResNet to the sparse format of LArTPC data in contrast to regular images where each pixel contains information [93]. After deghosting, the reconstructed charge from Cluster3D is redistributed to enforce conservation of the overall charge. This is done by counting the number of times a given 2D hit is associated with a 3D space point and distributing the charge equally. The charge assigned to each space point is the average of the charge measured on each plane that contributed to the hit. An example image showing the 3D space points after the deghosting step is shown in Figure 6.4.

Semantic segmentation and point proposal network The next step in the point classification task is to classify the deghosted 3D space points into categories based on the type of parent activity that created them. The neural network used for this is a U-ResNet architecture, this time with five semantic classes of space points: Michel electron, delta ray, electromagnetic shower, track, and low-energy deposition. Each point is assigned a score for each class, and the class with the highest

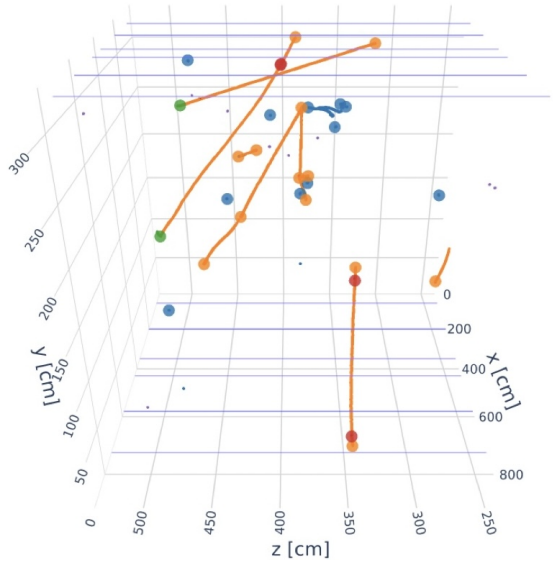


Figure 6.5: Example showing the 3D space points after the semantic segmentation and PPN stage. The space points are colored by the semantic type assigned by this stage (notably orange represents tracks, blue represents showers, green represents Michel electrons, and red represents delta rays). Also shown are the points of interest proposed by the PPN colored by the semantic class. The image shown is a data event from Run 1 and the same as in Figures 6.3 and 6.4.

score is taken as the prediction. The semantic type of the space points is used in later stages of the reconstruction.

The point proposal network (PPN) is a neural network that is tasked to propose the 3D space points that are the initial and terminal points of track-like objects and the initial points of shower-like objects [94]. The PPN shares the same encoding backbone as the semantic segmentation network as shown schematically in Figure 6.2. The points of interest proposed by this network are valuable for high-level physics analyses (e.g., particle start/end points and interaction vertex-finding) and are used in the clustering tasks in the later stages of the reconstruction. Figure 6.5 shows the 3D space points colored by semantic class and the points proposed by the PPN.

Point clustering The final step in the point classification task is to cluster the 3D space points in a set of points belonging to the same activity, i.e. track and shower fragments. This network, named Graph-SPICE, also uses a U-ResNet architecture and is tasked with learning an embedding of the space points in a higher-dimensional space where the points belonging to the same activity are

close together. In this embedded space, the points belonging to the same semantic class are used to build a k-nearest neighbors graph, which connects the points that are close together. Each edge is given a score based on the feature vectors of the two points it connects and the edge is activated if the score is above a threshold. Space points are clustered into fragments by finding the connected components of the graph. The output of this stage is a set of point clusters, each representing a track or shower fragment. A paper detailing the Graph-SPICE network is in preparation.

6.3.3 Formation of Particles and Interactions

Particle instance clustering At this point in the reconstruction, groups of space points corresponding to particle fragments have been identified. The next step is to cluster these groups into particles. The neural network used for this task is a GNN that is tasked with a binary (on/off) edge classification, where an activated edge indicates that the two instances belong to the same particle. Two distinct GNNs are used for particle instance clustering: GrapPA-Track and GrapPA-Shower for tracks and showers, respectively. It is worth noting that these two GNNs have identical architecture and only differ in the tasks for which they have been trained. The details of the architecture of these GNNs are described in [95].

The node features in each GNN are mostly geometric quantities such as the particle fragment direction vector and the most-likely PPN point. Edge features are also geometry-based and include the distance between the two fragments. The initial state of the graph has activated edges for all node pairs up to a maximal distance. This limitation is motivated by both the computational cost associated with fully connecting an image in a detector as large as ICARUS and by physics concerns: beyond a certain length scale there cannot be legitimate connections. The result of these two GNNs is a set of complete particles of both the track and shower categories. Small gaps in tracks, for example due to track-breaking at the cathode, are expected to be resolved in this stage. The shower-clustering step also identifies the primary shower fragment that is used to identify the shower start point, a critical feature for the interaction-clustering step.

Interaction clustering The final stage of the event reconstruction is to cluster the particles into interactions, defined as a group of particles originating from the same vertex. Particles that are directly connected to the primary vertex are considered primary particles, meaning they are immediate products of the interaction, whereas other particles are labeled as secondaries. The clustering of particles into interactions is easily cast as an edge-classification task, where an activated edge indicates two particles belong to the same interaction. In addition to the interaction clustering, the network is also tasked with predicting for each node the particle type (photon, electron, muon, pion, or proton) and the primary/secondary label. The neural network used for this task is a GNN and is described extensively in [95]. The architecture of this GNN differs from GrapPA-Track and GrapPA-Shower only in that it has more hidden features, which are useful for the extra tasks of particle identification (PID) and primary ID classification. The output of this stage is a full list of interactions in the event, each with a list of primary and secondary particles and their types.

6.4 Post-Processing

After the full chain of neural networks has been applied to the event, the output is a list of interactions with their associated particles. Each particle has been designated as a primary or secondary particle and has had a particle type assigned. In addition to this high-level set of information, a full physics analysis requires further information to be extracted from the event. This includes kinematic information such as the momentum and energy of the particles, as well as other geometric information such as the interaction vertex location. Association of interactions with the corresponding optical flash provides precise timing that can be leveraged in analyses. This information is extracted using traditional reconstruction algorithms that are well-established in the field of neutrino physics.

6.4.1 Kinematic Reconstruction

The kinematic reconstruction of the particles in the event is performed using traditional reconstruction algorithms. Of relevance for the analysis presented in this thesis are the following reconstructed quantities:

- **Particle direction reconstruction**: The direction of the particle at the start (tracks and showers) and end (tracks only) points is reconstructed using a principal component analysis (PCA) of the space points in the local neighborhood of the point. When combined with the kinetic energy of the particle, the momentum (direction and magnitude) can be reconstructed.

- **Calorimetric energy reconstruction**: The energy of the particle is reconstructed using the total charge deposited by the particle in the detector. The charge is converted to energy using an effective formula that accounts for the average amount of electron-ion recombination, the TPC electronics gain, the work function of liquid argon, the attenuation due to electron lifetime, and a scale factor that accounts for the missed shower fragments that have fallen below the detection threshold. This is most useful for showers, for which other methods of energy reconstruction are not available.

- **Track length reconstruction**: The length of the track is reconstructed by segmenting the track and summing the length of each segment along their respective principal axes.

- **Range-based energy reconstruction**: The energy of the track is estimated using the well-known relationship between the kinetic energy of a particle and its range in a material according to the Bethe-Bloch formula. This relation is known as the Continuous Slowing Down Approximation (CSDA) and it is used to estimate the energy of the particle. This method assumes that the particle ranges out in the argon without nuclear interactions and that the track is fully contained in the detector.

- **Vertex reconstruction**: The interaction vertex is reconstructed as the point joining the start points of the primary particles in the interaction.

- **Fiducial volume**: The interaction vertex is used to determine if the interaction occurred within the fiducial volume of the detector. This is defined as the region 25 cm from the edges of the detector in the x and y directions, 30 cm from the upstream edge in the z direction, and 50 cm from the downstream edge in the z direction. These numbers were assumed in the SBN proposal [14] and are chosen to avoid reconstruction inefficiencies near the edges of the detector (e.g. exiting particles that are not seen).

- **Containment**: The containment of the particles in the detector is determined by requiring that all the space points of the particle are at least 5 cm from the edges of the detector. At the interaction level, the interaction is considered contained if all of the particles that comprise it are contained. The containment cut also applies the constraint that the space points must be in the same TPC that created them; out-of-time tracks may be reconstructed partially in the wrong TPC and therefore be removed by this cut. The value of 5 cm was chosen based on the expected magnitude of space charge effects in the detector, which tends to displace space points away from the detector boundaries.

6.4.2 Optical Flash Association

The optical flash is a prompt signal in the PMTs associated with the scintillation light produced by the neutrino interaction in the detector. As part of the standard reconstruction chain described in Section 5.2, the optical flash is formed from optical hits, which themselves are reconstructed from the PMT waveforms. This flash object can be thought of as a set of optical hits each with a well-defined position, time, and photoelectron count. The job of the flash association algorithm is to leverage these characteristics and match them to the interactions in the event that are imaged by the TPC through detection of ionization charge signals.

The tool used for flash association is a package originally developed for MicroBooNE called OpT0Finder. OpT0Finder employs a likelihood-based approach to match the optical flash to charge activity associated with an interaction in the event. For each interaction, as defined by the upstream ML reconstruction, the algorithm calculates an allowed range of x offsets that are allowed by the

geometry. This reflects our inability to know beforehand the exact time of the interaction. For each of these allowable positions[1], the algorithm calculates a hypothesis flash based on the expected light yield of the interaction, the distribution of the ionization charge in the interaction, and the position of the PMTs in the detector. The hypothesis flash and the observed flash each have a likelihood computed using a joint probability distribution of the photoelectron count in each PMT given the x position of the flash. Maximizing the likelihood between the two allows the assignment of a score to each potential interaction-flash pair.

A matching algorithm then selects the best pairings based on the scores while simultaneously skipping pairs with an already matched flash or interaction. The result of this association process is the assignment of a time stamp to each interaction which was successfully matched to a flash. This time stamp can be used to precisely tag an interaction as being in-time or out-of-time with respect to the beam spill, and therefore can be leveraged as a tool for cosmic rejection.

[1]This is not actually a brute-force algorithm, but rather an optimization algorithm that profiles across the drift direction.

Chapter 7

Muon Neutrino Interaction Selection and Performance

This chapter provides an overview of the chosen signal definitions for the analysis presented in this thesis. The signal definitions are chosen to reflect the goals of the ICARUS collaboration and the SBN Program, and are designed to be simple and robust to the imperfections in the reconstruction and simulation. An overview of the signal channels and the selections are given in Section 7.1. The chapter also presents the variables used to characterize the performance of the selections (Section 7.2) as well as the performance of the selections on simulation (Section 7.3).

7.1 Neutrino Signal Definitions and Selections

The signal definitions chosen for the analysis presented in this thesis reflect both the goals of the ICARUS collaboration and the SBN Program more broadly. One of the near-term goals of the ICARUS collaboration involves a single-detector search for an eV-scale sterile neutrino, both as a contribution to the world knowledge of neutrino physics and as a demonstration of the capabilities of the ICARUS detector. Achieving this early physics result, before analysis tools have been fully optimized, has motivated the choice of relatively simple final states to minimize impacts to the analysis sensitivity from imperfect event reconstruction and detector modeling.

The joint oscillation analysis of the SBN Program, as outlined in the SBN proposal [14] and discussed in Chapter 3, will involve an inclusive ν_μ CC channel that encompasses a wide range of neutrino interaction topologies. This channel is therefore a key component of the SBN Program's physics reach, and the ICARUS collaboration has a strong interest in understanding the performance of the detector in terms of this signal selection.

7.1.1 Signal Channels

Three signal definitions are chosen to accomplish the goals described above: a ν_μ CC $1\mu1p$ channel, a ν_μ CC $1\mu\text{N}p$ channel, and a ν_μ CC inclusive channel. In common to all three of these signal channels are a few key features:

- **Containment**: The entirety of the interaction is required to be contained within a volume that is defined as 5 cm from the edges of the detector in all three dimensions. This requirement ensures that the range-based energy reconstruction, which has significantly better resolution than other methods, can be applied to tracks in the final state.

- **Fiducial Volume**: The interaction vertex must be located within the fiducial volume of the detector, which is defined as the region 25 cm from the edges of the detector in the x and y directions, 30 cm from the upstream edge in the z direction, and 50 cm from the downstream edge in the z direction. This requirement is motivated by the desire to avoid the reconstruction inefficiencies and biases that can arise near the edges of the detector.

- **Single Muon**: Each signal definition requires the presence of a single primary muon in the final state. This muon is required to have a length of at least 50 cm, principally to reduce the contamination from NC interactions with a pion in the final state. This length threshold is also consistent with the proposed selection in the SBN proposal.

- **Proton Threshold**: Protons are required to be primary particles and to have deposited at least 50 MeV of energy to count towards the final state of the interaction. This threshold is chosen to ensure that the proton is sufficiently long to be reconstructed as a track and to ensure that the proton is energetic enough to be well-separated from the muon in the final state. Interactions with one proton above threshold and one proton below threshold are classified as $1\mu1p$ final states under this definition.

- **Other Particles**: The final state may contain additional protons, pions, and showers. These particles are required to be primary particles and to have deposited at least 25 MeV of energy

to be counted towards the final state of the interaction. This threshold is chosen to ensure that the particles are sufficiently energetic to be visible in the final state.

The $1\mu1p$ channel requires exactly one proton above threshold in addition to the muon, while the $1\mu\text{N}p$ channel requires at least one proton above threshold in addition to the muon. The ν_μ CC inclusive channel requires only the presence of a single muon in the final state and any number of other final state particles above threshold. The structuring of these signal definitions means that the $1\mu1p$ channel is a subset of the $1\mu\text{N}p$ channel, which is in turn a subset of the ν_μ CC inclusive channel.

While the $1\mu1p$ and $1\mu\text{N}p$ channels are inherently more statistically limited than the ν_μ CC inclusive channel, they provide simple final states that are well-suited to the early stages of the ICARUS physics program. The energy reconstruction for these two channels is also more precise due to the simple final state, which may be beneficial for neutrino oscillation physics measurements. Additionally, these final states are particularly low in cosmic background, especially in combination with the containment requirement. In contrast, the ν_μ CC inclusive channel is more statistically powerful and provides a more complete picture of the detector performance due to the wider range of neutrino interaction topologies. Final states in this channel may also contain showers, the precise reconstruction of which is of significant importance for any future oscillation analysis.

7.1.2 Additional Selections

In addition to the requirements described above, the following selections are applied to the reconstructed interactions to further reduce the background and improve the purity of the selected samples:

Flash Time: Each interaction is required to be matched with an optical flash detected by the PMTs that is consistent with the time of the neutrino beam spill. The association of the interaction with the flash is performed with OpT0Finder, as described in Section 6.4. This analysis

targets neutrinos from the BNB, which has a spill length of 1.6 us. The requirement that interactions be associated with a flash in-time with the beam provides significant background rejection of cosmic-induced interactions that are not in-time with the beam. This cut is not expected to provide significant background rejection for cosmic-induced interactions that are in-time with the beam and trigger the detector readout.

CRT Veto: Each event has a collection of PMT flashes and CRT hits. As part of the standard reconstruction chain, a matching algorithm is run to associate the CRT hits with the PMT flashes using timing information. This analysis makes use of this association by requiring that selected events have at least one flash within the beam spill window that is not matched to a CRT hit. This requirement strongly suppresses cosmic-induced interactions, and is complementary to the flash time cut discussed above in that it rejects many of the cosmic-induced interactions that are in-time with the beam. This veto uses only the Top and Side CRT subsystems, as the Bottom CRT subsystem was not fully commissioned prior to the data used in this analysis.

7.1.3 Neutrino Sample Composition

The primary sample used in this analysis to set the central value expectation for the selections is a sample of simulated neutrino interactions produced according to the predicted BNB flux and overlayed with the expected out-of-time cosmic flux (see Section 5.1.1 for more details). The size of this sample is 2.68×10^{20} POT, which is more than a factor of ten larger than the recorded dataset of 1.92×10^{19} POT for this analysis. The statistical uncertainty of this larger simulated sample is negligible for the analysis presented in this thesis. It is worth noting that the neutrinos in this sample are required to interact in the active volume of the detector. The contribution to the expected event rates from neutrinos that interact outside of the active volume is expected to be negligible and thus is not included in this analysis.

Most of the triggered events in data are due to cosmic muons in-time with the beam, which are not present in the simulated neutrino sample. Rather than attempting to model the cosmic background with a simulated dataset, the cosmic background is estimated using the off-beam dataset

(see Section 4.5). This method is expected to provide a more accurate estimate of the cosmic background than a simulated dataset, which would be subject to uncertainties in the cosmic flux and the detector response to cosmic-induced interactions. The scaling of the off-beam dataset to the on-beam dataset is based on matching the number of opened trigger gates in the two datasets. Though the number of opened trigger gates is one-to-one, and therefore the number of on-beam and off-beam events is near equal, there is a limited statistical penalty due to both the smallness of the background and the ability to use the entire off-beam dataset rather than the 10% that is used for the on-beam dataset due to blinding considerations. The on-beam dataset used in this analysis will be described in more detail in Chapter 9 and the uncertainty associated with the data-driven cosmic background estimate will be assessed in Chapter 8.

7.2 Variables of Interest

This section presents an overview of the variables used to characterize the performance of the three selections described in the preceding section. The variables are divided into three categories: interaction kinematic variables, kinematic imbalance variables, and PID variables. The distributions of these variables can be used to validate the performance of the selections and to identify potential sources of bias or inefficiency in the reconstruction and simulation. In each of the following plots, the distributions are made with truth information to show the true distributions of these variables for each of the signal channels prior to any reconstruction or selection effects. The distributions are also scaled to the recorded dataset of 1.92×10^{19} POT to show the predicted event rates.

7.2.1 Interaction Kinematic Variables

These variables reflect the kinematics of the neutrino interaction and the final state particles. The total energy of the interaction is a key variable for understanding the energy of the parent neutrino, while the kinematics of the final state particles can provide information about the nature of the interaction and the performance of the reconstruction. The total energy of the interaction

is a necessary input to oscillation analyses. Additionally, these variables are often used in cross section measurements to probe the underlying physics of neutrino interactions. For the chosen signal definitions, there is always a single primary muon and there may be additional primary protons, pions, and showers above the energy threshold of the signal definition.

Muon Kinetic Energy ($\mathbf{T}_\mu$): The kinetic energy of the muon as calculated using the Continuous Slowing Down Approximation (CSDA) [96] to relate the muon track length (range) to its kinetic energy. This method is typically known as "range-based" kinetic energy reconstruction.

Leading Proton Kinetic Energy ($\mathbf{T_p}$): The kinetic energy of the most energetic primary proton as calculated using the CSDA to relate the proton track length (range) to its energy.

