Domain Generalization with Machine Learning in the NOvA Experiment

Andrew Thomas Cleve Sutton
Graham, North Carolina

B.S. Physics, University of North Carolina at Charlotte, 2016
B.S. Mechanical Engineering, University of North Carolina at Charlotte, 2016

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
February 2022

Acknowledgements

I'd like to thank my parents, Tym Cleve and Lisa Sutton, who never pressured me to pursue any one path and allowed me the freedom to explore and choose my own way. Thanks are also owed to my brother, Alex, who has pushed me to be my best by making nearly everything a competition, and my sister, Amanda, who can always make me laugh.

My fellow graduate students at UVA, particularly Grace Cummings and Akin Morrison, were the best friends and peers I could ask for. They helped me immensely to tackle my first two years of classes and get through the initial milestone that was the Qual. My advisor, Craig Group, allowed me the maximum freedom to pursue the topics that I was interested in and was always available to discuss both physics-related and personal topics. I owe a significant amount of gratitude to Michael Baird for being the greatest mentor and teacher that anyone could have.

I'd like to express my appreciation to the NOvA collaboration for assembling a wonderful array of physicists to share ideas and work to produce the best results possible. Young NOvA in particular has been the most welcoming and friendly group that I've ever been a part of. I'd also like to thank the Test Beam group; I always looked forward to a day out at FTBF, even when we were attacked by the Blob.

My deepest gratitude is owed to my wife, Hayley, whose has provided unparalleled understanding and support for my path in physics. Hayley's efforts have only been matched by those of my cat, Phoebe, whose companionship made the task of writing this manuscript both easier and more difficult with her desire to sit on my lap throughout much of it. My other cat, Cosmo, has been a large orange stress ball and gracefully put up with me rubbing his ample belly. Lastly, I'd like to thank my beautiful daughter, Sophia, whose birth in December provided an excellent motivation to complete my analysis and writing as soon as possible.

Contents

List of Tables

List of Figures

1. Neutrinos: A Desperate Remedy

Classical physics treats systems as continuous objects that can be broken down into infinitesimal arbitrarily small bits. In this picture, those infinitesimal bits behave exactly the same as the macroscopic object from which they are derived. The behavior of these objects is generally considered to be fully deterministic; that is, the evolution of a system can be completely determined with enough prior knowledge. However, through the 19^{th} century, various experiments brought this classical view into question, and by the turn of the century, the quantum hypothesis began to take hold. In this paradigm, the microscopic level becomes discrete rather than continuous. A particle physicist is concerned with understanding the nature of and interactions between the discrete quanta that make up the macroscopic matter that we observe daily.

Throughout the 20^{th} century particle physicists developed the "Standard Model" which aims to be a complete theory describing the properties and interactions of the fundamental constituents of the universe. The Standard Model has proven to be quite robust and has repeatedly provided theoretical predictions that have been borne out by experiments. In this model there are three types of particles: fermions (spin = $1/2$), gauge bosons (spin = 1), and scalar bosons (spin = 0). Gravity aside, particles can interact via the electromagnetic, weak, and strong forces, which are mediated by the gauge bosons: photons, W and Z bosons, and gluons respectively. The fermions can be further classified into quarks (these combine to make protons and neutrons) which participate in all three interactions, and leptons (e.g. electrons and neutrinos) which do not participate in strong interactions. The scalar bosons only have one known fundamental member, the Higgs boson, which is a result of the Higgs mechanism through which some of the other particles acquire mass.

Many open questions lie beyond the Standard Model (BSM), and in response, many hypotheses have been developed though there is still much to test. For a long time, the Standard Model has assumed neutrinos to be massless particles; however, we know for sure that this is not the case in reality. Additionally, the prevalence of matter over antimatter in our universe is the subject of many theoretical and experimental papers. Neutrinos offer a unique window as they provide a clear and direct probe into both of these phenomena the resolution of which will likely require some adjustment to the current Standard mModel.

1.1 Little Neutral One

Standard Model neutrinos are massless leptons with no electromagnetic charge, hence the name which comes from "neutral" + the diminutive Italian suffix "ino". Neutrinos, being uncharged leptons, only participate in weak interactions making them some of the most elusive particles to observe. Doing so typically requires very large detectors, a high-intensity source, or preferably both. As stated above and discussed later, the massless assumption has been shown to be incorrect. Neutrino masses are indeed quite small, being $O(10^6)$ times smaller than the next lightest fundamental particle, the electron.

The history of the neutrino can be traced back to the discovery of spontaneous radiation by Henri Becquerel in 1896. Over the course of numerous experiments, Becquerel showed that uranium salts emitted some form of intrinsic radiation that needed no additional energy input [1, 2, 3]. This discovery won Becquerel along with Marie and Pierre Curie the 1903 Nobel Prize in physics. Ernest Rutherford began investigating this new spontaneous radiation and showed in 1899 [4] that uranium released a "complex" radiation made of two components, α and β. In his studies, Rutherford showed that β-radiation was more penetrating but was less ionizing than α-radiation. It was later shown that the ionizing particles constituting β-radiation were identical to cathode rays, which are now called electrons.

Investigations into the nature of the types of radiation continued, and in 1914 James Chadwick, of neutron fame, showed that the energy spectrum of β-radiation was not discrete like α-radiation but was instead continuous with some peaks (the peaks were an artifact of the detection method) [5]. This feature was a mystery that clashed with the quantized energy levels that explained other types of radiation leading many to disbelieve Chadwick's results. The outbreak of the Great War halted Chadwick's research when he was arrested as a prisoner of war while in Germany in November 1914. While captured, Chadwick organized a small lab where he performed small-scale experiments and lectured on radioactivity [6]. A young English cadet named Charles Ellis developed an interest in the field under Chadwick's tutelage, and in 1927, with his student William Wooster, showed definitively that the β spectrum was continuous [7] (see Figure 1.1). There were many early hypotheses to explain the continuous β spectrum, one of which being a suggestion that a photon was also released along with the electron of definite energy. The supposition here being that secondary collisions occurred which generated the continuous spectrum. This possibility was ruled out by Ellis and Wooster in 1927 when they showed that no interacting secondary radiation

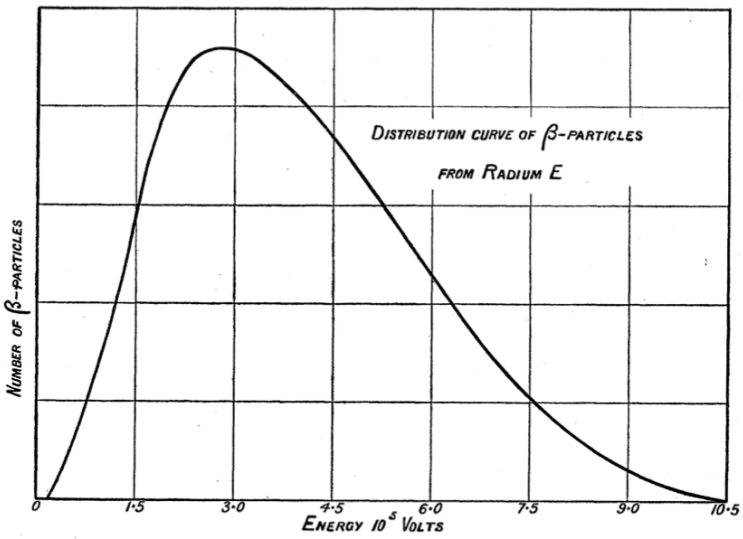


Figure 1.1: β spectrum of ^{210}Bi, historically called Radium-E [7].

was released along with the electron [7].

A continuous β spectrum was a problem. If a β-emitting nucleus starts off with a definite mass, then releases an electron, it then turns into a different nucleus with some lower mass. This requires the emission of an amount of energy exactly equal to the mass difference; however, with a continuous spectrum, the mass difference was not fully accounted for. The upper limit of the β spectrum matches the energy difference between the mother and daughter nuclei, but the average energy of the escaping electron is significantly lower. Some energy is missing and the equations are not balanced.

Wolfgang Pauli in 1930 proposed a "desperate remedy" in a letter he sent to a conference in 1930 that he could not attend [8]. Pauli suggested the existence of an electrically neutral, spin $1/2$ particle that he dubbed the "neutron" which resided in the nucleus. The desperation of physicists at this time can be summarized by advice which Pauli quotes in his letter: "Oh, it's better not to think about this at all, like new taxes." Pauli's neutron did not travel at the speed of light but had a mass on the same order of magnitude as the electron and was highly penetrating. Clearly, this particle would be nearly impossible to observe. Nonetheless, by adding a light neutral particle emitted alongside the electron, the continuous spectrum of β decay would be explained by each particle taking some amount of the total energy which would sum to be constant. This proposal was driven primarily by a desire to "save" the laws of energy and angular momentum conservation which are fundamental building blocks of physics.

With Chadwick's discovery of the neutron in 1932, Pauli's proposed particle was re-dubbed the "neutrino" owing to its hypothesized low mass. Enrico Fermi then set about developing a quantitative theory of β decay incorporating Pauli's neutrino [9]. Fermi's formulation was revolutionary. Rather than supposing that the emitted electrons and neutrinos resided in the nucleus, he took the view that the nucleus was made of only heavy protons and neutrons. Electrons and neutrinos are then be created during β emission and can also be annihilated through absorption similar to the successful theory of photon emission. With this framework, Fermi predicted the relative lifetime of various radioactive elements that undergo β decay and, more importantly, reproduced the observed continuous spectrum.

1.2 Ghost Hunting

Fermi's β theory provided a solid foundation, but the issue which prevented Pauli from initially publishing his hypothesis still remained: neutrinos had not been directly observed. In fact, using Fermi's theory and dimensionality arguments, Bethe and Peierls showed that the cross-section for a neutrino interaction would be on the order of $10^{-44}\,\mathrm{cm}^2$ giving a penetrating power of 10^{16} km in solid matter [10]. The neutrino's elusive nature persisted for nearly 20 years after Fermi formulated his theory. Various experiments attempted to measure the recoil of the nucleus after decay to fully account for the energy spectrum but were all unsuccessful. Even if these experiments had succeeded they would not yield incontrovertible evidence that neutrinos indeed existed since they would just show that the missing energy is indeed not given to the nucleus. In order to definitively prove the existence of the neutrino, the particle had to be directly observed.

Before moving forward it is important to note two things: (1) β decay comes in two types and (2) particle interactions can be shuffled around by moving particles from the right-hand side to the left and vice versa. The two types of β decay are: β^- where a neutron converts to a proton with emission of an electron and an antineutrino and β^+ where a proton turns into a neutron emitting a positron and a neutrino:

$$n \rightarrow p + e^- + \bar{\nu}, \tag{1.1}$$

$$p \rightarrow n + e^+ + \nu, \tag{1.2}$$

where the $\bar{\nu}$ denotes an antineutrino which is emitted in β decay to preserve the overall lepton number, L in the system (electrons and neutrinos have $L = 1$ while positrons and

antineutrinos have $L = -1$). So far the β decay we've been discussing has been specifically β^-, which is a common decay mode for neutron-rich heavy nuclei like those involved in nuclear fission.

Luis Walter Alvarez, in 1949, proposed an experiment to finally observe the direct effect of neutrinos [11] (Bruno Pontecorvo initially proposed a similar experiment in 1946, but his lecture was quickly classified by the U.S. government [12]). This proposal involved reversing the β decay process by exposing atoms to a neutrino source (nuclear reactor). A nucleus would then absorb a neutrino and change an internal neutron into a proton in a reversal of Eq. 1.1. The resulting nuclei with an increased atomic number would then be extracted and the amount of activity would determine if a neutrino-induced reaction took place, thus showing whether or not neutrinos exist.

Along similar lines, Clyde Cowan and Frederick Reines proposed an inverse β decay experiment in 1953 [14]. Rather than counting the daughter nuclei from a neutrino-induced reaction, they instead chose a method to observe the full reaction in situ. Their plan was to expose a large source of protons to antineutrinos produced by nuclear fission. The protons

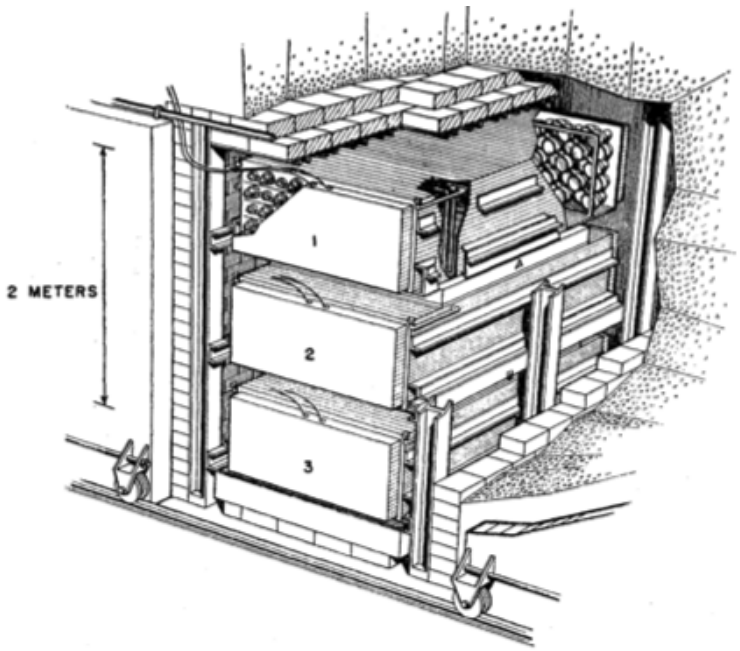


Figure 1.2: Sketch of the detector used by Cowan and Reines to first detect neutrinos. Tanks 1-3 were filled with liquid scintillator and viewed by PMTs. Tanks A and B in between the scintillator tanks contained the water and cadmium targets. The system was then encased in lead for shielding [13].

would absorb the antineutrinos, convert into neutrons, and release positrons in the process:

$$p + \bar{\nu} \rightarrow n + e^{+} \tag{1.3}$$

To detect this interaction, they built large liquid scintillation detectors doped with cadmium [15]. The positron would quickly annihilate with an electron in the scintillator releasing two prompt photons while the neutron would bounce around for some time until being captured by the cadmium releasing a delayed photon signal occurring about 5 µs after the positron signal. The detector design can be seen in Fig. 1.2. To be sure that they were measuring only neutrino interactions Cowan and Reines placed their detectors underground, encased them in lead shielding, and employed multiple tanks to act as veto counters for incoming cosmic rays that might mimic the neutrino capture signal. Their initial results in 1953 were promising [16] and with an improved detector and higher intensity reactor, they reported the first conclusive observation of the neutrino in 1956 [17].

1.3 Something New Under the Sun

For the purposes of this review, the next "big" event in neutrino physics was the so-called "Solar Neutrino Problem". Raymond Davis Jr. had previously attempted to use Alvarez's inverse beta decay technique to measure reactor antineutrinos. This method relied on capturing neutrinos on $^{37}_{17}$Cl which converts to $^{37}_{18}$Ar via the process in Eq. 1.4.

$$n + \nu \rightarrow p + e^{-} \tag{1.4}$$

The created argon, which has a half-life of 35 days and decays via electron capture, would then be chemically separated and the counted. Davis' original attempt to use this process on reactor antineutrinos failed since it relies on a neutrino interaction. However, it did show that the neutrino from Eq. 1.4 is distinct from the antineutrino in Eq. 1.3 [18].

In considering the backgrounds for the reactor experiment, Davis recognized that this technique could be used to measure the rate of neutrinos being emitted from the sun. The fusion reactions that power the sun emit neutrinos through either the p-p or carbon-nitrogen-oxygen cycle. In each cycle, neutrinos are produced during the fusion process and during the subsequent β decay of fusion products. Theorist John Bahcall teamed up with Davis to perform rigorous calculations of the neutrino flux originating in the sun. In 1964 Bahcall

and Davis simultaneously published two letters that acted as a proof of concept showing that meaningful theoretical calculations could be performed and that the experimental setup could be scaled to obtain the sensitivity needed to experimentally measure these reactions [19, 20].

The efforts of Bahcall and Davis continued over the next few years [21, 22]. Bahcall produced updated predictions which incorporated ongoing nuclear measurements and attempted to further understand and constrain the largest uncertainties (chief among which was the internal temperature of the sun). Davis devised a larger scale experiment with $\times 100$ more liquid tetrachloroethylene, C_2Cl_4, and located in the Homestake mine with over twice the overburden[1] compared to the 1964 experiment. The increased overburden, in turn, decreased the rate of cosmic-ray-induced background events from 3.5 counts per day to about 0.2 counts per day, well below Bahcall's predicted solar neutrino event rate of 2-7 counts per day. From this 1968 experiment, Davis measured the solar neutrino flux to be $\leq 0.3 \times 10^{-35}$ s^{-1} per ^{37}Cl atom [22], between 2-6 times lower than Bahcall's prediction of $2.0 \pm 1.2 \times 10^{-35}$ s^{-1} per ^{37}Cl atom [21]. Additionally, if all background counts were treated as signal, Davis' experimental bound would only increase to 0.6×10^{-35} s^{-1} per ^{37}Cl atom, still only about a third of the central value predicted by Bahcall.

The deficit of solar neutrinos was an interesting problem, though, it was not taken very seriously at the time. Although various hypotheses existed which could explain the neutrino deficit, physicists were reluctant to accept them owing to the significant change they would entail in the burgeoning Standard Model. Additionally, with a 60% error on the theoretical prediction, many physicists were willing to believe that the solar models were incorrect or at least were missing crucial inputs.

1.4 And Now For Something Slightly Different

Before discussing the resolution of the solar neutrino problem, I will first review some additional developments in particle physics that will come into play. Among these are the discovery that parity is not conserved in weak interactions, the theoretical development and observation of neutral current (NC) interactions, and the discovery of different neutrino flavors associated with the charged leptons.

[1] Overburden is the material above an experiment, this can be simply rocks on top of the building housing the detector (eg. the NOvA far detector), or it may be the Earth above the detector as a result of placing it in a mine or under a mountain. The purpose of overburden is to capture unwanted particles originating interaction in the atmosphere or in outer space.

As seen from the neutrino observation of Reines and Cowan, Fermi's β theory was quite successful. Fermi's theory was also branded as the "weak interaction" theory owing to the small value of the coupling constant, G_F, associated with these interactions. In Fermi's original paper, he estimated the value to be $O(10^{-6}\,\mathrm{GeV}^{-2})$, while later measurements would yield the current value of $1.16 \times 10^{-5}\,\mathrm{GeV}^{-2}$. "Weak" is a bit of a misnomer. When a unit-less coupling constant is derived for these interactions, we see that $\alpha_W \sim \frac{1}{30}$, which is in fact larger than the electromagnetic coupling constant of $\alpha_{EM} = \frac{1}{137}$. The smallness of G_F is instead caused by the high mass of the W boson, the mediator of these interactions, with a mass of $80.4\,\mathrm{GeV}/c^2$.

Conservation laws are integral to understanding physics. Many of our favorite conservation laws can be derived using Noether's theorem, which states that any differentiable (continuous) symmetry has an associated conserved quantity. For example, physical processes unfold in the same manner regardless of whether they occur today or two weeks from now; this time-translation symmetry gives rise to conservation of energy. Likewise, a physical process is the same whether it occurs right here or over there, which is space-translation symmetry that is responsible for the conservation of momentum.

Along with continuous symmetries, there are also certain interesting discrete symmetries as well. Three discrete operations are particularly notable: charge conjugation, C, exchanging particles for antiparticles; parity inversion, P, a sign-flip of the spatial coordinates (i.e. $x \to -x$); and time reversal, T, where $t \to -t$. The first successful quantum field theory was quantum electrodynamics (QED), which explained electromagnetic interactions, and therefore, obeyed Maxwell's equations. It is relatively easy to show that Maxwell's equations are invariant under each of the C, P, and T transformations, and thus, any combination thereof. The invariance of QED under individual C, P, and T transformations naturally led many physicists to believe that all physical quantum field theories, like the theory of weak interactions, would also be symmetric under each individual transformation as well.

In the 1950s the $\tau - \theta$ problem arose in weak decays. Here the τ meson decays via one mode and the θ decays via another. The problem being that both particles have all of the same characteristics and were identical except for their decay modes. Further, the decay modes are such that the θ has a parity of +1 while the τ has a parity of −1. Tsung-Dao Lee and Chen-Ning Yang proposed a solution in 1956, where the two particles are indeed the same K^+ meson which is able to decay via both modes as long as the weak interaction did not obey parity conservation [23]. Before publication, they shared their

solution with Chien-Shiung "Madame" Wu, who promptly designed an experiment to test their hypothesis. Madame Wu's famous 1957 experiment showed that parity inversion symmetry was not only violated in β decays, but that it was violated as much as it possibly could be [24].

Skipping forward by about a decade, both Steven Weinberg and Abdus Salam, nearly simultaneously, drawing on the parallels of QED and weak interaction theory, realized that it was possible to combine the two into a new electroweak theory [25, 26]. By coupling four massless fields to a scalar (spin-0) doublet then utilizing invariance properties, they could eliminate one of the doublet states and generate four new fields. One field is massless and can neatly be identified with the photon from QED, while the other three are all massive fields. Two of the new massive fields are electrically charged and were identified as the mediators of all of the weak interactions that were known up to that point. These interactions all involved either charged particles being created or destroyed; thus the mediator would also need to have an electric charge to transfer it through the interactions. The third massive field had no electric charge and was one of the main predictions of this new formulation. The new interactions mediated by this field were dubbed as weak neutral current (NC) interactions as opposed to the charged current (CC) interactions that were already observed.

NC weak interactions are naturally difficult to observe as they require there to be no exchange of electric charge to occur. Similar interactions occur quite frequently via electromagnetic interactions because the mediating photon transfers no electric charge. However, because the NC mediator has a large mass and the photon is massless, the majority of these interactions proceed via a photon mediator. The lowest background method for observing NC interactions would be to utilize a particle which cannot interact electromagnetically. A natural candidate was found in the neutrino, and in 1973 the collaborators from the Gargamelle experiment at CERN found the first evidence [27]. Their observations relied on an incoming beam of neutrinos interacting on a target without the presence of an outgoing charged lepton. This was a huge success and validation of a novel prediction from the new electro-weak theory.

The final pieces of the neutrino puzzle were falling into place alongside the development of the electro-weak theory. As we've seen, neutrinos are created in conjunction with a charged lepton. In β decay they accompany electrons, while in other interactions, like the decay of a charged pion, they are released along with a muon. It quickly became obvious that neutrinos and the charged leptons were linked, and in 1962 Danby *et al.* [28] discovered that muon-related neutrinos were distinct from electron-related ones. They developed an

experiment at the Brookhaven accelerator in which they generated a beam consisting of charged pions which would decay into muons and neutrinos. By shielding the muons, they were able to isolate a beam consisting primarily of neutrinos. They then observed the results of these neutrinos interacting via the CC mode, and found that in all cases they only produced muons and never electrons. The few electron-like interactions they did observe could be tracked down to backgrounds inherent to the beam. Further, if any neutrino could produce either muons or electrons, then they would have at the very least observed equal numbers of muon and electron events which they did not see. This was the first evidence that neutrinos had an intrinsic "flavor" associated with the charged leptons.

Over the last century, the particle zoo was going through periods of expansion and contraction. Many new particles were being observed, and theoretical developments were then able to classify and explain the variety, thus collapsing the zoo down to a few constituent pieces. Muons and electrons along with their accompanying neutrinos had been observed, and at the beginning of the 1970s physicists began to search for a new heavy lepton. Antonio Zichichi and his collaborators undertook one of the first searches using the accelerator at CERN [29]. The energy of their accelerator turned out to be too low, though, they were able to exclude a large amount of mass space for a new charged lepton. Inspired by Yung-Su Tsai's 1971 paper [30], Martin Perl and his collaborators found the first evidence of a new lepton having a mass between $1.6 - 2.0$ GeV [31]. This lepton would be dubbed the τ since it was the third charged lepton and like the electron and muon, also has an associated neutrino.

1.5 Neutrinos: Lost and Found

As with everything in neutrino physics thus far, the resolution of the solar neutrino problem would have to wait for some time. It took another 30 years after Davis' discovery of the solar neutrino problem for the Super Kamiokande and Sudbury Neutrino Observatory experiments to reduce the signal-to-background ratio enough to be sure of the result. However, the theoretical framework began to develop as early as 1957 and was largely fleshed out in the decade after Davis' results.

Mirroring developments on the K^0 meson, Bruno Pontecorvo proposed, in 1957, that the observed neutrino and antineutrino could be linear superpositions of two other underlying states [32]. Pontecorvo's hypothesis showed that this would cause neutrinos to oscillate into antineutrinos over a time period based on the mass difference of the underlying neutrino states. In 1962, Ziro Maki, Masami Nakagawa, and Shoichi Sakata proposed a similar two-

neutrino theory which distinguished electron-type neutrinos from muon-type neutrinos, allowing each type to oscillate into the other [33]. The discovery of the uniqueness of neutrino flavors by Danby *et al.* ruled out Pontecorvo's original idea, but still allowed for the proposal of Maki *et al*. With the publication of Davis' results in May of 1968, Pontecorvo, along with Vladimir Gribov, showed how the two-neutrino theory could explain the solar neutrino problem [34]. When the τ lepton was discovered, it was natural to expand the two-neutrino theory to three generations, with the three observable neutrino states (one coupled to each charged lepton) being a superposition of three underlying neutrino states. The existence of three and only three "light" neutrino states was further confirmed in 1989 by the ALEPH and OPAL detectors at the Large ElectronPositron (LEP) collider at CERN by measuring the width of the neutral weak Z boson [35, 36].

The mixed neutrino hypotheses put forth separately by Pontecorvo and Maki, Nakagawa, and Sakata require neutrinos to have some mass even if it is quite small. This conflicted with some of the theoretical developments discussed in the last section. Because weak interactions were shown to maximally violate parity conservation, the only neutrinos that have been observed were left-handed neutrinos (i.e. their spin is in the opposite direction to their movement) or right-handed antineutrinos. What's more, with neutrinos only interacting via the weak force, it is impossible to directly observe right-handed neutrinos or left-handed antineutrinos. Maximal parity violation is simple enough to incorporate into the weak interaction model but does necessitate removing right-handed neutrinos from the picture. Additionally, the electro-weak unification discussed above also provides a method for particles to gain mass through the Higgs mechanism. To do so, both left and right handed particle need to be included. Therefore, massive neutrinos, even ones with very low mass, did not quite fit into the picture that was unfolding.

In 1977, Lincoln Wolfenstein provided a path through which neutrinos could still oscillate even if they had zero mass [37]. He showed that oscillations could be induced (or modified if neutrinos have mass) as neutrinos pass through matter as long as the index of refraction is different for the neutrino states. In the massless case, depending on the particular characteristics of NC interactions, he showed that up to 40% of the electron neutrinos coming from the sun could be transformed to another flavor just due to the matter effects inside of the sun. If neutrinos indeed have mass and can oscillate in the vacuum from one flavor to another, then the probability of neutrino oscillation would be modified owing to the CC scattering of electron neutrinos on electrons.

It is important to note that the neutrinos created via solar processes are all electron-type

neutrinos, and their energies are a maximum of ~1 MeV with most neutrinos being at a lower energy. This means that if solar neutrinos do oscillate from electron-type to muon or tau, then it would be impossible to see them because they must undergo CC interactions and have enough energy to produce the associated charged lepton. At low energies muons and taus cannot be created, thus solar neutrino oscillations would only manifest as a reduction in the CC electron neutrino rate.

Armed with the framework of neutrino oscillations and matter effects, experimentalists went to work. In 1998, the Super Kamiokande experiment published the first results showing that neutrinos created by cosmic rays interacting in the Earth's atmosphere oscillated from muon-type to tau-type [38]. Their results were consistent with the oscillation hypothesis at the 90% level, and future measurements would confirm this discovery. Since Super-K was observing atmospheric neutrinos, this observation did not directly explain the solar neutrino problem. However, only 4 years later, the Sudbury Neutrino Observatory (SNO) published direct evidence of neutrino oscillations occurring with solar neutrinos [39]. Armed with 30 years of measurements and interaction models, along with a greatly reduced background, SNO showed that the non-electron-type neutrinos made up ~66% of the solar neutrino flux at the 5.3σ level, and thus, resolved the solar neutrino problem.

1.6 What Neutrino Oscillations Tell Us

As discussed above, the Standard Model of Particle Physics is currently formulated with neutrinos being exactly massless; however, neutrino oscillations necessitate them having some mass. Neutrinos are still unique, in that their masses are so much smaller than all of the other fundamental fermions. From Figure 1.3 we can see that the largest possible neutrino mass is seven orders of magnitude lower than the next lightest fermion. It is possible to extend the Standard Model in such a way that neutrinos obtain their masses in the same manner as the other fermions, but, this fails to explain the large gulf between neutrino masses and the other fermions.

Neutrino oscillations in the 3-flavor paradigm are "controlled" by six parameters, see Section 2.3 for the details. Two of these parameters are directly related to the neutrino masses; three are so-called "mixing angles", which determine how the underlying neutrino states combine to create the observed neutrinos; and the final parameter allows for the violation of the combined CP symmetry. Recall that C transformations turn particles into antiparticles and P is spatial inversion.

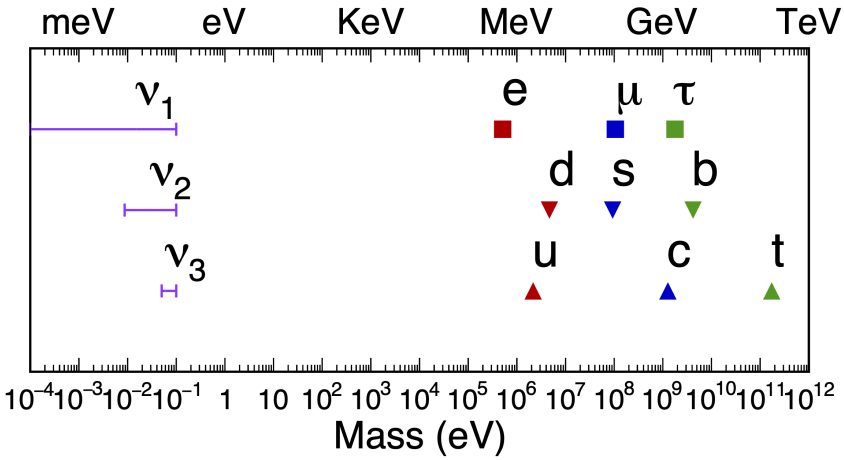


Figure 1.3: Masses of the Standard Model fermions. The squares denote the charged leptons and the triangles are the quarks. Color represents the fermion generations. The three neutrino mass states are shown as purple lines denoting the possible mass ranges.

The CP violating parameter is one of the most interesting components of neutrino oscillations. The universe as we know it is made up primarily of matter with very little antimatter being present. However, matter and antimatter should have been created in the same quantities at the beginning of the universe. This means we need some mechanism that prefers matter over antimatter. That process is often termed as leptogenesis for the electron-positron asymmetry and baryogenesis for the proton-antiproton (and neutron-antineutron) asymmetry. The CP violating parameter in the neutrino sector provides the sort of mechanism needed for leptogenesis and this can be transferred over to baryogenesis through "sphalerons" [40].

Determining the mass spectrum of neutrinos is clearly interesting to understand the nature of neutrino mass generation. There are many potential models to generate neutrino masses (some of which are discussed in brief in the next chapter) and accurately measuring the two oscillation parameters directly related to the neutrino masses will help to cut back the model forest. Of course, these models are all extensions to the Standard Model, and therefore, allow for new physics with many potential knock-on effects (none of which I will go into here).

The final piece of neutrino oscillations is the mixing angles. These have the obvious impact related directly to the nature of the neutrino states. Additionally, there is a potential symmetry that is somewhat apparent in the exact values of these angles. If this symmetry

remains after a precise measurement of the mixing angles, it could have implications beyond just the makeup of the neutrino states. These effects include: models of neutrino mass generation, GUT-scale physics, as well as baryogenesis [41].

2. A Review of Neutrino Physics

As discussed in the previous chapter, the theoretical development of neutrino physics has taken nearly 100 years and there are still more open questions. Intertwined with the growth of neutrino physics was the creation of the Standard Model, including electroweak unification. This unification provides a nice method for generating particle masses, though neutrinos, as we currently understand them, notably do not benefit from this mass generation. In the following sections I will give a quick review of the Standard Model as it relates to neutrinos with a review of the Higgs mechanism and potential methods of extending the Standard Model to produce neutrino masses. I will then discuss the various neutrino interactions of particular interest to the NOvA experiment. Finally, I will present neutrino oscillation physics and give a review of the current experimental results measuring the oscillation parameters.

2.1 The Standard Model

The Standard Model of particle physics is the most complete theoretical description we have regarding three of the four fundamental forces. It provides astonishingly accurate predictions for the weak and electromagnetic forces and also provides a framework for the strong force[1]. Since I am mainly concerned with neutrinos, I will not go into much further detail on the strong interactions except to mention their effects in the final state interactions from neutrino-nucleus scattering.

In general, the Lagrangian density, $\mathcal{L}$, of a "physically interesting" quantum field theory (QFT) can be built up from the requirements that it obeys some gauge symmetry (e.g. $U(1)$, $SU(2)$, $SU(3)$ or combinations thereof) and is renormalizable. Starting with a fermion field, ψ, and enforcing local phase invariance leads to a new vector field, A_μ, which connects ψ at nearby points and is necessary in order to obtain derivative (kinetic energy) terms. From the symmetry transformation and the vector field, a covariant derivative, D_μ, can be defined which can be included in the Lagrangian density while maintaining gauge

[1]Quantum chromodynamics (QCD) which describes the strong force is non-perturbative at low energies. Therefore simplifying assumptions cannot be made to directly calculate some values. This non-perturbative nature is due to the fact that at lower energy transfer the QCD coupling constant grows larger due to "color confinement" [42]

invariance. Finally, one can add a potential term, V, and kinetic terms for the field A_μ as well based on its derivatives to arrive at a final Lagrangian density:

$$\mathcal{L} = \bar{\psi}(i\not{D})\psi + \mathcal{L}_{KE}(A) + V + \text{h.c.} \tag{2.1}$$

It is also tempting to include mass terms, $m\bar{\psi}\psi$, for the particle in question, since we know that particles can have mass. However, depending on the symmetry group which we are trying to describe and features of the interaction, these mass terms may not be gauge invariant. If we start with a scalar field, ϕ, then $\bar{\psi}(i\not{D})\psi - > |D_\mu\phi|^2$. This model building procedure is described quite well in chapter 15 of Peskin and Schroeder [43].

Quantum electrodynamics (QED) can be well described by a field theory with $U(1)_{EM}$ symmetry, where the conserved quantity is identified as the electric charge and the single generator couples to the photon. Originally, Fermi's β theory represented weak interactions as contact processes, that is, without some mediating particle analogous to the photon. However, the "weakness" of these interactions could instead be understood as being mediated by very massive bosons. This would result in the short-range nature that allowed the contact interaction model to be so accurate. Recasting Fermi's theory as a non-contact force highlights the similarities between electromagnetism and the weak force that Fermi originally drew on to create his β theory.

Recall that there are two forms of β decay producing either electrons or positrons. This means that there must be at least two mediating bosons of opposite electric charge in order to transfer the charge of the decaying nucleon to the outgoing electron or positron, as shown in Figure 2.1. With at least two mediating bosons, the weak force must be formulated from a symmetry group containing at least two generators. This means that $U(1)$ cannot work and instead we look to the next group up, $SU(2)$, which has three generators. In 1960, Sheldon Glashow, building on previous work by Julian Schwinger, attempted to unify electromagnetism and the weak interaction and formulated a Lagrangian that was only partially symmetric [44]. Then in 1967 and 1968 respectively, Steven Weinberg [25] and Abdus Salam [26] developed a unified theory drawing on developments by Peter Higgs; François Englert and Robert Brout; and Gerald Guralnik, Carl Hagen, and Tom Kibble.

