FERMILAB-THESIS-2023-04

Data-Driven Validation of NOvA's Convolutional Neural Network for Electron (Anti)Neutrino Selection

Anna Maureen Hall

B.S. Physics, University of Florida, 2017

A Dissertation presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Doctor of Philosophy

Department of Physics

University of Virginia
April 2023

Acknowledgements

From when I was little, through school, and all the way to deciding to go to grad school, my parents and brothers have been nothing but encouraging and supportive, and for that I am always grateful. Dad, you always encouraged my love of science, and Mom, you always taught me to be true to myself, both of which got me to where I am today. Allen, you always remind me to relax and not take life too seriously, advice that I need to remember to take more often. Geoff, thanks for always being there as the best little brother I could ask for.

My first year of grad school was a rough one, and I would not have gotten through it without my trusty study and lunch group. Thomas Krahulik, Marybeth Beydler, Miller Eaton, Jacob Higgens, Arthur Conover, and Matt Wampler spent many hours in the library with me that year, and together we dragged each other through passing our classes and (eventually) the qual. Ever since then, hanging out with you guys was always the highlight of my week. Marybeth and Thomas in particular, your friendship through the last few years has meant everything to me.

Once I started research and moved to Fermilab, I have Michael Baird and Andrew Sutton to thank for getting me situated in NOvA, and helping me get started. Michael in particular was always making sure I understood the project I was working on, and why it was important. Thank you for always taking the time to talk through any aspect of NOvA with me and for the many Star Wars discussions at lunch.

In fact, everyone at Fermilab was just the best. I miss being able to just turn around and ask for debugging help, seeing if it was bread o'clock yet, and Cake Thursdays. The Test Beam folks, in particular, all helped me beat my imposter syndrome and realize I am a real part of the experiment with all of our debugging of the various Test Beam issues and long days and nights on shift. Maria, Abhilash, Faiza, and Amy, thank you for making me have a social life throughout my whole time at Fermilab. I miss our game nights and trivia nights.

Throughout my whole time at UVa, from that first year to getting this thesis finally

written, my advisor Craig Group has always been there for me, making sure I had everything I needed. Thank you for making sure I made time every week to study for the qual my first summer, and letting me move to Fermilab. I am grateful for you checking in on me every week, but also giving me the space to just work and figure things out on my own. I knew I could always reach out if I got stuck. Thank you, too, for your understanding and support when I was struggling during the pandemic lock-downs. I couldn't have asked for a better advisor.

I have to give the biggest of Thanks You's to Alex Sargent, my partner through all of this. You have always been my rock and biggest cheerleader. Thanks for moving all over the country for me, and being there for me these last few months especially. I could not have done any of this without your love and support.

Finally, a big thanks to Carol Flumerfelt, my high school physics teacher. I may have driven you crazy with never finishing my homework, but your class was one of my favorites. A few weeks before senior year, when everyone in class was talking about what we going to study in college, you asked me why I wasn't studying physics. It had never occured to me until then, and you making a point to ask me that question set me onto the path that has led me here, writing my PhD thesis. So thank you for seeing the physicist in me and encouraging me to explore that possibility.

Contents

List of Tables

List of Figures

1. Introducing the Neutrino

Meet the neutrino: a nearly massless, electrically neutral particle that tends to generate as many questions as it helps answer. Making up about 0.3% of the universe, it is the most abundant massive particle known to us. It got a reputation as the "ghost particle" in media due to its dislike of being observed. This comes from the fact that it only interacts via the weak force and gravity, with the latter being to an extremely small degree due to the neutrino's tiny mass. Because the weak force is a short-range force and neutrinos are so small, they usually pass straight through everything they come by, rarely interacting - hence the "ghost particle" moniker. In fact, there are trillions of neutrinos passing through your body every second, but chances are, at most one will interact with an atom in your body in your lifetime.

Neutrinos are spin 1/2, left-handed fermions included in the Standard Model of Particle Physics as part of the lepton family. There are three known flavors of neutrinos — electron neutrinos ν_e , muon neutrinos ν_μ , and tau neutrinos ν_τ — named for the lepton associated with that neutrino's weak force interactions. Each flavor also has an associated antineutrino $\bar{\nu}$.

When a neutrino *does* weakly interact with matter, it occurs in the form of either a charged current or neutral current interaction. Charged Current (CC) interactions involve a neutrino coupling to a W boson to produce a lepton matching the flavor of the incident neutrino. Neutral Current (NC) interactions involve the incident neutrino coupling to a Z boson with a neutrino as the outgoing particle.

The original Standard Model included neutrinos as massless particles, so when it was proven that neutrinos do have a mass, they no longer fit perfectly into the Standard Model. This opened the door to the possibility of more neutrino flavors, neutrino interactions with new physics, and a lot of new questions physicists around the world are trying to answer.

1.1 History

About a hundred years ago, there was a lively debate in the physics community about the nature of beta decay. By the mid 1910s, three types of radiation had been identified: alpha, beta, and gamma where the outgoing particles were a helium nucleus, an electron,

and photons respectively. Measurements of alpha and gamma radiation energy distributions were providing narrow peaks that corresponded to the energy and mass differences between the parent nucleus before and after decay as would be expected under conservation of energy. Beta decays were thought to be emitting a lone electron, so it was expected that measurements of the electron energy would show similar results. However, various different experiments showed that to not be the case. Instead, by 1927, it had been definitely concluded that the outgoing beta particles were being created with a continuous spectrum of energies, seemingly violating conservation of energy[1]. Various theories were posed, debated, and tested to explain this phenomena, but to no avail. Neils Bohr even went so far as to propose that conservation of energy was just not valid for nuclear decays[1].

In 1930, in a letter to a conference of nuclear physicists, Wolfgang Pauli suggested a new particle to solve the mystery of the missing energy in beta decay. This particle, which he called a neutron[2], would be electrically neutral, have spin 1/2, and have a mass on the order of an electron but not larger than 0.01 that of a proton. Despite Pauli's insistence throughout the letter that this idea was probably a stretch (he called it a "desperate remedy" and stated that his remedy "may seem almost improbable because one probably would have seen those neutrons, if they exist, for a long time" [2]), the idea was taken up by others, especially Enrico Fermi. In 1934, Fermi proposed his complete theory of beta decay that hinged on the existence of Pauli's particle, now called a neutrino to differentiate it from the heavier neutron recently discovered by James Chadwick [3]. This theory was also the introduction of the weak force that we now know to be the only force that neutrinos are sensitive to, other than gravity. Fermi first tried publishing his theory in *Nature* but was rejected "because it contained speculations too remote from reality to be of interest to the reader." Luckily for us, he was able to instead publish in several European journals, and the neutrino was officially born.

Understandably, it was largely believed that we would never be able to detect a neutrino due to its size and reluctance to interact with matter. However, in 1942, Wang Gangchang proposed a way to detect neutrinos via beta-capture. This method proved to show some evidence of neutrinos, but was not definitive. Instead, the 1956 experiment by Frederick

[1] At the time, this was not as extreme an idea as it might seem to us today. Remember that the late 1920s was in the middle of the rise of quantum mechanics as a paradigm shift in physics, so it wouldn't have been totally outlandish to consider that maybe the law of conservation of energy as we know it, doesn't hold up at all levels of physics.

[2] The neutron that we know and love today had yet to be discovered, so Pauli named his neutrally charged particle a neutron to follow the pattern set by protons and electrons.

Reines and Clyde Cowan, called Project Poltergeist, is credited with the first detection of a neutrino through inverse beta-decay. This is when the study of neutrinos truly exploded.

A few other key moments in the history of neutrinos include

- 1962 - Discovery of the muon neutrino at Brookhaven National Laboratory, which the 1988 Nobel Prize was awarded for
- 1970 - First observation of a solar neutrino
- 1973 - First observation of a neutral current scattering off an electron (which then led to the discovery of the Z boson) in a bubble chamber at CERN
- 1987 - First observation of a supernova neutrino by Kamiokande
- 1995 - Rienes recieved the Nobel Prize for the first detection of a neutrino (Cowan had passed away by then)
- 1998 - First strong evidence for neutrino flavor oscillations meaning that neutrinos have a mass, a fact that does not fit into the Standard Model
- 2000 - Discovery of the tau neutrino by DONUT, a particle theorized to exist in the 1970s
- 2002 - Ray Davis and Masatoshi Koshiba receive the Nobel Prize for "pioneering contributions to astrophysics, in particular for the detection of cosmic neutrinos"
- 2015 - Nobel Prize is given to members of SNO and Super-K for "discovery of neutrino oscillations which shows that neutrinos have mass"

1.2 The Scope of Neutrino Physics

Nowadays, the field of neutrino physics is a rich one with a wide reach in its implications. We are still working to understand the basic properties of neutrinos, from how many different types there are, to the size of their masses, to whether they can be their own anti-particle. Studying neutrino interactions with matter, such as measurements of neutrino cross sections on different target materials, is a big part of the field as well, both for furthering understanding of neutrino physics, but nuclear physics as well.

The discovery that neutrinos oscillate between their flavor states means that they must have a mass to provide the mechanism for the oscillations (See Chap.2). As already mentioned, neutrino masses do not fit into the Standard Model of Particle Physics, making it solid evidence of new physics Beyond the Standard Model (BSM). Experiments, such as the one this thesis is based on, are working to characterise these oscillations as well as using them as an opportunity to look for even more new physics like the possible existence of sterile neutrinos[3]. Oscillation experiments are also capable of probing to what extent neutrinos violate CP symmetry, if they do at all.

Other experiments are looking for neutrino-less double beta decay, which is an extremely rare theorized process that, if observed, would indicate that neutrinos are Majorana particles, ie. are their own antiparticle [4]. If this is found to be true, it could be a clue to solving the matter-antimatter assymetry mystery, as could a non-zero CP violating phase. Still other experiments look for neutrinos to interact in non-standard ways indicating new physics, with many of these experiments, like IceCube, aiming to study neutrino interactions at extremely high energies (on the PeV - EeV scale) where BSM physics is most likely to occur.

Neutrino physics extends beyond studying neutrinos themselves, but also to using neutrinos to study other phenomena. In astrophysics, neutrinos are an excellent messenger of far away events like supernovas and star mergers. Neutrinos produced by these sources not only offer information of the events themselves, but can also point directly back to the location of the source. The full list of astrophysics topics that neutrinos are used to study is too long to include here, but some highlights include using solar neutrinos to study the processes inside the Sun, using high energy cosmic neutrinos to study cosmic background radiation, and looking for elusive "primordial neutrinos" which would have been created early in the lifetime of the universe.

More locally, neutrinos created by radioactive decay inside the Earth can help study the Earth's interior, as can neutrinos that pass through the entire Earth's diameter. Prototypes are also being tested to use the detection of neutrinos coming out of nuclear reactors as near real-time monitors of fuel consumption. These detectors have the potential to eventually be used in nuclear non-proliferation efforts.

Finally, the innovation required to detect neutrinos at all, let alone conduct full analyses

[3]Sterile neutrinos wouldn't interact with any force including the weak force. So if they exist, they would only be able to be detected through oscillation analyses that account for oscillations into all three known flavor types, and then look to see if there is still a deficit of observed neutrinos compared to the source flux of neutrinos. Those "missing" neutrinos could be unseen sterile neutrinos that the source neutrinos oscillated into.

with the data, has provided connections to many different fields of physics and beyond. Particle accelerators are the newest sources of neutrinos used in studies, and the technologies surrounding them are always evolving to provide more intense, higher energy beams. The need for very large detectors with very small backgrounds has led to great advances in the creation of large underground lab facilities, often in still active mines. Detector types are extremely varied, with new ideas and designed always being proposed and explored. On the data analysis side, reconstruction tools are consistently being created and improved, as are simulation methods. The introduction of AI into particle physics research has also vastly increased our ability to use our data to its fullest extent.

Although the scope of neutrino physics is deep and wide, this thesis will focus on the study of neutrino oscillations in the context of NOvA, and in particular, on NOvA's validation and use of a convolutional neural network to identify and classify neutrino interactions.

2. Neutrino Oscillations

Neutrinos interact with the weak force — and are therefore observed — via a set of flavor states. There are three known flavors named for the leptons associated with each neutrino's weak interaction: electron neutrinos ν_e , muon neutrinos ν_μ , and tau neutrinos ν_τ . One of the most interesting things about neutrinos, though, is that they propagate not in these flavor states, but in a set of mass states creatively called ν_1 , ν_2 , and ν_3 . The flavor states do not correspond to the mass states in a one-to-one fashion, but rather as coherent superpositions of the mass states. This gives rise to the phenomena we call neutrino oscillations, where a neutrino can be produced as a certain flavor, travel some distance, and be observed as a different flavor than before.

Hints that neutrinos oscillate between flavors started to show up in the late 1960's with Ray Davis and his Homestake Experiment based off calculations done by John Bahcall [5][6][7]. This experiment placed a tank of cleaning fluid in the Homestake Gold Mine in South Dakota with the intent to observe the rate that electron neutrinos coming from the Sun interacted with the chlorine in the cleaning fluid. Interestingly, the observed rate of solar neutrinos was a fraction of what was expected. This became known as the Solar Neutrino Problem, and was the first experimental indication that neutrinos oscillated between flavor states. Ray Davis and Masatoshi Koshiba would later receive the 2002 Nobel Prize "for pioneering contributions to astrophysics, in particular for the detection of cosmic neutrinos" because of this 24 year-long experiment.

Theories about neutrino oscillations predate the Homestake Mine Experiment, however, with Bruno Pontecorvo in 1957, suggesting neutrino-antineutrino oscillations based off of neutral Kaon mixing [8]. The discovery of the muon neutrino in 1962 led instead to the idea of flavor mixing as was outlined in a paper by Ziro Maki, Masami Nakagawa, and Shoichi Sakata [9]. Pontecorvo then turned this into full flavor oscillations between electron neutrinos and muon neutrinos (tau particles themselves wouldn't be proposed for a couple more years, let alone tau neutrinos) with a proposal paper in 1967 [10]. This paper included a prediction that the observation of solar neutrinos could be about a third of that predicted due to oscillations, ie. an anticipation of the Solar Neutrino Problem. Following the first results of the Ray Davis Homestake experiment, Pontecorvo and Vladimir Gribov published a full formulation for 2 flavor oscillations in 1969 [11].

These formulations were extended to 3 flavor oscillations in vacuum in the 1970's, and is known as the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) Matrix in honor of all four of its conceptualists.

The first strong evidence of neutrino oscillations occured in 1998 by the Super-Kamiokande experiment with their observation of the disappearance of muon neutrinos [12]. The Sudbury Neutrino Observatory (SNO) experiment, however, gets the credit for the first conclusive evidence for neutrino oscillations in 2002, with their observation of electron neutrinos produced in the Sun oscillating into tau neutrinos [13]. These two experiments share the 2015 Nobel Prize in Physics for the "discovery of neutrino oscillations which shows that neutrinos have mass."

2.1 The PMNS Matrix

As already mentioned, the PMNS matrix provides the mathematical framework for neutrino oscillations. It builds upon the basic quantum mechanics principle that any set of eigenstates that are coherent superpositions of another set of eigenstates, can be described as the application of a transformation matrix U onto those eigenstates. For neutrino oscillations, the PMNS matrix performs the role of the transformation matrix. With it, the the flavor states and mass states can be related such that

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \tag{2.1}$$

where for any flavor state ν_α and the mass states ν_i, we can write

$$|\nu_\alpha\rangle = \sum_i U_{\alpha i} |\nu_i\rangle \tag{2.2}$$

The PMNS matrix itself is often parameterized into

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{23} & \sin\theta_{23} \\ 0 & -\sin\theta_{23} & \cos\theta_{23} \end{pmatrix} \begin{pmatrix} \cos\theta_{13} & 0 & \sin\theta_{13} e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -\sin\theta_{13} e^{i\delta_{CP}} & 0 & \cos\theta_{13} \end{pmatrix} \begin{pmatrix} \cos\theta_{12} & \sin\theta_{12} & 0 \\ -\sin\theta_{12} & \cos\theta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.3}$$

where the angles θ_{ij} are the mixing angles between the mass states[1] and δ_{CP} refers to the CP violating phase.

For standard neutrino oscillations, the PMNS matrix is considered unitary and points towards neutrinos having a Dirac mass. In the regime where neutrinos instead have a Majorana mass, the PMNS matrix is no longer unitary, and would have a fourth piece on the end that would account for two imaginary Majorana phases. These phases would not affect neutrino oscillations, however. Other regimes and theories extend the PMNS matrix to include additional parameters to account for additional neutrino flavors and/or mass states, as in the case of the see-saw model of neutrino mass [2], or in the event that light sterile neutrinos exist as some experiments seem to suggest[3]. This thesis will assume the standard oscillations model of only 3 flavor states and 3 mass states, all with Dirac mass assumed.

2.2 Oscillation Probabilities

Following basic Quantum Mechanics and using the PMNS matrix from above, we can set up the following to find the probability of a neutrino oscillating from flavor state α into flavor state β while traveling in a vacuum for time t

$$P(\nu_\alpha \rightarrow \nu_\beta) = |\langle \nu_\beta | \nu(t) \rangle|^2 = \left| \sum_i U_{\beta i} U^*_{\alpha i} e^{-iE_i t} \right|^2 \tag{2.4}$$

Here, we are assuming that neutrinos are travelling as a plane wave to simplify the problem[4]. Additionally, since neutrinos are ultra-relativistic and have extremely small masses, we can

[1] Mixing angles determine the proportions of the mass states in the composition of each flavor state.

[2] This GUT motivated model pairs each of the known light, left-handed neutrinos with a very heavy, right-handed neutrino with a Majorana mass. The heavier the right-handed neutrino, the lighter the left-handed, observable neutrino would be. This would provide an avenue to explain why the neutrino masses are so small compared to quarks and the other leptons [14].

[3] The idea of light sterile neutrino came from some early experiments that saw an excess of electron neutrinos compared to what was expected in a short baseline. An addition fourth neutrino with a small mass that participated in neutrino oscillation, but not in weak force interactions (hence the name "sterile") can explain this anomaly quite well [15]. However, other experiments have either failed to replicate this phenomena or have been unable to definitively contribute their own observed excesses to new physics instead of poorly constrained background event rates and systematic uncertainties.

[4] Truly, we should treat the oscillation probability as a quantum field theory problem complete with the neutrino acting as a wave packet, and do the calculation as such. However, we will arrive at the same general conclusions and results whether we do the proper QFT treatment or the simplified version presented here.

reasonably assume that $E \simeq p$. Thus, we can expand E such that

$$E_i = \sqrt{p^2 + {m_i}^2} \simeq p + \frac{{m_i}^2}{2E} \tag{2.5}$$

These assumptions alone can help simplify the calculation of this probability, but the result can still become quite complex, quite quickly. Instead, when in the experimental context and for illustrative purposes, it is common to reduce the problem down to a 2-flavor system instead of a 3-flavor system. For 2-flavors, the PMNS matrix reduces to

$$\begin{pmatrix} \cos\theta_{12} & \sin\theta_{12} \\ -\sin\theta_{12} & \cos\theta_{12} \end{pmatrix} \tag{2.6}$$

With all these assumptions and simplifications, Eq.2.4 can then be written as

$$\begin{aligned} P(\nu_\alpha \to \nu_\beta) &\approx \sin^2 2\theta \sin^2\left(\frac{\Delta m^2 L}{4E}\right) \\ P(\nu_\alpha \to \nu_\alpha) &\approx 1 - \sin^2 2\theta \sin^2\left(\frac{\Delta m^2 L}{4E}\right) \end{aligned} \tag{2.7}$$

where Δm^2 refers to the difference of the squares of the masses of the two mass states involved (called the mass splitting), and L refers to the length of oscillation, assumed to be approximately equal to ct. Being in a 2-flavor scenario also means that we can define the probability that a neutrino *won't* oscillate into another flavor as one minus the probability that it *will* oscillate, as reflected by the second line of Eq.2.7. Proper treatment of this oscillation probability calculation with the full 3-flavor system comes up with this approximation as the dominant term.

Looking closely at this oscillation probability, we can pull it apart into an amplitude driven by oscillation parameter θ, and an oscillating piece driven by the travel distance L, neutrino energy E, and the mass splitting. E and L are determined by the experimental design, while the mass splitting is determined by nature, making it another oscillation parameter. Proper choices of E and L by an experiment can therefore allow us to measure these two oscillation parameters. However, looking at Eq.2.7, it is clear that the oscillation probabilities are insensitive to the octant of the mixing angles[5] and the sign of the mass splitting.

[5]In the context of neutrino oscillations, "octant" refers to which side of maximal mixing the mixing angle falls on. We know the mixing angle must be within the range of 0 and $\pi/2$ with maximal mixing at $\theta = \pi/4$. The upper (lower) octant then refers to the values of θ that are greater (less) than this maximal mixing angle.

However, subdominant effects/terms seen when looking at the full 3-flavor treatment of oscillations, along with the inclusion of matter effects, could allow us to disentangle these degeneracies. Further discussion of experimental techniques used to measure the various oscillation parameters can be found in Section 2.4.

2.3 Matter Effects

As neutrinos travel through a vacuum, it is relatively straightforward to define the oscillation parameters, as was demonstrated in the previous section. However, when neutrinos travel through matter, the effects of coherent forward scattering on the electrons and nucleons that make up that matter must be taken into account. This effect is known as the MSW (Mikheyev-Smirnov-Wolfenstein) effect and can greatly alter the shape of neutrino oscillation probabilities [16][17].

Typically, these matter effects are treated in a similar manner to that of light refracting in a medium, with a designation of a potential V that drives the effects. In this case, where V is used to account for neutrino interactions in matter, it has contributions both from neutral current interactions of the neutrino colliding with nucleons, and charged current weak interactions between the neutrino and any leptons it encounters. Because neutrinos travelling through ordinary matter typically will only encounter electrons, only the charged current interaction channel is usually included in matter effect formalisation[6]. Additionally, the NC interactions occur at an equal rate for all three flavor states, so that part of the matter effect potential can be subtracted out as a universal contribution for convenience.

For neutrinos in a vacuum, Eq.2.4 was built using a Hamiltonian H_0 representing a freely propagating particle traveling through a vacuum. To include matter effects, the Hamiltonian used should instead be $H = H_0 + V$. The eigenstates and eigenvalues of this Hamiltonian are no longer those from Sec.2.2 as they now need to have a dependence on the matter density and the neutrino energy to account for the addition of V to the Hamiltonian. This means that the effective mixing angles and mass splittings also depend on the matter density and neutrino energy, instead of being independent parameters as they are in the vacuum case.

In this new regime, the impact of the matter effects on oscillation probabilities is greatest

[6]This isn't to say that a neutrino wouldn't ever interact with a muon or a tau, but the the short lifetimes of those particles combined with the extremely small cross sections for such interactions make these types of interaction rare enough to have negligible effect in this context.

when V is such that we can get resonance within the oscillations. Resonance can be achieved when the density of the matter that the neutrinos propagate through is such that there is maximal mixing ($\sin^2 2\theta_m = 1$ where θ_m is the mixing angle in matter, not vacuum). The corresponding resonance condition in the 2-flavor regime is

$$\begin{aligned} \Delta m^2 \cos 2\theta > 0 \quad & for\ neutrinos \\ \Delta m^2 \cos 2\theta < 0 \quad & for\ antineutrinos \end{aligned} \tag{2.8}$$

It shouldn't be surprising that there are different (and opposite) conditions for resonance for neutrinos and antineutrinos since they interact differently in matter. This will lead to different oscillation probabilities in matter as well, regardless of the δ_{CP} phase. It follows that this difference would be maximal when at resonance. Experiments that take advantage of these matter effects, particularly when at or near resonance, could then potentially observe an enhancement of either neutrino or antineutrino oscillations. Whichever is enhanced would then tell us the sign of Δm^2, as long as the matter effects can be separated from any potential CP violation effects.

For a full treatment of the MSW effect and the differences that arise from varying matter densities, see [16][17]. For the context of this thesis, it is enough to understand that an asymmetry in neutrino oscillation rates versus antineutrino oscillation rates would have its roots in both matter effects and the CP violating phase, and a properly designed experiment can use this to probe both δ_{CP} and the neutrino mass hierarchy problem.

2.4 Measurement of Oscillation Parameters

Different combinations of baseline, neutrino energy, and neutrino source in an experiment, leads to sensitivity to different oscillation parameters. Sec.2.3 highlighted how matter effects can help determine the mass hierarchy. This section will discuss how different style experiments have led to our current global set of oscillation parameters, and how current and planned experiments plan to further constrain these values, and probe at the remaining unknowns.

There are two main types of oscillation analyses: disappearance and appearance. They both start with a neutrino source of a single flavor, but disappearance analyses look for where events are "missing" compared to the initial flux after allowing the neutrinos to travel. In other words, disappearance analyses aim to measure $\nu_\alpha \rightarrow \nu_\alpha$ oscillation probabilities.

Appearance analyses instead aim to measure $\nu_\alpha \rightarrow \nu_\beta$ oscillation probabilities. Experiments are designed to be able to run one or both of these analyses types, with the first generations of experiments being disappearance searches.

We can further categorize oscillation experiments by their sources: solar, atmospheric, reactor, and accelerator. The sun outputs a flux of electron neutrinos, and nuclear reactors provide a flux of electron antineutrinos. Accelerators provide sources of muon (anti)neutrinos. Atmospheric neutrinos are created when cosmic rays interact with our atmosphere, creating cascades of particles and decays that result in the production of muon (anti)neutrinos and electron (anti)neutrinos. The ratio of the rate of production of the muon neutrinos to the electron neutrinos is 2:1.

Lastly, each experiment is either considered a short baseline or a long baseline experiment based on the distance between the neutrino source and detector, with solar and atmospheric experiments in the long baseline group. Reactor and accelerator experiments can be either since we can chose both the source and detection locations.

Careful choices of combinations of the above types of oscillation experiments, along with a choice of neutrino energy, allow us design experiments that effectively condense the 3-flavor oscillation probabilities into the 2-flavor version discussed in Sec.2.2 perhaps with just a couple extra terms from the full 3-flavor treatment. For example, low energy solar neutrino experiments and long baseline reactor experiments where the primary oscillation channel observed is $\nu_e(\bar{\nu}_e) \rightarrow \nu_e(\bar{\nu}_e)$, have baselines such that $\Delta m^2_{21} L/E \gtrsim 1$ and $\Delta m^2_{31} L/E \gg 1$. In these experiments, because $\Delta m^2_{31} L/E \gg 1$, the oscillations driven by Δm^2_{31} are averaged out, making the Δm^2_{12}-driven oscillations dominant. Therefore, when we write the oscillation probability in the full 3-flavor regime, we can ignore the terms associated with oscillations driven by Δm^2_{31}. These probabilities look like the following

$$P(\nu_e \rightarrow \nu_e) = P(\bar{\nu}_e \rightarrow \bar{\nu}_e) \simeq sin^4(\theta_{13}) + cos^4(\theta_{13}) \left(1 - sin^2(2\theta_{12}) sin^2\left(\frac{\Delta m^2_{21} L}{4E}\right)\right) \quad (2.9)$$

It also turns out that θ_{13} is much smaller that the other mixing angles, so we could further ignore the terms that are suppressed by $sin^2(\theta_{13})$, leaving just the 2-flavor approximation (Eq. 2.7) behind. θ_{12} and Δm^2_{12} are often referred to as the solar mixing angle and mass splitting because the solar neutrino oscillation observations (at least historically) were dominated by these two parameters.

Atmospheric neutrino and short baseline reactor experiments have $\Delta m^2_{21} L/E \ll 1$, meaning that we can ignore the terms dependent on Δm^2_{21} by setting $\Delta m^2_{21} = 0$. This leaves ν_μ disappearance probabilities reliant on just θ_{23} and Δm^2_{23}, which we label as the atmospheric parameters. Results show that $|\Delta m^2_{sol}| \ll |\Delta m^2_{atm}|$, and that θ_{23} is near maximal mixing.

The atmospheric and solar parameters cover both of the mass splittings, θ_{12}, and θ_{23} and have some sensitivity to θ_{13}. Short baseline reactor $\bar{\nu}_e$ disappearance experiments are able to further constrain θ_{13} rather well. Accelerator experiments, like NOvA, were introduced with the current generation of experiments, but what is left for them to measure?

First of all, current constraints of θ_{23} place it near maximal mixing, but it's currently unclear in which octant that parameter lies, if it's not indeed maximal mixing. This constraint was primarily driven by atmospheric experiments that unfortunately can only get so precise due to our inability to truly choose a baseline and the sizable uncertainties associated with the measurement of the initial neutrino energies and flux (they are both highly model dependent). Accelerator experiments, however, have choice in both the source and detection locations, as well as the ability to choose the neutrino energy range in the neutrino beam. Furthermore, a second detector can be placed near the accelerator source - like NOvA has - to better understand the incoming neutrino flux. This all allows accelerator experiments to perform much more precise measurements of θ_{23} while being more sensitive to the subdominant terms present in the full 3-flavor oscillation probability that do not fall prey to octant degeneracy. This will hopefully allow us to determine the octant of θ_{23}.

When it comes to the mass splittings, we are allowed to choose the convention by which we label the mass eigenstates. We have chosen to label the mass eigenstates that make up the Δm^2_{21} mass splitting such that $m_2 > m_1$ and $\Delta m^2_{21} > 0$, but we cannot also choose where in the ordering m_3 belongs. In other words, we don't know the sign of the Δm^2_{23} mass splitting. As mentioned in Sec.2.3, matter effects can cause the rate of neutrino and antineutrino oscillations to differ based on the sign of the mass splittings. Accelerator neutrino beams that can produce both a neutrino and a separate antineutrino beam while travelling through a large amount of matter (such as the Earth's crust) would be able to compare oscillation rates between the two beams to potentially determine the mass hierarchy. The Normal Hierarchy would be if m_3 has a larger mass than the other two mass eigenstates, and the Inverted Hierarchy would be if the opposite is true.

Finally, there is another oscillation parameter that has been largely ignored thus far in this thesis: δ_{CP}. This phase would cause neutrinos and antineutrinos to oscillate at

different rates if it is not 0. Accelerator experiments are the only way to probe this CP violating phase as it requires an appearance search as well as neutrino from antineutrino separation. Atmospheric experiments can do appearance analyses, but cannot reliably distinguish neutrino events from antineutrino events, leaving accelerator experiments the only champions of δ_{CP}.

There are a couple reasons only an appearance analyses can probe the CP violating phase[7], but I find it easier to explain why disappearance experiments can't probe δ_{CP}. It all comes down to keeping CPT conservation. The following would show the transformations of the appearance oscillation probabilities under CP, T, and CPT transforms:

$$\begin{aligned} P(\nu_\alpha \to \nu_\beta) &\xrightarrow{CP} P(\bar{\nu_\alpha} \to \bar{\nu_\beta}) \\ P(\nu_\alpha \to \nu_\beta) &\xrightarrow{T} P(\nu_\beta \to \nu_\alpha) \\ P(\nu_\alpha \to \nu_\beta) &\xrightarrow{CPT} P(\bar{\nu_\beta} \to \bar{\nu_\alpha}) \end{aligned} \tag{2.10}$$

To instead look at a disappearance probability, we would just swap out β for α in Eq.2.10. It becomes obvious that T conservation cannot then be violated in this case. If T conservation cannot be violated and CPT must always be conserved, then there is no choice but to also have CP be conserved. Therefore, there just is no CP violation in disappearance experiments. This leaves appearance experiments the only choice for exploration, since we know the CP violating phase has to be somewhere in neutrino oscillations according to the PMNS matrix, even if that phase turns out to be 0.

Overall, accelerator experiments, such as NOvA, have the potential to further probe the already constrained oscillation parameters, especially the currently not well constrained θ_{23}, determine the neutrino mass ordering, and provide the first probes of the CP violating phase in the neutrino sector.

[7]The most concrete of these reasons come from just looking at the terms in the full 3-flavor oscillation probability that have δ_{CP} included in them, and looking at whether they are associated with disappearance and/or appearance. Turns out that the way that the math works out for disappearance searches has all the δ_{CP} terms cancel out.

3. The NOvA Experiment

The NOvA (NuMI Off-axis ν_e Appearance) experiment is a long base-line neutrino oscillation experiment based at Fermilab that uses a two detector design. NOvA utilizes the NuMI (Neutrinos at the Main Injector) neutrino beam provided by Fermilab which is composed of mainly muon neutrinos (ν_μ) or antineutrinos ($\bar{\nu_\mu}$) depending on the beam mode. The experiment primarily studies ν_μ ($\bar{\nu_\mu}$) disappearance and ν_e ($\bar{\nu_e}$) appearance to primarily measure θ_{23}, $|\Delta m^2_{32}|$, and δ_{CP} and determine the neutrino mass hierarchy. The NOvA collaboration consists of over 260 experimenters from 49 institutions located across 8 different countries.

3.1 NuMI Beam

To be able to observe and then study neutrino events in our detectors, we must have a high intensity neutrino beam source. NOvA uses Fermilab's NuMI beam which stands for Neutrinos at the Main Injector. This beam is created by accelerating protons to 120 GeV, and steering them into a graphite target in spills lasting 10 μs, to create a beam of hadronic particles. A pair of magnetic horns then focus the beam into a more forward direction while also selecting for an electromagnetic charge sign. These hadronic particles, mostly pions and kaons, are then allowed to decay into primarily (anti)muons and muon (anti)neutrinos. The beam then travels through a hadron absorber followed by an expanse of rock to absorb and remove any hadrons that didn't decay in time and all the muons that were created during the decay process, leaving behind just the neutrinos. This beam finally arrives at our Near Detector before continuing on through the earth's crust to our Far Detector and beyond.

There are two modes that the beam can run in, and they are dictated by the direction of the current in the magnetic focusing horns. Forward Horn Current (FHC) mode selects for positively charged hadrons which produces a primarily ν_μ beam, while Reverse Horn Current (RHC) mode selects for negatively charged hadrons to produce a primarily $\bar{\nu}_\mu$ beam. Both beam modes have some contamination consisting of wrong sign neutrinos coming from wrong-sign hadrons that were not focused away before the decay pipe, and a small number of electron (anti)neutrinos resulting from subdominant decay modes of K^+ and pions, as well as the decay of K^0s and muons. Fig.3.2 shows the simulated flux of

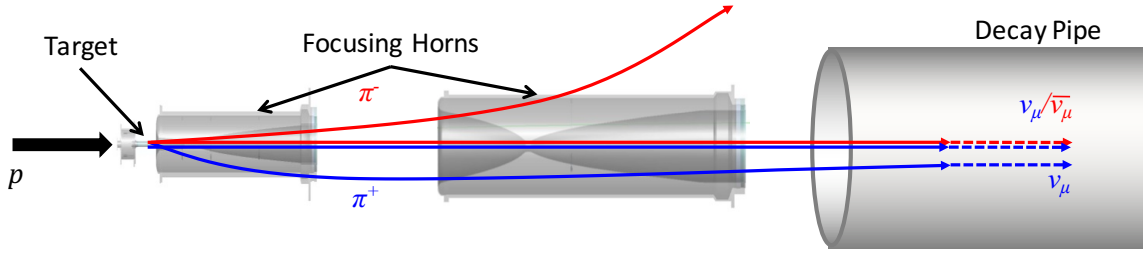


Figure 3.1: Diagram of the NuMI beam with the focusing horn set to the Forward Horn Current (FHC) mode to produce a primarily ν_μ beam. Note this diagram is not to scale.

neutrinos through the Near Detector broken down by neutrino type.

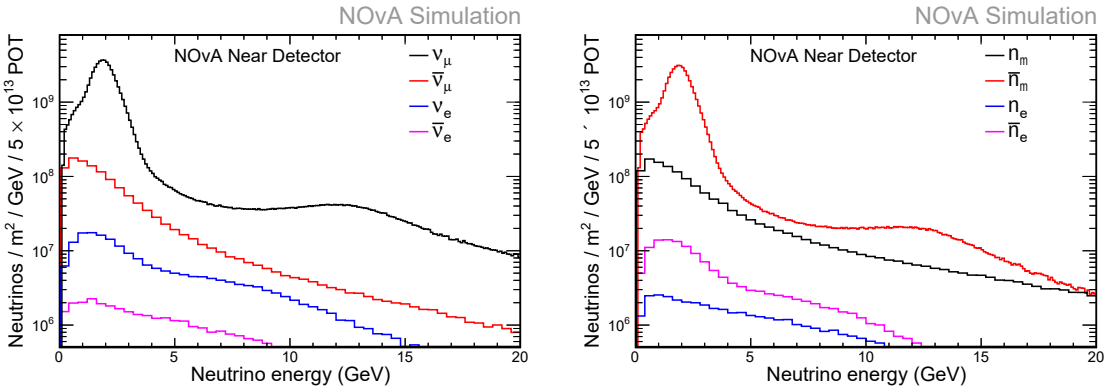
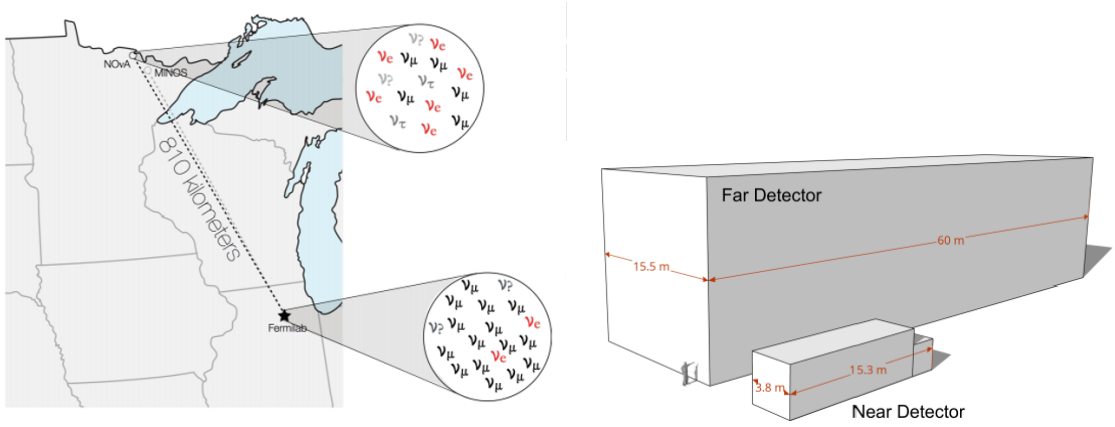


Figure 3.2: Simulated Beam flux at the ND. Left (right) shows the flux in the (anti)neutrino beam mode. PPFX weights have been applied (See 3.5)

A combination of beam simulations and hadron and muon monitors located throughout the beamline help to determine and constrain the flux of beam neutrinos through our detectors, as does data from other experiments. These measures of total flux is dependent on the number of Protons on Target (POT) that are used to create the NuMI beam, making it a convenient metric to report and calculate our beam exposure. Our most recent main analysis used an accrued exposure of 13.6 (12.5) x 10^6 POT for the FHC (RHC) beam mode delivered over 555.3 (321.1) seconds of integrated beam-pulse time. The beam power ranged from 650 - 756 kW during this time. We are planning on doubling the FHC exposure for our next main analysis.

3.2 The NOvA Detectors



(a) (b)

Figure 3.3: (a) Map showing the locations of the ND and FD. (b) To scale comparison of the two detectors.

NOvA consists of two functionally equivalent detectors with the Near Detector (ND) located at Fermilab and the Far Detector (FD) located 810 km away in Ash River, Minnesota. The detectors are situated 14 milliradians off axis from the NuMI neutrino beam provided by Fermilab. This was done so that our detectors would capture neutrinos with a narrow band of energies around 2 GeV. This energy combined with our baseline length should place our experiment at the first maximum for oscillations of ν_μ ($\bar{\nu}_\mu$) into ν_e ($\bar{\nu}_e$) as seen in Fig.3.4.

The 300 ton ND is the smaller of the two detectors and is situated 100 m underground and 1 km from the neutrino beam source. The much larger 14 kT FD sits above ground and is therefore subject to a large amount of cosmogenic activity at a rate of approximately 140 kHz. The ND(FD) is made up of 192 (896) planes which are each in turn comprised of 94(384) cells. As you can see in Fig.3.5, the orientation of the cells in each plane alternate between vertical and horizontal to allow for 3D reconstruction of events. The vertical cells provide a top view while the horizontal cells provide a side view. The ND has an additional 22 planes at the back of the detector to make up a region called the muon catcher. This region has a 4 inch steel plane after each pair of vertical and horizontal planes.