Interaction Visible Energy ($\mathbf{E_{vis}}$): The best estimate of energy deposited in the detector by the neutrino interaction. The energy deposited by primary showers are estimated by summing the deposited charge and converting it to energy using an effective constant that accounts for electron-ion recombination and shower clustering efficiency. Primary muon, pion, and proton energies are estimated using the CSDA to relate the track length to the energy deposited. In addition to the kinetic energies of the particles, the mass of the particle is also included in the energy calculation for the muon and pion, which are created rather than ejected during the interaction. Neutrons leading to recoiling protons are not currently included in the energy calculation as they are classified as secondaries and only primary particles are included for now. The interaction visible energy is therefore calculated as:

$$E_{\text{vis}} = \sum_{\text{Showers}} E^{calo}_{\text{shower}} + \sum_{\text{Muons}} E^{range}_{\text{muon}} + \sum_{\text{Pions}} E^{range}_{\text{pion}} + \sum_{\text{Protons}} T^{range}_{\text{proton}} \tag{7.1}$$

Muon Transverse Momentum ($\mathbf{p^\mu_T}$): The component of muon momentum perpendicular to the neutrino beam direction. This should generally be balanced by the transverse momentum of the

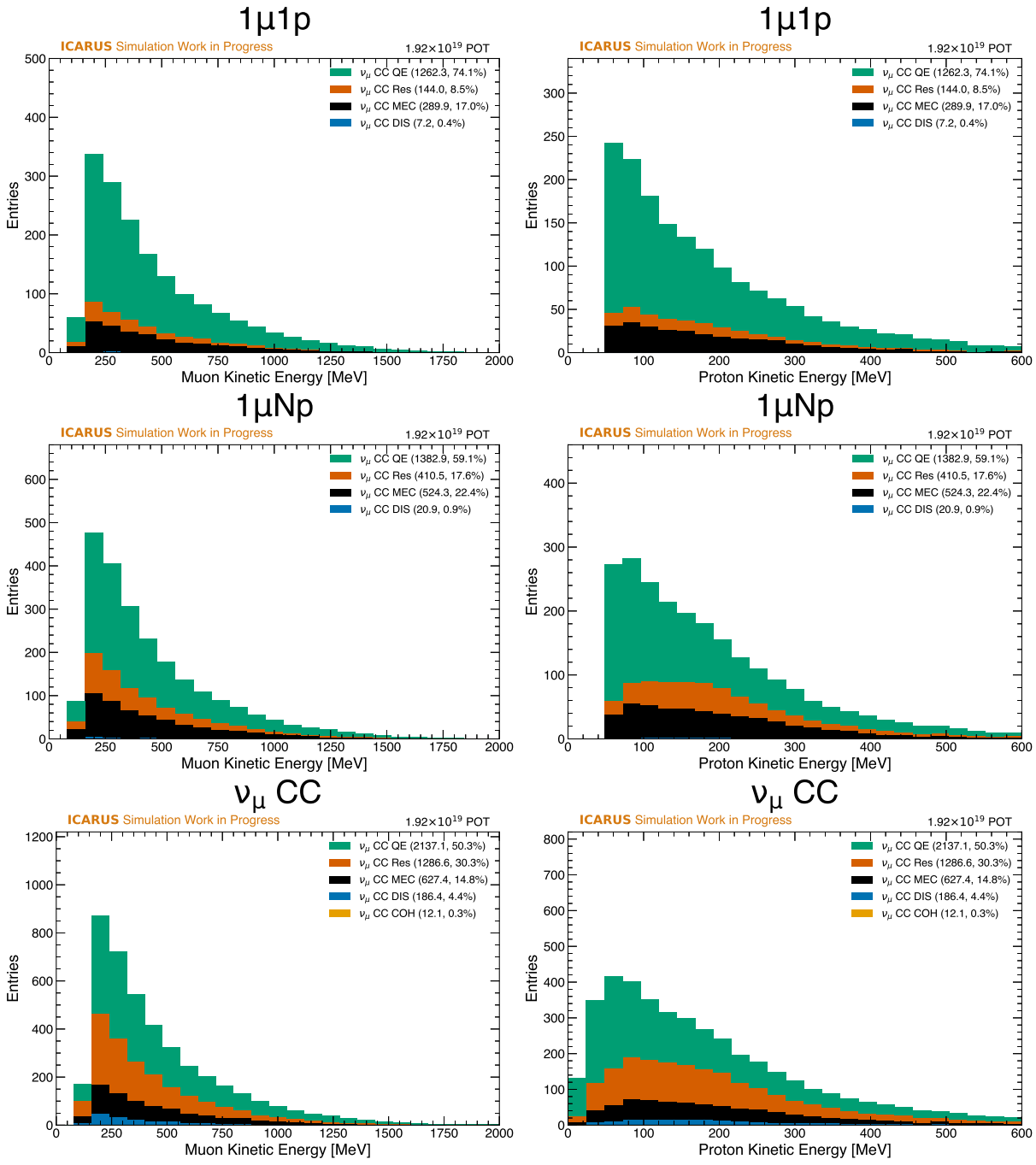


Figure 7.1: Kinetic energy of the muon (left) and the most energetic proton (right) for each of the three signal channels: from top to bottom $1\mu 1\mathrm{p}$, $1\mu \mathrm{Np}$, and ν_μ CC inclusive.

hadronic system in the absence of final state interactions, as described in more detail below in the text on kinematic imbalance variables.

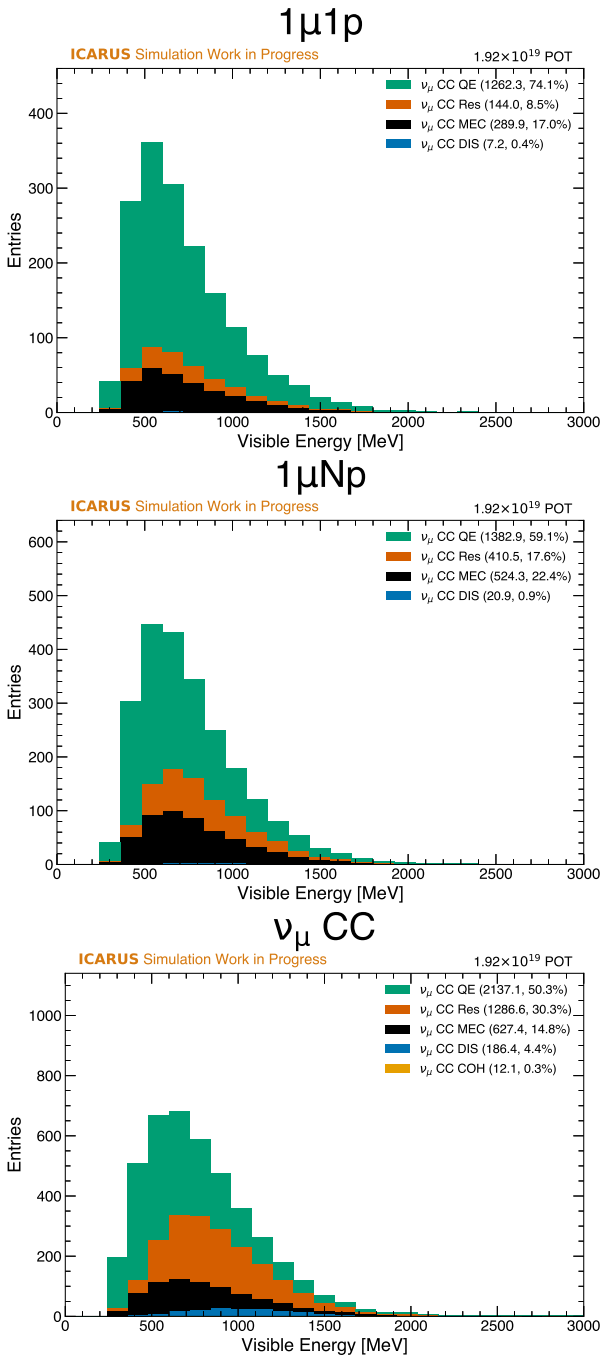


Figure 7.2: Total visible energy of the interaction for each of the signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive.

Leading Proton Transverse Momentum ($\mathrm{p}_\mathrm{T}^\mathrm{p}$): The component of the momentum of the most energetic primary proton perpendicular to the neutrino beam direction.

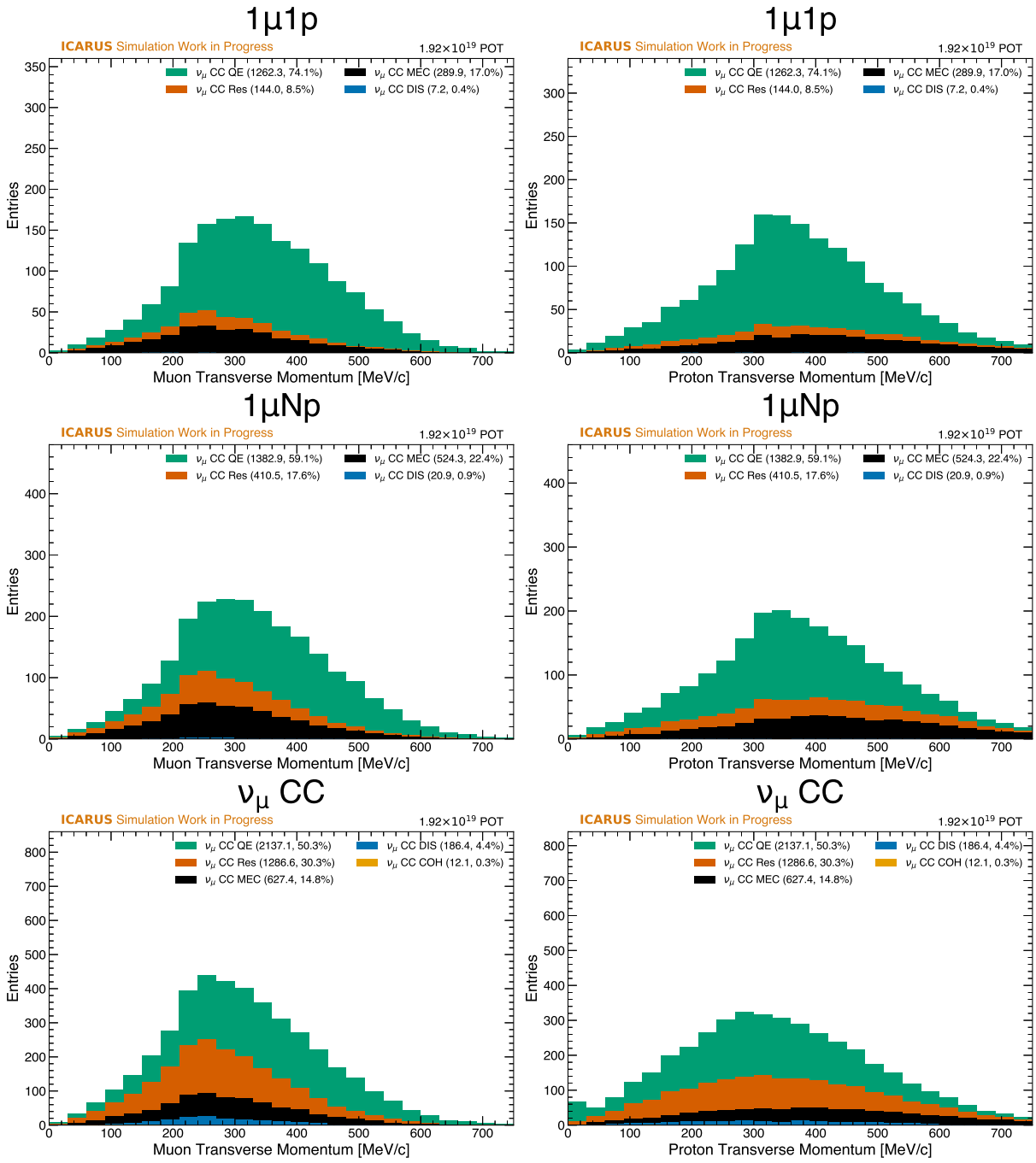


Figure 7.3: Transverse momentum of the muon (left) and the most energetic proton (right) for each of the three signal channels: from top to bottom $1\mu1$p, 1μNp, and ν_μ CC inclusive.

Muon Polar Angle (θ_μ): The polar angle of the muon calculated with respect to the neutrino beam direction (z-axis):

$$\theta_\mu = \arccos\left(\frac{p_z^\mu}{p^\mu}\right) \tag{7.2}$$

Muon Azimuthal Angle (ϕ_μ): The azimuthal angle of the muon calculated with respect to the neutrino beam direction (z-axis):

$$\phi_\mu = \arccos\left(\frac{p_x^\mu}{p_T^\mu}\right) \tag{7.3}$$

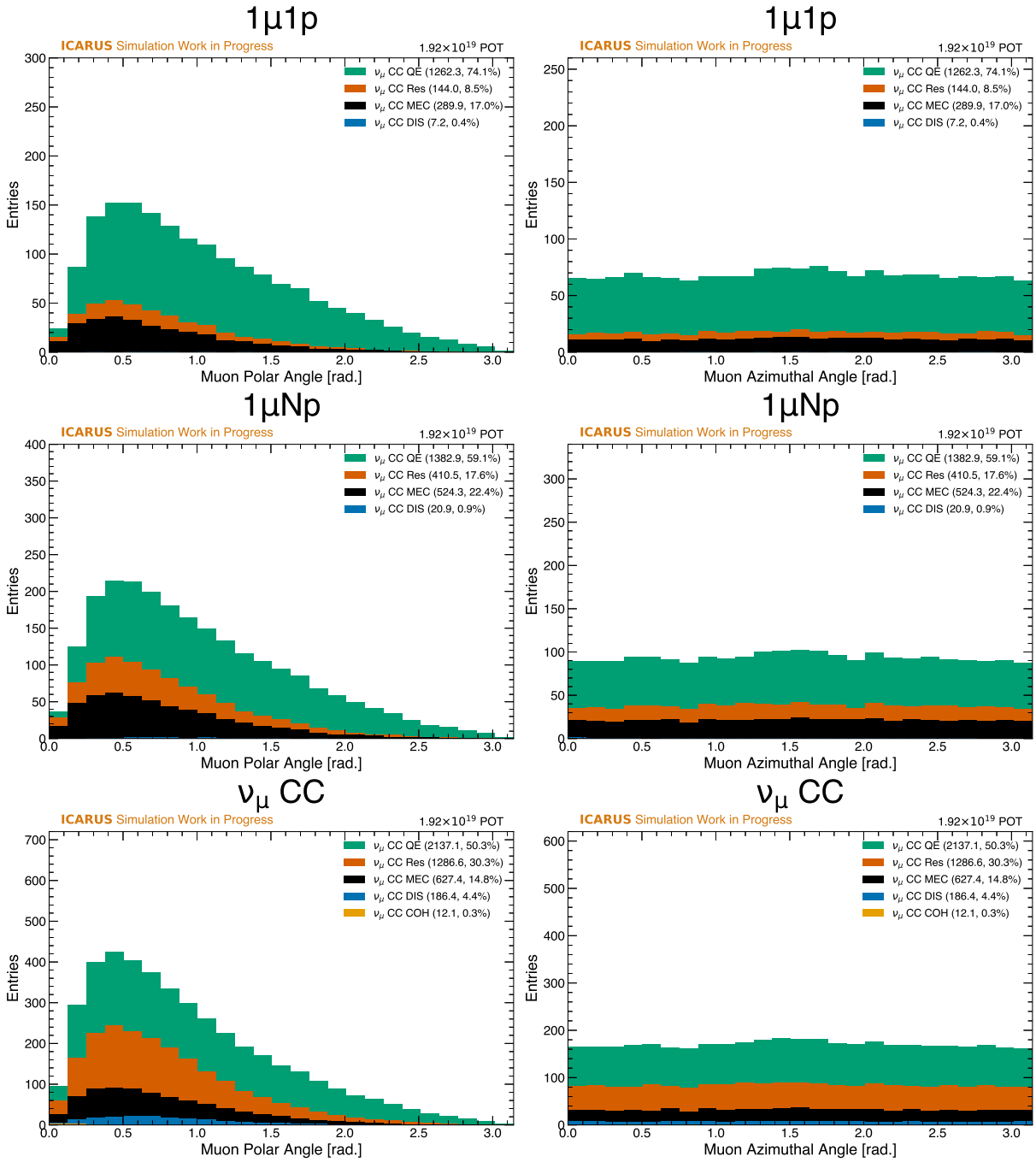


Figure 7.4: Polar angle (left) and azimuthal angle (right) of the muon for each of the three signal channels: from top to bottom $1\mu1\text{p}$, $1\mu\text{Np}$, and ν_μ CC inclusive.

Opening Angle ($\theta_{\mu\mathbf{p}}$): The angle between the muon and the leading proton.

$$\theta_{\mu p} = \arccos\left(\frac{\vec{p}^{\,\mu}\cdot\vec{p}^{\,p}}{|\vec{p}^{\,\mu}||\vec{p}^{\,p}|}\right) \tag{7.4}$$

7.2.2 Kinematic Imbalance Variables

The axis of the beam defines the direction along which the momentum of the interacting neutrino is oriented. The transverse plane is defined as the plane perpendicular to the neutrino beam direction, while the longitudinal direction is defined as oriented along the neutrino beam direction. From conservation of momentum, the momentum in the transverse plane should sum to zero in the absence of initial state nucleon transverse momentum and final state interactions (FSI). Fermi motion inside the nucleus produces non-zero transverse momentum even for quasi-elastic (QE) interactions. More complex interactions such as resonant pion production and deep inelastic scattering can produce final states similar to QE interactions, but with additional transverse momentum that populates the region above the Fermi momentum. These variables are natural choices for probing the effects of FSI and for measuring neutrino interaction cross sections.

In the ICARUS geometry, the beam axis is defined as the z-axis, and the transverse plane is correspondingly the x-y plane. For the signal definitions considered in this analysis, there exists a single muon and a combined hadronic system, respectively having momentum $\vec{p}^{\,\mu}$ and $\vec{p}^{\,h}$ and transverse components $\vec{p}^{\,\mu}_{\,T}$ and $\vec{p}^{\,h}_{\,T}$. The transverse component of the momentum transfer to the nucleus, $\vec{q}_T$, is defined as equal and opposite to $\vec{p}^{\,\mu}_{\,T}$. Schematically, this is shown in Figure 7.6 within the transverse plane. This formalism, presented in detail in [97], leads to the introduction of several variables:

$\delta\mathbf{p_T}$: The magnitude of the vector difference between the transverse momentum of the muon and the hadronic system. This variable quantifies the amount of transverse momentum that is not accounted for by the reconstructed particles. Large values of this quantity exceeding the Fermi

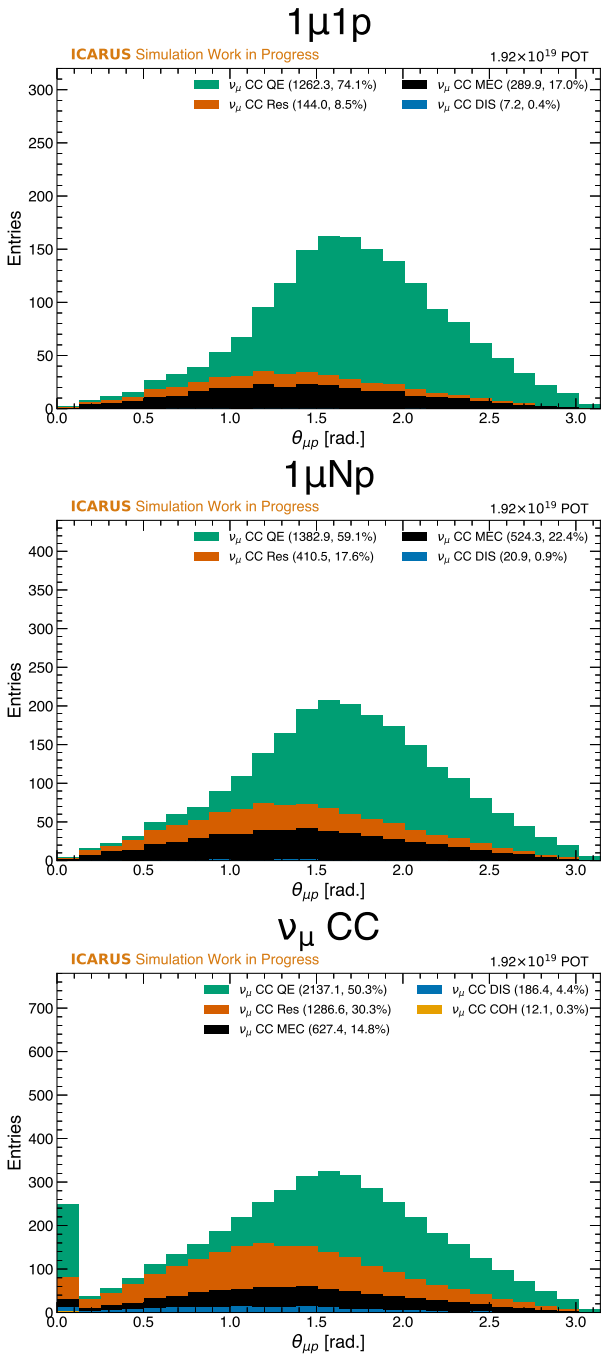


Figure 7.5: The muon-proton opening angle for each of the three signal channels: from top to bottom $1\mu 1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive. The peak at the lowest bin in the inclusive channel is from interactions where no proton is present.

momentum are indicative of FSI.

$$\delta p_T = \left| \vec{p}_T^{\,\mu} - \vec{p}_T^{\,h} \right| \tag{7.5}$$

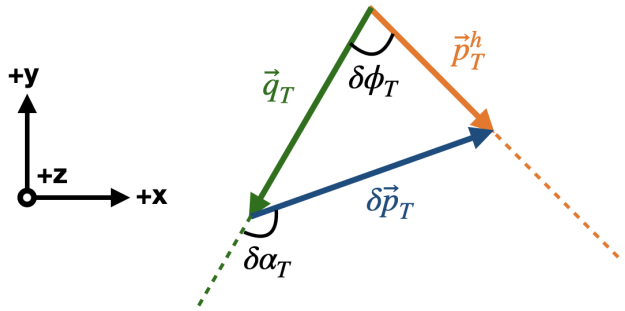


Figure 7.6: Diagram showing the each of the chosen kinematic imbalance variables in the transverse plane.