Starting off with an $SU(2) \times U(1)$ symmetric theory containing four massless bosons, then introducing a new "Higgs" field, the original bosons turn into two charged massive bosons, one neutral massive boson, and one neutral massless boson. The charged bosons can be identified as the two charged weak bosons we needed; the neutral massless boson

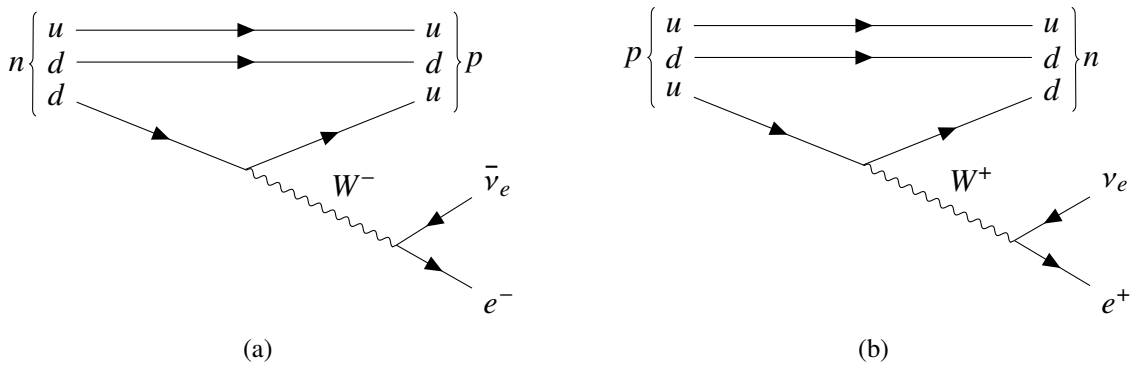


Figure 2.1: Feynman diagrams for (a) β^- and (b) β^+ decays showing the need for two oppositely charged mediating bosons.

is the familiar photon, and the neutral massive boson is a new weak mediator which hadn't been observed before. The process of going from a symmetric theory with massless bosons to a new model with massive bosons is termed "spontaneous symmetry breaking". The exact details are worked out in Appendix B.

Based on Madam Wu's experiment showing that the charged weak interaction maximally violates parity conservation, we know that it only acts on left-handed fermions. Therefore, when we couple this theory to fermions, we can use the left-handed projector, $(1-\gamma^5)\psi = \psi_L$, to select out only left-handed particles. Since neutrinos only interact via the weak force, they can only be observed as left-handed and so we can arrange the fermions into left-handed doublets and right-handed singlets with the right-handed neutrino singlet not appearing in the theory:

$$\chi_L = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \quad \psi_R = e_R^- \tag{2.2}$$

while for quarks we have:

$$\chi_L = \begin{pmatrix} u \\ d \end{pmatrix}, \quad \psi_R = u_R \text{ or } d_R \tag{2.3}$$

The fallout of this is that typical mass terms, $m(\bar{e}_L e_R + \bar{e}_R e_L)$, cannot be formulated because the left-handed field belongs to the doublet representation while the right-handed field is in the singlet. However, by coupling the fermions to the Higgs field (which is a doublet), we can again use spontaneous symmetry breaking to produce gauge invariant mass terms. Since the right-handed neutrino singlet is not observed in this theory, this method of mass generation leaves the neutrino massless.

Alternative methods to generate neutrino masses must, therefore, do away with gauge invariance of the underlying model, introduce new particles to couple to the left-handed neutrino, or introduce some new physics at some high energy scale. With gauge invariance being so central to our current understanding of particle physics, it is obviously difficult to abandon. Introducing new particles, on the other hand, has been done repeatedly in the past and, thus, is a natural extension. This is done in the neutrino sector by postulating some number of new "sterile" right-handed neutrinos which do not interact via any of the three forces encoded in the Standard Model. With a right-handed neutrino singlet, we can then produce a mass term via spontaneous symmetry breaking in the same manner at the rest of the fermions.

With the inclusion of sterile neutrinos, we may also formulate another invariant mass term by taking these neutrinos to be "Majorana" particles[2]. A model with only the "normal" mass term is considered unsatisfactory by some because it does not explain the smallness of neutrino masses as compared to the other leptons, since they all obtain mass via the same process. The addition of the Majorana mass term, on the other hand, produces mass eigenstates that are proportional to the inverse of the mass of the sterile states. Thus, if the sterile states are heavy, then this method, also called the "see-saw mechanism", naturally produces very light neutrino masses. More detail about these models can be found in the PDG [45] and a nice article by Raby and Slansky [46], but no more will be said here since NOvA cannot access the Majorana versus Dirac nature of neutrinos.

2.2 Neutrino Interactions

As shown in Appendix B, neutrinos interact via the weak currents as:

$$J_W^{\mu+} = \frac{1}{\sqrt{2}}\bar{\nu}_{iL}(U^\dagger)_{il}\gamma^\mu e_{lL} = \frac{1}{\sqrt{2}}\bar{\nu}_i(U^\dagger)_{il}\gamma^\mu\frac{1}{2}(1-\gamma^5)e_l; \tag{2.4}$$

$$J_W^{\mu-} = \frac{1}{\sqrt{2}}\bar{e}_{lL}\gamma^\mu U_{li}\nu_{iL} = \frac{1}{\sqrt{2}}\bar{e}_l\gamma^\mu\frac{1}{2}(1-\gamma^5)U_{li}\nu_i; \tag{2.5}$$

$$J_Z^\mu = \frac{1}{2\cos\theta_w}\bar{\nu}_{iL}(U^\dagger)_{il}\gamma^\mu U_{li}\nu_{iL} = \frac{1}{2\cos\theta_w}\bar{\nu}_l\gamma^\mu\frac{1}{2}(1-\gamma^5)\nu_l; \tag{2.6}$$

[2]The wave function of a Majorana particle does not change under charge conjugation. That is to say for a Majorana particle the particle state is identical to its antiparticle. This is in contrast to Dirac particles which change their quantum numbers. Only electrically neutral particles can fulfil the Majorana criteria since they have no electric charge to change on conjugation.

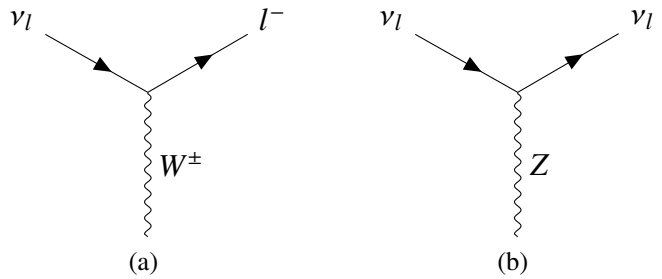


Figure 2.2: Feynman diagrams for neutrinos interacting via the weak (a) charged current and (b) neutral current.

where the index i runs over the neutrino mass states $(1, 2, 3)$, the index l runs over the lepton flavors (e, μ, τ), and the subscript L identifies the left-handed particles. These vertices are summarized by the Feynman diagrams in Figure 2.2. The weak interaction amplitude, $\mathcal{M}$, for some process is given by:

$$\mathcal{M} = gJ^\mu \frac{-i(g_{\mu\nu} - q_\mu q_\nu / M_{Z/W}^2)}{q^2 - M_{Z/W}^2} gJ^\nu \tag{2.7}$$

where g is the weak coupling constant, J^μ are the relevant currents listed above, $M_{Z/W}$ is the boson mass, and q is the four momentum transfer carried by the $W^\pm$ or Z boson. In the limit where q is much smaller than the mass of the bosons, which is accurate for NOvA, this becomes:

$$\mathcal{M} = \frac{-ig^2}{M_{Z/W}^2} J^\mu J_\mu \tag{2.8}$$

In NOvA the "important" processes are charged current (CC) since these will produce a charged lepton which then identifies the incoming neutrino flavor. Neutral current (NC) interactions are completely background processes for the investigations performed here. There are, therefore, two main CC interaction types to consider, ν on electron or ν on nucleon. To obtain some understanding of these processes, we can analyze the simple elastic and quasi-elastic (QE) interactions shown in Figure 2.3. The differential cross section of these types of interactions can be written as:

$$\left.\frac{d\sigma}{d\Omega}\right|_{\text{CM}} = \frac{1}{64\pi s} \frac{\boldsymbol{p}_{2\text{CM}}}{\boldsymbol{p}_{1\text{CM}}} |\mathcal{M}|^2 \tag{2.9}$$

where we are working in the center of mass reference frame and $\boldsymbol{p}_{1,2\text{CM}}$ are the magnitudes

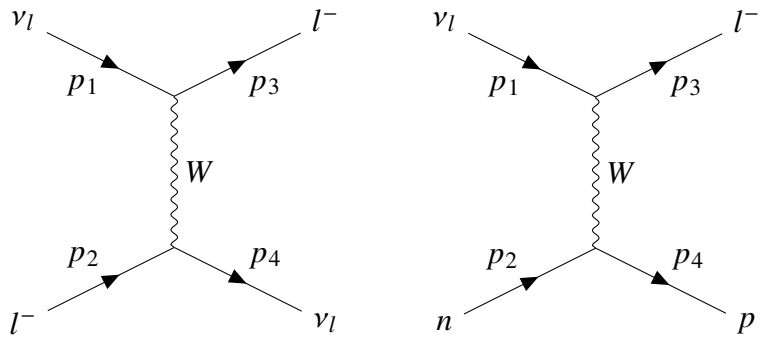


Figure 2.3: charged current (a) elastic and (b) quasi-elastic scattering off of an electron or neutron respectively. For antineutrinos, the neutron and proton of figure (b) will be interchanged.

of three momenta of the neutrino and outgoing lepton. Additionally, s is one of the of the Mandelstam variables which are convenient Lorentz-invariant quantities describing interaction kinematics

$$s = (p_1 + p_2)^2 = (p_3 + p_4)^2 \tag{2.10}$$

Here we are interested in a lab frame where neutrinos interact with particles that are taken to be at rest, thus we have

$$\boldsymbol{p}_{1\mathrm{CM}} = \frac{E_{1\mathrm{lab}}\, m_2}{\sqrt{s}} \tag{2.11}$$

It is relatively simple, if a bit tedious, to show that at tree level for high neutrino energy Eq. 2.9 becomes:

$$\frac{d\sigma}{d\Omega} \sim \left(\frac{g^2}{8\pi M_W^2}\right)^2 E_{1\mathrm{lab}} m_2 \tag{2.12}$$

From this result we can immediately see that, due to the m_2 dependence, the cross section for scattering off a nucleon is much greater than scattering off an electron. Therefore, we will only consider neutrino-nucleus interactions from here on.

2.2.1 Interaction modes

There are four primary interaction modes that occur in NOvA's energy range:

- Quasi-elastic (QE)
- Resonance (RES)

- Deep inelastic scattering (DIS)
- Multi-nucleon knockout, also referred to as two-particle-two-hole (2p2h), which is dominated by the meson exchange current (MEC)

The first three of these are shown in Figure 2.4 and cross sections as a function of neutrino energy are shown in Figure 2.5.

QE interactions are the simplest of the four and were discussed above. These are the inverse of the β decays mentioned earlier, and here a neutron (proton) absorbs the incoming neutrino (antineutrino) and produces an outgoing charged lepton and a proton (neutron). For free nucleons, this interaction mode is straightforward to calculate, though when it occurs inside of a nucleus a proper treatment becomes laden with form factors [47].

Starting at a neutrino energy of around 0.3 GeV, the target nucleon can be excited into a Δ resonance state, which then decays to produce more hadrons. At lower neutrino energies, the most common resonance is $\Delta(1232)$. Specifically, the possible states are $\Delta^{+}(\Delta^{-})$ or $\Delta^{++}(\Delta^{0})$ if a (anti)neutrino has a CC interaction with a neutron or proton, respectively. The lowest energy resonance will most commonly decay into a nucleon and a single pion (BR 99.4%) while at higher energies the resonance states will decay into multiple pions or possibly kaons or η particles, though these branching ratios are much smaller than the pion mode. All Δ resonances can also decay into a photon, though these modes have the lowest branching ratios. Resonance interactions are quite important for NOvA since it is possible for NC interactions to mimic the signature of a CC event. If the NC interaction results in

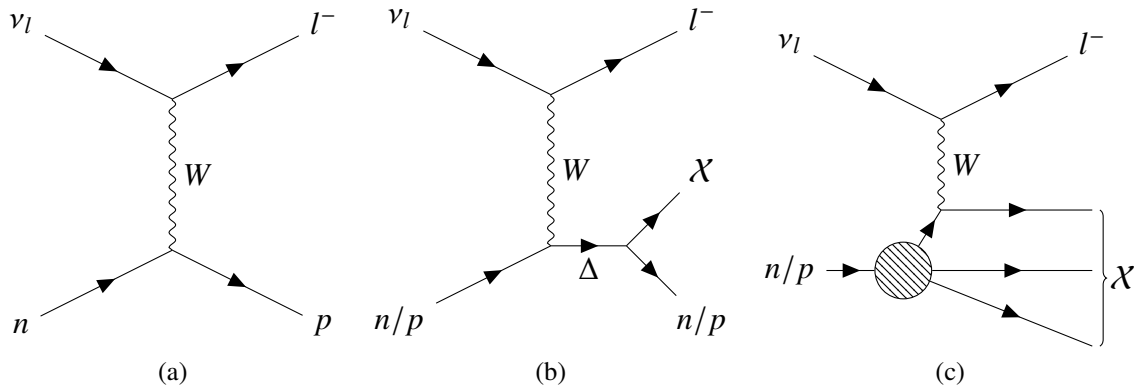
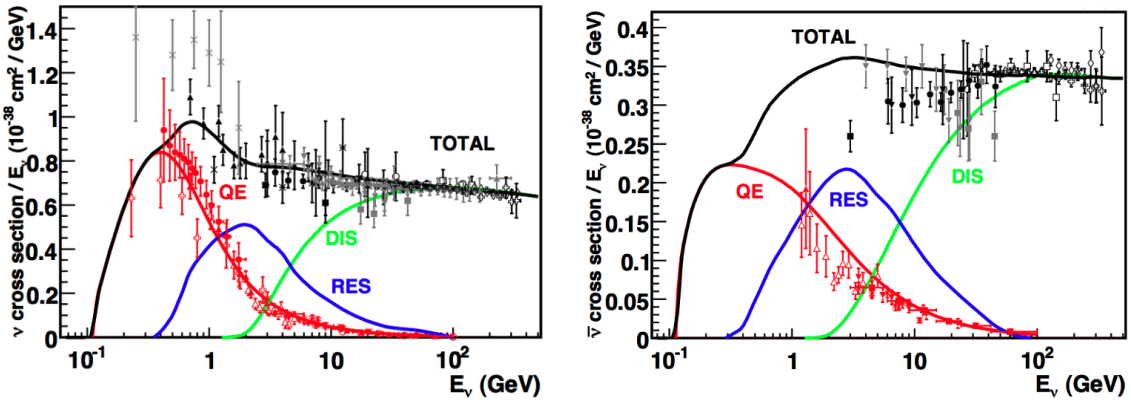


Figure 2.4: Three important neutrino-nucleus interaction modes. (a) Quasi-elastic scattering or QE, (b) resonance production or RES, and (c) deep inelastic scattering or DIS. The outgoing X particles are undefined hadrons that will depend on final state interactions and, for RES, on which resonance state was exited.



(a) (b)

Figure 2.5: Cross sections of (a) neutrino and (b) antineutrino interactions with nucleons as a function of neutrino energy. The solid lines show predictions from models for QE, RES, and DIS interactions while the data points are taken from various experiments [48]. The multi-nucleon knockout interactions are missing from these plots but would peak in the "dip" region between the QE and RES peaks.

a π^0 being produced, this will decay into two photons which produce an electromagnetic shower similar to an electron, while if a charged pion is produced, the topology is similar to a muon.

At energies of about 2 GeV or more, the neutrino will scatter off of a constituent quark inside of the nucleon. This deep inelastic scattering can knock out the quark at low energies or "blow up" the nucleon entirely at higher energies. In either case, the nucleon remnants will hadronize with a hadronic jet forming at higher energies.

The final interaction mode of interest for NOvA is the case of multi-nucleon knockout. Most frequently, this occurs when the W boson is absorbed by two nucleons which are then knocked out of the nucleus, leaving behind two holes; thus this mode also termed as two-particle-two-hole or 2p2h. Further, the 2p2h effect is dominated by the meson exchange current (MEC) where the two nucleons are interacting with each other via a meson (typically taken as a π). Multi-nucleon knockout, 2p2h, and MEC are all used nearly synonymously in the literature surrounding neutrino interactions. As you may guess from the numerous accepted names, this interaction mode is particularly poorly understood with a peak in the cross section occurring between QE and RES.

Since all of these interaction modes occur inside of the nucleus, at least when the target material is anything but hydrogen, the outgoing particles can re-interact and produce

different particles than those initially predicted. These final state interactions, or FSI, rely heavily on nuclear modeling, and while one model choice may work well for some nuclei or interaction energies it may not be suitable for others. FSIs are less of an issue for an oscillation analysis since the neutrino flavor is tagged by the outgoing lepton. FSIs can, however, change the hadronic energy deposition, and therefore, alter the estimation of the incoming neutrino's energy.

2.3 Oscillations

As discussed in Appendix B, if we allow for neutrinos to have mass then the mass eigenstates can be different from the weak interaction states. Therefore, we have unitary matrix which mixes the mass states, $|\nu_i\rangle$, to for the weak states, $|\nu_\alpha\rangle$:

$$|\nu_\alpha\rangle = \sum_{i=1}^{3} U^*_{\alpha i} |\nu_i\rangle \tag{2.13}$$

where ν_α are the flavor states (ν_e, ν_μ, ν_τ), ν_i are the mass states (ν_1, ν_2, ν_3)[3], and U is a unitary mixing matrix. This matrix is commonly referred to at the PMNS matrix, named after Pontecorvo, Maki, Nakagawa, and Sakata, four physicists who formulated the initial neutrino mixing theory [32, 33].

With the flavor states being superpositions of the mass states, as a neutrino propagates in time and space, the mass states will interfere with each other. The states will evolve as:

$$|\nu_\alpha(\boldsymbol{x}, t)\rangle = \sum_{i=1}^{3} U^*_{\alpha i} |\nu_i(\boldsymbol{x}, t)\rangle \tag{2.14}$$

thus the probability of a neutrino of flavor α to oscillate into flavor β over some time interval, t, and distance, L, is found from:

$$P(\nu_\alpha \to \nu_\beta; L, t) = \left| \sum_{i,j} U_{\alpha i} U^*_{\beta i} \left\langle \nu_j \middle| \nu_i(L, t) \right\rangle \right|^2 \tag{2.15}$$

In the usual oscillation derivation, three assumptions are typically made:

[3]There can be more than three light neutrino mass states in some models, but we will restrict ourselves to the three-neutrino paradigm since that is the minimum number required.

1. The mass states are taken to be plane waves or stationary states: $e^{-i\boldsymbol{p}_i\cdot\boldsymbol{x}}\left|\nu_i\right\rangle$ or $\left|\nu_i(\boldsymbol{x},t)\right\rangle = e^{-iE_it}\left|\nu_i\right\rangle$

2. Neutrino masses are small so the energies (or momenta) of the three mass eigenstates can be taken to be equal.

3. Neutrino masses are small, and therefore, we can replace the time propagation, t, with a distance L if we use stationary states.

With these in hand, we can plug in the evolving wave function for the mass states, take the energies and momenta of all states to be equal, and Taylor expand about a mass of zero ($E_i = E = \sqrt{p^2 + m_i^2} \sim p + \frac{m_i^2}{2E}$). Thus we easily find:

$$\begin{aligned} P(\nu_\alpha \to \nu_\beta; L) &= \left| \sum_{i,j} U_{\alpha i} U^*_{\beta j} e^{-i\frac{m_i^2 L}{2E}} \left\langle \nu_j \middle| \nu_i \right\rangle \right|^2 \\ &= \sum_{i,j} U_{\alpha i} U^*_{\beta i} U^*_{\alpha j} U_{\beta j} e^{-i\frac{\Delta m_{ij}^2 L}{2E}} \end{aligned} \tag{2.16}$$

where $\Delta m_{ij}^2 = m_i^2 - m_j^2$ are called the mass splitting terms. Making the above assumptions is equivalent to stating that the Hamiltonian in the mass basis is: $H_m = \frac{1}{2E}\mathrm{diag}(m_1^2, m_2^2, m_3^2)$.

We can exploit the unitarity of the matrix U to arrive at the following form:

$$\begin{aligned} P(\nu_\alpha \to \nu_\beta; L) = \delta_{\alpha\beta} &- 4\sum_{i<j} \mathrm{Re}[U_{\alpha i} U^*_{\beta i} U^*_{\alpha j} U_{\beta j}] \sin^2\left(\frac{\Delta m_{ji}^2 L}{4E}\right) \\ &+ 2\sum_{i<j} \mathrm{Im}[U_{\alpha i} U^*_{\beta i} U^*_{\alpha j} U_{\beta j}] \sin\left(2\frac{\Delta m_{ji}^2 L}{4E}\right) \end{aligned} \tag{2.17}$$

while for antineutrinos $U \to U^*$, therefore, the imaginary term violates CP conservation since the sign flips for antineutrinos.

The assumptions above raise a few problems:

1. Plane waves do not propagate in space while stationary states do not propagate in time, and we observe oscillations occurring over both time and space.

2. A least two of the neutrino states must have mass for oscillations to occur. This means that if the energies of all three states were equal in one reference frame, you could

boost to a different frame where this no longer applies. Therefore, the oscillation probability would be dependent on the frame.

For these reasons, a more correct derivation[4] would take the mass states as wave packets. However, in a wonderful twist, the wave packet formulation also results in Eq. 2.16 (as long as two reasonable conditions are met) so I will reserve that derivation for Appendix C.

The PMNS matrix, in the 3-flavor paradigm, is a unitary 3×3 matrix, and is therefore parameterized by four independent values when only considering Dirac neutrinos[5]. The most common parameterization uses three mixing angles $(\theta_{12}, \theta_{13}, \theta_{23})$ along with an exponential phase, δ_{CP}, that violates CP symmetry:

$$U_{PMNS} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \times \begin{pmatrix} c_{13} & 0 & s_{13}\, e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}\, e^{i\delta_{CP}} & c_{13} & 0 \end{pmatrix} \times \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.18}$$

or, fully expanded:

$$\begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{\text{CP}}} \\ -s_{12}c_{23} - c_{12}s_{13}s_{23}e^{i\delta_{\text{CP}}} & c_{12}c_{23} - s_{12}s_{13}s_{23}e^{i\delta_{\text{CP}}} & c_{13}s_{23} \\ s_{12}s_{23} - c_{12}s_{13}c_{23}e^{i\delta_{\text{CP}}} & -c_{12}s_{23} - s_{12}s_{13}c_{23}e^{i\delta_{\text{CP}}} & c_{13}c_{23} \end{pmatrix}$$

where $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$.

2.3.1 Matter effects

As neutrinos propagate through matter, the flavor states will interact with the surrounding material and scatter [37, 50]. This scattering can occur either incoherently, where the states of the initial particles change in the process, or coherently, where the states remain the same. Incoherent scattering would cause the underlying mass states to decohere, and so are not of interest here. Coherent forward scattering can either occur through charged or neutral current processes. The amplitude of NC scattering on electrons is the same for all three neutrino flavors. However, since most matter contains electrons and not muons or taus, only the ν_e states will undergo CC elastic scattering as shown in Figure 2.3a. Similarly,

[4]The most consistent derivation would dive fully into quantum field theory, which I won't attempt; but a nice review can be found at Ref. [49].

[5]For Majorana neutrinos, two additional parameters are needed, bringing the total to 6 independent parameters.

all three flavor states will undergo neutrino-nucleus scattering with the same amplitude, so these interactions will not affect neutrino oscillations. The coherent forward scattered wave packets will interfere with the unscattered waves resulting in matter mass eigenstates that differ from those in a vacuum. The impact of matter on neutrino oscillations was initially formulated by Lincoln Wolfenstein and later expanded by Stanislav Mikheyev and Alexei Smirnov. For this reason, neutrino matter effects are also called the MSW effect.

The MSW effect can be quantified by adding a new potential term to the neutrino propagation Hamiltonian. The only term in this potential will be provided by the CC coherent forward scattering of ν_e, since this is the only scattering which ν_μ and ν_τ cannot partake in. This process produces a term depending on the electron number density of the matter, N_e, and on the weak coupling constant, G_F, resulting in a potential of:

$$V = \text{diag}(\pm\sqrt{2}G_F N_e(x), 0, 0) \tag{2.19}$$

where the plus sign is for neutrinos and the negative is for antineutrinos and the electron density of a material can vary as a function of position. We now have the effective Hamiltonian in matter

$$H_M = H_m + U^\dagger V U \tag{2.20}$$

with the subscript M denoting that this is the Hamiltonian in matter, U is the standard PMNS matrix, and $H_m = \frac{1}{2E}\text{diag}(m_1^2, m_2^2, m_3^2)$ is the vacuum Hamiltonian, as described above. By diagonalizing this H_M, we can find the effective mass eigenstates and mixing angles in matter. Doing so is quite messy with three neutrino states to worry about, though, using only two neutrinos is straightforward and instructive.

In the case of only two neutrino states, the mixing matrix is parameterized by a single mixing angle as:

$$U = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \tag{2.21}$$

In matter, the neutrino flavor states are related to the new effective mass eigenstates by a different mixing matrix, $\tilde{U}$:

$$|\nu_\alpha\rangle = \sum_{i=1} \tilde{U}^*_{\alpha i} |\nu_i\rangle_M \tag{2.22}$$

with its own matter mixing angle, θ_M. Now we want to find the effective masses and mixing angle in terms of the vacuum masses and angle. The masses are found by diagonalizing the effective Hamiltonian. Here we will take the potential to be associated with only one flavor

state, $V = \text{diag}(\pm\sqrt{2}G_F N_\alpha(x), 0)$, which makes the effective Hamiltonian:

$$H_M = \begin{pmatrix} \frac{1}{2E}m_1^2 \pm \sqrt{2}G_F N_\alpha \cos^2\theta & \pm\sqrt{2}G_F N_\alpha \cos\theta \sin\theta \\ \pm\sqrt{2}G_F N_\alpha \cos\theta \sin\theta & \frac{1}{2E}m_1^2 \pm \sqrt{2}G_F N_\alpha \sin^2\theta \end{pmatrix} \tag{2.23}$$

the eigenvalues of which give the effective masses:

$$\begin{aligned} m_M^2 = &\frac{m_1^2 + m_2^2}{2} + \sqrt{2}EG_F N_\alpha \\ &\mp \frac{1}{2}\sqrt{(\Delta m^2 \cos 2\theta - 2\sqrt{2}EG_F N_\alpha)^2 + (\Delta m^2 \sin 2\theta)^2} \end{aligned} \tag{2.24}$$

where I've suppressed the $\pm$ associated with the change from neutrinos to antineutrinos so as to avoid confusion with the $\mp$ for the two effective mass states. From these masses we get the effective mass splitting in matter as:

$$\Delta m_M^2 = \sqrt{(\Delta m^2 \cos 2\theta \mp 2\sqrt{2}EG_F N_\alpha)^2 + (\Delta m^2 \sin 2\theta)^2} \tag{2.25}$$

with the $\pm$ for neutrino $\leftrightarrow$ antineutrino restored. We can also determine the effective mixing angle by finding the rotations that diagonalize the effective Hamiltonian. Doing so leads to:

$$\tan 2\theta_M = \frac{\Delta m^2 \sin 2\theta}{\Delta m^2 \cos 2\theta \mp 2\sqrt{2}EG_F N_\alpha} \tag{2.26}$$

For two neutrino states, the vacuum appearance probability is given by:

$$P(\nu_\alpha \to \nu_\beta) = \sin^2 2\theta \sin^2\left(\frac{\Delta m^2 L}{4E}\right) \tag{2.27}$$

Oscillations in matter give the same result but with $\theta \to \theta_M$ and $\Delta m^2 \to \Delta m_M^2$. Using the above results for the mass splitting and angle in matter, we see that $\Delta m_M^2 \sin^2 2\theta_M$ does not rely on the matter density and is invariant. Therefore, we can restate the two flavor oscillation probability in matter as:

$$P_M(\nu_\alpha \to \nu_\beta) = \frac{\Delta m^2 \sin^2 2\theta}{\Delta m_M^2} \sin^2\left(\frac{\Delta m_M^2 L}{4E}\right) \tag{2.28}$$

2.3.2 Types of Oscillation Experiments

The characteristic oscillation length, $l_{osc} = 2E/\Delta m^2_{ij}$, defines the type of oscillation experiment that can observe a particular mass splitting. The sensitive mass splitting range for various types of neutrino experiments are shown in Table 2.1.

Experiment		Channel(s)	L (m)	E (MeV)	$\lvert\Delta m^2\rvert$ (eV^2)
Solar		ν_e disappearance	10^{10}	1	10^{-10}
Atmospheric		ν_μ ($\bar{\nu}_\mu$) disappearance	$10^4 - 10^7$	$10^2 - 10^5$	$10^{-4} - 10^{-1}$
Reactor	Short	$\bar{\nu}_e$ disappearance	$10^2 - 10^3$	1	$10^{-3} - 10^{-2}$
	Long		$10^4 - 10^5$		$10^{-5} - 10^{-4}$
Accelerator	Short	ν_μ ($\bar{\nu}_\mu$) disappearance	10^2	$10^3 - 10^4$	$> 10^{-1}$
	Long		$10^5 - 10^6$		$10^{-3} - 10^{-2}$

Table 2.1: The sensitive mass splitting range based on typical lengths and neutrino energies of specific types of neutrino oscillation experiments in a vacuum. Short and long are indicative of the oscillation baselines used in reactor and accelerator experiments. Also shown are the particular channels that each experiment type can observe [45].

Solar and long baseline reactor experiments observe ν_e and $\bar{\nu}_e$ disappearance respectively, therefore, these experiments probe the θ_{12} and θ_{13} mixing angles along with small mass splittings. Owing to the MSW effect within the Sun, the product $\Delta m^2_{21} \cos 2\theta_{12}$ must be positive. We are then free to set either the sign of Δm^2_{21} or the octant of the angle θ_{12}. The standard convention is to take θ_{12} to be in the lower octant ($0 < \theta_{12} < \pi/4$). Doing so means that we also choose Δm^2_{21} to be positive; in other words, the ν_1 state is lighter than ν_2. Combining results from solar and reactor experiments we've also found that the mass splitting Δm^2_{21} is of the order 10×10^{-5} eV^2, $\sin^2\theta_{12} \sim 0.3$, and $\sin^2\theta_{13} \sim 2 \times 10^{-2}$.

Both long baseline accelerator and short baseline reactor experiments probe similar mass splitting ranges, which are two orders of magnitude higher than Δm^2_{21}. This introduces a hierarchical nature to the mass splittings: $\Delta m^2_{21} << |\Delta m^2_{31}| \sim |\Delta m^2_{32}|$. We have to include the absolute values in this expression because we cannot tell if ν_3 is the largest or smallest mass eigenstate. This means that there are two possible orderings for the neutrino mass spectrum:

1. Normal Ordering (NO): $m_1 < m_2 < m_3$, with $\Delta m^2_{32} > 0$.

2. Inverted Ordering (IO): $m_3 < m_1 < m_2$, with $\Delta m^2_{32} < 0$.

These are also commonly referred to as the mass hierarchies but "hierarchy" typically means that one value is much smaller than the others which may not necessarily be the case since all three neutrino masses could be nearly the same size and $O\left(\sqrt{\Delta m^2_{32}}\right)$. Therefore, I will used the therm "mass ordering" instead.

NOvA, being a long baseline neutrino experiment, observes ν_μ ($\bar{\nu}_\mu$) disappearance. These neutrinos will turn into both electron and τ flavor neutrinos. As discussed in the next chapter, NOvA views a neutrino beam that is peaked at about 2 GeV; thus, ν_τ CC events will not occur very frequently since the beam energy is only just above the τ mass. Additionally, if a ν_τ CC event does occur the τ will rapidly decay and the event signature will not be easily separable from backgrounds. Thus, in addition to ν_μ ($\bar{\nu}_\mu$) disappearance, NOvA also observes ν_e ($\bar{\nu}_e$) appearance.

The oscillation probability for ν_μ ($\bar{\nu}_\mu$) disappearance is a bit messy to calculate and isn't particularly informative in its full form. However, we can see that since the PMNS matrix is unitary, when both α and β are the same, the product $|U_{\alpha i}|^2|U_{\alpha j}|^2$ will be real and we can avoid working out the imaginary component of the oscillation probability. Using the fact that Δm^2_{21} is small we can ignore all $i, j = 1, 2$ contributions, and further, express the oscillation probability in terms of only one mass splitting. Since we are dealing with trigonometric terms, there are many ways you can combine things to simplify the expression, and consequently, you may see various different expressions in the various literature. In any case the ν_μ ($\bar{\nu}_\mu$) survival probability is

$$P(\nu_\mu\ (\bar{\nu}_\mu) \rightarrow \nu_\mu\ (\bar{\nu}_\mu)) \approx 1 - \left(\cos^2\theta_{13}\sin^2 2\theta_{23} + \sin^2 2\theta_{13}\sin^4\theta_{23}\right)\sin^2\Delta_{32} \tag{2.29}$$

where $\Delta_{ij} \equiv \frac{\Delta m^2_{ij} L}{4E}$. Owing to the smallness of $\sin^2\theta_{13}$, the second term is suppressed and $\cos^2\theta_{13} \simeq 1$, leading to:

$$P(\nu_\mu\ (\bar{\nu}_\mu) \rightarrow \nu_\mu\ (\bar{\nu}_\mu)) \approx 1 - \sin^2 2\theta_{23}\sin^2\Delta_{32} \tag{2.30}$$

and we see that to good approximation the ν_μ ($\bar{\nu}_\mu$) oscillation probability is independent of δ_{CP} and only relies on one mixing angle. From this expression it is clear that, as a function of neutrino energy for a set baseline, the amplitude of ν_μ disappearance is controlled by θ_{23} while $|\Delta m^2_{32}|$ determines where the oscillation dips occur, as can be seen in Figure 2.6.

Next, we focus on ν_e appearance following closely with Ref. [51]. Using Eq. 2.16

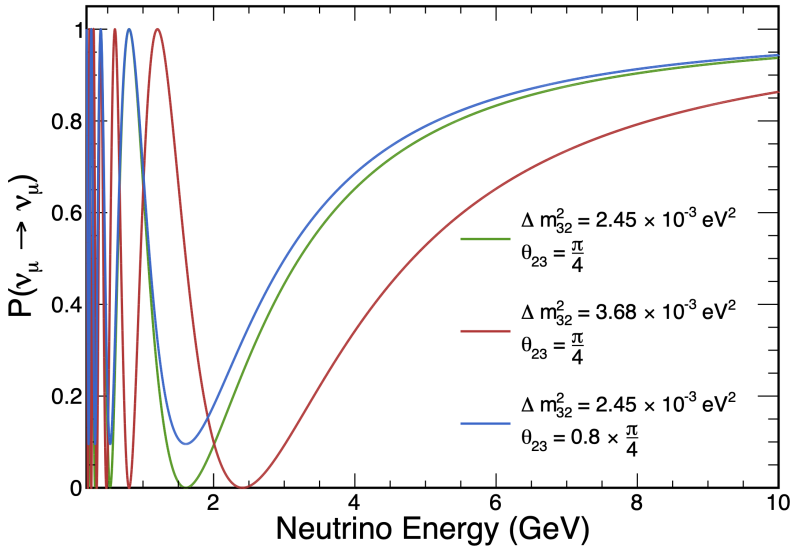


Figure 2.6: ν_μ appearance oscillation probability for three different sets of Δm^2_{32} and θ_{23} showing how Δm^2_{32} controls the position of the first oscillation dip and θ_{23} controls the amplitude.

directly we have:

$$P(\nu_\mu \to \nu_e) = \left| U_{\mu 1} U^*_{e1} e^{-i\frac{m_1^2 L}{2E}} + U_{\mu 2} U^*_{e2} e^{-i\frac{m_2^2 L}{2E}} + U_{\mu 3} U^*_{e3} e^{-i\frac{m_3^2 L}{2E}} \right|^2 \tag{2.31}$$

Then, taking advantage of the unitarity of the PMNS matrix[6] and with some rearrangement of the exponentials we get:

$$P(\nu_\mu \to \nu_e) = \left| 2U_{\mu 2} U^*_{e2} \sin \Delta_{21} + 2U_{\mu 3} U^*_{e3} \sin \Delta_{31} e^{-i\Delta_{32}} \right|^2 \tag{2.32}$$

Plugging in the matrix elements and using one of the double angle formulas[7] gives:

$$P(\nu_\mu \to \nu_e) = \Big| \cos\theta_{13} \cos\theta_{23} \sin 2\theta_{12} \sin\Delta_{21} - \frac{1}{2} \sin^2\theta_{12} \sin 2\theta_{23} \sin 2\theta_{13} \sin\Delta_{21} e^{i\delta_{\mathrm{CP}}} + \sin 2\theta_{13} \sin\theta_{23} \sin\Delta_{31} e^{-i(\Delta_{32}+\delta_{\mathrm{CP}})} \Big|^2 \tag{2.33}$$

[6] In particular we can use: $U_{\mu 1} U^*_{e1} + U_{\mu 2} U^*_{e2} + U_{\mu 3} U^*_{e3} = 0$.
[7] $\sin 2\theta = 2 \sin\theta \cos\theta$

The second term in this equation will be particularly small since it involves $\sin\theta_{13}$ and $\sin\Delta_{21}$. Further, owing to the smallness of Δ_{21}, we can use the small angle approximation to write $\sin\Delta_{21} \approx \Delta_{21}$, which gives us:

$$\begin{aligned} P(\nu_\mu \rightarrow \nu_e) &\approx \left|\sqrt{P_{sol}} + \sqrt{P_{atm}} e^{-i(\Delta_{32}+\delta_{\mathrm{CP}})}\right|^2 \\ &= P_{sol} + P_{atm} + \sqrt{P_{sol}}\sqrt{P_{atm}}\cos\left(\Delta_{32} + \delta_{\mathrm{CP}}\right) \end{aligned} \tag{2.34}$$

with $\delta_{\mathrm{CP}} \rightarrow -\delta_{\mathrm{CP}}$ for antineutrinos, where:

$$\sqrt{P_{sol}} \equiv \cos\theta_{13}\cos\theta_{23}\sin 2\theta_{12}\sin\Delta_{21} \approx \cos\theta_{13}\cos\theta_{23}\sin 2\theta_{12}\Delta_{21} \tag{2.35}$$

$$\sqrt{P_{atm}} \equiv \sin\theta_{23}\sin 2\theta_{13}\sin\Delta_{31} \tag{2.36}$$

These are subscripted with sol and atm because they are dominant components in solar and atmospheric neutrino oscillation experiments, respectively.