The design of the cells are the same between the two detectors, other than the length - 4.1 (15.6)m for the ND (FD). Each cell has cross-sectional dimensions of 3.6 cm x 5.6 cm and is made of extruded PVC optimized to have a highly reflective inner surface. The cells

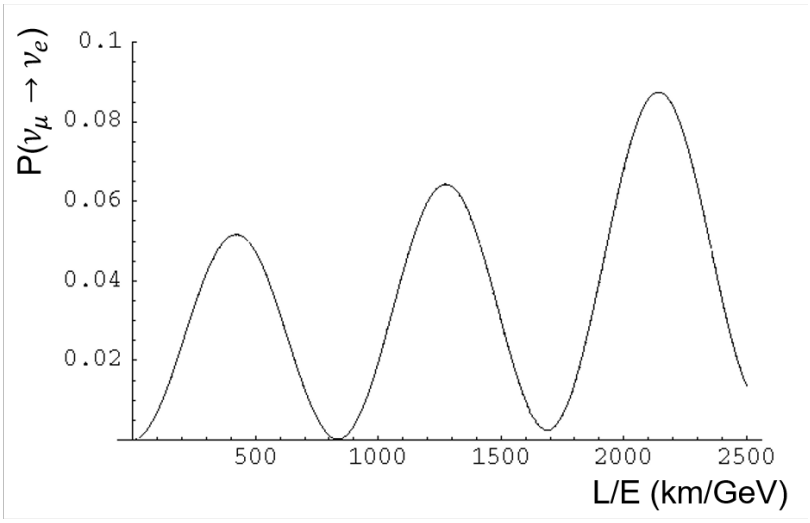
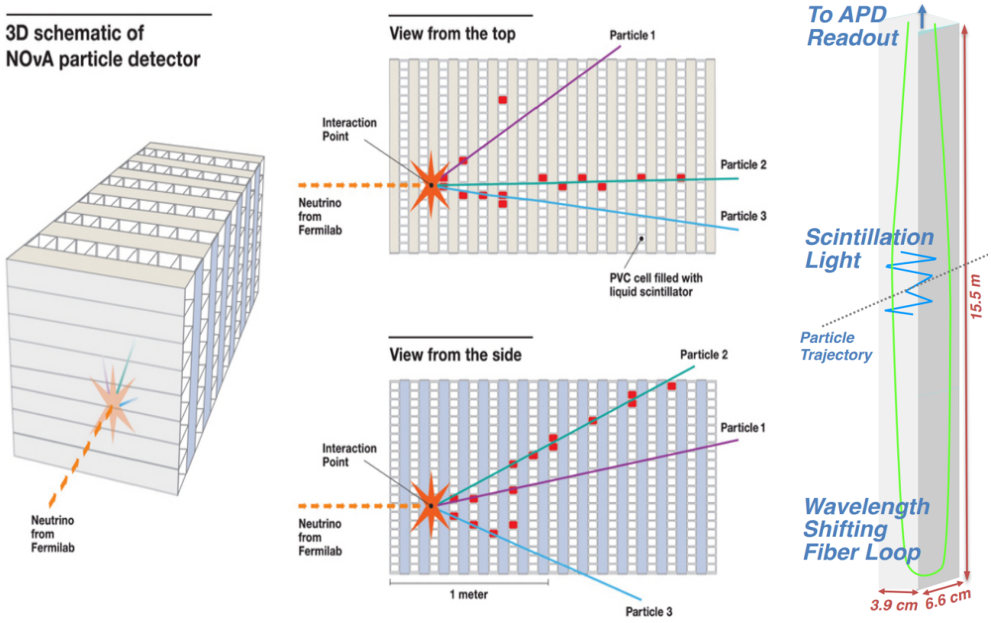


Figure 3.4: Probability of a ν_μ oscillating into a ν_e as a function of L/E. NOvA has a baseline of 810 km and neutrino energy around 2 GeV, placing it at the first maxima.



(a) (b)

Figure 3.5: (a) Schematic of the NOvA detectors showing the alternating planes made of vertical and horizontal planes. These alternating planes create a top view and a side of view of each particle interaction inside the detector. (b) Schematic of a cell icluding the wavelength shifting fiber.

are filled with a scintillating oil which is roughly 95% mineral oil and has been doped with pseudocumene as the primary scintillant. Each cell also has a loop of wavelength shifting fiber to help with the collection of the scintillating light produced by charged particles inside the detector, and to aid that light's propagation to the readout electronics at the end of the cell.

There is also a third, much smaller detector as part of our Test Beam effort. This detector is of the same design as our main detectors, but has been placed in a charged particle beam instead of a neutrino beam. This charged particle beam is accompanied by a set of beamline detectors allowing for particle tagging before they enter the Test Beam detector. We will use data from this set up to better study and constrain several of our systematic uncertainties, especially those related to detector response and calibration. More details on this detector and the Test Beam effort in general can be found in Appendix A.

3.3 DAQ System

3.3.1 Hardware

The cells in each plane are grouped into 32-cell modules with a set of readout electronics for each module. Both ends of the wavelength shifting fibers from each of the cells in a module connect to an avalanche photodiode (APD). These APDs convert the light signal coming out of the cells to an electronic signal via the photoelectric effect. A cooling system is in place to keep the APDs at roughly -15C to reduce thermal noise. A dry gas system is also in place to keep condensation from forming on the cold APDs and surrounding electronics.

Electronic signals are sent from the APDs to a front-end board (FEB) to be digitized and time stamped. The signal is digitized by integrating and shaping the signal by an ADC (analog to digital converter) inside the FEB. Now digitized, the signal is in the form of a pulse-like distribution of ADC[1] over time. Four time samples from this distribution are taken and fit to the pulse shape to determine the exact timing of the signal. These 4 samples are also used to determine if the signal is above the threshold to be considered a "hit." These thresholds are tuned individually for each APD to remove background electronic noise. Hits above threshold will have the ADC values for the 4 time samplings recorded along with

[1]We tend to use ADC to refer to both the FEB component that does the analog to digital conversion as well as the resulting pulse-like signal.

the calculated timestamp. These values are used in later reconstruction to determine the amount of photoelectrons (PE) to associate to each hit[2]. Different FEBs are used in the ND than the FD, but the difference is only in the firmware used. The ND firmware is optimized for a faster readout to accommodate the larger influx of detector activity due to its proximity to the neutrino beam source.

The next link in the DAQ chain is a data concentrating module (DCM). The DCMs are responsible for consolidating the data received from the FEBs into 50μs to 5ms data packets via internal buffers, and then transfer those data packets out to the larger DAQ network. There is one DCM for every 64 FEBs.

It is essential that the clocks on the FEBs and DCMs be in sync with each other and across each detector, and that each detector is in sync with the other so that the times stamps we attach to each signal is meaningful. We accomplish this with a timing system comprised of timing distribution units (TDU) daisy chained together at each detector. These chains store calibrated timing delay values on them that account for travel time of signals moving along the cables connecting all the electronics on the detectors. Each chain has a master TDU that is responsible for driving the time increment counters for each detector and issuing commands to the rest of the chain. Each master TDU is also connected to an external GPS which is considered our global clock. This GPS drives a 10 MHz oscillator, resulting in our over 12,000 FEBs to all be able to sync within 0.1 μs of each other.

The master TDUs are also responsible for receiving and forwarding along a 1 Hz signal and a beam spill signal from Fermilab's accelerator system. The TDUs convert the input signals into NOvA local time so that they can be used to tag coinciding events. The signal is sent to the ND and then forwarded on to the FD, with redundancies in place to ensure that no signal is dropped due to network traffic overload.

3.3.2 Data Handling

Data packets from the DCMs are sent to a buffer node farm of about 200 computers, where packets with similar time stamps are sent to the same node to assists in event building done by the buffer nodes. The buffer nodes create one circular buffer allowing for several different triggering algorithms to each run over the data to make decisions about whether or not each data packet should be kept. Some triggers are simply time dependent, such as the NuMI spill trigger which looks for data that coincides with the beam spill signal. Other

[2]Sec.3.4 goes into detail on the calibration process to convert this PE value to energy.

triggers are data-driven triggers (DDT). These rely on algorithms that look for the basic hallmarks of various event types, or for a certain level of overall detector activity to be met.

Once a trigger flags some data to be saved, an event is built from all the detector activity associated with that trigger and time stamp. These events typically cover 550 μs of time. Events are organized by trigger and time, and finally saved to file, while anything that wasn't triggered on is allowed to disappear. Data files are then copied to physical tapes to be archived for later use.

3.4 Detector Calibration

As noted in Sec.3.3.1, NOvA's data comes off the detector as collections of hits that each have a number of photoelectrons (PE) attached to it. A key step in processing our data is to convert that PE value into the amount energy deposited in the detector in units of GeV. This is done using a two step calibration process. The first step, relative calibration, aims to correct the amount of PE seen per hit so it can be a standard unit throughout the entire detector. Absolute calibration then aims to calculate the conversion rate of those corrected PE into GeV.

The Bethe-Bloch equation (Eq.3.1) describes the stopping power of a material as a particle travels through it. In other words, it describes how much energy a charged particle can be expected to loose as a function of distance travelled through some target matter ($\frac{dE}{dx}$).

$$-\left\langle \frac{dE}{dx} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} ln \frac{2m_e c^2 \beta^2 \gamma^2 W_{max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right] \tag{3.1}$$

where

$K =$ coefficient for dE/dx = $4\pi N_A r_e^2 m_e c^2$

$z =$ charge number of incident particle

$Z =$ atomic number of absorber

$A =$ atomic mass of absorber

$\beta =$ incident particle velocity / c

$\gamma =$ Lorentz factor

$W_{max} =$ maximum possible energy transfer to an electron in a single collision
$= \frac{2m_e c^2 \beta^2 \gamma 2}{1+2\gamma m_e/M+(m_e/M)^2}$

$I =$ mean excitation energy

$\delta(\beta\gamma) =$ density effect correction as a function of $\beta\gamma$

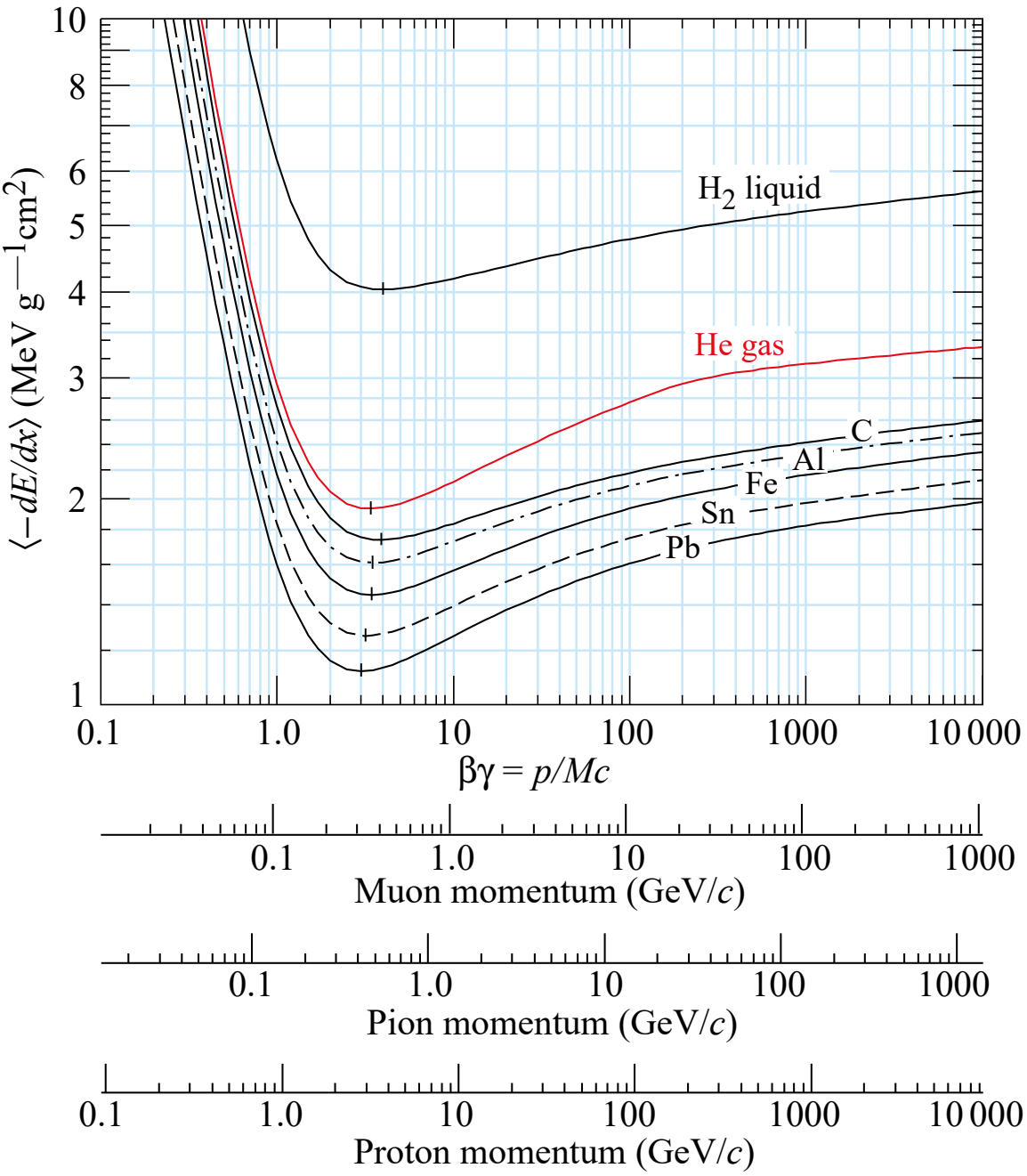


Figure 3.6: Stopping power for several particles in several different materials. Most relevant to calibration in NOvA is the carbon curve paired with the muon momentum. Minimum ionizing point for each curve is marked by a vertical line [18].

Fig.3.6 shows this equation for several different particles incident on several different materials. Note that for muons incident on carbon and H_2, the minimum ionizing region (where it is relatively flat) occurs around the 0.2-2 GeV/c range. The flux of cosmic muons incoming into our mostly hydrocarbon filled detectors predominantly have energies within this range. Therefore, cosmic muons provide a standard candle from which to do our calibration with.

To be able to take advantage of cosmic muons as a standard candle, we need samples of muons in both detectors that are independent of the NuMI beam. For the FD and Test Beam, a sample is created by a minimum bias trigger set to capture detector snapshots at a 10 Hz rate which we refer to as our cosmic trigger. As the ND is underground, the cosmic muon sample is less abundant making a similar trigger inefficient. Instead, a data-driven trigger was designed to save total detector activity that is above a certain threshold to create a cosmic muon sample. Both samples are filtered to remove any activity that coincides with a beam spill timestamp, and are therefore mostly comprised of muonic activity. Similar samples are also simulated for each detector for use in the calibration process.

Once these samples are obtained, reconstruction is run to identify and characterize individual muon tracks (which are typically long and straight) including reconstructing their trajectories. From these tracks, tricell hits are located. These are defined as hits along a muon track that have a hit in the cells on either side of it in the same plane. This means we know through which side the muon entered the cell and that it then exited the cell through the opposite side. As Fig.3.7 shows, this requirement allows us to then use simple geometry to calculate the pathlength of the muon through the cell by dividing the length of the side of the cell the muon doesn't pass through by the cosine of the angle of the muon track. Being able to calculate the pathlength then allows us to set up PE/dx ratios to then compare to dE/dx during the calibration process. This collection of tricell hits, which we call Pre-Calibrated (PC) hits, form the samples that are used for the calibration process.

In low statistics areas, the tricell requirement is loosened to allowing the neighboring cells to be in a different plane, but a similar pathlength calculation is still done. In other cases, like cells on the edge or corner of the detector, where even the looser tricell requirements do not give enough statistics, the assigned pathlength is an average of all possible pathlengths through that cell.

Lastly, the PC hit samples are divided into through-going muons (muons that originate outside the detector and travel all the way through it) and muons that stopped inside the detector. The former is used for the relative calibration and the latter is used for the absolute

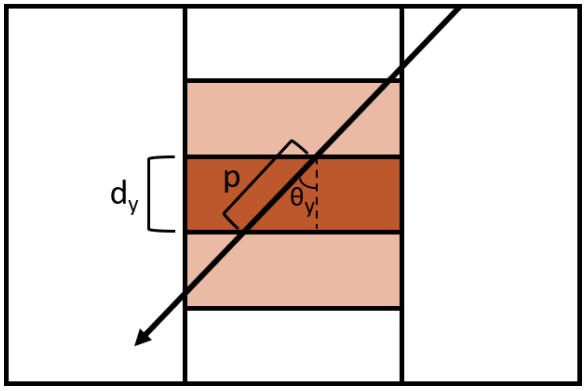


Figure 3.7: Depiction of tricell selection. The muon track goes through all three red cells with the dark red cell being the tricell hit. In this example, the pathlength p of the muon through the cell would be calculated by $d_y/cos(\theta_y)$.

calibration.

3.4.1 Relative Calibration

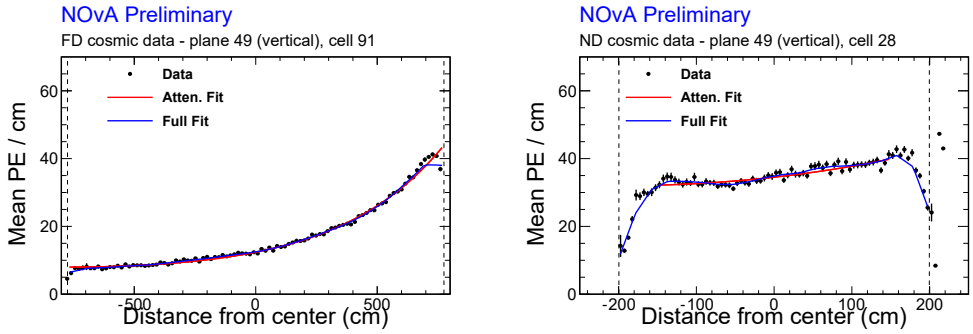
The relative calibration step aims to make the PE unit standard across the whole detector so that a single conversion rate from PE to GeV can be used for the entire detector. There are three main things that are accounted for: self-shielding of the detectors, signal attenuation as light travels along the length of cells, and threshold bias.

Self-shielding refers to the phenomena we observe of the lower sections of the detectors tending to see lower values of PE/cm than the upper sections. This arises from the muons losing energy as they travel through the detector via the ionization of the scintillator, resulting in lower average visible energy depositions the further the muon travels. Because most of the cosmic muons enter from the top of the detectors and travel downwards, the upper sections of the detector will see the higher energy depositions compared to the lower sections.

Signal attenuation can most clearly be seen by organizing the PC hits of through-going muons into profiles, one for each cell in each plane. These profiles are 2D plots of position along cell length (W) by photoelectrons per pathlength (PE/cm) like the examples in Fig.3.8. As can be seen in these example cells, there is a decrease in PE/cm as you move down the cell away from the readout electronics (positive W to negative W). The size of this effect varies according to the size of the detector, with the FD and its 15m long cells seeing by far the most drastic signal attenuation, and the much smaller ND and TB detector having only some attenuation. In addition to the attenuation, there are visible drop offs in these profiles at either end of the cells. This is caused by the physical structure of the cell ends making it hard to properly capture ionizing light in those regions.

For our electronics to register a hit, the PE it sees needs to be above a certain threshold (individually determined for each cell). This means that attenuation effects will not only cause hits to have a lower PE than otherwise expected, but it might cause hits to not register at all. This creates a threshold bias in that hits near threshold require an upwards fluctuation of PE to be seen, especially for hits far from the readout electronics. Furthermore, different cells in the detector also have different hit efficiencies, regardless of where along the cell the hit was, due to physical differences in the cells themselves. This also creates a threshold bias.

Threshold bias and self-shielding are incorporated into the relative calibration in a somewhat roundabout way. Brightness maps for each detector are created based off cosmic muon data. They map out where the hot spots and cool spots of the detectors are, prior to



(a) (b)

Figure 3.8: Examples of single cell profiles of through-going PC hits in the (a) Far Detector and (b) Near Detector. Each profile is fit using Eq.3.3 (red curve) which is then corrected with a LOWESS fit (blue curve) to account for the behavior at the ends of each cell.

any corrections. These brightness maps are used as input to our simulation to weight the PE of simulated hits so that they better reflect the reality of the detectors. Therefore, each hit in simulation will have a true energy and an "observed" energy stored. A threshold and shielding correction factor T is calculated using the simulated through-going PC hit sample and the following:

$$T = \frac{PE}{\lambda} \times \frac{E_{true}}{E_{MIP}} \tag{3.2}$$

where PE is the number of photoelectrons that the simulated readout electronics "see", λ is the number of photoelectrons that the readout would have seen in the absence of statistical fluctuations (making it only dependent on the simulated threshold), and E_{true} is the true energy deposited in the cell for that hit. $E_{MIP} = 1.78 \times pathlength$ and represents what we would expect a minimum ionizing particle (MIP) to deposit based on the Bethe-Bloch equation and our detector composition. The first part of this ratio accounts for threshold effects while the second part accounts for shadowing effects.

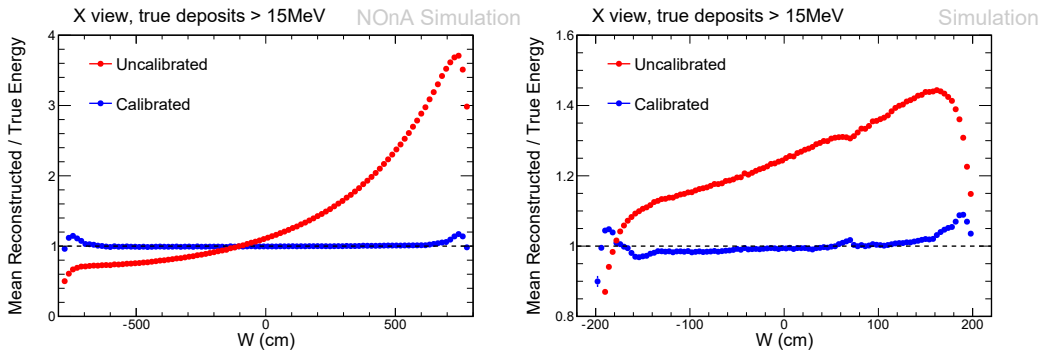
To better account for cell-to-cell differences caused by manufacturing differences in the wavelength shifting fiber rather than differences caused by cell location, each cell is assigned a fiber brightness score which is essentially a measure of how bad a cell's attenuation is. Using 12 bins of fiber brightness, for each combination of fiber brightness and view (X/vertical and Y/horizontal), T is plotted as a function of cell number and W. A fit is then done to a polynomial that includes W to the 11th order, cell number to the fifth order, and cell position in the detector to the 3rd order to get the final correction factors. These corrections are then applied to the attenuation profiles for both the data and the MC.

The final step of the relative calibration is for the attenuation profiles to be fit themselves after the threshold and shielding corrections have been applied. This fit is done according to the functional form

$$y = C + A\left(exp\left(\frac{W}{X}\right) + exp\left(-\frac{L+W}{X}\right)\right) \tag{3.3}$$

where y is the detector response (PE), and L is the cell length, and X is the attenuation length. A and C are constants that, along with X, are allowed to freely float in the fit. This fit is then corrected with a locally weighted scatter plot smoothing (LOWESS) to account for the behavior at the cell ends, and any other major deviations from the attenuation fit. These deviations are thought to be caused by the fiber not sitting inside the cell as expected or from air bubbles inside the cells. A χ^2 is also calculated for each fit. Any cell is considered to still be uncalibrated if the χ^2 is over 0.2. An example of a well calibrated cell can be seen in Fig.3.8 where the red line is the initial fit and the blue is the fit after the LOWESS procedure.

The constants from these fits can be used to correct the PE of hits based on which cell the hit was in and how far down the cell the hit was located. Fig.3.9 shows the before and after of applying the relative calibration to a simulated sample. The flattening of the uncorrected distribution indicates that we now have a unit of energy that we call PECorr that we can say is uniform across the whole detector.



(a) (b)

Figure 3.9: Before and after of relative calibration for the X-views of simulated cosmic muons for the (a) Far Detector and (b) Near Detector. Each curve is the ratio of the reconstructed energy to the true energy. Note the flatness of the calibrated curve indicating a uniform value across the detector.

The results of the fits and the χ^2 values are input into a single csv file for the whole

detector which is then put into the database to be available to be used by any process that calls for calibrated hits.

3.4.2 Absolute Calibration

Once the relative calibration has corrected for cell to cell differences and attenuation effects, the observed detector response should be uniform across the whole detector. Now the absolute energy scale of the detector can be set by finding the conversion rate of a PECorr (corrected PE) into reconstructed GeV. This is done with the absolute calibration phase.

For absolute calibration, the stopping muon sample is used, and PC (tricell) hits are found the same way as before, only with added filters to look at only muons that stop within the detector. These PC hits are then organized into 2D histograms of PECorr/cm by distance to track end, one for each view (x-view from the vertical cells and y-view from the horizontal cells). An example can be seen in Fig.3.10. By plotting the attenuation corrected detector response (PECorr/cm) against the distance to track end, we can see that it resembles the Bethe-Bloch equation. In general, the muons that enter the detectors start out as MIPs with uniform energy deposition dE/dx (the flat region of Fig.3.10), but then start to deposit more energy in a shorter travel distance as the particle starts to slow down. For the NOvA detectors, this starts when the stopping muon is about 100 cm from its track end. Finally, the particle comes to rest at the end of the track and stops depositing any energy. Since we can distinguish the MIP region from the stopping region by when the curve stops being flat, we can isolate the MIP region to use as the basis for the absolute calibration.

PC hits considered for the absolute calibration calculation come from this MIP region and pass several other cuts to ensure that we are not introducing other biases into the sample. Table 3.1 gives the full set of cuts used.

Cut	Purpose
-80 < W < 80	remove edge effects and threshold effects
100 < distance to track end < 200	select for the MIP region
PECorr > 0	ensure hits were calibrated
path length > 0 PE > 0 PECorr/path length < 100 PE/cm	other basic quality cuts

Table 3.1: Absolute Calibration Cuts

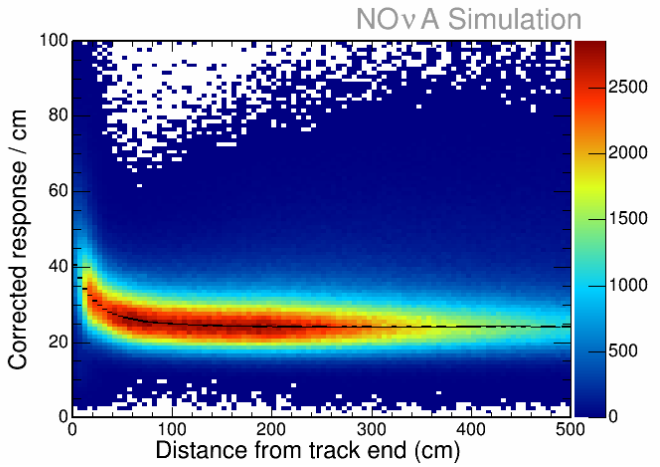


Figure 3.10: PECorr/cm vs. distance to track end for tricell hits from FD MC stopping muons. The black fit curve shows the mean of a fit to the peak of the PECorr/cm at particular distances from the track end.

Once the hits are selected and separated out into the vertical and the horizontal views, three values are extracted called Muon Energy Units (MEU). These are ultimately the numbers that make up the energy scale. From the simulation, we get both a MEU_{truth} and a MEU_{reco} while we get just a MEU_{reco} from the data. MEU_{truth} represents the true simulated energy deposition in the detector and is the mean distribution of MeV/cm from the selected simulation hits. The MeV values are extracted from the truth values embedded in the simulation. MEU_{reco} represents the detector response and is the mean distribution of PECorr/cm for both the data and simulation. The MEU_{reco} values are found separately for each view since the vertical view gets much lower statistics. Then the two values are averaged together. The final calorimetric energy scale is the following

$$\text{Calorimetric energy scale} = \frac{MEU_{truth}}{MEU_{reco}} = \frac{\frac{GeV}{cm}}{\frac{PECorr}{cm}} \tag{3.4}$$

where the same MEU_{truth} value is used for both the data and MC.

3.4.3 Calibration Epochs

Ideally, the entire calibration procedure would only need to be done once and the resulting constants used for the experiment's lifetime. In reality, though, that is impractical. Updates to our detector response model improves simulation which would in turn affect

both the absolute calibration and T correction factors. The detectors themselves are also not stagnant, with maintenance being required and performed regularly which would affect detector response. In addition, we have observed not insignificant detector aging in our detectors (see App.B) which means that the absolute energy scale shouldn't be the same over time to reflect the decrease in detector response. So instead, the calibration process is done separately for each epoch in our data taking.

3.5 Simulation

NOvA's simulation (commonly referred to as our Monte Carlo or MC) is done in several stages, with each stage providing the inputs to the next stage. This includes simulation of the beam, simulation of the neutrino interactions, simulation of those interactions inside the detector, simulation of the detector response, and simulation of the work done by our readout electronics for our actual data.

The simulation of the neutrino beam is done with a Geant4-based [19] simulation of 120 GeV protons incident on a target material matching the actual beam conditions. The resulting hadrons, along with their travels through the focusing horns and eventual decay into neutrinos, are also simulated. All of the target hall materials, including the focusing horns and their support mechanisms, are included in these simulations. There is a mismatch between the resulting flux of neutrinos from this simulated beam and observed rates in our detectors, so the final flux is reweighted using the Package to Predict the Flux (PPFX) [20] originally created by the MINERvA group. This takes the data from several experiments that all use Fermilab's NuMI beam to create weights which are then applied to our simulated flux.

GENIE 3.0.6 [21] provides the base for our simulation of neutrino interactions. While this version of GENIE provides a good place to start, the models included do not necessarily provide great agreement to our observed data, when used in their out-of-the-box form. Instead, a NOvA specific tune of the models included in GENIE is used so the resultant MC better matches our ND data [22]. Generation of cosmic rays, primarily for use on the FD, is done with the Cosmic-ray Shower Library (CRY)[23].

Whether it's particles from the simulated beam, or cosmic rays from CRY, the next link in the MC chain is to simulate the propagation of those particles in our detectors. This is done with GEANT4 which also produces energy depositions in the active material of the detectors. These energy depositions are then converted into scintillation light and

Cherenkov light, and the transport of those photons to the end of the cells is simulated. From there, a parameterized simulation of our readout electronics simulates the conversion chain of photon to photoelectron to ADC pulse to hit [24]. These hits are saved to file as raw data to match our data.

All along this process, key "truth" values are also saved to file like particle types, histories, energies, trajectories, etc., and are propagated through the file processing to be accessible alongside any reconstructed objects and values.

3.6 Event Reconstruction

3.6.1 Clustering

The first step to event reconstruction is to actually identify collections of hits that are near each other in space and time and together make up individual particles moving through the detector or particle interactions occurring inside the detector. Each of the events that come off of the buffer nodes will in truth be made up of many, many particle interactions, and separating out those interactions into individual slices is the only way to do any meaningful physics with the data and simulation. Slicing algorithms will aim to make these slices by finding clusters of hits that are near each other in space and time. The algorithm that NOvA is currently using is called the Time Density Slicer (TDSlicer) and it works by finding local maxima in the density of hit times.

First, hits are looked at in each view separately[3]. Centroids in the time density of the *xzt* and *yzt* sets of hits are found using the algorithm laid out by Rodriguez and Liao [25]. The local density ρ_i of hit i is found by

$$\rho_i = \sum_j exp(-d_{ij}^2/\tau^2) \tag{3.5}$$

where τ is a scaling parameter set to 16(18) ns for the ND(FD) to reflect the timing resolution between hits that each detector achieves. d_{ij} is the Euclidean distance between points and is further defined as

$$d_{ij} = \frac{||dt| - R/c|}{\tau} \tag{3.6}$$

where dt is the difference in time between hits and R is the spatial separation. Once the

[3]Planes with horizontal cells make up the y-view while planes with the vertical cells make up the x-view. The plane number relates to the z-coordinate of each hit.

density of each is found, an isolation score δ_i is also assigned to each hit according to

$$\delta_i = \min_{j|\rho_j > \rho_i} (d_{ij}) \tag{3.7}$$

This finds the the Euclidean distance to the closest hit spatially with a higher density. The centroids are finally then defined as the hits with the highest density and isolation scores.

Next, clustering around these centroids is started by first grouping hits to their nearest centroids whose time stamp is within 10τ. Then Prim's algorithm [26] is used to continue building out the 3D clusters in each view. This algorithm successively adds unclustered hits to whichever cluster has the closest hit to the unclustered hit as long as that distance is not more than $d_{ij} = 8(5)\tau$ for the ND(FD). Hits that are further than this from any clustered hit are simply not clustered.

The final slices are made by taking these 3D clusters from each view and matching them up by comparing average zt values for the clusters in each view to create 4D slices. Slices that have zt values too different between the two clusters that make it up, or clusters that do not find any match are counted as noise slices and ignored by analyses. These slices now form the basis for the rest of event reconstruction.

3.6.2 Vertexes and Prongs

The next step in event reconstruction is to identify each individual particle present in each slice and map their trajectories. This is done by identifying an interaction vertex, forming prongs that start at the vertex, and finally tracking the trajectory of each prong.

Interaction vertices are found by first applying a multi-Hough Transform [27]. This algorithm iteratively identifies straight lines in the slice. Each possible pair of hits in a slice is used to create a straight line running through them parameterized in polar coordinates (ρ, θ). A Gaussian smear vote is calculated for each pair as

$$\text{vote} = \exp\left(\frac{-(\rho - \rho_0)^2}{2\sigma_\rho^2}\right) \exp\left(\frac{-(\theta - \theta_0)^2}{2\sigma_\theta^2}\right) \tag{3.8}$$

where σ_ρ and σ_θ are dependent on the distance between the two hits and the detector resolution. A map is formed of the votes of all the possible hit pairs and the peak is defined as a Hough line. The hits that lie along that line are labelled as such and removed from consideration. The next highest peak forms the next Hough line, and so on until there are

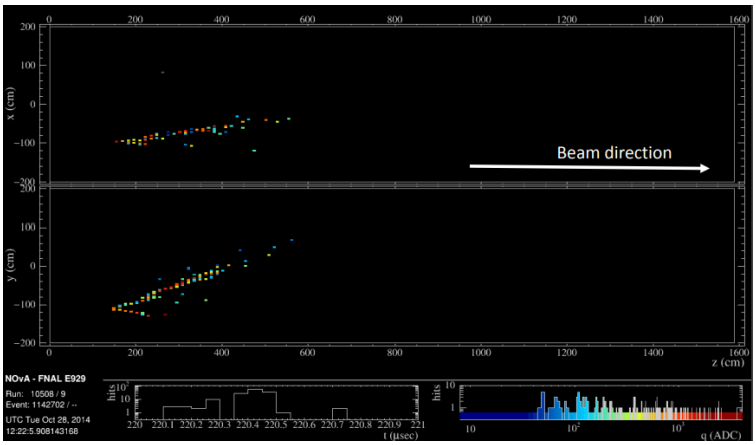
no more peaks above a tuned threshold.

Once all of the Hough lines have been found in a slice, an elastic arm algorithm [28] is used to find the interaction vertex based on where all the Hough lines intersect. The "elastic arms" are straight lines coming from an origin seeded from the Hough lines just found. The parameters defining the arms are found by minimizing an energy function based on the distance between hits in the slice and the arms and how well each hit associates with the fitted arm lines. The origin associated with the set of arms that best minimizes the energy function becomes the vertex of the interaction.

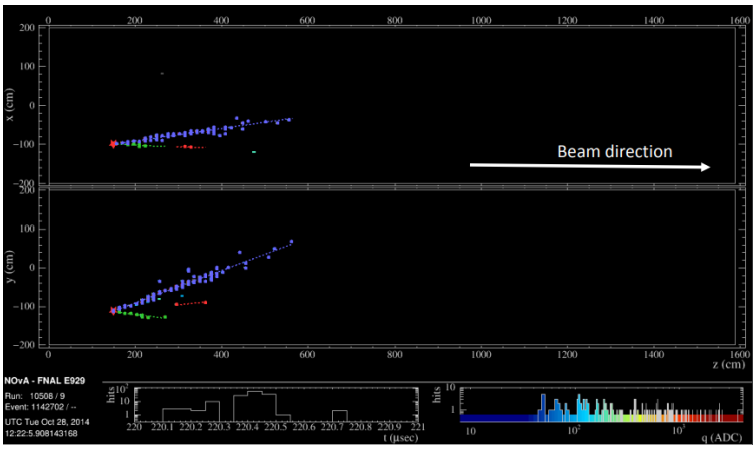
Once a vertex has been found, prongs representing individual particles can be found. Prongs can be identified as clusters of hits with a start point and a direction and are found using a possibilistic fuzzy-k means algorithm [29][30]. the "fuzzy" descriptor here refers to the algorithm's ability to allow hits to belong to multiple clusters. Here, we define "possibilistic" to mean that not every hit needs to belong to a cluster, allowing for outliers to be treated as noise. Like TDSlicer, this clustering is again initially done separately for each view, with view matching based on cluster characteristics done at the end of the process.

Using the found vertex as the origin, hits in each view within the slice have their 1D angle relative to the z-direction through the detector, mapped out with the hit's distance to the vertex determining the uncertainty. Clusters are built from centers of density within this 1D angle distribution. We start with a single prong coming from the densest angle. A membership factor is calculated for each hit to that prong. This factor is based on the hit's angular distance to the prong and the current number of prong centers. Then another prong center is added based on the next densest angle, and the membership factors for each hit in each prong are updated/calculated. This process is repeated until all hits have at least a 1% membership in a prong. Then, prongs that have a large degree of overlapping member hits are merged into a single prong and prongs with a large spatial gap, indicating two colinear particles, are split into 2 prongs.

The matching of the 2D prongs for each view to the other view is done based on the prong's deposited energy profile vs. the z-position of the hits in the prong. Prongs from either view that aren't matched with a prong from the other view are not thrown away, but rather kept as a 2D prong. Fig.3.11 shows the results of reconstructing the vertex and prongs of a sample ν_μ event.



(a)



(b)

Figure 3.11: Event Display of a ν_μ event with (a) showing the hits in the event with the colors of hits scaling with energy deposited. (b) shows the individual prongs found by the fuzz-k algorithm and the star indicates the reconstructed vertex location.

3.6.3 Tracks

Particles like electrons or hadrons can create showers as they travel through the detectors, and the prong making process just described does well at describing them. Other particles like muons, protons, and pions do not create shower-like structures, but rather create narrow tracks potentially with some Coulomb scattering. It is possible to fully define these particles' trajectories beyond just the particle starting point and direction, and there are two main algorithms used on NOvA to do that.

Break Point Fitter (BPF) uses the fuzzy-k 3D prongs and vertex to walk down a track

in segments, allowing for coulomb scattering at appropriate intervals. At each of these inserted scattering planes, the amount of coulomb scattering is determined by the particle mass and energy deposition rate. Starting at the vertex of the interaction, each segment walking outwards to the end of the track is added to the final track, resulting in a fully reconstructed 3D track. For every fuzzy-k prong, this process is repeated using each of muon, proton, and pion particle mass assumptions in the Coulomb scattering calculation.

Kalman Tracker is also used to reconstruct particle tracks. This algorithm, however, doesn't require prongs, and instead uses the hit clusters from TDSlicer directly. Done initially in each view separately, tracks are found by seeds created from pairs of hits less than 4 planes apart. A modified Kalman filter [31] is then used to project that seed forward, adding nearby hits to the track based on the current track position and angle. The track characteristics are updated to include information from the new hits. This is iterated until no more hits can be added. View matching is then done to turn the 2D tracks from each view into a single 3D track where possible, by looking for similar z-coordinates. This algorithm in NOvA was optimized for the longer muon tracks typical of ν_μ CC events.

There is also a tracking algorithm that was created and optimized for use on our cosmic muon data called cosmic tracker. This tracker makes the assumption that muons travelling through the detector move strictly in a straight line. This then allows us to perform very quick fits of a straight line to hits in a slice that deposited energy. This works well for the cosmic data that is overwhelmingly made up of cosmic muon tracks where no vertex or prong finding would be necessary. Trajectory information like position and direction can easily be taken from the fitted straight line. However, as already noted, the real paths of muons travelling through the detector are not straight. A second cosmic muon tracker called window tracker is also often used to account for the curving of muon tracks. This algorithm will look at the hits in a slice in segments that are a tuned number of planes wide. A straight line is fit to the hits in the first segment of the track, creating an initial track. Then the window moves down by a plane, adding any hits consistent with the current track. A new straight line fit is done to the hits in the current window before the window moves over again. This process is repeated until all the planes with hits in the slice have been evaluated. This algorithm results in tracks that will allow for changes in direction, but is still relatively quick to run over the large cosmic muon sample.

4. NOvA 3 Flavor Analysis

NOvA's main analysis is its 3-flavor neutrino oscillation analysis that aims to take advantage of the experiment's two detector design to measure oscillation parameters such as θ_{23} and Δm^2_{32} with a ν_μ disappearance search, as well as probe the CP violating phase δ_{CP} and determine the neutrino mass hierarchy with a ν_e appearance search. This is all done by first using the NuMI beam data at the ND to inform our simulated sample at the ND. Then that simulated sample is extrapolated to the FD and fit to the FD neutrino data to get our final results. This method allows us to greatly reduce various systematic uncertainties that are correlated between the detectors, such as flux and cross-section uncertainties. This chapter will go through each of these steps in detail, and discuss NOvA's most recent analysis results.

4.1 Event Selection

Fig.4.1 shows the three main types of neutrino interactions seen in our detectors. The top panel shows a typical ν_μ event characterized by its long straight muon track and accompanying hadronic activity (here, a proton). The middle panel shows a typical ν_e event that again has a hadronic component, but instead of a muon track, has an electromagnetic (EM) shower as its defining characteristic. Finally, the bottom panel shows a Neutral Current (NC) event characterized by heavier particles creating shorter but higher dE/dx tracks and showers. Notably, because in a NC interaction, the outgoing lepton is a neutrino, we cannot distinguish the flavor of the incident neutrino. Because of this, at least in our main 3-flavor analysis, we consider NC events as background events while ν_μ ($\bar{\nu}_\mu$) CC and ν_e ($\bar{\nu}_e$) CC events make up our signal.

Data must go through several rounds of cuts[1] before they can be considered a candidate neutrino event. These cuts can be categorized as controlling for data and reconstruction quality, removing background events (primarily from cosmic ray activity), and containment cuts to ensure good energy resolution. Events that pass all of these cuts are then passed to our Particle Identifier (PID) for the final event identification and selection.

[1]Here and throughout, we use cuts to mean the criteria used to make selections in the same way that casting directors will make cuts at auditions.