$\delta\phi_\mathbf{T}$**:** The angle between the the transverse momentum transfer vector and the transverse momentum of the hadronic system. In the case of a free and stationary nucleon target, this angle would be zero. Initial state motion leads to small values of this quantity, while FSI can lead to larger values.

$$\delta\phi_T = \arccos\left(\frac{\vec{p}_T^{\,h}\cdot\vec{q}_T}{|\vec{p}_T^{\,h}||\vec{q}_T|}\right) \tag{7.6}$$

$\delta\alpha_\mathbf{T}$**:** The angle between the transverse momentum transfer vector and transverse missing momentum vector. This angle is less sensitive to the initial state motion of the nucleons, but is sensitive to FSI. In the absence of FSI, this quantity does not have a preferred orientation.

$$\delta\alpha_T = \arccos\left(\frac{\vec{q}_T\cdot\delta\vec{p}_T}{|\vec{q}_T||\delta\vec{p}_T|}\right) \tag{7.7}$$

7.2.3 PID Variables

Each particle is assigned scores by a GNN, as described in Section 6.3, that provide an indication of how likely the particle is to be a muon, proton, or pion for tracks, or an electron or photon for showers. The PID scores are used to assign a particle type to each reconstructed particle. A difference in the distribution of these scores between data and simulation may indicate a bias in the simulation with respect to data, so it is important to validate that these distributions are in agreement with data.

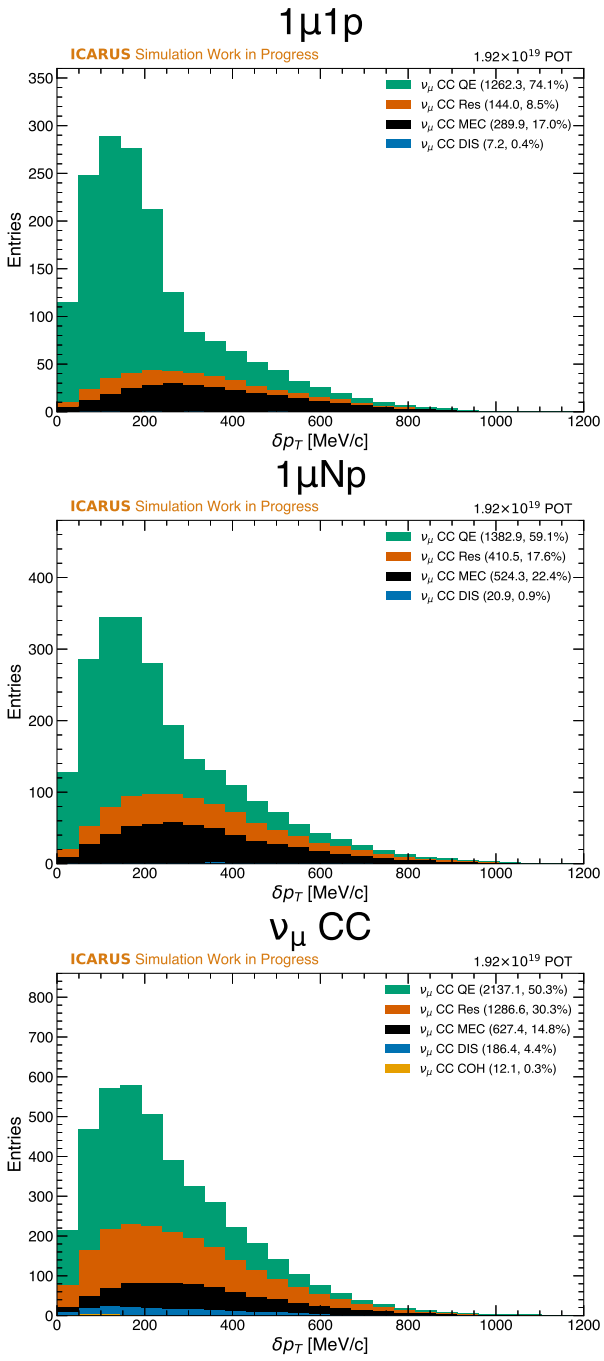


Figure 7.7: Transverse momentum of the interaction for each of the three signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive.

Muon "Softmax" PID Score: The muon score assigned to the primary muon of the interaction by the reconstruction. A high score near 1 indicates that the particle is likely to be a muon, while a low score near 0 indicates that the particle is not likely to be a muon.

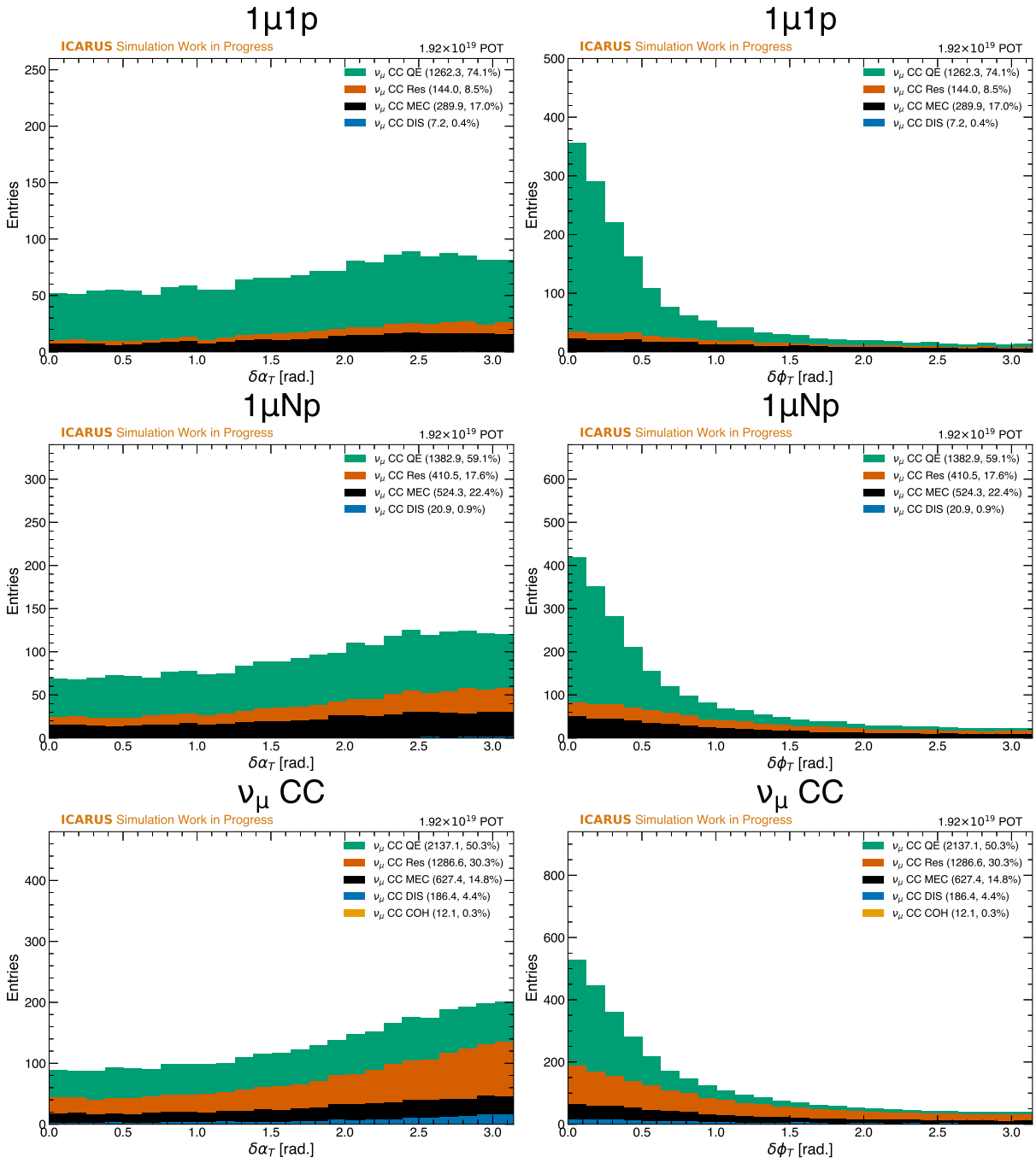


Figure 7.8: Kinematic imbalance variables $\delta\alpha_T$ (left) and $\delta\phi_T$ (right) for each of the three signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive.

Proton "Softmax" PID Score: The proton score assigned to the leading proton of the interaction by the reconstruction. A high score near 1 indicates that the particle is likely to be a proton, while a low score near 0 indicates that the particle is not likely to be a proton. This variable is significantly

more peaked than the muon score, presumably because the muon and pion are more similar in terms of their energy deposition and track length, so the range of the plot has been restricted.

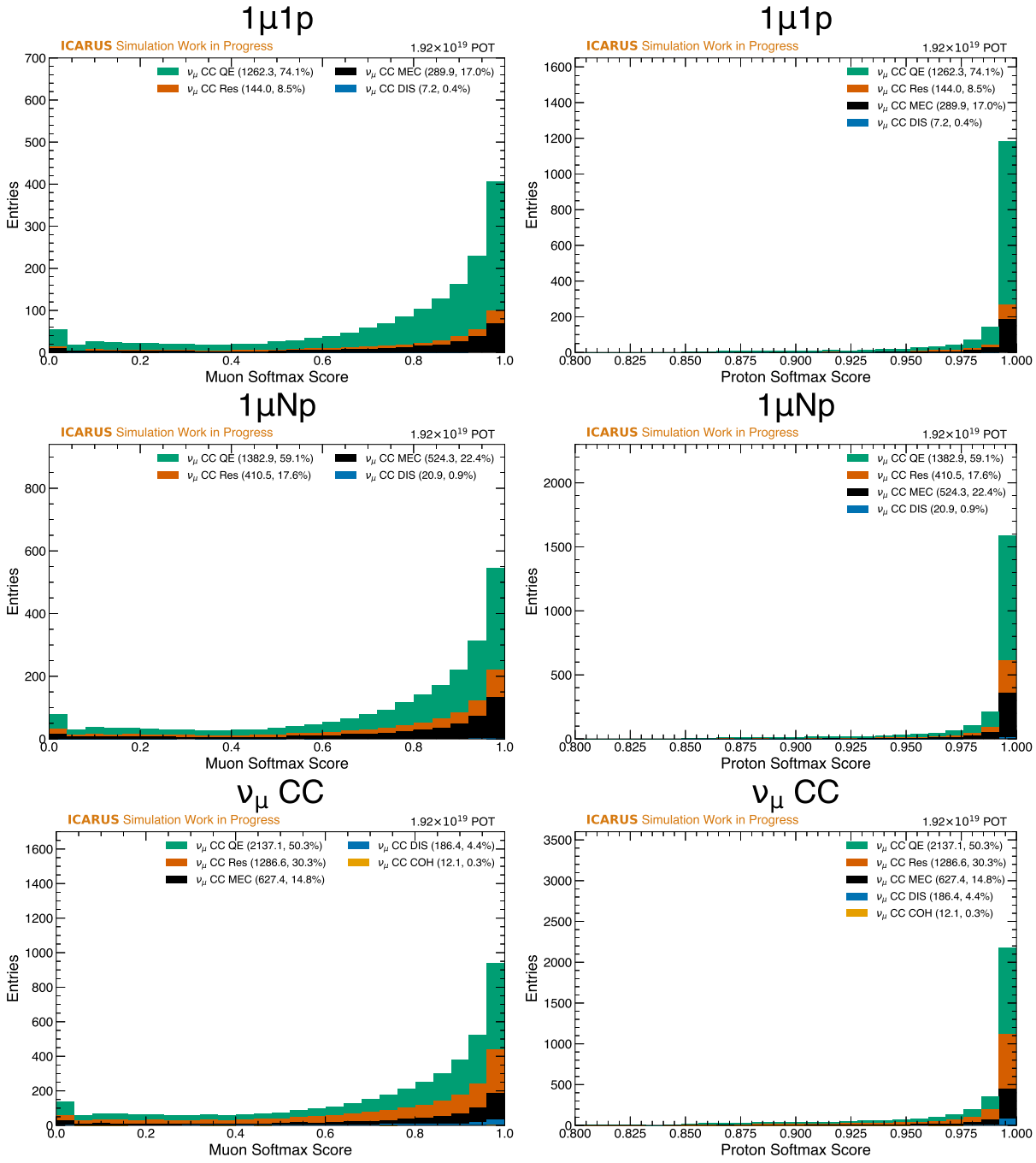


Figure 7.9: Softmax PID scores for the muon candidate (left) and the leading proton candidate (right) for each of the three signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive.

Table 7.1: Summary of the purity and efficiency of the selections for each of the signal channels.

Selection Cut	$1\mu1p$ Purity [%]	$1\mu1p$ Efficiency [%]	$1\mu Np$ Purity [%]	$1\mu Np$ Efficiency [%]	ν_μ CC Purity [%]	ν_μ CC Efficiency [%]
No Cut	0.0	99.9	0.1	100.0	0.1	100.0
Fiducial Volume	0.1	98.8	0.1	98.8	0.3	98.2
Containment	1.1	94.9	1.5	95.0	3.5	94.1
Final State	66.2	73.9	71.2	77.9	9.5	86.3
Flash Time	80.1	72.4	83.0	76.4	87.8	84.5
CRT Veto	80.3	71.3	83.3	75.4	90.4	83.3

7.3 Selection Performance

The performance of a selection can be characterized by its purity and efficiency. The purity of a selection is defined as the fraction of selected events that are true signal events, while the efficiency is defined as the fraction of true signal events that are selected. It is worth noting that the definition of efficiency used in this thesis is relative to the signal definition and not to the total number of neutrino interactions as may be used in other contexts.

A table summarizing the purity and efficiency of the selections for each of the signal channels is shown in Table 7.1. The purity and efficiency are calculated using the definitions described above using a Bayesian method [98] that produces more robust error bars in cases where the purity or efficiency fraction is near zero or one. Each row denotes the successive application of the selection cuts, providing a clear way to understand the impact of each cut on the purity and efficiency of the selection. From this table, several conclusions can be drawn:

- The application of the containment cut provides a significant improvement in purity for all signal channels. This is most pronounced for the ν_μ CC inclusive channel, where cosmic backgrounds are larger. This cut also has a non-negligible impact on the efficiency of the selection. This is largely due to cases where an un-contained cosmic-induced interaction is clustered into the neutrino interaction.

- The application of the cut on the final state of the candidate interactions has a significant impact on the purity of the selection for all three channels, though the impact is not as pronounced for the ν_μ CC inclusive channel due to the presence of larger cosmic backgrounds. This is also where we see the biggest impact on the efficiency of the selection - an effect that

can be directly attributed to imperfections in the reconstruction. This efficiency loss is larger for the two exclusive channels, where the final state is more constrained.

- The flash time cut has a negligible impact on the efficiency of the selection for all three channels, and only provides an additional improvement in purity for the ν_μ CC inclusive channel. This is consistent with the expectation that the flash time cut primarily targets cosmic-induced interactions that are not in-time with the beam. These backgrounds are largest for the ν_μ CC inclusive channel, which is why the impact of this cut is most pronounced for this channel.

- The CRT veto cut also has a $\sim$1% impact on the efficiency of the selection for all three channels. Purity is not significantly improved except for the ν_μ CC inclusive channel, where remaining in-time cosmic backgrounds are largest. The cost to the efficiency, though small, may not be worth it for the two exclusive channels. Since this analysis is focused on demonstrating the performance of all three selections, it is included in the final selection for all three channels to maintain consistency. It is worth highlighting that this cut may be significantly more important if the signal was extended to include interactions that exit the detector.

The purity and efficiency of a selection can also be visualized as a function of the variables of interest to understand the relationship between the selection and each variable. This can help to identify regions of phase space where the selection is particularly efficient or pure, and to identify regions where the selection may be underperforming. In each of the following plots, the purity and efficiency are calculated per-bin using the same binning as the variable itself. The selected candidates are binned according to the *reconstructed* variable, as opposed to the true quantity presented previously, and broken down into categories based on their true final state. These plots are shown in Figures 7.10 through 7.16. There are several notable features worth highlighting:

- Generally, the purity and efficiency drop off at regions of phase space correlated with lower energy. Smaller particles are more likely to be mis-reconstructed and may not be as separable from other particles in the final state.

- Angles where tracks are parallel to the wire planes are more likely to be mis-reconstructed. This is especially visible in the muon and proton transverse momentum variables, where the lower values are associated with a particle that is more aligned with the z-axis. This is a known side effect of the coherent noise filtering algorithm used upstream in the reconstruction chain, which tends to remove signal from tracks that are parallel to the wire planes.

- There is a drop in purity and efficiency for the two exclusive channels at higher values of the muon-proton opening angle. This is due to the difficulty in separating the muon and proton tracks when they are highly collinear.

- The values of δp_T outside of the Fermi momentum are associated with a drop in purity and efficiency for the two exclusive channels. This is a region of phase space with fewer QE interactions and more complex interactions that are more difficult to reconstruct.

At this stage of the analysis, the selections are performing well in terms of purity and efficiency on simulation. Benchmarking the performance of the selections on data is a key step in the validation of the selections and the reconstruction chain, but prior to this step the full set of uncertainties must be understood. This will be the focus of the next chapter.

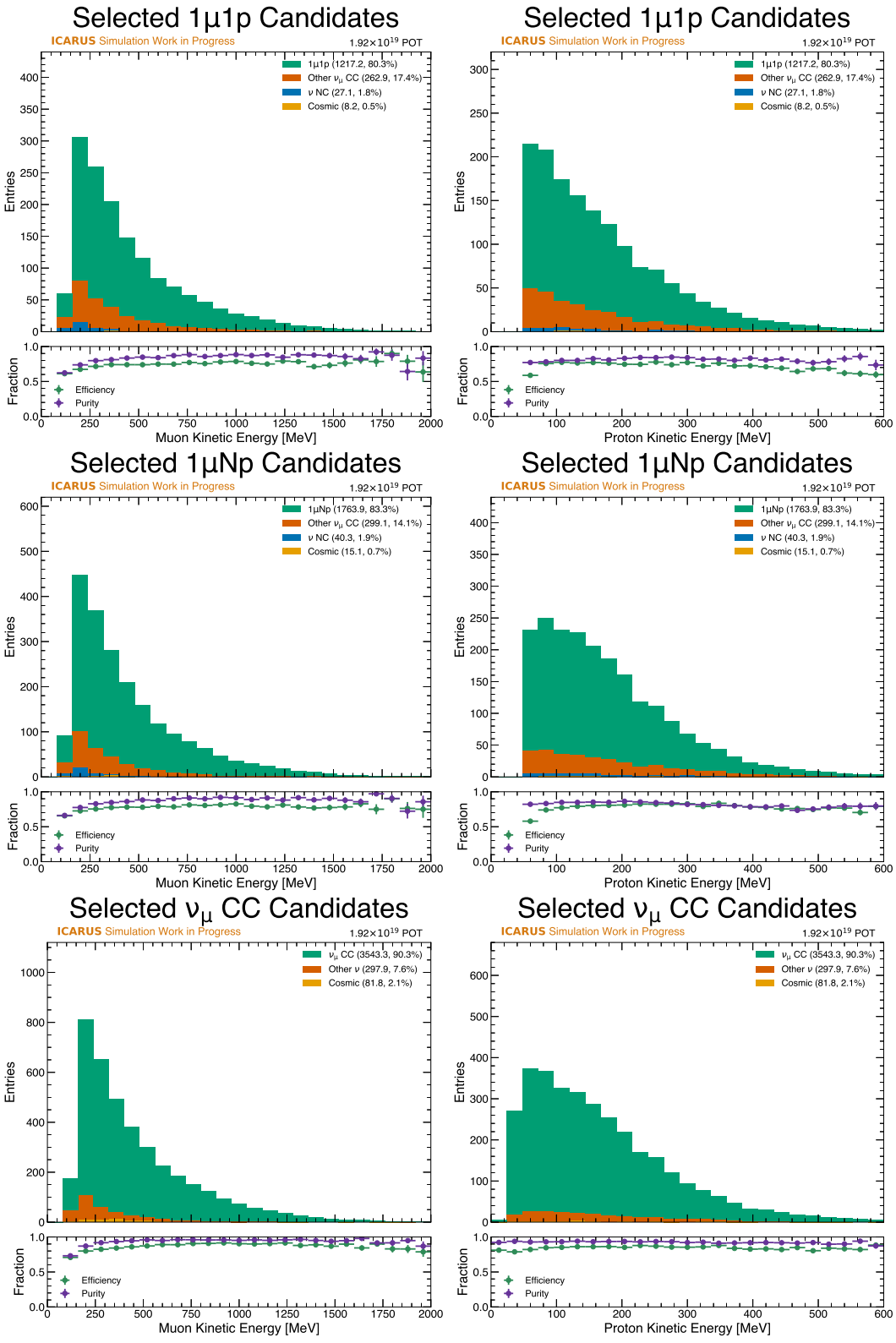


Figure 7.10: Purity and efficiency as a function of the kinetic energy of the muon (left) and the most energetic proton (right) for each of the three signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive.