Eq. 2.34 is the ν_e appearance probability in a vacuum. Since Δm^2_{21} is much smaller than Δm^2_{31} and $\sin^2\theta_{13} << 1$, we can treat the three-neutrino picture as the product of two-neutrino cases. The two-neutrino cases are: e-μ with $\Delta m^2 \approx 0$, e-τ with $\theta \approx \pi$, and μ-τ which is unaffected by matter, corresponding to the Δm^2_{21}, Δm^2_{31}, and Δm^2_{32} mass splittings, respectively. Therefore, using Eq. 2.25 gives us the modified mass splittings in matter:

$$\begin{aligned} \Delta m^2_{21M} &\approx \mp 2\sqrt{2}EG_F N_e \\ \Delta m^2_{31M} &\approx \Delta m^2_{31} \mp 2\sqrt{2}EG_F N_e \\ \Delta m^2_{32M} &\approx \Delta m^2_{32} \end{aligned} \tag{2.37}$$

Using Eq. 2.28 and $\cos\theta_{13} \simeq 1$ then gives:

$$\sqrt{P_{sol}}_M = \cos\theta_{23}\sin 2\theta_{12}\frac{\sin\left(aL\right)}{aL}\Delta_{21} \tag{2.38}$$

$$\sqrt{P_{atm}}_M = \sin\theta_{23}\sin 2\theta_{13}\frac{\sin\left(\Delta_{31} \mp aL\right)}{\left(\Delta_{31} \mp aL\right)}\Delta_{31} \tag{2.39}$$

where $a = \frac{1}{\sqrt{2}}G_F N_e$. At an average density of $2.84\,\mathrm{g/cm^3}$ and a baseline of 810 km, $aL \approx 0.218$, meaning that the MSW effect is important for ν_e appearance in NOvA. In general, the ν_μ disappearance probability is also modified when traveling through matter, but these corrections will be small and Eq. 2.29 is quite accurate for NOvA.

There are a few more things to note about the behavior of the ν_e appearance probability

in the presence of matter. With Δm^2_{31} appearing in the denominator of $\sqrt{P_{atm\,M}}$, we have a convenient handle to determine the mass ordering. The ν_e appearance probability will be enhanced for the normal ordering while $\bar{\nu}_e$ appearance will be suppressed (the roles are reversed for the inverted ordering). The value of δ_{CP} will also alter the relative rate of ν_e and $\bar{\nu}_e$ appearance, with $0 < \delta_{\text{CP}} < \pi$ enhancing ν_e appearance and $\pi < \delta_{\text{CP}} < 2\pi$ suppressing it. The MSW effect introduces a great way to measure both the mass ordering and δ_{CP}; however, since both of these measurements rely on a comparison of ν_e and $\bar{\nu}_e$ appearance, certain combinations of δ_{CP} and mass ordering will be degenerate with others for NOvA's oscillation baseline.

2.4 Oscillation Measurements Thus Far

Based on the oscillation probabilities worked out above, along with characteristic baselines and neutrino energies, we can determine which types of neutrino experiments are best able to measure certain oscillation parameters. This is summarized in Table 2.2.

Experiment		Dominant	Important
Solar		θ_{12}	Δm^2_{21}, θ_{13}
Atmospheric			θ_{23}, θ_{13}, $\lvert\Delta m^2_{31,32}\rvert$, δ_{CP}
Reactor	Short	θ_{13}, $\lvert\Delta m^2_{31,32}\rvert$	
	Long	Δm^2_{21}	θ_{12}, θ_{13}
Accelerator	ν_μ ($\bar{\nu}_\mu$) Disapp.	θ_{23}, $\lvert\Delta m^2_{31,32}\rvert$	
	ν_e ($\bar{\nu}_e$) App.	δ_{CP}, sign($\Delta m^2_{31,32}$)	θ_{13}, θ_{23}

Table 2.2: Dominant and important oscillation parameters for various experiment types [45].

As discussed above, solar and reactor oscillation experiments have been able to quite precisely measure the mixing angles θ_{12} and θ_{13} as well as the smaller mass splitting, Δm^2_{21}. Similar to ν_μ disappearance, the $\bar{\nu}_e$ disappearance channel observed by reactor experiments with a baseline of $\sim$1 km is primarily dependent only on θ_{13} and a single effective mass splitting. These experiments have provided a very accurate measure of $\sin^2\theta_{13}$ yielding a global fit at $(2.18 \pm 0.07) \times 10^{-2}$. Using that result to interpret the results of solar experiments, gives best-fits of $\sin^2\theta_{12} = 0.307^{+0.013}_{-0.012}$ and $\Delta m^2_{21} = (7.53 \pm 0.18) \times 10^{-5}$ eV^2. All best-fit values are taken from the PDG [45].

With half of the oscillation parameters being well constrained by solar and reactor experiments, there are still three more left to be measured, namely Δm^2_{32}, θ_{23}, and δ_{CP}. From Table 2.2 we see that these three parameters all come into play for atmospheric and accelerator experiments through the ν_μ ($\bar{\nu}_\mu$) disappearance and ν_e ($\bar{\nu}_e$) appearance channels. Atmospheric experiments have a high statistics sample of neutrinos to observe; however, neither the baseline nor energy can be altered and these experiments rely on the modeling of neutrino production in the atmosphere to extract oscillation parameters. What's more, both δ_{CP} and the sign of Δm^2_{32} are difficult to measure in atmospheric experiments since it is hard to obtain clean samples of only neutrinos or antineutrinos. The high statistics of atmospheric experiments in combination with short baseline reactor experiments and accelerator experiments have allowed for an accurate measure of $|\Delta m^2_{32}|$ with a best fit of $(2.500 \pm 0.065) \times 10^{-3}$ eV^2.

2.4.1 Prospects for NOvA

The angle θ_{23} is dominant only in the long baseline ν_μ ($\bar{\nu}_\mu$) disappearance channel probed by accelerator experiments. These experiments only started in earnest around 2006 and are severely limited by statistics when compared to other neutrino sources. As a result, θ_{23} is the least well constrained of the three mixing angles with a best-fit of $\sin^2\theta_{23} = 0.55 \pm 0.02$. This value is suspiciously close to "maximal mixing", which occurs at $\sin^2\theta_{23} = 0.5$ and means that the ν_3 mass state couples equally to the ν_μ and ν_τ states. If θ_{23} is not maximal, then it can lie in either the upper or lower octant ($\theta_{23} > \pi/4$ or $\theta_{23} < \pi/4$, respectively). As mentioned above, NOvA can measure θ_{23} and $|\Delta m^2_{32}|$ from the dip amplitude and location in the ν_μ ($\bar{\nu}_\mu$) disappearance energy spectrum.

Thanks to the MSW effect, accelerator experiments are able to probe both δ_{CP} and the mass ordering (sign of Δm^2_{32}). However, as we discussed above, the mass ordering and δ_{CP} can work against each other when comparing ν_e to $\bar{\nu}_e$ appearance. Depending on the oscillation baseline and neutrino energy, it is possible for certain parameter combinations to lead to results that are indistinguishable from one another. As you can see in Figure 2.7, in a vacuum and with θ_{23} at maximal mixing (top left), the two mass ordering lie on top of each other and only CP conservation ($\delta_{\text{CP}} = 0, \pi$) or violation ($\delta_{\text{CP}} = \pi/2, 3\pi/2$) can be distinguished with the choice of mass ordering being fully degenerate. If θ_{23} were non-maximal (top right) then the choice of octant moves the ellipses along the CP conservation diagonal. The inclusion of the MSW effect (bottom plots) pulls the δ_{CP} ellipses apart

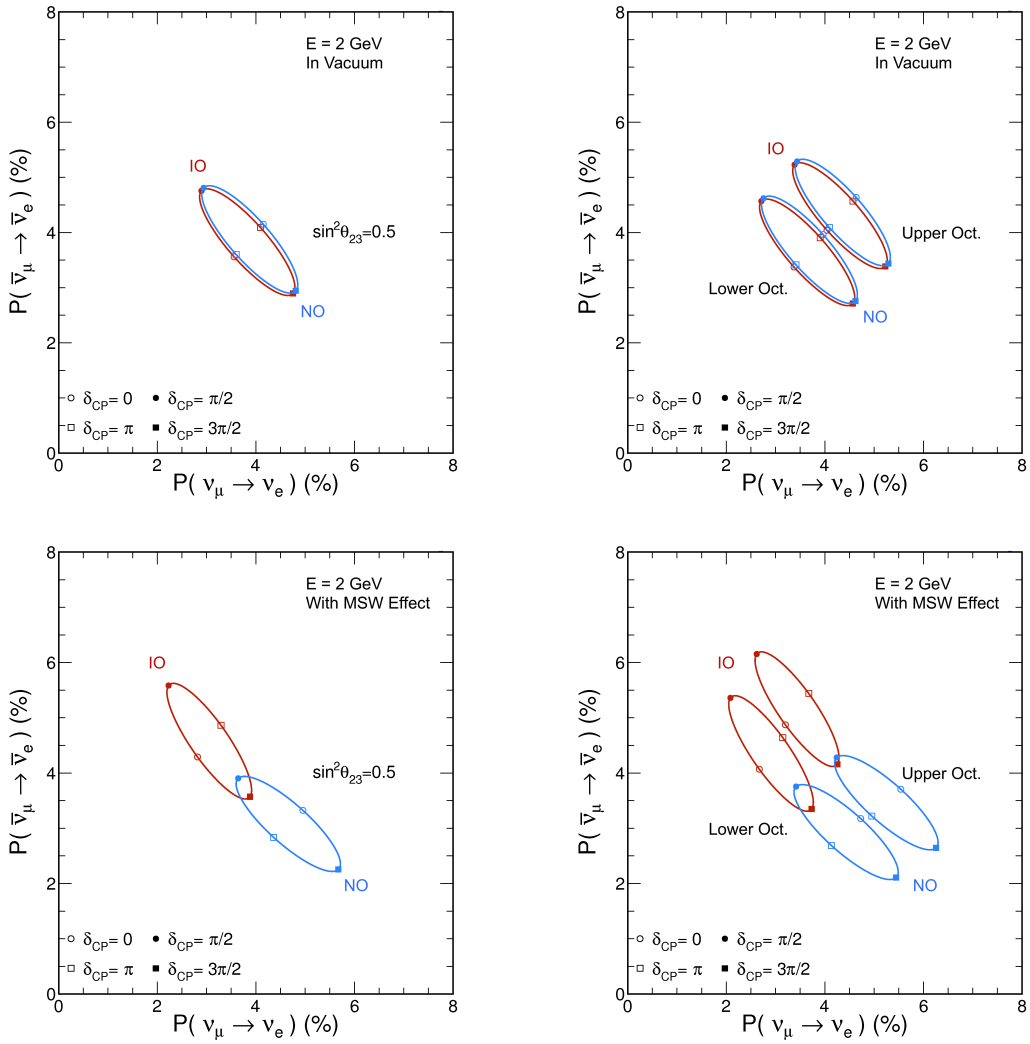


Figure 2.7: Comparison of the oscillation probabilities of ν_e and $\bar{\nu}_e$ appearance as a function of δ_{CP} under various conditions: (top) vacuum oscillation, (bottom) oscillations in matter with the MSW effect, (left) with θ_{23} at maximal mixing, and (right) with θ_{23} in the upper or lower octant. A neutrino energy of 2 GeV and baseline of 810 km was used.

along the CP violation direction with the only degenerate location being right in the middle (depending on the error bars of the measurement). If the oscillation baseline through matter were longer, then the ellipses would be pulled further apart.

By combining ν_μ ($\bar{\nu}_\mu$) disappearance and ν_e ($\bar{\nu}_e$) appearance, NOvA is able to probe both the value and sign of Δm^2_{32}, the mixing angle θ_{23}, and the CP violating phase, δ_{CP}. However, owing to the degeneracies noted above, only certain combinations can be distin-

guished. Thus NOvA can effectively separate between NO with $\delta_{CP} = 3\pi/2$ and IO with $\delta_{CP} = \pi/2$. If nature has chosen parameters in the highly degenerate region, NOvA will be limited by the measurement uncertainties such that determination of the mass ordering, δ_{CP}, and the octant of θ_{23} may not be possible to a high level of significance.

3. The NOvA Experiment

NOvA, the **N**uMI **O**ff-axis ν_e **A**ppearance experiment (here the Greek ν is pronounced here as a "v"), is a long-baseline neutrino experiment based at the Fermi National Accelerator Laboratory (FNAL, aka Fermilab). Fermilab provides a beam consisting of primarily muon neutrinos (or antineutrinos) and NOvA observes ν_μ ($\bar{\nu}_\mu$) $\rightarrow \nu_e$ ($\bar{\nu}_e$) appearance along with ν_μ ($\bar{\nu}_\mu$) $\rightarrow \nu_\mu$ ($\bar{\nu}_\mu$) disappearance. Using a near detector (ND) to sample the initial neutrino composition and a far detector (FD) situated ~809 km away to measure the composition after oscillations, the NOvA collaboration is seeking to resolve key questions in neutrino physics, namely the values of δ_{CP}, Δm^2_{32}, and θ_{23}.

By observing both neutrino and antineutrino oscillations, NOvA can compare the rates to make a measurement of δ_{CP}. Additionally, because the NuMI beam travels through the Earth, NOvA can leverage the MSW effect to, depending on the value of δ_{CP}, probe the mass ordering by measuring the sign of Δm^2_{32} . Finally, with a baseline of ~809 km, neutrino energy peaked at ~2 GeV, and a primarily ν_μ ($\bar{\nu}_\mu$) initial beam, NOvA can measure the values of Δm^2_{32} and θ_{23}.

3.1 The NuMI Beam

To generate a beam of neutrinos, protons are first accelerated to high energy and collided into a stationary target to create a beam of outgoing hadrons. The charged hadrons are magnetically focused to select out either positive or negatively charged particles. The selected hadrons are then allowed to decay, producing mainly muon neutrinos along with other leptons, hadrons, and photons. Finally, the beam passes through absorbers to remove the majority of these other particles, leaving a beam consisting primarily of muon neutrinos. Figure 3.1 shows a schematic representation of a neutrino beam.

NOvA utilizes the NuMI (**Ne**u**t**rinos at **M**ain **I**njector) beam provided by Fermilab. Starting with a 35 keV H^- source, a radio-frequency quadrupole accelerates these ions to 750 keV and sends them into the linear accelerator (Linac) [52]. The Linac further accelerates the ions to 400 keV before sending the ions through a carbon foil, that strips the two electrons from the H^- ions, leaving a proton beam [52]. The proton beam is then injected into the Booster ring, a synchrotron that accelerates the beam to 8 GeV. The

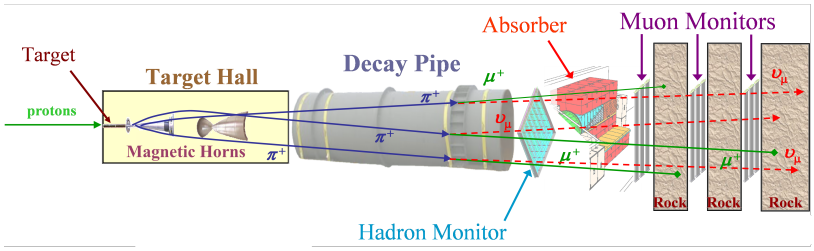


Figure 3.1: Schematic representation of a neutrino beam with high energy protons entering from the left and neutrinos exiting to the right.

Booster "bunches" the beam into 52.8 MHz RF "buckets". Based on its size, the Booster ring can fit 84 bunches (called a "batch") at once [52]. Booster batches are finally injected into the Recycler and Main Injector rings.

The Recycler and Main Injector are two synchrotons housed in the same tunnel. The Recycler receives twelve 8 GeV proton batches from the Booster [52]. It then "slip-stacks" the 12 batches into 6 double-intensity batches. Slip-stacking is achieved by injecting two batches one after the other and slowing the first batch down slightly. When the batches overlap, they are captured and merged into a single batch containing twice the protons. After slip-stacking, the batches are extracted into the Main Injector and accelerated to a final energy of 120 GeV.

The Main Injector was originally designed to deliver 400 kW of beam power, but with the addition of slip-stacking it is reaching average powers of 700 kW or greater, equivalent to $\sim 5 \times 10^{13}$ protons on target (POT) per spill. NuMI spills last for $\sim$10 µs with a current repetition rate is $\sim$1.3 s. The 120 GeV proton beam is collided with a graphite target to produce a neutrino beam as described above. The NuMI beam utilizes two magnetic focusing horns that can be run in forward horn current (FHC) mode to select π^+, in turn producing ν_μ, or reverse horn current (RHC) mode, selecting π^- which decay into $\bar{\nu}_\mu$. An example of the FHC mode is shown in Figure 3.2.

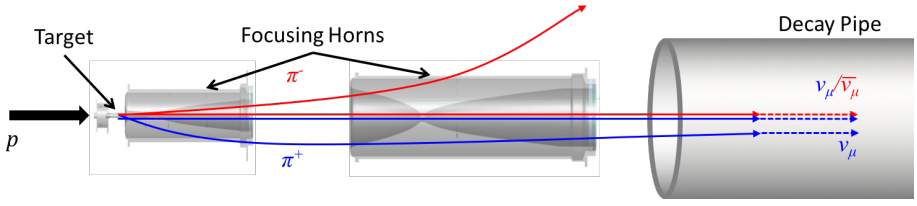


Figure 3.2: Diagram of the NuMI focusing horns in forward horn current mode showing the focusing of π^+ and defocusing of π^- to produce a ν_μ beam.

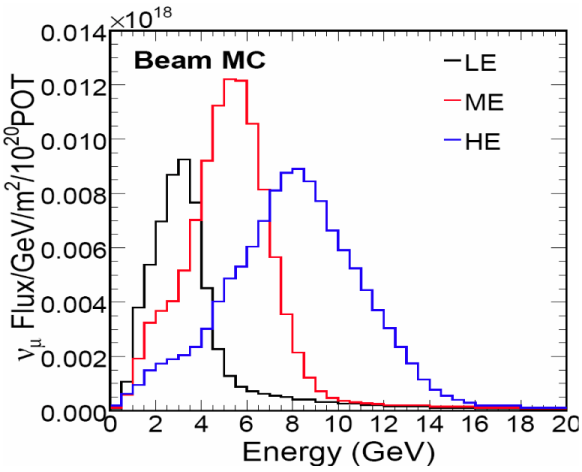


Figure 3.3: Neutrino energy spectrum at 1.04 km from the NuMI production target in the low energy (LE), medium energy (ME), and high energy (HE) configurations [53].

In addition to changing the horn current, both the target and horns can be moved relative to each other to produce different beam energies. Three configurations were conceived: low energy (LE), medium energy (ME), and high energy (HE), each corresponding to the resulting peak in the neutrino spectrum. NOvA utilizes the ME configuration, which provides an on-axis beam peaked at ~5.5 GeV, as shown in Figure 3.3.

To increase the measurement sensitivity of neutrino oscillations, it is important to tune the neutrino energy spectrum. With the oscillation baseline established, typically by feasibility constraints, the most desirable neutrino spectrum is one that is peaked at the energy that lies at a maximum oscillation probability. This choice must also be weighed against the neutrino flux since a high event rate is clearly preferable. To this end, NOvA has decided to place its detectors 14 mrad off of the beam axis. The neutrino energy from pion decay, $\pi^{\pm} \rightarrow \mu^{\pm} + \nu_{\mu}$, can be determined as a function of the pion energy, E_{π}, and angle with respect to the pion direction, θ, and is approximately (see Appendix D for details):

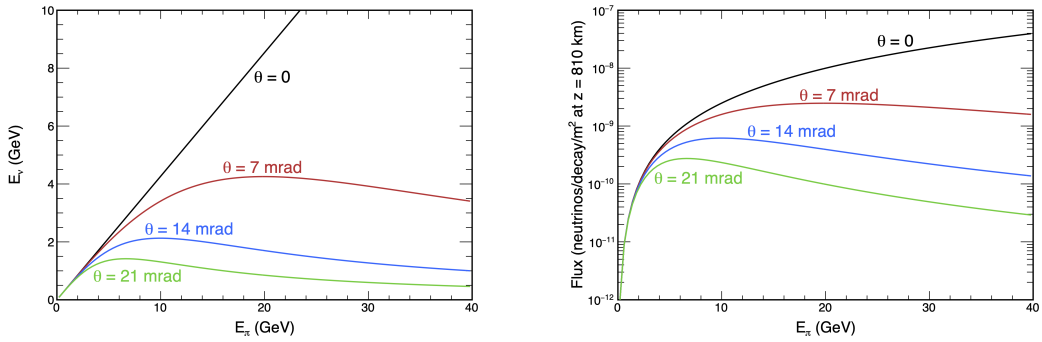
$$E_{\nu} \approx \frac{(1 - m_{\mu}^2/m_{\pi}^2)E_{\pi}}{1 + \gamma^2\theta^2} \tag{3.1}$$

where m_{μ} and m_{π} are the muon and pion masses respectively and γ is the standard Lorentz factor, which here is equal to E_{π}/m_{π}. Likewise, the neutrino flux over some area, A, at a

distance, z, from the pion decay, can be approximated as:

$$\Phi_\nu \approx \frac{A}{4\pi z^2}\left(\frac{2\gamma}{1+\gamma^2\theta^2}\right)^2 \tag{3.2}$$

As you can see from Figure 3.4a, an angle of 14 mrad provides a beam that is peaked at ~2 GeV which is near the oscillation maximum of 1.6 GeV for a 810 km baseline at the current Δm^2_{32} best-fit. Increasing the angle further would favorably shift the beam peak closer to the oscillation maximum, though, as shown in Figure 3.4b this would have a negative impact on the neutrino event rate.



(a) (b)

Figure 3.4: (a) Neutrino energy and (b) flux as a function of pion energy for various angles with respect to the pion direction.

3.2 The NOvA Detectors

NOvA utilizes a two detector approach typical of long-baseline neutrino experiments. The smaller near detector (ND) is located at Fermilab and samples the initial composition of neutrinos. The large far detector (FD) samples the neutrinos at Ash River, Minnesota, and the rate of oscillations can be determined. In order for the beam to pass from Batavia, Illinois to Ash River, Minnesota, the NuMI beam is angled 58 mrad (3.3°) downward through the Earth, which allows NOvA to leverage the MSW effect to probe the neutrino mass ordering.

The detectors are functionally-identical segmented tracking calorimeters. "Functionally-identical", means that the detectors utilize similar designs, materials, and electronics, which allows for the cancellation of many large systematic uncertainties when using the observed

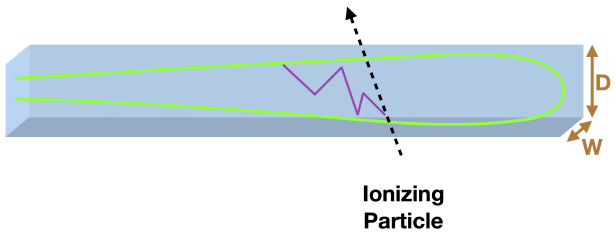


Figure 3.5: Illustration of a NOvA cell showing an incoming ionizing particle (black dotted arrow) giving off scintillation light (purple) which is collected by the loop of wavelength-shifting fiber (green).

ND spectra to predict composition at the FD. The detectors are made from extruded PVC planes containing individual 6.6 cm deep in the beam direction by 3.9 cm wide "cells". The cells are filled with mineral oil that has been doped with pseudocumene, the scintillating agent. The scintillator is ~94.9% mineral oil, ~5% pseudocumene, with the remainder being wavelength shifters, an anti-static agent, and an antioxidant [54]. Each cell is also contains a loop of wavelength-shifting fiber, which collects the scintillation light and directs it up to the readout electronics while also changing the wavelength of the light to the sensitive range of the electronics. The PVC mixture contains 15% titanium dioxide to increase the reflectivity, thus allowing the fiber to capture more of the deposited light. An illustration of a NOvA cell is shown in Figure 3.5.

A single "module" is made up of 32 cells in a single PVC extrusion. Multiple modules are aligned to create single planes. The planes are arranged with the cell lengths lying alternately vertical or horizontal to create two detector views. The top and side views can then be combined to generate three-dimensional events. A schematic representation of the alternating plane structure is shown in Figure 3.6.

The ability to distinguish the outgoing charged leptons types is critical to identifying the neutrino flavors that NOvA observes. Muons, being heavier particles, leave long, relatively straight tracks in the detectors, allowing them to be easily distinguished from other particles. However, the electron signal can be mimicked more easily by other particles. As electrons travel through the detectors, they create EM showers that begin to spread out, while a π^0 quickly decays into two photons, which go on to create a similar shower appearance. To distinguish electrons from other particles, the NOvA detectors are built from "low Z" materials, resulting in a radiation length[1] of ~36 cm or about 5.5 planes. This means that

[1] A radiation length is the mean distance that a high energy electron will travel before losing ~63% of its energy, or 77% of the mean free path that a high energy photon will travel before undergoing pair production.

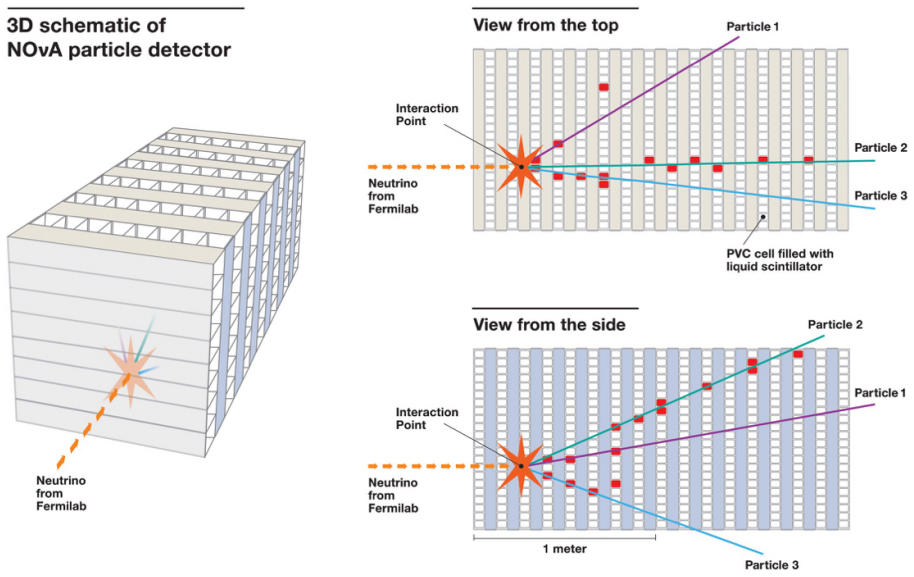


Figure 3.6: A schematic representation of the alternating planes and resulting detector views used for three dimensional particle tracking.

the two photons produced by π^0 decay will travel a significant distance into the detector before they develop EM showers, in contrast to the electron which immediately initiates an EM shower. What's more, the low Z material produces a a Molière radius[2] of about 11 cm, meaning that EM showers will take up 2–3 cell widths, allowing them to be clearly distinguished from tracks left by heavier particles.

3.2.1 The Far Detector

The FD is located 810 km from the NuMI production target and is on the Earth's surface with a modest rock overburden of 3 meter water equivalent (mwe). Being on the surface introduces a significant background of cosmic particles at a flux of ~130 kHz. However, over a 10 µs beam spill this only 1–2 particles, meaning the majority of cosmogenic events can be removed via precision timing for the readout electronics. The small overburden provides about 12 radiation lengths of shielding to reduce the rate of cosmogenic photons, which would mimic a ν_e signal.

While the FD is placed near the oscillation probability maximum, the majority of ν_μ have oscillated into ν_τ and not into the ν_e that NOvA can observe (the tau lepton is too

[2]The Molière radius defines a cylinder about an EM shower, containing 90% of the shower's energy.

massive for the NuMI neutrinos to produce very many and the tau lifetime is very short making that signal difficult to distinguish). This, along with the small neutrino cross section, requires a very large far detector in order to observe enough ν_e events to make a precision measurement. To this end, the FD measures in at about $60 \times 15.5 \times 15.5\,\mathrm{m}^3$ and 14 kton in mass, making it the largest free-standing plastic structure ever built. The mass is approximately 65% scintillator and 35% PVC.

The FD is constructed from ~15.5 m long modules with 12 modules, or 384 cells, per plane. These are grouped into 32 plane "blocks", with 2 blocks constituting one "diblock". The full detector is then made from 14 diblocks (grouping detector elements into blocks and diblocks is mainly a factor of construction and electronic outfitting). Altogether, the FD has 896 planes and 344,064 cells.

3.2.2 The Near Detector

The ND is placed 1015 m from the production target onsite at Fermilab, allowing ample room for the pions (and kaons) to decay into neutrinos. Additionally, the ND is 105 m underground, 225 mwe, which reduces the rate of cosmogenic particles to ~40 Hz. Such a low cosmic rate means it would take ~2500 beam spills on average to capture one coincident cosmic-ray. However, being so close to the beam source means that the ND observes 5–10 neutrino events per spill; the implications of this will be discussed in detail later.

The ND is about four times smaller than the FD in each dimension and only has 20,192 cells. The ND consists of 3.9 m long modules, three of which form one plane. The main detector body has 192 planes for a total length of 12.9 m. A muon with 1–3 GeV of kinetic energy will travel 4.8–14.1 m on average in the NOvA detectors. Therefore, many ν_μ events would not be completely contained by in a 12.9 m detector, since they could interact anywhere within the detector volume and not just at the front. To solve this issue, the main detector body is followed by the "muon catcher", which has 11 plane pairs with 10 cm steel plates between each pair to increase the stopping power.

In the muon catcher, the horizontal planes are only two modules tall and the vertical planes are 2.6 m long. The height of the muon catcher was determined by the steel plates that were readily available at the time of construction. However, being only two thirds the height of the main detector poses a minimal issue. With the beam being angled downward by 3.3° a neutrino entering at the very top front of the detector (3.8 m high) will be at 3.1 m high by the time it reaches the back of the main detector body. Therefore, a 2.5 m tall muon

catcher will still provide good coverage for muons in the few GeV range. A scale model of the NOvA detectors is shown in Figure 3.7.

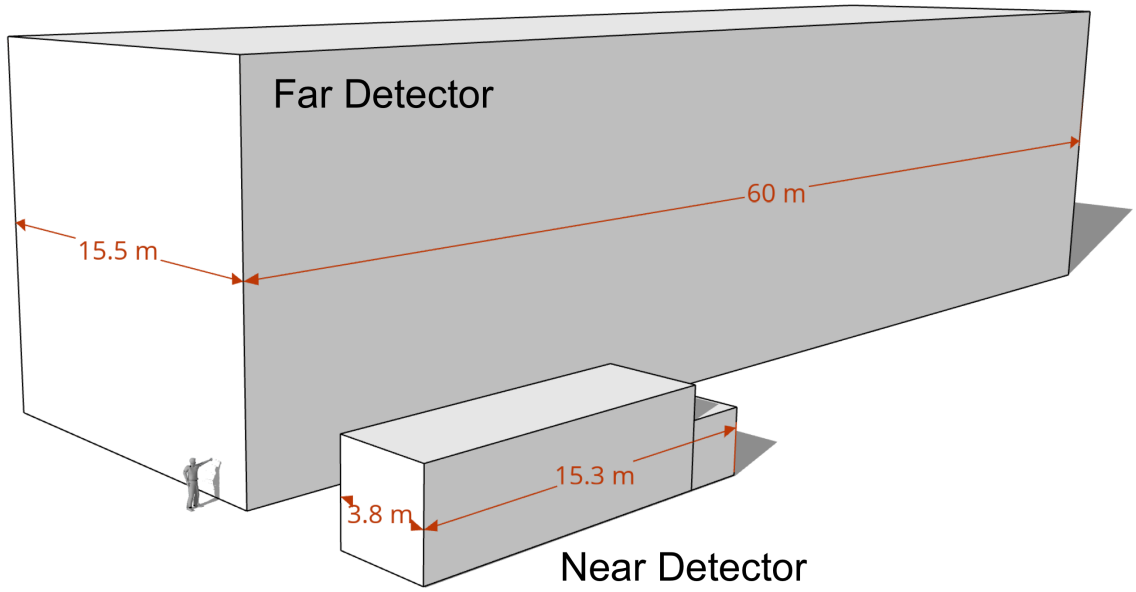


Figure 3.7: Scale models of the NOvA detectors.

3.3 The Data Acquisition System

The data acquisition (DAQ) system is responsible for reading out and saving interesting physics events in the NOvA detectors. This is comprised of avalanche photodiodes, front end boards, data concentrator modules, and timing distribution units, along with power distribution boxes and a host of servers to manage the components and process the data. Since the primary purpose of NOvA is to observe weakly interacting neutrinos, the DAQ was designed for continuous readout where the data is constantly collected and separate servers later make the decision whether to keep the data or drop it.

3.3.1 Avalanche Photodiodes

Custom 32-channel Hamamatsu avalanche photodiodes (APDs) are used to readout all the cells of one module. Both ends of the loop of wavelength shifting fiber are optically connected to a single APD pixel. The APDs have 85% quantum efficiency for the

500–550 nm light emitted by the fibers. "Hits" are recorded when the energy deposited in a cell is greater than a preset ADC threshold (ADC stands for analog to digital conversion and is the amplified signal output by an APD). Thresholds are determined individually for each APD pixel based on the amount of observed electronic noise. To reduce noise, the APDs are cooled to −15 °C by a thermoelectric cooler. This cooling, however, increases the possibility of condensation forming on the APDs which is prevented by constantly flowing dry air over the boards.

The APDs are supplied with a maximum bias voltage of 425 V which can be regulated down to tune the gain. Originally the APDs were running with a gain of 100 based on initial studies showing that this value allowed for the lowest possible hit threshold while keeping the noise rate to a minimum. Once the large FD was built it was found that hits occurring further down the long cells were falling below the ADC thresholds due to the increased light attenuation. At a gain of 100, the efficiency of recording hits at the far end of a cell dropped down to about 84% compared to the 98% and higher efficiency closer to the APDs. The highest possible gain that could be achieved by all APDs on both of the detectors was calculated to be 150[3], which was set after the first run at full operation. This higher gain increases the hit efficiency to greater than 90% along the entire length of an FD cell.

3.3.2 Front End Boards and Data Concentrator Modules

Each APD is connected to one front end board (FEB). FEBs perform the integration, shaping, digitization, and packaging of the APD signals according to the preset APD pixel thresholds mentioned above. The APD signals are read by a custom application-specific integrated circuit (ASIC). The ASIC takes in 32 input channels (one for each APD pixel) that are each amplified, integrated, and shaped before being "multiplexed", where multiple channels are grouped together and sent out sequentially to the analog-to-digital converter (ADC). The multiplexing occurs at 16 MHz meaning each channel in the group is sent to the ADC once every 62.5 ns.

As mentioned previously, the FD cosmic rate results in 1–2 cosmogenic particles passing through the detector over a 10 µs time window. This allows the FD multiplexing ratio of 8:1 (8 input channels are grouped into 1, resulting in 4 groups total), so at 16 MHz each channel is multiplexed once every 500 ns. With the significant pile-up at the ND, a multiplex ratio of 2:1 is used, resulting in a much faster sampling rate of 125 ns. The decreased multiplexing

[3] Simulations have shown that the actual gain at the determined voltages is closer to 140.

ratio means that the ND FEBs require 16 ADCs while the FD boards only need 4.