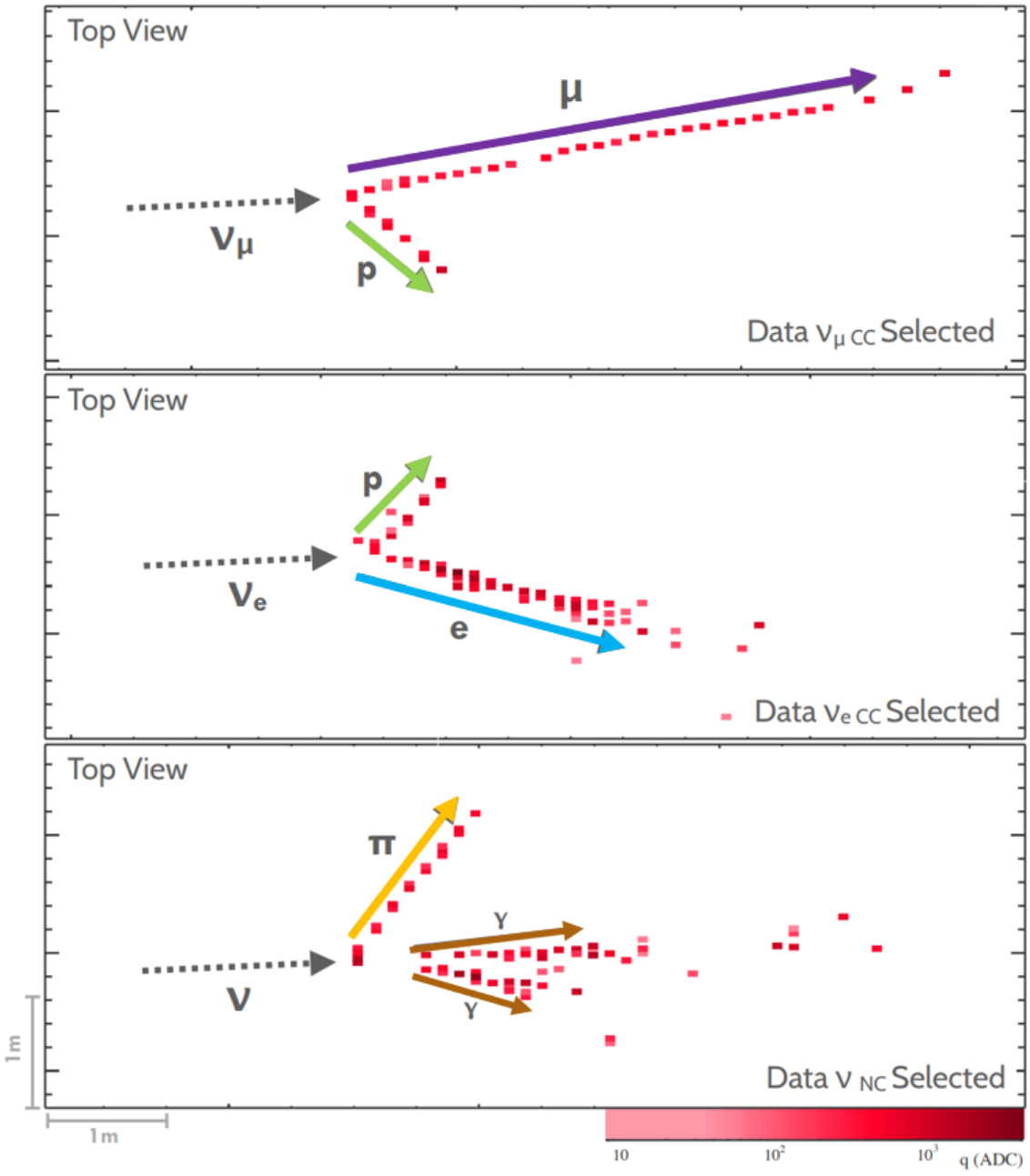


Figure 4.1: Examples of typical neutrino interactions as seen by the NOvA detectors. Top shows a typical ν_μ event, middle shows a typical ν_e event, and bottom shows a typical Neutral Current event. Hit color scales with energy deposition.

4.1.1 Data Quality

These cuts are mostly applied not on an event by event basis, but rather on a beam spill by beam spill basis, or by data run. These cuts remove data from consideration when we know that either the quality of the neutrino beam was not within established parameters, or we know that the detector was not performing at an acceptable level. The beam quality cuts include a cut on the beam width, beam position, focusing horn current, spill POT, and beam intensity. Only allowing events from beam within these parameters helps ensure that we are using a consistent source of neutrinos. On the detector side of things, we remove data that corresponds to instances where parts of the detector were not online when they should have been, the detector was not in sync with the global timing system, or when some or all of the detector is noisy above our usual accepted threshold (for example, when the APDs are in their warm state instead of their cooled state). Additionally, if any of the monitoring for any of the above parameters is missing for a spill or run, that spill or run will also be cut as we cannot know if the quality of the data was good during that time.

4.1.2 Event Quality

The next round of cuts are quality checks on the event level. The first cut here is a timing cut that selects events that occurred within just the beam spill window. Reminder that events are saved in 550 μs packets, and we know that the beam spill occurs during the 217 - 229 μs window in the middle. On its own, a timing cut to select out just this beam spill window is able to reduce our cosmic ray background at the FD from $O(70)$ cosmics per event to just 1 to 2 per beam spill window. The rest of the event quality cuts ensure that the event was reasonably reconstructed and that it meets some basic criteria to be considered either a ν_μ or ν_e event.

For a ν_μ candidate event, the following requirements need to be met

- There is at least 20 hits in the event
- The event spans at least 5 consecutive planes
- There is at least one reconstructed 3D Kalman track
- There is at least one cosmic track (meaning a track found using the cosmic tracker algorithm, not a track coming from a cosmic background event)

- The event's reconstructed energy is greater than 0

There is an additional cut on the event's reconstructed energy to be below 5 GeV as events above that have been found to not contribute to our oscillation results and are often harder to ensure containment on (ie. they are less likely to fit inside the detector).

The ν_e requirements are similar in nature, but are optimized for looking for an EM shower instead of a muon track. They are as follows

- There must be less than 8 hits per plane to avoid APD flashers[2]
- A valid vertex was found
- At least one fuzzy-k prong was reconstructed
- There is between 30 and 150 hits in the event slices in the FD or between 20-200 hits in the ND
- The longest reconstructed prong is between 100 cm and 500 cm in length

Like the ν_μ event selection, we also limit the energy range of interest for the ν_e selection to be between 0 and 4.5 GeV for the ND and between 0.5 and 4 GeV for the FD.

4.1.3 Containment

Our analysis relies on our ability to reconstruct the interacting neutrino's energy, and that can only be done by looking at the energies deposited by the outgoing particles. To ensure good resolution of the neutrino energy, we require that the entirety of a candidate neutrino event be "contained" inside the detector so that there is minimal energy escaping the detector. These containment cuts not only look to see if the event ends inside the detector, but also that it started inside the detector to help remove cosmic and rock muons that would have originated outside the detector.

These cuts are found using a signal to background ratio test, energy resolutions, and the fraction of escaping energy to balance signal selection efficiency with good energy resolution and background rejection. The ND cuts are also able to use Data/MC comparisons to further

[2] APD flashers refer to when a single APD gets over saturated, often by a high energy muon, and causes the electrical ground in the FEB it's connected to, to drop. This momentary lowering of the hit threshold then creates hits in the surrounding APDs that share that FEB, looking like hits in several cells in several planes at once. This can then look like an EM shower to our reconstruction algorithms, but will have a much higher hit density than a real EM shower, hence this cut on hits per plane.

create and validate these cuts. Our FD data is blinded in the beam spill window until "box opening" when we look at the results of our analysis, so no Data/MC comparisons can be made for tuning the FD containment cut in the same way as they are for the ND.

In the ND, the main values considered in the containment cuts of either ν_μ or ν_e events are the coordinates of the start and stop points of every prong in the event as well as every Kalman track in the ν_μ events. Here, one Kalman track is allowed to be inside the Muon Catcher, but no other part of the ν_μ events, or any part of the ν_e events, are allowed to enter it. Additionally, for the ν_μ Kalman tracks, the end point is projected forward based on the track trajectory, and the number of cells it passes through until it reaches the detector edge is counted. A limit is placed on this number. In the same vein, the ν_e events see a cut on the number of planes between the vertex of the event slice and the front of the detector to help remove events that originated outside the detector, like rock muons.

The FD cuts rely on calculations of the closest distance to any part of the event slice to each detector edge, with a much harsher cut applied to the distance to the top face where we expect an excess of cosmic muons. The ν_μ events additionally have the cut on the number of projected cells between any Kalman track end and the edge of the detector, mirroring the ND cut, while the ν_e events also retain their cut on the number of planes between the vertex and front of the detector.

4.1.4 Background Rejection

There are three main backgrounds that we want to remove from our final sample: cosmic ray muons, rock muons, and NC events. We use a convolutional neural network (CNN) for our PID (discussed in detail in Sec.4.2) and it is able to distinguish between the three neutrino event types discussed at the beginning of this section, as well as cosmic muons. The NC events are therefore mostly removed by simply only selecting events that score high in either the ν_μ or ν_e categories, although they cannot be fully removed this way. Sec.4.4.1 discusses how we account for the NC that still find their way into the final selected samples.

Cosmic muons are largely removed, to first order anyway, by the timing cut previously discussed, but this still leaves a sizable background rate. To help deal with the remaining cosmic background, another CNN was developed and trained to select out cosmic muon-like activity from our FD data[3]. This network is used as a filter to remove cosmic-like activity before the regular reconstruction is even run on the FD NuMI beam data, reducing not only

[3]See Sec.4.2 for details on the cosmic CNN.

the cosmic background, but NOvA's consumption of computing resources.

Rock muons (muons that were created by beam interactions with the rock surrounding the ND) are also largely removed prior to full event reconstruction. Here, the position of the event's vertex relative to the detector faces, particularly the front face, is used to remove events that start too close to the detector edges. This cut was optimised to remove the most rock muons possible while still maintaining 99% of the true ν_μ , ν_e , and NC events in the ND sample. This is only done at the ND as rock muons are not a real concern at the above ground FD.

Although the cosmic CNN and the rock muon filter remove a large amount of these backgrounds, there is still some leftover by the time we get past all the other event selection cuts prior to the final PID cut. These are removed with a set of Boosted Decision Trees (BDT) trained to remove the cosmics that look the most like signal, as those are what are be left at this point in the event selection process. There are a total of 4 BDTs each for the ν_μ and ν_e selections.

The ν_μ based BDTs were trained separately for the FHC and RHC beam modes, with additional BDTs trained for each of Period 1 FHC data and Period 2 FHC data as the detector conditions were sufficiently different for those two data-taking runs than for the rest of the data-taking runs. The training samples for these BDTs all had loose PID cuts applied to them to remove any events that were obviously not ν_μ -like. The following were the input variables

- The cosine of the angle with respect to the beam direction
- The track length
- The larger of the y-coordinates of the track start and stop position
- The projected extension length of either the cosmic track or Kalman track end to the detector wall
- The cosine of the track angle to the y-axis
- The fraction of the hits in the slice that were in the track
- The p_T/p (transverse momentum divided by total momentum) of the track

For the ν_e selection, the BDTs were again trained separately for the FHC and RHC beam modes, as well as separate trainings for the Core and Periphery samples. As the

next section will further describe, the ν_e sample is split into two main samples with the Core sample being comprised of fully selected ν_e events, and the Periphery sample being comprised of events that were not contained, but were otherwise selected as ν_e events. The training samples for all four of these BDTs have loose PID cuts applied to them, as well as require there to be a prong with length greater than 0. The FHC and RHC samples are split up into those events in the detector's fiducial volume and those that are not, to train the Core and Periphery BDTs respectively. The following input variables are used for the training

- Number of hits in the slice
- p_T/p of the event
- Direction of the EM shower relative to the z direction
- Distances between any prong and the walls of the detectors
- Fraction of event energy that is in the EM shower
- Shower width
- Vertex location

Together with the timing cut, containment cuts, and final PID selection, these cosmic rejection efforts are able to reduce the cosmic background by 6 orders of magnitude. We use the FD NuMI beam data that falls outside the beam spill window (and is therefore all cosmic data) to estimate the rate of cosmic background events that still potentially make it into our final sample.

4.1.5 Reconstructed Muon Identification Algorithm

While our main PID in both the ν_e and ν_μ selection process is a convolutional neural network, the ν_μ sample selection also utilises an algorithm called the Reconstructed Muon Identification (ReMID) algorithm to distinguish muon tracks coming from ν_μ CC interactions from other tracks coming from background CC events and NC events. In particular, ReMID is designed to primarily distinguish muon tracks from charged pion tracks.

Muons and pions have similar rates of energy loss dE/dx from ionization as they travel through a medium due to their similar masses and charges. However, ionization is not the only way for a particle to experience energy loss. Pions, due to their strong force

interactions, will experience hadronic scattering with an interaction length of 82 cm in our detectors (which is much shorter than a typical track length). Muons will also see some scattering due to weak force interactions, but those occur at a much lower rate. The ionizations from the outgoing particles of the pion scatterings add to the dE/dx recorded by the detectors for that reconstructed track[4]. This creates a difference in shape of the overall dE/dx for pion tracks as compared to the muon tracks. This pion scattering will also present as a higher rate of trajectory deflections with distinct deflection angles. ReMID takes advantage of these differences by using log-likelihood based values for both dE/dx, and the scattering, as inputs.

Muon and pion tracks in our detectors will be of similar lengths up to a certain energy. After that, high energy muons will have tracks that extend much further through the detector than the pions will. Therefore, track length is also used as an input for ReMID. Additionally, because a muon coming from a true ν_μ event will have much less hadronic activity near it (outside of the vertex region, anyway) than a pion will, the fraction of planes that a track runs through that also contain hadronic contamination is used as the final ReMID input.

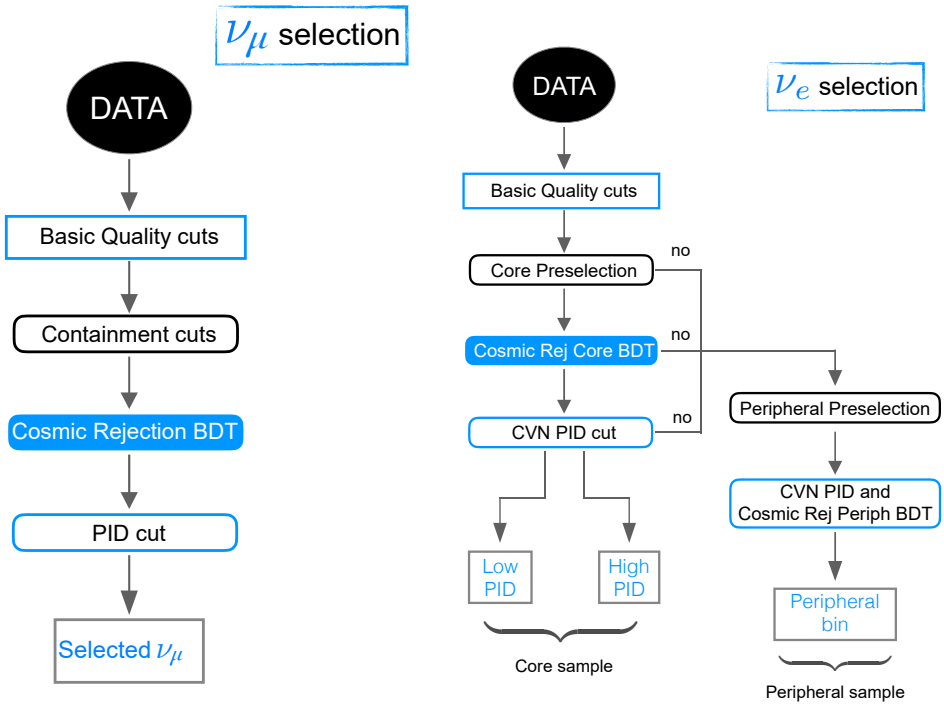
These inputs are all used to create and train a BDT that runs over every track in an event. The track with the highest ReMID score is considered the most muon-like and that score is used as a PID for identifying how much an event overall resembles a ν_μ CC event. Seperate ReMIDs are trained for each of the beam modes and detectors.

4.1.6 Final Event Selection and Binning

Fig.4.2 shows the cut flow for both the ν_μ and ν_e event selection. In both, the final stage is a CNN-based PID cut. However, we do not leave the samples at that, but rather split each of the samples up to increase the sensitivity of the analysis. The final fit of our predictions to data is done to all of these samples simultaneously.

Even though most of the initial muon neutrinos have oscillated away by the time they reach the FD, very few have oscillated into electron neutrinos. Most of them are instead now tau neutrinos that we can't see in our detector. This leads to the final selected ν_e sample to be a small one easily overwhelmed by statistical uncertainties. To combat that, we split the

[4]In reality, at this point, instead of dE/dx, we really mean total deposited energy. The algorithm itself calculates the dE/dx by going plane by plane down the reconstructed track, and setting each plane's dE/dx as the total energy of the hits in the entire plane, ignoring planes where there is obvious hadronic activity like near the event vertex. This means hits from multiple cells in the same plane will be included in the calculation. This in turn allows for all the outgoing particles from pion scattering to be included in the overall dE/dx calculation for that track.



(a) (b)

Figure 4.2: Flowcharts for (a) muon neutrino event selection and (b) electron neutrino event selection

sample into two: the High PID sample and the Low PID sample. These both have a strict cut on the event's CNN-based PID score, but the Low PID sample has a looser cut than the High PID sample. The High PID sample is given more significance when the fit to data is performed, but the analysis' sensitivity is boosted by the additional statistics provided by the Low PID sample. The locations of the cuts on PID score to create the High and Low PID samples are determined alongside the location of the cosmic rejection BDT cut via a box cut in the 2D PID x cosmic BDT space using the following figure of merit (FOM)

$$FOM^2 = \sum_i \frac{S^2}{S+B} \tag{4.1}$$

where S(B) indicates the number of signal(background) events in each of the PID bins. These box cuts are tuned by letting an optimising algorithm make three samples, and then the lowest PID and lowest purity sample is dropped to create just our High and Low PID samples. This results in a FOM comparable to optimising for just two samples to begin with, while also reducing background rates. Each of these samples are placed into bins of neutrino energy E_ν with width 0.5 GeV.

The Periphery sample is yet another ν_e sample comprised of events that passed all the preselection cuts other than containment, as well as a PID cut. These events are ones that are likely to be ν_e events, but we can't guarantee a good energy resolution on them. The PID and comsic rejection BDT cuts on the Periphery sample is optimised in the same way as the Low and High PID samples. This sample is placed into a single bin meaning that their inclusion into any fit is with event rate only. This again allows us to boost the statistical power of the ν_e sample without compromising the sensitivity of the fit to data.

The final ν_μ sample is split into four different samples called quartiles based on each event's hadronic energy fraction $\frac{E_{had}}{E_\nu}$. The power of the FD ν_μ sample in measuring θ_{23} and Δm^2_{32} comes from being able to pinpoint exactly at what energy the majority of the muon neutrinos are oscillating away and at what rate those oscillations occur. This can be visualized in Fig.4.3 where you can see the comparison of what an unoscillated FD ν_μ sample would like next to an oscillated sample. The location and depth of the "dip" around 1.5 GeV in the oscillated sample is what we are looking for. The energy resolution of the sample events affects how well we can characterise this "dip" and we have worse energy resolution with events with a high hadronic energy fraction. Additionally, the higher hadronic energy fraction quartiles are also the most impure. The boundaries of the quartiles are determined so that each of the four samples has a similar number of events populating

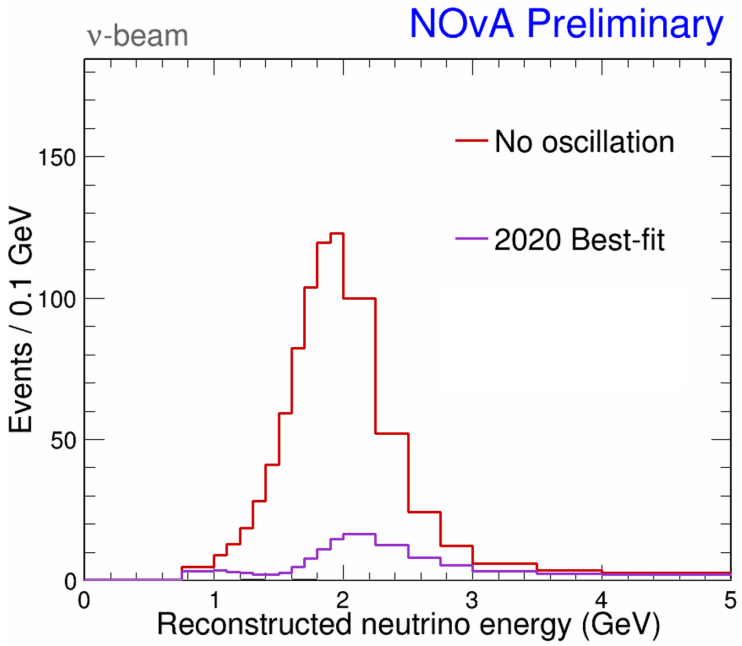


Figure 4.3: Comparison of FD FHC ν_μ prediction with and without oscillations applied. The location and depth of the "dip" in the oscillated distribution around 1.5 GeV is key to our measurement of θ_{23} and Δm^2_{23}.

it. Like the High PID ν_e sample, the lower hadronic energy fraction samples are allowed to pull a larger significance during the final fit.

The final PID cut for the ν_μ sample is a combination of a cut on an event's ReMID score and CNN-based PID score. The cut placement for these two scores, as well as for the cosmic BDT score, is optimized simultaneously using the following FOM

$$FOM = \frac{S}{\sqrt{S+B}} \tag{4.2}$$

These cuts are optimised for the best FOM in the "dip" region as that was found to give us the greatest reduction in background with the smallest negative impact on the signal. These cuts are the same for both beam modes and for all four quartiles.

Like the ν_e samples, the binning for the ν_μ samples is done as a function of reconstructed neutrino energy E_ν, but here, the binning is not uniform. Instead, a variable binning was optimised for the greatest sensitivity in the "dip" region, with finer binning in the "dip" region and much wider bins far away from there. This binning can be seen in Fig.4.3.

4.2 NOvA's Convolusional Neural Networks

As mentioned several times already, NOvA's main PID is in the form a convolutional neural network (CNN), as is one of our cosmic rejection methods. In fact, NOvA was the first to use a convolutional neural network in a HEP result with our paper on the CNN itself in 2016 [32], followed by a results paper in 2017 [33]. That first implementation increased our effective exposure by 30% compared to the more traditional ID methods we were previously using. Since then, we have improved our training methods, and developed the first convolutional neural network for particle identification which uses context information [34]. This section will discuss how these CNNs work and how they are used in NOvA.

Our CNNs were developed in the "image recognition" style and was initially based on GoogLeNet[35]. Since the NOvA detectors are segmented in such a way that it is easy to visualize the data in sets of 2D images where every cell is a pixel, it made sense to choose a network already optimised to handle images as a starting point for developing the structure of our CNN.

The inputs for our CNNs are in the form of pixel maps. These maps are grids of pixels where each pixel represents a cell in the detector and the intensity of the pixel is the calibrated energy deposited in that cell, in the spirit of a grayscale image[5]. Each event slice will have two associated pixel maps, one for each view of the detector. The pixel intensities are encoded using 8-bits that saturate at 278 MeV to help with data storage and transfer during the training and application of the CNNs. The advantage of this pixel map set-up, as well as using a CNN in general, is that these pixel maps are made with very little data processing and reconstruction. Other machine learning methods, like a BDT or kNN, require high-level inputs such as fully reconstructed muon tracks and particle momentums. There is always a margin of error in these high-level variables which can then cause skews in the training or evaluation of the network. We avoid this by using just the low-level pixel maps as our CNN input. Furthermore, this method allows the network to make the decisions on what features are important in the event identification, as opposed to us making that choice via which high-level variables get used as inputs in other identification methods.

Traditional neural networks consist of an input layer, one or more hidden layers, and an output layer. In general, the input and hidden layers are made up of multiple nodes, and the nodes of one layer are fully connected to all the nodes in the next layer, as depicted in Fig.4.4. However, as with most image recognition neural networks, this quickly becomes

[5]To be clear, however, our pixel maps are *not* images.

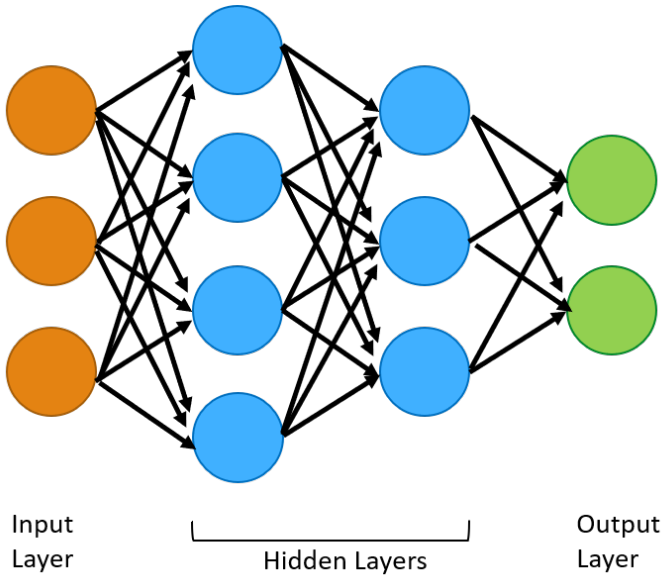


Figure 4.4: Depiction of a simple neural network with two hidden layers.

computationally impractical as the number of input nodes equals the number of pixels in the image (or, in our case, pixel map). Convolutional neural networks include layers that are not fully connected, but rather only connect pixels within a certain radius of each other, to reduce the computational load. These convolution layers do this by applying a filter (also commonly called a kernel) to a part of the pixel map, calculating the result of the filter, and using the result of that filter to create a pixel in a new pixel map. Then this filter is slid over a set number of pixels and repeated until the entire original pixel map has been processed, and the new one populated. These convolutions pull out features from the pixel maps that the network can make further decisions on, like object edges in a traditional, image-based CNN. In NOvA's CNNs, these convolution layers produce feature maps of the same dimensions as the pixel maps. Pooling layers are introduced that functionally work the same as the convolution layers, but work to reduce the dimension of resultant maps. Fig.4.5 shows the basic structure of a CNN like the one NOvA uses.

Unlike most images, NOvA's pixel maps actually come in sets of two, one for each view of the detector. Correspondingly, our CNN is set up to actually start out as two parallel networks, called towers, that work to identify features from each view independently. The two views are connected in the final layers of the network, with the last layer being a fully connected layer.

NOvA actually has three different CNNs it uses in its 3-flavor analysis: Event CNN,

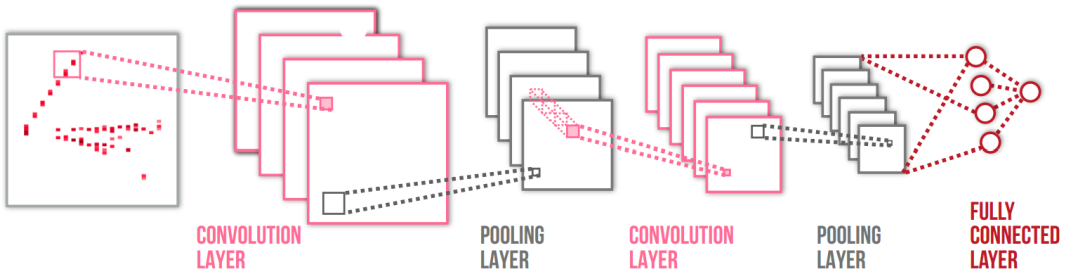


Figure 4.5: Basic structure of a Convolutional Neural Network. The pink and grey layers do the feature extraction while the final red layers resemble more traditional neural networks.

Prong CNN, and Cosmic CNN. The Event CNN (also called CVN for Convolutional Visual Network) follows the network architecture just described, and is the CNN used as our primary PID. The output layer is comprised of four nodes that output scores between 0 and 1 for the likelihood that the event can be classified as either a ν_μ CC event, ν_e CC event, NC event, or cosmic activity. Fig.4.6 gives a visual representation of the results of this CNN. Note the level of separation achieved between the different event types, although there is non-negligible overlap of the NC events into both the ν_μ CC and ν_e CC regions.

The training of the Event CNN is supervised learning, meaning that the training sample is tagged with its true event type. Currently we can only do this by using simulated data as the training, testing, and validation samples[6]. While we do overlay this sample with real cosmic data to improve the performance on selecting cosmic muons, no other data is involved in the training process. So while we can evaluate the performance level and biases of the CNN on MC in the same way that one would typically test and validate a CNN, we must validate the performance of the CNN on data by other means. For the ν_e selection, this is done via a suite of studies called the Muon Removed studies. These studies, described in detail in Chap.5, are the primary focus of this thesis.

The second CNN used is the Prong CNN. This network aims to classify the individual particles that make up an event as being either a muon, electron, photon, pion, or proton. The architecture for this network differs from the Event CNN in that it has two additional towers for the top and side view of just the prong being evaluated, with our fuzzy-k reconstruction providing the prongs. All four towers are concated together in the final layers of the network

[6]In the future, we may be able to incorporate real tagged data coming from our Test Beam effort, where charged particles are tagged by beamline instrumentation before they even enter the Test Beam detector. See App.A for more information.

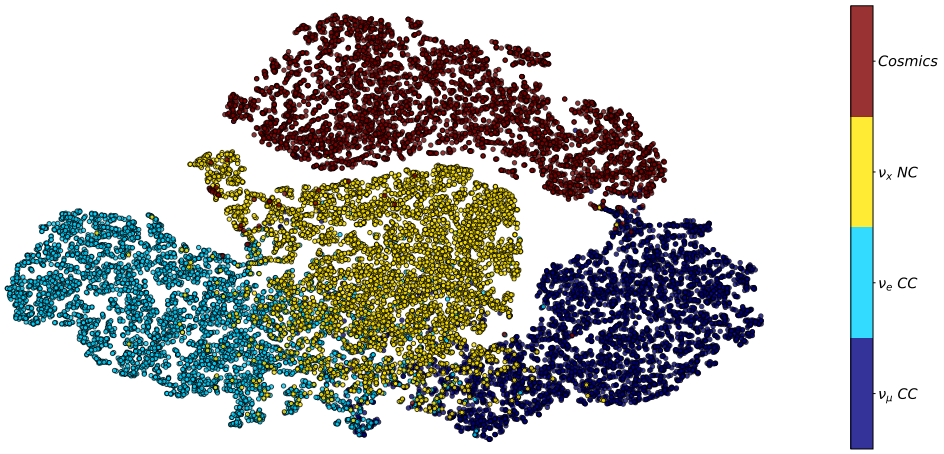


Figure 4.6: t-SNE transformation of the feature layer of NOvA's Event CVN showing its ability to separate the different event types. Note the pronounced seperation of the cosmic events from the neutrino events, and the expected overlap of the NC events on the CC events in the middle region.

just like the two towers in the Event CNN. The particle classification done with Prong CNN is primarily used in aiding with energy estimation as described in the next section.

The final CNN is the Cosmic CNN as mentioned in Sec.4.1.4. This network has a similar general structure to the Event CNN, but the input pixel maps are different. Because the purpose of this Cosmic CNN is to filter out cosmics from the FD NuMI sample before reconstruction is run, in order to reduce the computational load of the experiment, the pixel maps will not have calibrated energy information. Instead, the pixels each are assigned a 1 or 0 to indicate if there was any energy deposition there or not. Furthermore, these maps are the size of the entire FD in 16 μs windows instead of just focusing in on an event slice like the other CNNs do. These windows overlap by 1 μs to ensure that events are not being chopped up by the time boundary, and the entirety of the beam spill window is always ensured to be fully inside one of these windows. This CNN is able to achieve over 90% cosmic rejection.

4.3 Energy Estimation

Since the neutrinos travelling through our detectors are invisible until they interact with the matter in the detector, producing visible charged particles, we cannot directly

measure the incident neutrino energies. Instead, we must reconstruct it from the energies and momentums of the outgoing particles of the interaction. The 3-flavor analysis uses two different energy estimators to reconstruct the neutrino energy E_ν, each optimised for either ν_μ or ν_e events.

4.3.1 Electron Neutrino Energy Estimation

The ν_e energy estimator first uses Prong CNN to identify the prongs in a ν_e candidate event that are associated with electromagnetic activity including photons, electrons, and neutral pions. The rest of the prongs are assumed to be hadronic activity and grouped together as such. The calorimetric energy (the calibrated energy derived from observed PE as described in Sec.3.4) of each group is summed into E_{EM} and E_{HAD}. These are inputted into

$$E_\nu = \alpha_0(\alpha_1 E_{EM} + \alpha_2 E_{HAD} + \alpha_3 E_{EM}^2 + \alpha_4 E_{HAD}^2) \tag{4.3}$$

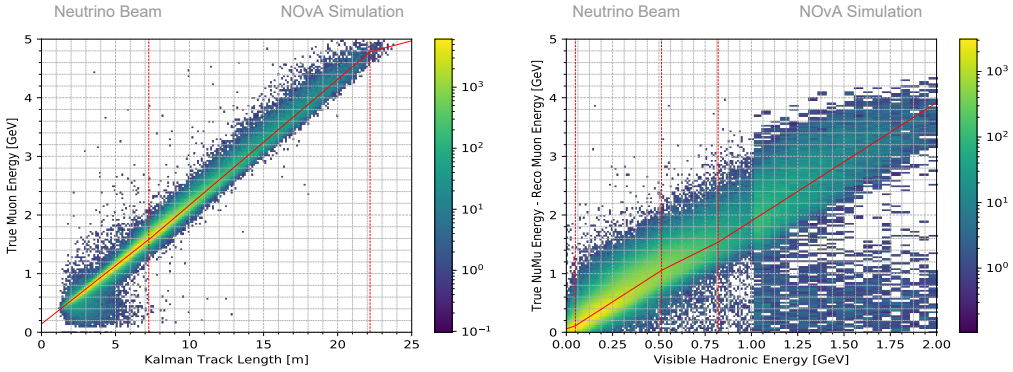
where the coefficiencts α_i were found by fitting this function to simulated events while minimizing variance of E_ν from the true neutrino energy. There is a separate set of coefficients for each beam mode. The resolution of this energy estimator is given as the Root Mean Square (RMS) of $(\frac{E_\nu^{True} - E_\nu^{Reco}}{E_\nu^{True}})$ and was 10.7 (8.8)% for the FHC (RHC) in NOvA's 2020 analysis.

4.3.2 Muon Neutrino Energy Estimation

The ν_μ energy estimator finds the muon and hadronic energy separately using the muon track length and the visible calorimetric hadronic energy[7] respectively as inputs into piece-wise linear spline functions fit to simulated ν_μ events. The final neutrino energy is defined as the sum of these two energies $E_\nu = E_\mu + E_{Had}$.

To perform the spline function fit, we first construct 2D histograms of true muon energy vs. Kalman track length for the muon and (true neutrino energy - reconstructed muon energy) vs. visible hadronic energy for the hadronic activity, as seen in Fig.4.7. Here, the muon track is identified using ReMID. For each piece in the histograms, a Gaussian profile is created and the mean is fit to the spline function. This fit is done while optimising the energy resolution.

[7]In other words, all the calorimetric energy left in the slice after removing the muon track from consideration



(a) (b)

Figure 4.7: The 2D histograms used to fit the (a) muon and (b) hadronic piece-wise linear spline functions in the FHC beam mode. The vertical dashed red lines demarcate the boundaries for each linear segment. The solid red lines are the result of the fits to the Gaussian distributions of each piece.

Mode	Detector	RMS$\left(\frac{E_\nu^{True}-E_\nu^{Reco}}{E_\nu^{True}}\right)$
FHC	ND	10.8%
	FD (P1)	9.6%
	FD (P2)	9.2%
	FD (P3+)	9.0%
RHC	ND	9.4%
	FD	8.2%

Table 4.1: Energy Resolutions of the muon neutrino energy estimator

The ν_μ energy estimator is trained separately for each beam mode and each detector. Additionally, the FD FHC estimator was trained independently for the first period as well as the second period of data taking as the detector conditions were sufficiently different during those times relative the rest of the FD's data-taking. Table 4.1 gives the energy resolutions achieved by the ν_μ energy estimator for the 2020 analysis.

4.4 Extrapolation

Because NOvA has a two detector design, we can use the ND to inform us on the initial flux of neutrinos before they travel to the FD. We use the ND NuMI beam data to correct event rates in our ND MC and then extrapolate that simulation to the FD, taking into account the difference in detector sizes and oscillations. These predictions then fit to the FD data.

In this section, we will cover the processes used to first correct the ND simulation and then the extrapolation process to create our FD predictions.

4.4.1 Decomposition

In essence, our analysis works by counting up how many muon (anti)neutrino interactions we start out with at the ND, and see how many we still have at the FD as well as how many electron (anti)neutrinos have appeared by that point. The difference between the ND and FD event counts is what gives us information on the various oscillation parameters we are trying to probe. Obviously, the analysis is much more involved than just counting events, but being able to do proper accounting of event rates at both detectors is essential to the analysis. The ND ν_μ ($\bar{\nu}_\mu$) simulated event rate matches what is seen in data pretty well to begin with thanks to our cross section group's work in providing data-driven tunes to the cross section model used to generate the events in our simulation. On the other hand, the rate of ν_e -like background events at the ND does not see the same level of agreement. These backgrounds include intrinsic beam ν_es as well as true ν_μ CC and NC events that look like ν_e events. After applying the ν_e selection cuts, as described in a previous section, to both the ND data and MC, there is a 12.5% deficit of simulated events compared to the data events. We use data-driven processes called decomposition to correct these MC ν_e background event rates so that we can accurately account for these backgrounds when we make our appeared ν_e prediction at the FD.

There are three different decompositions we use, each focusing on a different background event type. BEN decomposition focuses on correcting the rate of intrinsic beam ν_es that are the result of pion and kaon decays. The dominant decay modes of these mesons actually produce ν_μ ($\bar{\nu}_\mu$)s, so we can use the rates of contained and uncontained ν_μ candidates in the ND to get a handle on the amount of the parent particles that were present. The contained ν_μ sample is well understood to be made up of primarily ν_μs coming from pion decay. For each bin in neutrino energy, the difference between the number of contained candidate ν_μ events in the data and MC sample is used to create a set of weights that can be applied to this piece of the ν_e sample. The uncontained ν_μs come from both pions and kaons, so the weights found from the contained ν_μs are applied to this sample to account for the pion differences first. Then, new weights are made from the remaining difference in the MC and data event rates. These weights can then be applied to beam ν_es coming from kaon decay.

The NC and ν_μ CC background pieces of the ND selected ν_e MC is corrected using Michel decomposition. Here we use Michel electrons[8] to identify the ν_μ CC and NC events from within the selected ν_e ND sample. For the ν_μ CC events, the muon typically ends with a Michel electron. Meanwhile, these electrons are usually coming from the hadronic shower of the NC events (usually from charged pion decay) at a lower rate than the ν_μ CC events. An algorithm is used to find these Michel electrons in the selected ν_e data and MC samples, and a template log likelihood fit to the data is used to separate out the ν_μ CC contribution from the NC and true ν_e CC contribution[9]. This fit includes scale factors for the ν_μ CC and NC pieces of the Michel electron samples for each neutrino energy bin, which can then be used as correction weights.

The final decomposition used is proportional decomposition and is much simpler than the other decompositions. For each analysis energy bin of the selected ν_e ND samples, we find the difference between the data and MC samples. These differences are used to weight each component of the MC so that the proportions of each component in the bin stays the same, but the overall event rate of the bin matches the data.

In past analyses, including our most recent 2020 analysis this thesis is based on, the FHC ND simulation has been corrected by applying the BEN decomposition, Michel decomposition, and proportional decomposition in that order. The RHC simulation has only had the proportional decomposition applied. There are plans to apply a version of the BEN decomposition to the RHC in future analyses, though. Fig.4.8 shows the results of applying the BEN and Michel decompositions to the FHC for the 2020 analysis.

[8]Electrons produced when a muon decays at rest. In our detectors, these electrons do not have enough energy to induce an EM shower and instead appear as a few extra hits at the end of a muon track that are easy enough to distinguish from the muon track itself.

[9]Michel electrons in both the NC and ν_e CC events are coming from hadronic showers in similar ways, so it is near impossible to separate the NC and true ν_e CC events using Michel electrons. However, we will have already corrected the amount of true ν_e CC events using BEN decomposition, so this correction is still primarily going only to affect the NC and ν_μ CC event rates, especially since we can use truth information in the corrected events to force it so.

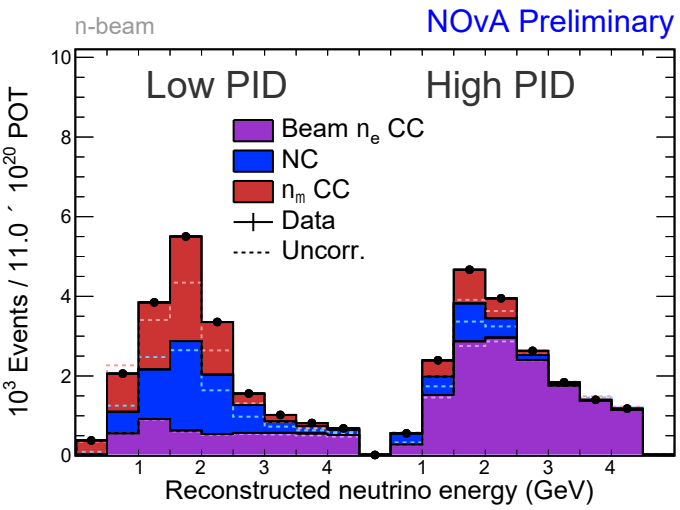


Figure 4.8: The 2020 analysis ND FHC ν_e selected samples. The solid shaded stacked histograms show the MC after the BEN and Michel decompositions were applied to the MC and the dotted lines show the uncorrected MC for each component. Note the general agreement of the corrected MC to the data.

4.4.2 Predictions

Once we have the ν_e ($\bar{\nu}_e$) and ν_μ ($\bar{\nu}_\mu$) event selections set, and we have the decomposition correction weights calculated, we are ready to make our FD signal and background predictions using an extrapolation process. Fig.4.9 shows the general extrapolation process of the ν_μ component as an example. The steps performed for each component are as follows

1. Select the candidate neutrino events from the ND MC to get a spectra of events as a function of reconstructed neutrino energy.
2. Apply any appropriate decomposition corrections to the selected MC sample.
3. Convert the MC spectrum to be a function of true energy.
4. Apply a Far to Near ratio to scale the spectra to the FD.
5. Apply any appropriate oscillation probabilities.
6. Convert the spectrum back into a function of reconstructed energy.