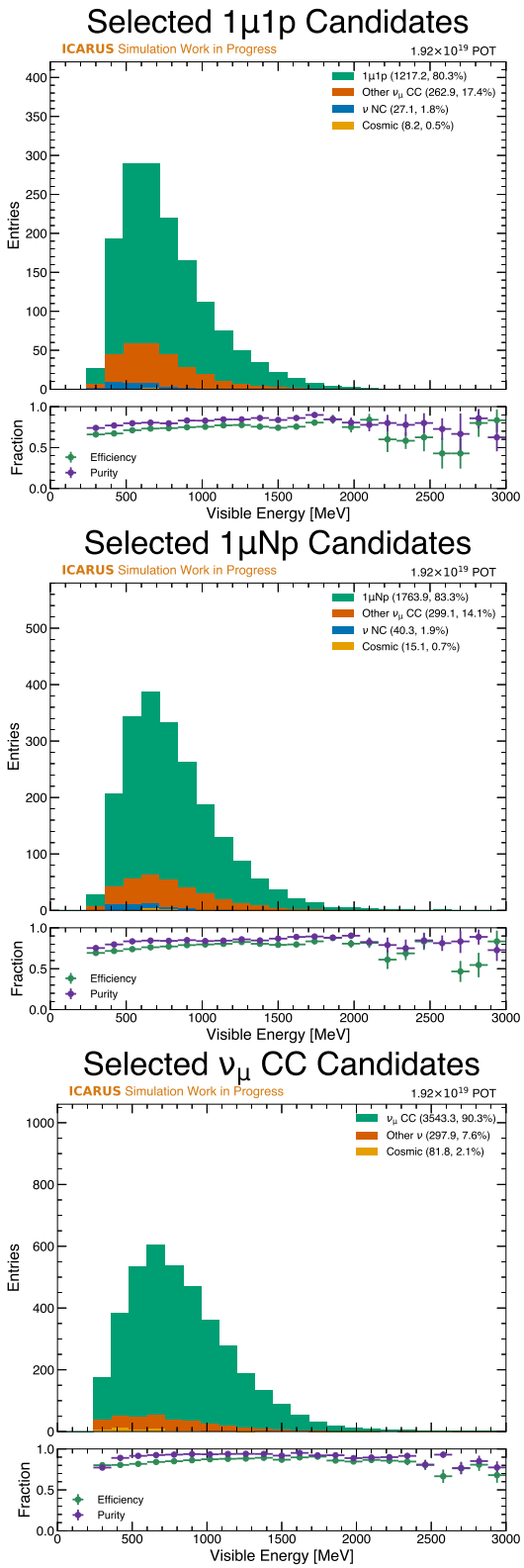


Figure 7.11: Purity and efficiency as a function of the total visible energy of the interaction for each of the signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive.

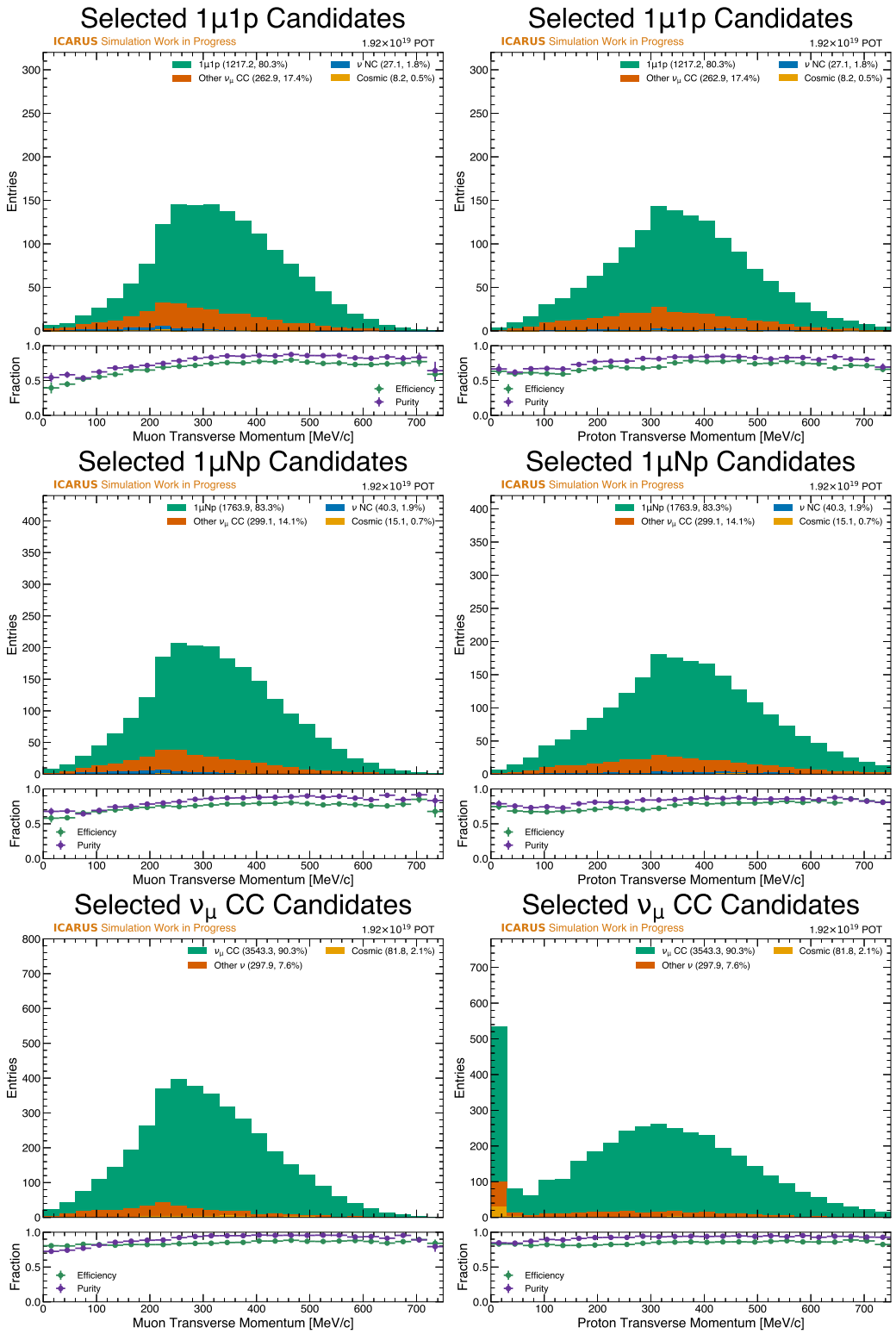


Figure 7.12: Purity and efficiency as a function of the transverse momentum of the muon (left) and the most energetic proton (right) for each of the three signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive. The peak at the lowest bin in the inclusive channel in the proton transverse momentum variable is from interactions where no proton is present.

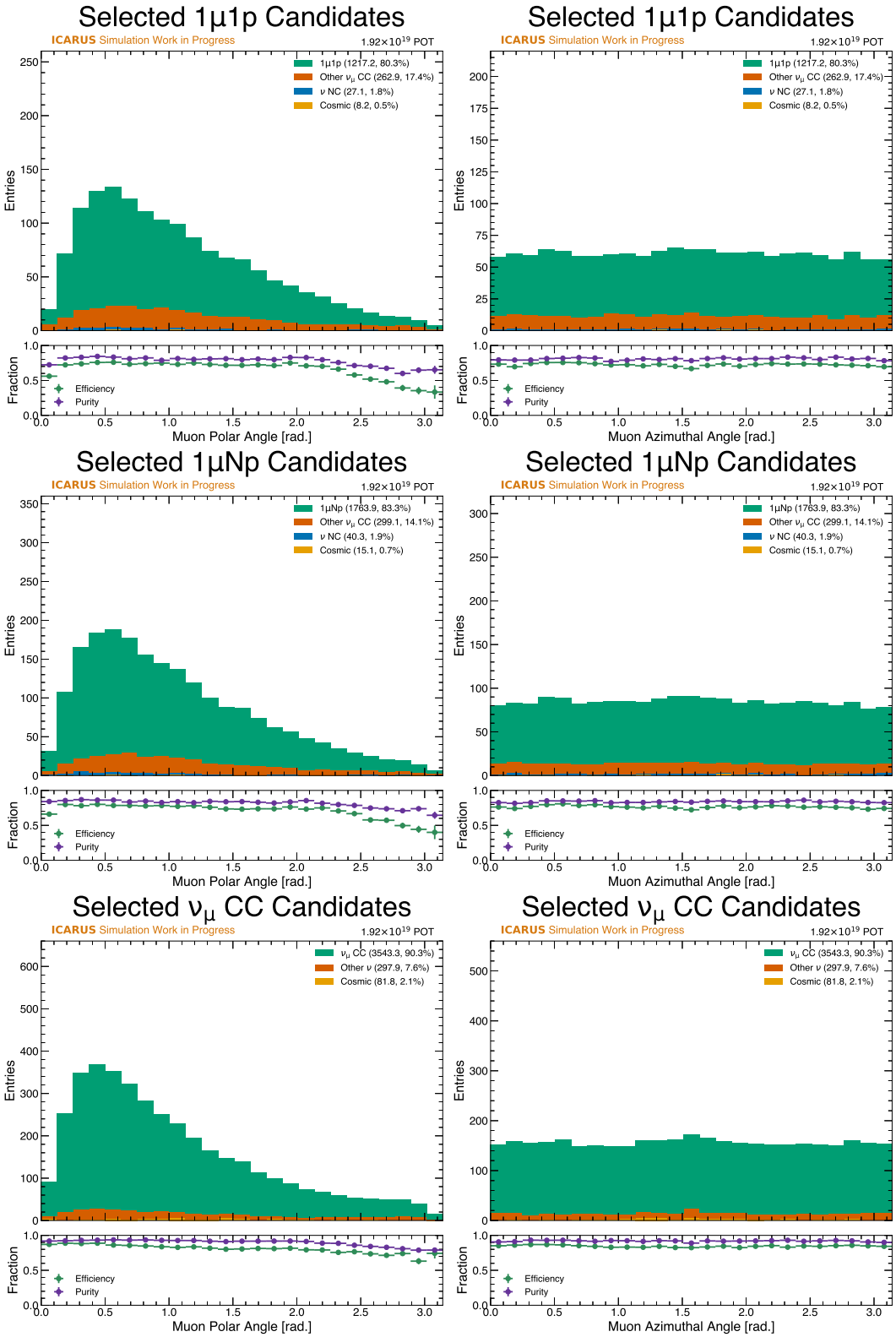


Figure 7.13: Purity and efficiency as a function of the polar angle (left) and azimuthal angle (right) of the muon for each of the three signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive.

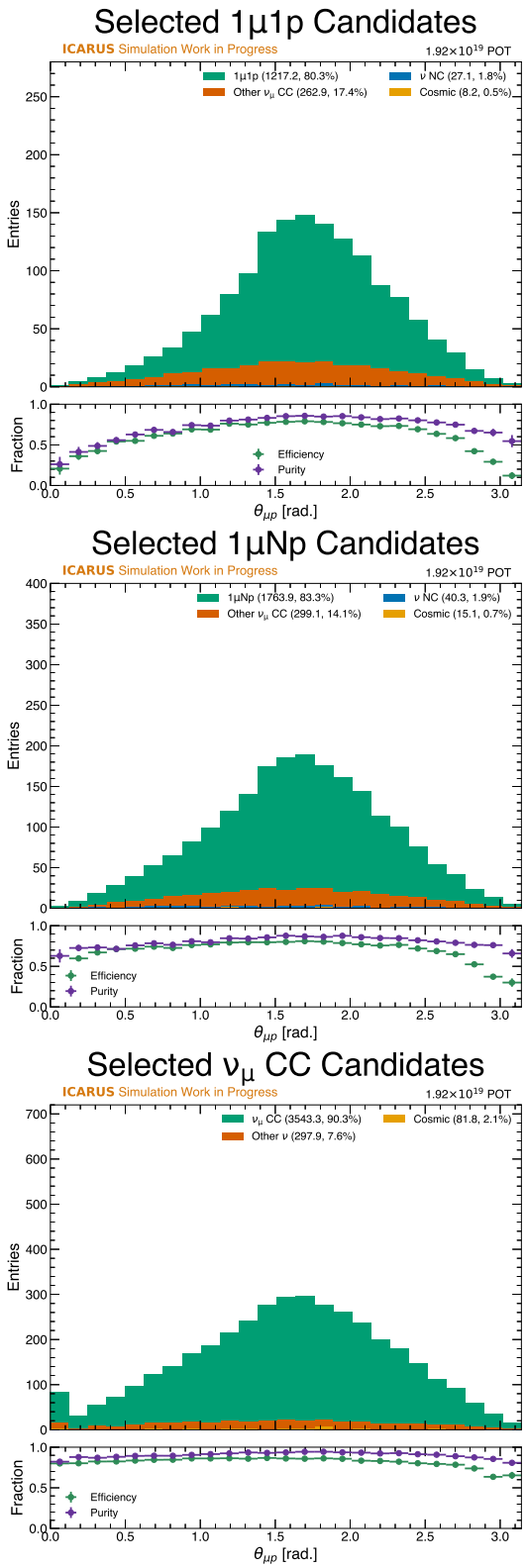


Figure 7.14: Purity and efficiency as a function of the the muon-proton opening angle for each of the three signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive.

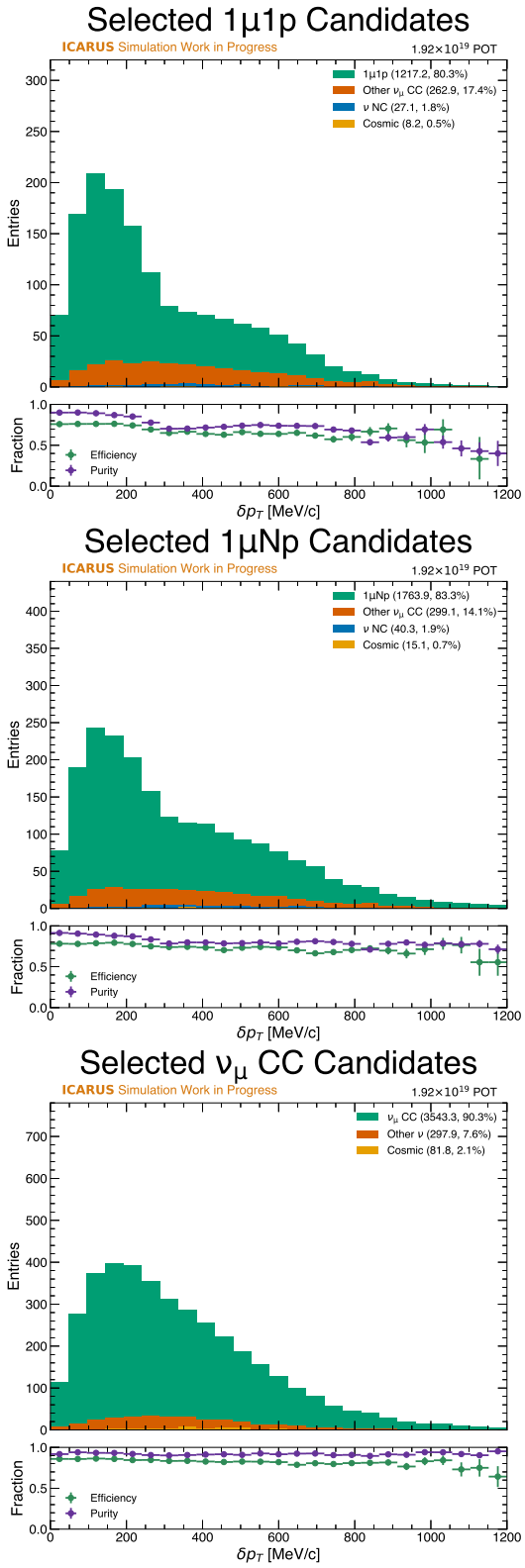


Figure 7.15: Purity and efficiency as a function of the transverse momentum of the interaction for each of the three signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive.

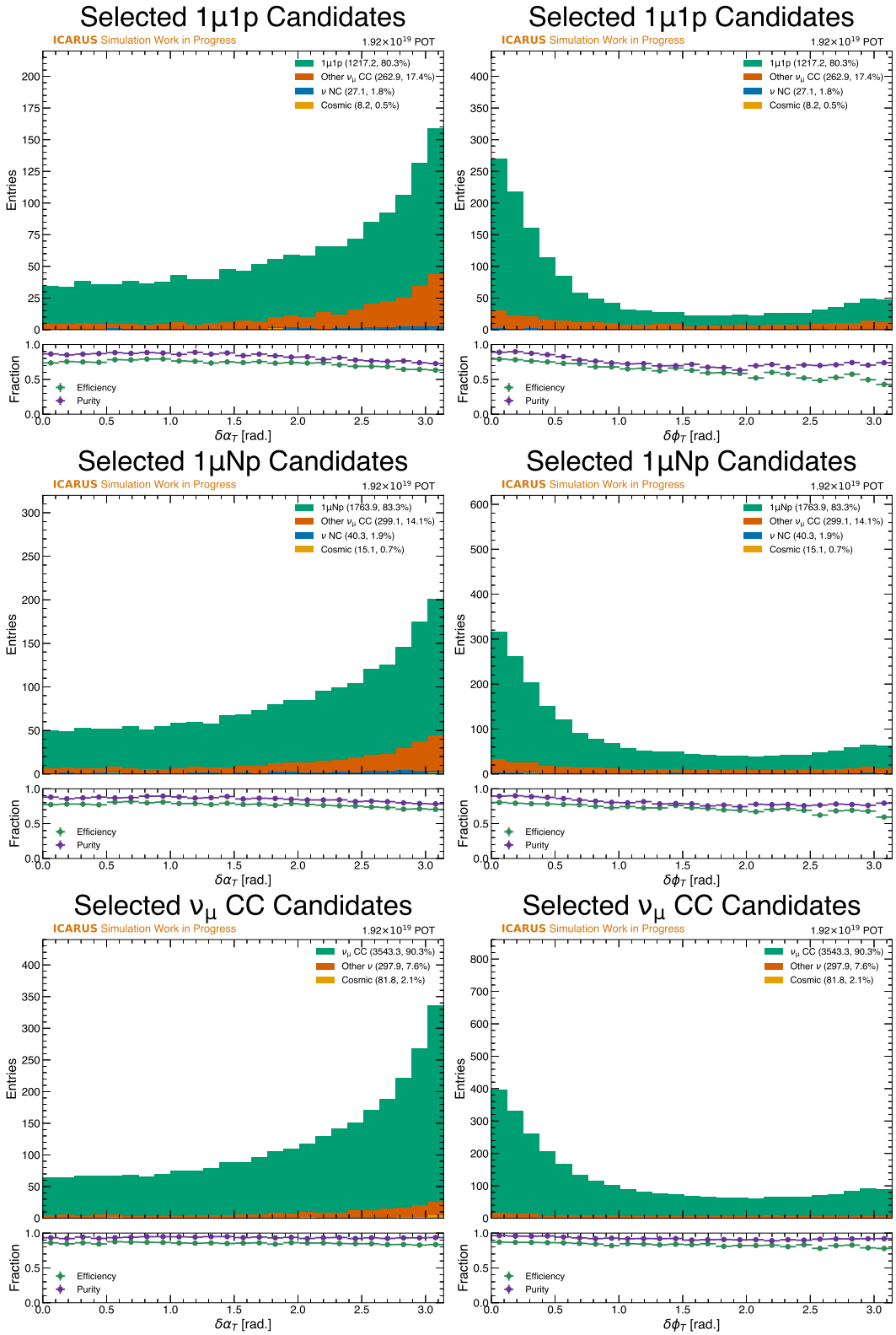


Figure 7.16: Purity and efficiency as a function of the kinematic imbalance variables $\delta\alpha_T$ (left) and $\delta\phi_T$ (right) for each of the three signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive.