After the ADC, the digitized signals are sent to a field-programmable gate array (FPGA) on the FEB. The FPGA is loaded with the pixel thresholds and decides whether to keep the signals or drop them. This is done via dual-correlated sampling (DCS) with 3 time steps, where for the nth sample: $DCS_n = ADC_n - ADC_{n-3}$. This method removes low frequency noise and prevents a wandering baseline from impacting the threshold. If the DCS value for a sample is above the preset threshold, then either the single sample is recorded (single-point readout) or that sample along with the three preceding samples are recorded (multi-point readout). For single-point readout, the best possible timing resolution can be determined from a uniform distribution between two successive samples:

$$\sigma^{\text{ND}}_{\text{single}} = \frac{125\,\text{ns}}{\sqrt{12}} = 36.1\,\text{ns}$$
$$\sigma^{\text{FD}}_{\text{single}} = \frac{500\,\text{ns}}{\sqrt{12}} = 144.3\,\text{ns}$$

While in multi-point readout, a fit of the pulse shape from the four recorded samples is performed to more precisely determine the hit time. Empirical studies of multi-point cosmic data have found that this method improves the timing resolution to a maximum[4] of 10–20 ns at the FD and 5–10 ns at the ND. Like the increased gain, multi-point readout was implemented after the initial data run in 2014.

The data concentrator modules (DCMs) are responsible for collecting data from up to 64 FEBs, combining it all, and sending the data packets to a computing farm for further processing. A DCM consists of an FPGA and a small Linux computer which allows for user control and transmission to the computing farm. The FPGA receives the hit packets from each connected FEB and collates them into 50 µs groups called "microslices". These microslices are then grouped again into 5 ms "millislices", which are sent off to a farm of servers which then decide whether to save the data to disk or drop it.

3.3.3 Timing System

Precise, accurate, and synced timing is paramount to capture beam spills occurring over 800 km away and events across large detectors. Additionally, the precise time of the beam spill can vary between 1–1.5 s; therefore a simple repeating trigger is not feasible. When a

[4]The timing resolution is dependent on the number of photoelectrons produced by a particle interaction in a cell.

beam spill occurs a signal from the accelerator division is received by a timing unit at the ND. This signal is then sent to the FD timing system while correcting for the time-of-flight between the detectors (~2.7 ms).

The timing system consists of GPS receivers and timing distribution units (TDUs). A primary TDU at each detector obtains a clock signal from the GPS system, ensuring that both detectors are synced to the same global reference clock. This signal is passed to a set of secondary TDUs which connect to the DCMs. The secondary TDUs are arranged in a chain with the final TDU in the line having a loop-back to pass the signal back up the chain. Each secondary TDU is then connected to a chain of DCMs, with the final DCM also having a loop-back. These loop-backs are used so the TDUs can account for delays from the lengths of cable by measuring the time difference between the initial signal and the echoed signal that passes back up the chain. To sync the timing chains, the primary TDU sends a signal stating what the future timestamp will be. Then the sync signal is sent down the line and each component begins counting down their delay times. When the countdown ends, each component begins counting from the timestamp that was sent and all components are now in sync.

3.3.4 Triggering

The last stage in the DAQ is the farm of servers called buffer nodes. Each DCM sends the same interval millislice to a single buffer node, and the next millislice is then sent to the next node. In this way, each buffer node receives a 5 ms snapshot of the entire detector. Once all nodes have one millislice, the next millislice goes back to the first node and cycle repeats until there is all nodes are completely filled. If a node is filled then the next incoming millislice pushes out the earliest one in memory in a first-in-first-out model. This means that the amount of look-back time in the buffer farm is dependent on the data rate coming from the detector and the number of buffer nodes in the farm. The typical look-back time is about 20 minutes on each detector.

The buffer nodes make the decision whether to save any "interesting" data or allow it to drop. Data is saved in the 50 µs microslice time windows. If an event crosses a microslice boundary, both slices are saved. When a trigger is received all buffer nodes look through their data and grab the millislices coincident with the trigger time. The buffer nodes then all send their interesting millislices to the data logger which saves the data to disc.

Each NuMI beam spill is saved regardless of the amount of activity in the detectors.

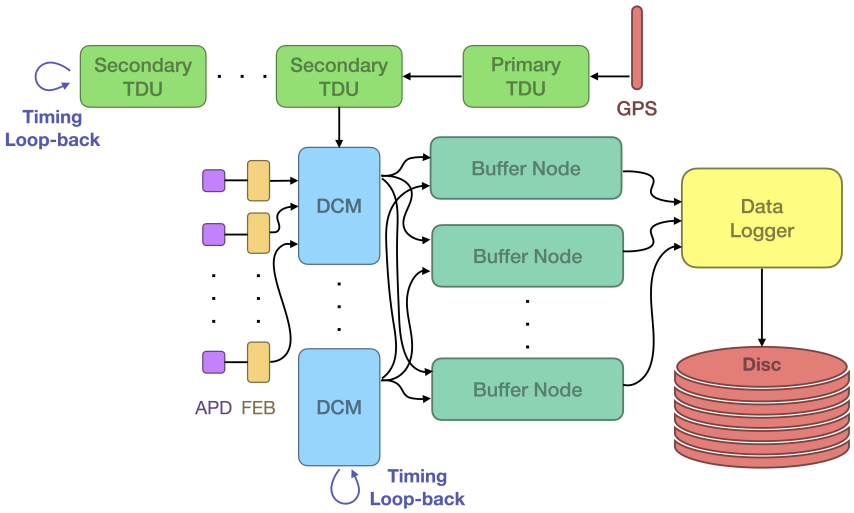


Figure 3.8: Overview of the NOvA DAQ showing the data paths from APD to disc and the timing loop-backs at the end of the timing chains.

Although the spills only last 10 µs, a larger 500 µs window[5], centered on the beam spill, is saved. This large window allows for some amount of drift in the timing system and also collects an out-of-beam side-band which is used to estimate in-spill backgrounds.

In addition to spill triggers, the DAQ can also save clock-based triggers and perform some minimal data reconstruction to generate data-driven triggers (DDTs) based on specific criteria in the detectors. A minimum-bias cosmic trigger at the FD is saved at 10 Hz which provides a beam-independent sample of particles for calibration and background estimation. At the ND, an activity based DDT which records events that have total detector activity above a threshold is also used for calibration and to monitor the detector health. Further DDTs are also employed to capture other interesting physics like supernovae, magnetic monopoles, and dark matter searches, but these are not relevant to the current dissertation. An overview of the DAQ and dataflow is shown in Figure 3.8.

3.4 Detector Calibration

Calibrating the detectors requires both a timing and energy component. The timing calibration acts as a check and correction on the delays calculated via the timing loop-backs on the detectors. The energy calibration converts the recorded ADC values for each hit into

[5]These time windows typically end up being 550 µs long since events occurring on a slice boundary are common.

an amount of deposited energy in MeV. Energy calibration occurs in two steps: relative and absolute calibration. The relative calibration corrects the ADC values such that some amount of deposited energy is the same regardless of when or where it occurred, while the absolute calibration sets the overall energy scale giving each hit a meaningful energy in MeV.

The key to energy calibration comes from the Bethe-Bloch formula (Eq. 3.3). This formula gives the mean rate of energy loss of a relativistic "heavy" charged particle as it travels through some target medium. It is also referred to as the "stopping power" of the target medium. For most target materials, an incident particle with $\beta\gamma$ between 2–10 will be at or near a minimum in this curve (shown in Figure 3.9), and is called a "minimum-ionizing particle" or MIP.

$$\left\langle -\frac{dE}{dx} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 W_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right] \tag{3.3}$$

with:

β:	Incident particle velocity relative to the speed of light	z:	Charge number of incident particle
γ:	Lorentz factor = $(1-\beta^2)^{-1/2}$	Z:	Atomic number of target
$m_e c^2$:	Electron mass $\times\, c^2$ = 0.511 MeV	A:	Atomic mass of target
r_e:	Classical electron radius = 2.818 fm	$W_{\max}$:	Maximum possible energy transfer to an electron in a single collision
N_A:	Avogadro's number = 6.022×10^{23} mol^{-1}	I:	Mean excitation energy of target
K:	dE/dx coefficient = $4\pi N_A r_e^2 m_e c^2$ = 0.307 MeV mol^{-1} cm^2	$\delta(\beta\gamma)$:	Correction due to the density effect as a function of β and γ

for an incident particle of mass M we have:

$$W_{\max} = \frac{2m_e c^2 \beta^2 \gamma^2}{1 + 2\gamma m_e / M + (m_e)/M)^2}$$

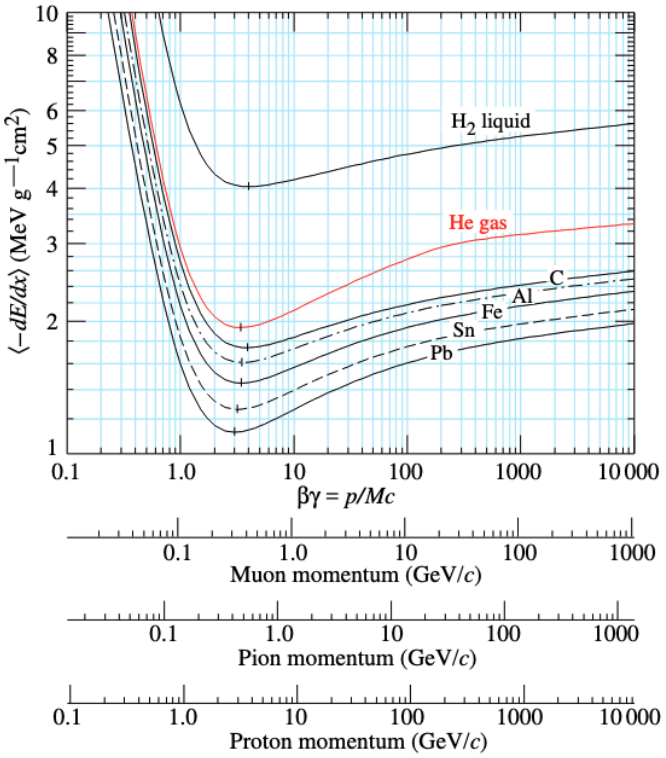


Figure 3.9: Mean energy loss rate in various materials for muons, pions, and protons. The minimum-ionizing point is shown as a vertical dash [45].

3.4.1 Timing Calibration

The first step in timing calibration is to select suitable tracks. At the FD, abundant cosmic muons are utilized; however, because the ND is underground, the rate of cosmic rays is severely reduced and muons originating from beam interactions in the surrounding rock are used instead. Highly vertical tracks are disregarded since they will produce many photons along the length of a single vertical cell and similar events are rare in the horizontal cells. Tracks are required to be very straight-going to ensure they were produced by muons rather than EM or neutron-induced events. Additionally, selected particles must enter and exit the detector so that the energy loss (and deceleration) can be ignored and speed of light can be safely assumed.

The hit time from the DCM is corrected for the distance of the hit from the readout and the time-of-flight of the muon through the detector. If the detector were fully synchronized then these corrections would produce the same time values for all hits on a single track. However, since the DCM times are not calibrated, we use the difference in hit times to

determine the necessary delays for all DCMs relative to one fixed DCM. This procedure produces timing calibration offsets for each DCM with an error of less than 10 ns.

3.4.2 Relative Energy Calibration

The goal of the relative calibration is to correct the observed photoelectrons (PE) to account for the attenuation of light as it travels along the fiber lengths. It also will account for natural variations in the electronics, scintillating oil, and cell reflectively. From Figure 3.9, we can see that muons with momentum between 0.2–2 GeV/c interacting with carbon or hydrogen have an approximately constant energy deposition and the dE/dx is relatively flat even out to ~100 GeV. Luckily, this momentum range coincides nicely with the peak of the cosmic-ray muon spectrum. This means that along an average cosmic muon track, we can assume that the same amount of scintillation light should have been produced in each cell. It is important to select through-going muons that traverse the detector to avoid the sharp rise in energy deposition that occurs as the muon comes to a stop.

Three things are needed to perform the relative calibration: a conversion of ADC to photoelectrons; the position along the length of the cell where a hit occurred, W; and the path length of the muon through the cell. A simple conversion factor is used to convert ADC to PE, while W is determined by using the adjacent planes in the other view (i.e. the position along a vertical cell can be estimated from which horizontal cells were hit in the neighboring planes). The path length through a given cell is not as simple to determine since many paths could produce the same signature. Therefore, only cells along the track with hits in both neighboring cells on the same plane are used. This "tricell" requirement means that we know the particle had to pass through the two opposite walls of the cell, and thus, the path length through the cell is the cell width divided by the cosine of the direction of the particle through the cell. With PE, W, and cell path length determined, the attenuation fit can be performed by plotting the average PE/cm vs W for each cell and fitting the profile with:

$$y = C + A\left(e^{\frac{W}{X}} + e^{-\frac{L+W}{X}}\right) \tag{3.4}$$

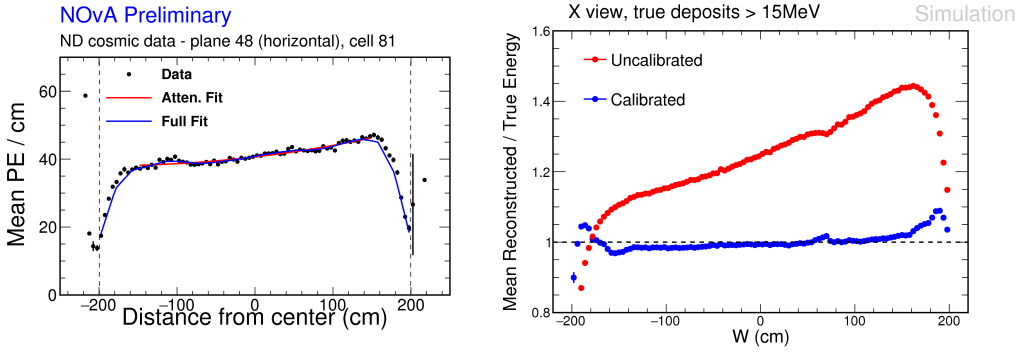
where y is PE/cm, L is the cell length, W is the position along the cell as measured from the center, and C, A, and X are the fit parameters. The second exponential accounts for the path of light traveling down the cell and along the loop back up to the readout.

Before performing the attenuation fit, corrections must be applied to account for "thresholding" and "shadowing". The further down a cell that a hit occurs, the more likely it is

to fall below the hit threshold due to attenuation. This means that more hits at the far end of a cell will require an upward fluctuation in the amount of scintillation light produced as compared to hits nearer to the readout. The shadowing effect is caused by the distance that a muon has to travel through the detector. As seen in Figure 3.9, the energy deposition is not quite flat, so as the muon travels through the detector and loses energy, the rate of energy loss changes. These two effects are accounted for using a sample of simulated cosmic events to determine a correction factor, T:

$$T = \frac{PE}{\lambda}\frac{E_{true}}{E_{MIP}} \tag{3.5}$$

where PE is the amount of simulated photoelectrons collected, λ is the amount of PE that would be generated without any statistical fluctuations, E_{true} is the true energy deposited in the cell, and E_{MIP} is the amount of energy deposited by an actual minimum-ionizing particle. Instead of being calculated for each individual cell, this correction is calculated as a function of view (vertical or horizontal), cell number, W, and (at the FD only) fiber brightness[6]. After applying Eq. 3.5 then PE/cm vs W is plotted and the fits are performed using Eq. 3.4. The attenuation fit and resulting relative calibration is shown in Figure 3.10.



(a) (b)

Figure 3.10: (a) The attenuation fit in a ND cell. An additional fit is applied to account for the "roll-off" at the ends of the cell. (b) The mean ratio of reconstructed and true energy as a function of W before and after calibration. The relative calibration flattens the curve and the absolute calibration moves it up or down.

[6]The fiber brightness is a measure of the attenuation length of each fiber. Manufacturing differences lead to different brightness values, which are accounted for in the simulation. Fiber brightness differences also occur in the ND, but the impact is less significant since the cells are much shorter.

3.4.3 Absolute Energy Calibration

The absolute energy calibration is the final step and takes the corrected PE per cm output by the relative calibration and determines the scaling factor to give a meaningful estimate of the energy deposited in GeV. As in the relative calibration, the absolute calibration utilizes tricell hits from cosmic-ray muons. However, here stopping muons are used as opposed to through-going ones. Again, the Bethe-Bloch formula is employed, but by using stopping muons, the distance of a cell hit from the end of the track can be determined and used to select only cells that are in the MIP region. For the NOvA detectors, the MIP region lies between 100–200 cm from the end of the muon track.

With the hits selected, the simulated energy deposition of a MIP in MeV is divided by the average corrected PE/cm to produce an absolute calibration scale factor. Since this is done over an average of many hits, there is a possibility of bias due to an asymmetry in the hit rate between the orthogonal detector views. The vertical planes will have fewer hits than the horizontals since cosmic-rays are predominately downward-going and can travel far along a single vertical cell. This means that the scale factor would be driven more by the horizontal planes than the verticals. To avoid this possible bias, the absolute calibration is performed separately for each view.

It is well known that as the fibers, scintillating oil, and electronics age, their light transmission, output, and response decrease. This aging effect has been seen to reduce the number of collected hits from an event by about 0.3% per year in each detector. To handle this drift, the absolute calibration procedure is performed over short timescales called "epochs". Epochs are defined when beam, file production, or analysis dictate, resulting in around 2–5 epochs per beam run[7]. Both the relative and absolute calibrations are performed similarly on data and simulation.

3.5 Simulations

This section will describe NOvA's simulation chain or Monte Carlo (MC). Specifically, the following describes the simulation used in the 2020 analysis, which is the basis for this dissertation. NOvA's MC can be broken into five parts, each serving as the input into the next:

[7]Beam runs typically start in the fall and last until the beginning of summer. In the summer the NuMI beam is shutdown for repairs, improvements, and to save money on cooling in the hot months.

- G4NuMI: A GEANT4-based simulation of the incoming proton beam interacting with the production target and the subsequent particle interactions to produce a simulation of the neutrino flux at the NOvA detectors.
- GENIE: A neutrino interaction generator which determines the interaction mode, kinematics, and the resulting final state particles [55, 56].
- GEANT4: Takes the final state particles from GENIE and simulates their passage through matter and the related energy depositions [57, 58].
- Photon production and transport: Converts the GEANT4 energy depositions into scintillation and Cherenkov photons, then transports them through the cells, up the fibers, and to the APDs.
- Electronics: A custom simulation of NOvA's front end electronics, taking the simulated photoelectrons from the APD and performing the pulse shaping, digitization, and noise simulation giving a simulated hit that is comparable to the recorded data.

G4NuMI starts with a beam of 120 GeV protons and simulates the interactions on the production target. The target geometry includes the full target assembly including the material for cooling, support, and the magnetic horns. The protons produce a cascade of hadrons, which then can also interact. All of these interactions are simulated using the FTFP_BERT model, which is the Fritiof model [59] for energies above 4 GeV and the Bertini model [60] for energies below 5 GeV along with "standard" EM physics [57]. Hadronic modeling, however, is difficult to get correct due to messy strong interactions. To correct for observed mis-modeling, NOvA uses the package to predict the flux (PPFX) which was originally designed for the MINERvA experiment [61]. The PPFX corrections are driven by data from various experiments and determines correction weights to the hadronic model between 12–120 GeV. Additionally, using a "multi-universe approach", the uncertain parameters are tweaked and the procedure is rerun many times to propagate uncertainties. The simulated, PPFX-corrected neutrino flux at the ND is shown in Figure 3.11 for both beam modes.

The flux determined by G4NuMI and PPFX is used as the input to the GENIE neutrino interaction generator. In the 2020 analysis, NOvA used GENIE version 3.0.6. This version of GENIE allows for the selection of groups of various interaction models that work well together. The following models are used:

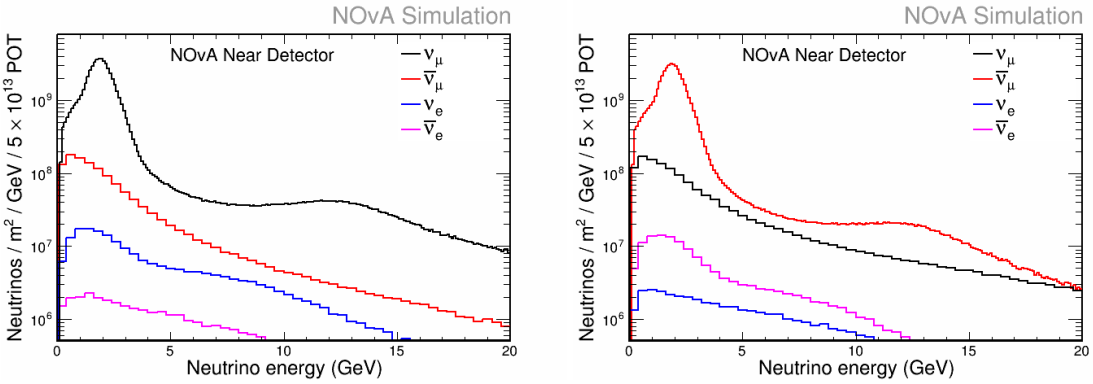


Figure 3.11: PPFX-corrected NuMI beam flux at the ND showing the various neutrino components in (left) the neutrino beam mode and (right) the antineutrino beam mode.

- Charged-current quasi-elastic and two-particle-two-hole: València [62, 63].
- Resonance and coherent pion production: Berger-Sehgal [64, 65].
- Deep inelastic scattering: Bodek-Yang with data-driven parameterization for hadronization [66, 67].
- Final state interactions (FSI): hN semi-classical intranuclear cascade [68].

A local Fermi gas is used for the initial nuclear states in CC QE and 2p2h interactions while a global relativistic Fermi gas is used in all other interactions.

Of course, no model is perfect and in the past NOvA has needed to perform a significant amount of tuning to the GENIE models in order to obtain better agreement with our observed ND data. With this version of GENIE, only the 2p2h and FSI models are adjusted. The 2p2h tune modifies the ν_μ CC interactions using a technique similar to the one developed by MINERvA [69]. NOvA, however, uses two 2D gaussians in the $(|\boldsymbol{q}|, q_0)$ phase-space, where q is the four-momentum transfer, to adjust the simulated 2p2h component to improve the data-MC agreement. The FSI hN model parameters were tuned to various π^+ on ^{12}C scattering experiments. The result of the tuning procedure is shown in Figure 3.12 as a function of reconstructed $|\boldsymbol{q}|$.

The particles coming from the GENIE interactions are then the input to another GEANT4 simulation, which propagates the particles through the detector geometry and determines how much energy is deposited and where. In modelling the density correction in the Bethe-Bloch formula ($\delta(\beta\gamma)$ in Eq. 3.3) GEANT4 uses an approximate parameterization from 1984 based on tabulated parameters for different materials. However, NOvA has

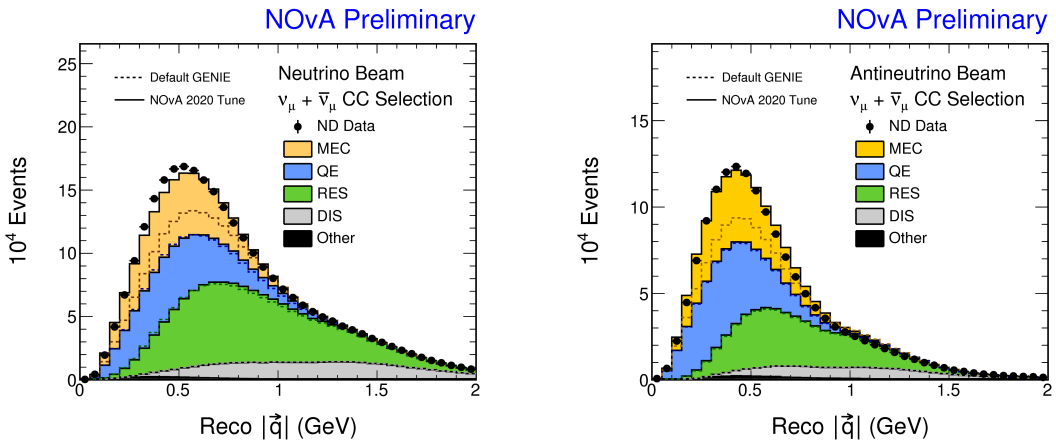


Figure 3.12: Spectra of ν_μ and $\bar{\nu}_\mu$ CC events in the ND as a function of three-momentum transfer. The dashed lines show the default GENIE configuration and the bold lines show the spectra after tuning.

implemented a patch to perform an exact calculation of this correction using Sternheimer's method [70]. The exact calculation increases the range of muons in by 1% for energies between 0.1–2 GeV in the ND muon catcher steel and 0.3% for a 1 GeV muon in the main detector body. What's more, this method decreases the amount of collectable energy deposition, with about 0.9% more energy being deposited in the PVC cell walls rather than the scintillator.

The energy depositions determined by GEANT4 are converted to scintillation photons on a cell-by-cell basis using a model that includes scintillation as well as Cherenkov light:

$$N_\gamma = F_{view}\left(Y_s E_B + \epsilon_C C_\gamma\right) \tag{3.6}$$

where:

N_γ:	total number of photons produced	E_B:	energy deposited according to Birks' law [71]
F_{view}:	a scale factor depending on the detector view (x or y)	ϵ_C:	scintillator efficiency for Cherenkov photons
Y_s:	scintillation photons produced per unit energy deposition	C_γ:	number of Cherenkov photons produced by the charged particle

and E_B is determined using Birks' law:

$$\frac{dE_B}{dr} = \frac{dE/dr}{1 + k_B dE/dr} \tag{3.7}$$

which gives the modified energy deposition in a scintillator by parameterizing the amount of quenching[8] using the Birks constant, k_B. NOvA uses an in-house measurement of k_B ($0.011\,55\,\mathrm{g/cm^2/MeV}$ and performs a fit to determine the remaining parameters in Eq. 3.6. The light model fit is performed in the 2D space of distance to end of track and corrected PE/cm. Four sets of data and MC samples are used: ND and FD cosmic rays, ND muon tracks, and ND proton tracks, where the muon and proton tracks come from ν_μ events generated by the NuMI beam. The MC cosmic ray samples are generated using the CRY simulator [72].

GEANT4 can be used to transport the photons in the cells and up the fibers, though, this would be a very time consuming process. Instead, NOvA uses a simulation-derived template which gives the photon collection rate as a function of position along the length of a cell. This template also provides an arrival time based on the many possible paths that light could travel up the fibers. The photon transport also includes light attenuation along the fibers drawn from a set of quality control tests performed on the actual fibers before installation. Fluctuations in the number of collected photons are modeled using Poisson statistics.

The final stage of simulation is to model the response of the APDs along with the shaping and digitization of the FEBs. The collected photons are adjusted for the APD efficiency and smeared by sampling a log-normal distribution to account for excess APD noise [73]. The FEB pulse shaping is performed by a CR-RC circuit and is modeled using two exponentials, one for the pulse rise and the other for the fall. ADC conversion is the performed on the shaped pulses while accounting for the ballistic deficit[9]. Finally, electronics noise is modeled with correlated noise occurring between adjacent samples and among all pixels of a single APD since they all share a common reference voltage.

It would be technically possible to completely simulate cosmic ray events in the FD and events originating in the surrounding rock for the ND, but this would be prohibitively time consuming. Therefore, cosmic ray events are collected by a 10 Hz trigger and this cosmic activity is overlaid onto the MC events to produce a realistic cosmic background. At the

[8] Internal absorption of photons or other reduction in energy transfer inherent to a scintillator.

[9] Ballistic deficit is the reduction in pulse amplitude as a result of a finite shaping time constant which causes less than the full amount of charge to be collected.

ND, special samples of rock events were simulated using interactions occurring in the rock in front of the detector. These rock samples are then overlaid at a rate determined by the average POT/spill for the associated time period.

4. Event Reconstruction

In order to use the information recorded by the detector for a physics analysis, it is necessary to perform some amount of event reconstruction. The reconstruction chain takes in calibrated cell hits and runs various algorithms to group together the hits originating from a single particle interaction and extract kinematic and identification information. These algorithms are developed and tested either on simulated events or using well understood data as a standard candle. With MC events you have access to the underlying "truth" information (i.e. where exactly the interaction occurred, what the incoming particle energy was, etc.); however, simulations rely on underlying models, none of which are perfect representation of the true physics. Utilizing actual data has the benefit that the algorithms are developed with the real detector response and physics, but you must rely on well-understood events which may be limited.

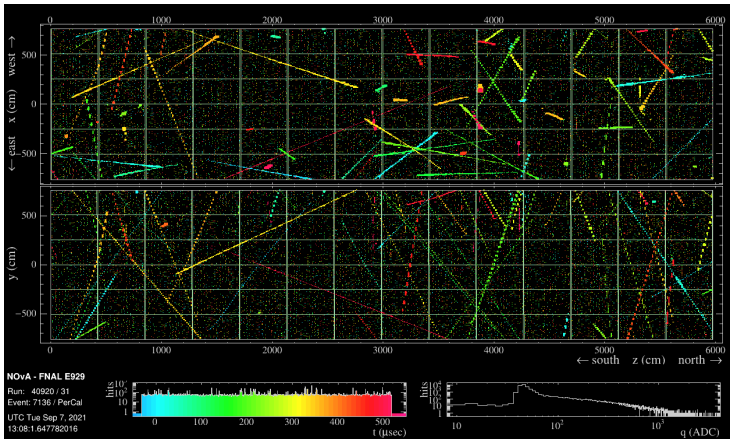
NOvA's reconstruction is performed in three parts. First, hits originating from a single interaction are identified and the tracks or showers of the outgoing particles are formed. Using this information, both the incoming and outgoing particle types can be determined along with the interaction type. Finally, based on the reconstructed event and the particle/event classification, the energies of the particles are estimated. Since NOvA is observing interacting neutrinos, there is no track left by the incoming particle. Instead, NOvA relies solely on the final state to extract the neutrino flavor and energy which are used to probe neutrino oscillation physics.

4.1 Reconstruction Algorithms

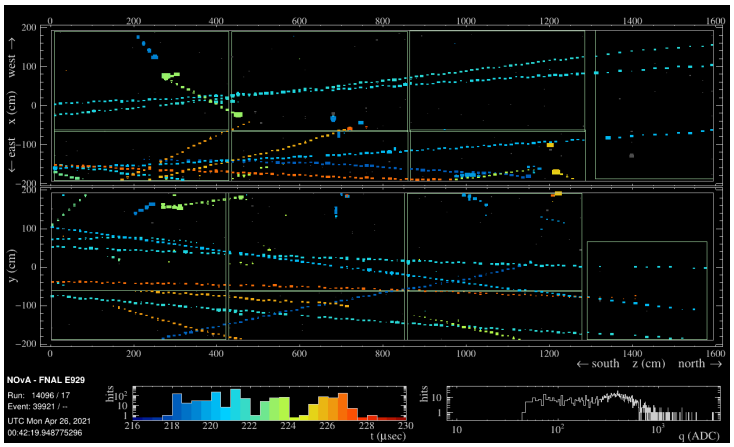
As mentioned above, the first part of extracting physics information is to group hits into single interaction events. Then, from these events, further grouping is performed to form tracks and showers associated with the outgoing particles. With this information, basic kinematic quantities can be extracted such as outgoing particle angle and energy, and the energies of the outgoing particles can be summed to determine the energy of the initial neutrino.

4.1.1 Event Slicing

For a given readout window, there can be many overlapping particle interactions as shown in Figure 4.1. From timing alone, it is clear that you can separate out the majority of the individual particle interactions. However, timing alone is not enough since particles may arrive simultaneously, as seen in the FD (Figure 4.1a). Therefore, both temporal and spatial information is used to perform the event "slicing", or initial hit grouping.



(a)

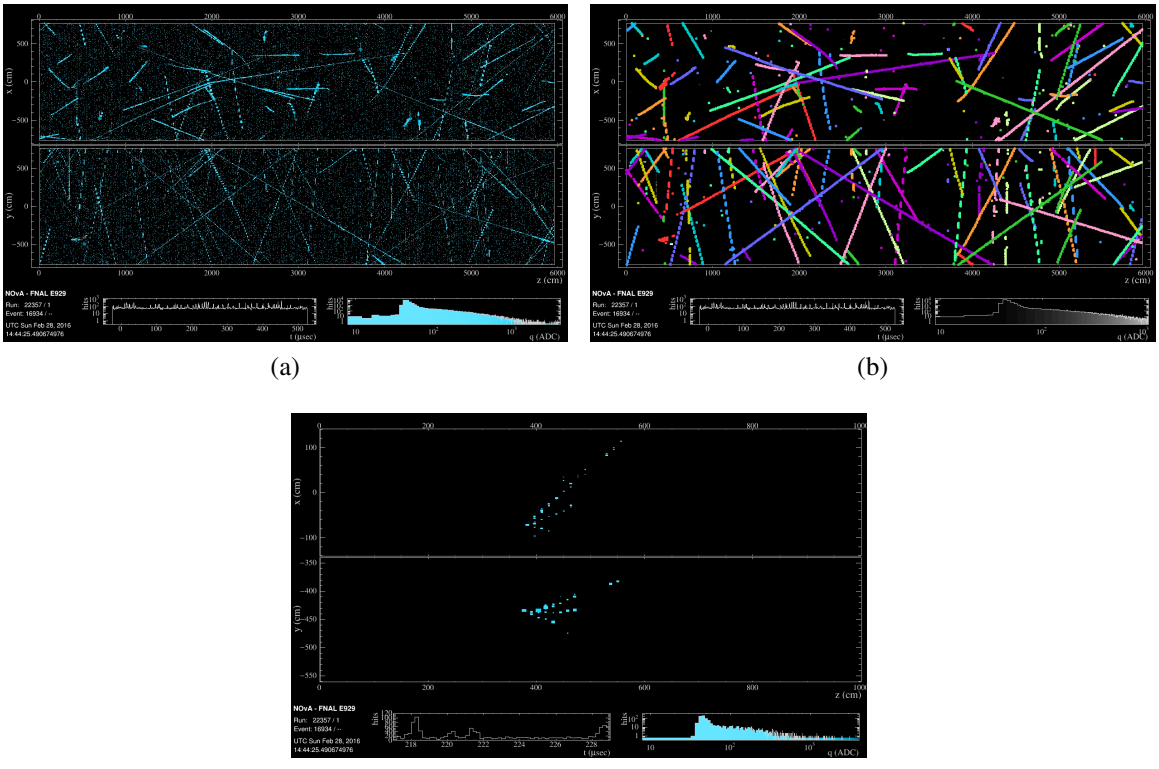


(b)

Figure 4.1: Trigger readouts from the NOvA (a) FD cosmic trigger and (b) ND beam spill with the beam entering from the left. Hit color corresponds to readout time. The top pane of each display shows the x dimension (left/right position from the vertical cells) and the bottom pane shows the y dimension (up/down position from the horizontal cells).

The particular slicing algorithm employed is called "TDSlicer", where TD stands for time-density. This method combines a density clustering algorithm from Rodriguez and

Liao [74] with Prim's minimum spanning tree [75]. For each hit, the density of the surrounding hits is found using the time difference and the spatial separation between hit pairs. Next, each hit is given an isolation score based on the distance to the nearest hit with a higher density. Slice centroids are then determined by plotting density versus isolation, with centroids being hits of both high density and isolation. From the centroids, slices are built up by adding nearby hits within a minimum distance from a hit in the slice. These algorithms are run separately for the x and y views, which are then combined into 4D slices using the average z-position and time values between possible pairs of slices between views. An example of the resulting slices is shown in Figure 4.2.



(a) (b)

(c)

Figure 4.2: A single FD readout window (a) before slicing with color showing deposited charge, (b) after slicing with color denoting hits belonging to the same slice (N.B. due to a limited color pallet the same color has been used to represent separate slices), and (c) a zoom in of the slice containing the neutrino interaction.

4.1.2 Vertex Identification

An accurate determination of the interaction location, or "vertex", is necessary as a starting point to make clusters representing single particles contained in a slice. If the reconstructed vertex is further forward than the actual one, then the resulting particle clusters would be shorter than they should be, and in the worst case a non-existent backward-going particle could be reconstructed. The result of the vertexing procedure, described below, is shown in Figure 4.3.

Vertex reconstruction begins with a 2D Hough transform with a modified voting scheme [76]. The Hough transform fits a line through each pair of points in one view (x or y) of a slice. Polar coordinates (ρ, θ) are used instead of rectangular (x/y, z) in order to parameterize vertical lines. Each line casts a vote in ρ-θ space with a smeared gaussian as described in Ref. [76] with the variance on ρ and θ determined by the detector resolution. The most frequent lines show up as peaks in ρ-θ space, and represent major features (particle paths) in the slice. The line with the greatest number of votes is taken as the first Hough line. The hits lying along that line are removed from consideration, and the next Hough line is found by repeating the voting process with the remaining hits.