The Far to Near (F/N) ratio mentioned in step 4 is the ratio of the corrected ND MC true energy spectrum to the uncorrected ND MC true energy spectrum, multiplied by the FD

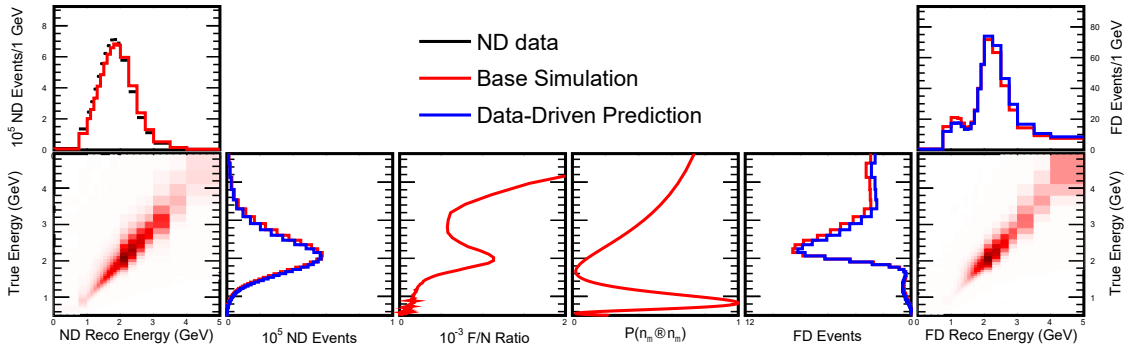


Figure 4.9: A cartoon example showing the general extrapolation process. Here, we are showing the extrapolation process for the $\nu_\mu \rightarrow \nu_\mu$ oscillation mode.

MC true energy spectrum. Applying this ratio allows us to correct the event rates to the FD to account for the differences in the environment of the two locations such as detector size and the effect of beam spread, as well as differences between ν_e and ν_μ cross sections.

This whole process is done separately for each component of the initial ND sample and for each possible oscillation channel, and the final predictions can be built by adding all the necessary pieces for each analysis together into one spectrum. While this general process is followed for all parts of the final prediction, there are aspects unique to each component and oscillation channel.

$\nu_\mu \rightarrow \nu_\mu$ signal

For this oscillation channel, both ν_μ and $\bar{\nu}_\mu$ events are considered signal as differentiating between the two does not affect the analysis' sensitivity to θ_{23} and Δm^2_{23}. Additionally, unlike the ν_e samples, the beam based backgrounds to the ν_μ disappearance channel are very small, especially compared to the cosmic background rate. That means that we can ignore their contributions to the oscillations, and derive the entire $\nu_\mu \rightarrow \nu_\mu (\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu)$ FD signal from just the ν_μ ($\bar{\nu}_\mu$) ND selected sample.

Sec.4.1.6 already outlined how we split the ν_μ selected samples up into quartiles on the basis of hadronic energy fraction. For the extrapolation process, we further divide each quartile into three bins based on the reconstructed transverse momentum of the muon $|\vec{p}^{\,\mu}_T| = |\vec{p}_{mu}|\sqrt{1 - cos^2\theta_\mu}$ where θ_μ is the angle of the muon with respect to the beam direction. Each of these now twelve samples for each beam mode are extrapolated independently, and then the bins of $|\vec{p}^{\,\mu}_T|$ for each quartile are added back together to make the final ν_μ disappearance signal prediction. This separation into momentum bins is done because we

have multiple systematic uncertainties that are dependent on how $|\vec{p}_T^{\,\mu}|$ couples to the smaller ND's acceptance of events compared to the FD[10]. It was found that extrapolating in these $|\vec{p}_T^{\,\mu}|$ bins helps improve the robustness of the analysis by reducing some of those systematic uncertainties.

$\nu_\mu \rightarrow \nu_e$ signal

Here, we must start with the ND ν_μ sample, but end up with a ν_e sample. This is done by taking the same initial ν_μ sample as for the ν_μ disappearance channel, merge the quartiles into one spectrum, and then split it into $|\vec{p_T}|$ bins as before for extrapolation to the FD. The F/N ratio here is constructed with the ND ν_μ corrected and uncorrected MC spectra and the FD ν_e MC spectra. Like the ν_μ disappearance channel, we again add the $|\vec{p_T}|$ bins back up into a single spectrum after extrapolation. Unlike the ν_μ disappearance channel, however, we must extrapolate the $\nu_e \rightarrow \nu_e$ signal separately from the $\bar{\nu}_e \rightarrow \bar{\nu}_e$ signal as the observed difference between these channels is key to probing δ_{CP} and the neutrino mass hierarchy in our analysis.

$\nu_\mu \rightarrow \nu_e$ backgrounds

The most important backgrounds here are the same that the decompositions correct for: intrinsic beam ν_es and ν_e -like ν_μ CC and NC events. These go through the extrapolation process in a similar way as the ν_e appearance signal, but with the ND ν_e selection as the starting spectrum. Because this sample has a sizable portion of misidentified ν_μ CC and NC current events as previously described, we choose not to do the conversions of the spectra to and from functions of true energy since the misidentified events would provide true energies that diverge from the reconstructed energy provided by the ν_e energy estimator. This means that the F/N ratio is calculated in terms of reconstructed energy instead of true energy like what is done for the other extrapolations.

The other important background to the ν_e appearance oscillation channel that must be extrapolated, is the wrong sign oscillation channels defined as the channels that produce $\bar{\nu}_e$ (ν_e) events at the FD for the FHC (RHC) beam modes. Remember that although we train our PIDs separately on the different beam modes, they cannot actually distinguish between a neutrino and an antineutrino, so accounting for the wrong sign contribution to the ν_e appearance predictions is done by including those oscillation channels as their own extrapolated predictions.

[10] Smaller detector means the ND cannot contain larger, more energetic neutrino interactions to the level that the much larger FD can.

Other oscillation channel backgrounds

We expect there to potentially be a small amount of ν_τ CC present in both the ν_μ and ν_e candidate samples, but the size of this background is considered to be extremely small. Therefore, instead of extrapolating the relevant oscillation channels, we take this background straight from the FD MC.

While we mostly only consider the oscillation channels of ν_μ survival and ν_e appearance from a ν_μ source, we also need to account for background events coming from oscillations of ν_τs into either a ν_μ or ν_e , and beam ν_es oscillating into a ν_μ . These background rates for the final prediction are all taken from the FD MC as well.

Backgrounds not related to oscillations

The cosmic background prediction is estimated from the dedicated cosmic trigger data. Spectra are created from this dataset using the same selection cuts and binning as the ν_μ and ν_e signals and is very statistically rich. These spectra are then scaled to match the event rate of the NuMI beam data outside the beam spill window (often called the NuMI sideband). The final cosmic prediction is then found by scaling this sample once more, this time to match the time-wise exposure of the FD NuMI beam data.

Like the rock muons at the ND, we also expect there to be some contribution to the background of muons coming from particle interactions with the FD building and its limited amount of overburden. Although this rate is considered to be quite small, especially when compared to the rock muon rate at the ND, it still comes to a non-negligible amount of background for the ν_e sample. This background prediction is found using a specially simulated sample to estimate the event rate.

After all the relevant extrapolations have been made and the cosmic and rock muon rates calculated, the final predictions are built by adding all the relevant extrapolated spectra together with the cosmic and rock muon spectra for each beam mode.

4.5 Systematic Uncertainties

No experimental analysis is complete without consideration of the systematic uncertainties present, and NOvA is no different. For all of our 100+ identified sources of systematic uncertainties, their impacts are evaluated by either reweighting events in our nominal MC or by producing new simulation with systematic shifts applied at event generation. We are then able to create sets of predictions with these systematic uncertainty samples following the same methodology as the nominal prediction for use in the fits to data and in evaluating

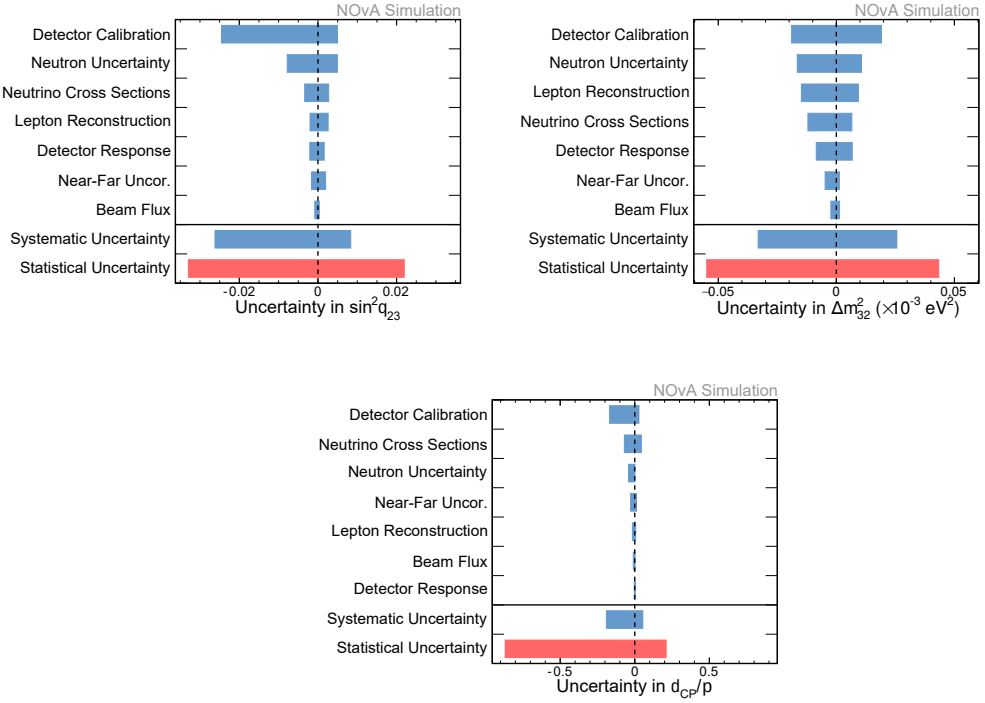


Figure 4.10: Estimated impacts of the systematic uncertainties on the measured oscillation parameters for a joint ν_e + ν_μ fit to both the FHC and RHC beam mode data. Here, the oscillation parameters are taken to be $sin^2\theta_{23} = 0.568$, $\Delta m^2_{32} = 2.40 \times 10^{-3} eV^2$, $\delta_{CP} = 0.82\pi$, and $sin^2\theta_{13} = 0.085$.

the analysis sensitivities.

For the most part, our systematic uncertainties can be categorized into being related to detector calibration and response, neutron uncertainty, neutrino interaction modelling (cross sections), near-far differences, lepton reconstruction, and beam flux. Fig.4.10 shows the impact of each of these uncertainties on $sin^2(\theta_{23})$, Δm^2_{32}, and δ_{CP} as estimated at the published 2020 best fit point[36].

The detector calibration uncertainties are the most impactful. These are mostly driven by the 5% uncertainty assigned to the calorimetric energy scale described in Sec.3.4. Other calibration related uncertainties come from the overall relative calibration as well as the modelled attenuation shape at the ends of the cells. While the detector calibration systematic uncertainties are currently the largest source of uncertainty, we are hoping to reduce this in the future with our ongoing Test Beam effort (See App.A).

The detector response uncertainties include uncertainties such as those assigned to our

light level model and Cherenkov light model. New to the 2020 analysis was a systematic uncertainty developed to simulate the effect of the detectors aging by creating a sample where the overall light level of the detector was reduced as a function of time. This new systematic uncertainty was developed as a result of my work on characterising the detector aging observable in the FD as outlined in App.B.

The neutron response uncertainty comes from our inability to well constrain the response of detector materials to fast (roughly $\geq$ 1 Mev KE) neutrons with our data. We consistently see a discrepancy between our predictions and data in neutron-rich $\bar{\nu}_\mu$ samples. The source of this discrepancy in our simulation modeling remains unknown, so the systematic uncertainty is developed by adding or subtracting energy from simulated events depending on the number of neutrons that are present with visible energy less than 20 MeV.

The neutrino interaction and cross section uncertainties are multifaceted and make up a majority of the individual sources of uncertainties. A full and detailed description of these uncertainties can be found at [22]. The sheer amount of uncertainties that fall under this umbrella of neutrino interactions becomes computationally prohibitive, so we instead identify the largest impact uncertainties from this set (about 25 uncertainties in total) and perform a Principle Component Analysis (PCA)[37] on the rest. To further reduce the computational load of the PCA application, we only actually employ the twelve most impactful principle components of the hundreds calculated from this method.

Detector specific uncertainties make up the next category of uncertainties. These encompass the effects of the differences between the detectors in terms of detector mass, beam exposures (the beam spreads out before reaching the FD), and acceptance differences coming from the difference in size of the detectors. Additionally, there is an uncertainty related to rate of event pile-up observed in the ND due to its proximity to the beam source, as well as an uncertainty related to the ν_e selection in the ND. The FD has an uncertainty derived from the large amount of cosmic rays that enter it, that the ND does not receive. The most recent analysis worked to improve several of these uncertainties, such as the inclusion of the the $|\vec{p_T}|$ binning in the extrapolation to reduce the ND acceptance uncertainty. The TDSlicer described in Sec.3.6.1 was an improvement on previous slicers and helped reduce the ND pile-up and selection efficiency uncertainties for the 2020 analysis.

Lepton reconstruction uncertainties include uncertainties related to variables such as the muon energy scale and lepton angle. The muon energy scale uncertainty has five components, three of which are uncorrelated uncertainties associated with the fundamental uncertainties in GEANT and the elements of the Bethe-Bloch equation that are relevant to

each of the FD, the ND, and the ND muon catcher. The fourth piece is an uncertainty on neutron pile-up in the ND arising from the observation of lengthening muon tracks in the muon catcher portion of the ND. The final piece is a correlated error on the muon energy scale between the FD, ND, and muon catcher. This correlated uncertainty is the dominant piece.

The last set of uncertainties are the flux and beam uncertainties. The sources of uncertainties here include those coming from the modelling of hadron production from the proton beam on the graphite target, as well those coming from beam transport parameters such as beam position, horn current, and target position. The large number of uncertainties in this category make it useful to run a PCA on them as was done for the neutrino interaction uncertainties. Here, all of the uncertainties are included in the principle component calculations, and the five most impactful are chosen to be used in the analysis.

In the end, we use a set of 64 systematic uncertainties in the fitting procedure for the 3-flavor analysis. It should be noted that for every measured oscillation parameter, the statistical uncertainty is larger than the total systematic uncertainty, as seen in Fig.4.10. While we will continue to take data for several more years, we unfortunately don't project that we will reach the point where our statistical uncertainty will become smaller than the systematic uncertainty before the end of our data taking.

4.6 Results

4.6.1 Fitting and Sensitivities

Once we have all of our predictions made, both the nominal and those for the systematic uncertainties, and we have done the necessary pre-processing on our blinded FD NuMI beam data, we have reached the point of "box opening." This is where we can finally look at our FD data and perform fits of the predictions to the selected ν_μ and ν_e data to derive the measurement of the target oscillation parameters.

The main fit is performed as a joint ν_μ disappearance and ν_e appearance fit over both the FHC and RHC predictions and data. This means performing a simultaneous fit over a total of 14 samples including the ν_μ quartiles for each beam mode, as well as the Core and Periphery ν_e samples, again in both beam modes. The fit itself is performed by minimizing a binned log-likelihood between the expected prediction $E(\vec{\theta}, \vec{\delta})$ and the observed data O. The systematic uncertainties are included as a sum of Gaussian penalty terms for each

systematic degree of freedom $\vec{\delta}$. This log-likelihood is defined as follows

$$\ln \mathcal{L} = -2 \sum_{i=1}^{N} \left[E_i(\vec{\theta}, \vec{\delta}) - O_i + O_i \ln \frac{O_i}{E_i(\vec{\theta}, \vec{\delta})} \right] + \sum_{j=1}^{N} \frac{\delta_j^2}{\sigma_j^2} \tag{4.4}$$

where i indicates the bin index and j indicates each of the 64 systematic uncertainties. The expected prediction $E(\vec{\theta}, \vec{\delta})$ is found using the full oscillation probability (not just the 2-flavor approximation) including matter effects. The following are set values used in the fit

- L = 810 km (baseline length)
- ρ = 2.84 g/cm^3 (average matter density)
- $\Delta m^2_{21} = 7.53 \times 10^{-5} eV^2$
- $\sin^2 2\theta_{12} = 0.851$
- $\sin^2 2\theta_{13} = 0.085 \pm 0.003$

where the last three were taken from the 2019 edition of the PDG[38] for use in the 2020 analysis. These are updated for every new analysis to keep up with global fits of the oscillation parameters which NOvA does not measure. Δm^2_{21} and $\sin^2 2\theta_{12}$ are fixed while $\sin^2 2\theta_{13}$ is allowed to float within its error range as a nuisance parameter like the systematic uncertainties. The values that we are trying to measure- Δm^2_{23}, $\sin^2 \theta_{23}$, and δ_{CP} - are allowed to float without constraints during the fit. The systematic uncertainties are also allowed to float, and their variance is tracked as the number of standard deviations they vary from their nominal value.

Our best fit point is the point that minimizes Eq.4.4 and includes values for the oscillation parameters as well as pulls on the systematic uncertainties. From this point, we can build confidence contours and significances using a Profiled Feldman-Cousins (FC) approach[39]. This approach allows us to build up confidence intervals that can be interpreted in the typical Frequentist manner while accounting for the fact that NOvA does not meet the requirements of Wilks' Theorem[40] which is necessary for a traditional Neyman[41] construction of

Frequentist confidence intervals[11].

The FC method provides a non-parametric approach to defining confidence intervals by throwing a large number of pseudoexperiments at sampling points throughout the parameter space of interest. More specifically, for NOvA, this method entails first defining the sampling points that will be used to probe a parameter space. Because computational resources are not infinite, we choose to sample 60 points distributed along the axis of a 1D significance plot, or a 30x30 grid of points distributed through a 2D space to build confidence contours. At each point, we first fit to the data with the parameters of interest fixed at that point in parameter space to find the set of nuisance parameters (systematic uncertainty pulls and the oscillation parameters not being mapped in the plot) that best minimize the log-likelihood at that point. Then, a large number of pseudoexperiments are thrown by generating Poisson-fluctuated neutrino energy spectra based on the predictions created by the oscillation parameters at this point and their associated set of fitted nuisance parameters. Two likelihoods are found for each experiment, one where only the nuisance parameters are allowed to float during minimization of the likelihood function, and one where all the parameters are allowed to float. A likelihood ratio of these two values is then found and added to the distribution of all such ratios for that point in parameter space.

For 1D significance plots, these distributions of likelihood ratios are used to find a p-value at each grid point by counting the fraction of pseudoexperiments that had a ratio larger than that of the data at that same point in parameter space. This p-value is then converted to a significance σ and plotted. For the 2D plots of confidence contours, after the likelihood ratios were found at each grid point, we find and map the critical likelihoods corresponding to the traditional 1-σ, 2-σ, and 3-σ significance levels. These three maps are subtracted from the map of standard likelihood found from the fits to data at each point. The final contours are then drawn where each of these resulting surfaces intersect with the 0 plane. We can then report our final results as a set of best-fit parameters with their accompanying significance contours and 1D significance plots. This Profiled FC method can also easily be extended to run hypothesis testing for the mass hierarchy and/or the θ_{23} octant.

One other plot that is often useful for conveying the results of our analyses is the bi-event

[11] NOvA's violations of Wilk's theorem come in a few forms. First, Wilk's theorem wants the maximum likelihood estimators of the parameters have ellipsoidal distributions. However, the parameters NOvA is measuring have several physical boundaries, such as the limited range that sine functions can produce, or the cyclical nature of the CP violating phase. Additionally, our event count is small meaning that we can't ignore $O(1/\sqrt{N})$ level deviations from the χ^2 distributions.

plot for ν_e appearance, as seen in Fig.4.11. The axes of this plot are the number of observed electron neutrino and antineutrino events. The ellipses are built from the predicted number of observed ν_e and $\bar{\nu}_e$ events expected for every value of δ_{CP} given a set combination of the other oscillation parameters. Each ellipse represents a different combination of mixing angles and mass splittings that correspond to each combination of mass hierarchy (normal or inverted) and θ_{23} octant. We can put a data point on the plot representing our count of observed ν_e and $\bar{\nu}_e$ events with both vertical and horizontal error bars, to show which combinations of hierchy, octant, and δ_{CP} are compatible and/or favored/disfavored by our data. Note, however, that the center of the plot is a region with a lot of degeneracies. This means that if nature is unkind and places our data point in this region, it could be difficult to disentangle the effects from the mass hierarchy, δ_{CP}, or octant beyond being able to disfavor combinations associated with the outer ends of the ellipses. If that is the case, however, all is not lost as new experiments can be designed that would create greater separation between these ellipses. For example, DUNE is an upcoming experiment whose baseline will be 1300 km, the entire length of which will be cutting through the Earth's surface. This increased amount of exposure to matter effects will pull the inverted hierarchy ellipses up and to the left, while the normal hierarchy ellipses would be pulled own and to the right, creating a larger separation between them.

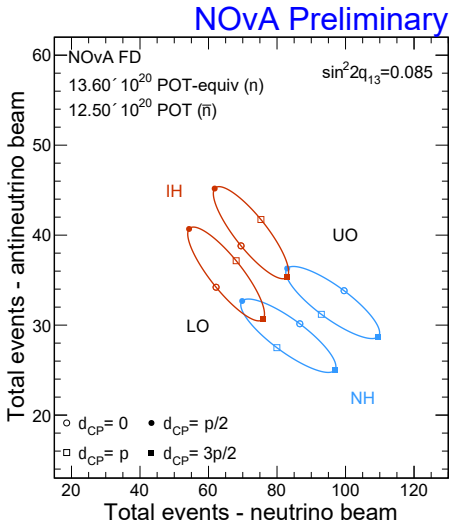


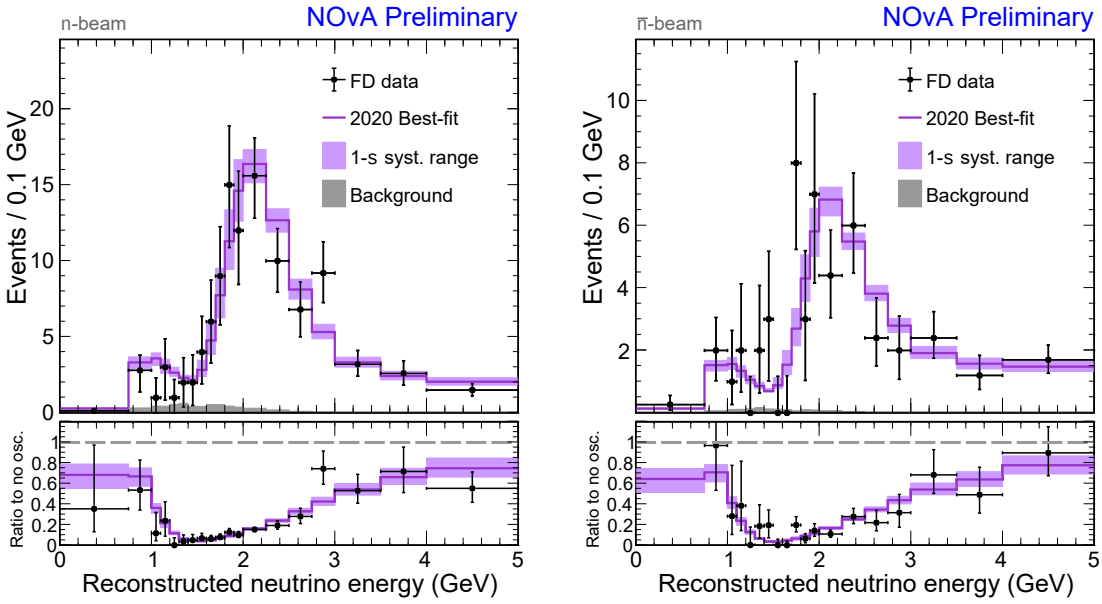
Figure 4.11: Example of a bi-event plot for the ν_e appearance analysis without the best fit or data point shown. Here, NH(IH) stands for Normal (Inverted) Hierarchy and UO(LO) stands for Upper (Lower) Octant. Each ellipse was constructed with the associated values of Δm^2_{32} and $\sin^2\theta_{23}$ fixed and allowing δ_{CP} to vary.

4.6.2 2020 Results

Our most recent major 3-flavor analysis was in 2020 and analysed data taken from February 6, 2014 to March 20, 2020 for a total exposure of 13.6×10^{20} POT of the Numi beam in the FHC neutrino beam mode and 12.5×10^{20} POT in the RHC antineutrino beam mode[36]. Table 4.2 details the events counts both observed and predicted at our best-fit point which in turn is detailed in Table 4.3. Figures 4.12 and 4.13 show the predicted ν_μ and ν_e spectra, again at the best fit point, along with the observed spectra from data. Fig.4.14 shows the bi-event plot from the previous section, but with our data point and best-fit point added. As you can see, nature was indeed unkind and placed is in the degenerate region. Regardless, our best-fit point was found in the normal mass hierarchy and upper θ_{23} octant with a significance of 1.0 σ and 1.2 σ, respectively, and an overall p-value of 0.705.

Looking at both the bi-event plot and the contours shown in Fig.4.15, we can see that we exclude the inverted mass hierarchy with $\delta_{CP} = \pi/2$ at more than 3σ and the normal mass hierarchy with $\delta_{CP} = 3\pi/2$ is disfavored at the 2σ confidence level. Other than these specific regions, we neither favor nor disfavor CP violation over conservation due to the level of degeneracy in this region.

These results agree with those of other oscillation experiments[43][44][45][46] as seen

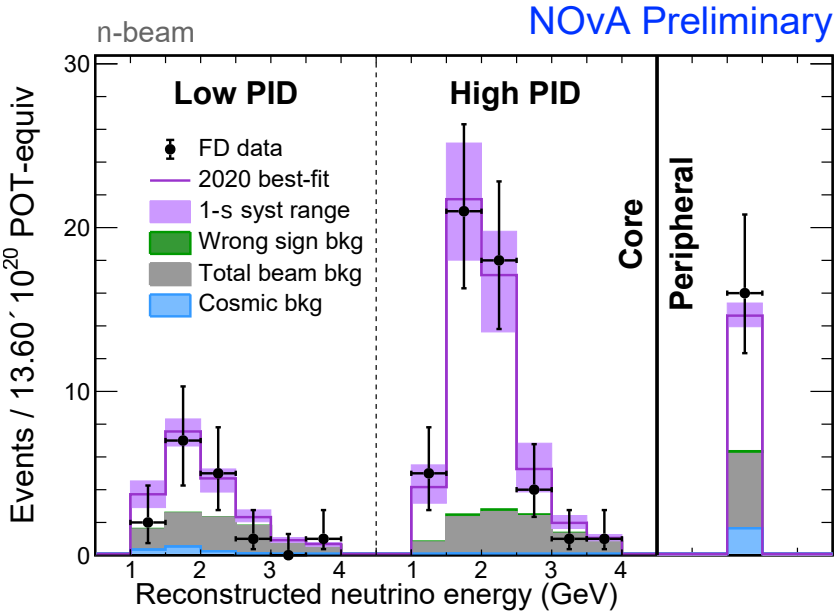


(a) (b)

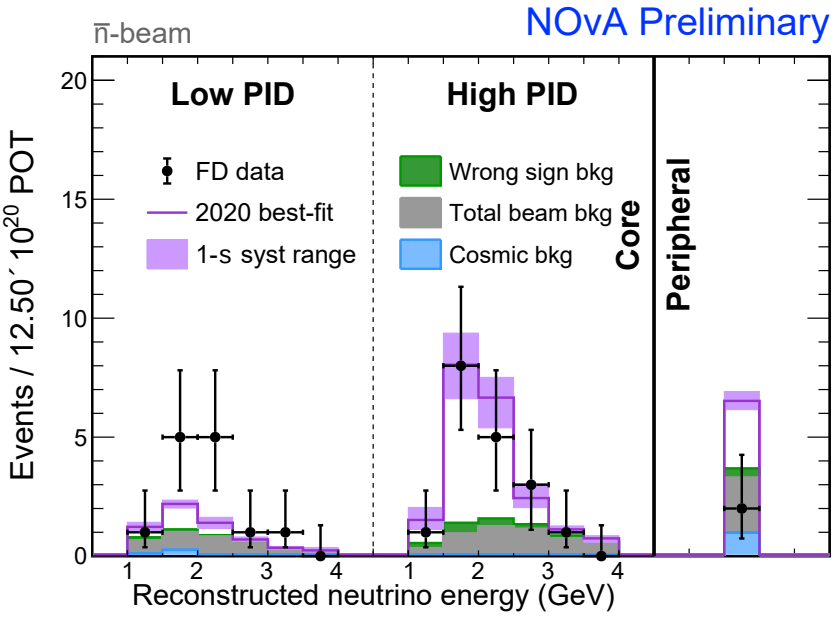
Figure 4.12: The predicted and observed ν_μ spectra coming from the (a) FHC neutrino beam and (b) RHC anti-neutrino beam. Here, the four quartiles have been combined back together into a single spectrum. The prediction was made at the best-fit point and includes a shaded systematic error band. The lower plots show the ratio of the ν_μ spectra to the unoscillated prediction.

	Neutrino beam		Antineutrino beam	
	ν_μ CC	ν_e CC	ν_μ CC	ν_e CC
$\nu_\mu \rightarrow \nu_\mu$	201.1	1.7	26.0	0.2
$\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu$	12.6	0.0	77.2	0.2
$\nu_\mu \rightarrow \nu_e$	0.1	59.0	0.0	2.3
$\bar{\nu}_\mu \rightarrow \bar{\nu}_e$	0.0	1.0	0.0	19.2
Beam $\nu_e + \bar{\nu}_e$	0.0	14.1	0.0	7.3
NC	2.6	6.3	0.8	2.2
Cosmic	5.0	3.1	0.9	1.6
Others	0.9	0.5	0.4	0.3
Signal	$214.1^{+14.4}_{-14.0}$	$59.0^{+2.5}_{-2.5}$	$103.4^{+7.0}_{-7.0}$	$19.2^{+0.6}_{-0.7}$
Background	$8.2^{+1.9}_{-1.7}$	$26.8^{+1.6}_{-1.7}$	$2.1^{+0.7}_{-0.7}$	$14.0^{+0.9}_{-1.0}$
Best fit	222.3	85.8	105.4	33.2
Observed	211	82	105	33

Table 4.2: Event counts at the FD for NOvA's 2020 analysis. Includes both the observed events and those predicted at the best-fit point.



(a)



(b)

Figure 4.13: The predicted and observed ν_e spectra coming from the (a) FHC neutrino beam and (b) RHC anti-neutrino beam. The prediction was made at the best-fit point and includes a shaded systematic error band.

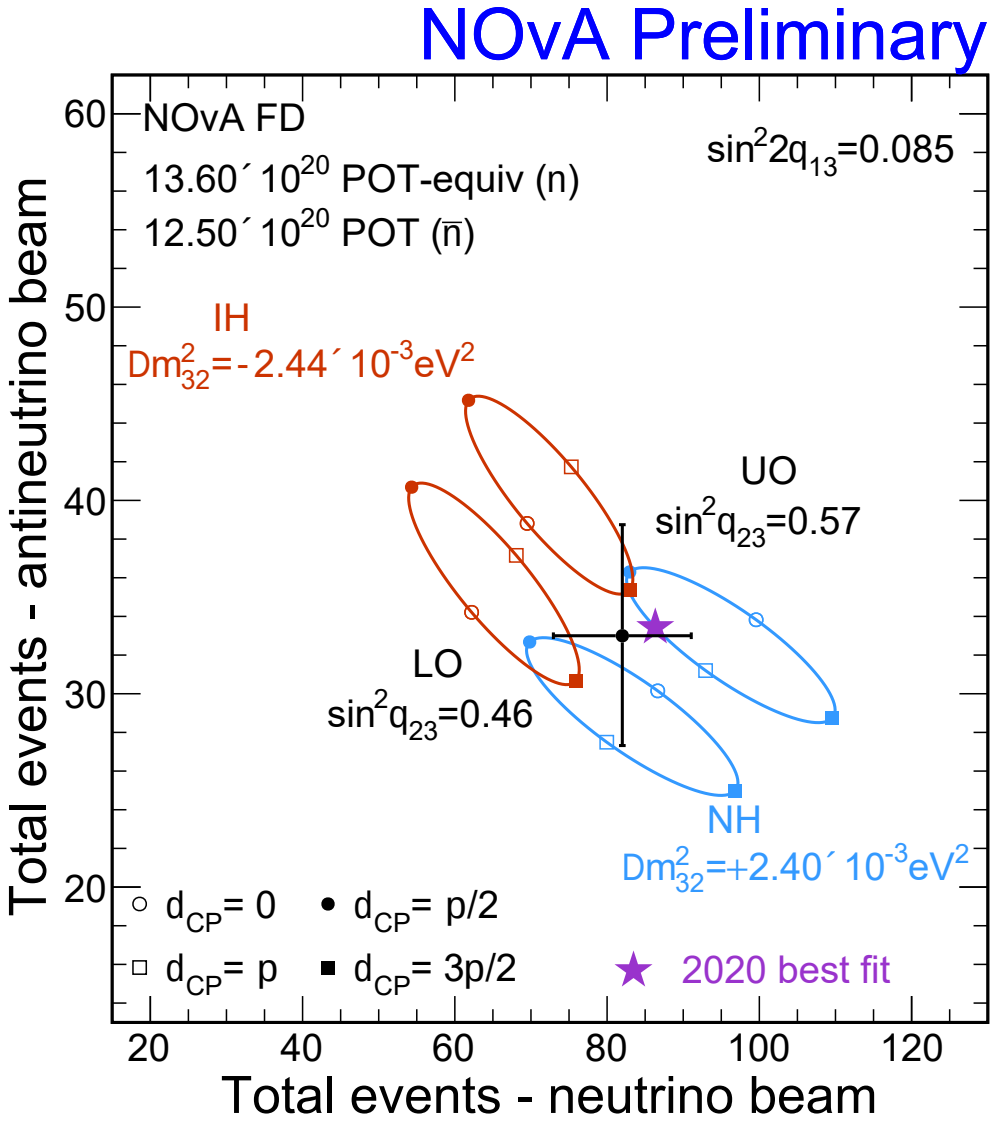


Figure 4.14: Bi-event plot showing the ellipses corresponding to the running conditions of NOvA's 2020 analysis. Also shown is the 2020 data and best-fit point.

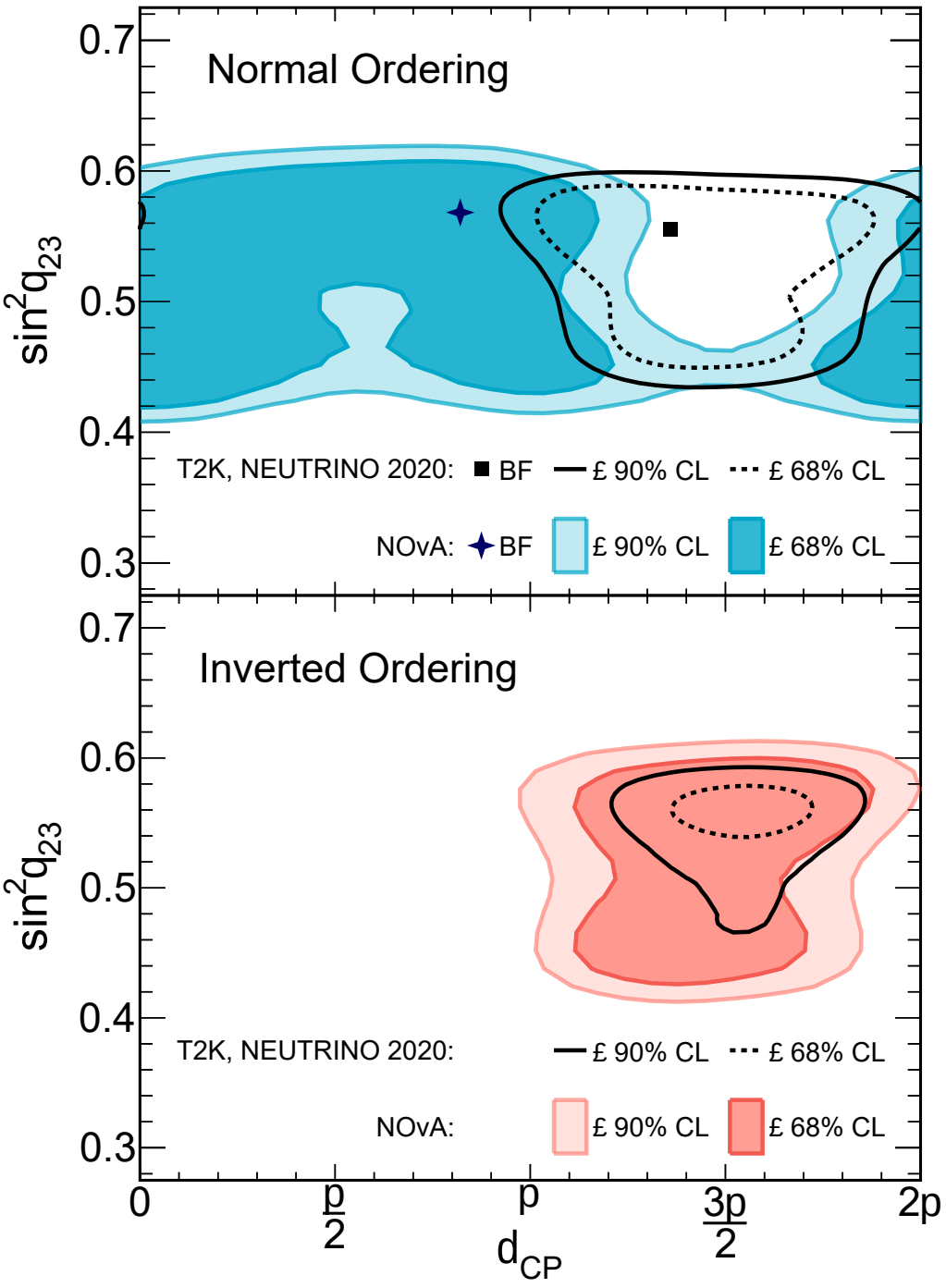


Figure 4.15: Confidence contours for θ_{23} vs. δ_{CP} for NOvA's 2020 analysis. Also included is the results that T2K presented at Neutrino 2020[42]. The top plot show the confidence intervals for the normal hierarchy while the bottom plot shows the inverted hierarchy.

Parameter	Normal Hier.		Inverted Hier.	
	UO	LO	UO	LO
$\Delta m^2_{32}(10^{-3}eV^2)$	+2.41 ± 0.07	+2.39	-2.45	-2.44
$\sin^2\theta_{23}$	$0.57^{0.03}_{-0.04}$	0.46	0.56	0.46
$\delta_{CP}(\pi)$	$0.82^{+0.27}_{-0.87}$	0.07	1.52	1.41
Rejection Significance	-	1.1σ	0.9σ	1.1σ

Table 4.3: Summary of oscillation parameters best0fit results for different choices of mass hierarchy and θ_{23} octant: either Upper Octant (UO) or Lower Octatn (LO). Also included is the uncertainties for the overall best-fit (normal hierarchy, UO) and the FC corrected significance with which we reject the other combinations.

in Fig.4.16. Notably, however, is that although T2K shows a preference for the normal mass hierarchy as we do, their best-fit point lies in our excluded region as can be seen in Fig.4.15. This apparent tension does not mean that our results are incompatible, as our regions of highest confidence in the inverted hierarchy overlap. A joint fit between the two experiments is being conducted to try to resolve this apparent tension[12].

[12]T2K's baseline is 295 km and their neutrino beam is peaked at 600 MeV, making it most sensitive to slightly different parts of parameter space than we are with our 810 km baseline and 2GeV peaked neutrino beam. Beyond the potential tension between our results, these complementary differences in sensitivities also motivate interest in a joint fit.

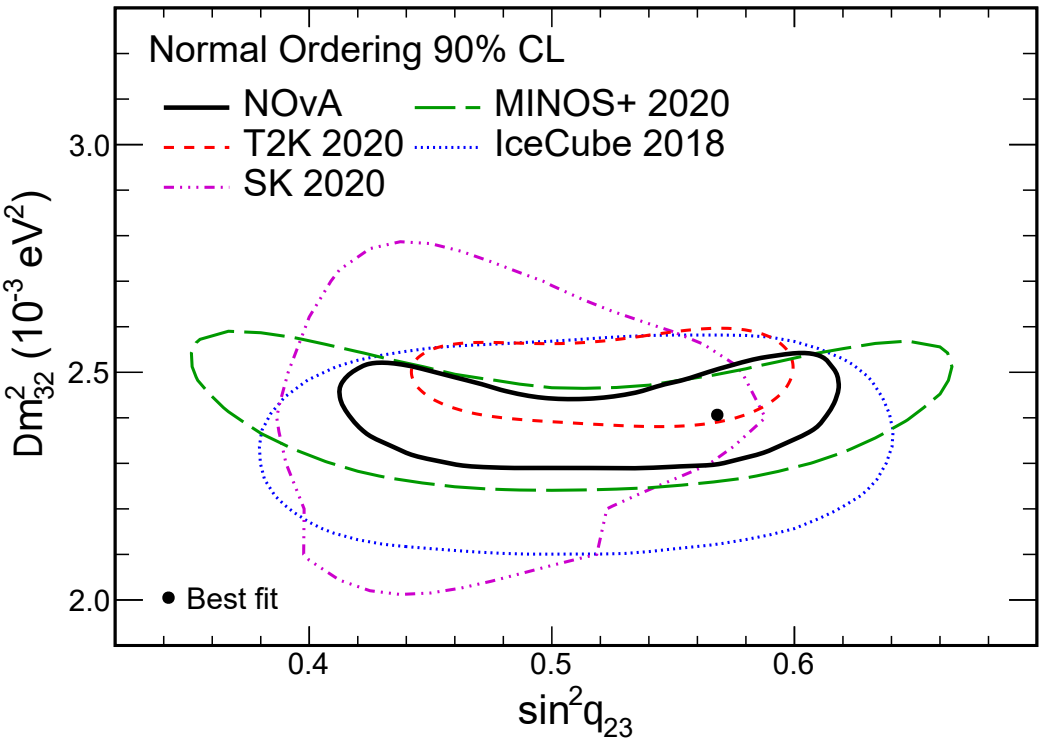


Figure 4.16: Normal Hierachy 90% confidence level contour for NOvA and other oscillation experiments[43][44][45][46], as well as NOvA's current best-fit point

5. Muon Removed (MR) Studies

NOvA's CNN used for electron (anti)neutrino identification and selection, is trained and tested on simulated MC samples. However, we don't use this CNN on just MC, but our data as well. To be able to use our CNN in our analyses, we must be confident that it performs the same on our data as it does on our MC. Otherwise comparisons of selected ν_e event counts between the two samples would be meaningless, and any fits of predictions to data would be skewed. Potential differences in performance would primarily come from mismodelling neutrino interactions when creating our MC, and would present as the CNN being better at identifying simulated ν_e events than real ν_e events coming from data.