Chapter 8

Systematic Uncertainties

Comparisons of the neutrino selection between data and Monte Carlo simulation only gain meaning within the context of properly characterized uncertainties on both the data measurement and the Monte Carlo simulation. The data measurement inherently contains statistical uncertainty reflecting the size of the data set. The expectation from the Monte Carlo simulation has, in addition to statistical uncertainty, other sources of systematic uncertainty arising from the uncertainty in the BNB flux, the neutrino interaction model, and the response of the detector to neutrino interactions.

The covariance matrix formalism [99] is used to accurately capture the scale and bin-to-bin correlations of each source of uncertainty. Under the assumption that the underlying sources of uncertainty are independent, the bilinearity of the covariance matrix allows for the independent calculation of covariance matrices for each source and the calculation of the total covariance matrix as the sum over all covariance matrices. It is worth noting that the effect of the sources of uncertainty are in general correlated to some degree, but the underlying sources are often independent. Section 8.1 describes the estimation of uncertainties associated with the BNB flux prediction. Section 8.2 describes the estimation of uncertainties associated with the modeling of ν-Ar interactions using the GENIE event generator. Section 8.3 describes the procedure for estimating systematic uncertainties associated with the modeling of detector effects. Finally, a summary of the overall uncertainties in the analysis is presented in Section 8.4.

8.1 BNB Flux Uncertainties

The BNB flux prediction used for ICARUS is based on the flux prediction computed for MiniBooNE [13]. The production rate of hadrons through p-Be interactions in the target is modeled using a GEANT4 simulation [74] and includes some additional re-weighting of pion production based on hadron production data as measured in the HARP experiment [100]. In addition to the uncertainties arising from hadron production, there are also uncertainties associated with the re-

scattering of pions and protons in the beryllium and aluminum horn through quasielastic and inelastic processes and the mis-modeling of current flow through the magnetic focusing horn. Additionally, there is a flat 2% uncertainty associated with the total number of POT due to the uncertainty in the measured POT from the BNB [13].

The estimation of the flux covariance matrix follows a "many-universe" technique, where the flux parameters associated with these systematic uncertainty sources are varied following a Gaussian distribution centered at the predicted central value and width. A total of $M = 1000$ universes are simulated in this way for each systematic parameter and each neutrino interaction is assigned a weight reflecting the probability of the true parameter living in that universe. After binning the selected interactions for each universe, the covariance E_{ij}^{α} between bins i and j under the effect of systematic parameter α is computed as:

$$E_{ij}^{\alpha} = \frac{1}{M} \sum_{m=1}^{M} (N_{CV}^{i} - N_{\alpha,m}^{i})(N_{CV}^{j} - N_{\alpha,m}^{j}) \tag{8.1}$$

where N_{CV}^{i} is the central value prediction for bin i and $N_{\alpha,m}^{i}$ is the content of bin i in the universe m. The total covariance matrix is then computed as the sum of the covariance matrices for each systematic parameter. This process is then be repeated for each reconstructed variable discussed in Section 7.2. A breakdown of the flux uncertainty is shown in Table 8.1.

8.2 Neutrino Interaction Uncertainties

As described in Section 5.1.2, neutrino interactions are simulated using the GENIE event generator [75] for the central value prediction. Within GENIE, there are a large number of parameters available for re-weighting that are used in the simulation of neutrino interactions. To estimate the uncertainty in the neutrino interaction model, the many-universe technique is used to vary these parameters in order to individually characterize their effect on the downstream analysis. The covariance matrix for each parameter is computed with Equation 8.1 as done for the flux uncertainties. The total covariance matrix is then computed as the sum of the covariance matrices for each parameter. A breakdown of the neutrino interaction model uncertainty is shown in Table 8.2.

Table 8.1: A breakdown of the overall scale of each flux uncertainty for each of the three signal definitions. Beamline uncertainties are those associated with the re-scattering of hadrons in the target and the modeling of the magnetic focusing horn. Hadron production uncertainties are those associated with the production of hadrons in the target.

Uncertainty	Type	$1\mu 1p$ [%]	$1\mu Np$ [%]	ν_μ CC [%]
Skin Depth	Beamline	2.2	2.7	3.2
Horn Current	Beamline	0.4	0.5	0.5
π^+	Hadron Production	4.7	4.7	4.8
π^-	Hadron Production	0.1	0.0	0.1
K^+	Hadron Production	0.1	0.1	0.2
K^-	Hadron Production	0.0	0.0	0.0
K^0_L	Hadron Production	0.0	0.0	0.0
Nucleon Inelastic	Beamline	0.9	0.9	0.9
Nucleon QE	Beamline	2.5	2.5	2.5
Nucleon Total	Beamline	0.8	0.8	0.8
π Inelastic	Beamline	1.2	1.2	1.2
π QE	Beamline	0.8	0.9	0.8
π Total	Beamline	0.7	0.8	0.8
POT	Beamline	2.0	2.0	2.0
Total Flux		6.7	6.9	7.3

Many of these individual parameters have a small effect on the overall uncertainty, but there are a few worth highlighting:

- The parameter associated with the Z-expansion of the axial form factor has a large effect on the overall uncertainty. This is especially true for the two exclusive channels, which are more dominated by QE interaction modes.

- The RPA suppression parameter has an effect in the range of 3-7%. This is largest for the inclusive channel.

- The normalization for CC MEC processes is large for all channels, but is especially large for the 1μNp channel. This is due to the larger contribution of MEC processes to the 1μN$p$channel.

Table 8.2: A breakdown of the overall scale of each neutrino interaction model uncertainty for each of the three signal definitions.

Uncertainty	Description	$1\mu1p$ [%]	$1\mu Np$ [%]	ν_μ CC [%]
ZExpAVariationResponse	Z-expansion description of the axial-vector form factor on CC-QE	9.6	8.1	6.3
RPA CCQE	RPA suppression of CC-QE	4.7	3.3	6.8
Coulomb CCQE	The strength of the electromagnetic potential for the Coulomb corrections on CC-QE	0.5	0.4	0.3
Norm CCMEC	Normalization of CC-MEC	7.5	10.2	7.1
Norm NCMEC	Normalization of NC-MEC	0.0	0.0	0.0
NCEL Variation Response	Variation of the dipole form factor on NCEL	0.1	0.1	0.1
CCRES Variation Response	Variation of the dipole form factor on CC-RES	2.2	4.5	6.7
NCRES Variation Response	Variation of the dipole form factor on NC-RES	0.4	0.4	0.5
NonRESBGvpCC1pi	Scale factor for the non-resonant background level of ν-p CC + 1π	0.0	0.1	0.1
NonRESBGvpCC2pi	Scale factor for the non-resonant background level of ν-p CC + 2π	0.1	0.2	0.6
NonRESBGvpNC1pi	Scale factor for the non-resonant background level of ν-p NC + 1π	0.0	0.0	0.0
NonRESBGvpNC2pi	Scale factor for the non-resonant background level of ν-p NC + 2π	0.0	0.0	0.1
NonRESBGvnCC1pi	Scale factor for the non-resonant background level of ν-n CC + 1π	0.1	0.2	0.4
NonRESBGvnCC2pi	Scale factor for the non-resonant background level of ν-n CC + 2π	0.1	0.2	0.7
NonRESBGvnNC1pi	Scale factor for the non-resonant background level of ν-n NC + 1π	0.2	0.2	0.2
NonRESBGvnNC2pi	Scale factor for the non-resonant background level of ν-n NC + 2π	0.0	0.0	0.0
NonRESBGvbarpCC1pi	Scale factor for the non-resonant background level of $\bar{\nu}$-p CC + 1π	0.0	0.0	0.0
NonRESBGvbarpCC2pi	Scale factor for the non-resonant background level of $\bar{\nu}$-p CC + 2π	0.0	0.0	0.0
NonRESBGvbarpNC1pi	Scale factor for the non-resonant background level of $\bar{\nu}$-p NC + 1π	0.0	0.0	0.0
NonRESBGvbarpNC2pi	Scale factor for the non-resonant background level of $\bar{\nu}$-p NC + 2π	0.0	0.0	0.0
NonRESBGvbarnCC1pi	Scale factor for the non-resonant background level of $\bar{\nu}$-n CC + 1π	0.0	0.0	0.0
NonRESBGvbarnCC2pi	Scale factor for the non-resonant background level of $\bar{\nu}$-n CC + 2π	0.0	0.0	0.0
NonRESBGvbarnNC1pi	Scale factor for the non-resonant background level of $\bar{\nu}$-n NC + 1π	0.0	0.0	0.0
NonRESBGvbarnNC2pi	Scale factor for the non-resonant background level of $\bar{\nu}$-n NC + 2π	0.0	0.0	0.0
RDecBR1gamma	Scale factor for the branching fraction of X + 1γ	0.0	0.0	0.0
RDecBR1eta	Scale factor for the branching fraction of X + 1η	0.2	0.3	0.4
COH Variation Response	Normalization of the Coherent production process	0.1	0.1	0.2
DISBY Variation Response	Normalization of the BY model of the DIS process	0.0	0.0	0.0
FSI π Variation Response	Variation for FSI involving pions	1.2	3.5	0.7
FSI N Variation Response	Variation for FSI involving nucleons	3.9	3.0	1.2
Total Cross Section		14.7	15.7	14.6

- The parameter controlling the dipole form factor for CC resonant production is progressively larger as the final state becomes less restrictive due to higher contributions from resonant production.

- Many of the parameters associated with anti-neutrino interactions have vanishingly small effects on the overall uncertainty. This is due to the low flux of anti-neutrinos in the BNB in the "neutrino-mode" horn configuration, which has been the operating mode for all physics datasets collected by the ICARUS detector.

Overall, all three channels have a resulting total uncertainty due to interaction model uncertainties of around 15%. The $1\mu\mathrm{N}p$ channel has a slightly larger uncertainty (around 1% more) due to the larger contribution of MEC processes to this channel, which have larger uncertainties associated with them.

8.3 Detector Response Uncertainties

The detector response uncertainty reflects many non-idealizations of the ICARUS detector and the underlying detector physics that can impact event reconstruction and signal selection. The effects considered can be broadly grouped into three categories:

1. Light yield, scintillation photon propagation, and optical response throughout the detector.

2. Electron-ion recombination, space charge effects, electron lifetime, diffusion, and other processes affecting drifting ionization charge.

3. Variable TPC channel gain, non-uniformities in response across the wire planes, electronic noise, and other effects impacting signal reconstruction associated with the TPC wires.

The detector response uncertainty is characterized through dedicated detector variation samples. Each detector variation sample implements a single variation in some underlying component of the Monte Carlo simulation that fully covers some aspect of the detector response that is not fully understood or known to be insufficiently modeled.

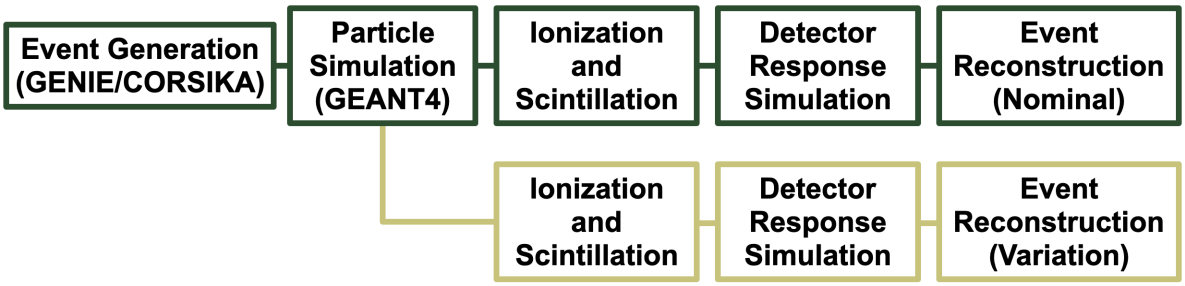


Figure 8.1: A diagram representing the full simulation chain for both the nominal simulation (green) and a detector variation sample (gold). The samples share the same set of generated events including particle trajectories and daughters, but downstream processes may vary according to statistical processes or the variation itself.

A sample of N_{CV} total events generated with the nominal simulation parameters and run through the full detector simulation and event reconstruction is used as the central value prediction. The same set of N_{CV} events is then run through the full detector simulation and event reconstruction with the detector variation applied. Random processes in the GEANT4 simulation of particle trajectories and secondaries may produce additional statistical differences between the nominal and systematic variation sample, so a workflow which shares the generated particles and their trajectories between both samples is used to characterize the detector systematics. A diagram representing the full simulation chain for both the nominal simulation and a detector variation sample is shown in Figure 8.1.

For this analysis, samples of $N_{CV} = 200,000$ events were generated for the central value and each detector variation sample. The choice of detector systematics to vary was made based on the expected impact of the variation on the final event selection. This list is not exhaustive and further detector systematics will be studied in the future. Due to the large time and computational resources required to generate each sample, the decision was made to use neutrino-only samples to characterize the detector systematics. Though the presence of cosmic rays does have an impact on the selection, the relative effect of the variation on the selection is expected to be similar between the neutrino-only and full samples. The detector variation samples produced for this analysis are:

TPC Signal Shape: This is a variation in the shape of the signal response on the wire planes. The central value sample uses a signal response that was tuned with cosmic muon data, whereas the variation sample uses the original signal response that was produced with a simulation. This is a conservative estimate of the uncertainty in the signal shape, as this is essentially treating the full effect of a known bias as a systematic uncertainty. Future detector systematic studies within ICARUS will aim to better characterize the uncertainty in the signal shape.

First Induction Plane Gain: This is a variation in the gain of the first induction plane of $\pm 15\%$, implemented separately as two distinct samples. These variations reflect the un-simulated variation of the gain across the first induction plane.

PMT Quantum Efficiency: There is a known data to simulation discrepancy in the optical model that is in the process of being understood. This variation is meant to cover this discrepancy by varying the quantum efficiency of the PMTs by 4%. In addition to potential impacts on the flash matching performance, this variation also directly impacts the emulated trigger decision and therefore may cause a shift in the selected events by changing the trigger efficiency.

TPC Coherent Noise: The coherent component of the TPC noise varied by 4.9% with respect to the nominal noise model over the duration of Run 2. A positive variation of this scale is applied to the nominal simulation to cover this discrepancy.

TPC Intrinsic Noise: Similar to the coherent noise variation, the intrinsic noise of the TPC varied by 3.7% with respect to the nominal noise model over the duration of Run 2. A positive variation of this scale is applied to the nominal simulation to cover this discrepancy.

Electron-ion Recombination: The recombination model used in the nominal simulation is the Modified Box model, as measured by ArgoNeuT [87]. Recently, the dependence of recombination on the angle of the track with respect to the electric field has been measured at ICARUS and encap-

sulated in a model referred to as the "Ellipsoidal" recombination model. The variation sample uses this model to estimate the uncertainty that may arise from improperly modeling recombination.

The comparison of the central value and variation samples is used to estimate the uncertainty in the detector response. The procedure used is covered in the following section.

8.3.1 Bootstrapping

The detector response uncertainty is estimated using a bootstrapping technique similar to the one employed in several differential cross section measurements performed by the MicroBooNE collaboration [56, 57, 101]. An overview of bootstrapping techniques can be found in [102]. The general idea is to estimate the uncertainty in a measurement by repeatedly sampling the data with replacement and computing the quantity of interest for each sample. The distribution of these quantities is then used to estimate the uncertainty in the measurement. This method inherently captures the bin-to-bin correlations in the uncertainty, which is important for the downstream analysis. The assumption made in this method is that the distribution being sampled from is sufficiently large enough to represent the true distribution of the data.

In the bootstrapping technique used for this analysis, the set of common events between the central value and a variation sample is sampled N_{CV} times with replacement to form one "universe." In each universe, the selected signal candidates from the central value sample and the variation sample are binned according to the reconstructed variable of interest. A difference vector $\vec{V}_{D,i}$ is computed for the i'th universe as the difference across the bins between the two samples. This process is repeated $N_{\text{bootstrap}} = 1000$ times to form a distribution of difference vectors. The average 1σ deviation $\vec{V}_D^{nominal}$ is then computed as the average of the difference vectors across the universes.

The uncertainty on $\vec{V}_D^{nominal}$ is then computed as the covariance matrix M_R of the difference vectors across the universes as is described in Equation 8.1. This covariance matrix can be understood as the matrix governing the Gaussian distribution of the difference vectors, or in other

Table 8.3: A breakdown of the overall scale of each detector response uncertainty for each of the three signal definitions.

Uncertainty	Type	$1\mu1p$ [%]	$1\mu Np$ [%]	ν_μ CC [%]
TPC Signal Shape	TPC	5.5	6.4	3.3
Increased Front Induction Gain	TPC	2.1	2.2	2.2
Decreased Front Induction Gain	TPC	2.3	2.2	2.4
Ellipsoidal Recombination	TPC	2.1	1.9	2.5
Increased Coherent Noise	TPC	2.1	2.2	2.2
Increased Intrinsic Noise	TPC	2.5	2.3	2.4
PMT Quantum Efficiency	PMT	2.1	2.2	2.2
Total Detector		8.1	8.6	6.9

words the matrix that describes how to perturb $\vec{V}_D^{nominal}$ according to the fluctuations seen in the bootstrapping procedure.

The overall covariance matrix for the variation is then computed using both $\vec{V}_D^{nominal}$ and M_R in a way that fully captures the detector response uncertainty. A new set of universes is generated in this step, with $N_{\text{universes}} = 1000$. In each universe, the matrix M_R is used to generate a perturbation $\delta\vec{V}_D$ by sampling from a Gaussian distribution centered at $\vec{V}_D^{nominal}$ with covariance M_R. This is realized by computing the Cholesky decomposition of M_R and using this to generate a random vector with elements correlated according to M_R from a vector with normally distributed, uncorrelated elements. The perturbation $\delta\vec{V}_D$ is then added to the central value vector $\vec{V}^{nominal}$ and scaled by a unit Gaussian r_i to form the final vector $\vec{V}_D^{final}$, as shown in Equation 8.2.

$$\vec{V}_{D,i}^{final} = r_i \left(\vec{V}^{nominal} + \delta\vec{V}_{D,i} \right) \tag{8.2}$$

This set of vectors $\vec{V}_{D,i}^{final}\ \forall i \in [1, N_{\text{universes}}]$ is then used to compute the covariance matrix M_D that describes the uncertainty on the event selection due to the variation in the detector response. This process is then repeated for each of the variation samples and for each of the reconstructed variables discussed in Section 7.2. These covariance matrices can then be combined to form the total covariance matrix for the detector response uncertainty. A breakdown of the detector response uncertainty is shown in Table 8.3

From these results, it is clear that the variation of the TPC signal shape is the most impactful on this analysis. This is not surprising as the signal shape is a fundamental aspect of the signal processing and affects all downstream reconstruction. Additionally, as noted in the summary of the variations above, this variation treats the full effect of a known bias as a systematic uncertainty. Work is ongoing within ICARUS to establish post-calibration uncertainties to more accurately model this effect. In contrast, the other variations have reached a floor of around 2% uncertainty. A few things are likely contributing to this floor: the statistical nature of the bootstrapping technique, the size of the sample used to estimate the uncertainty, and the remaining statistical processes that are not fully shared between the central value and variation samples. The actual uncertainty due to these variations is likely smaller than the 2% floor, so the TPC signal shape uncertainty can be treated as lower bound on the uncertainty due to the detector response.

8.4 Summary of Event Selection Uncertainties

The overall uncertainty in the event selection is computed as the sum of the covariance matrices for each source of uncertainty described in the previous sections. This total covariance matrix can then be used to place uncertainty bars on the histogram entries for each reconstructed variable used in the analysis, thus allowing for a visual representation of the uncertainty in the event selection. A more quantitative use of the covariance matrix is in the calculation of the χ^2 test statistic between data and the prediction. This will be discussed in more detail in the next chapter. A summary of the overall uncertainty in the event selection, broken down by category, is shown in Table 8.4. The total correlation matrix for the $1\mu1p$, $1\mu\text{N}p$, and ν_μ CC inclusive channels is shown in Figure 8.2.