Intersections of Hough lines are then used to seed an elastic arms-based vertex finder [77]. This algorithm searches for the optimal vertex by finding a set of lines, or arms, that describe the hits in the slice. The optimal arms, and therefore vertex, is the set that minimizes an energy cost function which is based on the distance and association factor between hits and arms [78]. The number of arms used is determined by how many lines

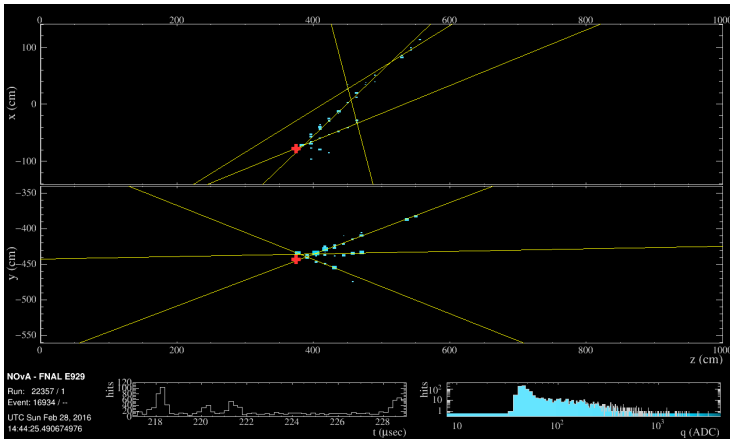


Figure 4.3: Result of vertex finding procedure showing the Hough lines (yellow) and vertex location (red cross). Uses the same event shown in Figure 4.2.

were found by the Hough algorithm. Arm directions and vertex candidates are seeded by the Hough results along with "minimally biased" seeds determined geometrically or from combinations of slice hits. Using the "simulated annealing" process described in Ref. [77], the "temperature" is slowly decreased, with the association of far-away hits to a given arm decreasing with temperature. This procedure helps to avoid local minima since long-range associations help to smooth the energy landscape and slowly decreasing these associations will then drive the fit to the global minimum, thus finding the optimal vertex location.

4.1.3 Particle Clustering

From the slice hits and the interaction vertex we can group hits into individual particle "prongs", which is a cluster of hits having a starting point and direction. This is done using a probabilistic clustering algorithm [79] based on the Fuzzy k-means technique [80]. The method, called fuzzyk by NOvA, allows hits to be members of more than one cluster, hence "fuzzy", and is probabilistic in that it does not require a hit to have a total membership probability of one. This probability feature allows for isolated hits to be treated as noise rather than being associated with a particle cluster.

Fuzzyk works by looking out from the vertex and sweeping around 360° to map out the angular hit distribution with an angular uncertainty determined as a function of the distance of a hit from the vertex. The angle with the greatest hit density is used as the first cluster center. Slice hits are associated with the cluster center as described in Ref. [78]. Based on the members hits, the cluster center location is updated and hits with less than 1%

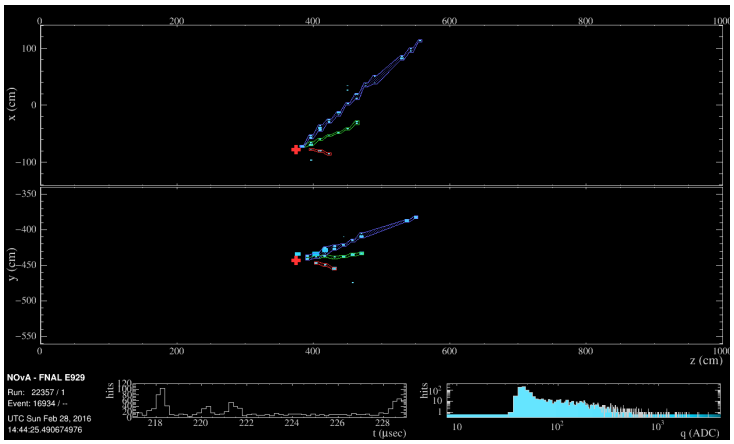


Figure 4.4: Result of the fuzzyk clustering procedure showing three prongs found in the event from Figure 4.2.

membership in that cluster are used to remap the angular density. The process is repeated with the next densest angle providing a new cluster center. This procedure is performed separately for each detector view and Kuiper's test [81] in the space of deposited energy versus z-position is used to view-match the prongs. Any prongs that cannot be matched remain as 2D prongs. The result of this algorithm is shown in Figure 4.4.

4.1.4 Particle Tracking

One step beyond the prongs produced by fuzzyk is a track which, in addition to having a starting point and a direction, also has a full reconstructed trajectory along the particle path. Where prongs are good for describing showering particles like electrons or hadrons, tracks provide more information for MIPs, such as muons, which travel more or less on straight paths through the detectors with some slight Coulomb scattering. NOvA has developed two tracking algorithms: Kalman tracker, and Break Point Fitter (BPF).

Kalman tracker has historically been the primary track reconstruction method. It employs a modified Kalman filter [82, 83, 84], which is a method of estimating some true value in discrete steps, here the position of the particle along its path, using a measurement at each step. Kalman filters assume that the current state of a system is linearly related to the previous state. The implementation of the Kalman tracker does not rely on the vertex or prongs produced by elastic arms and fuzzyk. Starting at the downstream end of the slice, a straight line through a pair of nearby hits is used to predict where the next upstream hit will be. If a hit is consistent with the prediction, allowing for Coulomb scattering, the it is added to the track and the prediction for the next hit is updated. This process is repeated until no more hits can be added. The Kalman tracker works on the detector views independently and merges them into a 3D track using a simple score based on the z-positions.

BPF was developed as an alternative tracking method which breaks the particle path at various points to allow for Coulomb scattering dependent on the particle mass and energy deposition rate [85]. Using a fuzzyk prong as the input, the radiation lengths of the traversed material are tabulated and the break points are placed either at a multiple of the radiation length or if the estimated scattering angle of the particle is large enough. From the scattering planes and the angle estimates, BPF then determines the full particle trajectory. Since energy deposition and scattering angles are dependent on the particle's mass, the procedure is performed three times, assuming that the particle is a muon, proton, and charged pion.

4.2 Particle Identification

In order to perform a physics analysis, it is necessary to know the type of particle being observed. Particles can be identified using various reconstructed quantities such as path length, interaction signature, or mass estimation. NOvA's particle identification algorithms use numerous inputs and then output a score, typically from zero or negative one to one, representing a sort of confidence the algorithm has in the particle identification[1]. Both the algorithm and the resulting scores are commonly referred to by the acronym "PID". To select particles of interest, various "cuts" are applied to the PIDs and to some basic reconstructed variables.

In the 3-flavor analysis, the PIDs must be able to separate between ν_e CC, ν_μ CC, NC, and other background events; examples of the first three types are shown in Figure 4.5. CC ν_μ events are most easily identified by a long straight muon track, although the charged pion from an NC event can mimic the topology of a low energy muon. In a similar manner, the two photons produced by a π^0 decay could also be mistaken for an electron if their opening angle is small enough that the showers overlap and the π^0 decay occurs close to the interaction vertex. Multiple PIDs have been developed to discriminate between these and other interaction types.

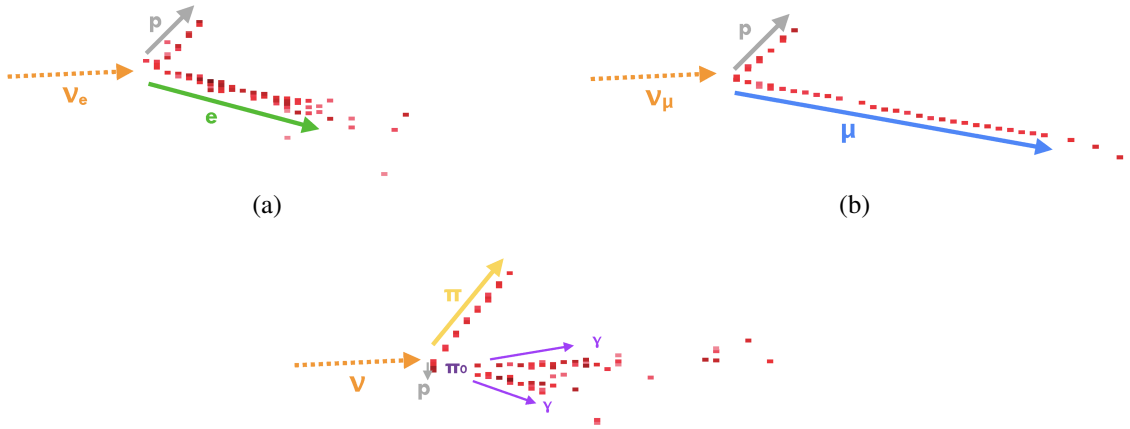


Figure 4.5: Three interaction topologies of interest for the 3-flavor oscillation analysis. (a) ν_e CC, (b) ν_μ CC, and (c) NC interaction with a π^0 detached from the vertex.

[1]A score of one means the algorithm is pretty sure that the particle is of a specific type and a score of zero (or negative one) means that the algorithm is pretty confident that it is not of that type.

4.2.1 Boosted Decision Trees

A decision tree is a machine learning technique that takes some inputs and makes binary decisions based on the values in order to predict some feature. The act of boosting creates multiple classifiers where the next classifier is trained to perform better on samples that the previous one mis-classified. In this way, boosted decision trees (BDTs) combine many weak trees into a much stronger one [86].

NOvA uses BDTs to identify muons based on the Kalman track and to reject background events from cosmic rays (from here on these will be referred to simply as "cosmics"). The reconstructed muon identifier (ReMID) only considers events with a reconstructed Kalman track, since the Kalman tracker was designed specifically for long, straight, muon-like tracks. This BDT uses four inputs:

- The difference of the log-likelihood that the energy deposition (dE/dx) is consistent with a muon versus a charged pion. Only MIP-like energy depositions are considered.
- The difference of the log-likelihood that the amount of scattering is consistent with a muon versus a pion.
- Track length.
- The fraction of planes along the track with MIP-like energy depositions.

The log-likelihood evaluations are done with respect to a set of simulated muons and pions. The MIP-like requirements in the first and last inputs help to avoid the impact of hadronic activity overlapping with the track. The last input in particular also adds to the discrimination between muons and charged pions because the strongly interacting pions could have non-MIP-like hadronic energy depositions along their tracks.

The specifics on cosmic rejection are discussed in Chapter 5, but the BDTs are just one method employed. Because we achieve cosmic rejection in other ways, the purpose of the BDTs are to catch cosmics that get past those other methods. Therefore, the BDTs are trained on samples that contain more signal-like cosmics with many of the easy-to-identify cosmics already removed. This selection is done using containment cuts (discussed in 5) and loose cuts on the signal PIDs (discussed below). The cosmic rejection BDTs were trained independently for ν_μ, $\bar{\nu}_\mu$, ν_e, and $\bar{\nu}_e$.

4.2.2 Convolutional Neural Networks

Convolutional neural networks (CNNs) are a type of machine learning algorithm that are frequently used in computer vision applications for image recognition and analysis [87]. This type of network, therefore, is a natural choice to identify particle types and interaction modes in a highly-segmented, pixel-based detector. CNNs work by taking in pixel maps, where the amplitude is the calibrated energy deposition, and convolving them with various filters in order to extract topological features. Image convolution can be represented by a matrix multiplication where, for each pixel, some matrix or "kernel", is multiplied by a matrix representing the surrounding pixels. This operation then creates a new pixel value that is some combination of the initial pixel with its neighbors. For example, if a 3×3 grid of pixels were being convolved at once, and with the pixel in question being the center one, the identity kernel would be:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

which, when multiplied by a matrix of pixels will leave the central one unchanged. In NOvA's application, the pixel amplitude is the calibrated energy deposited in each cell.

NOvA uses three different CNNs in the 3-flavor oscillation analysis. The first, CNN_{cos}, uses the ResNet18 architecture [88, 89] and acts as an early form of cosmic rejection. CNN_{cos} identifies time windows containing cosmic-like tracks which are filtered out from consideration, thus greatly reducing the amount of "uninteresting" data that needs to be processed around the spill window. A second network, CNN_{evt}, is built from a modified mobilenet v2 [90, 91] architecture and serves as the primary PID and is trained on slices to scores events as ν_μ CC, ν_e CC, NC, or cosmic. The final network, CNN_{png}, is also built on the mobilenet v2 architecture and identifies the particle type of a reconstructed prong with labels for electron, muon, proton, neutron, charged pion, neutral pion, photon, and other. In the 3-flavor analysis, CNN_{png} is used to identify particle types for energy estimation, discussed below.

CNN_{evt} and CNN_{cos} take in two pixel maps from the slice, one for each view, while CNN_{png} uses the slice pixel maps along with two maps containing only the prong of interest, for a total of four inputs. For CNN_{png}, long prongs were removed from consideration since these are almost always muons and are easily identified by ReMID. The training set for CNN_{evt} contained a limited number of cosmics so that it could learn to distinguish more

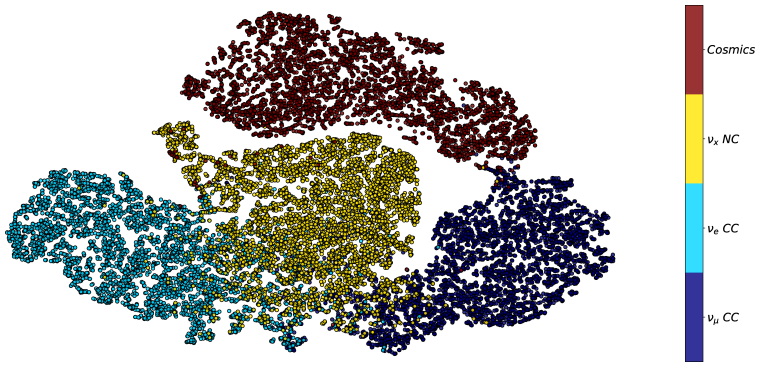


Figure 4.6: A t-SNE transformation [92] of the feature vector from the CNN_{evt} classifier trained on the neutrino beam mode sample with color representing interaction type. This shows good separation between the four interaction types with some overlap between NC events and both ν_e CC and ν_μ CC events.

between neutrino interactions since there are dedicated cosmic rejection algorithms in use. Additionally, CNN_{evt} was trained in a way to reduce the influence of calorimetric energy on classification decisions since the energy scale is a leading systematic uncertainty. This is done in the training step by multiplying all of the pixels in an event by a scale factor drawn from a gaussian distribution centered at 1 with a standard deviation of 0.1. By randomly modifying the energy depositions, the network learns to rely more on topology than on calorimetry. The performances of CNN_{evt} and CNN_{png} is shown in Figures 4.6 and 4.7.

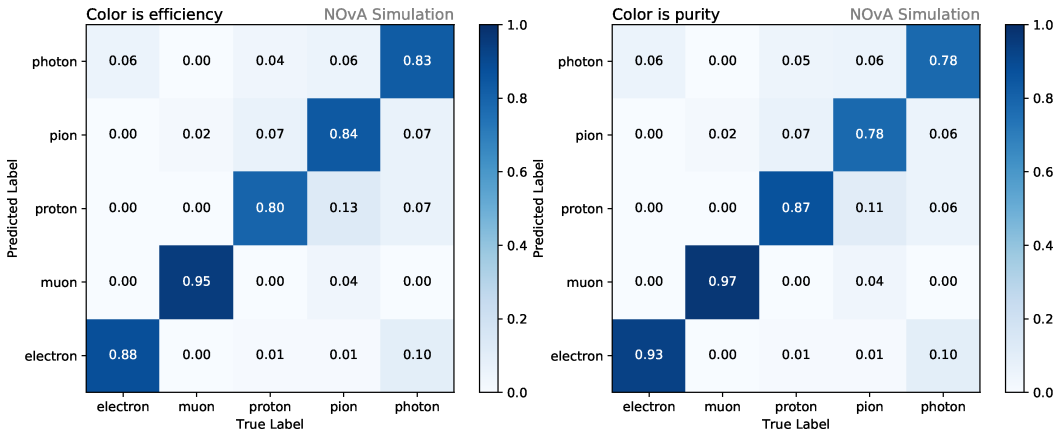


Figure 4.7: Performance evaluation matrices for the CNN_{png} network trained on the neutrino beam mode sample. The predicted label on the y-axis is the highest scoring label from the network. The diagonal shows the efficiency (left) and purity (right) of each particle type while the off-diagonal shows mis-classification and the background contamination.

CNN_{cos} uses pixel maps of the full detector over a 16 µs time window and runs on unprocessed files to provide an early cosmic tag. The time windows overlap each other by 1 µs to avoid particles being chopped up due to crossing a time boundary. Using unprocessed files means that CNN_{cos} has no access to any energy information and relies solely on event topology. If CNN_{cos} returns a cosmic score of 0.811 or higher, the time window is removed from consideration, thus reducing the size of the cosmic files that need to be processed. CNN_{cos} is also run on NuMI beam spill events, though in this case it just provides another data product which can be used in the analysis and time windows are not removed from consideration.

4.3 Energy Estimation

A precise and accurate estimate of the incoming neutrino energy is paramount to the measurement of the oscillation parameters, since the oscillation probability is directly dependent on the energy. Additionally, with neutrinos only interacting via the weak force, the only trace they leave in the detector are the tracks of the outgoing particles. Therefore, accurate neutrino energy estimation requires good reconstruction and accurate estimation of muon, electron, and hadronic energy deposits. To train an energy estimator, it is necessary to either know the energy of the incoming neutrino or to have a large sample of the various outgoing particle types with well-determined energies. The latter is hard to come by and NOvA is undertaking a test beam effort to compile these datasets. By instead training the energy estimators on simulation, we have access to the true neutrino energy. Of course, model dependencies must be understood and quantified when relying on MC, which will be discussed later.

4.3.1 Electron Neutrinos

Charged current electron neutrino events contain an outgoing electron shower along with a hadronic component, as seen in Figure 4.5a. The electromagnetic showers produced by an electron (as well as photons) result in a distinctly different detector response as compared to hadronic deposition due to the lack of strong interactions. This difference can be quantified by comparing the simulated detector response, the amount of deposited energy that produces detectable light, for EM-dominated events and hadron-dominated events. Doing so shows that the EM response is about 1.26 times greater than the hadronic

response [93].

Simply applying a response ratio to the hadronic events is not sufficient due to inefficiencies in the clustering and identification algorithms. Instead, a quadratic fit is performed, treating the EM and hadronic components separately:

$$E_\nu = A \cdot E_{EM} + B \cdot E_{had} + C \cdot E_{EM}^2 + D \cdot E_{had}^2 \tag{4.1}$$

where E_ν is the estimated neutrino energy, E_{EM} and E_{had} are the sum of the calibrated energy deposits of EM or hadronic prongs and A, B, C, and D are constants to be fitted for.

The identity of each prong is determined using the scores from CNN_{png}. A total EM score is calculated from the sum of the electron, photon, and π^0 scores determined by CNN_{png} and a total hadronic score is calculated similarly from the proton and charged pion scores. Prongs are identified as EM if their EM score is greater than their hadronic score. The EM energy, E_{EM}, is the sum of the calorimetric energy from all EM prongs, while E_{had} is taken to be the total calorimetric energy of the full slice minus E_{EM}.

The simulated beam flux is peaked at 2 GeV, therefore, the natural tendency of an energy estimator would be to drive the energy toward this peak since those events would represent the majority of the sample. To avoid this potential bias, the training sample was re-weighted into a flat flux as a function of energy. With the sample re-weighted, the quadratic fit is performed on the distribution shown in Figure 4.8.

While re-weighting the events to have a flat flux does reduce the energy resolution[2] slightly, this technique produces an energy estimation that is unbiased across most of the energy range of interest, 1–4 GeV, as shown in Figure 4.9a. The energy resolution is also relatively flat across this energy range, as shown in Figure 4.9b. The average energy resolution at the FD is 10.3% for the neutrino beam mode and 9.1% for antineutrino beam mode. Training separate energy estimators for neutrino and antineutrino beam modes is necessary due to differences in sample purity and interaction types that can occur.

[2]Energy resolution is defined here to be the RMS of the $(E_{reco} - E_{true})/E_{true}$ distribution where E_{reco} is the estimated energy and E_{true} is the true energy. The mean of this distribution quantifies the amount of bias in the energy estimator with a mean of zero being unbiased. The FD energy resolution has been evaluated on the oscillated sample composition at the 3-flavor best-fit point.

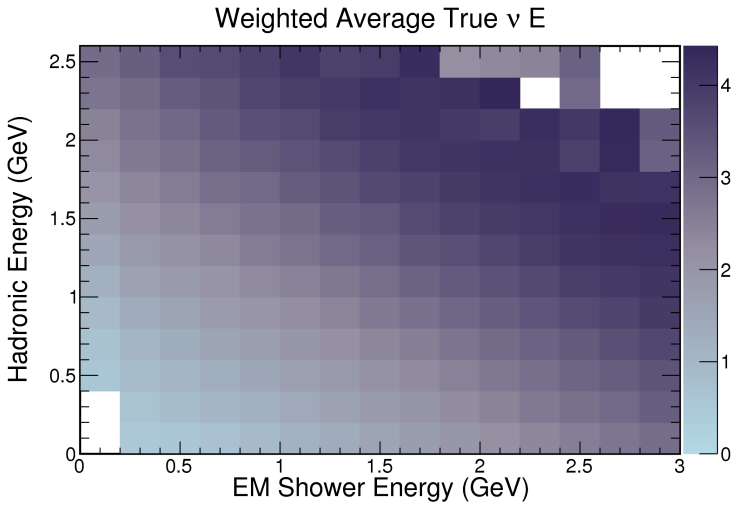
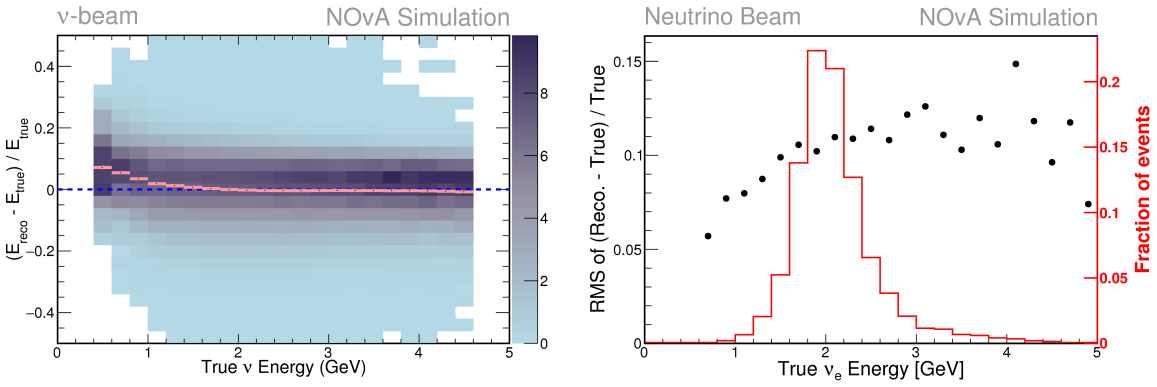


Figure 4.8: Distribution of E_{had} versus E_{EM} for simulated ν_e events from the neutrino beam mode where color denotes the average true neutrino energy re-weighted to a flat flux. Projecting this out of the page reveals a roughly quadratic surface.



(a) (b)

Figure 4.9: Results of the ν_e energy estimator. (a) 2D plot of energy resolution versus true neutrino energy where the pink line shows the mean in each bin. (b) RMS of the energy resolution as a function of true neutrino energy where the red histogram shows the simulated neutrino flux. Both plots show the FD energy estimator trained on the neutrino beam mode sample.

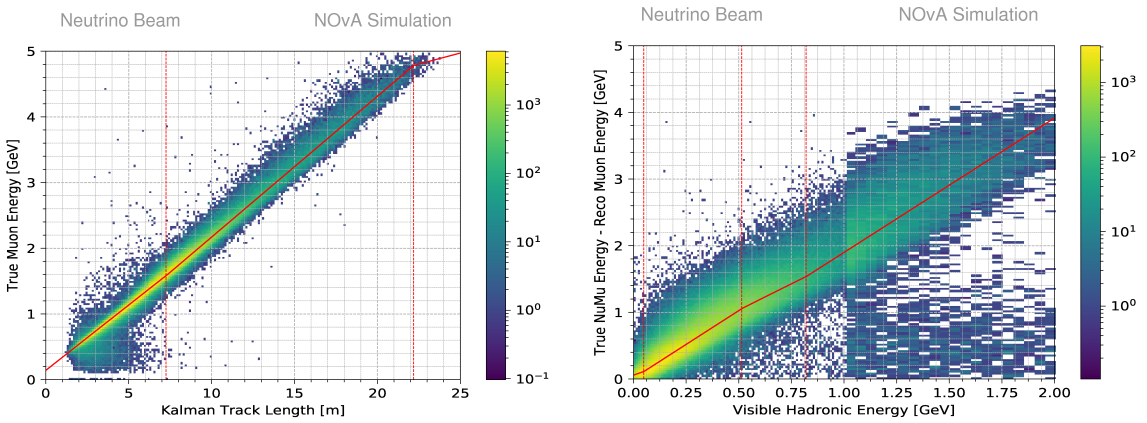
4.3.2 Muon Neutrinos

As discussed earlier muons between 0.2–2 GeV/c have a relatively constant energy deposition as MIPs. This means that the muon track length is highly correlated with its initial energy. Thus, ν_μ energy estimation is performed by first determining the muon contribution via its track length, then the energy of the hadronic system is added to sum up to the neutrino's energy:

$$E_\nu = E_\mu + E_{had} \tag{4.2}$$

In a ν_μ event, the muon track is found using ReMID. Then a piece-wise linear spline fit is performed in the space of reconstructed Kalman track length versus true muon energy drawn from simulation. As seen in Figure 4.10a, the spread in this distribution is relatively small for true muon energies above 1 GeV. For the spline fit, the segment slopes and intercepts along with the stitch points are all fitted using the gaussian mean of the distribution for each track length bin. This produces a muon energy resolution of about 3%.

A similar fitting procedure is performed to estimate the hadronic composition. Hadronic depositions are not clean like muon tracks, so the total calorimetric energy of the hadronic system is used instead of track length. Rather than utilize the hadronic prongs directly, the hadronic system is taken as all the hits in the slice that are not associated with the muon track. In this definition, the possible track overlap near the interaction vertex is taken into account when summing the calorimetric contributions. Since the goal is to estimate the neutrino energy and not the true hadronic energy, the y-axis in the fit is defined as the true neutrino energy minus the reconstructed muon energy as determined above. From Figure 4.10b the spread in the distribution is significant, leading to a hadronic energy resolution of about 26%, though combining this with the precise muon energy estimation results in an average energy resolution of 9.1% (8.2%) for FD neutrinos (antineutrinos).



(a) (b)

Figure 4.10: Results of the ν_μ energy estimator. (a) True muon energy versus Kalman track length distribution used to determine E_μ and (b) the remaining event energy vs visible (calorimetric) hadronic energy distribution used to determine E_{had}. The spline fits are shown as solid red lines and spline break points are shown as red dashed lines.

5. The 3-Flavor Analysis

As was mentioned a few times in earlier chapters, NOvA is making a measurement of three neutrino oscillation parameters. The ν_μ and $\bar{\nu}_\mu$ disappearance channel provides sensitivity to the θ_{23} mixing angle and the magnitude of the Δm^2_{32} mass splitting. By comparing the rate of ν_e and $\bar{\nu}_e$ appearance NOvA can access the CP violating phase, δ_{CP}, and also gain sensitivity in determining the octant of θ_{23}. What's more, by taking advantage of the MSW effect as neutrinos pass through the Earth, the ν_e and $\bar{\nu}_e$ samples provide the ability to probe the neutrino mass ordering (the sign of Δm^2_{32}).

To make these measurements means to fit simulation to data. With simulated neutrinos following the spectrum and composition of the source beam, we can apply the oscillation probabilities with a set of oscillation parameters to predict the energy spectrum of the ν_μ ($\bar{\nu}_\mu$) CC and ν_e ($\bar{\nu}_e$) CC events that we would expect to observe at the Far Detector. We then compare the predicted spectra to the observed data and adjust the applied oscillation parameters to find the best agreement. Up to now I've spoken about the appearance and disappearance samples separately, though in reality all of the samples are fit simultaneously to determine the oscillation parameters that best match all of the observed data.

The analysis can be broken down into a few major parts. First, the neutrinos candidates must be identified. Recall that we only care about CC events since these have an associated lepton which allows us to tag the incoming neutrino. We need to be able to choose the CC neutrino signal events while removing backgrounds such as NC interactions or particles created by cosmic rays. Next, the events are split into subsamples to improve the sensitivity of the oscillation fit. Background events still infiltrate our samples, so we split them down into samples with varying signal-to-noise ratios, allowing "clean" data to have more influence on the fit. Ideally, we would toss out samples that are highly contaminated, but neutrinos do not interact very often, so even with a high intensity beam NOvA is still limited by the sample statistics. With two functionally-identical detectors, NOvA uses the ND samples to correct the simulated events and "extrapolate" the ND observation to provide a better prediction of the FD samples. This procedure greatly reduces the impact of the largest systematic uncertainties like cross-section modeling and beam flux simulation. Finally, the FD predictions are compared to the data and a fit is performed to determine which set of oscillation parameters best matches the observed neutrino spectra.

5.1 Event Selection

First, let's get some jargon out of the way; event selections are also commonly referred to as "selection cuts" or just "cuts" for short. To narrow down the events of interest, cuts are applied on various quantities, from basic detector information to high level reconstructed values or even external factors like beam conditions. For example, a simple cut may be something like: `minx > 10` (i.e. you only want to consider events where the minimum x position of all the hits is greater than 10). Typically, many such cuts are combined to perform some specific function and these are then further combined with other cuts to select samples with a high signal-to-noise ratio.

In NOvA, the event selection is split into four main functions:

- Quality cuts, which make sure that, among other things, the detector was working properly and the beam was being delivered as expected.
- Containment, which ensures that the events we consider do not lose energy due to particles exiting the detector, and further reduces background from particles that interact outside of the detector.
- Cosmic rejection and rock veto to remove the high rate of cosmogenic particles present at the Far Detector and events originating in the surrounding rock at the Near Detector.
- Charged current neutrino selection tuned for either ν_e ($\bar{\nu}_e$) or ν_μ ($\bar{\nu}_\mu$) CC topologies.

5.1.1 Quality Assurance

The quality cuts are all relatively straightforward, though there is a slight difference between the quality cuts used for the ν_μ and ν_e[1] selections in both beam modes. For both selections we remove events that occurred when the detectors were overly noisy or if large sections of the detectors were not reading out. Additionally, events occurring when the delivered beam is out of spec are also rejected. The various handles we use for the beam quality are: current and polarity in the magnetic focusing horns; beam intensity, measured in protons on target (POT); and beam width in the two transverse dimensions. We also have

[1] I will forego explicitly writing ν_μ ($\bar{\nu}_\mu$) or ν_e ($\bar{\nu}_e$) when there is no need to distinguish between neutrino and antineutrino beam modes. In these instances I will simply use the un-barred form to mean both neutrino and antineutrino cases.

the ability to manually mark runs as "bad", which we do if a monitoring variable is not reporting properly or if known work is occurring which is not caught by the other quality cuts.

In the ν_μ selections we limit the maximum reconstructed energy to 5 GeV, since events beyond this range do not contribute significantly to the oscillation measurement. As a simple first step for cosmic rejection, events that are confined to the transverse dimensions are removed since muons are highly penetrating and muons created by a beam neutrino will cross many planes in the longitudinal direction. Finally, we require that there be more than twenty total hits in an event and that the muon-related tracking and PID algorithms were able to run on the event.

The philosophy for the ν_e quality cuts are similar to the ν_μ; however, here we make sure that an event vertex was identified and that the fuzzyk clustering algorithm successfully found at least one particle candidate. In addition, we apply a cut on the maximum number of hits per plane that were recorded to remove "flasher events". Recall that a single FEB reads out 16 cells with two APD pixels per cell. The pixels all share a common electrical ground and when a large amount of light is collected by a single pixel, it can cause the common ground to drop lower than its typical level. As the ground quickly recovers, the threshold DCS signature mimics a real particle and a hit is recorded for the adjacent pixels on that APD. The resulting topology is problematic for ν_e events since these can produce wide particle showers which also cover many pixels on a single plane. Therefore, if more that eight hits per plane were recorded in an event, then it is rejected from consideration as a ν_e candidate.

5.1.2 Containment Criteria

Neutrinos, due to their lack of electric charge, leave no tracks within our detectors. This is inconvenient and requires us to rely solely on hits generated by the outgoing particles from the interaction in order to estimate the energy of the incoming neutrino. This means that it is imperative that nearly all of the energy carried away by the outgoing particles is deposited within the detector. To that end, we use simple cuts on basic reconstructed quantities to ensure that we only consider fully contained events. Additionally, background events generated by cosmic rays or neutrinos otherwise interacting outside of our detectors can also be removed using these same simple cuts.

Four metrics were considered in order to tune the containment cuts: ratio of signal to

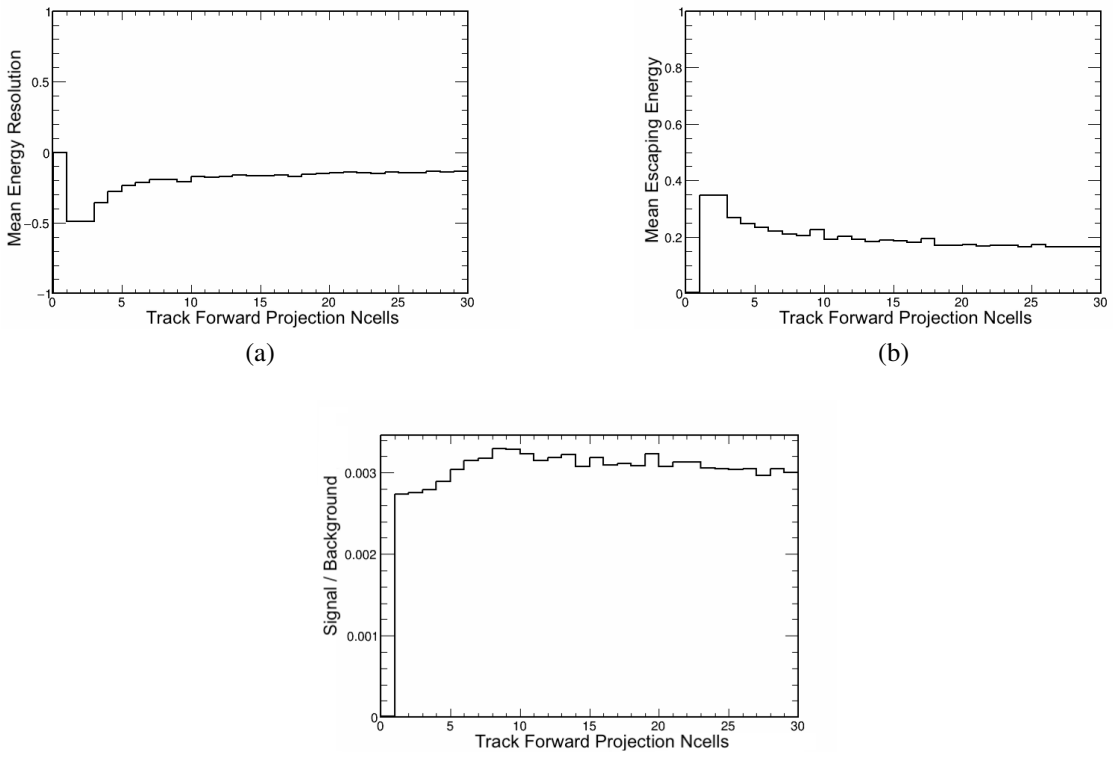


Figure 5.1: Plots of (a) mean energy resolution, (b) mean escaping energy, and (c) signal divided by background versus the number of cells between muon-like tracks and the edges of the far detector. These plots were used to determine the optimal cut point of > 5 for this reconstructed quantity.

background, energy resolution, the proportion of event energy that escapes the detector, and, in the ND only, the ratio of the event rates from data versus simulation. Each of the tuning metrics will change rapidly as you move from detector regions with high background to those further away from the walls. Optimum cut values are found where the metrics level off, indicating that many of the uncontained events have been rejected. An example of these metrics is shown in Figure 5.1 for the ν_μ FD sample. A cut of: `projected Ncells > 5` was found to be optimal for the quantity shown in that figure.

In the FD, the main containment criteria are based on the distances of all hits in an event from each detector wall; the result being that the active volume is 7% smaller than the total detector volume. Additional FD containment cuts are made for the ν_μ samples based on the number of cells pass through before reaching the detector walls as projected from the ends of the most muon-like track. This set of cuts slightly reduces the active volume

further, but these cuts are evaluated on an event-by-event basis so an exact active volume cannot be determined.