The ideal way to check for performance differences would be to first get pure samples of ν_e events from both the MC and data. Then, we would compare the selection efficiency (ratio of PID selected events to all events passing preselection) of the CNN PID[1] for the MC sample, to that of a data sample. If the selection efficiencies match, any potential bias is apparently negligible and the CNN can be used in the analysis. On the other hand, if there is a significant discrepancy, further action would need to be taken, such as adjusting the simulation that the CNN was trained on, the creation or expansion of a systematic uncertainty to cover the discrepancy, or the creation of a correction to the MC selected sample based on the selection efficiency differences.

Unfortunately, getting a pure sample of events is easier said than done, especially from our data samples. For a ν_e sample in particular, there is no good source of ν_e events in our data that can be used for such a study, even if we could identify them without using the CNN we are trying to test. There are intrinsic beam ν_es in the ND samples, but they have an energy spectrum different from what we would expect from appeared ν_es in our FD, and they are a rather low statistic sample as well. The appeared ν_es at the FD are off-limits as NOvA is a blinded analysis, and those ν_es are our signal. Instead, we must create samples from MC and data comprised of events that are meant to imitate a ν_e or $\bar{\nu}_e$ event. This is where the Muon Removed (MR) Studies come into play. These studies take easily identifiable and abundant muonic events and remove the muon hits (hence the name). Further modifications are applied to create events that resemble a ν_e ($\bar{\nu}_e$) event. These processes are done identically with both data and simulation and selection efficiency

[1]For the rest of this thesis, the PID being referred to is the Event CNN PID, so I will use the terms "CNN" and "PID" interchangeably.

comparisons are made. Currently, there are three of these studies used in our analysis: MR Electron-Added to isolate and check the hadronic component of a typical ν_e event, and MR Bremsstrahlung and MR Decay-in-Flight to isolate and check the EM component.

5.1 MR Decay-in-Flight

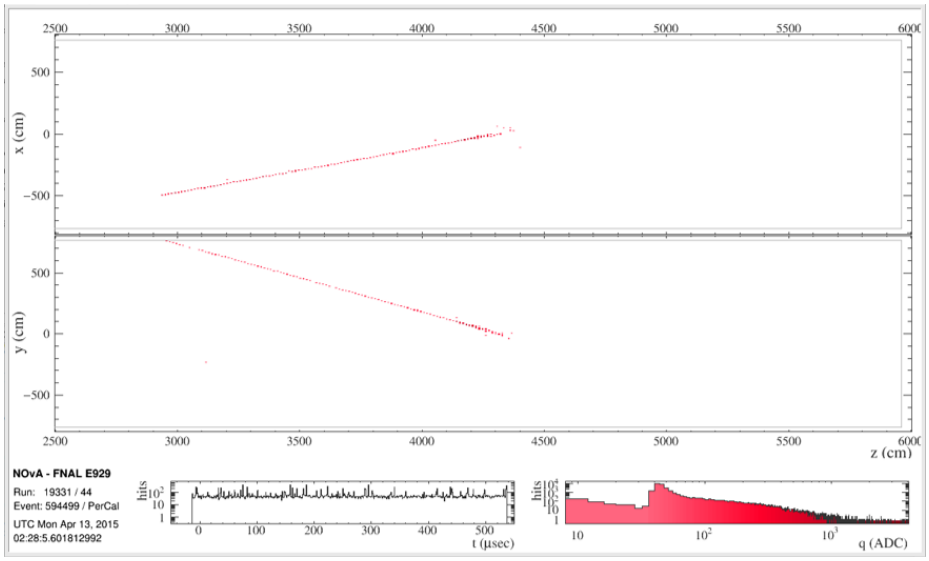
The MR Decay-in-Flight (MRDiF) study uses FD cosmic samples. The goal is to find cosmic muons that decayed in flight inside the detector, producing a pair of neutrinos and an electron which in turn induces an electromagnetic shower. Removing the muon hits prior to the decay leaves behind just the electron induced EM shower. This shower alone is enough for the CNN to recognize the event as a potential ν_e . Fig.5.1.1 shows a before and after event display of this process.

5.1.1 MRDiF Sample Creation

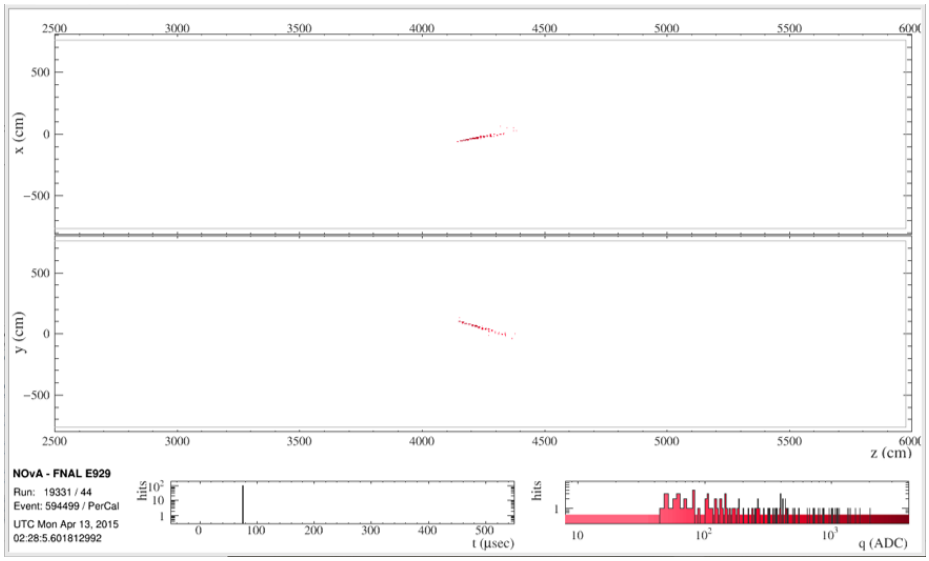
There are two sources of cosmic ray events that are pulled from to create the MRDiF samples. The data sample is created by selecting out muons from the large sample of cosmic activity that gets recorded by our FD. The FD records approximately 140 kHz of muons year round due to its size and location on the earth's surface. The simulation sample is pulled from a dedicated cosmic muon simulation that uses CRY as its event generator. Two full sets of each sample are created, one to correspond with the neutrino mode beam and one for the anti-neutrino beam mode. This is done not because the MRDiF samples themselves differ between the two modes, but the CNN does and we want to cross check both modes.

The selection of the MRDiF sample from either source is done using a algorithm called DiFShowerFinder. This algorithm looks at event slices that contain a single track that is considered 3D (has sufficient hits in both the x and y view) and applies some preliminary cuts. These are as follows:

- The end of the tracks are required to be contained inside the detector, ending no closer than 20cm to the detector edge so that we can capture the whole EM shower if there is one.
- There is an additional containment cut of the track end must be no closer than 50cm to any dead region of the detector to avoid false track ends.
- Tracks are required to have the cosine of the angle of the track to the beam direction



(a)



(b)

Figure 5.1: Event display of a cosmic muon that decayed in flight (a) before and (b) after of the Muon Removal process. The resulting event in (b) is an EM shower that the CNN could identify as a ν_e event. The color of the hits scales with the amount of energy deposited.

to be greater than 0.5 to have as horizontal events as we can to be closer to what the actual ν_e signal looks like.

- The track must traverse over 15 planes along its length. This cut was optimized via signal significance $\frac{s}{\sqrt{s+b}}$ using the truth data from the CRY sample to determine which events were signal or background.

Tracks that pass these preliminary cuts are then passed onto the actual DiF finding algorithm. This algorithm first assigns each plane along the track an average energy deposition value that is the average energy of that plane and the next two planes on either side resulting in a five plane average energy (5pE). Additionally, an estimate of what energy we would expect a MIP particle to deposit in that plane is calculated based on a MIP dE/dx of 0.00157 GeV/cm and an estimate of the path length through each cell based on the angle of the track and cell dimensions (similar to what is done during the tricell finding process in Calibration). After that, the algorithm goes plane by plane starting with the track start and looks for two consecutive planes that have an energy deposition that is 2.5 times that of the estimated MIP energy. This signifies the beginning of a shower and is marked as such. If a shower start is found, continuing along the track, the shower end is marked after five consecutive planes are found to all have a 5pE between 0.1 and 1.5 times the estimated MIP energy deposition.

If a shower is found with the shower start and end planes identified, some final clean-up and checks are done. Because the electron that resulted from the muon decay can still be considered a MIP for some distance before a shower structure appears, the shower start point needs to be adjusted to account for this. Starting from the original shower start point, we scan up the track until we hit a plane where the average energy of that plane with the next two planes upstream is ≤ 0.9 that of the expected MIP energy deposition. This plane becomes the new shower start point.

A similar adjustment process is done to the endpoint of the shower to help with the removal of background events consistent with Bremsstrahlung radiation. These background events have a muon track continuing through a photon induced shower and will have that muon track continue past the end of the shower. After the shower end has been adjusted, a further cut is done to remove events where the end of the shower is more than 26 planes away from the original track end, again to remove Bremsstrahlung radiation events. This cut was optimized via signal significance $\frac{s}{\sqrt{s+b}}$ using the truth values from the CRY sample to determine which events were signal or background.

Finally, all the energy deposited in the event between the adjusted starting plane and ending plane is added up, and a cut is applied requiring that the total energy be greater than 0.5 GeV.

Once a shower has been identified and the shower start and end planes properly identified, all the hits between those two planes that are less than 80 cm from the original cosmic track are saved to file. These hits form the MRDiF events, ready to be evaluated by the CNN. Again, Fig.5.1 shows the before and after of this process in an event display.

DiF events can also be identified in the simulated sample by looking at the truth values and saved particle histories saved during the event generation process. Doing this means we can identify DiF events by looking for events where the final daughter particle was an electron, that the electron was created via decay, that the parent was a muon, and that the energy of the electron was greater than 0.6 GeV to filter out Michel electrons. We can compare the number of events that are selected with this method to the amount selected with the first method described above to estimate a purity and efficiency of the created DiF samples. It was found that the nominal purity is roughly 97% and the nominal efficiency of being able to select DiFs from the cosmic sample is roughly 95%.

5.1.2 MRDiF Analysis

The goal of any of the MR* analyses are to compare the selection efficiency of the ν_e selection process between data and simulation. So after creating the MR* samples for both data and simulation, the next step is to define and apply preselection and full selection cuts. These cuts follow closely with the main analysis cuts with the preselection cuts including quality cuts, containment cuts, and other cuts designed to select out potential ν_e candidates. The full selection cuts consist of all the preselection cuts and then the final PID cut based on the CNN evaluation of each event. Additionally, because NOvA is a blinded analysis and this is a study that would be run prior to any box opening, a timing cut is applied to make sure we are not looking at FD events inside the beam spill time window as they could be our actual signal events. Once these cuts are applied, we have two sets of events: preselected events and fully selected events, of which the latter is a subset of the former.

One final step needs to be done before the selection efficiencies can be found and compared. Even though there is a cut on the track angle to try to get events that are more horizontal than vertical during the MRDiF finding process, the angle distribution of showers is still quite different for DiFs than that of the FD signal ν_e 's. This is just due to the fact that

cosmic muons are coming down from above the detector while the signal events are coming in from the side of the detector from the nearly horizontal beam. We know that the CNN is sensitive to the angle to the beam, so to further make this MRDiF sample resemble the signal ν_e sample, we reweight the events by $cos\theta_{beam}$ so that the preselected event distributions match the shape of simulated ν_e signal in the MC. Separate weights are created for the data and the CRY and these weights are then also applied to the fully selected events. Fig.5.2 shows the angle distribution of all three samples (data MRDiF, CRY MRDiF, and MC ν_e signal) before and after the reweighting process for the fully selected events for the NOvA's 2020 analysis. Note that the final bin was excluded (< 0.9667) for this analysis. This was due to overly low statistics in the CRY sample.[2]

After reweighting, various plots are made to compare several variables between the MRDiF and the MC ν_e signal. These plots can help identify fundamental differences between the data, MC, and CRY. A selection of these plots from the 2020 analysis can be seen in Fig.5.3. From these plots, it is clear that the data events and the FD MC events are similar in size and energy distribution, but the CRY sample trends towards smaller but more energy dense EM showers. This discrepancy is rooted in differences in how CRY models and generates muons in our detectors as compared to GENIE generated MC and the actual data. These differences do not keep this MRDiF study from being meaningful, but does limit its strength and scope. Different generation methods for the cosmic muons sample are being explored for future analyses.

Finally, the selection efficiencies are made by taking the ratio of the fully selected event spectrum to the preselected event spectrum for both the data and the CRY. These can be seen in Fig.5.4 for Ana2020. The selection efficiencies agree within error for both the neutrino beam and anti-neutrino beam mode, meaning that is not significant bias in our ν_e selector between data and simulation coming from mismodeling in our EM shower simulation [3].

[2]The CRY sample is a highly computing resource intensive sample to make and is not very efficient, which is the reason that the MRDiF simulated sample was much smaller than desired. This 2020 analysis was the first large scale production of the MRDiF sample, and we had overestimated the rate of the DiF events being generated (this rate cannot be directly controlled), but could not afford the time or resources to generate more. There is work being done in the collaboration to try to implement and understand other generators and new ways to simulate specific muon events in the detector that could help with this problem in the future.

[3]At the time that these plots were made for the 2020 3-flavor analysis, the treatment of the statistical error on the efficiency was done using Poisson statistics. As discussed in App.C and Sec.6.3.2, this is not the correct treatment of the statistical uncertainty of an efficiency as it does not take into account that the numerator is correlated to the denominator. This was discovered while writing this thesis and the necessary files were not available to remake these plots with the proper uncertainty treatment. However, the total event counts were available, and they show that the statistical errors would be reduced to about a fifth of what is shown in the plot. The relevant code has been updated so future analyses will do the correct uncertainty calculations.

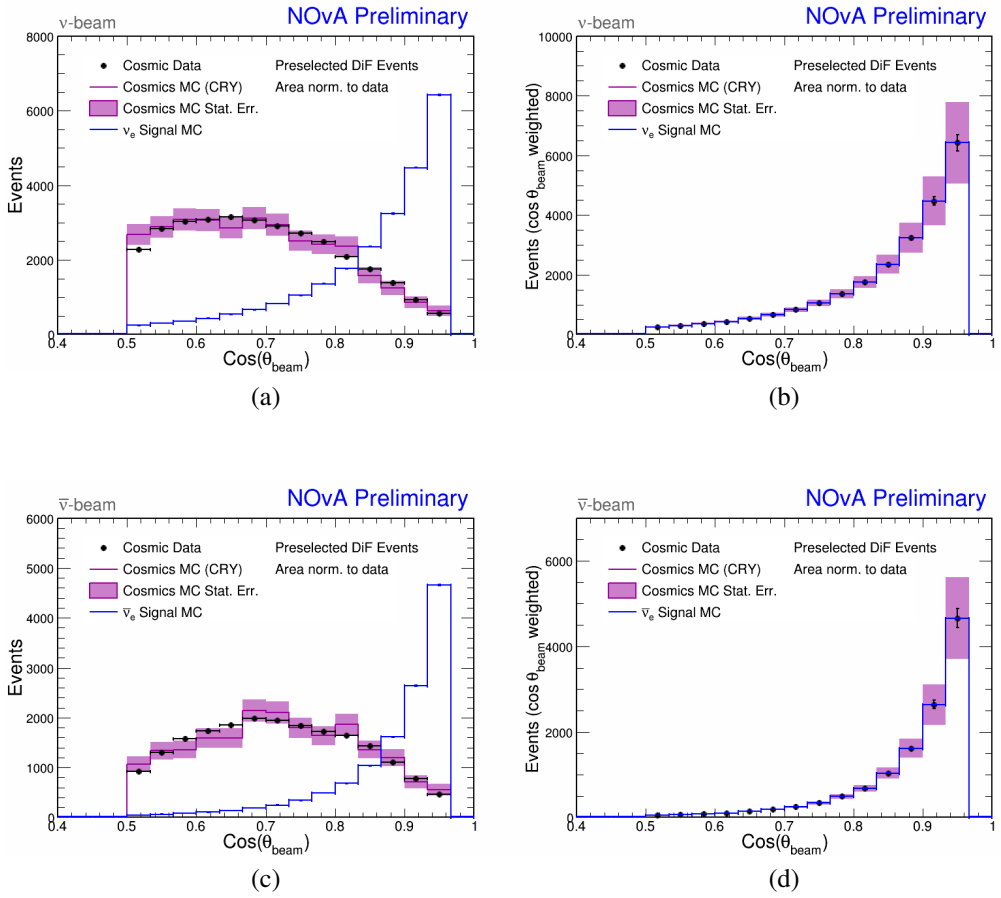


Figure 5.2: Plots of $cos(\theta_{beam})$ of the preselected DiF showers from both data and simulation (CRY) and preselected simulated FD ν_e events (MC). The top (bottom) row is the FHC (RHC) selections while the left (right) column shows the before (after) of the reweighting process for the DiF events. The simulated samples were area normalized to the data sample.

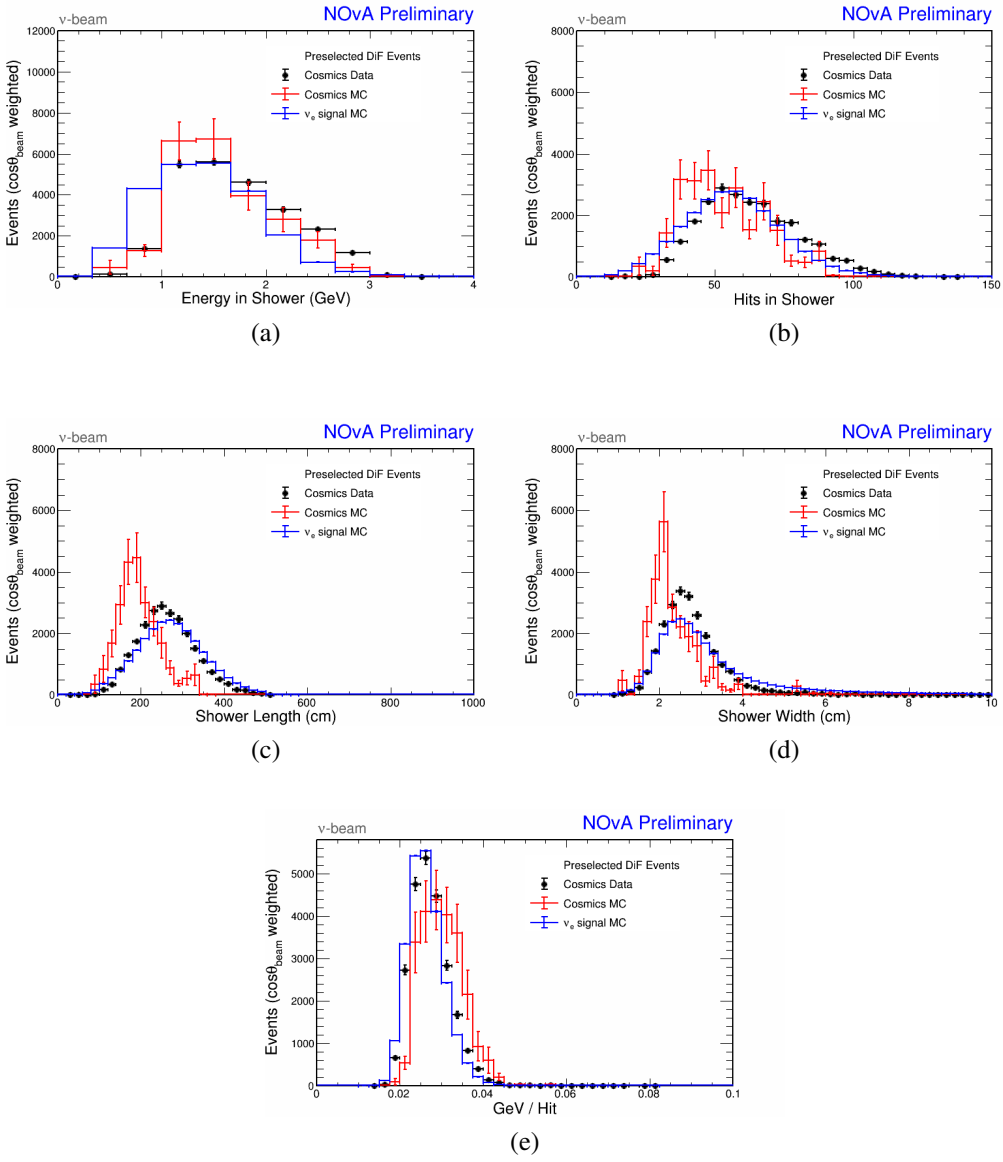
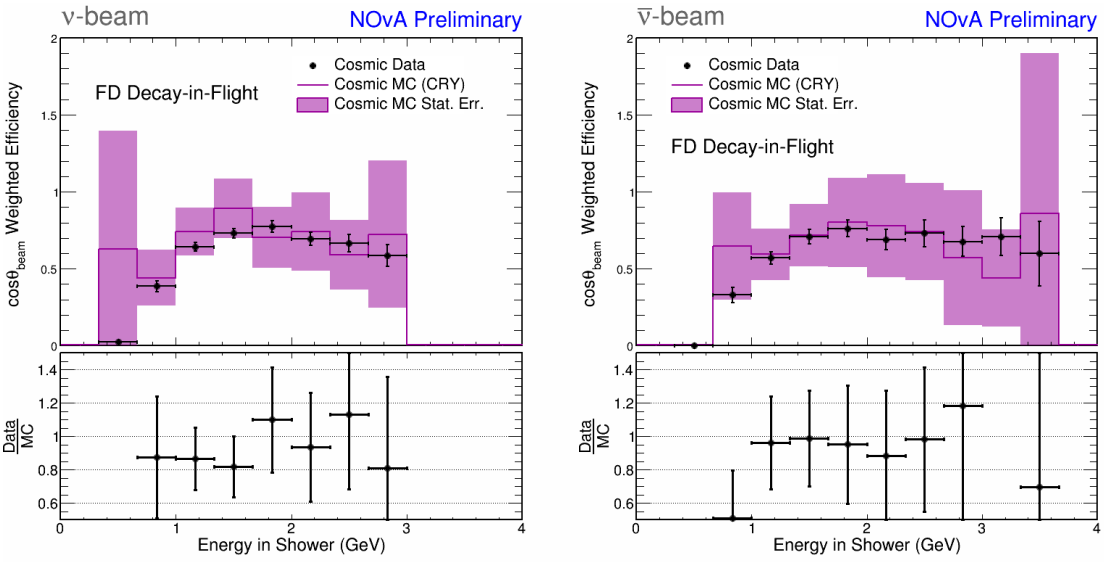


Figure 5.3: Comparisons of the preselected events from the FHC DiF data (black), DiF simulation (red), and FD MC (blue) of various variables. All plots have the DiF samples reweighted by $cos(\theta_{beam})$. All plots were area normalized to the DiF data to highlight shape differences. Error bars shown are statistical.



(a) (b)

Figure 5.4: Selection Efficiencies of the 2020 MRDiF samples with the data samples in black and the simulation (CRY) samples in purple. The $cos(\theta_{beam})$ weights were applied prior to the efficiency calculation. (a) shows the FHC beam mode results and (b) shows the RHC beam mode.

5.1.3 MRDiF at the ND

Although MRDiF was created to be conducted using cosmics at the FD, there is, in theory, no reason why it can't also be conducted using rock muons in the ND. Rock muons refer to muons coming into the near detector that were created by beam interactions in the rock surrounding the detector. These tend to fall within the required energy spectrum and have the ability to decay in flight within the detector. If a MRDiF sample could be made from these rock muons, we could not only compare the selection efficiency of the CNN between data and simulation, but also between the near and far detectors.

For the 2020 analysis, the MRDiF code was modified to allow it to run on ND files. This primarily consisted of turning hard coded values into adjustable parameters and adding new options to account for the differences in geometry and size in the two detectors. The algorithm was run over ND simulation and included a flag to say if an event was a true DiF based on simulation truth values. Using this flag to calculate the purity of the sample, it was determined that the out-of-the-box running of the MRDiF process on ND files resulted in a sample with vastly less than 1% purity. Looking at event displays for the background events, it became clear that the module was selecting almost all the EM shower-like events it saw regardless of whether it included a long muon track or not. The events that make up the data and simulation used at the ND are primarily beam neutrino interactions, most of which have a shower-like component which the MRDiF was picking up. Interestingly, there was also a second source of DiFs found when looking through the events displays like the one in Fig.5.5. Here, the muon coming out of a ν_μ event travels for a short while before decaying in flight, creating a large EM shower. These types of events would be useful to include in a ND DiF sample, but their selection process may need to run parallel to the selection of rock muon DiFs so the selection process can be optimised to both.

To attempt to reduce the overwhelming background found in the out-of-the-box running of the DiF module on the ND, a series of plots were made to look at the signal selected events to the background events in the MC. Signal events were events selected by the DiF algorithm that could be determined to be "true DiFs" via the truth values inside the MC, and the background events were all the rest of the events selected by the DiF algorithm. This led to a retune of some of the initial quality cuts described in Sec.5.1.1, the addition of a cut on the number of planes between the start of the track and the start of the shower, a requirement that the track starts outside the detector, a limit on the number of prongs allowed in the parent slice, and a cut on the percentage of the parent slice hits are included

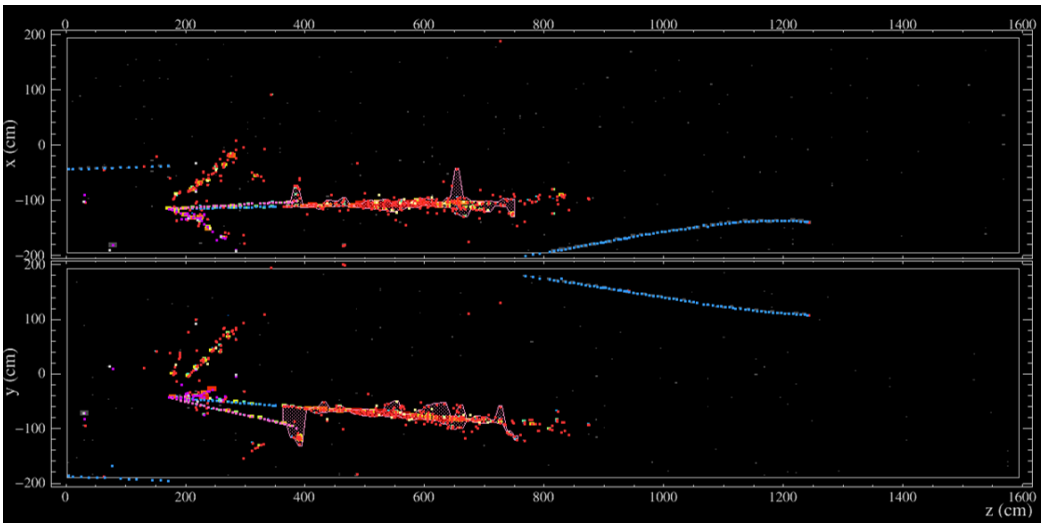


Figure 5.5: Event display of a ν_μ event where the outgoing muon decayed in flight to produce an EM shower that was in turn selected as a DiF. Here the event is overlayed with the "true" hits that are colored to represent the particle making those hits. The blue hits represent a muon, red an electron and its shower, and pink is hadronic activity.

in the DiF slice. the following table shows the cut flow of the application of each of these additional cuts.

Cut	Signal Events	Background Events	Signal/Background
Base cuts	5619	2.40×10^6	0.0023
Planes in parent slice > 15	5618	2.40×10^6	0.0023
Prongs in parent slice < 3	4958	1.23×10^6	0.0040
Track tail < 55 planes	4897	8437441	0.0058
Hits ratio < 0.75	3499	451082	0.0078
Track start outside detector	3310	336797	0.0098
Track neck > 5	3035	285634	0.0106

Table 5.1: Cutflow of the additional cuts applied to the ND DiF sample in an attempt to improve the sample purity. Track tail (neck) refers to the number of planes between the track end (start) and the shower end (start). Hits ratio refers to the ratio of hits included in the DiF slice to the number of hits in the parent slice.

Table 5.1 shows that even after the addition of these cuts and tunes, the purity of the sample was unfortunately only increased to be about 1%. Furthermore, it appeared as if the efficiency of the selection process was potentially as low as 3% by comparing the number of events that should have been flagged as a DiF according to MC truth values to the number of these "true DiFs" that were actually selected by the algorithm. It was determined that

a different approach would be required to be able to conduct the MRDiF study at the ND and was left to future studies. Potential approaches would be to either redesign the MRDiF module entirely to identify and ignore neutrino events at the start of the DiF finding process, or to create a BDT to better define the phase space with the highest purity of true DiF events. The drawbacks of both of these processes, however, would be that we would no longer have identical processes between the near and far detectors which would complicate (but not make impossible) comparisons of selection efficiencies between the two.

5.2 MR Bremsstrahlung

MR Bremsstrahlung (MRBrem) is similar in nature and procedure to the MRDiF study. It also takes advantage of the high rate of cosmic muons in the FD, but instead of looking for muons that decayed in flight, MRBrem looks for muons that experience Bremsstrahlung radiation. This radiation - caused by deceleration of the muon - comes in the form of photons which then create an electromagnetic shower inside the detector. MRBrem finds these showers and removes the muonic hits from within and from either side of the shower to create another sample of pure EM events. The advantage of MRBrem over MRDiF is that the rate of occurrence of Bremsstrahlung radiation is greater than that of DiF events, making this a statistically strong sample. The main disadvantages are twofold. First, the EM showers are photon induced instead of electron induced which our CNN could be sensitive to. Second, we have to remove the muon hits from within the shower itself, a step that isn't needed for the MRDiF sample. Nonetheless, the MRBrem sample is a great tool to probe for bias within our CNN, especially when done in conjunction with the other MR* studies.

The MRBrem sample creation process is initially very similar to that of the MRDiFs. Samples are made from the cosmic muon data and from CRY for both beam modes. Like the MRDiF process, the MRBrem process starts with looking for candidate tracks. These tracks must satisfy the following:

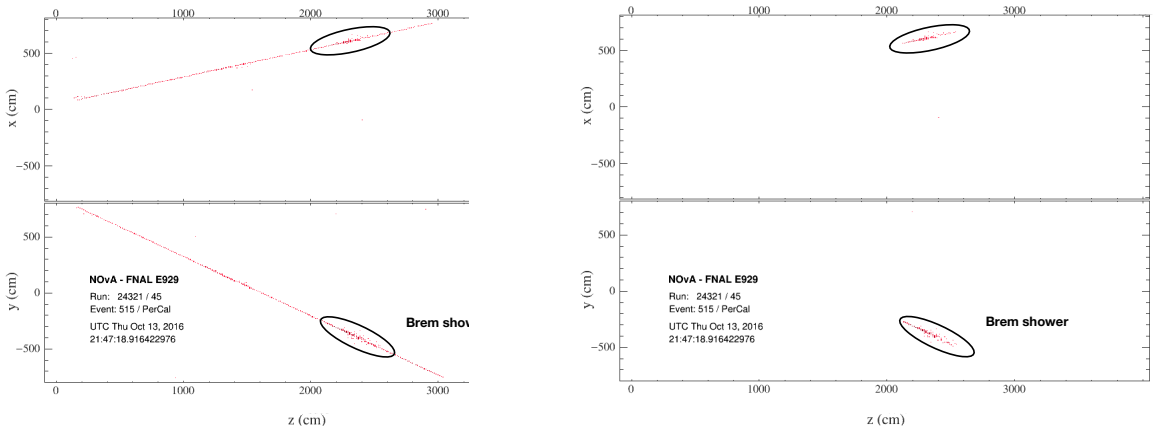
- Must be the only track in the event and must be a 3D track.
- The track must start outside the detector to ensure it is a cosmic muon.
- The tracks must have the cosine to the angle of the track to the beam be greater than 0.5 to select mostly horizontal events.
- The track must traverse at least 30 planes along its length.

After these initial track quality cuts, the same energy mapping process is done as that of the DiFs. However, to look for a MRBrem shower, slightly different parameters are used to find the showers and by extension, their start and end points. First, the module looks for 5 consecutive planes that have energy deposition equal to at least twice that of what we would expect a MIP to deposit. It then looks for 5 consecutive planes whose five plane average energy is between 0.5 and 1.5 that of a MIP. If a shower is found, then a final cut is applied requiring the shower energy to be greater than 0.5 GeV.

5.2.1 Muon Removal Process

Once showers have been identified, the muon hits are removed from within the shower. This is done with a separate package that was developed to remove muon hits from within various event types. This MuonRemove package first identifies the region of interest (here this is the EM shower) and maps out energy deposition similar to what has been described before. Going plane by plane in the region of interest, the energy deposition is compared to that of a MIP. If the plane has energy less than or equal to that of a MIP, the hits in that plane are all given a weight of 1 which is the signal for it be removed later in the process. Planes that have energy greater than a MIP are looked at cell by cell in order from the closest to the track outward. If the closest cell has energy equal to that of a MIP, that hit is given a weight of 1 and the rest a weight of 0 which is signal to be left alone as is. If the cell is less than that of a MIP, it is given a weight of 1 and the next closest cell is looked at. If the combined energy of these two cells are at least that of a MIP, they are both flagged for removal and the rest of the cell's hits are flagged to be left as is. Lastly, if the first cell has energy greater than that of a MIP, it is given a weight equal to the ratio of a MIP's expected energy in that cell to the energy actually deposited in that cell. The rest of the hits in that plane are flagged to be left as is. Hits outside the region of interest are assumed to be muon track hits and also flagged to be removed.

After all the hits have been assigned a weight, they are looped through one last time. If they have a weight of 1, that are skipped. If they have a weight of 0, they are added into the final product as is. If they have a weight between 0 and 1, the ADC values of the hit (which is what is used to calculate the energy in the hit as described in 3.3.1) is adjusted by multiplying it by (1-weight). This effectively removes the amount of energy attributed to the muon from the hit, leaving behind the energy attributed to the EM shower. The adjusted hit is then added to the final product. Finally, we are left with just the EM shower. Fig.5.6



(a) (b)

Figure 5.6: Event display of a cosmic muon that experienced Bremsstrahlung radiation (a) before and (b) after of the Muon Removal process. The resulting event in (b) is an EM shower that the CNN could identify as a ν_e event. The color of the hits scales with the amount of energy deposited.

shows the before and after of this process for a sample event in the and event display.

5.2.2 MRBrem Analysis

Once the samples have been created, the analysis procedure for MRBrem is the same as that for MRDiF. Fig.5.7 shows the effects of reweighting by shower angle for NOvA's 2018 analysis which was the most recent analysis that successfully included MRBrem. Fig.5.8 compares the selection efficiencies of the MRBrem samples. The selection efficiencies agree with each other, especially in the signal region of around 2 GeV. This again shows a lack of significant bias in our CNN caused by mismodelling in the EM sector. Although MRBrem was not used in our 2020 analysis due to some technical issues, it is planned to be used again in future analyses. Additionally, there are plans to adapt it to the ND in a similar fashion to how the attempt to adapt the MRDiF to the ND. The thought is that because the muon tracks extend out from the end of the shower for the Bremsstrahlung events in a way that they do not for the DiFs, it may be easier to remove the large background from the Bremsstrahlung sample at the ND where we could not for the DiF sample.

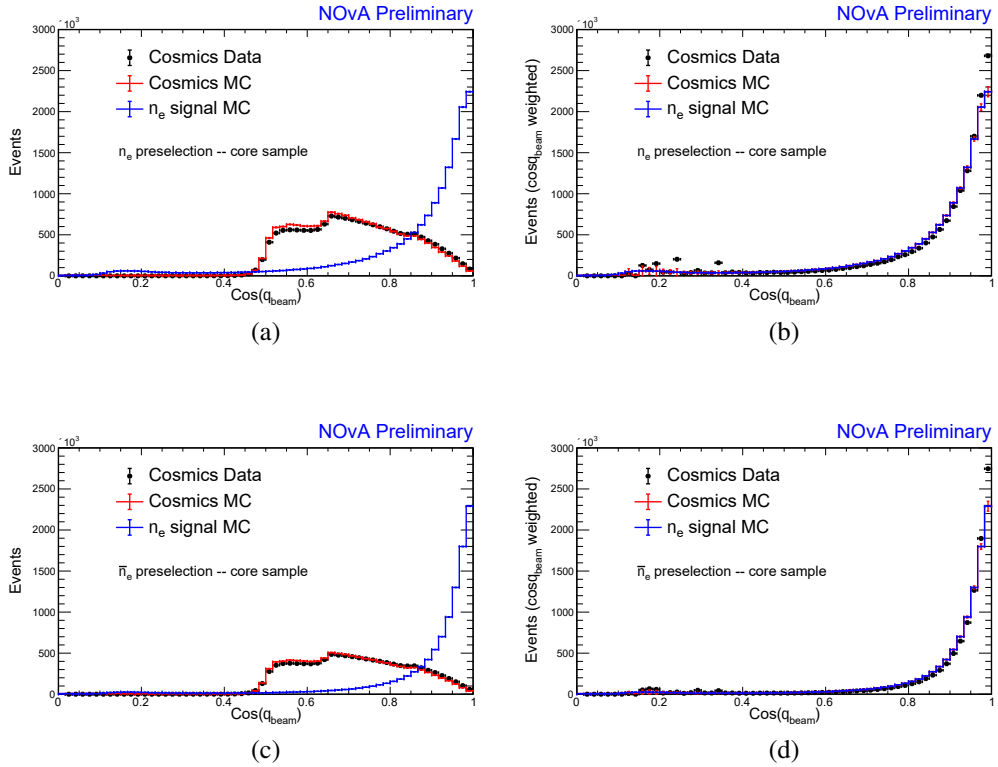
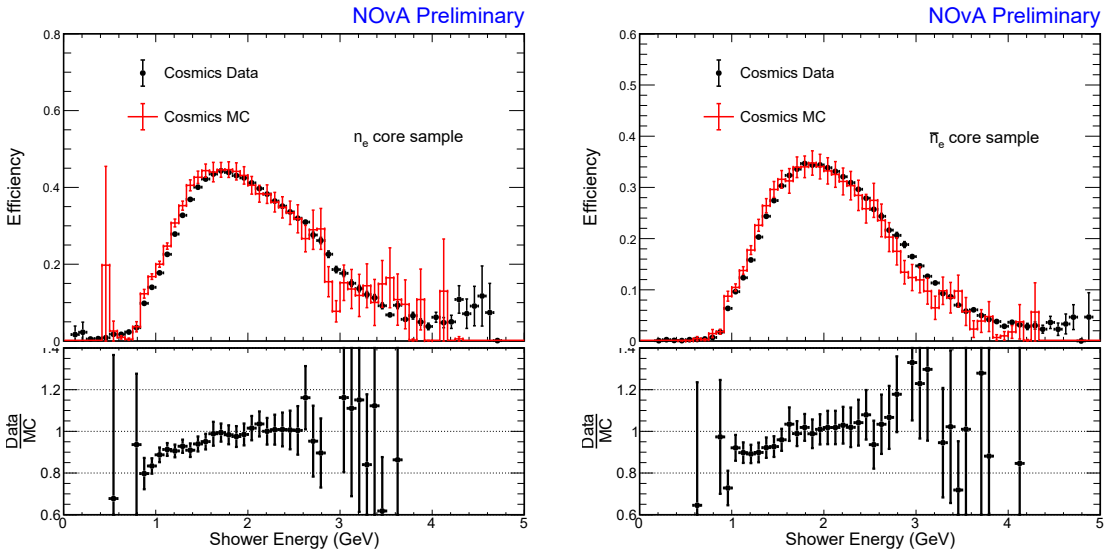


Figure 5.7: Plots of $cos(\theta_{beam})$ of the preselected Bremsstrahlung showers from both data and simulation (CRY) and preselected simulated FD ν_e events (MC). The top (bottom) row is the FHC (RHC) selections while the left (right) column shows the before (after) of the reweighting process for the Bremsstrahlung events. The simulated (MC) samples were normalized to the cosmic data sample.



(a) (b)

Figure 5.8: The selection efficiencies of the 2018 MRBrem samples as a function of EM shower energy and the ratio of the data to the simulation efficiencies on the bottom plots. The $cos(\theta_{beam}$ weights were applied prior to the efficiency calculations. (a) shows the results for the FHC beam mode (b) shows the RHC beam mode.

5.3 MR Electron-Added

Whereas the MRDiF and MRBrem studies probed the EM shower component of a typical ν_e event, MR Electron-Added (MRE) looks to probe the other component: the hadronic activity. This study turns to the ND and its abundant supply of ν_μ charged current events as its source of parent events. These ν_μ events are characterized by a long muon track and a hadronic shower. MRE takes these events, removes the muon and adds in a simulated electron of the same energy and direction as the removed muon. The result is an event that resembles a ν_e charged current event, but where differences between data and simulation have been isolated to just the hadronic component of the event.