By far the largest source of uncertainty in the event selection is the neutrino interaction model uncertainty. This will have the biggest impact in ICARUS-only analyses, where cancellation of systematic uncertainties between the near and far detectors is not possible. The detector response uncertainty is the next largest source of uncertainty, but is expected to be reduced in future analyses as the detector response simulation is improved and more precise variations of the detector response are implemented. Looking towards joint SBN analyses where the flux and neutrino inter-

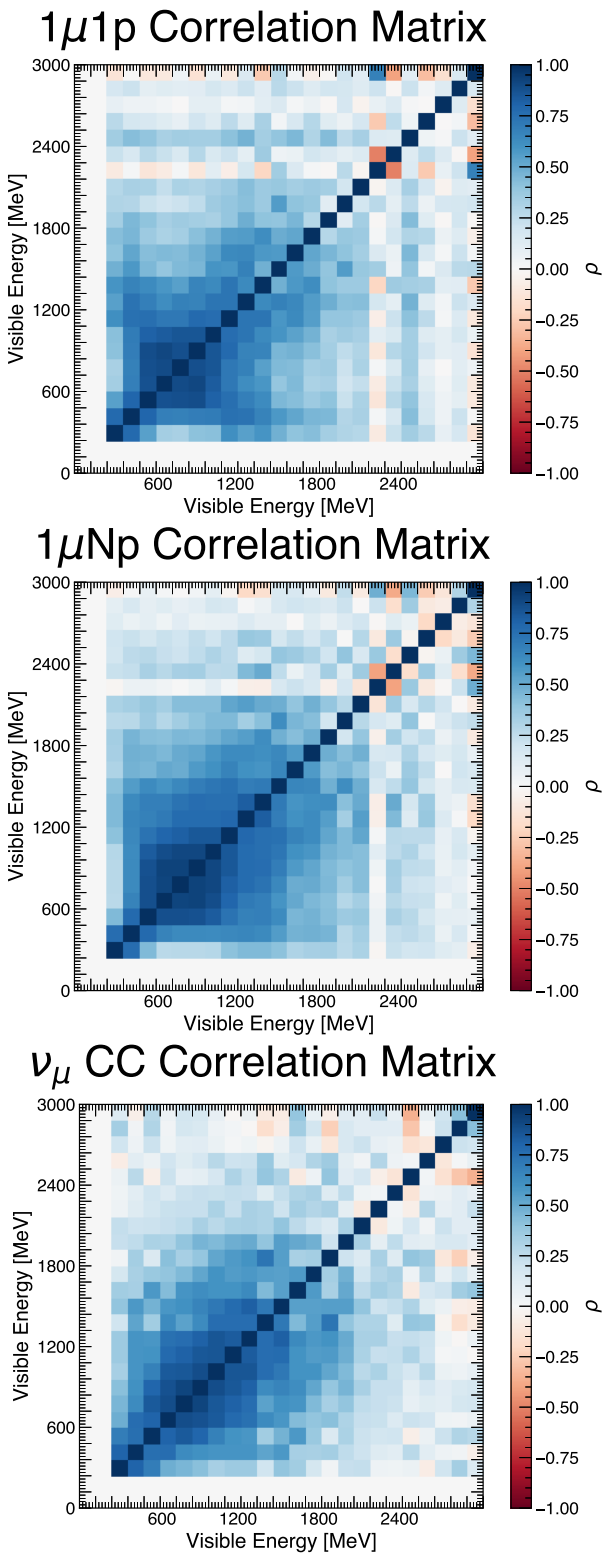


Figure 8.2: The total correlation matrix for the $1\mu 1p$ (top), $1\mu \mathrm{N}p$ (middle), and ν_μ CC inclusive (bottom) channels. The variable shown is the reconstructed visible energy.

Table 8.4: A breakdown of the overall scale of each uncertainty source by category for each of the three signal definitions.

Uncertainty	Description	$1\mu1p$ [%]	$1\mu Np$ [%]	ν_μ CC [%]
Detector	Uncertainties related to the modeling of the detector response	8.1	8.6	6.9
Flux	Uncertainties related to the modeling of the neutrino flux	6.7	6.9	7.3
Cross Section	Uncertainties related to neutrino interaction modeling	14.7	15.7	14.6
Statistical	Statistical uncertainties associated with the Monte Carlo simulation sample	2.4	2.0	1.6
Off-beam Statistical	Statistical uncertainties associated with the off-beam sample used to estimate cosmic backgrounds	0.3	0.4	0.5
Total		18.6	19.6	18.3

action model uncertainties cancel to first order, the detector response uncertainty will become the dominant source of uncertainty. The statistical uncertainty associated with the simulated central value sample and the off-beam dataset are both negligible in comparison to the other sources of uncertainty.

Chapter 9

Data/Simulation Comparisons

This chapter presents a comparison of Run 2 ICARUS data and the corresponding expectations from simulation. Section 9.1 describes the dataset used for this analysis, including the rejection of subsets of the dataset due to non-ideal detector conditions and the steps taken to comply with the ICARUS blinding policy. Section 9.2 presents the comparison of data and simulation in each of the reconstructed variables along with some discussion on the observed data/simulation agreement.

9.1 Run 2 Dataset

The dataset used for this analysis consists of the ICARUS Run 2 data collected between December 20, 2022 and July 16, 2023. This period of data collection represents the largest dataset with stable detector conditions collected by the ICARUS T600 detector at Fermilab to date, corresponding to approximately 2.05×10^{20} POT. More details about this physics run can be found in Section 4.5.

9.1.1 Data Quality

Though the detector was overall stable throughout the duration of this physics run, there were instances of detector instabilities that resulted in portions of DAQ runs being flagged as bad. These instabilities were typically due to single-component failures, such as the power supply of a TPC readout crate, or due to issues with the DAQ, trigger, cathode, or wire bias systems. These issues were logged and used as part of the data processing to identify and remove bad runs from the dataset. This results in a final dataset of around 1.92×10^{20} POT.

9.1.2 The ICARUS Blinding Policy

The ICARUS collaboration has an official data blinding policy in place to prevent experimenter bias in the analysis of the data. The philosophy behind this policy is to ensure that an analyzer will

not prematurely access the data in a statistically significant way that could influence the development of the analysis. By default, only 10% of the data is fully unblinded for the purpose of developing the analysis, while the remaining 90% is kept blinded until the analysis is finalized. The blinding policy also sets a process for accessing the blinded data, which is typically done in stages following the submission of technical documentation of the analysis and review by the collaboration.

In practical terms, this policy is enforced at the level of the ICARUS data processing and analysis software. A decision is made per-event to determine if the event is blinded or unblinded using a random number service that is seeded with a configurable parameter. This allows for reproducible blinding decisions, which is important for maintaining the same set of blinded events across all data-processing campaigns. The files ("Common Analysis Framework" files, or CAF files) containing analysis-level information for the 10% of the data that is declared unblinded are kept entirely separate from the CAF files containing the blinded data. It is worth noting that this policy only applies to data streams that contain neutrino interactions and not to cosmic-only data streams such as the off-beam data.

The goals of this analysis are to demonstrate the performance of the ICARUS machine-learning analysis chain on the Run 2 dataset in signal channels of relevance for ICARUS and the SBN Program. This can be done with the 10% of the data that is unblinded and does not require the increased statistics of the full Run 2 dataset. The final dataset size for this analysis is therefore 1.92×10^{19} POT, after accounting for data processing inefficiencies and removing data from periods of detector instability. The data used for this analysis was processed in full compliance with the ICARUS blinding policy. At no point during the development of this analysis was any portion of the blinded data accessed.

9.2 Data/Simulation Comparisons

This section presents the comparison of Run 2 ICARUS data and the corresponding expectations from simulation across a wide range of variables. The comparison is performed for the three

signal channels of interest for ICARUS and the SBN Program: $1\mu 1\mathrm{p}$, $1\mu \mathrm{N}\mathrm{p}$, and ν_μ CC inclusive. Agreement between data and simulation is crucial for demonstrating the readiness of the ICARUS machine-learning analysis chain for future analyses.

Each of the subsections below shows a set of distributions for a particular set of variables. In each distribution, the data is shown as black points with error bars representing the statistical uncertainty on the data. The expectation from the combination of simulation and off-beam data is shown as the filled histogram, with the error bands representing the full uncertainty as described in Chapter 8. The lower panel of each plot shows the ratio of data to the expectation with the same error bars and bands superimposed.

To quantitatively assess the agreement, the χ^2 test statistic is calculated for each distribution. The χ^2 is calculated using the sum of the overall covariance matrix for the expectation and the statistical covariance matrix for the data:

$$\chi^2 = \sum_{i,j} (O_i - E_i) \left(\Sigma^{-1}\right)_{ij} (O_j - E_j) \tag{9.1}$$

where O_i and E_i are the observed and expected values in bin i, respectively, and Σ is the covariance matrix. The χ^2 is then normalized by the number of degrees of freedom, which is the number of nonzero bins minus one. The χ^2 per degree of freedom is reported for each distribution.

9.2.1 Comparisons for Interaction Kinematic Variables

Comparisons of data and simulation are presented in Figures 9.1 through 9.4 for the interaction kinematic variables defined in Section 7.2.1. Generally, an excess of events is observed in the data compared to simulation, but the difference is near the expected uncertainty. The muon and proton energy distributions seem to have shifts in the distribution that are opposite: the muon energy distribution is shifted to higher energies in data compared to simulation, while the proton energy distribution is shifted to lower energies. The total visible energy distribution shows a generally higher amount of events in data compared to simulation, but the shape of the distribution matches well.

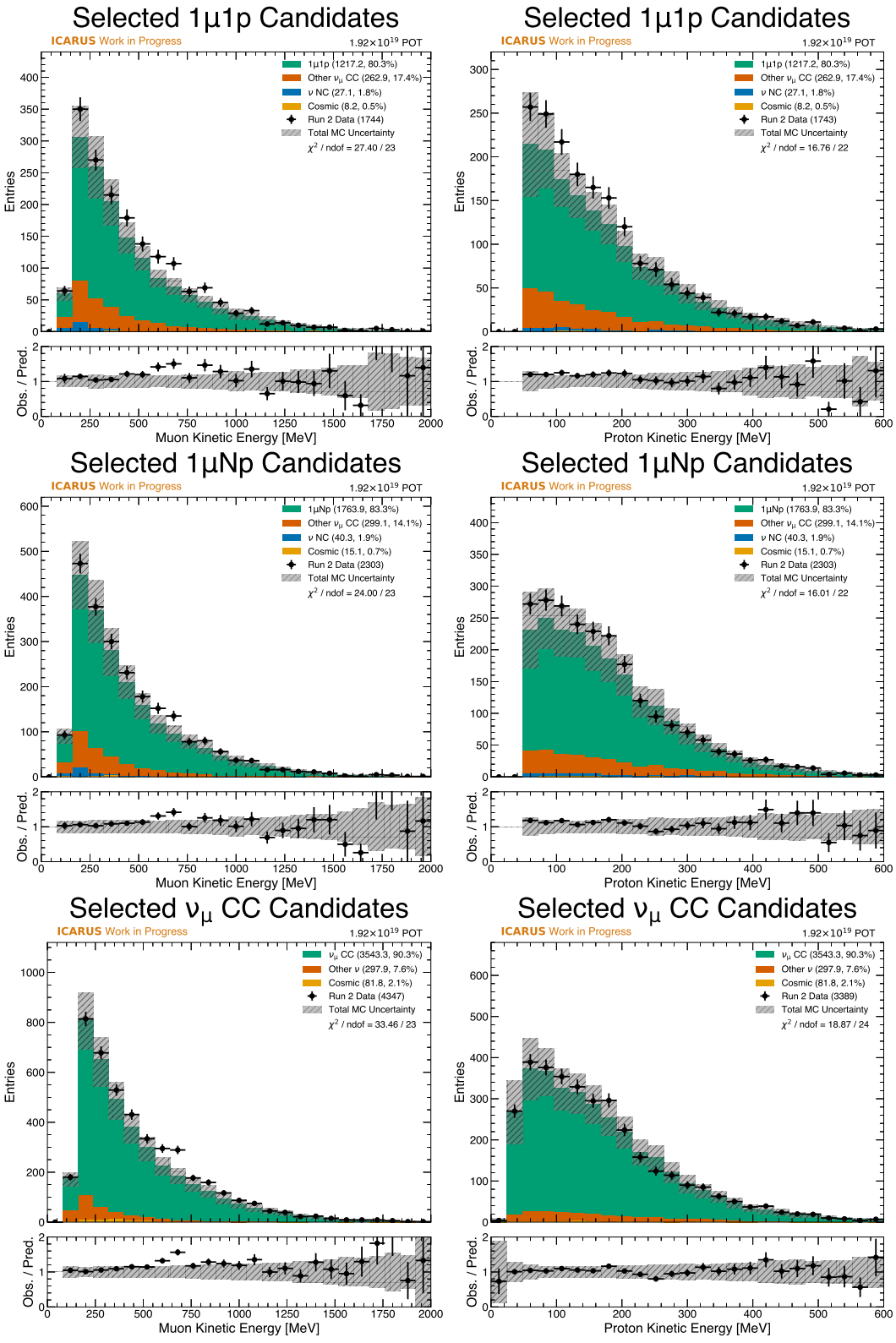


Figure 9.1: Comparison of data and simulation for the kinetic energy of the muon (left) and the most energetic proton (right) for each of the three signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive.

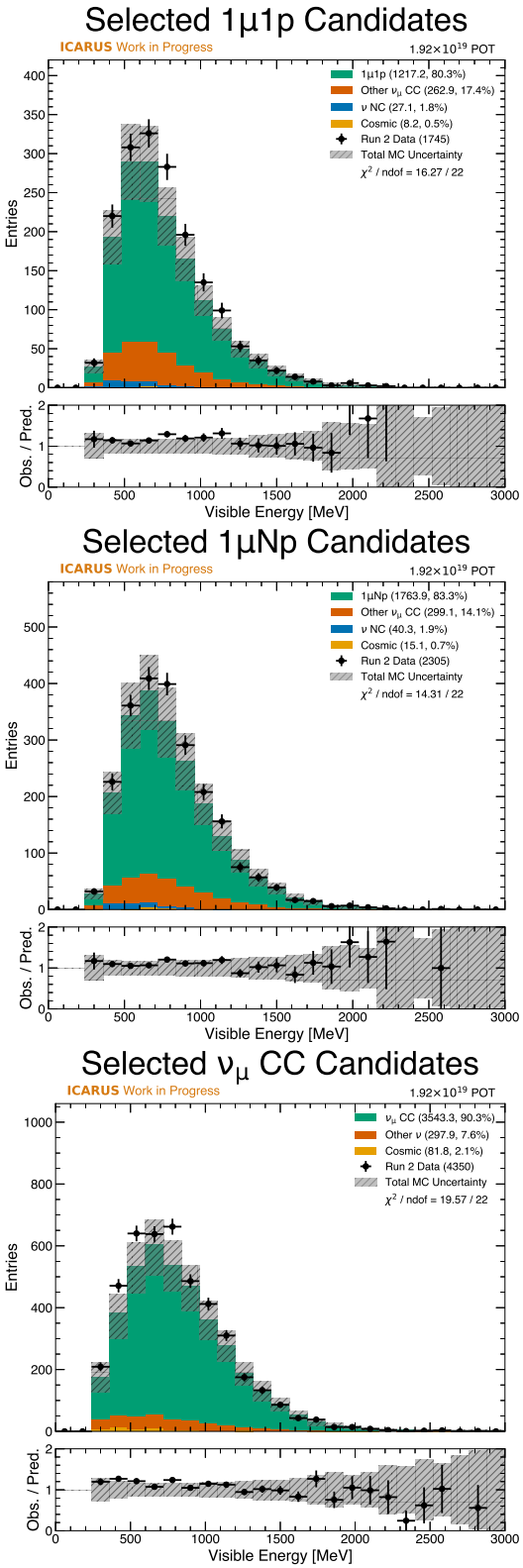


Figure 9.2: Comparison of data and simulation for the total visible energy of the interaction for each of the signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive.

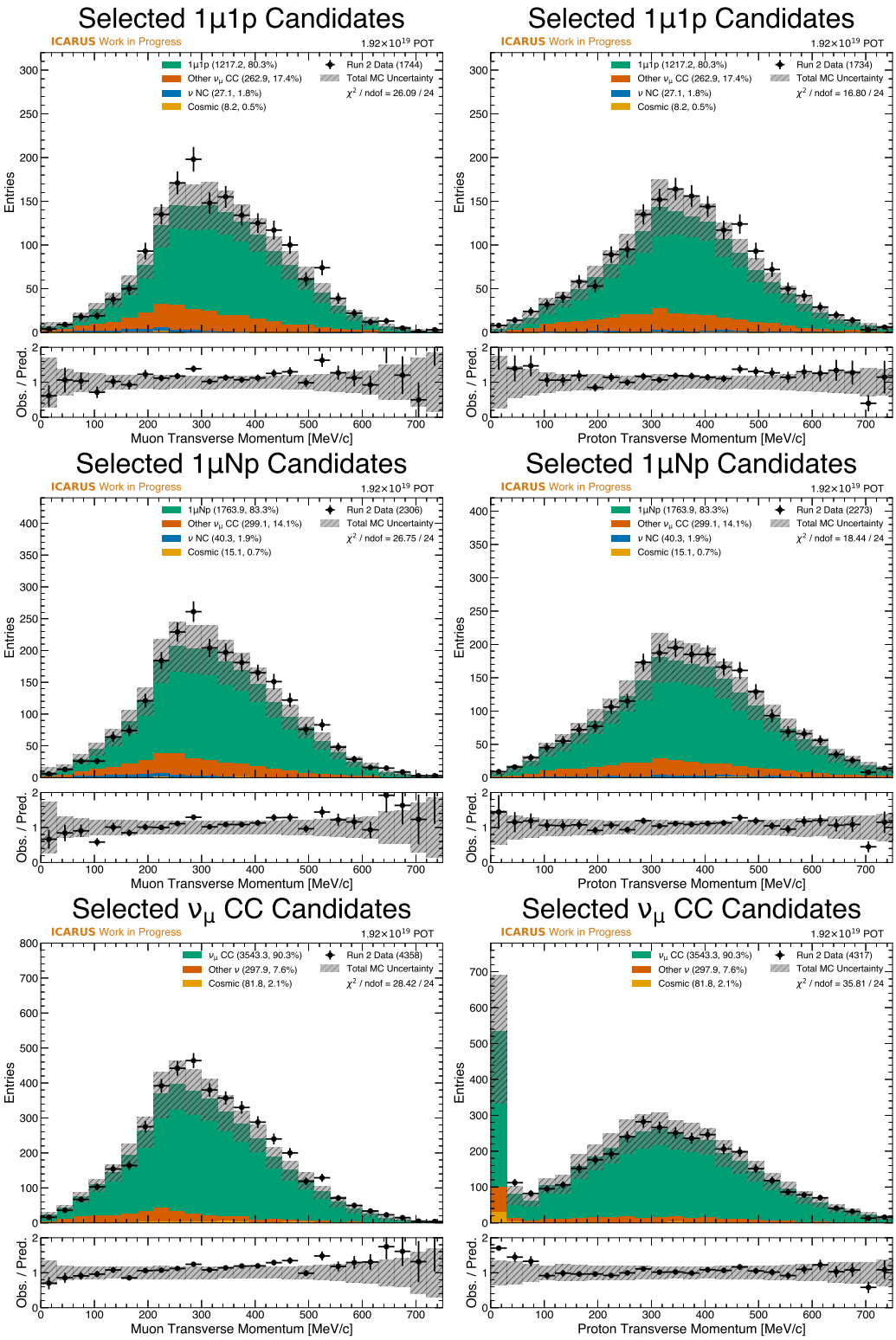


Figure 9.3: Comparison of data and simulation for the transverse momentum of the muon (left) and the most energetic proton (right) for each of the three signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive. The peak in the lowest bin of the proton transverse momentum distribution in the inclusive channel is from interactions that do not have a proton in the final state.