The containment criteria at the ND also employ a basic set of cuts on the extent of events in each direction, although the situation is slightly more complicated due to the muon catcher at the downstream end. For the ν_e samples, these cuts reduce the active volume of the detector by about 38%, while it's only reduced by about 16% for the ν_μ samples. The ν_μ samples again utilize the track-projected cells to verify that the muon component is far enough from the detector edges. There is also a requirement that at most only one muon-like track is present within the muon catcher. To round out the containment criteria, the ND ν_e samples have a final set of cuts on the position of the reconstructed interaction vertex. These vertex cuts are slightly harsher than the active volume cuts and ensure that the events originate within the detector volume.

5.1.3 ND Rock Veto and FD Cosmic Rejection

Particles entering the detector can appear to be the CC neutrino interaction signals we're searching for. A cosmic ray muon or an ND beam neutrino which interacts in the surrounding rock may look like ν_μ CC events. These muons may also undergo bremsstrahlung radiation or decay-in-flight and thus produce electrons which could be confused with ν_e CC events. Generally, these types of backgrounds will have hits which occur quite close to the edge of the detector and will therefore usually be caught by the containment cuts. However, there are scenarios where hits may not be recorded; for example, if the entering particle travels along the dead PVC material for some distance before passing through the scintillator or the produced light may just have a low fluctuation and not reach the detection threshold. Additionally, processing these frequent and uninteresting events (at least uninteresting with regards to the 3-flavor oscillation analysis) is computationally time consuming, so the earlier they can be removed from consideration the better.

Rock muons at the ND are rejected early in the file production process to reduce processing time. Recall that many of the reconstruction algorithms build on each other by taking inputs from a lower level of reconstruction. The ND rock filter utilizes a simple set of cuts on the vertex position produced by elastic arms. Specifically, the reconstructed vertex is required to be >21 cm from each edge of the detector. This cut removes $\sim$ 58% of the total event slices while retaining 99% of the interesting ν_μ events. The reduction in slices equates to 40% or more decrease in CPU time, providing a for a faster analysis turn-around.

Along similar lines, a cosmic veto was also developed for the FD. NOvA records large amounts of cosmic data using a 10 Hz trigger and uses these events to provide the shape of the cosmic ray background spectra for the ν_μ and ν_e CC samples. Many of these, cosmics are easy to remove from consideration using the containment criteria, with only a small fraction actually imitating our signal events. Therefore, CNN_{cos} filter, as described in Section 4.2.2, was developed to reduce the amount of data the needs to be processed for the background measurement. CNN_{cos} also provides cosmic rejection for in-spill events where it removes 95% of cosmics while retaining 99% of relevant signal events.

To complement CNN_{cos} and the containment cuts, we also employ several cosmic boosted decision trees, BDT_{cos}. While CNN_{cos} works well to remove a vast majority of cosmic backgrounds, retaining even 5% of these backgrounds result in $O(10^5)$ cosmic events in each samples. These events will also be much more signal-like since they were able to evade CNN_{cos}. To identify signal-like cosmics, the BDTs were trained on events that first passed a loose set of PID cuts based on the output of CNN_{evt} (and ReMID for the ν_μ samples). In total, six separate cosmic BDTs, one each for ν_μ and $\bar{\nu}_\mu$, and an additional two each for ν_e and $\bar{\nu}_e$.

Rather than considering full event slices, BDT_{cos} for the ν_μ samples assess the most muon-like track in the event. The input variables for each track were:

- Cosine of the track angle w.r.t. the beam direction.
- Cosine of the track angle w.r.t. vertical.
- The highest vertical position.
- Track length.
- Distance of closest approach to each side of the detector.
- Ratio of the number track hits to total hits in the slice.
- Reconstructed transverse momentum divided by total momentum.

After optimization using the typical figure-of-merit (FOM), $\frac{S}{\sqrt{S+B}}$, where S is the number of signal events and B is the number of background events, a cut value of 0.45 was found for both ν_μ and $\bar{\nu}_\mu$ samples. By itself, this cut on BDT_{cos} preserves >95% of the signal while removing ~70% of the background.

For the ν_e samples, four separate cosmic BDTs were trained to classify full event slices. As will be discussed later, we increase the statistics of the ν_e samples by including a set of events that fail either cosmic rejection or the containment criteria. This "peripheral" sample will, therefore, have a significant amount of cosmic contamination and require a separately trained BDT_{cos} which can handle signal events which exit the detector. The training methodology for core and peripheral BDT_{cos} are the same as we used for the ν_μ BDTs, with the major change being the input variables. The core BDTs use the following reconstructed values:

- Number of hits in the event.
- Distance of closest approach for any prong to each side of the detector.
- Width of the EM shower in the event.
- Direction of the EM shower development determined by the relative hit density between the upstream and downstream ends of the slice.
- The percentage of total slice energy carried by the EM shower.
- Reconstructed transverse momentum divided by total momentum.

while the peripheral BDTs use a smaller set of variables:

- Reconstructed momentum in the horizontal transverse direction divided by total momentum.
- Reconstructed momentum in the vertical transverse direction divided by total momentum.
- Distance from the start or end position of any prong to the top of the detector.
- Event vertex location.

Together, CNN_{cos} and BDT_{cos} reduce the cosmic contamination in the selected samples to $< 5\%$, a total reduction of 6 orders of magnitude. However, the containment criteria and PID algorithms work in conjunction with the cosmic rejection to reduce the impact of cosmic backgrounds even further.

5.1.4 Electron Neutrinos

The number of ν_e events is quite low in the FD. The beam begins as primarily ν_μ or $\bar{\nu}_\mu$ and by the time they reach the FD they have mostly oscillated away at the first oscillation maximum. However, at this location the majority of ν_μ have turned into ν_τ rather than ν_e which we can observe. The result being that we cannot just use the harshest possible cuts on the CNN_{evt} classifier to increase the signal-to-background ratio. Doing so would cause systematic uncertainties to be so dominant that no significant determination of mass ordering or δ_{CP} could be made. Instead, we do two things: (1) separate the sample of contained events into high and low PID regions, and (2) include another sample of uncontained events as a rate-only addition to the oscillation fit.

The "core" ν_e ($\bar{\nu}_e$) CC sample is made up of events which pass the containment criteria, CNN_{cos}, and a cut on the core-trained BDT_{cos}. In addition, we also implement PID cuts on CNN_{evt} to ensure we are choosing ν_e-like events. Typically, one would implement a single PID cut which maximizes the FOM $\frac{S}{\sqrt{S+B}}$, however, due to the low statistics, we also introduce another PID cut and include those events as another sample. The high PID cut is fairly tight at a CNN_{evt} score of ≥ 0.97 for each beam mode and produces a FOM of 4.69 (2.91) and purity of 79% (69%). The low PID cut accepts events between $0.84\ (0.85) \leq \text{CNN}_{\text{evt}} < 0.97$, giving a FOM of 1.87 (0.92) and purity of 51% (36%). If we were to simply use the looser cut to increase the statistics of ν_e samples then we would have slightly higher total FOM, however, the sample purity would only be 68% (58%). Instead, we separate the two PID samples so that the high purity one can maintain a significant pull on the fit without being watered down by the low purity bin while still gaining from the increased statistics.

To further bolster the statistics, we also include a "peripheral" sample. Events in this sample must first fail all of the core sample requirements. Therefore, these events might not be fully contained, and will be more background-like in topology. For these reasons we include this sample as a single bin statistics-only sample (the energy spectrum is not considered, though the reconstructed energy must be between 0–4.5 GeV). Because these events are more background-like, we employ a different BDT_{cos} which was trained specifically for this sample and employ a set of box cuts. The first cut is at a CNN_{evt} score or ≥ 0.97, just like the high PID core sample, and a BDT_{cos} score of > 0.60 (0.61). We gain even more events by loosening the BDT_{cos} cut to 0.56 (0.57) while tightening the CNN_{evt} cut to ≥ 0.995. With these two cuts combined the ν_e ($\bar{\nu}_e$) CC peripheral sample has a

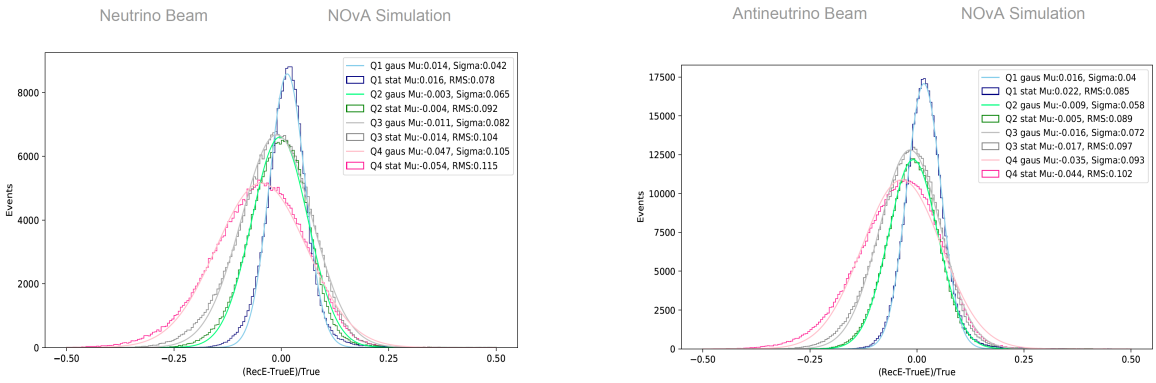
purity of 57% (43%), and by including this sample we increase the FD selection efficiency from 54% (64%) with just the core events up to 63% (75%).

5.1.5 Muon Neutrinos

For the ν_μ ($\bar{\nu}_\mu$) CC samples, we again employ $\mathrm{CNN_{evt}}$ as the main PID algorithm along with the ReMID BDT, which can provide a slight boost in sample purity. These PIDs are combined with $\mathrm{BDT_{cos}}$ for cosmic rejection. Cuts on these three classifiers were optimized simultaneously using the FOM described above. Additionally, since the power of the ν_μ samples lies primarily in the "dip" region of the energy spectrum at the oscillation maximum, the FOM for events between 1–2 GeV was also evaluated. The determined cut values were $\mathrm{BDT_{cos}} > 0.45$, ReMID > 0.3, and $\mathrm{CNN_{evt}} > 0.8$ for both neutrino and antineutrino mode. This produces samples that are 96% (98%) pure with a selection efficiency of about 33%.

Two techniques are used to increase the sensitivity of the oscillation fit with the ν_μ samples: (1) bin width optimization and (2) splitting into subsamples. Owing to the importance of the dip region, we employ variable bin widths for the ν_μ energy spectra with fine binning in the dip and coarser binning elsewhere. Along the same lines as the ν_e samples, we split the ν_μ events into smaller subsamples to isolate "good" events from "bad" events in the oscillation fit. Instead of using high and low PID samples the ν_μ selections are split based on the fraction of the neutrino energy that is carried away by the hadronic system, $\frac{E_{had}}{E_\nu}$, also referred to as hadronic energy fraction. By the nature of particle interactions in the NOvA detectors, the muon component of ν_μ CC events is much cleaner than the hadronic system and therefore, an estimation of E_μ will be much more accurate than E_{had}. Thus, events which have a higher hadronic energy fraction will have a poorer neutrino energy resolution. Hadronic energy fraction also correlates closely with background contamination from both NC neutrino interactions and cosmogenic particles.

To utilize the power of the hadronic energy fraction, events are split into multiple equal population quantiles. Each quantile serves as a separate sample in the oscillation fit, similar to the ν_e PID subsamples. The number of quantiles was optimized by checking the change in the sensitivity of the oscillation fit to measure $\sin^2\theta_{23}$ and Δm^2_{32}. It was found that the sensitivity improvement leveled off at four quartiles. The energy resolution of the standard spline-based energy estimator is shown in Figure 5.2 for each quartile in both neutrino and antineutrino beam modes.



(a) (b)

Figure 5.2: Energy resolution plots for the ν_μ spline energy estimator. Plots show each hadronic energy quartile for (a) neutrino beam-mode and (b) antineutrino beam-mode. The solid lines denote Gaussian fits to the resolution histograms.

5.2 Data-driven corrections

The NOvA analysis, like many particle physics experiments, relies heavily on a comparison of the recorded data to Monte Carlo simulation. As discussed in Section 3.5, the simulations require some amount of extra tuning to improve data-MC agreement. To recap, we implement PPFX corrections to the upstream beam simulations, adjust the poorly understood MEC interaction components, and tune the FSI model parameters to external π^+-on-^{12}C scattering data. The MEC tune brings the ND ν_μ simulation into close agreement with the observed data by attributing most of the difference to MEC interactions. After tuning, some data-MC disagreement still remains for ν_μ events and the ν_e simulation benefits very little. Therefore, we employ data-driven corrections to the ND simulation to account for any residual mis-modeling. To apply these corrections, the spectrum of the ND samples are decomposed into the constituent neutrino flavors (i.e. ν_μ, $\bar{\nu}_\mu$, ν_e, and $\bar{\nu}_e$ CC along with NC) and the corrections are applied on a component-by-component basis.

The simplest correction is performed on the $\bar{\nu}_e$ CC selection. In this case, all neutrino components are scaled up in a proportional manner:

$$N_X^{corr.} = \frac{N_X^{sim.}}{N_{total}^{sim.}} * N^{data} \tag{5.1}$$

where N is the number of events in a bin, X denotes any of the neutrino components, $corr.$

is short for corrected, and *sim.* is short for simulated.

The ν_μ and $\bar{\nu}_\mu$ ND selections are corrected in almost the same manner as the $\bar{\nu}_e$ selections, though here we only modify the ν_μ and $\bar{\nu}_\mu$ components, while the portion of ν_e, $\bar{\nu}_e$, and NC events are untouched from the simulation. The non-muon neutrino components are subtracted from the data and the two ν_μ components are scaled up to agree with that value on a bin-by-bin basis while maintaining their relative proportion. Expressed mathematically we have:

$$N^{corr.}_{\nu_\mu} = \frac{N^{sim.}_{\nu_\mu}}{N^{sim.}_{\nu_\mu} + N^{sim.}_{\bar{\nu}_\mu}} \times \left(N^{data}_{all} - N^{sim.}_{\nu_e,\bar{\nu}_e,NC} \right) \tag{5.2}$$

The $\bar{\nu}_\mu$ component is corrected in a similar manner with $N^{sim.}_{\nu_\mu} \rightarrow N^{sim.}_{\bar{\nu}_\mu}$ in the numerator.

The final selection, ν_e CC, has the most complex correction scheme and uses two techniques to adjust the various neutrino components. The ν_e component is adjusted by measuring the production of ν_μ from pions and kaons (both of which may also decay into ν_e) and the subsequent muon decay. Since this method corrects the beam electron-neutrinos, it is referred to as the BEN correction. The other method uses the measurement of Michel electrons produced by muons from ν_μ CC interactions or the hadronic shower from NC events.

The BEN correction takes advantage of the very well understood pion and kaon decays and adjusts the simulated rate of each to get better data-MC agreement. At low energies, the beam ν_e component originates primarily as a pion which either directly decays to a positron and ν_e or decays into a muon, which then decays to an electron. Starting around 4 GeV kaons begin to contribute a significant portion of the ν_e. The low energy, pion-dominated sample is made up of fully contained events that pass the ν_μ analysis selection criteria. The simulated ν_μ events originating as π^+ is then scaled to match the observed data with the background and non-pion ν_μ events removed. The resulting weight is then propagated to a scale factor for the rate of π^+ in the beam as a function of transverse and longitudinal momentum. Finally, that weight is then applied to all simulated ν_e CC events that came from π^+ decay. The kaon sample proceeds in a similar way after the pion scaling, with the main difference being that uncontained events are also included due to the high energy range of the kaon to ν_e contribution.

It should be possible to apply a BEN-type correction to the $\bar{\nu}_e$ selection as well, however, the antineutrino beam-mode has a much greater amount of "wrong-sign" contamination (i.e. neutrinos in the antineutrino beam or vice versa) particularly at higher energies in the kaon region (see Figure 3.11). In the neutrino-mode beam, the $\bar{\nu}_e$ component can be pretty

safely ignored since it so much smaller than the ν_e component. A proper treatment in the antineutrino beam-mode will require accounting for both ν_e and $\bar{\nu}_e$ contributions from kaon decays in order to scale each component.

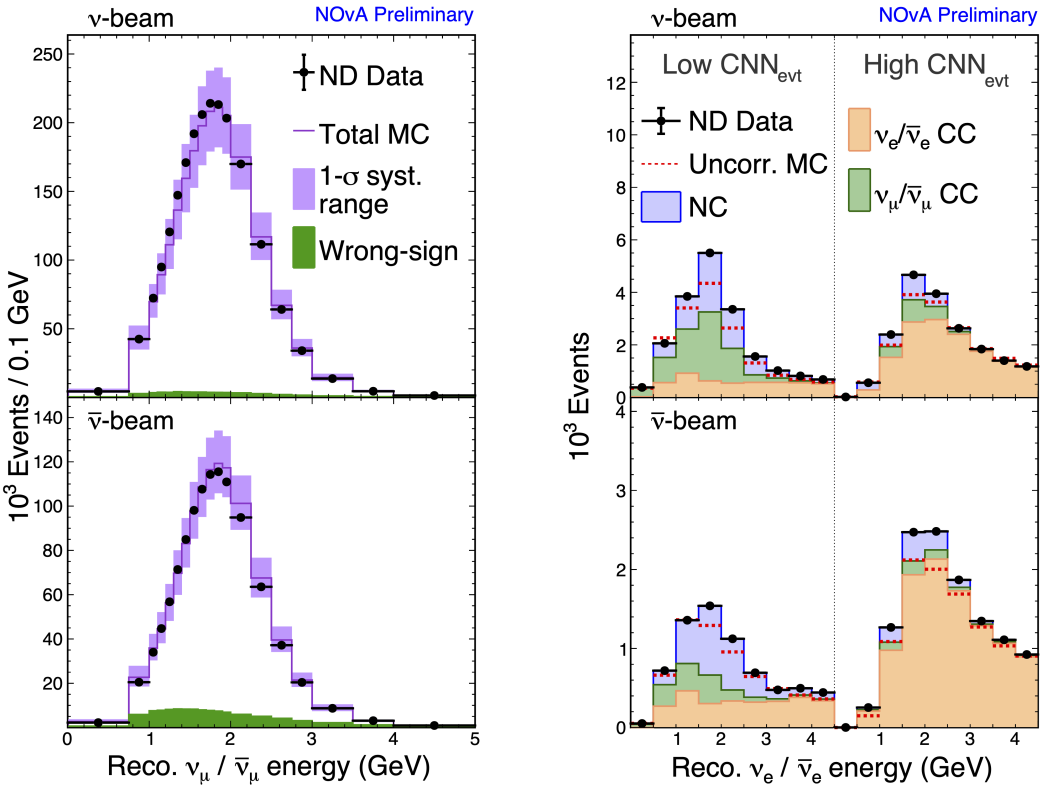
The final step is to use the Michel electrons to adjust the rate of ν_μ and NC events in the ν_e CC selection. Michel electrons are simply the electrons produced when a muon decays which were extensively studied by Louis Michel [94]. More specifically, we use the term "Michel electron" to refer to those produced after a muon as lost most of its energy. Michel electrons are thus a delayed signal, occurring hundreds of μs after the neutrino interaction, and can be produced either in ν_μ CC events or the hadronic system from NC or ν_e events where a pion decays into a muon which then decays to produce an electron (and a couple of neutrinos). Events containing a Michel electron are found using a simple log-likelihood identifier which uses calorimetric energy and number of hits along with the spatial and temporal distance of a suspected Michel cluster from the parent event slice.

With amount of ν_e events adjusted by the BEN correction, the only source of Michel electrons is from ν_μ and NC events, which are scaled to match the number of events in the Michel data sample. The event scaling relies on a template fit to the ν_μ component and therefore only works in analysis bins containing at least 20% ν_μ events. In all other bins, the proportional correction procedure used for the $\bar{\nu}_e$ selection is employed instead.

5.3 Near-to-Far Extrapolation

Owing to the event rate, data-driven corrections are only performed at the ND. However, by extrapolating the corrected samples from the ND, we can produce an improved prediction of the FD spectra. Not only does this generate data-corrected FD predictions, it also reduces the impact of systematic uncertainties that are correlated between the detectors. The most significant systematic uncertainties are related to the neutrino cross section model and the simulated beam flux. Luckily for NOvA, the two detectors are nearly identical in layout, material, and electronics, which means that any cross section modeling errors in the ND will be very similar to those in the FD and, of course, both detectors are sampling the same neutrino beam.

The ND selection criteria yield the spectra shown in Figure 5.3. We can split each FD sample into backgrounds and signal, using different extrapolation techniques for each. The ν_μ ($\bar{\nu}_\mu$) CC survival signal is composed of both right and wrong sign events; that is, $\nu_\mu \rightarrow \nu_\mu$ and $\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu$. Both neutrinos and antineutrinos are treated as signal here because,



(a) (b)

Figure 5.3: ND reconstructed energy spectra for (a) ν_μ CC and (b) ν_e CC with the neutrino beam-mode on the top and antineutrino on the bottom. The ν_μ MC histograms do not have the simple data-driven corrections applied to show the good agreement after the cross section tuning is performed. Hadronic energy fraction quartiles have been combined in the ν_μ plots. The red dotted line in the ν_e spectra show the total simulation before data-driven corrections and the solid histograms are after.

as we saw in Chapter 2, the disappearance probability is largely independent of δ_{CP}. These signals are extrapolated from the ND ν_μ CC selected events. The beam-based backgrounds for the FD ν_μ samples are quite minor, particularly when compared to cosmic backgrounds. Thus, we make no attempt to extrapolate them from the ND and instead directly use the FD simulation.

In contrast to the ν_μ samples, the ν_e appearance signal, ν_μ $(\bar{\nu}_\mu) \rightarrow \nu_e$ $(\bar{\nu}_e)$, must distinguish between neutrinos and antineutrinos owing to the significant δ_{CP} dependence. The ν_e and $\bar{\nu}_e$ appearance components are extrapolated from the ND ν_μ CC events, with the wrong sign appearance being treated as a background in the oscillation fit. The ν_e

samples have two further types of backgrounds: major and minor in impact. The major backgrounds are survival events from the intrinsic beam ν_e and ν_μ particles along with NC events. Major backgrounds must also be extrapolated from the ND to carry through the data-driven corrections applied to each of these components. These backgrounds are extrapolated using the ND ν_e CC selected events from Figure 5.3b. Minor backgrounds to the ν_e appearance samples are: $\nu_e \rightarrow \nu_\mu$, $\nu_e \rightarrow \nu_\tau$, $\nu_\mu \rightarrow \nu_\tau$, $\bar{\nu}_e \rightarrow \bar{\nu}_\mu$, $\bar{\nu}_e \rightarrow \bar{\nu}_\tau$, and $\bar{\nu}_\mu \rightarrow \bar{\nu}_\tau$ along with $\bar{\nu}_e \rightarrow \bar{\nu}_e$ and $\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu$. The last two components are only treated as minor backgrounds in the ν_e appearance samples. Again, the minor backgrounds are taken directly from the FD simulation.

For the signal samples, the first step of extrapolation is to transform the simulated ND ν_μ ($\bar{\nu}_\mu$) CC selected events from reconstructed neutrino energy into true neutrino energy. To do this, we apply the data-driven correction weights to each bin of reconstructed energy and sum up the number of events in bins of true energy. This produces corrected distributions of ND events as a function of true neutrino energy. The ratio of the corrected to uncorrected ND true-energy distributions provides a weight that is applied to the FD simulation to produce a data-corrected prediction of the FD samples as a function of true neutrino energy. The simulated FD events can have neutrino oscillations applied and, when re-weighted by the ND distributions, produce a prediction of the FD spectrum as a function of true neutrino energy. Finally, the FD prediction is transformed back into reconstructed energy and in this way we can generate predictions of the FD signals at any combination of oscillation parameters.

The major backgrounds for the FD ν_e ($\bar{\nu}_e$) samples are extrapolated from the simulated ND ν_e ($\bar{\nu}_e$) CC selected events. In this case, however, we do not transform the distributions into true neutrino energy because the ν_μ ($\bar{\nu}_\mu$) and NC events, which make up a significant portion of these samples, were mis-identified as ν_e ($\bar{\nu}_e$) events. Therefore, the energy estimation technique cannot be expected to perform well for many of the events in this sample and a conversion to true neutrino energy will not be accurate. Instead, we simply use the ratio of corrected to uncorrected ND event distributions as a function of reconstructed energy. Oscillations are applied to each component of the FD simulation and the ND ratio is applied to produce a prediction of the FD background in bins of reconstructed energy. Owing to the larger proportion of wrong-sign events in the antineutrino beam-mode, the neutrino and antineutrino components are fully extrapolated, while in the neutrino beam-mode the $\bar{\nu}_e \rightarrow \bar{\nu}_e$ and $\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu$ events are treated as minor backgrounds. We can further improve the ability of the extrapolation to constrain systematic uncertainties by isolating

ND subsamples that more closely correspond to FD samples. For example, the low energy bins in the ND are more useful than high energy bins in predicting the content of low energy bins of the FD spectra. Likewise, ND RES events can better constrain uncertainties on similar events in the FD. To take advantage of this fact, we expand the extrapolation procedure along two other dimensions to compliment the neutrino energy bins. In particular, we utilize the aforementioned hadronic energy fraction bins in addition to bins of transverse lepton momentum, p_{T}.

Hadronic energy fraction bins not only separate ν_μ interactions based on energy resolution and background contamination, each quartile also contains a different makeup of interaction modes. The first quartile, with most of the neutrino's energy going to the muon, is dominated by QE events along with a good amount of MEC interactions. The third quartile is almost entirely made up of RES interactions with only trace amounts of anything else. Similarly, the fourth quartile has a lot of RES but also contains nearly as much DIS events. Finally, the second quartile splits the difference between the first and third with nearly equal portions of MEC and QE and slightly more RES. Extrapolation using these hadronic energy fraction quartiles significantly reduces cross sectional uncertainties by grouping interaction modes into separate bins. This extrapolation variable only benefits the ν_μ samples since we do not separate the ν_e events based on hadronic energy fraction. Therefore, all of the quartiles are summed before performing the ν_e signal extrapolation. However, the ν_e backgrounds are extrapolated in the two $\mathrm{CNN}_{\mathrm{evt}}$ bins separately to provide some constraint there.

The ND is significantly smaller than the FD, meaning it will naturally require events to be more forward-going in order to pass the containment criteria. As a result, the two detectors are sampling slightly different regions of the kinematic phase space leading to differences in selection efficiency and an increased impact of cross section uncertainties which are less well constrained. This is alleviated by performing the extrapolation in bins of outgoing lepton transverse momentum. The p_{T} distributions for the ND and FD $\nu_\mu + \bar{\nu}_\mu$ events and the FD $\nu_e + \bar{\nu}_e$ events can be seen in Figure 5.4. From that figure we see that the FD allows for a larger amount of p_{T} owing to its large size. Like the hadronic energy fraction bins, we generate three equal population quantiles and extrapolate each one individually. Unlike the hadronic energy quartiles, the p_{T} bins are only used in the extrapolation and all p_{T} bins are re-summed before performing the oscillation fit. This procedure is only performed for the ν_μ and ν_e signal events while the background events are predicted as stated above.

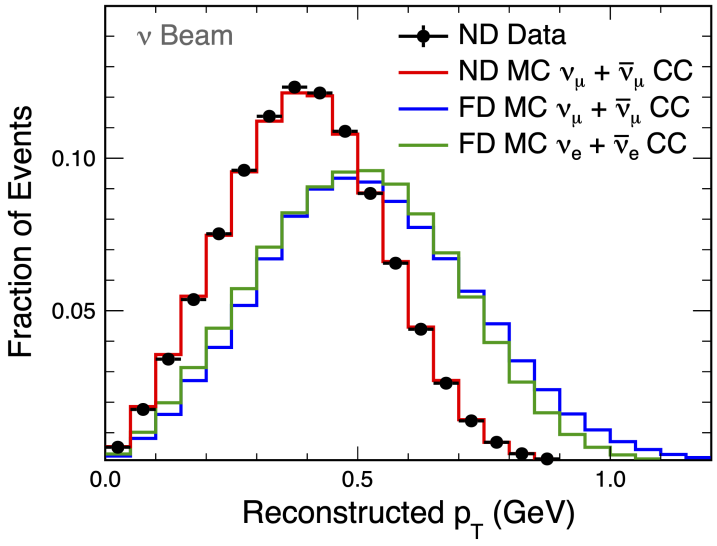


Figure 5.4: Distributions from the neutrino-mode beam of the fraction of selected events versus reconstructed p_{T} of the final state lepton, for the ND ν_μ CC data and simulation, and for the simulated FD ν_μ and ν_e signal events. The corresponding distributions from the antineutrino-mode beam are similar.

5.4 Cosmic Backgrounds

Cosmic ray events have come up a few times in earlier sections and they can contribute significantly to the FD background owing to their ~130 kHz rate. As discussed in Section 3.3.4, we capture cosmogenic events in two ways: (1) by capturing a 500 µs data window centered on the 10 µs beam spill and (2) with a dedicated 10 Hz trigger with the beam spill removed. The timing sidebands before and after the beam spill provide a snapshot of the cosmic rate coincident with the beam and will inherently include any bad runs or other data quality issues that are marked in the spill files. The sideband samples, however, have relatively low statistics, particularly after going through the various selection criteria. Therefore, we use the minimum bias 10 Hz trigger to determine the shape of the cosmic ray energy distributions and scale them to the rate determined by the sidebands. The cosmic ray events are added into the FD prediction after extrapolation and the related uncertainty is drawn from a Poisson distribution due to the low rate of events that pass the selection.

5.5 Systematic Uncertainties

There are a lot of moving parts in NOvA's 3-flavor neutrino oscillation analysis resulting in many components to which systematic uncertainties, or systematics for short, can and must be assigned. In all, we assign over one hundred systematics, with the most dominant ones being related to neutrino interaction cross sections, beam flux, and calibration. Some systematic effects are relatively easy to assess by simply reweighting the energy spectra according to some modification of an underlying component. Others require alterations to basic reconstructed quantities in order for their effects to be properly propagated to the neutrino energy spectra. The most in depth systematics can impact fundamental components in the simulation or calibration processes and require additional sets of simulated events to fully capture all of the changes that they may cause.

5.5.1 Neutrino Interaction Models

The GENIE neutrino interaction generator provides an event reweighting framework to propagate certain model uncertainties [56]. This framework provides various systematic "knobs" allowing for quick event reweighting to produce new true neutrino energy distributions based on changes to specific model parameters. The available knobs include NC and CC vector and axial masses for RES events, NC elastic axial mass and form factor, normalization on the CC QE z-expansion, parameters associated with the Bodek-Yang DIS model, branching ratios and pion angular distribution in RES decays, and low-pion-multiplicity DIS hadronization parameters.

The knobs provided in the GENIE reweighting framework do not cover all of the models used in GENIE v3.0.6, therefore, it was necessary for NOvA to develop additional weights to cover the possible model uncertainties. These include:

- Z-expansion: The z-expansion is a conformal mapping of the axial form factor in CCQE interactions to an analytical unit circle. This method is independent of the nuclear model used for the target and provides more conservative uncertainties [95].
  - parameter knobs were included in the GENIE reweight package but were not properly correlated. NOvA manually calculated the uncertainty affect with correlated parameters.
- QE random phase approximation (RPA): RPA accounts for long-range nucleon-

nucleon interactions which modify the kinematic phase space in CCQE interaction. This enhances the cross section at large Q^2 and suppresses it at low Q^2.

  - Knobs controlling the amount of enhancement and suppression applied by the RPA.

- RES Q^2 suppression: external and internal data show a suppressed rate of RES events at low Q^2. The Berger-Sehgal model we use was tuned to external data to account for this.

  - We apply a MINOS-like Q^2 suppression [96] to the previous simulation (which used a different model) and the ratio of the suppressed old model with the new model provides the uncertainty.

- 2p2h/MEC: events produced by the 2p2h model are adjusted using data-driven corrections and can be more "QE-like" or "RES-like".

  - The relevant QE and RES systematic are applied with 1 σ shifts to modify 2p2h fit which is then taken as the 1 σ 2p2h shape uncertainty.
  - An energy dependence uncertainty was developed for these cross sections by comparing various 2p2h models to the one we used and taking the maximum ratio at each energy as an uncertainty weight.
  - Interactions occur on nucleon pairs one of which depends on the particular model. Asymmetric uncertainties on the nucleon pair fraction were developed by various models to the València model we use.

- DIS: GENIE knobs are provided but are limited in application. have an abrupt cutoff at an invariant hadronic mass of $W = 2\,\text{GeV}$.

  - For multi-pion production we extend the 50% GENIE uncertainty up to $W = 3\,\text{GeV}$ an linearly decrease it to 5% at $W = 5\,\text{GeV}$.
  - The hadron formation zone within the nuclear medium was manually adjusted and re-simulated then compared to the base model to produce uncertainty weights.

- Final state interactions: the hN FSI model parameters required tuning to achieve agreement with data.

  - Using π^+ on ^{12}C scattering data [97, 98, 99, 100, 101, 102, 103] the parameters were adjusted and re-simulated then compared to the tuned model to produce uncertainty weights.

- ν_e ($\bar{\nu}_e$) vs ν_μ ($\bar{\nu}_\mu$) cross section ratio: radiative corrections and second class currents can affect the ν_e ($\bar{\nu}_e$) cross section while leaving the ν_μ ($\bar{\nu}_\mu$) cross section unchanged [104].

  - Following T2K we apply a 2% uncertainty on the ν_e/ν_μ and $\bar{\nu}_e$ /$\bar{\nu}_\mu$ ratios for the radiative corrections and a separate 2% uncertainty for possible second class currents [105].

In all, this produces 71 separate systematic uncertainties, with each one needing to be applied at multiple $\pm\sigma$ levels. This becomes computationally expensive and requires a prohibitively long amount of time. We leverage the fact that many systematics can produce correlated effects on the neutrino energy distribution to combine them into "principal components" to cover many systematics at once. Details on the method can be found in Ref. [106], but in short by randomly varying each of the systematics many times to build up an ensemble of "universes" we can produce a covariance matrix which, when diagonalized, will produce a new vector space of principal components of varying uncertainty impact. Rather than representing all 71 systematics as principal components, we instead separate out the 25 uncertainties with the largest impact and perform the principal component analysis (PCA) on the remaining "small" systematics. The PCA method we generates hundreds of PCs owing to the number of bins used to produce the covariance matrix, and utilizing hundreds of PCs as systematic uncertainties is no better than just applying all of the original systematics separately. Therefore, we keep only the most impactful PCs to estimate the total effect of these systematics on our analysis. In this case we use the twelve largest PCs which, when combined, provide a systematic shift that covers 95% of the total shift when all PCs are included across 97% of the energy bins.

The final neutrino interaction model uncertainty is on the ν_τ cross section. This systematic was derived from a measurement by the OPERA experiment, which compared the measured cross section to that from GENIE [107]. Their measurement was statistically limited but found the CC cross section to be $\sigma_{\nu_\tau} = 1.2^{+0.6}_{-0.5} * \sigma_{GENIE}$. Applying their error bars to the 88% of ν_τ events which are CC results in an uncertainty of 57% which we round up to a symmetric and conservative 60% that we apply as a flat weight to the rate of all ν_τ CC events.

5.5.2 Beam Flux

Like the neutrino interaction model uncertainties, the beam flux uncertainties are also applied as a reweighting of events. Both the hadron production corrections from the PPFX and beam transport components (i.e. magnetic horn current and position) are grouped under the beam systematics. Recall that the PPFX applies data-based corrections to the G4NuMI simulation. To do so, the PPFX has many different knobs which can be adjusted and the uncertainties on each one are propagated through to the final weights via ensemble simulation of many universes. For the beam transport components, we consider the following uncertainties:

- Horn current: ± 2 kA.
- Horn one and two positions in x and y: ± 3 mm.
- Measured beam position on target in x and y: ± 1 mm.
- Measured beam spot size in x and y: ± 0.2 mm.
- Horn cooling water layer: ± 1 mm.
- Target z position: ± 7 mm.
- Beam divergence: $+54$ µrad.
- Impact of using a more detailed horn geometry: had a negligible impact on the beam flux.

With twelve beam transport uncertainties and numerous PPFX knobs, the beam flux systematics are also a great candidate for the PCA method. This is performed in the same manner as with the neutrino interaction systematics, only here we only need to retain the five most impactful PCs in order to have 99% coverage of the total uncertainty across 85% of the energy bins. In this case, the uncertainty weights of the five PCs is also scaled up by 25% to increase the coverage bin fraction up to 95%.