5.3.1 Creating the MRE Sample

MRE looks at reconstructed NuMI beam files from the ND (both from data and from our nominal MC) to select out candidate ν_μ and $\bar{\nu}_\mu$ charged current events depending on the beam mode. The selection criteria require that the event have a valid vertex and at least one prong, and that there is a muon track with energy greater than zero. Further selections to ensure that the event is indeed a ν_μ CC event are not applied until the actual analysis is being conducted to allow for later adjustments to the ν_μ selection criteria as that selection may change from analysis to analysis. These events will typically resemble the first frame in Fig.5.9 after the ν_μ selection is applied. In this example, a short proton track makes up the hadronic portion of the ν_μ event.

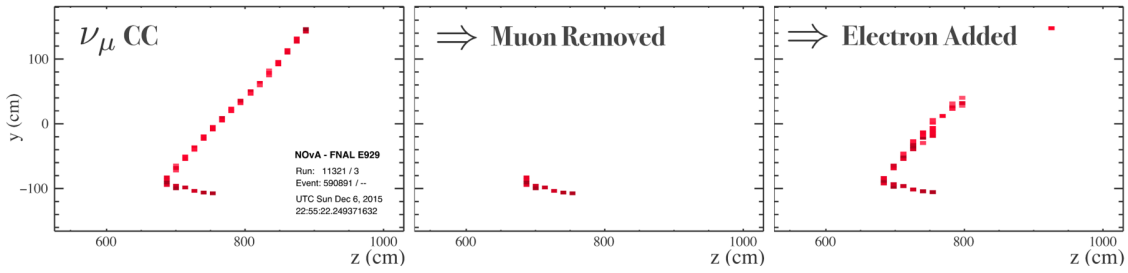


Figure 5.9: Before, during, and after of the MRE sample creation process. The first frame shows a typical starting ν_μ event. The middle frame shows the effect of removing the muon hits, and the final frame shows the addition of a simulated electron. Color of the hits scale with energy deposition.

If a slice passes the above preselection and a filter for rock muons[4], the next step is to select the most muon-like track. This can be done by a few different methods, but they all rely on some reconstructed value (ie. the longest track) or on some other specialized selection algorithm (ie. Prong CNN or RemID). This track is removed following the same method described in Sec.5.2.1 except this time, the area of interest is the vertex region of the event where the hadronic activity occurs. Additionally, all the information about the removed muon is stored for use in the next step of the MRE process: adding in a simulated electron (hence the name of this study).

GEANT is used to simulate electrons in each event that had a muon removed. The electron is generated with the following inputs taken from the saved information about the removed muon: the muon track starting point, the muon's direction, and the muon's energy. These hits are then overlayed onto the muon-removed event, combining overlapping hits into a single hit by creating a new hit with a carefully constructed ADC pulse that effectively added the energies together. We now have a sample of MRE events.

5.3.2 MRE Analysis

Similar to the MRDiF and MRBrem studies, the actual analysis part of the MRE study defines a set of preselection cuts and a set of full selection cuts to be applied to both the data and MC samples. These samples are then used to make various data to MC comparisons as well as selection efficiencies by taking the ratio of the fully selected event spectra to the preselected event spectra. No reweighting to the event angle to the beam is required, however, since MRE starts out as beam events and are therefore already oriented in the correct direction.

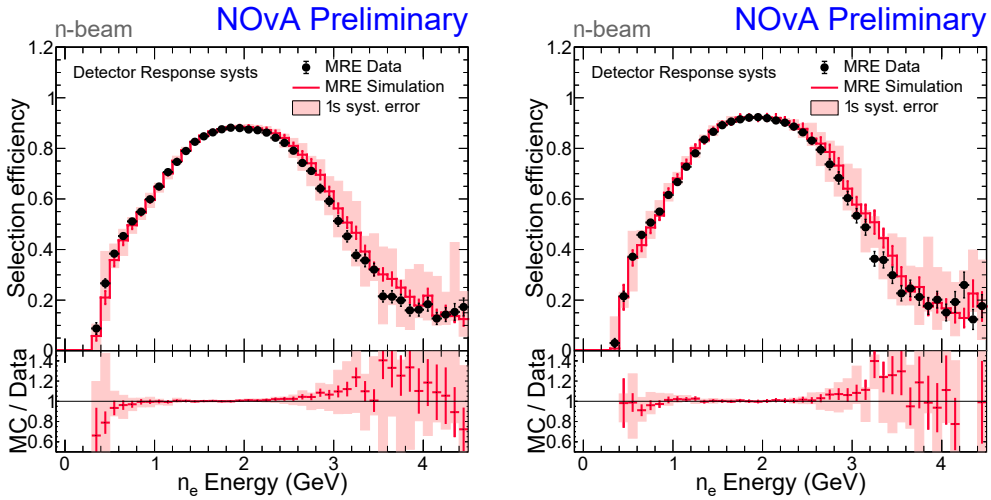
In addition to the main analysis' ν_e ($\bar{\nu}_e$) preselection and full selection cuts, MRE also needs to apply cuts to ensure that only events that started out as ν_μ ($\bar{\nu}_\mu$) events are included. These ν_μ cuts follow the main analysis cuts to select for ν_μ events and must be applied to the original parent slice, not the new MRE event. This is done by matching the MRE event to its parent event via a chain of associations that were created throughout the entire MRE process and saved for each event.

Lastly, it is possible to create systematically shifted versions of the MRE MC sample, meaning that the MRE analysis can include systematic uncertainties as well as statistical

[4]Rock muons are muons that were created by particles from the beamline interacting with the rock surrounding the ND that then enter the detector. They are usually considered a background similar to cosmic muons.

uncertainties in its analysis. These samples are created in the same way as the nominal MC, except that the initial MC and the simulated electron are generated with a systematic shift applied to the simulation generation process, such as by shifting calibration constants up or down by 1-σ. The systematic uncertainties used to make these additional samples are all associated with uncertainties in our detector response including calibration, light level modelling, and Cherenkov light modelling.

Fig.5.10 and Table 5.2 report the MRE results for NOvA's 2020 analysis. As the plots show, in both RHC and FHC, there is great agreement between data and simulation, particularly in the region of interest around 2 GeV. The table shows that the overall difference in both beam modes was less than 1%, again showing that there is no significant difference in the performance of our ν_e PID between data and simulation.



(a) (b)

Figure 5.10: Data and MC selection efficiency (full selection / preselection) for the MRE events in (a) FHC beam mode and (b) RHC beam mode. The MC sample is shown in red with the the shaded band indicating the detector response related systematic uncertainties. The data sample is in black. Below id the ratio of the MC efficiency to the data efficiency.

		Preselection	Full Selection	Efficiency	Difference
Neutrino Beam	Data	709112	564669	0.796	-0.76 %
9.5×10^{20} POT	MC	772566	619908	0.802	
Antineutrino Beam	Data	418245	348151	0.832	-0.46%
11.8×10^{20} POT	MC	475300	397454	0.836	

Table 5.2: Table with MRE integrals for RHC and FHC modes for NOvA's 2020 analysis. The ratio Full Selection/Preselection produces efficiencies for Data and MC and their differences are calculated with respect to MC.

6. MRE Corrections

Because the MR* samples are designed to be highly pure ν_e -like samples, they can be useful for even more studies than the already described validation of our PID performance. Early analyses used the MRE sample to help identify differences in the MC and the data, and the MRDiF and MRBrem studies have the potential to provide detector to detector cross checks once they are successfully run on ND data and simulation. The most impactful potential that the MR* studies have, however, is the ability to extend their use from cross checking CNN performance to actually correcting the predicted ν_e signal. This chapter will focus on the process of using the MRE sample to do just that for the FD ν_e signal prediction used in our main 3-flavor analysis.

6.1 Motivation

As described in Sec.4.4.1, NOvA's main 3-flavor analysis includes data-driven corrections to the ND simulation sample that is then extrapolated out to the FD to become our predicted signal and background. These corrections allow us to be confident that we are starting out with a representative amount of ν_μ events, beam ν_e events, and other background events, compared to the data, that we can then apply our Far-to-Near ratio and oscillations to, resulting in our prediction. However, one aspect of the FD predicted signal doesn't show up until after that process: the appeared ν_es . Because there is no data sample that can be pulled from to directly correct this sample in the same way as the decomposition process, we turn to the MRE to instead. In a similar way that the decompositions use the difference between the data and MC samples to create correction weights to apply to the ND prediction, we can turn the difference between the selection efficiencies of the MRE data and MC samples to create corrections to be applied to the FD prediction.

The MRE sample was chosen instead of the MRDiF and MRBrem studies for several reasons. First, the MRE events are effectively indistinguishable from actual ν_e events whereas the cosmic muon based MR samples are just EM showers without an accompanying hadronic component. Secondly, the MRE sample is a statistically rich sample, especially in the signal region around 2 GeV. The MRE sample also has the added bonus of focusing on the hadronic component of ν_e events, the simulation of which we assign a higher level of

uncertainty to than to our EM shower simulation. While the MRE sample is the preferred sample to make this correction with, it does not mean that future studies could not also include the MRDiF and MRBrem studies, even if significant care would need to be taken to properly account for the shortfalls of those samples.

For this thesis, there were two different sets of MRE samples used to make and apply corrections. The first sample was the one used for NOvA's 2020 analysis and will be referred to as the Prod5 sample. The second sample will be referred to as the Prod5.1 sample. This is a sample that was made with 0.2 times the POT as the Prod5 sample and includes a new light level tune, improvements to several aspects of the simulation and reconstruction, and the addition of new variables in the final output files. This updated production of all our data and simulation files is planned to be used for our next analysis. For the MRE samples themselves, the actual differences between the shapes of distributions for the Prod 5 and Prod5.1 MRE sample sets are slight, with the main difference (for the purposes of this thesis) being in the available statistics.

6.2 Acceptance Differences

The main drawback to using the MRE sample to make this correction comes from the fact that it is a sample made from exclusively ND data and we want to apply the correction to FD events. While the actual topology and appearance of the events should not differ between the two detectors, the kinematics do vary in several key ways. The ND is much smaller than the FD meaning that it is less likely to be able to contain larger events as illustrated in Fig.6.1. These FD events tend to have a larger angle between the EM shower

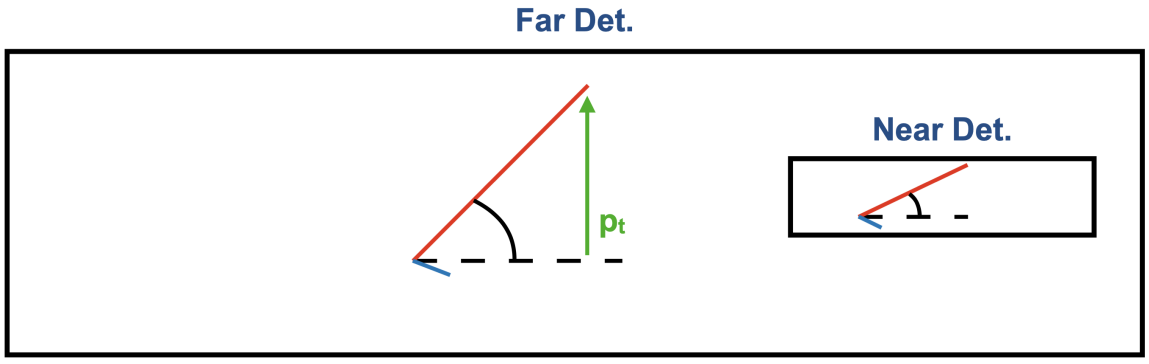


Figure 6.1: Scale drawing of the FD and ND with examples of neutrino events to highlight the acceptance differences between the two detectors caused by their difference in size.

and the beam direction, larger transverse momentum, and/or larger outgoing lepton energy resulting in a longer EM shower or muon track. Because these are all things that the PID can potentially be sensitive to, we want to make sure that any PID-based corrections are only applied to events that are in the same regions of kinematic phase-space as the events that the corrections were derived from.

The first step to taking into account these differences in the event kinematics, is to characterise them. While we do characterise these acceptance differences for ν_μs and beam ν_es quite well as part of the extrapolation procedure, we can't just assume that the differences would be the same for the MRE sample, as it is kinematically a mix of both. Remember that the MRE events are made from selected ν_μs in the ND. Because the length of a muon track is typically longer in our detectors than an EM shower of the same energy, the parent ν_μ events are less likely to be contained than the resultant MRE event. If the parent event fails containment, the resulting MRE event is not included in the final sample, even if that MRE event did pass containment[1]. This results in an MRE sample that effectively has a more harsh containment applied to it than even the beam ν_es in the ND, let alone the signal ν_es at the FD. This is illustrated in Fig.6.2.

As mentioned already, the angle of the EM shower with respect to the beam direction is another indicator of acceptance differences. Fig.6.3 shows the distribution of the cosine of this angle for the MRE samples and the FD MC ν_e sample. In this case, all distributions peak in the same place, but the FD spectrum has a larger tail as is expected.

In general, the FD is able to accept events with larger transverse momentum of the outgoing lepton. Fig.6.4 shows the MRE and FD distributions of p_T/p which is the fraction of the total outgoing lepton momentum that is in the transverse direction. There was an unexpected level of agreement between the MRE and the FD MC, which was chalked up to the fact that the MRE is made with the kinematics of a ν_μ CC event, but the reconstructed variables are evaluated using ν_e optimizations, the combination of which may have caused these distributions to look more alike than they would have otherwise. Regardless, this variable will still be used to create a kinematic phase-space boundary by excluding the higher p_T/p tail as this is a variable usually associated with acceptance differences between

[1]This isn't an oversight. If we used MRE events that had parents where the muon left the detector, we could not be sure that we had an accurate muon energy with which to simulate the electron. This could upset the balance of energy distribution in the event between the hadronic and EM components, potentially making the event no longer resemble a true CC event.

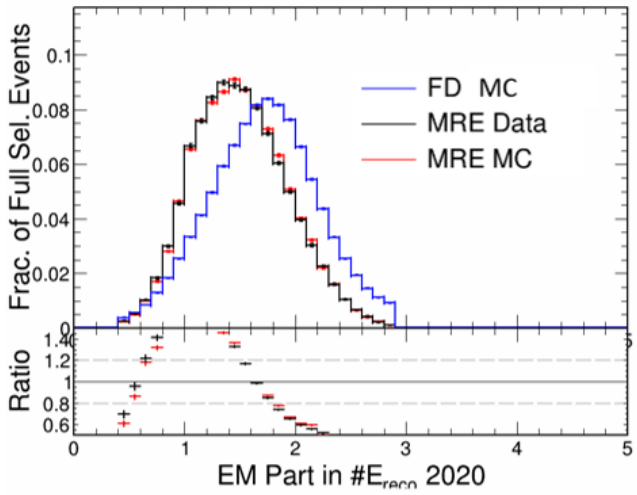


Figure 6.2: Total calorimetric energy of the EM showers of events in the MRE data sample (black), MRE MC sample (red), and the FD simulated ν_e signal sample (blue) that passed preselection. Distributions were area normalized to 1. Bottom plot is the ratios of the MRE distributions to the FD distribution.

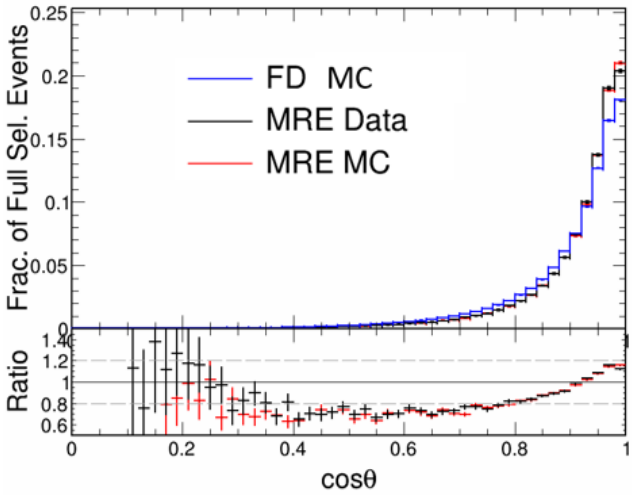


Figure 6.3: Distributions of $cos\theta$ where θ is the angle between the EM shower and the beam direction for selected ν_e events in the MRE data sample (black), MRE MC samples (red), and the FD simulated ν_e signal sample (blue). Distributions were area normalized to 1. Bottom plot is the ratios of the MRE distribution to the FD distribution.

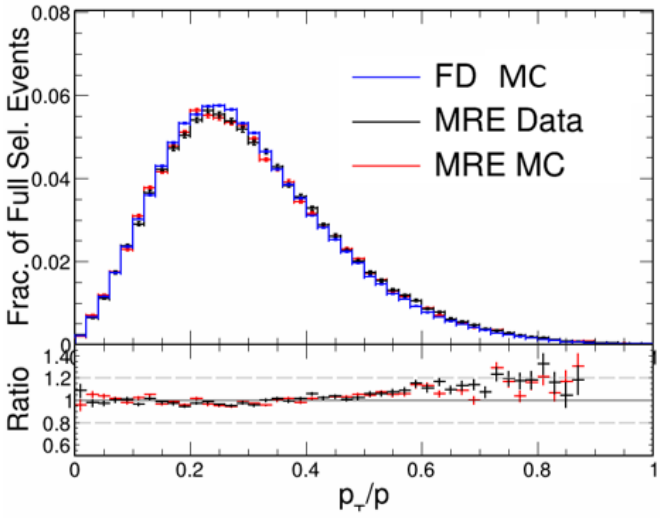


Figure 6.4: Distributions of p_T/p, the transverse momentum of the EM shower divided by the total momentum, for selected ν_e events in the MRE data sample (black), MRE MC samples (red), and the FD simulated ν_e signal sample (blue). Distributions were area normalized to 1. Bottom plot is the ratios of the MRE distribution to the FD distribution.

the detectors[2].

Fig.6.5 shows a few select other variables. The top row show variables that also showed divergences between the MRE samples and the FD MC due to either acceptance differences directly, or indirectly via the difference in EM shower energy. The bottom row shows some variables that do not appear to differ between the MRE and FD samples. While experimenting with which variables provided the most effective kinematic phase-space boundaries, the discrepancies like those in the top row, were all found to be largely covered by a combination of boundaries set by the EM shower angle to the beam, p_T/p, and the EM shower energy. Because of this, I chose to set the phase-space boundaries based only on those three variables.

[2]In hindsight, I should have looked at and used the same $|\vec{p_T}|$ variable used to make the transverse momentum bins in the extrapolation procedure instead of this fractional transverse momentum. However, while this would have reduced the impact scope of the corrections, as the resultant boundary likely would have cut more events from consideration than I otherwise did, the final conclusion of the effects of these corrections would remain the same. Furthermore, there is some amount of overlap between the events that would have been removed from consideration with a stricter $|\vec{p_T}|$ boundary and those that were actually removed via boundaries on the fractional transverse momentum, shower angle and shower energy.

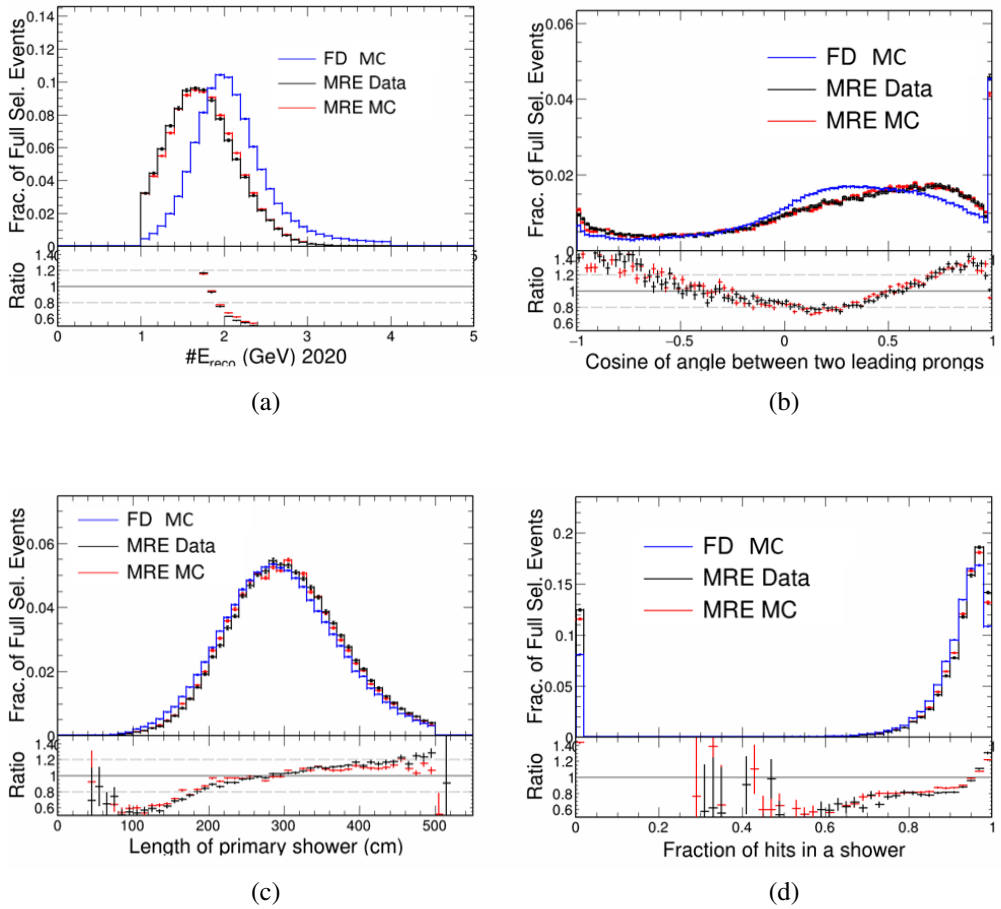


Figure 6.5: Distributions of (a) the reconstruction ν_e energy, (b) the cosine of the angle between the firtwo prongs in the event, (c) the EM shower length, and (d) the fraction of total slice hits that are in the EM shower for selected ν_e events in the MRE data sample (black), MRE MC samples (red), and the FD simulated ν_e signal sample (blue). Distributions were area normalized to 1. Bottom plots are the ratios of the MRE distribution to the FD distribution.

6.3 Correction Design

Now that the MRE vs. FD ν_e sample kinematic differences have been characterized, there are two ways we could go about actually calculating and applying the correction. First we could modify and reweight the MRE sample so that it more closely resembles the FD sample. Then, the ratio of the selection efficiencies between the data and simulation pieces of this modified sample would be used to create weights as a function of reconstructed neutrino energy, excluding bins with low statistics. These weights could then be applied directly to the whole predicted spectra at any point during or after the extrapolation process. This method would allow the corrections to be applied to most of the predicted signal, potentially only excluding the energy bins where the prediction would have low statistics anyway. However, the process of reweighting the MRE sample to make it resemble the FD would not properly preserve the differences in selection efficiency between the data and simulation without adding some amount of uncertainty. It wouldn't matter if we could apply the corrections to all the FD events, if they exaggerate or understate the true selection differences.

The other method doesn't rely on modifying the MRE sample to fit the FD, but rather just limits the scope of the corrections to regions of the kinematic phase space where the MRE sample has sufficient presence. This correction would be structured similarly to the first method in that it would still be a ratio weight, but it would also include "cuts" that would set the weight to 1 for events that fall outside the defined region of kinematic phase space. This method has the advantage of knowing that the true selection efficiency differences will remain intact as well as knowing that the corrections are only being applied to FD events that they actually correspond to. The downside is that the corrections would be applied to a smaller portion of the FD signal events than with the first method. Mechanically, this method also makes it harder to implement the correction as it would need to be evaluated and applied on an event by event basis during the prediction making process, instead of to the final prediction spectra as a whole after the fact. This last point is not that much of a concern, though, as the correction could be structured to easily be inserted into the prediction making code, even if the code for the correction itself is a bit more complex.

Despite the scope disadvantages, I have chosen to go this second route for this thesis. The transparency in what the corrections actually are and how they correspond to the events they are being applied to is worth more - at this stage anyways - than the larger scope that the first method would provide. The following sections will step through my method

of first defining the applicable region of phase space, calculating the correction, and then incorporating them into the existing prediction making process. Finally, I will show the effects of these corrections on the final predicted signal spectra and discuss the implications of those effects.

6.3.1 Defining the Applicable Kinematic Phase-Space

As mentioned in the previous section, I based the boundaries on the applicable kinematic phase-space on the the EM shower angle to beam (shwAngle), EM shower energy (shwE), and the fractional transverse momentum (p_T/p). These boundaries just need to placed to remove regions where there is not sufficient MRE representation but there is significant FD MC representation. This does not necessarily mean that I need to cut any region where there is more FD MC events compared to the MRE, but really just to exclude where the tails of the MRE sample lie and beyond.

Initially, I was going to set these boundaries by looking outwards toward the tail regions of the key distributions, and setting the boundary at the first bin that had less than 100 events in it. Using Poissonian histogram statistics, a bin with N events would have a statistical uncertainty of $\sqrt{N}$, so a bin with 100 events would have a 10% uncertainty attached to it. This should provide a good limit to determine boundary placement, and in fact, we have ND cross section studies that use this as their limit for finding where a MRE based ν_e selection efficiency correction can be applied, as did the early attempts of a MRE correction for NOvA's second 3-flavor analysis back in 2017. However, in the cross-section measurement case, the binning of the variable of interest is strictly set by the design of their analysis. That is not the case here, however, as the binning of my variables of interest is largely arbitrary. If I placed the boundary based only on making sure there was at least 100 events in every bin, the boundary would be in entirely different places with any change in binning.

Instead, I placed the boundaries by integrating over bins in the histogram, starting at either the first or last bin of the histogram and moving inwards towards the distribution peak, until the count reached 1% of the total number of events. While this 1% is still an arbitrary value itself, this method allows the placement to be made almost entirely independent of the binning chosen for each variable. Table 6.1 shows where these boundaries will be placed for the Prod5.1 version of the corrections. It also shows the "cut flow" of these boundaries for the Prod5.1 MRE sample and the FD MC, where each row adds the application of that boundary to the total event count. Overall, the FHC(RHC) MRE samples saw a total

	Boundary	Prod5.1 MRE		FD MC	
		Events	% of Initial	Events	% of Initial
FHC	No boundary	242650	100.0%	4.50×10^6	100.0%
	EM shower E < 2.6 GeV	240616	99.2%	4.21×10^6	93.7%
	p_T/p < 0.76	240616	99.2%	4.21×10^6	93.7%
	$\cos\theta_{beam}$ > 0.5	237857	98.0%	4.14×10^6	92.1%
RHC	No boundary	84250	100.0%	4.79×10^6	100.0%
	EM shower E < 2.6 GeV	83504	99.1%	4.32×10^6	90.1%
	p_T/p < 0.66	82639	98.1%	4.28×10^6	89.3%
	$\cos\theta_{beam}$ > 0.68	82102	97.5%	4.23×10^6	88.2%

Table 6.1: Definitions of the Prod5.1 version of the kinematic phase-space boundaries and their effects. First main column gives the boundary placements. Rest show the effect on the total number of PID selected events as each boundary is iteratively applied.

exclusion of 2%(2.5%) of events, and the FD MC saw a total exclusion of 7.9%(11.8%), meaning that at this stage, the corrections will still be applicable to most of the FD ν_e prediction. The boundaries found for the Prod5 MRE samples only differed in that the RHC boundary for p_T/p was set at 0.68 for Prod5 instead of 0.66 for Prod5.1 (the difference of a single bin), and therefore had a very similar rate of excluded events as the Prod5.1 version.

6.3.2 Calculating the Corrections

With the acceptance differences accounted for, the next step was to decide which variable basis should be used to calculate and then apply the corrections. Although it would seem to make the most sense to make the corrections with the same binning as the predictions they would be applied to (ie. as a function of E_ν split into the Low PID and High PID bins), this was decided against for a couple of reasons. First, we already saw that the EM shower energy for the MRE sample is shaped quite differently than the FD MC, leading to the total reconstructed E_ν to also have a differently shaped distribution between the MRE and the FD MC, as seen in Fig.6.5. The shape of the selection efficiencies also greatly differ, meaning that a correction based on the MRE as a function E_ν could be skewed when applied to the FD events.

The second cause to look elsewhere was seen when I looked to see what those corrections would be if we did use this binning. When looking at the smaller Prod5.1 sample, there were low enough statistics in the outer bins (I'm using the 100 events per bin metric here, as the binning is set) that the applicable range reduced down to [1, 3.5) GeV. Furthermore,

when looking at the uncertainty on the corrections, this range reduced down to [1, 2.5) GeV for the Low PID and removed the [2, 2.5) bin from consideration for the High PID, as these bins had correction uncertainty larger than the correction itself. This comes out to about a third of the events no longer be included in the corrections, in either calculation or application of the correction. Overall, it was determined that the analysis binning was not suitable to be the base for the corrections, and I needed to look elsewhere.

Learning from the analysis binning, there were three things I looked for in the new basis. The new basis would need to have general agreement, at least in overall shape, between the MRE samples and the FD MC, both in the distribution of selected ν_e events and the selection efficiency. Ideally, there would also be some philosophical motivation for the choice of basis, in a similar way that the analysis binning was initially chosen because it matched how the predictions were made. Lastly, the basis would also need to be able to be binned in such a way that the applicable range isn't decimated when removing bins with too low of statistics or too large an error on the correction.

Before continuing, I must take a quick diversion to talk about uncertainty calculations for the correction. The corrections will be applied to the FD prediction as weights made from the ratio of the data MRE selection efficiency ε_{data} over the MC MRE selection efficiency ε_{MC}. However, the true correction would not be these weights, but rather how far away from 1 that each of these weights are. Explicitly, this would be

$$\text{correction} = 1 - \left(\frac{\varepsilon_{data}}{\varepsilon_{MC}}\right) = 1 - \left(\frac{k_{data}}{n_{data}}\frac{n_{MC}}{k_{MC}}\right) \tag{6.1}$$

where n refers to the number of preselected events and k refers to the number of fully selected events for each sample. To find the uncertainty on the correction, we need to correctly find the uncertainty on the efficiencies first, and then propagate those through Eq.6.1. As discussed in [47, 48], this efficiency uncertainty must be found with care to account for the correlation between k and n, as well as ensure good behavior in the limiting cases of $k \to 0$ and $k \to n$. We will calculate the uncertainty σ_ε on the efficiencies as

$$\sigma_\varepsilon = \sqrt{\frac{(k+1)(k+2)}{(n+2)(n+3)} - \frac{(k+1)^2}{(n+2)^2}} \tag{6.2}$$

This σ_ε can then be propagated through Eq.6.1 to find the uncertainty on the correction

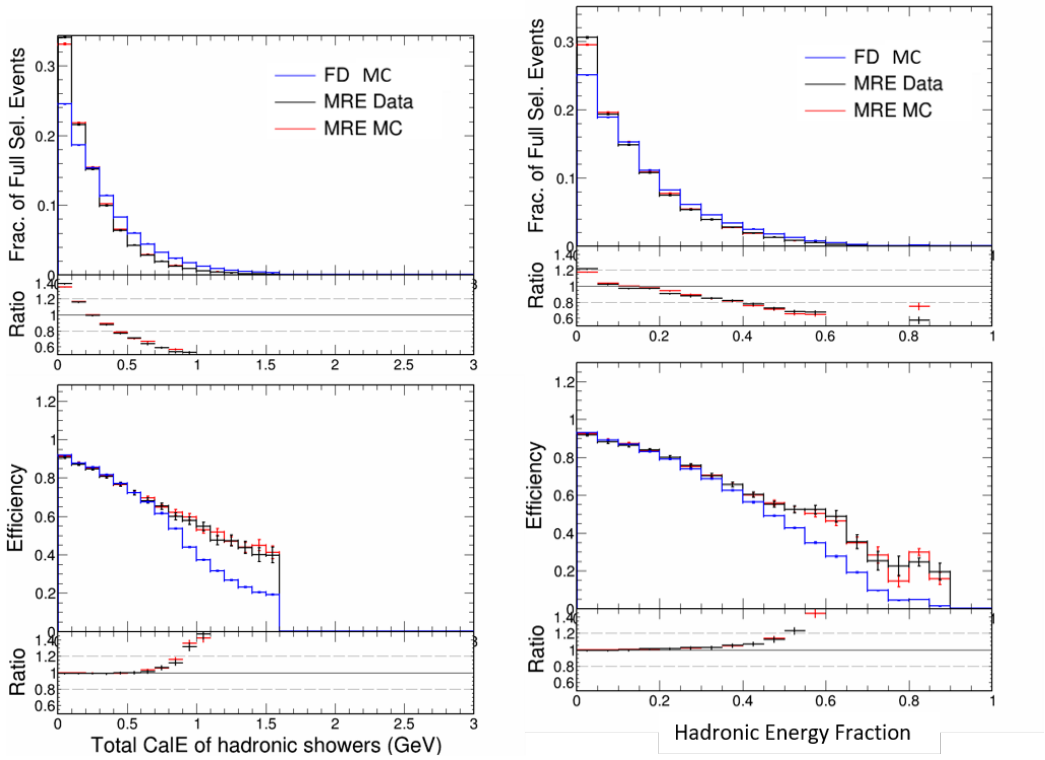
σ_{corr} to be

$$\sigma_{corr} = \frac{\varepsilon_{data}}{\varepsilon_{MC}} \sqrt{\left(\frac{\sigma_{\varepsilon}^{data}}{\varepsilon_{data}}\right)^2 + \left(\frac{\sigma_{\varepsilon}^{MC}}{\varepsilon_{MC}}\right)^2} \tag{6.3}$$

A full derivation of Eq.6.2 based on [47, 48] can be found in App.C. This estimation of the uncertainty holds well for this situation as long as n is sufficiently large, so we will bring back the requirement that any bin that a correction weight is made for must have at least 100 fully selected events.

With that aside finished, we can return to our search for a variable with which we can create our MRE corrections. Of the variables relevant to the ν_e selection process, the variables that showed the most agreement in both shape and selection efficiency between the MRE and the FD MC were almost all related to the size and shape of the EM shower such as shower length, width, and number of shower hits. This isn't terribly helpful, though, as the EM showers of the MRE are all simulated, regardless of whether we're looking at the data or MC sample. However, the distributions that dealt with the amount of hadronic activity present in each event also showed general agreement between the MRE and FD MC, meaning that they would meet the first condition for a new basis. Remember that the MRE was designed to probe for differences in the PID performance on data and MC coming from the hadronic sector, so we would expect the corrections to scale directly with the amount of hadronic energy in an event, making this class of variables philosophically motivated. Fig.6.6 shows the distributions for the total hadronic energy as well as the distributions for hadronic energy as a fraction of the total energy in the slice, both with the kinematic phase-space boundaries applied as cuts. Ultimately, I chose to use the hadronic energy fraction as the agreement between the MRE and the FD MC carried through more of the distribution and it also indirectly accounts for the differences in EM shower energy between the MRE and FD MC.

The binning was chosen by starting out with a very fine binning and then increasing the bin widths iterativly, checking every iteration the fraction of bins that I could get with uncertainty smaller than the corrections. I stopped re-binning once I had bin widths of 0.05 as going coarser than that wasn't improving results. In an attempt to include as many events as possible, I allowed for the last bin to be as wide as necessary to capture the end of the distribution tail where possible, while still requiring at least 100 events in that bin and that the uncertainty was smaller than the correction. Additionally, there were a handful of individual bins in the Prod5.1 version that had corrections so small that it would take an immense amount of data to be able to get the uncertainty smaller than correction, so those



(a) (b)

Figure 6.6: Top plots show distributions of (a) the total hadronic energy in the slice and (b) the fraction of total slice energy that is hadronic, for selected ν_e events in the MRE data sample (black), MRE MC samples (red), and the FD simulated ν_e signal sample (blue). Distributions were area normalized to 1. Bottom plots are the associated selection effiencies.

Hadronic Energy Fraction Lower Bin Edge	Prod5 Weights		Prod5.1 Weights	
	FHC	RHC	FHC	RHC
0.00	0.9929±0.0009	0.9935±0.0008	1	0.996±0.002
0.05	0.993±0.001	0.992±0.002	1.004±0.003	0.988±0.004
0.10	0.992±0.002	0.991±0.002	1.004±0.003	1
0.15	0.988±0.002	0.980±0.004	1.011±0.004	0.967±0.008
0.20	0.988±0.003	0.989±0.005	1.007±0.005	0.95±0.01
0.25	0.987±0.004	0.956±0.007	1.008±0.07	0.08±0.02
0.30	0.989±0.005	0.966±0.010	1	0.94±0.02
0.35	0.988±0.006	0.94±0.01	0.99±0.01	0.95±0.03
0.40	0.984±0.008	0.91±0.02	1	0.93±0.04
0.45	0.98±0.01	0.90±0.03	0.97±0.02	0.92±0.06
0.50	↑	1	0.96±0.03	↑
0.55	0.989±0.010	1	1	↑
0.60	↓	1	1.09±0.05	0.83±0.08
0.65	1	1	1.12±0.09	↓
0.70	1	1	0.93±0.05	↓
...	1	1	↓	↓
0.85+	1	1	1	1
Total	0.9970±0.0007	0.9987±0.0009	1.009±0.001	0.997±0.003

Table 6.2: Final weights for the Prod5 and Prod5.1 MRE corrections along with their uncertainties.

bins were manually set to have a weight of 1 (ie. no correction). Finally, every bin after a certain point was also manually set to have a weight of 1. This point would be where there were either not enough MRE events to guarantee a good correction or where there were no events at all.

Tab.6.2 shows the final weights and their uncertainties for both the Prod5 and Prod 5.1 samples in both beam modes. Since the PID was trained on MC, we expect that it would be marginally better at selecting events in the MC over the data, resulting in weights less than 1. This was indeed the case for both sets of RHC weights and the Prod5 FHC weights. The Prod5.1 FHC weights, however, have quite a few bins with a weight greater than 1, although the corrections themselves are mostly quite small. Since the Prod5 FHC samples does not see this behavior, and because the Prod5.1 sample has about 80% less statistics than the Prod5 sample, it is likely that these weights greater than 1 are being driven by upward statistical fluctuations. In the future, though, if a larger production of the Prod5.1 sample is made, this should be rechecked to ensure that it is not an indication of a larger

problem with the PID.

To actually apply these weights, a set of Weight objects in the NOvA software framework was made so that the they could be included in the spectra definitions being called by the prediction making macro[3]. Each Weight was made such that any event passed to it would first be checked to see if it was a FD event and a true simulated ν_e or $\bar{\nu}_e$. Then it checks if the event falls within the kinematic phase-space boundaries. If the event passes these checks, its hadronic energy fraction and beam mode is used to assign it a weight from the Tab.6.2. If it doesn't pass any of those checks, it just gets a weight of 1. This way, the Weight can be applied as a blanket weight to the whole prediction, but only affect the FD ν_e events to avoid double correcting the beam ν_es , while also applying to the other ν_e -like background events in addition to the ν_e ($\bar{\nu}_e$) signal.

6.4 Effects of the Correction

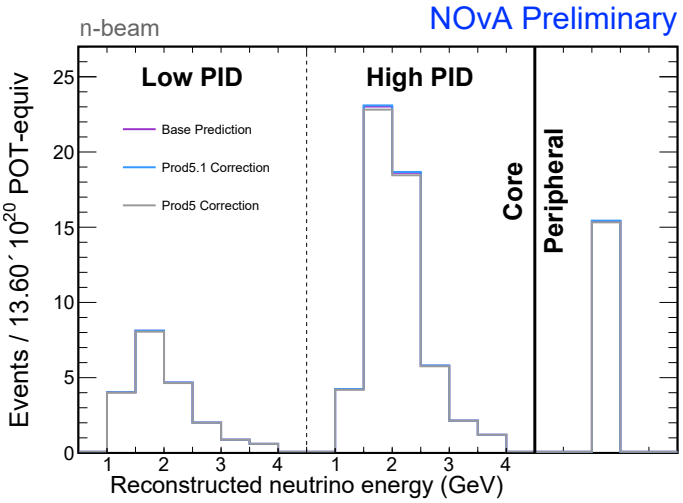
Three sets of predictions were made at the 2020 best-fit point to see the effects of the corrections. These predictions were all made with the same ND simulation and fake ND data as the base for the extrapolation and only differed in the MRE correction weights used: no weight, Prod5 weights, and Prod5.1 weights. The systematic uncertainty pieces of the predictions were also subject to the correction weights. Fig.6.7 shows the three resultant ν_e predictions plotted together for both beam modes. Tab.6.3 gives the corresponding event counts and how the corrected counts compare to the base prediction. Overall, the Prod5 total prediction only differed from the base by only -0.74%(-1.0%) for the FHC(RHC) while the Prod5.1 total prediction differed by 0.23%(-1.3%) for the FHC(RHC), all well within the uncertainties on the base prediction. These differences came mostly from the signal ν_es and $\bar{\nu}_e$s , although there was also contribution from the beam ν_es and NC current events.