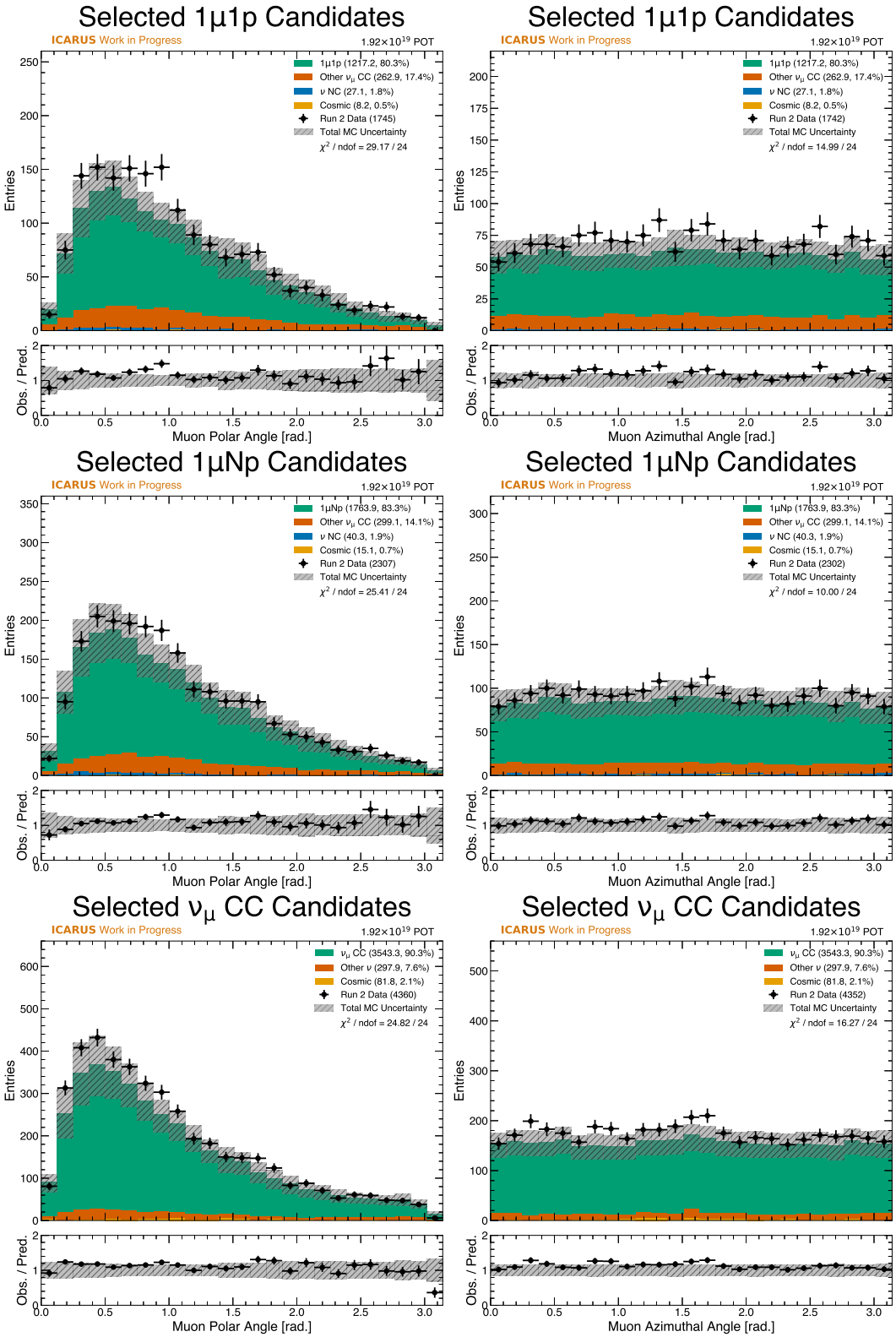


Figure 9.4: Comparison of data and simulation for the polar angle (left) and azimuthal angle (right) of the muon for each of the three signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive.

There are no significant deviations in the muon and proton transverse momentum distributions, the muon polar and azimuthal angle distributions, or the muon-proton opening angle distributions. Many reconstruction inefficiencies, such as signal removal from coherent noise filtering, could potentially impact these distributions, so it is especially important that we see good agreement in these variables.

9.2.2 Comparisons for Kinematic Imbalance Variables

The kinematic imbalance variables δp_T, α_T, and ϕ_T are shown in Figures 9.6 and 9.7. These variables are commonly incorporated into cross section measurements, so it is important for any future ICARUS cross section measurements that the data and simulation do not show major biases in these variables. The discrepancy in these variables seem to be only a normalization offset, with the shape of the distributions matching well. These variables are also sensitive to smearing from poor angular reconstruction, so the agreement in these variables is a good indication that the angular reconstruction is working similarly in data and simulation.

9.2.3 Comparisons for PID Variables

The PID variables are shown in Figure 9.8. These variables, perhaps more than any of the others, are where one would expect to see efficiency differences between data and simulation. If the PID scores are not well modeled, then it is possible that data and simulation may select interactions at different rates. There is largely good agreement in the proton PID score distributions, with only an overall normalization offset manifesting in the last bin. It is clear that the data and simulation are both confident that the leading proton candidate is a proton and to similar degrees.

The muon PID score shows a significant shift of the population from the peak to lower scores in data compared to simulation. Perhaps fortuitously, the cut placed on the muon PID score is quite low at 0.1, so this shift does not appear to affect the selection efficiency. Regardless, this discrepancy must be understood and will be the subject of upcoming investigations. One hypothesis is that the usage of the plane-averaged charge rather than collection-only charge in the space points may be introducing a smearing effect in data that is not present in simulation. The switch

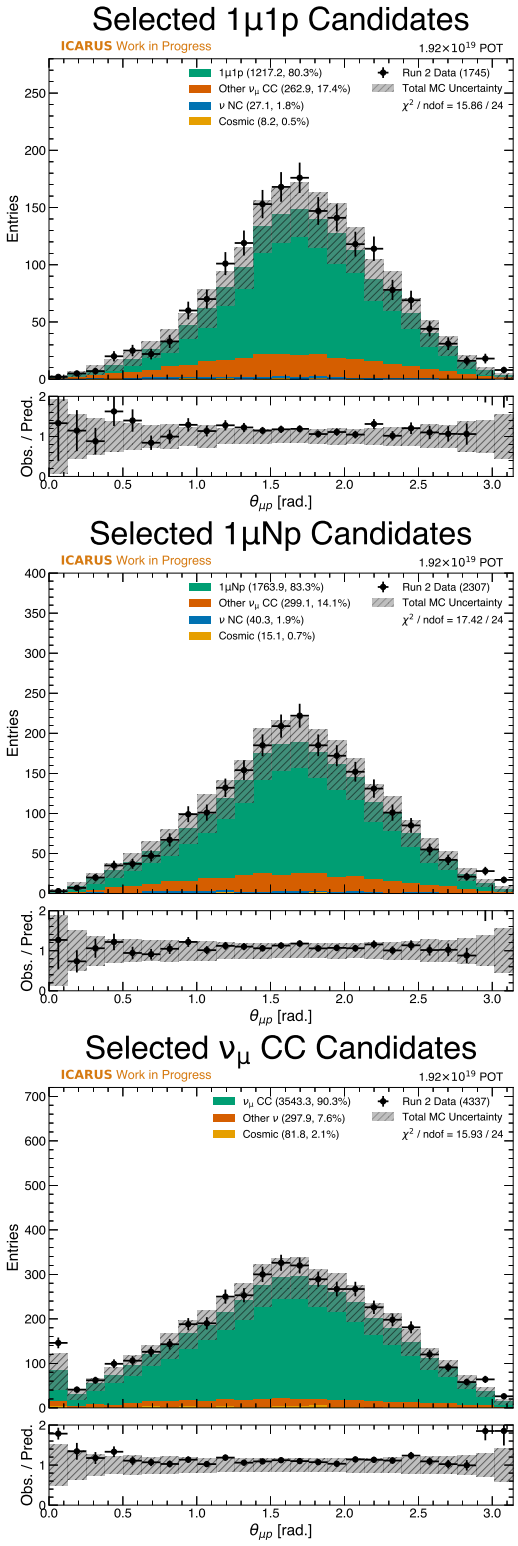


Figure 9.5: Comparison of data and simulation for the the muon-proton opening angle for each of the three signal channels: from top to bottom $1\mu1$p, 1μNp, and ν_μ CC inclusive. The peak in the lowest bin of the proton transverse momentum distribution in the inclusive channel is from interactions that do not have a proton in the final state.

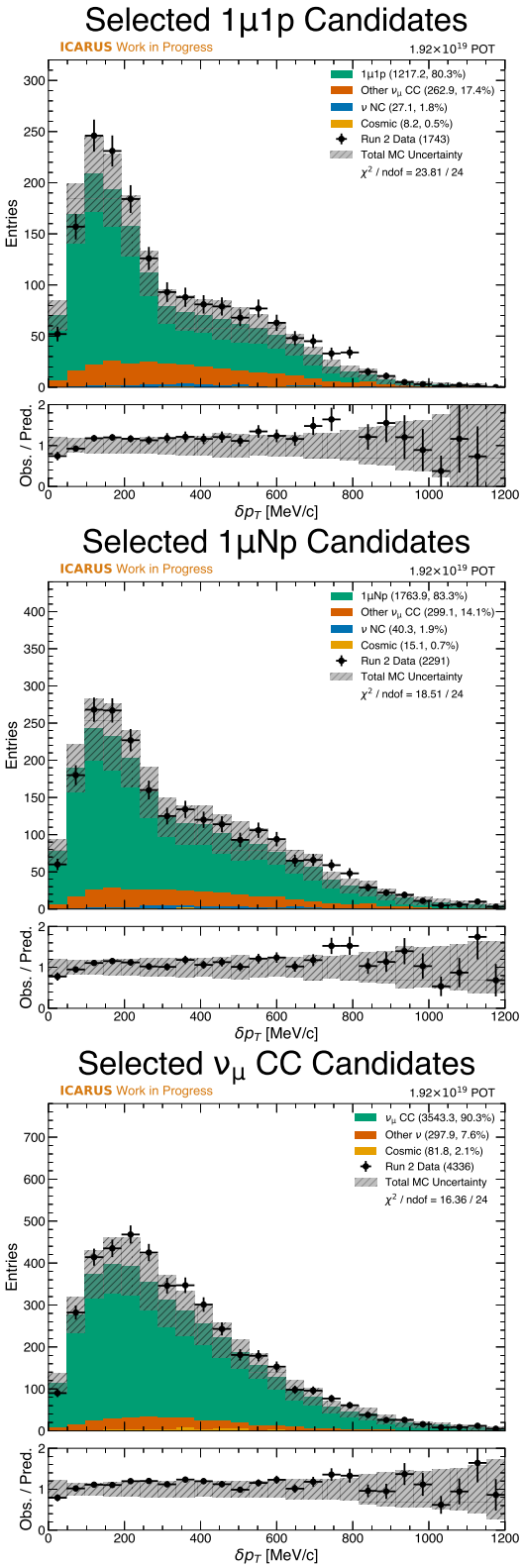


Figure 9.6: Comparison of data and simulation for the transverse momentum of the interaction for each of the three signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive.

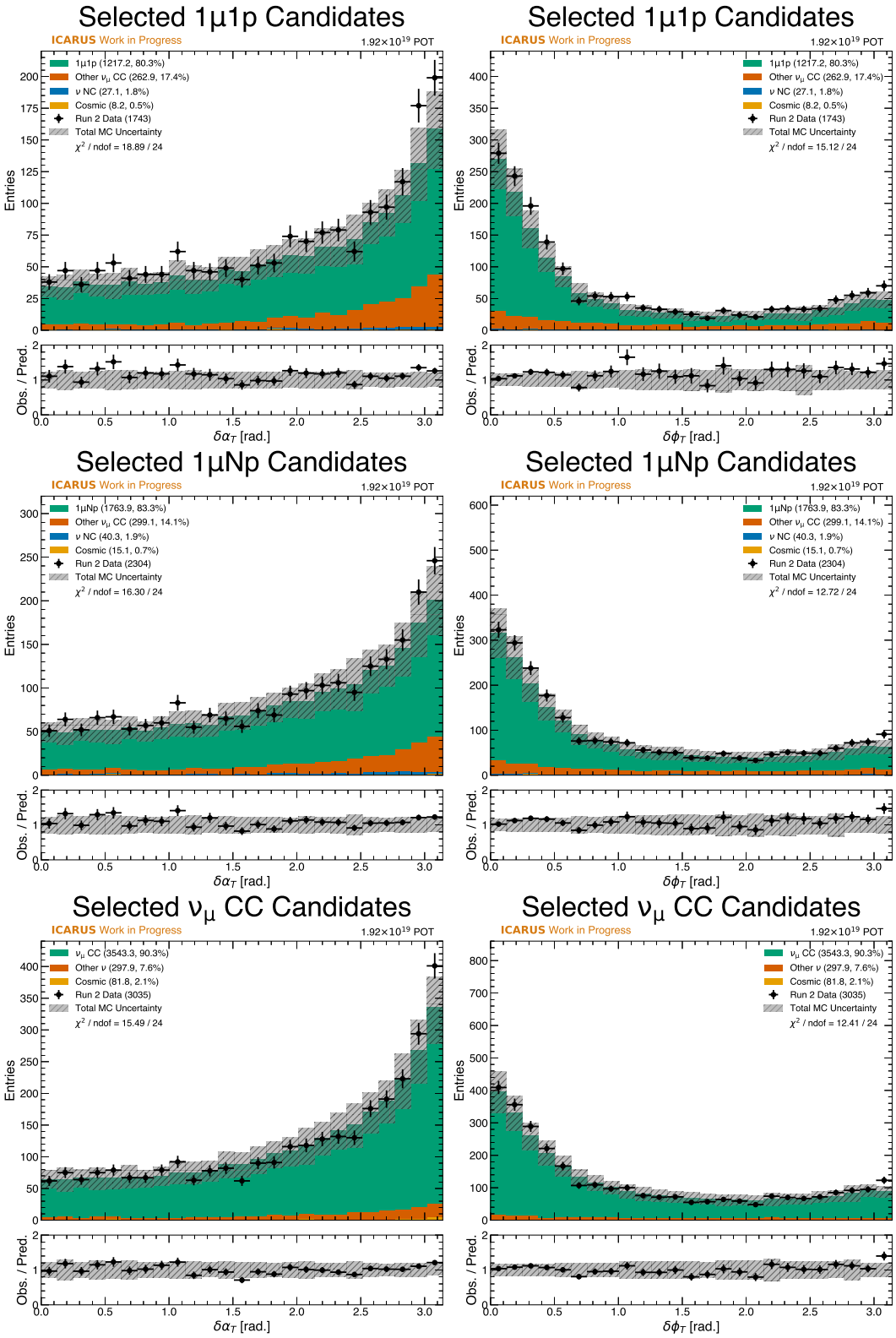


Figure 9.7: Comparison of data and simulation for the kinematic imbalance variables $\delta\alpha_T$ (left) and $\delta\phi_T$ (right) for each of the three signal channels: from top to bottom $1\mu1\mathrm{p}$, $1\mu\mathrm{Np}$, and ν_μ CC inclusive.

to collection-only charge in the space points will require re-training the network and re-running the analysis, but it is a necessary step to understand this discrepancy. Another possible source of this discrepancy is the different value of the electron lifetime used in the Monte Carlo simulation compared to the value observed in data. This will also be the subject of future investigations.

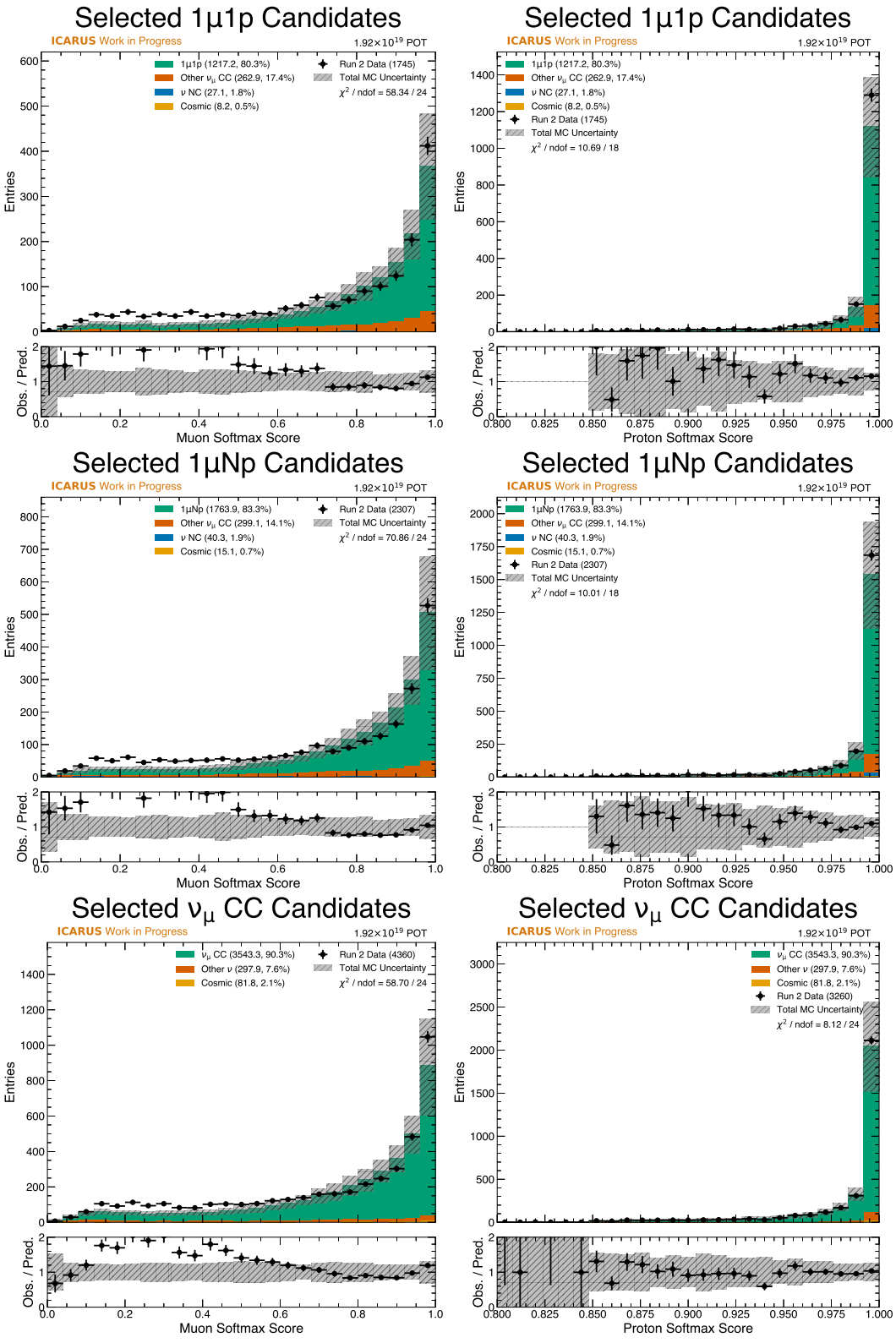


Figure 9.8: Comparison of data and simulation for the softmax PID scores for the muon candidate (left) and the leading proton candidate (right) for each of the three signal channels: from top to bottom $1\mu1\text{p}$, $1\mu\text{Np}$, and ν_μ CC inclusive.

Chapter 10

Conclusions

The analysis presented in this thesis has investigated the performance of the end-to-end machine learning reconstruction chain for the ICARUS T600 detector in three different muon neutrino signal channels: a $1\mu 1p$ channel, a $1\mu \mathrm{N}p$ channel, and an inclusive ν_μ CC channel. The first two exclusive channels are most relevant for the near-term single-detector analysis goals of ICARUS, while the inclusive channel is a more general test of the performance of the machine learning chain and a target for the SBN-wide joint oscillation analysis envisioned in the SBN proposal [14].

The performance of the selections in the three signal channels is quantified in terms of purity and efficiency. This analysis has demonstrated that the selections in the two exclusive channels have achieved a purity of at least 80% and an efficiency of at least 70% with all relevant backgrounds included. The inclusive channel has achieved a purity of 90% and an efficiency of 83%. In all cases, cosmic backgrounds are subdominant to mis-reconstructed neutrino backgrounds and the selections have achieved exceptional purity and efficiency. Given the 80% efficiency for automated reconstruction outlined in the SBN proposal, this analysis has demonstrated that ICARUS meets one of its core deliverables for the joint SBN analysis. Although the full joint SBN analysis includes both contained and exiting events, this analysis can be easily extended to exiting events and is a crucial step towards the ultimate goal. Of particular note on this topic is the fact that this reconstruction has been implemented for SBND as well, meaning that this reconstruction chain has no technical hurdles for use in the joint SBN analysis.

Detector systematics have been characterized in this analysis by producing simulation samples that implement single variations in the underlying parameters controlling the detector simulation. This analysis has shown that the only significant detector systematic among those studied is the systematic associated with the TPC signal response shape on the waveforms. Though this variation is expected to be conservative, more work is required within the ICARUS collaboration to more precisely characterize and mitigate detector systematics in order to reach 2-3% systematic

uncertainty from the detector model assumed in the SBN proposal. The current level of detector systematics is not expected to be a limiting factor in ICARUS-only analyses as the neutrino interaction model dominates the systematic uncertainty in the selection.