5.5.3 Other Weight-based Uncertainties

We have two final sets of systematic uncertainties which are applied via reweighting events: overall normalization and ν_e acceptance systematics. Effects that impact the total

event normalization include the measured POT, ratio of detector masses, and ND pile-up selection efficiency. The detector mass ratio uncertainty is only applied to FD events but affects both neutrino and antineutrino-mode data. The POT and pile-up uncertainties impact both detectors; however, they have different uncertainties for neutrino and antineutrino-mode data. As stated in the p_{T} extrapolation section, the different detector sizes results in different accepted kinematic regions for the detectors. In particular, the acceptance differences are more pronounced for the ν_e prediction since it is extrapolated from ND ν_μ events. A 0.4% normalization uncertainty was found for the acceptance systematic by reweighting the ND data and simulation to match the FD as a function of Q^2, total p_{T}, and event angle with respect to the beam direction. After reweighting and extrapolation, the new predictions were compared to the nominal predictions and the largest difference out of each of the three parameters was taken as the systematic uncertainty.

5.5.4 Non-weight-based Uncertainties

Some systematic uncertainties have detailed event-by-event impacts and cannot be propagated by simply applying some weights to the energy distributions. In these cases, we assess the uncertainties by altering reconstructed variables that are used in event selection, energy estimation, and extrapolation. These systematics include uncertainties that impact reconstructed muon energy, outgoing lepton angle, visible light deposited by neutrons, and tagging of Michel electrons for extrapolation.

The muon energy systematics can affect both detectors in a correlated manner or only one of the detectors. Additionally, owing to the steel planes, the ND muon catcher must be treated separately from the main detector body. The underlying causes of these systematics vary from mass accounting to uncertainties on the density term in the Bethe-Bloch formula. Regardless of the cause, the effects have been assessed as the impacts they have on the reconstructed muon length, since this is the only variable used to reconstruct the muon energy with the spline-based energy estimator. The determined muon length changes at the $1\,\sigma$ level are:

- Correlated effects: 0.74% in the FD, 0.74% in the ND, and 0.13% in the muon catcher.
- Uncorrelated effects: 0.15% in the FD, 0.13% in the ND, and 0.48% in the muon catcher.

- ND neutron pile-up: 0.46 cm in the ND and 1.3 cm in the muon catcher.

The pile-up systematic is asymmetric and can only increase the length of muon tracks, while the correlated and uncorrelated systematics can either increase or decrease track lengths. Further, the ND neutron pile-up systematic is an absolute increase rather than a relative percentage since thermalized neutrons can be disconnected from their parent interaction and appear at the beginning or end of a muon track thus increasing the apparent length by an amount independent of the actual muon length. This effect is more prevalent in data than in MC, thus the systematic is applied as an increase in track length to cover the discrepancy with data.

With the inclusion of p_{T} in the extrapolation procedure, it is important to consider the uncertainty on the reconstructed angle of the outgoing lepton with respect to the beam direction. In the ND this uncertainty was shown to be largely driven by misalignment of the detector planes, which give a 2.5 mrad uncertainty, while FD studies of the cosmic ray shadow cast by the moon has found a pointing resolution of $<1^{\circ}$. As a result, a conservative 10 mrad ($\sim 0.5^{\circ}$) uncertainty is taken on the reconstructed lepton angle in each detector view.

Data-simulation discrepancies have been found in $\bar{\nu}_{\mu}$ interactions, with the MC events producing more reconstructed prongs than data events. By the nature of $\bar{\nu}_{\mu}$ CC QE interactions, these events will have an outgoing neutron which is difficult to observe in our detector since neutrons must thermalize before depositing energy. Studies of prongs significantly disconnected from the interaction vertex (dominated by neutron daughters) have shown that MC events have an increased rate at low energy and a decreased rate at higher energy when compared to data events. Neutron models are widely varied and, as we've seen, final state interactions within the nuclear medium are poorly understood so the best candidate to explain these effects is a mis-modeling of fast neutrons leading to an increased number of daughter particles or an increase in the simulated neutron interactions. To cover the observed energy shift, a systematic was devised wherein one-in-three prongs related to the primary neutron with energy below 20 MeV have their energies scaled up to remove the observed discrepancy.

The final systematic uncertainty in this category is due to the efficiency of the Michel electron tagging procedure used in the ν_e extrapolation. Uncertainties on the tagging efficiency are due to uncertainties on the fraction of muons which produce Michel electrons, the Michel electron energy spectrum, and Michel electron hits that may overlap with the parent muon, making them unidentifiable. To cover these effects, 8% of Michel electron

candidates are manually forced to pass the selection criteria for the $+1\,\sigma$ shift, while 5% are forced to fail for the $-1\,\sigma$ shift.

5.5.5 Calibration and Light Model

The final category of systematic uncertainties are those that cannot be handled through reweighting or by adjusting reconstructed variables. The calibration and light model are so fundamental to all the subsequent parts of the reconstruction chain that in order to fully propagate any uncertainties requires a full set of newly simulated events. We call these types of systematics "file-based" and they cover uncertainties on the absolute and relative calibration, the calibration shape (i.e. calibration as a function of position in the cell), drift in the calibration over time, overall light level scaling, and the amount of Cherenkov light produced.

The absolute and relative calibration systematics were determined by studying data-MC differences in the energy response for various candidate particles. Selections were formulated to yield enhanced samples of muons and protons from contained QE events along with rock muons, π^0, and Michel electrons, all at the ND. Then data-MC comparisons were made for each sample by comparing the dE/dx as a function of position along the track. From these studies we found that the proton sample showed the largest discrepancy at 5%, which is taken as a conservative estimate for the calibration uncertainty across all samples and both detectors. The absolute calibration uncertainty is assessed by applying a 5% shift, correlated across both detectors, to the calibration scale factor. Since the relative calibration is local to each detector, the largest possible effect is if the ND and FD relative calibrations move in opposite directions. Thus, the relative calibration uncertainty is taken as a 5% shift, anti-correlated between the near and far detectors, on the calibration scale factor.

Recall that in the relative calibration procedure the attenuation fit, Eq. 3.4, is performed as a function of hit position along the length of the cell, W. Comparisons between the calibrated hit energy and true simulated hit energy as a function of W shows some remaining discrepancy that varies across the length of the cell. Through the central regions of the y-view cells a 3% discrepancy is seen while in the x-view the difference is less than 1%. However, in both views the edges of the cells have a greater discrepancy reaching above 5%. This calibration shape systematic is assessed by performing a W-dependent fit to the $cal_E/true_E$ ratio and applying that as the $+1\,\sigma$ uncertainty. The $-1\,\sigma$ uncertainty is simply

a reflection of the +1 σ across the nominal calibration curve.

The final calibration-related systematic is an uncertainty on the aging effect of the fibers, scintillator, and electronics which presents itself as an increase in the calibration scale factor over time caused by a reduction in light output and collection. To handle this effect, the calibration procedure is performed over short timescales determined by running conditions and file production efforts. Studies of the number of reconstructed hits in candidate events over time were performed to quantify the uncertainty on this drift effect. These studies corroborate the fact that we see a reduction in the number of hits as a result of degraded light output. The final uncertainty was found to correspond to a 4.5% decrease in the light model output per year. Owing to the definite direction of the effect, this systematic is assessed as a one-sided uncertainty that is correlated across both detectors.

NOvA's light model, Eq. 3.6, includes an overall scale factor for each detector view along with both scintillation and Cherenkov light production. The energy deposition measured by dE/dx for protons was seen to be 5% lower in simulation than in data, while it was only 1.5% lower for muons. This effect can be reproduced by adjusting the amount of scintillation and Cherenkov light yield (Y_s and ϵ_C in Eq. 3.6). The two parameters were adjusted to maintain the muon light-response while increasing that of protons. To cover the observed proton discrepancy requires the scintillation light output to be decreased by 4.4% and the Cherenkov efficiency to be increased by 83% over the nominal light model tune. Like the drift uncertainty, this systematic is also one-sided and fully correlated between detectors.

The nominal light model tune was found to have a bug in the simulated APD gain setting which resulted in a significant data-MC difference in the photoelectron distribution, but there was not enough time to implement a full fix. Therefore, to estimate the model uncertainty, a small sample was created with the bug fixed and the tuning procedure was re-performed. A comparison of the light scale factors (F_{view} in Eq. 3.6) produced with the improved tune versus the nominal tune was then used to set the systematic uncertainties. In the ND, a 10% increase in the scale factors of both views is taken as the +1 σ shift, while the FD requires a 16% increase in the x-view but a 6% decrease in the y-view. The -1 σ shifts are of the same size but in the opposite direction. Just changing the light model parameters would be almost entirely washed out by the calibration procedure; therefore, each shift in the light model is accompanied by an opposite shift in the calibration scale factor.

5.5.6 Additional Small Uncertainties

The final two systematic uncertainties are on the FD rock and cosmic event rates. The FD ν_μ rock rate is negligible; thus they are completely excluded from the analysis. The ν_e rock events, on the other hand, are included with a 100% systematic uncertainty since we have no way of constraining this background. The cosmic background rate has essentially no systematic uncertainty since it is pulled directly from the NuMI beam sidebands. It does, however, have a statistical uncertainty that is taken as the bounds of the central ±68.2% interquartile range of a Poisson distribution centered on the measured cosmic rate. The Poisson uncertainty is assessed on a bin-by-bin basis.

5.5.7 Impact of Uncertainties

The impacts of the major systematic uncertainties, with and without using the p_{T} extrapolation procedure, are shown in Figure 5.5. NOvA is clearly limited most by the size of our datasets, though the calibration uncertainties are the most limiting of the systematics. We can also see that neutron and cross section uncertainties are reduced by extrapolating in bins of p_{T} while the impact of the lepton reconstruction is increased as expected.

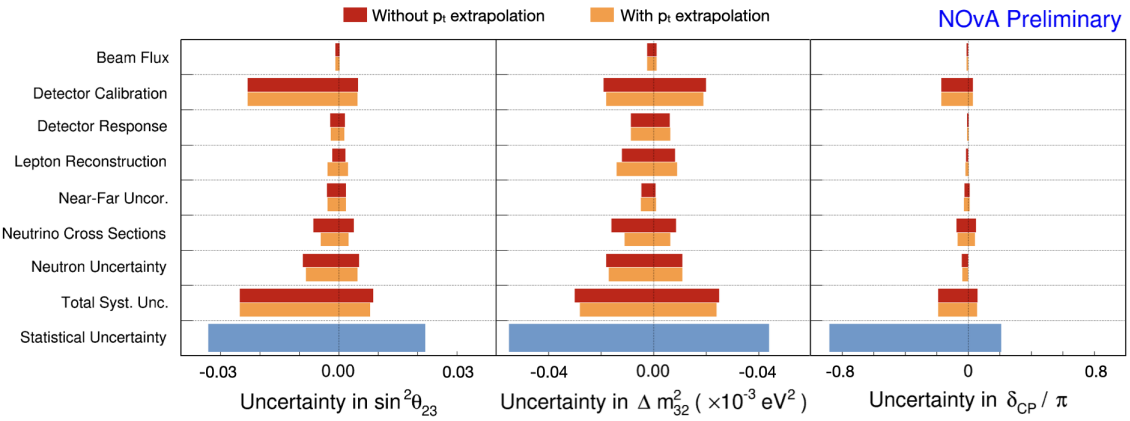


Figure 5.5: Systematic uncertainties on $\sin^2\theta_{23}$, Δm^2_{32}, and δ_{CP} evaluated at the best-fit point. Impacts with (orange) and without (red) p_{T} extrapolation bins are shown for comparison. Detector response includes the light-level and Cherenkov systematic uncertainties.

5.6 Oscillation Fit

The oscillation fit simultaneously uses all of the FD predictions and maximizes a binned Poisson log-likelihood ratio. Explicitly, the samples are four hadronic energy quartiles each for ν_μ and $\bar{\nu}_\mu$, along with two CNN_{evt} core samples and a single bin peripheral sample each for ν_e and $\bar{\nu}_e$ for a total of 14 individual samples. The log-likelihood ratio function is:

$$\ln \mathcal{L} = -\sum_{i=1}^{N} \left(E_i(\boldsymbol{\theta}) - O_i + O_i \ln \frac{O_i}{E_i(\boldsymbol{\theta})} \right) \tag{5.3}$$

where $E(\boldsymbol{\theta}, \boldsymbol{\delta})$ is the prediction for a given set of oscillation parameters, O is the observed data, and the index i runs over the bins from all samples. Often this is rewritten as a statistical χ^2 which is then minimized:

$$\chi^2 = -2 \ln \mathcal{L} \tag{5.4}$$

Systematic uncertainties can be included as nuisance terms, $\boldsymbol{\delta}$, which are systematic pulls measured in number of standard deviations. In this case, we change the prediction $E_i(\boldsymbol{\theta})$ to $E_i(\boldsymbol{\theta}, \boldsymbol{\delta})$ which is now dependent on the systematics. Additionally, we include a gaussian penalty term which sums over all of the possible systematic shifts giving us:

$$\chi^2 = 2 \sum_{i=1}^{N} \left(E_i(\boldsymbol{\theta}, \boldsymbol{\delta}) - O_i + O_i \ln \frac{O_i}{E_i(\boldsymbol{\theta}, \boldsymbol{\delta})} \right) + \sum_{j=1}^{M} \frac{\delta_j^2}{\sigma_j^2} \tag{5.5}$$

where σ_j are the number of events within a $1\,\sigma$ range of the nominal MC event count. For example, if a systematic is applied as a 60% uncertainty on a predicted event count of 10, then we would have $\sigma_j = 6$. If the fit determined that the ideal pull for this systematic is $\delta_j = 0.2$, then that means having this systematic account for 2 events best fits the measured data. During the fitting process many universes are "thrown" where the nuisance parameters are sampled from a gaussian distribution to account for all potential systematic shifts. Non-integer shifts are accounted for by interpolating the systematic effect between integer shift

values. We then profile over[2] the nuisance parameters to find the most probable systematic shifts given the measured data.

NOvA's samples are not particularly sensitive to all of the oscillation parameters, so we constrain some parameters by their global fits listed in the PDG. These are: $\Delta m^2_{21} = 7.53 \times 10^{-5}$ eV^2 and $\sin^2\theta_{12} = 0.307$ which are not varied and $\sin^2\theta_{13} = 0.0210 \pm 0.0011$ which is varied as a nuisance parameter like the systematic uncertainties [45]. The parameters we fit for (Δm^2_{32}, $\sin^2\theta_{23}$, and δ_{CP}) are all allowed to vary without constraint. Contours, either 1D or 2D, can be generated by profiling over whichever oscillation parameters are not being plotted.

As stated above, the best-fit point is the location that minimizes Eq. 5.5, therefore, when we plot the fit surfaces we will typically plot $\Delta\mathcal{X}^2$ where the $\mathcal{X}^2$ for the best-fit location is subtracted from all other points. $\sqrt{\Delta\mathcal{X}^2}$ is commonly referred to as significance, σ, which has the typical interpretation if the parameters follow a gaussian distribution (i.e. with one free parameter $1\,\sigma$ has a coverage probability of 68.27% and for $2\,\sigma$ it's 95.45%). The gaussian assumption, however, is explicitly incorrect for the oscillation parameters we consider here; $\sin^2\theta_{23}$ contains physical boundaries since $0 < \sin^2\theta < 1$, δ_{CP} is a poorly constrained cyclic parameter, and our event rates are relatively low and subject to Poisson fluctuations. For these reasons, the naive coverage values do not apply and empirical coverage values must be determined. We do this by utilizing the Unified Approach of Feldman and Cousins [108] (also referred to as FC corrections or just FC for short) which builds on Neyman's construction of confidence intervals [109]. This technique requires the construction of $\mathcal{O}(10,000)$ universes at every point in the parameter space. Each universe then has statistical fluctuations applied to generate sets of mock data. The $\mathcal{X}^2$ fit is performed for each set of mock data, which settles on some set of oscillation parameters. The best-fit $\mathcal{X}^2$, $\mathcal{X}^2_{\text{best}}$, is compared to the $\mathcal{X}^2$ computed from the true oscillation parameters used to generate that universe, $\mathcal{X}^2_{\text{true}}$, and the confidence intervals are constructed from the range that gives the desired percent coverage of the difference, $\mathcal{X}^2_{\text{true}} - \mathcal{X}^2_{\text{best}}$ (i.e. the "proper" $\Delta\mathcal{X}^2$ for a confidence interval of 90% is the value that includes $\mathcal{X}^2_{\text{true}} - \mathcal{X}^2_{\text{best}}$ for 90% of the mock data universes).

[2]Profiling is the replacement of the parameter with the value that best fits the data at the given point for the parameter(s) of interest. For example, to go from a 2D $\mathcal{X}^2$ surface over parameters a and b to a 1D surface over just parameter a, for each value of a, you will take the $\mathcal{X}^2$ from the point on the 2D surface with the minimum value for that a. In other words, for each value of a, you choose the b value that minimizes the 2D surface along that a slice.

5.7 Results

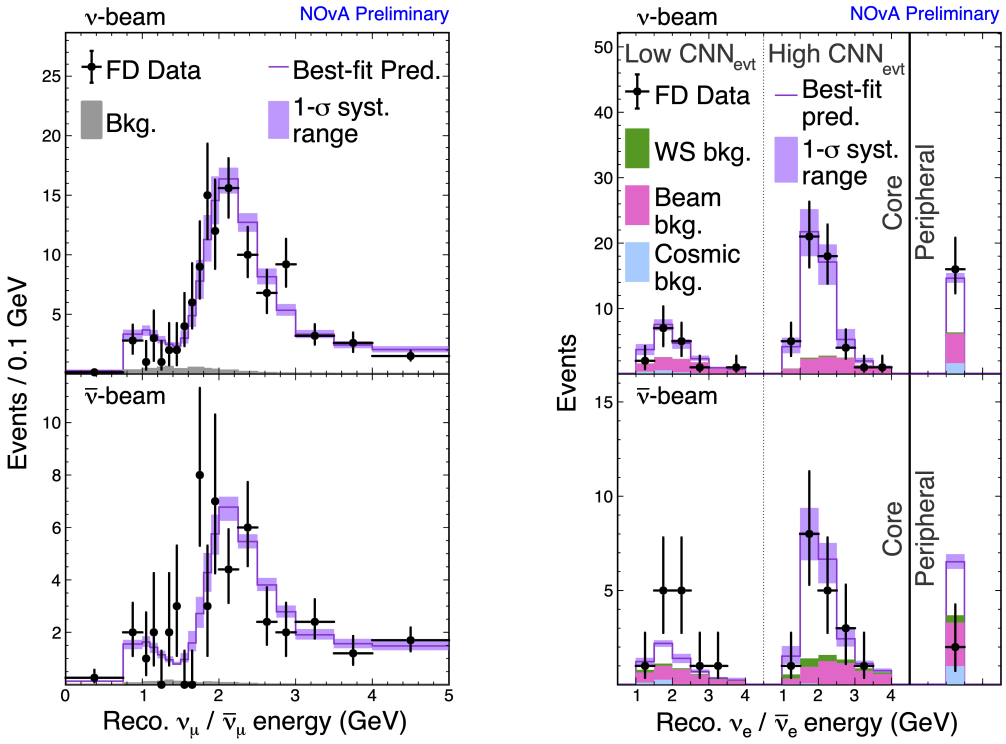
With all of the analysis components covered, we are now ready to look at the oscillation results. Utilizing the ND samples, shown earlier in Figure 5.3, and performing the Feldman-Cousins corrected fit yields the best-fit points in Table 5.1. We find the best-fit to lie in the normal mass ordering and upper θ_{23} octant (significance of $1.0\,\sigma$ and $1.2\,\sigma$, respectively), where $-2\ln\mathcal{L} = 173.55$ for 175 degrees of freedom (p-value of 0.705). The total event counts for each sample at the best-fit oscillation point, broken down by predicted event type, are shown in Table 5.2, and Figure 5.6 shows the FD predicted energy spectra.

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

Table 5.1: Summary of oscillation parameter best-fit results for different choices of the mass ordering (Normal or Inverted) and upper or lower θ_{23} octant (UO, LO), along with the FC corrected significance (in units of σ) at which those combinations are disfavored.

	Neutrino beam		Antineutrino beam	
	ν_μ CC	ν_e CC	$\bar{\nu}_\mu$ CC	$\bar{\nu}_e$ CC
$\nu_\mu \to \nu_\mu$	201.1	1.7	26.0	0.2
$\bar{\nu}_\mu \to \bar{\nu}_\mu$	12.6	0.0	77.2	0.2
$\nu_\mu \to \nu_e$	0.1	59.0	0.0	2.3
$\bar{\nu}_\mu \to \bar{\nu}_e$	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	$214.1^{+14.4}_{-14.0}$	$59.0^{+2.5}_{-2.5}$	$103.4^{+7.1}_{-7.0}$	$19.2^{+0.6}_{-0.7}$
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 5.2: Event counts at the FD, both observed and predicted at the best-fit point.



(a) (b)

Figure 5.6: FD reconstructed energy spectra for (a) ν_μ CC and (b) ν_e CC with the neutrino beam-mode on the top and antineutrino on the bottom. Hadronic energy fraction quartiles are combined in the ν_μ plots.

No strong asymmetry in the rates of ν_e versus $\bar{\nu}_e$ was observed. Therefore, these results disfavor combinations of oscillation parameters that would lead to such an asymmetry. Looking back at the bi-probability ellipses from chapter 2 (Figure 2.7), our data lie near the center of all the ellipses in the highly degenerate region, as shown in Figure 5.7. In particular, the inverted mass ordering with $\delta_{CP} = \pi/2$ is excluded at more than $3\,\sigma$ and the normal mass ordering with $\delta_{CP} = 3\pi/2$ is disfavored at $2\,\sigma$ confidence.

Finally, we can see the confidence contours in Figure 5.8. From those plots we see that while we can disfavor certain combinations of mass ordering and δ_{CP}, all values of δ_{CP} are allowed at the $1\,\sigma$ confidence level, given the appropriate choice of octant and ordering. Degeneracies aside, our results are still competitive with other long-baseline 3-flavor oscillation experiments like T2K and MINOS, as shown in Figure 5.9. Interestingly, the disfavored δ_{CP} region in Figure 5.8b lies almost exactly at T2Ks best-fit location; a tension which will require more data and a combined NOvA-T2K analysis to fully resolve.

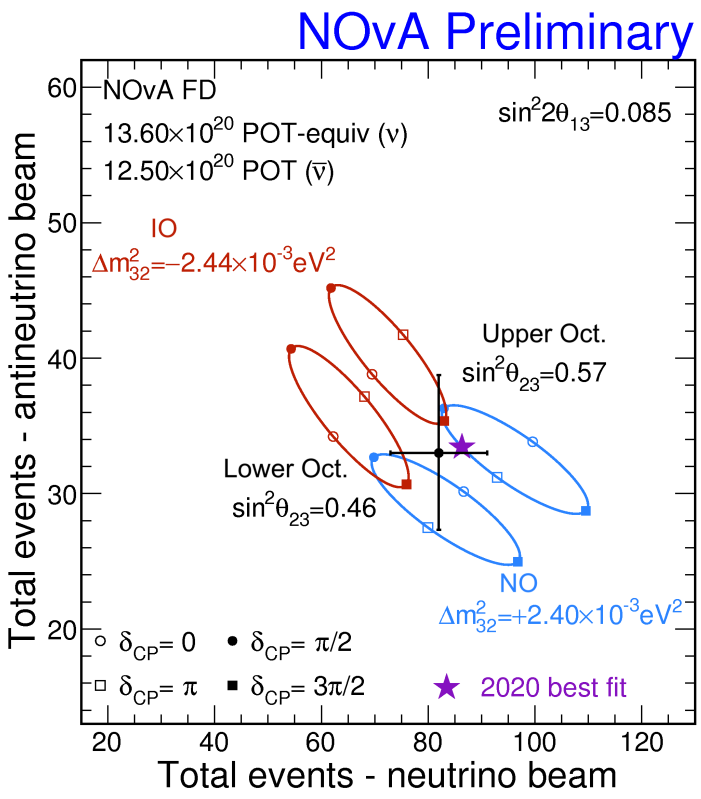


Figure 5.7: Bi-event plot with data and the 2020 best fit marker. The four ellipses correspond to the individual best fits in each quadrant of mass ordering and θ_{23} octant.

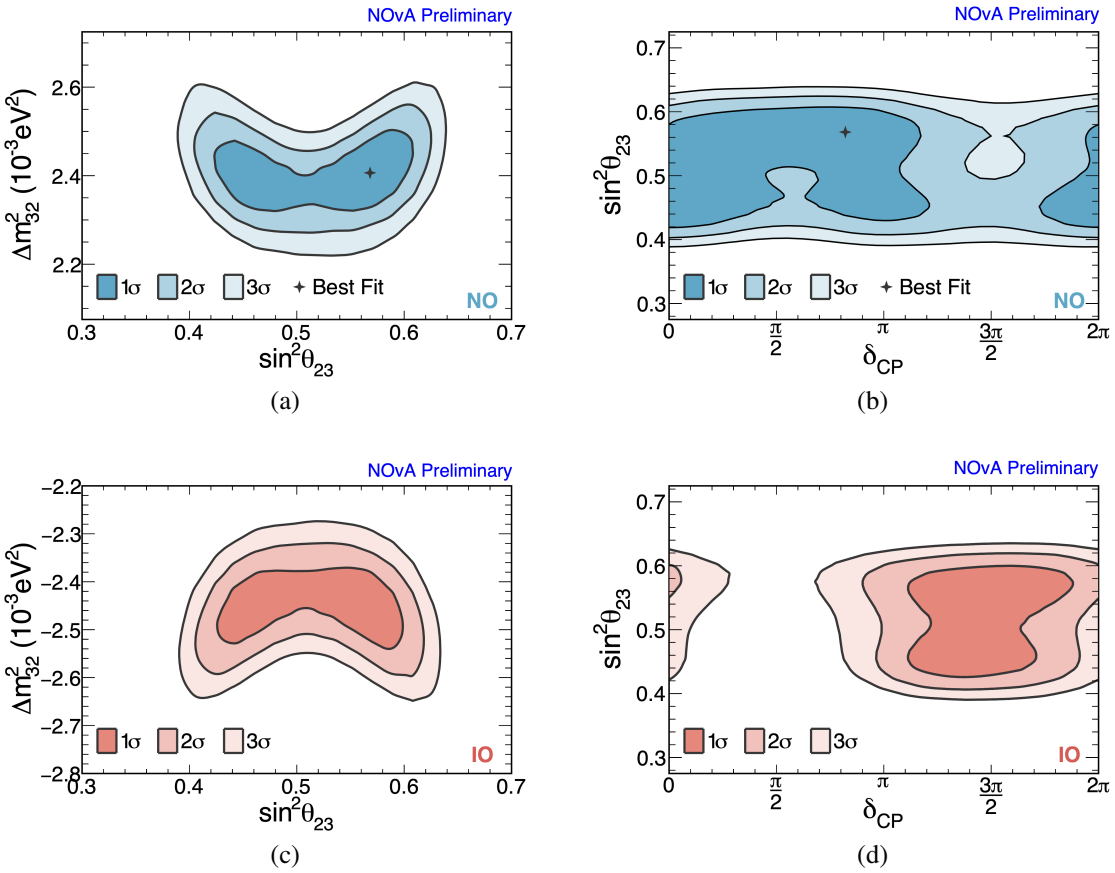


Figure 5.8: Confidence interval contours for (left) Δm^2_{32} vs $\sin^2\theta_{23}$ and (right) $\sin^2\theta_{23}$ vs δ_{CP} in the (top) normal and (bottom) inverted mass ordering. Plots are shown with 1, 2, and 3 σ Feldman-Cousins corrected contours. The best-fit point is in the normal ordering with $\delta_{CP} = 0.821 \times \pi$, $\sin^2\theta_{23} = 0.568$ (Upper Octant), $\Delta m^2_{32} = 2.406 \times 10^3$.

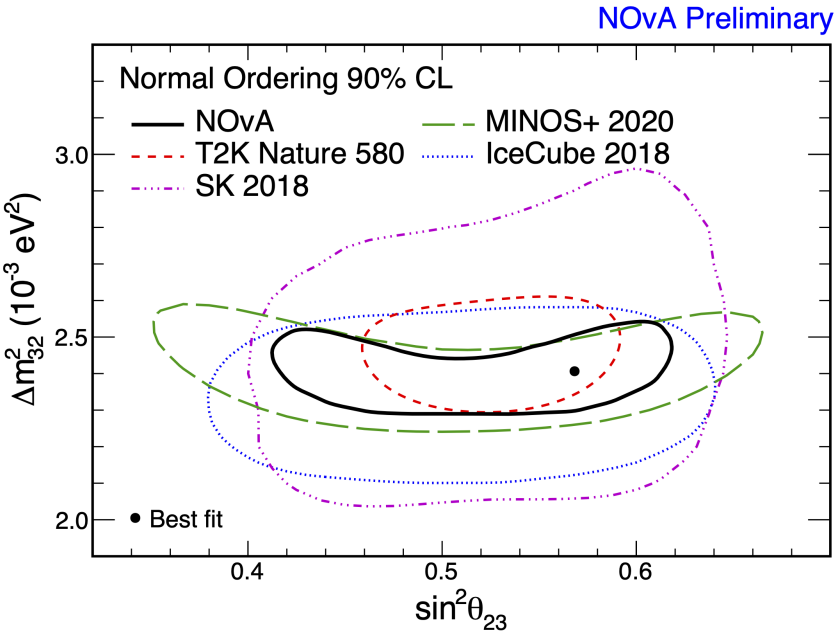


Figure 5.9: The 90% confidence level region for Δm^2_{32} versus $\sin^2\theta_{23}$, with the FC corrected allowed region and best-fit point for NOvA overlaid on contours from other experiments [110, 111, 112, 113].

6. A Long Short-Term Memory Neural Network

The "standard" spline-based ν_μ energy estimator (EE), described in Chapter 4, is straightforward and well motivated by event kinematics. The energy of the outgoing muon is highly correlated to its track length, so utilizing a spline fit yields a muon energy resolution of ~3%. The hadronic system, however, is much more messy and with a simple spline fit the best variable we have access to is the calibrated energy deposition from the hadronic hits. This results in a much worse energy resolution of ~26% since neutral and low energy particles are common and these may not deposit enough energy to overcome the trigger threshold. In total, the spline-based EE results in a total ν_μ energy resolution of ~9% while utilizing only two reconstructed variables from the neutrino event.

There are, of course, many reconstructed quantities which are completely ignored by the spline EE. These range from low level values such as start and stop positions of the various particles up to higher level things like CNN_{png} scores or particle momentum. A machine learning algorithm is a natural choice to make use of a large amount of input variables. These can range from BDTs to deep neural networks with a regression output rather than the classification tasks referenced in Chapter 4. In neutrino interactions, the number of outgoing particles is not set and therefore any number of particle prongs may be reconstructed by the fuzzyk algorithm. A variable number of inputs loans itself to a recurrent neural network (RNN) which takes in inputs sequentially and finds relationships between them [114].

Now for a quick overview of how deep neural networks (DNNs) work. I'll be limiting the discussion to "feed-forward neural networks". DNNs contain layers of neurons and take in some inputs to produce an output or outputs. The number of layers and neurons are both customizable parameters when creating a DNN. The neurons of one layer are typically fully connected to all of the neurons of the next layer, with the layers between the input and output being referred to as hidden layers. Networks are called "deep" when they contain more than one hidden layer. Each neuron produces some output vector which is a weighted combination of its inputs that are fed to an "activation function" to produce a standard output. There are many different types of activation functions, but here I'll briefly mention

four of them:

- Rectified linear unit (ReLU): 0 if $z < 0$, otherwise it is equal to z
- Sigmoid: $\sigma(z) = \frac{1}{1+e^{-z}}$, output is between 0 and 1 and is good for predicting probabilities
- tanh : Like the sigmoid in shape but ranges from -1 to 1. Provides more separation between negative and positive inputs. Good for classification tasks.
- Softmax: Used for N inputs , $\frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}}$. Like the sigmoid but returns probabilities for multiple classes.

where z is the input value. The weighted sum of the inputs is plugged into the activation function of a neuron, giving the output that is to be passed along to the next layer.

In training a DNN, we are finding, or rather the network is finding, the optimal weights between neurons in order to accurately predict the target quantity. To do so, a loss function is minimized via gradient descent. The initial weights are seeded, typically near zero, and the loss is calculated. Then the partial derivative of the loss with respect to each weight is determined in a process called backpropagation [115]. Since each layer is connected to the previous one, we can calculate all the partial derivatives for all weights by starting at the most downstream layer and working backward, making liberal use of the chain rule. The particular application of the backpropagated gradients can vary through the use of different optimization algorithms but essentially the weights are updated as some function of the gradient in order to minimize the loss function. Just as with the activation functions, there are various loss functions which are useful for different applications. For instance, multi-class classification problems would do well to utilize a softmax activation coupled with categorical crossentropy as the loss function [116].

Standard RNNs suffer significantly from the vanishing or exploding gradient problem [117]. This is particularly problematic because RNN layers not only pass outputs to the next layer, but they also pass them to the next neuron in the same layer as the inputs are sequentially added. If the weights between neurons are not near 1 then the repeated multiplication that occurs during backpropagation will cause the gradients to become progressively smaller or larger as the distance between inputs grow. This creates a natural block where inputs can only impact other inputs that are nearby in the input order and toward the end of the input chain. This is where long short-term memory (LSTM) networks come in, as

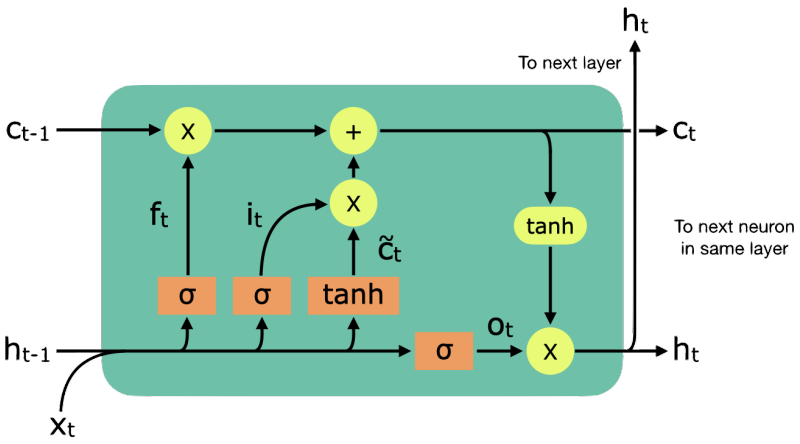


Figure 6.1: Illustration of an LSTM neuron [119]. The orange boxes show the four activation functions, one tanh and three sigmoid. The yellow circles are element-wise matrix operations, addition, multiplication, and tanh. The input vector is x_t, the hidden state is h_t, and the cell state is c_t.

they were specifically designed to overcome the vanishing gradient issue when connecting far-separated inputs [118].

Rather than using a single activation function, an LSTM uses four in a very specific configuration. Additionally, each LSTM neuron outputs both a hidden state which is passed to the next neuron and to the next layer, along with a cell state which is only passed to the next neuron. Therefore, each neuron takes in the two outputs from the previous one along with a new external input either from the previous layer or the overall set of input variables. The new input and previous hidden state are passed to four different activation functions. The first is a sigmoid "forget gate" which decides what things are irrelevant in the previous hidden state based on the new input variable. The second activation function is a sigmoid "input gate" that chooses which of the hidden state values will be updated. Third comes a tanh gate that creates the input for the new cell state. The outputs from the input gate are multiplied by the new cell state so that the interesting things are updated. The final gate for the new input and previous hidden state is a sigmoid "output gate" which determines what parts of the cell state will be output as the new hidden state.

The previous cell state is multiplied by the output of the forget gate to drop uninteresting quantities. Then the cell state is updated using the result of the input gate. Finally, a tanh is applied to the new cell state, which is then combined with the result of the output gate to produce the new hidden state. The freshly made hidden and cell states are passed to the next

neuron and the process repeats with the next input variable. LSTMs overcome the vanishing gradient issue because the cell states that are passed to the next neuron are unweighted so they cannot cause the gradient to decrease during backpropagation. This full procedure can be difficult to conceptualize in words, so an example neuron is shown in Figure 6.1 and the mathematical operations are:

$$\begin{aligned}
\text{forget gate:} \quad & f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) & \text{cell input:} \quad & \tilde{c}_t = \tanh\left(W_c x_t + U_c h_{t-1} + b_c\right) \\
\text{input gate:} \quad & i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) & \text{cell state:} \quad & c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \\
\text{output gate:} \quad & o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) & \text{hidden state:} \quad & h_t = o_t \circ \tanh\left(c_t\right)
\end{aligned}$$

where W_a and U_a are weight matrices for the input and hidden state vectors at each gate, b_a is a bias vector at each gate, and $\circ$ is the element-wise Hadamard product.