Using fake FD data, a fit was performed for all three sets of predictions. Although there were very slight differences in the best-fit points of these three predictions to the fake FD data, they were all well within the uncertainties quoted for the 2020 analysis. Additionally, confidence contours were made as seen in Fig.6.8, where the "just stats" contours did not include any systematic uncertainties in the contour building, whereas the " with systs" versions did. Note that these contours were not made with the full Feldman-Cousins method, but rather made with the typical Neyman construction. The full Feldman-Cousins method is too computationally expensive to run for this study, and the Neyman construction

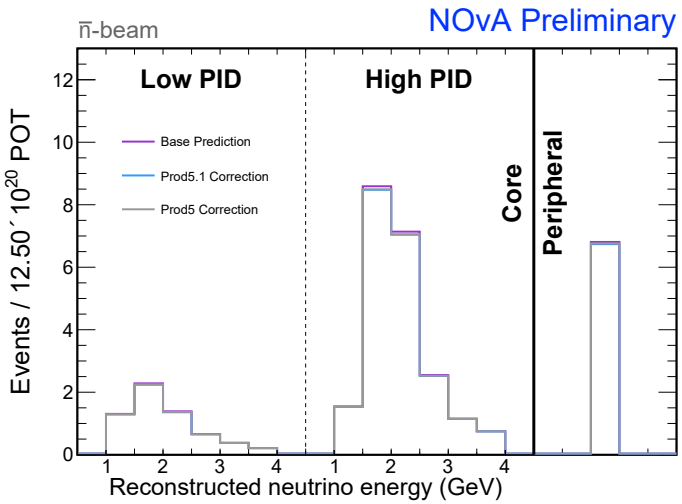
[3]In novasoft, the weight definitions are stored at CAFAna/Weights/MRECorrection.h

	Base Prediction Event Count	Prod5 Corrected Event Count	Diff.	Prod5.1 Corrected Event Count	Diff.
		FHC			
$\nu_\mu \rightarrow \nu_e$	64.4	63.9	-0.88%	64.6	0.28%
$\bar{\nu}_\mu \rightarrow \bar{\nu}_e$	1.07	1.06	-0.64%	1.07	0.20%
$\nu_\mu(\bar{\nu}_\mu) \rightarrow \nu_\mu(\bar{\nu}_\mu)$	1.5	1.5	-	1.5	-
Beam ν_e ($\bar{\nu}_e$)	14.85	14.76	-0.61%	14.87	0.17%
NC	5.28	5.27	-0.17%	5.28	0.03%
Cosmic	3.13	3.13	-	3.13	-
Other	0.43	0.43	-	0.43	-
Signal	$64.4^{+5.3}_{-5.8}$	63.9	-0.88%	64.6	0.28%
Background	$26.3^{+0.7}_{-0.9}$	26.2	-0.40%	26.3	0.11%
Total Prediction	$90.7^{+5.5}_{-5.9}$	90.1	-0.74%	90.9	0.23%
		RHC			
$\nu_\mu \rightarrow \nu_e$	2.47	2.43	-1.4%	2.42	-1.9%
$\bar{\nu}_\mu \rightarrow \bar{\nu}_e$	21.2	20.9	-1.2%	20.8	-1.4%
$\nu_\mu(\bar{\nu}_\mu) \rightarrow \nu_\mu(\bar{\nu}_\mu)$	0.35	0.35	-	0.35	-
Beam ν_e ($\bar{\nu}_e$)	7.0	6.95	-0.9%	6.92	-1.3%
NC	1.98	1.975	-0.3%	1.971	-0.6%
Cosmic	1.55	1.55	-	1.55	-
Other	0.26	0.26	-	0.26	-
Signal	$21.2^{+2.0}_{-2.1}$	20.9	-1.2%	20.8	-1.4%
Background	$13.6^{+0.5}_{-0.5}$	13.52	-0.8%	13.47	-1.1%
Total Prediction	$34.8^{+2.0}_{-2.3}$	34.4	-1.0%	34.3	-1.3%

Table 6.3: Event counts for MRE Corrected Predictions compared to the Base Prediction. The Diff. column refers to the percent difference between the base predictions and the corrected predictions.



(a)



(b)

Figure 6.7: The ν_e predictions made with fake ND data and at the 2020 best-fit point for (a) FHC beam mode and (b) RHC beam mode. Both the base prediction and the corrected predictions are plotted, but they are similar enough that they are often on top of each other.

provides a decent enough alternative[4]. Circled in grey are the regions where there are visible differences in the contours between the three predictions.

The biggest differences between the three sets of contours can be seen in the "stats-only" $\sin^2\theta_{23}$ vs. δ_{CP} contours, but they mostly disappear in the contours that included systematic uncertainties. While this is partially just indicative of how the impact of these corrections is overwhelmingly covered by the systematic uncertainties, it is also partially coming from how the MRE correction was actually applied. Remember that I applied the same corrections to each of the systematically shifted predictions as I did to the nominal prediction. This means that the effect of the corrections will be largely washed out when the log likelihood ratios are made at each grid point during the contour construction. It is expected that if the MRE corrections for the systematic uncertainties were unique to each corrections, we would see more of an effect in the contours, but still nothing drastically different.

Overall, while these MRE corrections do have a measurable impact on the amount of ν_e ($\bar{\nu}_e$) events present in a prediction, the difference is not enough to affect the results of our 3-flavor analysis and fall well within systematic uncertainties. The addition of these MRE corrections to the analysis, however, does increase the robustness of the analysis by providing a data-driven correction to our otherwise uncorrected ν_e ($\bar{\nu}_e$) predictions.

[4]In fact, these Neyman constructed versions are the inputs to the full FC treatment, informing both the sampling points of interest and the number of necessary pseudoexperiments that need thrown at each point.

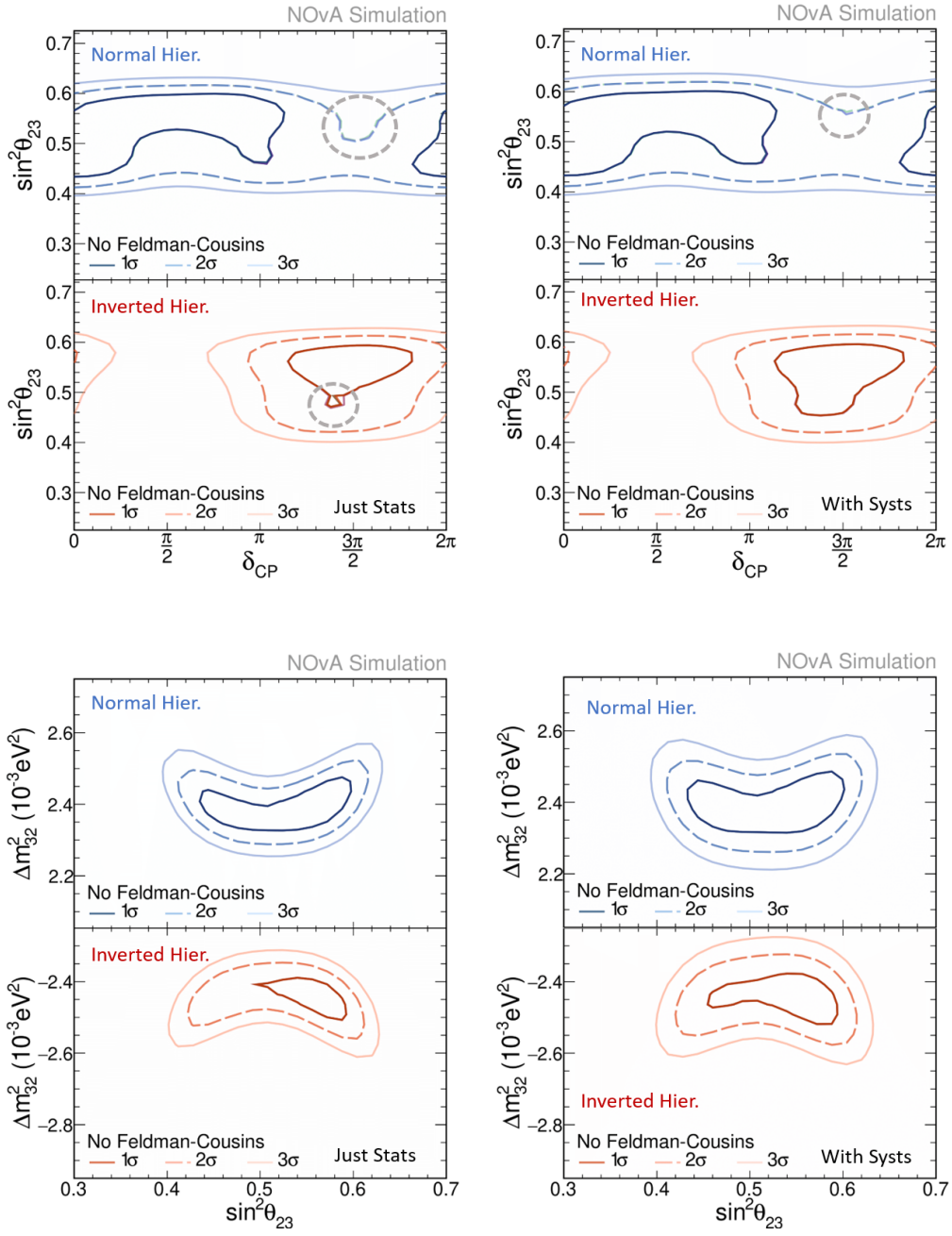


Figure 6.8: The confidence contours for the base prediction, Prod5 MRE corrected predictions, and the Prod5.1 MRE corrected prediction all plotted on top of each other. Top row is the contours for $\sin^2\theta_{23}$ vs. δ_{CP}. Bottom row is the contours for Δm^2_{32} vs. $\sin^2\theta_{23}$. Left column was made with just statistical uncertainties. Right column included the full set of systematic uncertainties. Although all three sets of contours are plotted, they are so similar that in most places, only the topmost contours (those for the base prediction) can be seen. The grey circles highlight where differences between the three predictions can be see.

6.5 Future Developments

This exercise in making and applying MRE based corrections for NOvA's appeared ν_e prediction has proven that these long awaited corrections are both possible and have the expected effect, but we don't have to stop there. As mentioned, the current treatment of the systematic uncertainties is likely not quite capturing the full power of these corrections. In this section, I will outline some ways that we could address this going forward along with the difficulties involved in each. In order for this section to be useful for a future NOvA researcher to use as a starting reference, I cannot avoid going into somewhat technical detail. Other readers, beware.

Ideally, since there is no reason to believe that the CNN selection efficiency would remain unchanged for a systematically shifted sample, we would create a MRE MC sample for each systematic uncertainty, where the sample was shifted by that systematic uncertainty. For example, when generating the initial ν_μ MC sample and the simulated electrons, we would shift all the calibration constants up or down by 5% to represent a 1-σ shift in that uncertainty. This is already done for our calibration and detector response systematic uncertainties and can be done for the rest of the 60 uncertainties on NOvA via reweights to the nominal MRE sample. These samples can then be used to create unique MRE corrections to correspond with each uncertainty. During the prediction making process, predictions are made that are also systematically shifted, and we could apply these unique MRE corrections to the corresponding prediction piece. These corrections would likely not wash out in the same way that we saw in this thesis when the systematic pulls are included in the contour building procedure. However, doing this is easier said than done.

Making the systematically shifted MRE samples is not really the issue, since we already have systematically shifted MRE samples for the major calibration and detector response uncertainties, and most of the other uncertainties can be created by applying weights to the already made samples. The issue is in how to incorporate the corrections. The most straight forward method would be to just define corrections for each of the uncertainties in the same way we find the nominal corrections, and then apply them individually when making the predictions. This would have the desired effect but would be incredibly tedious to do as we have around 60 systematic uncertainties for each analysis. This method would interrupt the overall workflow of our established analyses, making it not ideal by any means.

The other method considered would be to set up a system, similar to the decompositions, that would allow for the corrections from both the nominal and all the systematically shifted

MRE samples to be calculated on the fly before being applied. However, we cannot just replicate the decomposition system and swap things out for the MRE. First off, the decompositions are designed to be able to make their corrections by making a histogram where the bin counts are the corrections weight, and then multiply that histogram to the predicted sample. This works because the decomposition is done in the same binning as the predictions (as a function of E_ν), but the MRE corrections are not. Instead, a method of transformation between the hadronic energy fraction and E_ν would need to be found. This is possible, but introduce a new source of uncertainty on the corrections themselves. Additionally, the decomposition method as it currently stands, would not allow for the enforcement of the kinematic phase space boundaries on an event-by-event basis that the current version of the MRE corrections calls for. It is unclear if this issue can be solved within the context of the current prediction making code without going down the path of reweighting the MRE samples to account for the acceptance differences instead of the current method of excluding events. As already discussed earlier in this chapter, reweighting the MRE samples to kinematically match the FD MC would introduce new uncertainties on the corrections, as well as potentially fail to uphold the true differences in selection efficiency that this correction is supposed to account for.

Given the small size of the corrections and their limited impact on our analysis results, it very well may be that the addition of a full set of unique systematic uncertainty corrections would not actually be worth the amount of work, increased code complexity, and additional uncertainties that they would entail. Instead, I would recommend identifying the top five or so uncertainties whose unique corrections differ the most from the nominal corrections, or would have the largest impact elsewhere, and only incorporate those different corrections into the overall set of corrections.

7. Conclusions

NOvA's 3-flavor neutrino oscillation analysis has been able to provide constraints on the oscillation parameters θ_{23} and Δm^2_{32} that are competitive with other oscillation analyses around the world, as well as provide one of the first probes of δ_{CP} and the neutrino mass hierarchy. NOvA uses a convolutional neural network for identifying and categorizing neutrino events in our data, and this network is validated with a suite of data-driven studies. These Muon Removed studies confirm that the neural network performs similarly on our data as it does on the simulated samples it was trained on which is essential to our ability to appropriately use the network. Furthermore, I have shown that the extension of one of these Muon Removed studies into a data-driven correction to our predicted ν_e ($\bar{\nu}_e$) samples used in NOvA's 3-flavor analysis increases the robustness of the analysis without changing the results and conclusions drawn. This correction is expected to be used in future NOvA 3-flavor oscillation analyses, with the next analysis planned for when we have doubled our accrued FHC beam exposure later this year.

As NOvA enters into the last few year's of its run-time, decreases in our systematic uncertainties and increases in our ability to identify neutrino events will be the biggest sources of increased sensitivity in our analyses. The MR* studies and corrections will continue to play a large role in these efforts, with the potential impact of the corrections growing with every improvement of the overall analysis. Further exploration of the potential of the MRE correction should start with applying the corrections to the systematic uncertainties to either increase the applicability of the corrections or potentially constrain some of these systematic uncertainties. This is easier said than done, however, and a few potential options were outlined in the previous chapter.

The MRBrem and MRDiF studies also have potential to contribute to a joint MR* correction in conjunction with the MRE. These could provide an additional correction based on CNN performance deriving from the EM shower component of a ν_e event. Additionally, it should be possible for these two studies to be successfully run on ND data - albeit with some sizable adjustments to their event selection algorithms - as was explored already with the MRDiF algorithm. They could then provide checks not only on the CNN performance between data and MC, but also between the ND and the FD, as well as potentially help constrain the acceptance difference between the detectors. While these improvements may

not prove to have a sizable impact on the actual results of the 3-flavor analysis, they - like the MRE corrections - would help strengthen confidence in our event selection and analysis methodology.

Beyond the 3-flavor analysis, NOvA is also starting to include MRE corrections to our cross-section analyses using our ND data. Additionally, the concept of these MR* studies and corrections could be applied to other neutrino experiments like the upcoming DUNE experiment. This neutrino oscillation experiment will have a baseline of 1300 km through the Earth's surface as well as the world's most intense neutrino beam. DUNE's detectors will be extremely granular allowing for much more detailed looks at neutrino interactions. Because of this, the effects of any neutrino interaction mismodelling will be more pronounced in the simulation, making any CNN-like selection process even more prone to performance differences when run on the data and simulation. Studies and corrections similar to NOvA's MR* studies and MRE corrections will be essential to DUNE's confidence in their particle PID and by extension, their analyses.

A. NOvA's Test Beam Calibration

In addition to the two main detectors, NOvA has commissioned a third detector that is in the Fermilab Test Beam Facility. This smaller scale detector is functionally equivalent to the others and lies in a charged particle beam. NOvA hopes to use this Test Beam (TB) detector to decrease calibration uncertainties in the main detectors, create a library of tagged particles from data, and better understand and characterize detector response to charged particles.

A.1 Test Beam Detector

The TB detector is functionally equivalent to the main detectors, but on a much smaller scale. The TB detector is roughly 30 tons compared to the ND at 300 tons and consists of 63 planes. Each plane is comprised of 2 modules of 32 cells each, totalling 64 cells per plane and 4,032 cells total. The whole detector is considered a single diblock (DB1), but can be broken down into the front block (the first 32 planes) and the back block (the last 31 planes).

When it comes to the electronics on the detector, the TB detector differs slightly from the other detectors in that is uses a combination of front end boards (FEB). As noted in Sec.3.3.1, The ND and FD FEBs differ in their firmware with the ND firmware being optimised to better handle bursts of concentrated high activity due to its proximity to the beam source. TB uses a combination of these FEB versions. Most of the planes in the detector have the FD FEBs while 8 have the ND FEBs. These planes with the ND FEBs are situated in the middle planes of each block. The FD style FEBs in the front block connect to DCM 1, while those in the back block make up DCM 2. The 8 ND style FEBs make up DCM 3.

The liquid scintillator used to fill the detector is nominally the same as for the other detectors. However, the oil was sourced from three different places resulting in differing quality. The front block of the detector was filled first and our first data taking run only had this front block filled with oil and outfitted with electronics. This oil was from NDOS (the ND-like prototype detector) and had been stored at Fermilab in a tanker. After the first data taking run, the back block of the detector was filled. The first two-thirds was filled

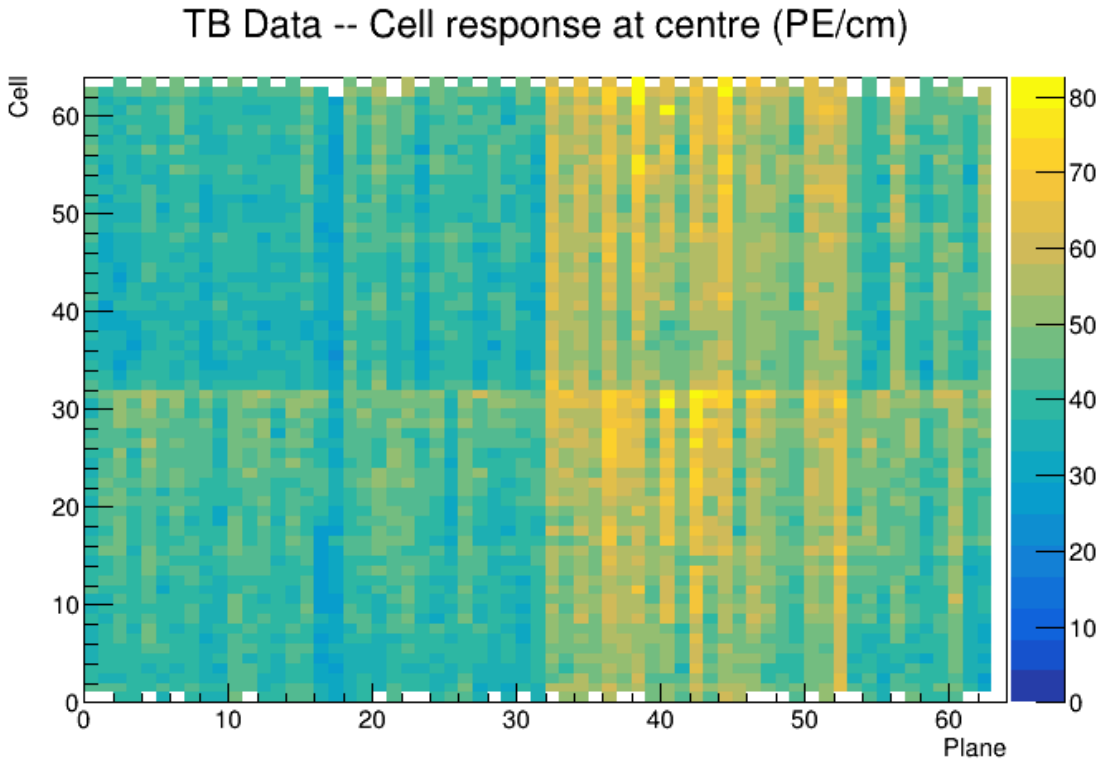
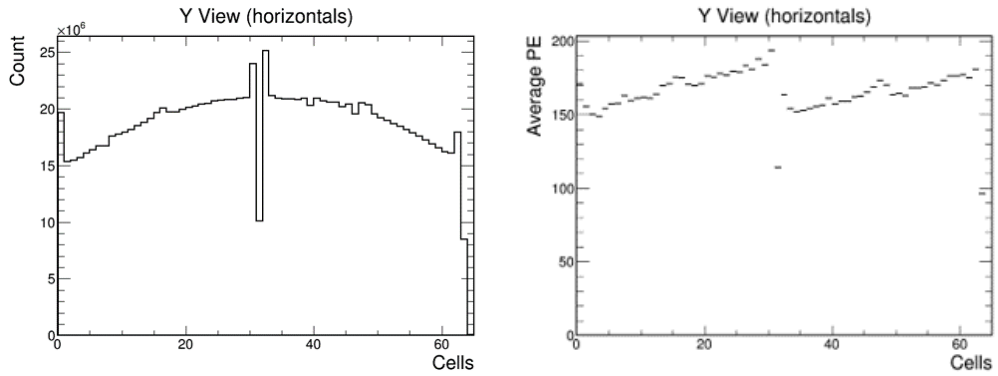


Figure A.1: Average cell response (PE/cm) at the center of each cell in the TB detector. Planes 0-31 have the NDOS oil, planes 32-52 have the Ash River oil, and planes 53-62 have the barrel oil.

with oil left over from filling the FD and that had been stored at Ash River. The final third of the back block of the detector was filled with oil barrels from one of the contributing universities of the collaboration. Testing showed that the Ash River oil was the best quality of the three oils in terms of light yield, with the oil barrels not much worse. However, due to the way that the NDOS oil was stored, there was some light leakage causing the oil to age. This resulted in the NDOS oil to be of noticeably lower light yield than the other oils. However, all the oil was determined to still be within the design specs, and relative calibration will account for the differences anyways. Fig.A.1 shows the visible effect of these different oil qualities in the average cell response (PE/cm) at the center of each cell.

During filling of the detectors, we came across another challenge. All three of NOvA's detectors were built so that they were slightly tilted so that the horizontal cells are slightly uplifted at the end from which they were filled. This way, during filling, air inside the cells could easily escape through the lifted filling end. However, for TB, the tilt was in the



(a) (b)

Figure A.2: Profile of the cells in the horizontal planes of the TB detector prior to correcting for the underfilled cells where (a) shows the number of tricell hits per cell and (b) shows the average PE per tricell hit. Note the dips in both at cells 31 and 63 where the air bubbles sit, and the spikes on either side of those underfilled cells.

wrong direction resulting in sizable air bubbles being trapped inside the topmost cells of each module. This corresponds to the topmost horizontal cells in the detector as well as the horizontal cells exactly half way up the detector. These underfilled cells have a much lower light yield than their neighbors as can be clearly seen in Fig.A.2 where a large drop in tricell hits can be seen for cell 31 where the air bubble sits. The cells on either side have a higher than expected number of tricell hits due to how the selection of hits is made.

The underfilled cells issue was mostly assuaged in April 2021 (middle of the third data taking run) by extending the length of those cells and putting in more oil, forcing the air bubble closer to the edge of the detector. However, data taken before this time will always have this artifact, so care has been taken to account for, or at minimum acknowledge, the presence of these cells in the calibration process and in the various TB analyses.

A.2 The Beam and Beamline Detectors

As previously mentioned, the TB detector lies in a charged particle beam rather than the NuMI beam that the main detectors lie in. This beam first starts out as a proton beam from the Main Injector (the same proton beam that also produces the NuMI beam) and is the primary beam. This protons are accelerated to 120 GeV and diverted to a copper target approximately 120m from the TB detector, creating a secondary beam. This secondary beam mostly comprises of protons and pions and the energy of these particles can be

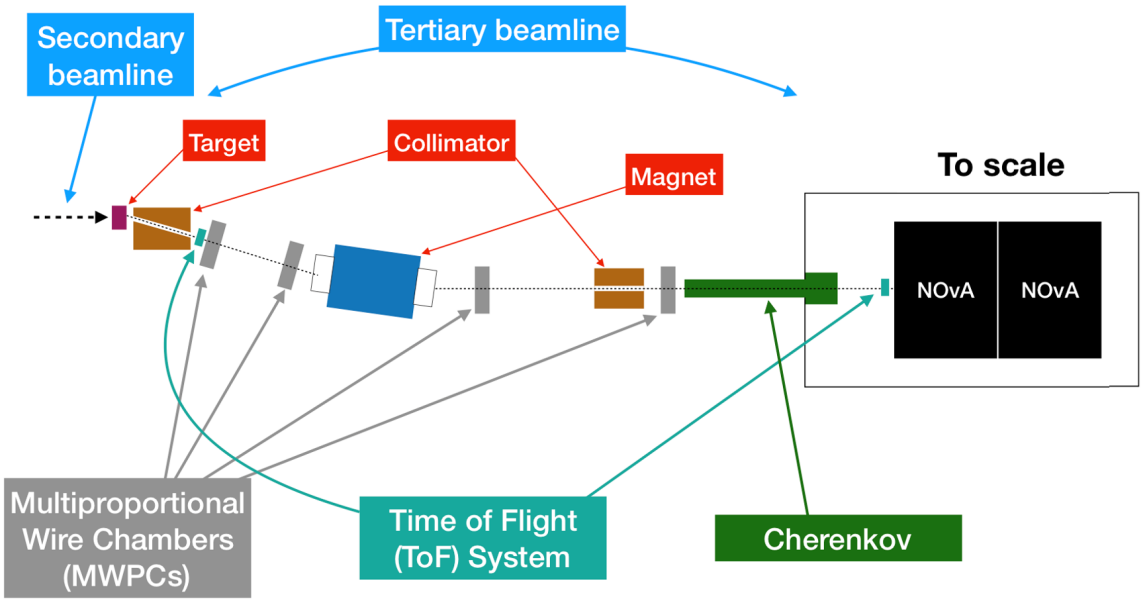


Figure A.3: Top view diagram of the TB beamline and its detectors.

tuned between 10-100 GeV. Finally, this secondary beam hits another copper target, this time located approximately 14m from the TB detector, creating the tertiary beam that our detector actually sees. The particles in this beam both have much less momentum (0.5 Gev/c to 2 GeV/c) than the secondary beam and the beam is much more diverse in its particle make up. The main particles seen from this beam are protons, pions, electrons, muons, and kaons, all with momentums comparable to the final state particles seen in the main detectors from neutrino interactions.

To actually separate the tertiary beam from the secondary beam and direct it towards the TB detector, an analyzer magnet is used with a collimator. This selects out a desired momentum from the tertiary particles and directs them towards the detector by correctly setting the strength of the magnetic field and width of the collimator opening.

To be able to reach the goals for TB laid out at the beginning of this chapter, the particles entering the detector need to be able to be tagged with their particle type and their momentum before entering the detector. This is done with a series of beamline detectors including four wire chambers, a cherenkov detector, and a time-of-flight (TOF) system.

The wire chambers are more specifically multi-wire proportional chambers. These are spaced out along the length of the beamline and allow for 3D tracking of the particles as they travel along the beamline to the detector.

The TOF system is comprised of three scintillating panels. One is upstream of the beam

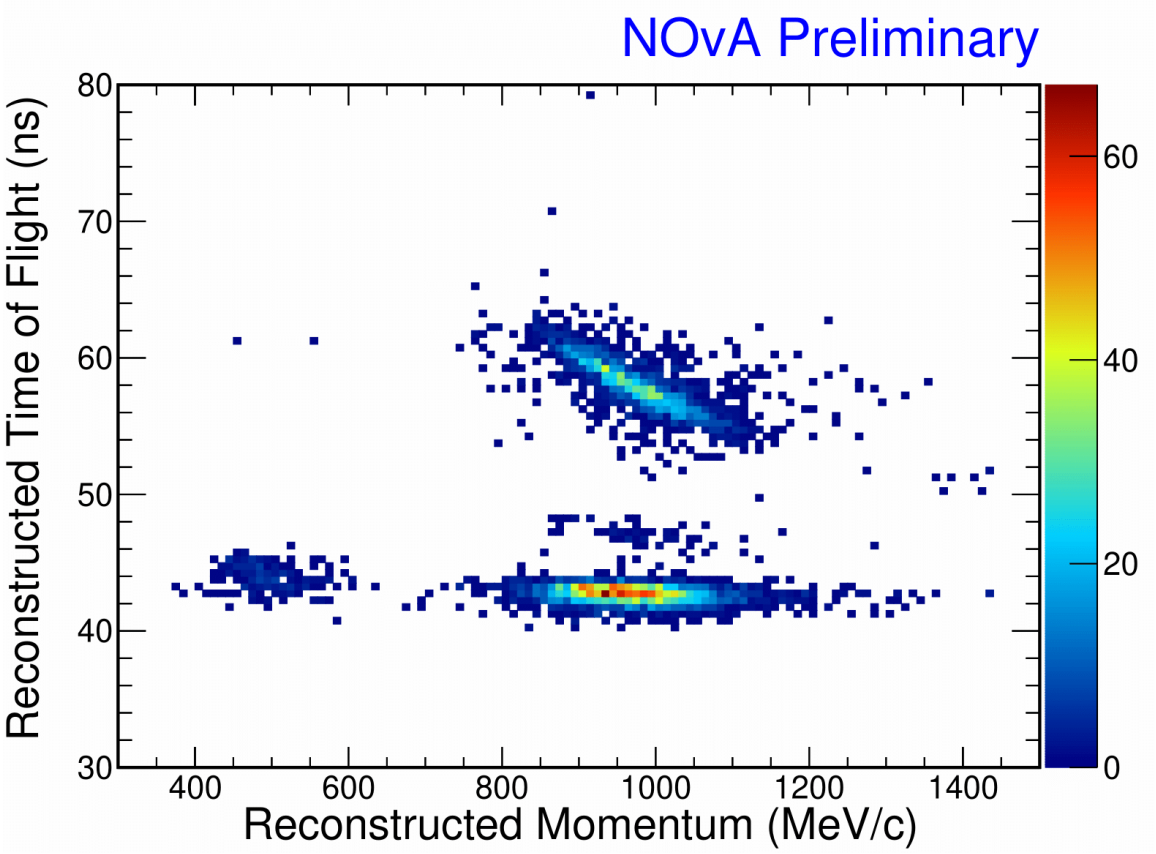


Figure A.4: Reconstructed Time of flight vs Momentum for Test Beam data. The Dipole Magnet current configurations here are 500 A and 1 kA, corresponding to particles with 500 MeV/c and 1000 MeV/c respectively.The plot shows the ability of the NOvA Test Beam beamline instrumentation to separate charged-particle types and measure their time of flight and momentum. The gradient colors display data accumulation in the parameter space.

while the other two are downstream. The TOF system allows for a particle's mass to be measured which helps in identifying what type of particle it is. Remember that the analyzer magnet makes it so that all the particles in the beamline at a given time are of roughly the same momentum, so any differences in velocity of the particles is caused by their differences in mass. The TOF system pairs up hits in the upstream panel to the downstream panels and simply calculates the difference in hit times. Combining that information with the baseline of the TOF system gives the particle's speed which can in turn be combined with the particle momentum to give the particle's mass.

Fig.A.4 shows reconstructed time of flight vs. momentum for a subset of the TB data

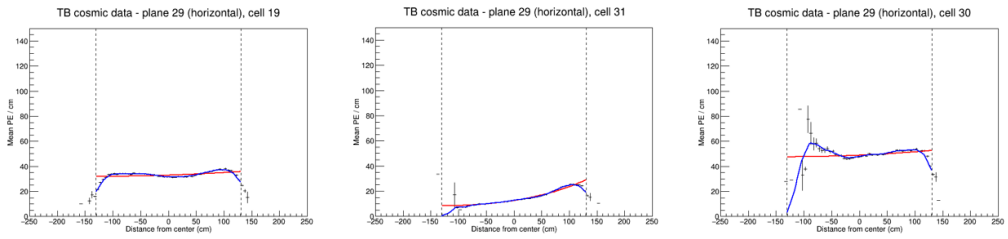
corresponding to particles with 500 MeV/c and 1000 MeV/c. The cluster of particles near the top of plot are primarily protons while the smaller cluster in the center is primarily kaons. The clusters on the bottom are a mix of electrons, pions, and muons. The electrons can be identified from these events via the Cherenkov detector at the end of the beamline. Currently, there is not a clear way to distinguish the pions from the muons, but we may be able to with future analyses.

These beamline detectors combined with the analyzer magnet mean that we can tag particles that travel through the beamline with their particle ID and momentum before they enter the detector. These tags can be compared to what is seen in the detector after standard reconstruction processes have been applied to the detector data. This is where the power of the TB program comes in.

A.3 TB Detector Calibration

One of the main goals of the TB project is to be able to better understand - and hopefully lower - NOvA's detector calibration uncertainty. This will be done by comparing what the beamline detectors tell us a particle's energy is to what our regular detector calibration process tells us what that particle's energy was.

In general, the TB calibration procedure followed that of the main detectors as described in Sec.3.4, with some changes made to address issues that are unique to TB.



(a) (b) (c)

Figure A.5: Selection of three horizontal cells from the same plane and module in the TB detector. The plots show the mean PE/cm for (a) a well behaved cell away from the underfilled cells, (b) an underfilled cell, and (c) a cell neighboring the underfilled cell. The red curves are the attenutation fit and the blue curves show the fit after the LOWESS fit corrections

For TB, we are generally able successfully run relative calibration on the whole detector with the exception of two regions. These are the top cells of each horizontal module, the

underfilled cells described previously, as well as the horizontal cells on either side of the underfilled cells. Fig.A.5 shows examples of the mean PE/cm of a regular cell away from the top of the module, an underfilled cell, and one of the neighbors of the underfilled cell. As the plots show, the underfilled cell not only has fewer hits and a lower PE rates as was already shown in Fig.A.2, but also has worse attenuation and threshold bias effects as compared to the regular cell. Unsurprisingly, the resulting attenuation and LOWESS fits for these underfilled cells did not have low enough χ^2 to be able to consider those cells as calibrated.

What was surprising was the behavior seen in the cells that neighbor the underfilled cells. These cells showed erratic spikes and dips especially in the far end of the cell, and large errors asociated with the points on the plot in these regions. This can be attributed to how tricell hits are selected, however. With the underfilled cells having lower hit counts, it follows that it would be harder for the neighbor cells to find true tricell hits since the tricell requirements include having hits in the track from both neighboring cells. This results in much less accurate pathlength calculations which in turn leads to the behavior we see in Fig.A.5, and these neighboring cells also being left uncalibrated. Calibration efforts for the data taken after the filling of the underfilled cells showed that while we still can't calibrate every one of these cells, a larger portion now behave more like the regular cell in the above plot.

A.3.1 Simulation for Calibration

Another aspect of the TB calibration process that had to be modified was the simulation used for the MC side of the calibration process. CRY is the main generator used to make the cosmic muon simulation for the main detectors, and was originally also used for the TB detector. However, CRY uses muon momentum and angle distributions that do not match what we expect for our detector locations, and it is highly inefficient process when applied to our detectors in that not every event that is generated ends up passing through our simulated detectors. This results in the final output files including large amounts of empty events. This isn't too much of an issue for the main detectors, especially the FD, since the sheer size of the detectors increase the percentage of the generated particles that touch the simulated detectors to a passable amount. However, due to the small size of the TB detector, it is a much less practical endeavor. Large amounts of computing resources were required to make the initial large batch CRY simulation for the TB, and the end result

still did not produce enough actual statistics to properly run the calibration.

Instead, an alternative simulation method was developed. To ensure that we are only generating muons that we know will enter the detector, we used the cosmic data to seed the direction and location of the muons. For every selected cosmic event in the data, the reconstructed muon kinematics can be pulled and used to create any amount of HEPEVT entries with statistical fluctuations applied to the kinematic values. A charge is also assigned with each muon having a flat 56% chance of being positive. Other models for assigning charge could be considered in the future, but this is currently good enough for calibration purposes. Finally, the momentum assigned to the HEPEVT entries can be chosen to be within a specific range depending on whether the sample we want to produce is full of through-going or stopping muons. These HEPEVT entries can then be read in and a text based event generator built into our software.

The results of this simulation method is a significantly fewer number of files needed to run calibration (ie. 500 compared to the over 95k needed with the CRY sample) where almost every event in each file can be used. These muons are also much more representative of our actual data as they were seeded from the data, which we hope will improve the calibration results. This muon generation method works well for TB and is being used for ongoing calibration efforts, but has not yet been adapted for use in the main detectors, although there is hope to sometime in the future.

B. Detector Aging

As we have entered into the second half of NOvA's proposed lifetime, we have started to see some signs of detector aging in our data. Understanding and quantifying this aging will help improve both our detector response simulation and shape certain systematic uncertainties included in analyses.

B.1 Motivation

The NOvA 2019 analysis was a top-up analysis where the only change since the last analysis in 2018 was the inclusion of FD data taken since the last analysis. The ND data and resulting predictions would not be changed. Therefore, we needed to check that the ND data taken since the 2018 analysis was consistent with previous periods indicating that the already made predictions would still be valid for the new analysis and increased FD exposure. Additionally, data from before and after the 2018 summer beam shutdown[1] were compared to ensure that there were no significant changes in our data due to changes in beam behavior seen by other experiments.

This study compared several key variables across three different datasets that were all taken during RHC beam running. The dataset used to represent the 2018 analysis were periods 4 and 6. Epoch 7d was used to represent the new data taken before the summer shutdown. Epoch 8b was used to represent the new data taken after the summer shutdown.

Figures B.1-B.2 show some of the key results of this study. Most notably, there was a slight 0.15% decrease in reconstructed neutrino energy between periods 4+6 and epoch 8b as well as a 0.5% decrease in the normalization of selected events between epochs 7d and 8b. Additionally, there was an observed 0.84% decrease in the number of hits seen per slice from periods 4+6 and epoch 8b. All three of these effects were seen both before and after the summer shutdown, suggesting that the effect has more to do with detector aging effects than changes in the beam. At the time of the study, the normalization systemic uncertainty was large enough to cover the change in normalization and the light level systematic uncertainty

[1]The NuMI beam stops running every summer for a few months for maintenance, repairs, and upgrade work. The summer time is chosen as the higher temperatures make it the most expensive time of the year to run the beam. NOvA continues to take data during this time to a lack of beam doesn't mean that cosmic activity ends, or other exotic events like supernova neutrinos that we have searches for can't happen.

was determined to cover the decrease in the number of hits per slice, and the analysis was conducted as planned. However, if these changes were indeed caused by the detector aging, as time goes on and the effects get worse, it's likely that these systematic uncertainties would no longer be enough, and we would need to start including detector aging into our simulation and/or create a new systematic uncertainty associated with the aging. The first step to either of those is to better characterise the aging and potentially identify which component(s) of the detectors are the most affected.