Agreement between data and simulation is generally good across variables that probe a wide region of phase space. Two notable exceptions are the overall higher number of candidate interactions selected in data with respect to the simulation and the disagreement observed in the muon PID score variable. The former disagreement is not well-correlated with any particular distribution and is currently thought to be a result of the chosen central value parameters used in the neutrino interaction model. The ICARUS collaboration is exploring a tune of the central value parameters to better match external data. The latter disagreement may be due to the use of plane-averaged charge rather than collection-only charge, thus causing a bias in data with respect to simulation. Moreover, the electron lifetime in the Monte Carlo simulation is different from the value observed in data and may contribute to this discrepancy. This is actively being investigated at the time of this thesis being written. Overall, there are no substantial disagreements between data and simulation that would suggest an inefficiency in the selection that is correlated with a particular variable.

In conclusion, this analysis has demonstrated that the ICARUS T600 detector is capable of achieving high purity and efficiency in the selection of contained muon neutrino interactions using an end-to-end machine learning reconstruction chain. This analysis meets the deliverables outlined in the SBN proposal for the BNB muon neutrino disappearance channel of the joint oscillation analysis, and can be readily extended to the NuMI beam at ICARUS and to the SBND detector. The inclusion of un-contained muon neutrino interactions in the selection to increase statistics is a natural extension of this analysis and is a target for future work. Work is also ongoing to develop analyses with this reconstruction chain targeting electron neutrinos from the BNB and the NuMI beam. This analysis represents a critical step forward in achieving the ultimate goal of the SBN Program: a joint oscillation analysis that combines the precise reconstruction and selection of muon neutrino and electron neutrino interactions from the BNB and NuMI beams at ICARUS and

the BNB at SBND. The ICARUS collaboration is well-positioned to contribute to this goal with the work presented in this thesis.

Bibliography

[1] Ivan Esteban, M.C. Gonzalez-Garcia, Michele Maltoni, et al. The fate of hints: updated global analysis of three-flavor neutrino oscillations. *Journal of High Energy Physics*, 2020:178, 9 2020.

[2] J. A. Formaggio and G. P. Zeller. From ev to eev: Neutrino cross sections across energy scales. *Reviews of Modern Physics*, 84:1307–1341, 9 2012.

[3] J A et al. Precision electroweak measurements on the z resonance. *Physics Reports*, 427:257–454, 5 2006.

[4] Borexino. Comprehensive measurement of pp-chain solar neutrinos. *Nature*, 562:505–510, 10 2018.

[5] Takaaki Kajita. Atmospheric neutrinos and discovery of neutrino oscillations. *Proceedings of the Japan Academy, Series B*, 86:303–321, 2010.

[6] Xin Qian and Jen-Chieh Peng. Physics with reactor neutrinos. *Reports on Progress in Physics*, 82:036201, 3 2019.

[7] Ubaldo Dore, Pier Loverre, and Lucio Ludovici. History of accelerator neutrino beams. *The European Physical Journal H*, 44:271–305, 11 2019.

[8] A. Aguilar, L. B. Auerbach, R. L. Burman, et al. Evidence for neutrino oscillations from the observation of $\bar{\{}\nu_e\}$ appearance in a $\bar{\{}\nu_\mu\}$ beam. *Physical Review D*, 64:112007, 11 2001.

[9] A. A. Aguilar-Arevalo, B. C. Brown, J. M. Conrad, et al. Updated miniboone neutrino oscillation results with increased data and new background studies. *Physical Review D*, 103:052002, 3 2021.

[10] Teppei Katori and Janet M. Conrad. Beyond standard model searches in the miniboone experiment. *Advances in High Energy Physics*, 2015:1–19, 4 2015.

[11] C. Adams, R. An, J. Anthony, et al. Ionization electron signal processing in single phase lartpcs. part i. algorithm description and quantitative evaluation with microboone simulation. *Journal of Instrumentation*, 13:P07006–P07006, 7 2018.

[12] P. Abratenko, A. Aduszkiewicz, F. Akbar, et al. Icarus at the fermilab short-baseline neutrino program: initial operation. *The European Physical Journal C*, 83:467, 6 2023.

[13] A. A. Aguilar-Arevalo, C. E. Anderson, A. O. Bazarko, et al. Neutrino flux prediction at miniboone. *Physical Review D*, 79:072002, 4 2009.

[14] R. Acciarri, C. Adams, R. An, et al. A proposal for a three detector short-baseline neutrino oscillation program in the fermilab booster neutrino beam, 3 2015.

[15] R. Acciarri, C. Adams, R. An, et al. Design and construction of the microboone detector. *Journal of Instrumentation*, 12:P02017–P02017, 2 2017.

[16] S. Amerio, S. Amoruso, M. Antonello, et al. Design, construction and tests of the icarus t600 detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 527:329–410, 7 2004.

[17] J Chadwick. Intensitätsverteilung im magnetischen spectrum der β-strahlen von radium b + c. *Verhandl. Dtsc. Phys. Ges.*, 16:383, 1914.

[18] Wolfgang Pauli. Pauli letter collection: letter to lise meitner. Typed copy, 1930.

[19] F. Reines, C. L. Cowan, F. B. Harrison, et al. Detection of the free antineutrino. *Physical Review*, 117:159–173, 1 1960.

[20] The ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 716:1–29, 7 2012.

[21] M. Goldhaber, L. Grodzins, and A. W. Sunyar. Helicity of neutrinos. *Physical Review*, 109:1015–1017, 2 1958.

[22] M. Aker, A. Beglarian, J. Behrens, et al. Direct neutrino-mass measurement with sub-electronvolt sensitivity. *Nature Physics*, 18:160–166, 2 2022.

[23] D. Akimov, J. B. Albert, P. An, et al. Observation of coherent elastic neutrino-nucleus scattering. *Science*, 357:1123–1126, 9 2017.

[24] P. Vilain, G. Wilquet, R. Beyer, et al. Coherent single charged pion production by neutrinos. *Physics Letters B*, 313:267–275, 8 1993.

[25] B Pontecorvo. Electron and muon neutrinos. *Zh. Eksp. Teor. Fiz.*, 37:1751–1757, 1959.

[26] Raymond Davis. A review of the homestake solar neutrino experiment. *Progress in Particle and Nuclear Physics*, 32:13–32, 1 1994.

[27] Masayuki Nakahata. Super-kamiokande. *Nuclear Physics B - Proceedings Supplements*, 87:125–134, 6 2000.

[28] Q. R. Ahmad, R. C. Allen, T. C. Andersen, et al. Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^8B solar neutrinos at the sudbury neutrino observatory. *Physical Review Letters*, 87:071301, 7 2001.

[29] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. Remarks on the unified model of elementary particles. *Progress of Theoretical Physics*, 28:870–880, 11 1962.

[30] J W F Valle. Neutrino physics overview. *Journal of Physics: Conference Series*, 53:473–505, 11 2006.

[31] X. Qian and P. Vogel. Neutrino mass hierarchy. *Progress in Particle and Nuclear Physics*, 83:1–30, 7 2015.

[32] S Gariazzo, C Giunti, M Laveder, et al. Light sterile neutrinos. *Journal of Physics G: Nuclear and Particle Physics*, 43:033001, 3 2015.

[33] C. Giunti. Short-baseline neutrino oscillations with 3+1 non-unitary mixing. *Physics Letters B*, 795:236–240, 4 2019.

[34] Y. Fukuda, T. Hayakawa, E. Ichihara, et al. Evidence for oscillation of atmospheric neutrinos. *Physical Review Letters*, 81:1562–1567, 8 1998.

[35] A. Letourneau, V. Savu, D. Lhuillier, et al. Origin of the reactor antineutrino anomalies in light of a new summation model with parametrized β^- transitions. *Physical Review Letters*, 130:021801, 1 2023.

[36] J. M. Gaillard. The brookhaven neutrino experiment, 2 1963.

[37] P. Adamson, I. Anghel, A. Aurisano, et al. Combined analysis of ν_μ disappearance and $\nu_\mu \rightarrow \nu_e$ appearance in minos using accelerator and atmospheric neutrinos. *Physical Review Letters*, 112:191801, 5 2014.

[38] Tadayoshi Doke, Kimiaki Masuda, and Eido Shibamura. Estimation of absolute photon yields in liquid argon and xenon for relativistic (1 mev) electrons. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 291:617–620, 6 1990.

[39] W.J. Willis and V. Radeka. Liquid-argon ionization chambers as total-absorption detectors. *Nuclear Instruments and Methods*, 120:221–236, 9 1974.

[40] Jay N. Marx and David R. Nygren. The time projection chamber. *Physics Today*, 31:46–53, 10 1978.

[41] Carlo Rubbia. The liquid-argon time projection chamber: A new concept for neutrino detectors, 5 1977.

[42] Steve Ritz, Hiroaki Aihara, Martin Breidenbach, et al. Building for discovery: Strategic plan for u.s. particle physics in the global context, 5 2014.

[43] S. N. Gninenko. Miniboone anomaly and heavy neutrino decay. *Physical Review Letters*, 103:241802, 12 2009.

[44] S. N. Gninenko. Sterile neutrino decay as a common origin for lsnd/miniboone and t2k excess events. *Physical Review D*, 85:051702, 3 2012.

[45] Peter Ballett, Silvia Pascoli, and Mark Ross-Lonergan. Mev-scale sterile neutrino decays at the fermilab short-baseline neutrino program. *Journal of High Energy Physics*, 2017:102, 4 2017.

[46] André de Gouvêa, Patrick J. Fox, Roni Harnik, et al. Dark tridents at off-axis liquid argon neutrino detectors. *Journal of High Energy Physics*, 2019:1, 1 2019.

[47] Brian Batell, Joshua Berger, and Ahmed Ismail. Probing the higgs portal at the fermilab short-baseline neutrino experiments. *Physical Review D*, 100:115039, 12 2019.

[48] Luca Buonocore, Claudia Frugiuele, and Patrick deNiverville. Hunt for sub-gev dark matter at neutrino facilities: A survey of past and present experiments. *Physical Review D*, 102:035006, 8 2020.

[49] Brian Batell, Joshua Berger, Luc Darmé, and Claudia Frugiuele. Inelastic dark matter at the fermilab short baseline neutrino program. *Physical Review D*, 104:075026, 10 2021.

[50] E.L. Snider and G. Petrillo. Larsoft: toolkit for simulation, reconstruction and analysis of liquid argon tpc neutrino detectors. *Journal of Physics: Conference Series*, 898:042057, 10 2017.

[51] H. Chen, G. De Geronimo, F. Lanni, et al. Front end readout electronics of the microboone experiment. *Physics Procedia*, 37:1287–1294, 2012.

[52] P. Abratenko, R. An, J. Anthony, et al. Measurement of the longitudinal diffusion of ionization electrons in the microboone detector. *Journal of Instrumentation*, 16:P09025, 9 2021.

[53] P. Abratenko, O. Alterkait, D. Andrade Aldana, et al. Measurement of ambient radon progeny decay rates and energy spectra in liquid argon using the microboone detector. *Physical Review D*, 109:052007, 3 2024.

[54] R. Acciarri, C. Adams, R. An, et al. Michel electron reconstruction using cosmic-ray data from the microboone lartpc. *Journal of Instrumentation*, 12:P09014–P09014, 9 2017.

[55] P. Abratenko, R. Acciarri, C. Adams, et al. Determination of muon momentum in the microboone lartpc using an improved model of multiple coulomb scattering. *Journal of Instrumentation*, 12:P10010–P10010, 10 2017.

[56] P. Abratenko, R. An, J. Anthony, et al. First measurement of inclusive electron-neutrino and antineutrino charged current differential cross sections in charged lepton energy on argon in microboone. *Physical Review D*, 105:L051102, 3 2022.

[57] MicroBooNE collaboration, P. Abratenko, O. Alterkait, et al. First double-differential cross section measurement of neutral-current π^0 production in neutrino-argon scattering in the microboone detector. 4 2024.

[58] P. Abratenko, R. An, J. Anthony, et al. Search for an excess of electron neutrino interactions in microboone using multiple final-state topologies. *Physical Review Letters*, 128:241801, 6 2022.

[59] P. Abratenko, R. An, J. Anthony, et al. Search for an anomalous excess of inclusive charged-current ν_e interactions in the microboone experiment using wire-cell reconstruction. *Physical Review D*, 105:112005, 6 2022.

[60] P. Abratenko, R. An, J. Anthony, et al. Search for an anomalous excess of charged-current quasielastic ν_e interactions with the microboone experiment using deep-learning-based reconstruction. *Physical Review D*, 105:112003, 6 2022.

[61] P. Abratenko, R. An, J. Anthony, et al. Search for an anomalous excess of charged-current ν_e interactions without pions in the final state with the microboone experiment. *Physical Review D*, 105:112004, 6 2022.

[62] P. Abratenko, R. An, J. Anthony, et al. Search for neutrino-induced neutral-current δ radiative decay in microboone and a first test of the miniboone low energy excess under a single-photon hypothesis. *Physical Review Letters*, 128:111801, 3 2022.

[63] C Rubbia, M Antonello, P Aprili, et al. Underground operation of the icarus t600 lar-tpc: first results. *Journal of Instrumentation*, 6:P07011–P07011, 7 2011.

[64] M. Antonello, B. Baibussinov, P. Benetti, et al. Search for anomalies in the ν_e appearance from a ν_μ beam. *The European Physical Journal C*, 73:2599, 10 2013.

[65] M. Antonello, B. Baibussinov, P. Benetti, et al. Experimental search for the "lsnd anomaly" with the icarus detector in the cngs neutrino beam. *The European Physical Journal C*, 73:2345, 3 2013.

[66] M. Antonello, P. Aprili, B. Baiboussinov, et al. Measurement of the neutrino velocity with the icarus detector at the cngs beam. *Physics Letters B*, 713:17–22, 6 2012.

[67] M. Antonello, B. Baibussinov, P. Benetti, et al. Precision measurement of the neutrino velocity with the icarus detector in the cngs beam. *Journal of High Energy Physics*, 2012:49, 11 2012.

[68] M. Antonello, B. Baibussinov, P. Benetti, et al. Experimental observation of an extremely high electron lifetime with the icarus-t600 lar-tpc. *Journal of Instrumentation*, 9:P12006–P12006, 12 2014.

[69] L. Bagby, B. Baibussinov, V. Bellini, et al. New read-out electronics for icarus-t600 liquid argon tpc. description, simulation and tests of the new front-end and adc system. *Journal of Instrumentation*, 13:P12007–P12007, 12 2018.

[70] L. Bagby, B. Baibussinov, B. Behera, et al. Overhaul and installation of the icarus-t600 liquid argon tpc electronics for the fnal short baseline neutrino program. *Journal of Instrumentation*, 16:P01037–P01037, 1 2021.

[71] H. de Kerret, Y. Abe, C. Aberle, et al. The double chooz antineutrino detectors. *The European Physical Journal C*, 82:804, 9 2022.

[72] K. Biery, C. Green, J. Kowalkowski, et al. artdaq: An event-building, filtering, and processing framework. *IEEE Transactions on Nuclear Science*, 60:3764–3771, 10 2013.

[73] R. Acciarri, C. Adams, R. An, et al. Noise characterization and filtering in the microboone liquid argon tpc. *Journal of Instrumentation*, 12:P08003–P08003, 8 2017.

[74] S. Agostinelli, J. Allison, K. Amako, et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506:250–303, 7 2003.

[75] Costas Andreopoulos, Christopher Barry, Steve Dytman, et al. The genie neutrino monte carlo generator: Physics and user manual. 10 2015.

[76] K. Sh. Egiyan, N. Dashyan, M. Sargsian, et al. Observation of nuclear scaling in the $a(e, e^{\{\prime}\})$ reaction at $x_b > 1$. *Physical Review C*, 68:014313, 7 2003.

[77] J. Nieves, J. E. Amaro, and M. Valverde. Inclusive quasielastic charged-current neutrino-nucleus reactions. *Physical Review C*, 70:055503, 11 2004.

[78] J. E. Amaro, M. B. Barbaro, J. A. Caballero, et al. Using electron scattering superscaling to predict charge-changing neutrino cross sections in nuclei. *Physical Review C*, 71:015501, 1 2005.

[79] S. Dolan, G. D. Megias, and S. Bolognesi. Implementation of the susav2-meson exchange current 1p1h and 2p2h models in genie and analysis of nuclear effects in t2k measurements. *Physical Review D*, 101:033003, 2 2020.

[80] J. Gonzalez-Rosa, G. D. Megias, J. A. Caballero, and Barbaro M. B. Susav2 model for inelastic neutrino-nucleus scattering. *Physical Review D*, 105:093009, 5 2022.

[81] Ch. Berger and L. M. Sehgal. Lepton mass effects in single pion production by neutrinos. *Physical Review D*, 76:113004, 12 2007.

[82] A. Bodek and U.K. Yang. Modeling deep inelastic cross sections in the few gev region. *Nuclear Physics B - Proceedings Supplements*, 112:70–76, 11 2002.

[83] Júlia Tena-Vidal, Costas Andreopoulos, Christopher Barry, et al. Hadronization model tuning in genie v3. *Physical Review D*, 105:012009, 1 2022.

[84] Julia Tena Vidal. Agky hadronization model tuning in genie 3. page 078. Sissa Medialab, 3 2022.

[85] S. Dytman, Federico Sanchez, M. Sorel, and Luis Alvarez-Ruso. Final state interaction models in neutrino-nucleus cross sections. pages 51–59, 2009.

[86] D. Heck, J. Knapp, J. N. Capdevielle, et al. Corsika: A monte carlo code to simulate extensive air showers, 2 1998.

[87] R Acciarri, C Adams, J Asaadi, et al. A study of electron recombination using highly ionizing particles in the argoneut liquid argon tpc. *Journal of Instrumentation*, 8:P08005–P08005, 8 2013.

[88] P. Abratenko, M. Alrashed, R. An, et al. Measurement of space charge effects in the microboone lartpc using cosmic muons. *Journal of Instrumentation*, 15:P12037–P12037, 12 2020.

[89] W. Shockley. Currents to conductors induced by a moving point charge. *Journal of Applied Physics*, 9:635–636, 10 1938.

[90] S. Ramo. Currents induced by electron motion. *Proceedings of the IRE*, 27:584–585, 9 1939.

[91] MicroBooNE collaboration, R. Acciarri, C. Adams, et al. The pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the microboone detector. 8 2017.

[92] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*, pages 234–241. 2015.

[93] Laura Dominé and Kazuhiro Terao. Scalable deep convolutional neural networks for sparse, locally dense liquid argon time projection chamber data. *Physical Review D*, 102:012005, 7 2020.

[94] Laura Dominé, Pierre Côte de Soux, François Drielsma, et al. Point proposal network for reconstructing 3d particle endpoints with subpixel precision in liquid argon time projection chambers. *Physical Review D*, 104:032004, 8 2021.

[95] François Drielsma, Qing Lin, Pierre Côte de Soux, et al. Clustering of electromagnetic showers and particle interactions with graph neural networks in liquid argon time projection chambers. *Physical Review D*, 104:072004, 10 2021.

[96] P A Zyla, R M Barnett, J Beringer, et al. Review of particle physics. *Progress of Theoretical and Experimental Physics*, 2020, 8 2020.

[97] MicroBooNE collaboration, P. Abratenko, O. Alterkait, et al. Measurement of nuclear effects in neutrino-argon interactions using generalized kinematic imbalance variables with the microboone detector. 10 2023.

[98] D Casadei. Estimating the selection efficiency. *Journal of Instrumentation*, 7:P08021–P08021, 8 2012.

[99] Morris L. Eaton. *Multivariate Statistics*. SPIE, 1 2007.

[100] G. Prior. The harp experiment: first physics results. *Nuclear Physics A*, 752:24–33, 4 2005. Proceedings of the 22nd International Nuclear Physics Conference (Part 2).

[101] P. Abratenko, R. An, J. Anthony, et al. First measurement of energy-dependent inclusive muon neutrino charged-current cross sections on argon with the microboone detector. *Physical Review Letters*, 128:151801, 4 2022.

[102] Michael R. Chernick, Wenceslao González-Manteiga, Rosa M. Crujeiras, and Erniel B. Barrios. *Bootstrap Methods*, pages 169–174. Springer Berlin Heidelberg, 2011.