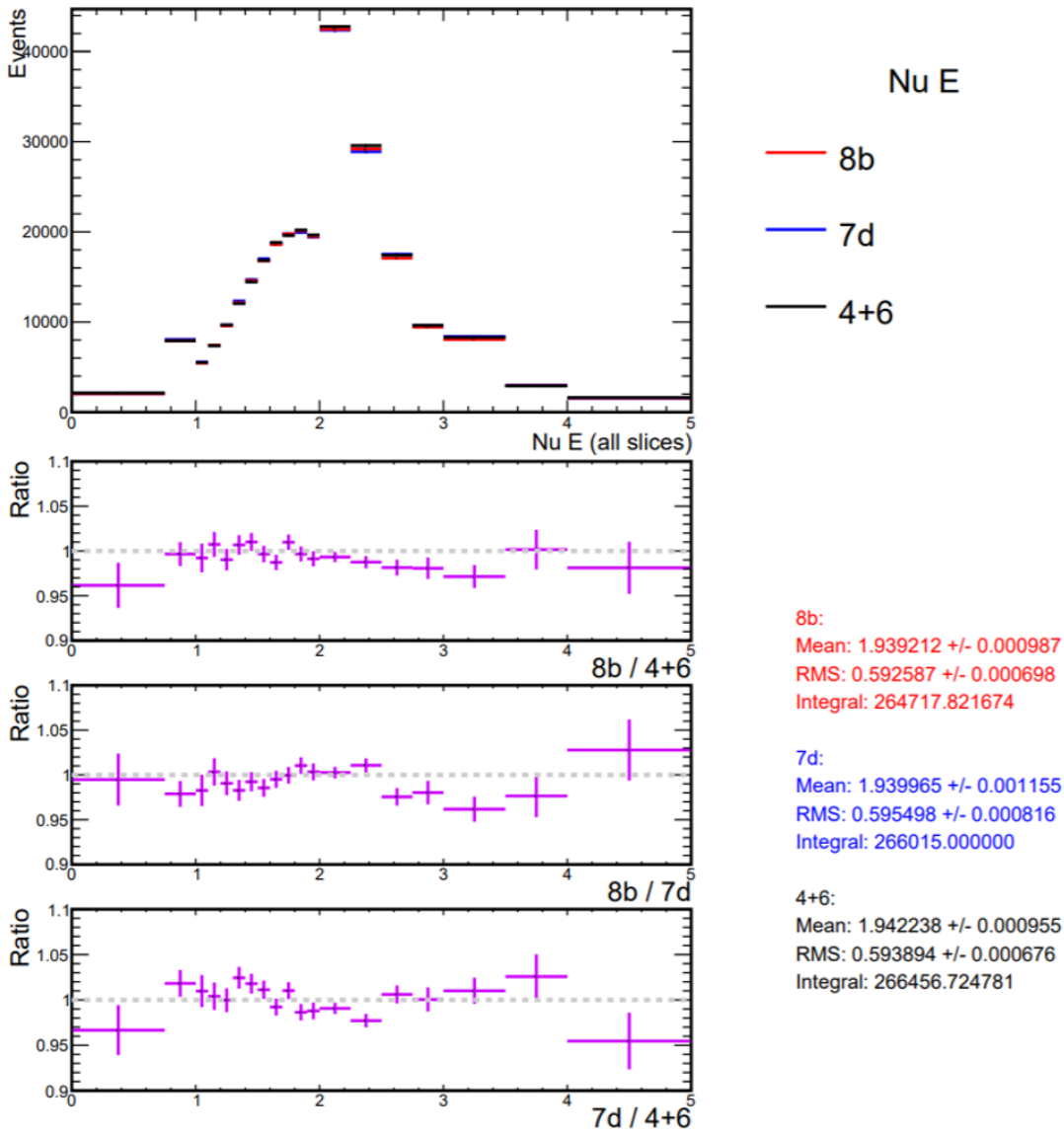


Figure B.1: Reconstructed ν_μ CC energy for events selected with the normal ν_μ CC cuts during epoch 8b (red), epoch 7d (blue) and periods 4+6 (black)

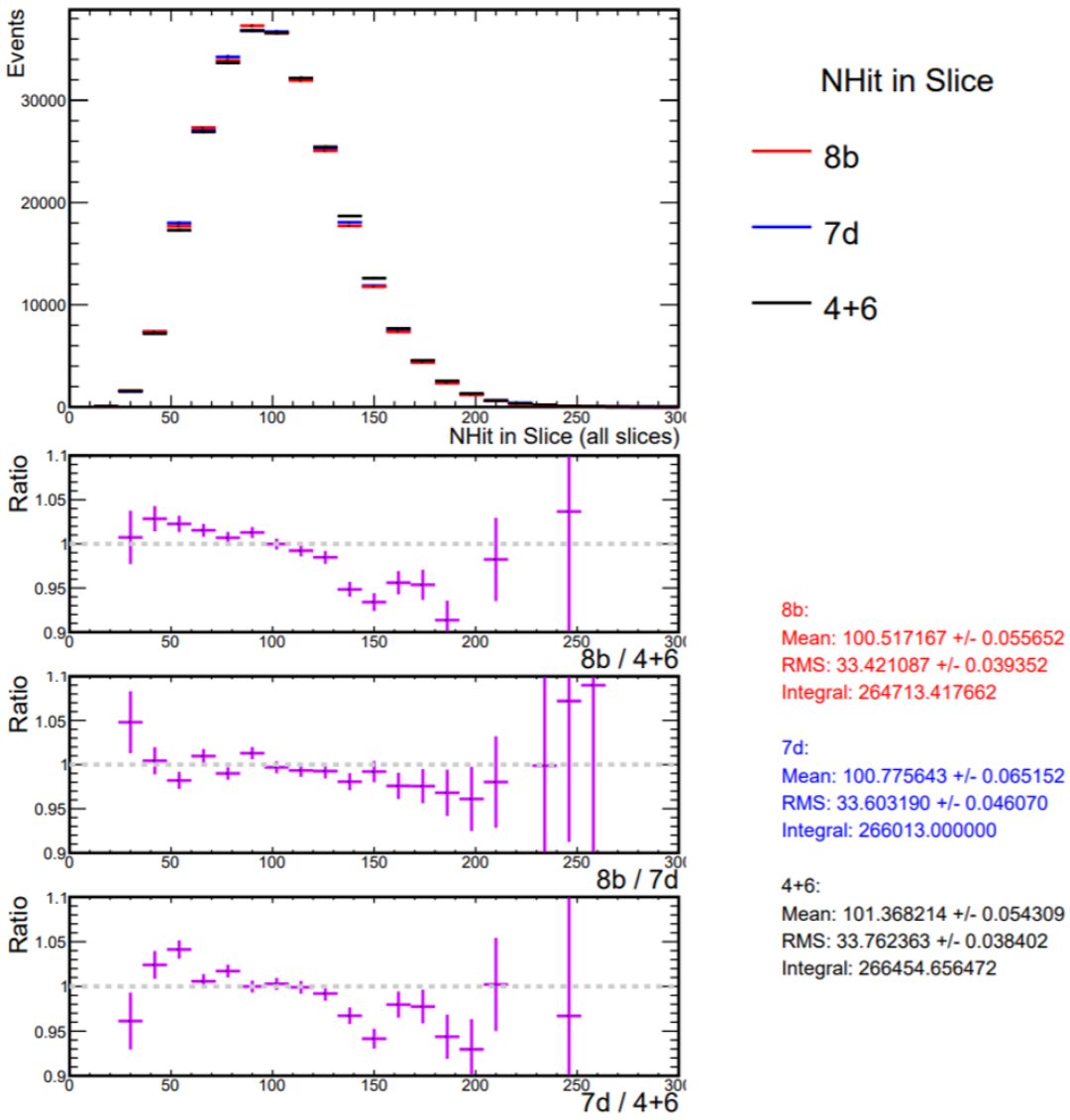


Figure B.2: Number of hits per slice for events selected with the normal ν_μ CC cuts during epoch 8b (red), epoch 7d (blue) and periods 4+6 (black)

B.2 Aging in the FD

To start quantifying and characterizing in a meaningful way, we turned to the FD due to the vast amount of cosmic muons seen there. These cosmics are more easily separable from beam events than at the ND, and provide a large enough sample to be able to see smaller effects with statistical significance. We only considered data from period 3 (starting in October 2015) and later as period 3 was when the FD switched from low to high gain making it hard to separate aging effects from gain differences.

Ideally, the dedicated cosmic muon sample made from our cosmic muon trigger would be used for any study like this. However, at the time of the initial study in 2019, that sample had only been processed for data through 2017 due to the intensive computational resources it requires, meaning there were 2 years worth of eligible data that would be excluded from the study. Instead, we used the NuMI beam data with a cut on the timing of the hits to remove the beam spill window. The downside of this is that we don't have data during the summer shutdown months, but we were able to track changes from October 2015 all the way through February 2019, ie. periods 3 through 8.

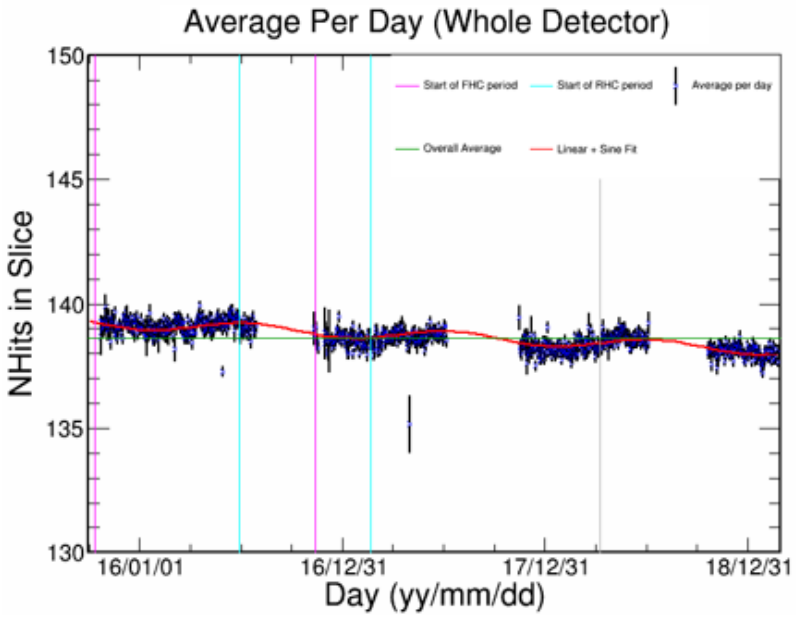


Figure B.3: Daily average number of hits in a slice for the FD cosmic data. The vertical lines demarcate changes in beam mode from FHC to RHC. The horizontal green line indicates the overall average number of hits in slice. The red line is a sinusoidal+linear fit to the data.

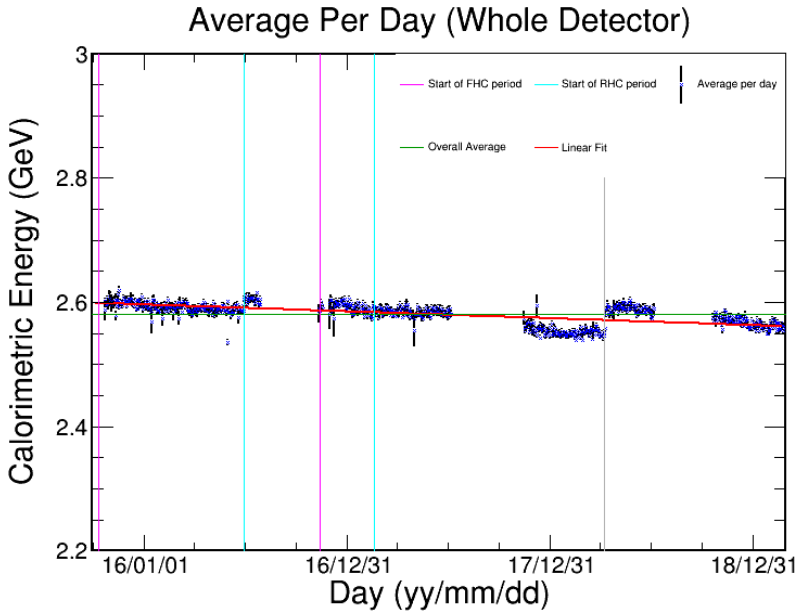


Figure B.4: Daily average calorimetric energy per slice for the FD cosmic data. The vertical lines demarcate changes in beam mode from FHC to RHC. The horizontal green line indicates the overall average number of hits in slice. The red line is a linear fit to the data.

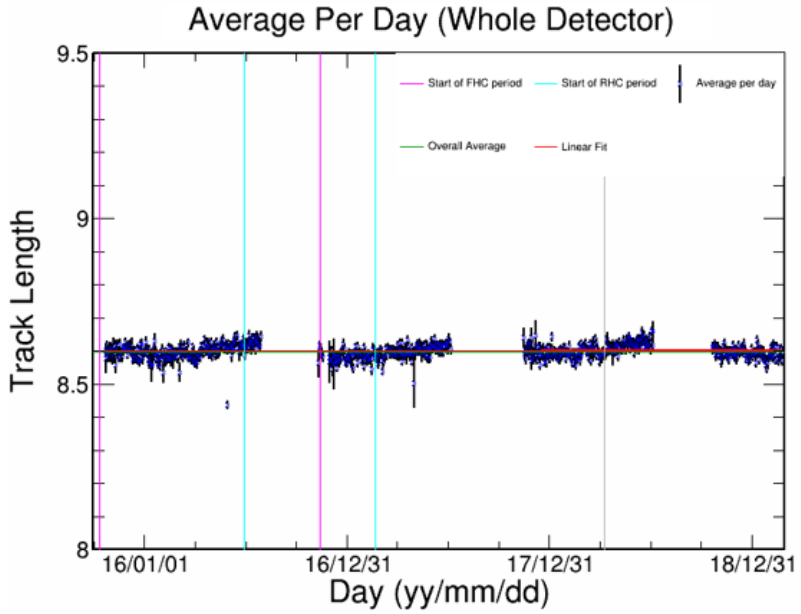


Figure B.5: Daily average cosmic track length for the FD cosmic data. The vertical lines demarcate changes in beam mode from FHC to RHC. The horizontal green line indicates the overall average number of hits in slice. The red line is a linear fit to the data.

Fig.B.3-B.5 show the change in hits per slice, calorimetric energy, and track length over time, with a sinusoidal plus linear fit applied to the hits in slice plot and just a linear fit applied to the other plots. The sinusoidal piece to the fit was to account for known seasonal patterns seen in the cosmic data at the FD.[49] The frequency was set to a year and all other fit parameters were allowed to float. Track length showed to remain constant over time which was reassuring as it meant we can attribute a decrease in hits in slice to signal loss rather than to changes in the nature of our cosmic sample. The calorimetric energy had several jumps to higher energies at the starts of some periods before decreasing throughout the rest of the period, as well as stark differences between FHC and RHC data. This was attributed to the fact that calibration constants were calculated separately for each of these time periods and that the energy estimators used for the RHC and FHC were different. It proved too difficult to separate these effect from any potential aging effects via a fit like what was done for the number of hits in slice distribution. Looking at the decrease in energy in each period still qualitatively gives evidence for some kind of aging, even if we could not use it for any kind of useful quantitative measurement.

We determined that the most useful measure of aging at this stage in the investigation was the number of hits in a slice as that quantity relies the least on reconstruction processes and can be most easily traced back to signal loss as the main cause[2]. Taking the slope of the linear piece of the fit as rate of loss of hits in slice, we can determine that there is a loss of 0.328 ± 0.007 hits per slice per year which is $0.237 \pm 0.005\%$ of the mean number of hits in slice (139 hits).

To get a sense of what this number actually means, we compared it to the effect of turning the light level of the detector down with the corresponding correcting upward adjustment in the calibration energy scale. The light level model for the detector is the model we've created and tuned to describe how we expect light signal to be generated in the scintillator and propagated through the wavelength shifting fibers to the readout electronics. By turning up or down the light level, we mean that we increase or decrease the normalization of the parameterisations for the light level model to effectively increase or decrease the efficiency of the detector to see and collect light signal from charged particle traveling through the detector. The adjustment in the calibration energy scale is just to make sure that this increase or decrease in detector efficiency does not change the energies of the particles simulated in the detector. Samples created with the light level turned up or down are used to create

[2]The detector aging would cause less signal to make it to the readout electronics meaning it will be more likely for a hit to be under threshold and not actually be classified as a hit.

systematic uncertainties to account for potential differences in our light level model and simulation and actual detector response.

Figures B.6-B.7 shows the comparison of the light level down sample to the nominal FD MC for both the RHC and the FHC. Table B.1 compares the difference in hits in slice lost from the light level down sample from the nominal MC to the loss over time in the data. It should be noted that this cannot be taken as a one-to-one comparison as the light level down and nominal MC samples are taken from simulated neutrino events whereas the data samples are purely cosmic muons and there is no evidence that the loss of hits in slice would be the same between cosmic muons and neutrino interactions. Therefore, all comparisons that can be made must be only between % differences and not absolute differences.

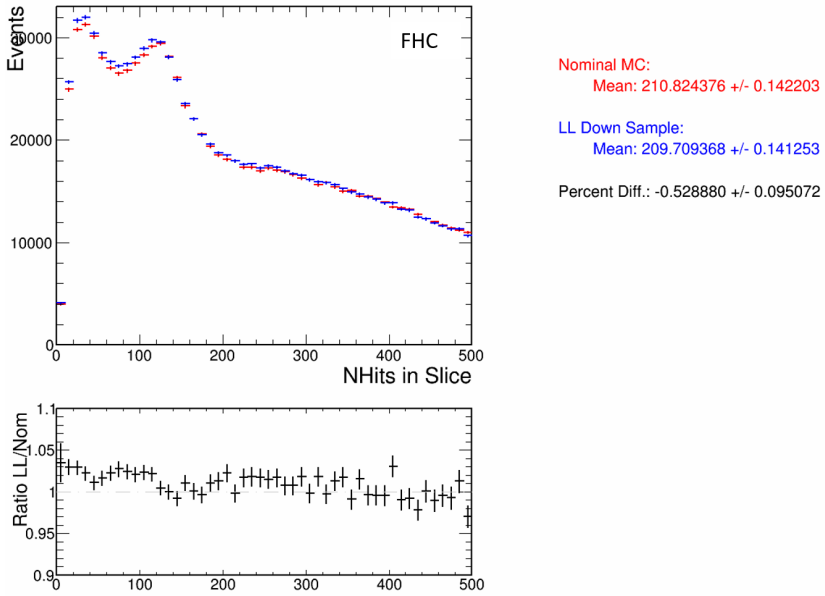


Figure B.6: Number of hits per slice for the FD FHC nominal MC and Light Level Down MC.

Overall, we can see from Table B.1 that a 10% loss in light level roughly corresponds to about 2 years worth of aging in the cosmic muon data sample, confirming that we are approaching the time (if we haven't reached it already) that the aging effects will be larger than the current Light Level systematic uncertainty allows for. Therefore, for the 2020 Analysis, a new systematic uncertainty was made to account for this drift in number of hits per slice using the approximation of roughly 2 years worth of aging corresponds to a 10% decrease in light level in the FD. This was done by creating a simulated sample by applying a linear downward gradient of 4.5% per year to the parameterisation of the light

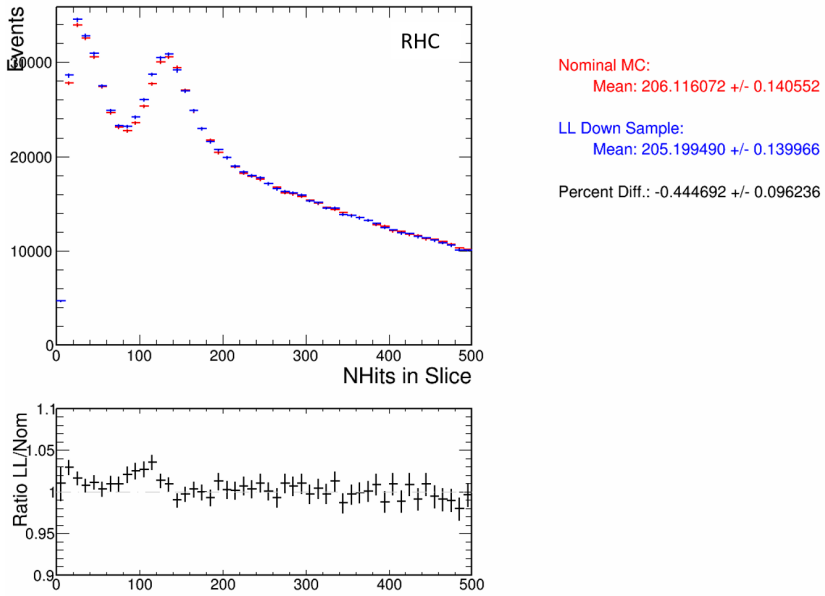


Figure B.7: Number of hits per slice for the FD RHC nominal MC and Light Level Down MC.

level model with respect to the midpoint date of January 1st, 2017. In the future, we want to develop a more sophisticated systematic uncertainty to handle this aging, requiring further characterization and quantifying of the effect.

Sample	Nominal NHits/slice	Absolute Shift (hits)	Relative Shift
Cosmic Muon Data	138.611 ± 0.007	−0.328 ± 0.007/year	−0.237±0.005%/year
FHC 10% Light Level Down	210.8 ± 0.1	−1.1 ± 0.2	−0.529 ± 0.95%
RHC 10% Light Level Down	206.1 ± 0.1	−0.9 ± 0.2	−0.445 ± 0.096%

Table B.1: Drift in hits per slice for data as well as for the light level 10% down sample compared to the nominal MC. The Relative Shift for the data was calculated by comparing the drift per year value to the overall average value for hits per slice. For the light level down samples, the relative shift was calculated by comparing the absolute difference in hits per slice from the nominal MC sample to the mean hits per slice in the nominal MC sample.

B.3 Potential Sources of Aging

There are several parts of the detector that are the likely culprits for the aging. Most likely is the scintillator and fibers inside each cell. The effect of either of those aging would directly create a lowering of the light level of the detector and therefore a decrease in the detector's efficiency. The other likely culprit would be the readout electronics themselves. It doesn't matter how great our detector is at collecting light signal if the electronics have a decreasing ability to convert that light into a digital signal that we can actually use in our analyses. We suspect that a majority of the aging is coming from the scintillator and fiber instead of the electronics as the electronics are constantly monitored and routinely swapped out when they start showing a decline in performance. That isn't to say that all of the electronics can't show some level of degradation over time, whether they are sitting on the shelf or are actively being used on the detector.

One further study that has been done was to start trying to separate out the effect of the scintillator and fiber aging to the electronics aging. This was done by looking at changes in the cell efficiencies along the cell length over time. We would expect for there to be a decrease in the cell efficiency over time if the detector is aging, but where along the cell those decreases are worse could tell us about what is aging. If just the electronics were aging, then the decrease in efficiency would be more or less uniform over the whole cell length as all of the signal loss would happen at the readout of cell, not while the signal was traveling. Aging in the scintillator and/or fibers would also lower the overall efficiency, but the decrease would be much worse in the far end of the cells as degradation in the scintillator and fibers would make attenuation worse alongside lowering the overall light level. We cannot totally disentangle the electronics aging from the scintillator/fibers aging by looking at the cell efficiencies, but we can at least qualitatively see which cause is the more prominent.

For this study, the nearline files for every December between 2015 and 2018 were used. These are files that are processed close to real time to the data being collected and are primarily used for close to real time monitoring data quality and detector health. These were used for this study because they were more readily available than regular files and already had the necessary cell efficiency variables calculated for this study. Only the month of December was used because the nature of the nearline files meant that the computational resources required to run over all of the files since Period 3 would have been beyond what this study could reasonably ask for. December was chosen as it was the month of the year

with the most data and is usually one of the most stable months as far the detector uptimes go.

For each December separately, two 2D histograms were collected that were cell by distance to readout. One was filled with the tricell hits that the detector actually saw on each cosmic track and the other was filled with the number of tricell hits that we had expected to see for each cosmic track. The cell efficiency is calculated by taking the ratio of the actual hits histogram to the expected hits histogram. To correctly find the average cell efficiency as a function of distance to readout from the resulting plot of this, a projection of the 2D efficiency plot was first made as a function of distance to the readout. Then each bin of distance to readout was divided by the number of cells that had contributed hits to the original expected hits plot for that bin of distance to readout.

Fig. X shows the results of this for each of the four Decembers that were included in this study, as well as the absolute and relative differences between December 2015 and December 2018. There is a clear overall decrease in cell efficiency over time, with the far end of the cells seeing a much more drastic drop in efficiency than the near end. This indicates that although the electronics are still likely to be contributing to the overall detector aging, the effect is dwarfed by the aging in the scintillator and fiber.

There were also initial looks at trying to separate out gain changes in the APDs from aging effects from the scintillator by looking at how the ADC (PE) distribution width changes as a function of time for each bin in W. The idea here is that the width should be proportional to $\sqrt{F * PE}$ where F is a factor dependent on gain. If F is changing over time, then that indicates a gain change over time - likely from electronics aging. Future detector aging studies should start there. It is also suggested that the production of the resulting plots and the cell efficiency plots be adopted into the regular data quality routine to make it easier to monitor the aging through the rest of the experiments run time.

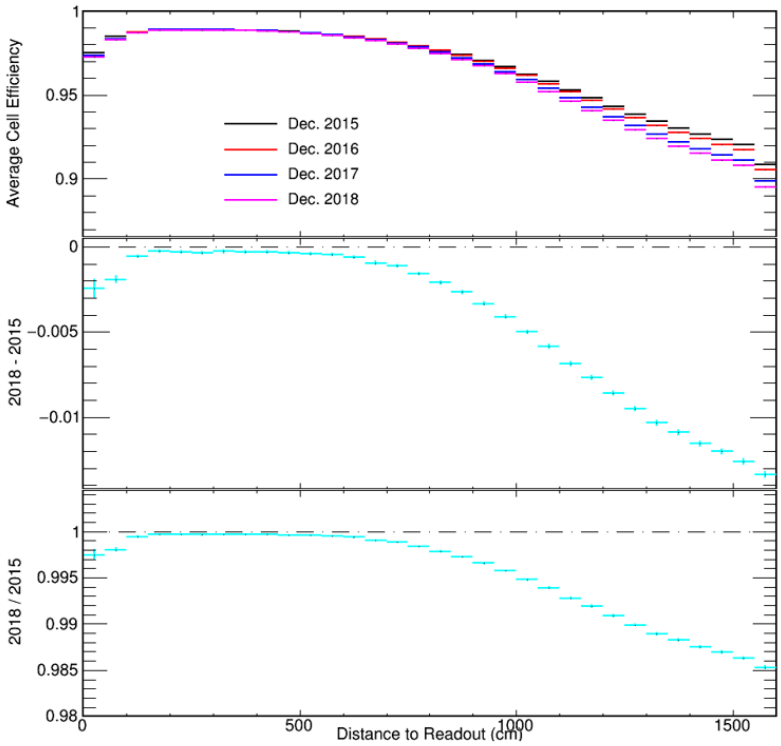


Figure B.8: Average cell efficiency as a function of distance to readout. Top plot shows the efficiencies for each Decemeber between 2015-2018. Middle(Top) plot shows the absolute(relative) difference between December 2015 and December 2018.

C. Efficiency Uncertainty Calculations

Finding the uncertainty on a cut selection efficiency must be done with some care. In this appendix, we will derive the selection efficiency uncertainty used in this thesis, as was described in [47, 48].

First, let's define the efficiency that we are trying to find the uncertainty of. For preselected events n, let the number of events that pass the selection cut be k such that the selection efficiency ε can be defined as $\varepsilon = \frac{k}{n}$.

As we are dealing with histograms in this case, we could stat with looking at using a Poissonian error for k and n, as would typically be done for histogram bin counts. We would then have the following

$$\sigma_k = \sqrt{k} \tag{C.1}$$

$$\sigma_n = \sqrt{n} \tag{C.2}$$

$$\begin{aligned} \sigma_\varepsilon &= \frac{k}{n}\sqrt{\left(\frac{\sigma_k}{k}\right)^2 + \left(\frac{\sigma_n}{n}\right)^2} \\ &= \frac{k}{n}\sqrt{\left(\frac{\sqrt{k}}{k}\right)^2 + \left(\frac{\sqrt{n}}{n}\right)^2} \\ &= \sqrt{\frac{k^2(n+k)}{n^3}} \end{aligned} \tag{C.3}$$

Already there are issues with this treatment when looking at the limiting cases of $k = 0$ or $k \to n$. If $k = 0$, then this uncertainty would give us an efficiency of 0 ± 0 regardless of what n is. However, we would expect there to be some error, even if there were no selected events in a sample. On the other end of the spectrum, as $k \to n$, the uncertainty becomes larger than 1. The resulting reportable range for the efficiency would then have a lower value below 0, which is unphysical.

These problems arise in part because k and n are not independent measurements since k is a subset of n. To account for this correlations, we could instead turn to a Binomial error treatment, where we reframe the question as finding the probability $P(k)$ that k events will pass the cut given prior knowledge I that this is a binomial process (either each event passes the cut, or they don't) and that ε must be between 0 and 1 to be physical. This probability,

given the true efficiency ε and preselected events n, can be written as

$$P(k|\varepsilon, n, I) = \binom{n}{k} \varepsilon^k (1-\varepsilon)^{n-k} \tag{C.4}$$

From here, we could substitute in our estimate for the efficiency $\varepsilon = \frac{k}{n}$, and find the standard of deviation from there. However, this again doesn't behave well in the limiting cases.

Instead, we should apply Baye's theorem to convert Eq.C.4 into $P(\varepsilon|k, n, I)$. Doing this gives us

$$P(\varepsilon|k, n, I) = \frac{P(k|\varepsilon, n, I)P(\varepsilon|n, I)}{Z} \tag{C.5}$$

where Z is a normalization constant, and $P(\varepsilon|n, I)$ is the probability that the true efficiency will be between ε and $\varepsilon + d\varepsilon$. Next, we can apply the prior knowledge we have that this a binomial process and that ε must be between 0 and 1. Th first means we can substitute Eq.C.4 into Eq.C.5 as well as

$$P(\varepsilon|n, I) = \begin{cases} 1 & \text{if } 0 \leq \varepsilon 1 \\ 0 & \text{otherwise} \end{cases} \tag{C.6}$$

We can set up the following to solve for the normalization constant Z

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} P(k|\varepsilon, n, I) \\ &= \frac{1}{Z} \frac{n!}{k!(n-k)!} \int_0^1 \varepsilon^k (1-\varepsilon)^{n-k} \end{aligned} \tag{C.7}$$

We can use the Beta function to solve for Z to arrive at

$$\begin{aligned} P(\varepsilon|k, n) &= \frac{\Gamma(n+2)}{\Gamma(k+1)\Gamma(n-k+1)} \varepsilon^k (1-\varepsilon)^{(n-k)} \\ &= \frac{(n+1)!}{k!(n-k)!} \varepsilon^k (1-\varepsilon)^{(n-k)} \end{aligned} \tag{C.8}$$

This probability now performs as expected in the limiting cases, with insurances that the efficiency cannot be outside of (0,1). Furthermore, we can now assign a non-zero probability to $\varepsilon = 0$ only if $k = 0$ as well. Similar for $\varepsilon = 1$ only has a non-zero probability if $k = n$.

From this probability, we can find the mean $\bar{\varepsilon}$ to be

$$\begin{aligned} \bar{\varepsilon} &= \int_0^1 \varepsilon P(\varepsilon|k,n) d\varepsilon \\ &= \frac{k+1}{n+2} \end{aligned} \tag{C.9}$$

and the mode of ε by solving $\frac{dP}{d\varepsilon} = 0$. This comes to be $\frac{k}{n}$, our estimate for the efficiency. Calculating the variance $V(\varepsilon)$ now follows as

$$\begin{aligned} V(\varepsilon) &= \bar{\varepsilon^2} - \bar{\varepsilon}^2 \\ &= \int_0^1 \varepsilon^2 P(\varepsilon|n,k) d\varepsilon - \bar{\varepsilon}^2 \\ &= \frac{(k+1)(k+2)}{(n+2)(n+3)} - \frac{k+1}{n+2} \end{aligned} \tag{C.10}$$

Given this variance and looking at how it behaves, Ullrich and Xu[48] argue that it is not unreasonable to assign the uncertainty on the efficiency to be the standard deviation $\sigma_\varepsilon = \sqrt{V(\varepsilon)}$, which is what was done for this thesis.

D. Glossary

ADC Analog to Digital Converter - refers to both the electronics and pulse signal coming off our detectors
APD Avalanche Photodiode
CC Charged Current, referring to the type of weak interaction involving a W boson and has an outgoing non-neutrino lepton.
Cell base unit of our detectors
CNN Convolusional Neural Network
CRY Cosmic-Ray Shower Library - generator used to make our cosmic ray simulations
CVN Convolusional Visual Network, the name for our CNN event PID
DCM Data Concentrator Module
DDT Data-Driven Trigger
dE/dx energy deposited by a charged particle per distance travelled. Also referred to as the energy loss rate.
Epoch subset of a Period
FD/ND/TB - Far Detector, Near Detector, Test Beam detector
FEB Front End Board
FHC - Forward Horn Current - beam mode that produces a mostly neutrino beam
Hit discrete signal in a cell as seen by our detectors
Light Level measure of how well a detector, particularly the scintillator, is able to produce light signal from a charge particle moving through it
LOWESS Locally Weighted Scatterplot Smoothing
MC Monte Carlo - NOvA's main simulation samples
MEU Muon Energy Unit
MIP Minimum Ionizing Particle
MRE Muon Removed Electron Added
MRBrem Muon Removed Bremsstrahlung
MRDiF Muon Removed Decay-in-Flight
NC Neutral Current referring to the type of weak interaction involving a Z boson and where the outgoing lepton is a neutrino
PC hits Pre-Calibrated hits - tricell hits before calibration is applied

PE, PECorr Photoelectrons, corrected photoelectrons - the amount of light signal coming off the APD in units of photoelectrons with PECorr accounting for signal attenutation
Period refers to a data taking run. Usually delineated by changes in run or beam conditions such as a switch from FHC to RHC beam mode
PID Particle Identifier
Plane layer in our detector made of up either horizontally or vertically aligned cells
POT Protons on Target - measure of exposure of the beam to our detectors
Prong cluster of hits that make up a single particle track or shower within a slice
RHC - Reverse Horn Current - beam mode that produces a mostly antineutrino beam
Slice cluster of hits that make up a particle interaction
TDU Timing Distribution Unit
ToF Time of Flight

Bibliography

[1] Charles Drummond Ellis and W. A. Wooster. The average energy of disintegration of radium e. *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences*, 117:109–123, 1927.

[2] W. Pauli. Dear radioactive ladies and gentlemen. *Phys. Today*, 31N9:27, 1978.

[3] F. L. Wilson. Fermi's theory of beta decay. *American Journal of Physics*, 36:1150–1160, 1968.

[4] B. J. P. Jones. The physics of neutrinoless double beta decay: A primer, 2021. URL https://arxiv.org/abs/2108.09364.

[5] John N. Bahcall, Neta A. Bahcall, and Giora Shaviv. Present status of the theoretical predictions for the ^{37}Cl solar-neutrino experiment. *Phys. Rev. Lett.*, 20:1209–1212, May 1968. doi: 10.1103/PhysRevLett.20.1209. URL https://link.aps.org/doi/10.1103/PhysRevLett.20.1209.

[6] Raymond Davis, Don S. Harmer, and Kenneth C. Hoffman. Search for neutrinos from the sun. *Phys. Rev. Lett.*, 20:1205–1209, May 1968. doi: 10.1103/PhysRevLett.20.1205. URL https://link.aps.org/doi/10.1103/PhysRevLett.20.1205.

[7] John N. Bahcall and Raymond Davis. Solar neutrinos: A scientific puzzle. *Science*, 191(4224):264–267, 1976. doi: 10.1126/science.191.4224.264. URL https://www.science.org/doi/abs/10.1126/science.191.4224.264.

[8] B. Pontecorvo. Inverse beta processes and nonconservation of lepton charge. *Zh. Eksp. Teor. Fiz.*, 34:247, 1957.

[9] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. Remarks on the Unified Model of Elementary Particles. *Progress of Theoretical Physics*, 28(5):870–880, 11 1962. ISSN 0033-068X. doi: 10.1143/PTP.28.870. URL https://doi.org/10.1143/PTP.28.870.

[10] B. Pontecorvo. Neutrino Experiments and the Problem of Conservation of Leptonic Charge. *Zh. Eksp. Teor. Fiz.*, 53:1717–1725, 1967.

[11] V. Gribov and B. Pontecorvo. Neutrino astronomy and lepton charge. *Physics Letters B*, 28(7):493–496, 1969. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(69)90525-5. URL https://www.sciencedirect.com/science/article/pii/0370269369905255.

[12] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, Aug 1998. doi: 10.1103/PhysRevLett.81.1562. URL https://link.aps.org/doi/10.1103/PhysRevLett.81.1562.

[13] Q. R. Ahmad et al. Direct evidence for neutrino flavor transformation from neutral-current interactions in the sudbury neutrino observatory. *Phys. Rev. Lett.*, 89:011301, Jun 2002. doi: 10.1103/PhysRevLett.89.011301. URL https://link.aps.org/doi/10.1103/PhysRevLett.89.011301.

[14] Tsutomu Yanagida. Horizontal Symmetry and Masses of Neutrinos. *Progress of Theoretical Physics*, 64(3):1103–1105, 09 1980. ISSN 0033-068X. doi: 10.1143/PTP.64.1103. URL https://doi.org/10.1143/PTP.64.1103.

[15] CARLO GIUNTI and MARCO LAVEDER. SHORT-BASELINE ACTIVE-STERILE NEUTRINO OSCILLATIONS? *Modern Physics Letters A*, 22(33):2499–2509, oct 2007. doi: 10.1142/s0217732307025455. URL https://doi.org/10.1142/s0217732307025455.

[16] L. Wolfenstein. Neutrino oscillations in matter. *Phys. Rev. D*, 17:2369–2374, May 1978. doi: 10.1103/PhysRevD.17.2369. URL https://link.aps.org/doi/10.1103/PhysRevD.17.2369.

[17] A. Yu. Smirnov. The msw effect and solar neutrinos, 2003. URL https://arxiv.org/abs/hep-ph/0305106.

[18] R. L. Workman et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022. doi: 10.1093/ptep/ptac097.

[19] S. Agostinelli et al. GEANT4–a simulation toolkit. *Nucl. Instrum. Meth. A*, 506: 250–303, 2003. doi: 10.1016/S0168-9002(03)01368-8.

[20] L. Aliaga et al. Neutrino flux predictions for the NuMI beam. *Physical Review D*, 94(9), nov 2016. doi: 10.1103/physrevd.94.092005. URL https://doi.org/10.1103/physrevd.94.092005.

[21] C. Andreopoulos, A. Bell, D. Bhattacharya, F. Cavanna, J. Dobson, S. Dytman, H. Gallagher, P. Guzowski, R. Hatcher, P. Kehayias, A. Meregaglia, D. Naples, G. Pearce, A. Rubbia, M. Whalley, and T. Yang. The genie neutrino monte carlo generator. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 614(1):87–104, 2010. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2009.12.009. URL https://www.sciencedirect.com/science/article/pii/S0168900209023043.

[22] M. A. Acero et al. Adjusting neutrino interaction models and evaluating uncertainties using NOvA near detector data. *The European Physical Journal C*, 80(12), dec 2020. doi: 10.1140/epjc/s10052-020-08577-5. URL https://doi.org/10.1140/epjc/s10052-020-08577-5.

[23] C Hagmann, D Lange, and D Wright. Cosmic-ray shower library (cry). 2012. *URL https://nuclear. llnl. gov/simulation/doc_cry_v1*, 7.

[24] A Aurisano, C Backhouse, R Hatcher, N Mayer, J Musser, R Patterson, R Schroeter, and A Sousa. The nova simulation chain. *Journal of Physics: Conference Series*, 664 (7):072002, dec 2015. doi: 10.1088/1742-6596/664/7/072002. URL https://dx.doi.org/10.1088/1742-6596/664/7/072002.

[25] Alex Rodriguez and Alessandro Laio. Clustering by fast search and find of density peaks. *Science*, 344(6191):1492–1496, 2014. doi: 10.1126/science.1242072. URL https://www.science.org/doi/abs/10.1126/science.1242072.

[26] R. C. Prim. Shortest connection networks and some generalizations. *The Bell System Technical Journal*, 36(6):1389–1401, 1957. doi: 10.1002/j.1538-7305.1957.tb01515.x.

[27] Leandro A. F. Fernandes and Manuel Menezes de Oliveira Neto. Real-time line detection through an improved hough transform voting scheme. *Pattern Recognit.*, 41: 299–314, 2008.

[28] Track finding with deformable templates — the elastic arms approach. *Computer Physics Communications*, 71(1):77–98, 1992. ISSN 0010-4655. doi: https://doi.org/10.1016/0010-4655(92)90074-9. URL https://www.sciencedirect.com/science/article/pii/0010465592900749.

[29] Miin-Shen Yang and Kuo-Lung Wu. Unsupervised possibilistic clustering. 39(1), 2006. ISSN 0031-3203. doi: 10.1016/j.patcog.2005.07.005. URL https://doi.org/10.1016/j.patcog.2005.07.005.

[30] Raghu Krishnapuram and James M. Keller. A possibilistic approach to clustering. *IEEE Trans. Fuzzy Syst.*, 1:98–110, 1993.

[31] P. Billoir and S. Qian. Simultaneous pattern recognition and track fitting by the kalman filtering method. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 294(1):219–228, 1990. ISSN 0168-9002. doi: https://doi.org/10.1016/0168-9002(90)91835-Y. URL https://www.sciencedirect.com/science/article/pii/016890029091835Y.

[32] A. Aurisano, A. Radovic, D. Rocco, A. Himmel, M.D. Messier, E. Niner, G. Pawloski, F. Psihas, A. Sousa, and P. Vahle. A convolutional neural network neutrino event classifier. *Journal of Instrumentation*, 11(09):P09001, sep 2016. doi: 10.1088/1748-0221/11/09/P09001. URL https://dx.doi.org/10.1088/1748-0221/11/09/P09001.

[33] P. Adamson et al. Constraints on oscillation parameters from ν_e appearance and ν_μ disappearance in NOvA. *Physical Review Letters*, 118(23), jun 2017. doi: 10.1103/physrevlett.118.231801. URL https://doi.org/10.1103/physrevlett.118.231801.

[34] F. Psihas, E. Niner, M. Groh, R. Murphy, A. Aurisano, A. Himmel, K. Lang, M. D. Messier, A. Radovic, and A. Sousa. Context-enriched identification of particles with a convolutional network for neutrino events. *Physical Review D*, 100(7), oct 2019. doi: 10.1103/physrevd.100.073005. URL https://doi.org/10.1103/physrevd.100.073005.

[35] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions, 2014. URL https://arxiv.org/abs/1409.4842.

[36] M. A. Acero et al. Improved measurement of neutrino oscillation parameters by the NOvA experiment. *Phys. Rev. D*, 106(3):032004, 2022. doi: 10.1103/PhysRevD.106.032004.

[37] Jonathon Shlens. A tutorial on principal component analysis, 2014. URL https://arxiv.org/abs/1404.1100.

[38] M. Tanabashi et al. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018. doi: 10.1103/PhysRevD.98.030001. URL https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[39] M. A. Acero et al. The Profiled Feldman-Cousins technique for confidence interval construction in the presence of nuisance parameters. 7 2022.

[40] S. S. Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Annals Math. Statist.*, 9(1):60–62, 1938. doi: 10.1214/aoms/1177732360.

[41] J. Neyman. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 236(767):333–380, 1937. ISSN 00804614. URL http://www.jstor.org/stable/91337.

[42] Patrick Dunne. Latest neutrino oscillation results from t2k, July 2020. URL https://doi.org/10.5281/zenodo.3959558.

[43] K. Abe others. Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations. *Nature*, 580(7803):339–344, apr 2020. doi: 10.1038/s41586-020-2177-0. URL https://doi.org/10.1038/s41586-020-2177-0.

[44] P. Adamson et al. Precision constraints for three-flavor neutrino oscillations from the full MINOS+ and minos dataset. *Phys. Rev. Lett.*, 125:131802, Sep 2020. doi: 10.1103/PhysRevLett.125.131802. URL https://link.aps.org/doi/10.1103/PhysRevLett.125.131802.

[45] K. Abe et al. Atmospheric neutrino oscillation analysis with external constraints in super-kamiokande i-IV. *Physical Review D*, 97(7), apr 2018. doi: 10.1103/physrevd.97.072001. URL https://doi.org/10.1103/physrevd.97.072001.

[46] M.G. Aartsen et al. Measurement of atmospheric neutrino oscillations at 6–56 GeV with IceCube DeepCore. *Physical Review Letters*, 120(7), feb 2018. doi: 10.1103/physrevlett.120.071801. URL https://doi.org/10.1103/physrevlett.120.071801.

[47] Marc Paterno. Calculating efficiencies and their uncertainties. 12 2004. doi: 10.2172/15017262.

[48] T. Ullrich and Z. Xu. Treatment of errors in efficiency calculations, 2007. URL https://arxiv.org/abs/physics/0701199.

[49] M. A. Acero et al. Seasonal variation of multiple-muon cosmic ray air showers observed in the NOvA detector on the surface. *Phys. Rev. D*, 104(1):012014, 2021. doi: 10.1103/PhysRevD.104.012014.