6.1 The Basic Network

The implementation of an LSTM-based ν_μ energy estimator (EE) for NOvA was first investigated by a postdoc, Alex Radovic, and a graduate student, Dmitrii Torbunov [120]. With a goal of utilizing much more of the event information, the LSTM EE takes inputs from reconstructed particle prongs as well as information from the event slice as a whole. Prongs were chosen as opposed to the Kalman tracks used by the spline-based EE due to waning collaboration support for the Kalman algorithm. Recall from Chapter 4 that the fuzzyk algorithm first performs 2D reconstruction in each detector view and then attempts to view-match prongs to obtain full 3D reconstruction. Depending on the particle characteristics or path in the detector, it is common for some prongs to remain unmatched. These 2D-only prongs still contain valuable event information and are, therefore, also utilized by the LSTM EE alongside the 3D prongs. The complete list of input variables is:

- 3D prongs:
  - `dir{X,Y,Z}`: cosine of the prong angle w.r.t. each direction.
  - `start{X,Y,Z}`: starting coordinates of the prong.
  - `len`: prong length.
  - `nhit`: number of hits in the prong.
  - `nhitx`: number of hits in the prong in the x-view.

  - `nhity`: number of hits in the prong in the y-view.
  - `nplane`: number of planes passed through by the prong.
  - `calE`: total calorimetric energy of all hits in the prong.
  - `weightedCalE`: total calorimetric energy of all hits in the prong taking into account hits that are shared between prongs.
  - `cnnpng`: Score from the CNN_{png} classifier.
    * `muonid`
    * `electronid`
    * `pionid`
    * `protonid`
    * `photonid`
  - `bpf{muon,proton,pion}`: variables determined by the BPF tracking algorithm for each of the three particle hypotheses.
    * `energy`
    * `momentum{X,Y,Z}`
    * `overlapE`: estimation of the amount of non-muon, pion, or proton energy lying on the track near the event vertex. This is based on the energy deposited by a minimum ionizing particle for each BPF particle hypothesis.

- 2D prongs:
  - Same as the 3D prongs excluding the CNN_{png} and BPF variables.

- Slice:
  - `calE`: total calorimetric energy of all hits in the slice.
  - `remPngCalE`: `calE` - sum of all 3D prong `calE`
  - `nHit`: number of hits in the slice
  - `orphCalE`: slice calorimetric energy not grouped in the prongs.
  - `coarseTiming`: a flag denoting whether multi-point readout was used for this event.
  - `lowGain`: a flag denoting whether low or high gain APD voltages were used for this event.

In total we have 33 inputs for each 3D prong, 13 for each 2D prong, and 6 for the event as a whole; much more than the 2 inputs per event used by the spline EE.

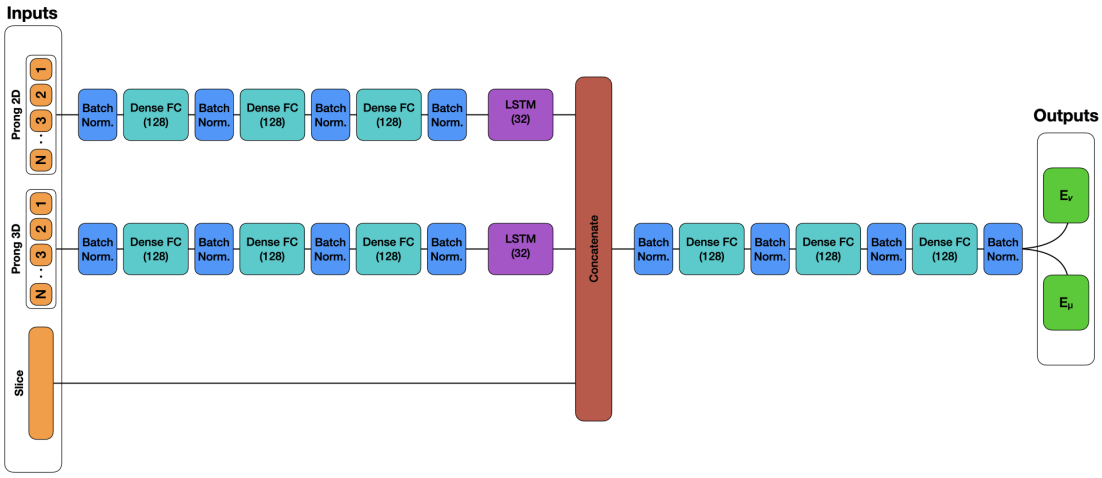


Figure 6.2: Base LSTM energy estimator architecture showing the three input paths. The batch normalization is shown by the blue boxes, cyan boxes represent fully connected dense layers containing 128 neurons each, the LSTM layers are purple, and the concatenation step is shown in red.

The network architecture is shown in Figure 6.2. Stepping through that figure, we begin with the inputs which are passed through a series of pre-processing stages before going to an LSTM layer. The LSTM outputs are concatenated and a series of post-processing layers are used before predicting the neutrino and muon energies. With only one set of variables per event the slice inputs will not benefit from an LSTM layer. Instead, the slice inputs bypass the pre-processing and LSTM stages and pass directly to the concatenation. Pre and post-processing is performed with a series of batch normalization and dense fully connected layers with 128 neurons each. DNNs are subject to an issue called "covariate shift" where alterations to upstream weights necessarily change the input distributions for downstream layers [121]. This causes the optimization process to continually chase after changing distributions. Batch normalization mitigates this issue by normalizing the inputs to maintain the same distributions throughout training [121]. The dense layers utilize the ReLU activation function. The LSTM layers each take input vectors that are 128 elements long and output vectors containing 32 elements. The number of LSTM neurons, or time

steps, varies from event to event based on the number of 2D and 3D prongs that were reconstructed. After post-processing, which is similar to the pre-processing, a pair of single-node dense layers without activation functions perform the regression to predict the muon and neutrino energies.

Simplified sample selections were used to avoid biasing the LSTM EE with non-signal-like events. The selection cuts include: the basic quality and containment cuts discussed in the last chapter along with requirements that the event originates as a muon neutrino, neutrino energy is below 7 GeV, and loosened cuts on the muon PIDs (remID > 0.15 and CNN_{evt} muon score > 0.5). This produces data sets containing between 1.5×10^6–4×10^6 events depending on the detector and beam mode. These samples are too large to process due to technical limits; thus they were decreased to be around 5×10^5–8×10^5 events. From these samples, 30% were used for testing with the remainder being used for training.

6.2 Sample Re-weighting

Recall that the NuMI beam is peaked at about 2 GeV which created a tendency for the ν_e energy estimator to bias itself toward that peak with a poorer energy resolution for events above and below the peak. The ν_μ LSTM EE is similarly susceptible[1] to energy biasing due to the peaked nature of the NuMI beam simulation used in training. This bias can be sidestepped by re-weighting the training distributions such that the neutrino flux is flat across the energy range [120]. The result of flat weight training can be seen in Figure 6.3 where the LSTM EE trained without flat weights displays a significant tendency to over(under)estimate the energy of events below(above) 2 GeV. With flat weights applied, the LSTM EE results in a mean that is much closer to zero over the majority of the energy range.

In the ND samples, the energy distributions are much more peaked than events at the FD owing to the proximity of the beam source. This results in low training statistics at energies away from the peak, which in turn require high weights in order to flatten the distribution. The end result is a poor convergence of the training loss in these samples. To rectify this, we could generate new training samples with much higher statistics, though this would be a large undertaking without significant benefit to other algorithms within the

[1]Interestingly, the standard spline ν_μ EE is less plagued by this bias due to the nearly linear response between muon track length and energy and the fact that the muon carries the majority of the energy from a ν_μ CC interaction.

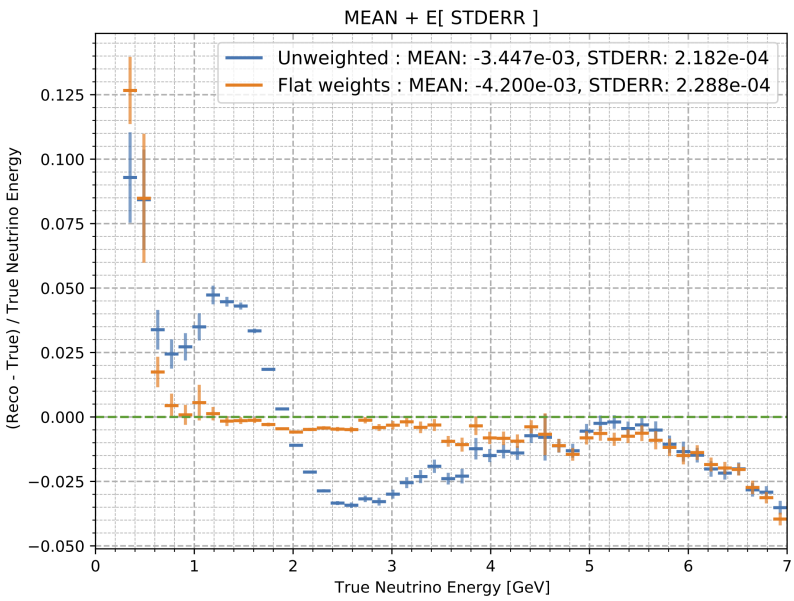


Figure 6.3: Mean of the LSTM EE energy resolution as a function of true neutrino energy for the far detector neutrino mode samples. The blue curve is an LSTM EE trained without weighting the sample and the orange curve is one trained with flat weights.

collaboration. Instead, it was decided that clipping the weights at a maximum value of 50, which will prevent the low statistics regions from being over-weighted in the training stage. This approach does slightly reduce the network performance for events above about 3 GeV, but it significantly improves the loss convergence during training [120].

6.3 Noise Injection

As discussed in Chapter 5.5, the calibration systematic uncertainties produce some of the most significant limitations to the precise measurement of the oscillation parameters. Additionally, because the LSTM EE directly uses the calibrated energy from the prongs and related variables like BPF overlap energy, it is particularly susceptible to changes in those inputs. We can evaluate which inputs are most important to the LSTM EE by injecting artificial noise into each of the inputs of the evaluation sample and assessing the resulting change in network performance, as described in Ref. [122]. This method will not provide an absolute measure of the importance of each input, but it does allow for the assessment of the relative importance of the inputs. In this case we evaluate the energy resolution (i.e. the RMS of the $(E_{reco} - E_{true})/E_{true}$ distribution). The most important inputs are those

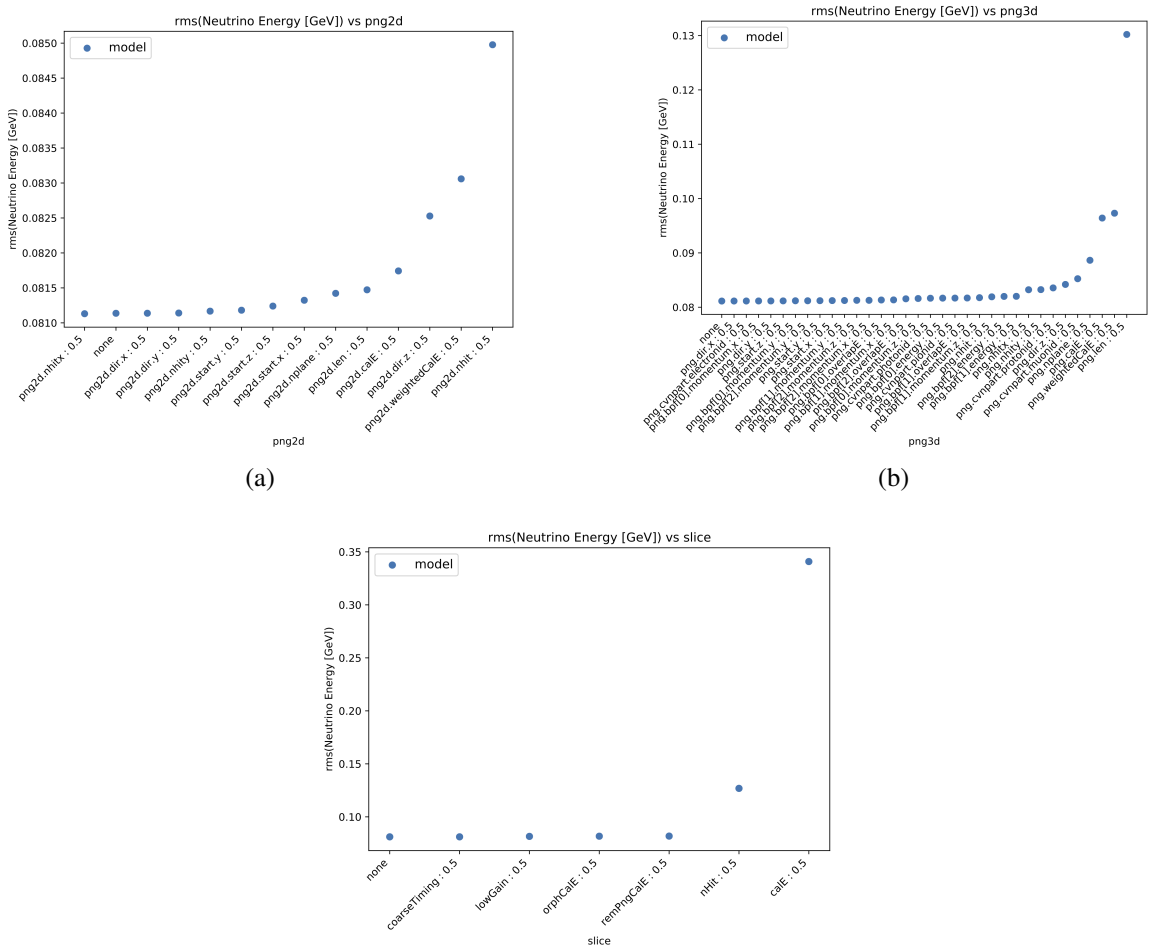


Figure 6.4: RMS of the energy resolution for the LSTM EE trained and evaluated on far detector, neutrino-mode data as a function of perturbed input. Each input was modified by sampling a gaussian with a 50% spread. Prong 2D inputs are shown in (a), prong 3D inputs are in (b), and slice inputs are in (c).

that correspond to larger energy resolutions which signals a degraded EE performance. An example of the result from the input perturbation on the slice inputs for the FD neutrino-mode network is shown in Figure 6.4 and the top five variables from each input type are listed in Table 6.1. Thus we see that the variables directly related to the calibrated energy depositions are consistently among the most important inputs relied upon by the LSTM EE. The other trained networks (i.e. antineutrino-mode at the FD and both neutrino modes at the ND) have very similar rankings, with calibrated energy and related variables being among the most important inputs.

rank	png2d	png3d	slice
1	nhit	len	calE
2	weightedCalE	weightedCalE	nHit
3	dir.z	calE	remPngCalE
4	calE	nplane	orphCalE
5	len	cvnpart.muonid	lowGain

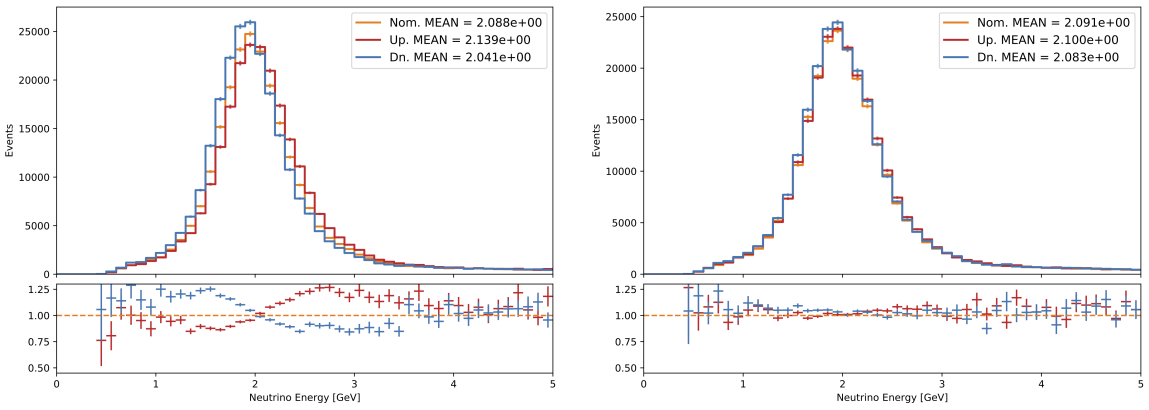
Table 6.1: Top five important inputs for the far detector, neutrino-mode network ranked using the perturbation method.

To remedy the reliance on calibrated energy, we can train the network in such a way that it "knows" that the calibrated energy inputs may be less reliable. This was done by injecting noise into the relevant inputs during the training phase. Three noise models were tested: discrete, uniform, and gaussian. The discrete model randomly applies some noise, ϵ, drawn from the set $\{-a, 0, a\}$ with equal probability. This mimics the calibration up and down systematic samples, but here we can set the value a to any percentage while the systematic samples are set at 5% based on data studies. The uniform noise model is similar but instead uniformly samples ϵ from the range $[-a, a]$. Finally, the gaussian model samples from a normal distribution with width a: $\mathcal{N}(0, a)$. For each noise model, the parameter a was tested between 0 and 50% in steps of 5%. Noise was injected into the following inputs:

- 3D prongs:
  - calE
  - weightedCalE
  - bpf{muon, proton,pion} overlapE
- 2D prongs:
  - calE
  - weightedCalE
- Slice:
  - calE
  - orphCalE
  - remPngCalE

The optimal noise injection was found to be the discrete model with $a = 20\%$ by comparing the energy distributions from the nominal, calibration up, and calibration down MC samples. The energy spectra for the basic LSTM and the noise injected LSTM networks are shown in Figure 6.5, and we see that the difference between the calibration samples and the nominal sample is significantly decreased in the network trained with noise injection.

Finally, we can compare the input importance via perturbation for these basic LSTM EE and the one trained with noise injection. From Figure 6.4, we notice that the most important input among all input types is the slice `calE` a 50% perturbation of which increases the energy resolution from about 8% up to about 35%. Performing the input perturbation on the noise-injected network greatly reduces the impact of the slice `calE` input such that the energy resolution only increases to about 18% when this input is perturbed, as seen in Figure 6.6. In all we see that noise injection greatly reduces the impact of the calibration systematic uncertainties on LSTM EE, and that although `calE` and related variables are still among the most important inputs the network becomes quite robust to the relatively small uncertainty of ±5% used in the analysis.



(a) (b)

Figure 6.5: Energy spectra and ratios of shifted to nominal calibration for the FD neutrino-mode (a) basic LSTM and (b) noise-injected LSTM evaluated on the (orange) nominal, (red) calibration up, and (blue) calibration down MC samples.

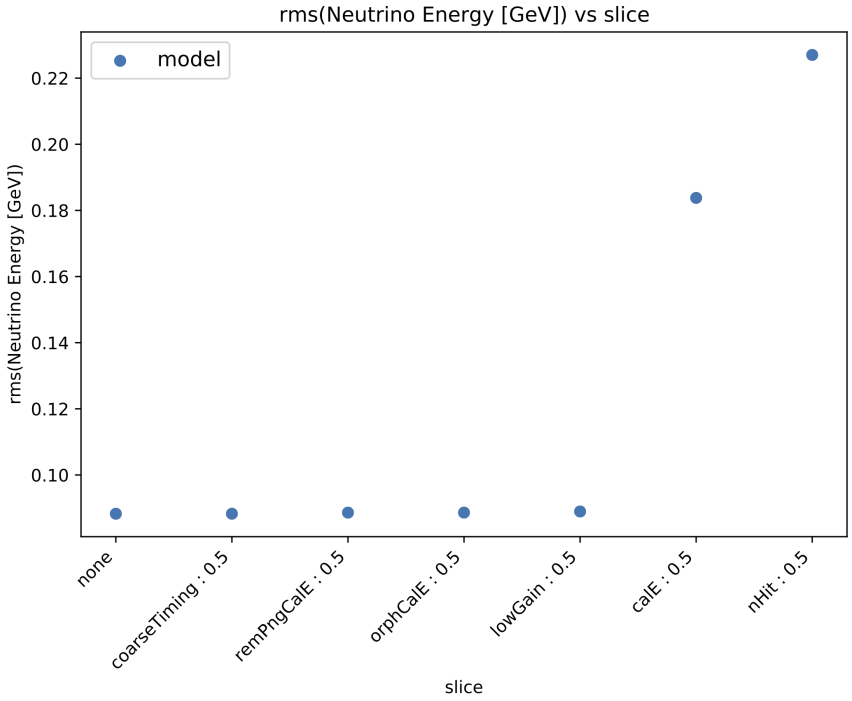


Figure 6.6: RMS of the energy resolution for the LSTM EE trained with discrete noise injection and evaluated on far detector, neutrino-mode data as a function of perturbed slice input.

6.4 Results

The LSTM architecture with flattened distributions and noise injection was trained on datasets containing about 1×10^6 events for the FD neutrino and antineutrino energy estimators and 1×10^6 events for ND neutrino mode and 7×10^5 events for the ND antineutrino mode energy estimators. The performance of the four LSTM networks is compared to the spline-based EE in Table 6.2. We see that the LSTM EE provides a comparable or better

Energy Estimator		Resolution (%)	Bias ($\times 10^{-3}$)
FD	FHC	8.83 (9.13)	-3.30 (-14.96)
	RHC	7.41 (8.20)	-6.16 (-11.68)
ND	FHC	11.13 (10.82)	-1.58 (-24.0)
	RHC	9.65 (9.41)	-1.30 (-21.9)

Table 6.2: Energy resolution and bias (RMS and mean, respectively, of the energy resolution distributions) of the LSTM EE and spline-based EE, in parentheses, for the FD and ND in neutrino (FHC) and antineutrino (RHC) beam modes.

overall energy resolution and is less biased than the spline-based method. Recall also that the LSTM EE has the benefit of a flatter energy resolution and bias across the analysis energy range.

It is not necessary to perform the full 3-flavor analysis in order to determine the impact of the LSTM EE. Instead, we can look at sensitivity contours made using Asimov data, named for the science fiction writer Isaac Asimov. An Asimov dataset for the 3-flavor analysis is one in which we generate fake data points directly from the predicted FD spectrum at some given oscillation parameters without systematic shifts applied [123]. Statistical fluctuations are ignored and the number of event counts in an Asimov dataset do not have to be integer values. Asimov data represent the median sample from the ensemble of all possible statistically shifted and systematically fluctuated universes. Performing the oscillation fit to Asimov data thus allows us to estimate the median significance we can expect[2] under the assumption of our chosen oscillation parameters [124].

Recall that we have a few systematic uncertainties that are applied by directly changing certain reconstructed quantities. The relevant ones for the LSTM EE are the systematics related to the reconstructed lepton angle, muon energy scale, and neutron energy deposition. The lepton angle and muon energy scale uncertainties are, unfortunately, applied as modifications to the muon-identified Kalman track and spline-based EE variables. Since the LSTM EE utilizes prongs, these changes are not propagated to the relevant input variables. From Figure 6.4 we know that the prong x and y directions are not very important, although the prong z direction is somewhat highly regarded by the LSTM EE. The current implementation of the muon energy scale uncertainty is more problematic since it changes energy-based variables but is not propagated to the individual prongs. Additionally, the effects that the muon energy scale systematic are covering will also have some impact on non-muon prongs as well. Initial studies have shown that the LSTM is robust against small changes in the non-muon prongs, but the proper systematic implementation has not been studied in great detail. For these reasons, when considering the Asimov sensitivities, the lepton angle and muon energy systematics have been removed. Luckily, the neutron energy systematic is applied directly to the prong level variables and will be picked up when running the LSTM EE concurrently with the systematic evaluations[3].

[2]Of course, once statistical fluctuations are included we know that, even with many repeated experiments, this median value will never actually be achieved.

[3]In the typical NOvA analysis chain, the DNNs are evaluated long before the systematics are applied in order to save processing time later. In my analysis, I've been able to evaluate the LSTM EE at run-time instead. This allows the neutron systematic uncertainty to be propagated correctly.

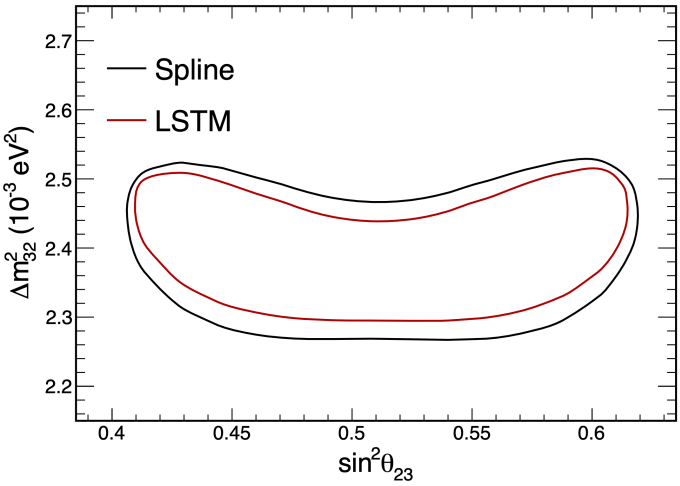


Figure 6.7: Asimov 90% sensitivity contours of Δm^2_{32} versus $\sin^2\theta_{23}$ with ν_μ-only data for the (black) spline-based and (red) LSTM EEs.

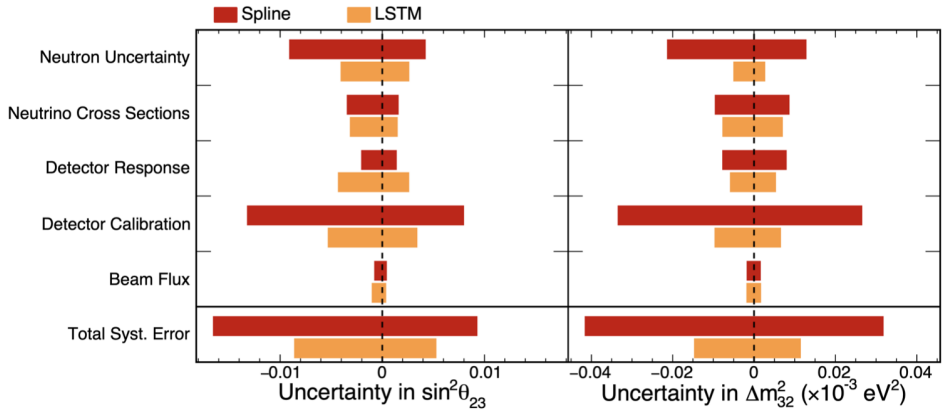


Figure 6.8: Impact of systematic uncertainties on the $\sin^2\theta_{23}$, Δm^2_{32} sensitivities with ν_μ-only data. Uncertainties for the (red) spline-based and (orange) LSTM EEs are shown for comparison. Detector response includes the light-level and Cherenkov systematic uncertainties.

The LSTM is an energy estimator for ν_μ ($\bar{\nu}_\mu$) events only; therefore, the sensitivity studies will be limited to the ν_μ disappearance channel. As discussed in chapter 2, the ν_μ channel is most sensitive to the Δm^2_{32} and $\sin^2\theta_{23}$ oscillation parameters and is relatively independent of δ_{CP}. Thus I will only consider the sensitivities to Δm^2_{32} and $\sin^2\theta_{23}$. When we evaluate the Asimov sensitivities for the LSTM and spline-based EEs, Figure 6.7, the benefit of the increased energy resolution is immediately obvious. Additionally, if we

investigate the impact of specific systematic uncertainty groups, then we see from Figure 6.8 that the LSTM EE is more robust in nearly all cases. The LSTM is less impacted by alterations to the hadronic prongs due to the neutron systematic and even shows some marginal improvement in response to the neutrino cross section uncertainties. The greatest improvement is to the calibration uncertainties, as we expected from applying the noise injection procedure. Finally, we see that the impact of the light-level and Cherenkov uncertainties (grouped as "Detector Response" in Figure 6.8) are slightly increased. This is likely due to an increased reliance on lower level inputs like the prong direction and number of hits that is associated with a decreased dependence on calibration-related inputs. As the light model is changed some simulated hits will go below or above the detection threshold, thus relying more on lower level reconstructed quantities will cause these systematic changes to become more important.

The impact of the LSTM EE can be further quantified by determining how much additional data would be required for the spline-based EE to produce comparable sensitivities. The results of this are shown in Figure 6.9. In summary, we see that the LSTM EE provides the equivalent sensitivity of 10–20% more data regarding the measurement of $\sin^2\theta_{23}$ and a much greater sensitivity to Δm^2_{32} than a 30% increase in data would provide with the spline EE.

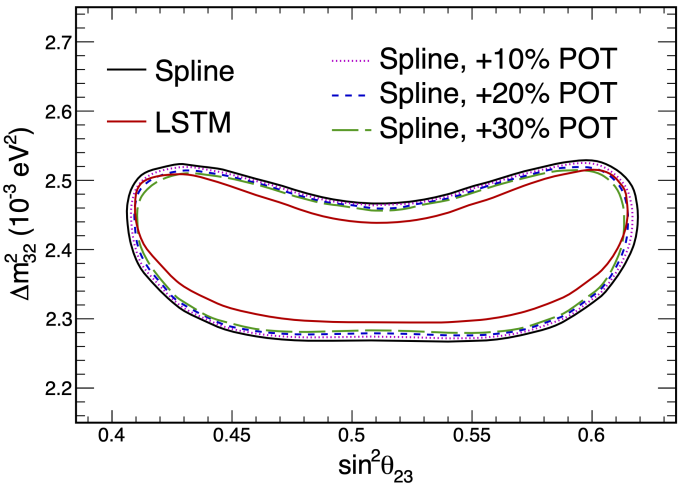


Figure 6.9: Asimov 90% sensitivity contours of Δm^2_{32} versus $\sin^2\theta_{23}$ with ν_μ-only data for the (black) spline-based and (red) LSTM EEs. The spline EE contours are also shown with 10%, 20%, and 30% more data.

7. Domain Generalization by Adversarial Training

I've mentioned multiple times throughout this thesis but it bears repeating: our models[1] are imperfect and at best represent one possible approximation of the real physical world that we are observing. The potential impact of this is evident in the necessary reliance on simulation in order to constrain possible background and ultimately make a measurement of the neutrino oscillation parameters. Ideally, we choose the most physically motivated arrangement of models and procedures as our nominal MC in order to most closely mirror reality. However, the fact remains that we must choose some imperfect representation and use that to develop our various analysis tools. From clustering to PID and energy estimation, the choice of model pervades the entire analysis chain. We attempt to bracket our ignorance by developing systematic uncertainties to cover many possible situations and we quantify the impact of model variations by propagating changes through the entire analysis. In developing new reconstruction tools, we often check their robustness against certain relevant systematics to ensure some level of model independence, with any remaining effects being passed into the oscillation fit as nuisance parameters.

In the machine learning literature, each model, and even the collected data, is referred to as a "domain". Often in machine learning you have the benefit of collecting real world data which you can then use to train your neural networks. Here, however, this is not an easy route since we must use our collected data to perform the measurement of oscillation parameters. While it is possible to collect large samples of real tagged particles, this has not been performed yet for NOvA and with the small cross section of neutrino interactions it is not feasible to obtain an independent, unbiased sample of fully representative event topologies. Therefore, we must train our algorithms on a "source domain" that does not quite mirror the "target domain" of our collected data.

The resolution of this issue of training a network on one domain (or domains) and using it on another is the main goal of "domain generalization" and "domain adaptation". The main difference between adaptation and generalization is that in adaptation you collect some amount of data from the target domain, either labeled or unlabeled, which is then

[1] Here I'm using the term "model" not only to refer to neutrino interaction models but also other procedures and components like calibration, light model, detector aging, etc.

directly used in the training process to adapt the network to the specific target domain, while in generalization you employ one or many source domains without any access to a target domain in order to generalize the network to some out-of-distribution domain [125]. In NOvA we sit squarely in the realm of domain generalization; however, there is some crossover in the techniques used to either adapt or generalize. Here I present a method for domain generalization applied to the ν_μ LSTM energy estimator.

7.1 Adversarial Training

There are many methods available to achieve domain generalization; a nice survey can be found at Ref. [126]. In fact, the noise injection technique discussed in the last chapter is just such a technique classified by Wang *et al.* as "data manipulation" [126]. Here we investigate an adversarial approach to domain-invariant representation-based learning devised by Ganin and Lempitsky for domain adaptation [127]. Kevin Mulder, a NOvA collaborator, has investigated applying this method to the CNN PIDs.

A typical neural network consists of inputs, some feature extractor layers (G_f), label predictor or regression layers (G_y), and outputs. Ganin *et al.* modify the network architecture by adding on a domain classifier that attempts to determine which particular source domain a given input belongs to [127]. They do this by introducing the domain classifier (G_d) after the feature extraction and in parallel with the main predictor. In between the domain classifier and feature extractor they place a gradient reversal layer (GRL) which passes inputs through unchanged during the forward pass through the network but flips the sign of the gradient during backpropagation. The total loss function becomes the sum of the losses from G_y and G_d. Since the gradient of G_d is flipped, the domain classifier acts adversarially on the feature extractor to push it away from learning things that can distinguish between domains. Attached to the GRL is a hyperparameter, λ, which is multiplied by the gradient to increase or decrease the adversarial strength of the domain classifier. The weights associated with the predictor, θ_y, and the domain classifier, θ_d, are updated normally according to the process of backpropagation. The particular usage of the gradients varies by the backpropagation method used, but taking stochastic gradient descent[2] as an

[2]Other optimization algorithms will generally look quite similar but with some function, usually depending on the gradient, also being multiplied by the second term along with the learning rate.

example, we have:

$$\theta_y \leftarrow \theta_y - \mu \frac{\partial L_y}{\partial \theta_y} \tag{7.1}$$

$$\theta_d \leftarrow \theta_d - \mu \frac{\partial L_d}{\partial \theta_d} \tag{7.2}$$

where μ is the learning rate and $L_{y,d}$ is the gradient of the main predictor or domain classifier loss functions. The feature extractor is updated in a modified manner:

$$\theta_f \leftarrow \theta_f - \mu \left(\frac{\partial L_y}{\partial \theta_f} - \lambda \frac{\partial L_d}{\partial \theta_f} \right) \tag{7.3}$$

The technique of Ganin *et al.* utilizes two inputs: the source domain, and unlabeled data from the target domain. This has been expanded on by Zhao et al. to include multiple source domains by creating multiple classifiers each between the target domain and a given source domain [128]. It bears repeating that in NOvA we do not have access to data from our target domain and instead must rely solely on various source domains to train our algorithms. However, the same principles from the domain adaptation methods by Ganin *et al.* and Zhao *et al.* still apply here. If the various source domains can bracket the possible ranges of the target domain, then they can be employed to train the network away from particular domain distinguishing features and toward an invariant representation that is applicable across many possible domains.

The particular implementation of the adversarial domain generalization used here for the LSTM EE is inspired by that of Tonutti *et al.* [129], termed a domain adversarial neural network (DANN), with some modifications[3]. The GRL is inserted parallel to the single-node regression dense layers of the LSTM EE, after the post-processing layers. The GRL then connects to a label-predicting dense layer which aims to determine the domain identifier. This implementation was built to accommodate any number of source domains, so it utilizes the softmax activation function to perform classification of multiple labels. The softmax activation function is paired with the categorical crossentropy as the loss function for the domain classifier. No additional inputs are needed beyond those used for the standard LSTM EE, though the training data must have domain labels attached toward which the domain classification layer will train. The input data are thoroughly shuffled before training to ensure they are not sequential in detector time and that events from multiple domains are

[3]The GRL code can be found in Appendix E

randomly interwoven.

I've applied the adversarial method to create domain generalized neural networks for two sets of training samples with the goal of reducing the impact of leading systematic uncertainties. To investigate the ability to reduce uncertainties associated with the neutrino interaction models, I've used the nominal GENIE simulation along with simulated events coming from the NEUT interaction generator in a two-source-domain approach. In contrast to the noise injection method, to generalize against the calibration systematic uncertainties, I've used the calibration up and down special samples along with the nominal calibration samples to train against three source domains.

7.2 Neutrino Interaction Generator Samples

Some time has been spent in Chapters 3 and 5 discussing the neutrino interaction models employed via the GENIE generator, and we saw that related uncertainties are among the leading systematics limiting the precision of our measurements. The significant impact of these systematic uncertainties is completely reasonable considering the wide choice of models, each generally focusing on predictions in some particular energy range, coupled with the inherent messiness of the nuclear landscape and final state interactions. Neutrino interaction models provide a natural place to employ adversarial domain generalization since so many possible source domains exist. I've chosen two source domains, one being NOvA's setup of the GENIE interaction generator and the other being T2K's setup of NEUT version 5.4.1.

NEUT is a neutrino interaction generator that was originally developed in the 1980s for the Kamiokande experiments in Japan (e.g. Super-K, T2K, and the upcoming Hyper-K) [130]. It's code was originally implemented in FORTRAN, and though some of it has been updated, a significant portion of the original FORTRAN code is still in use. There were two possible routes to implement NEUT for use in NOvA's framework. The first would be to fully model the NOvA detectors and plug the G4NuMI results into NEUT. The other option is to piggy-back on the framework already provided by the implementation of GENIE. In the second case, events generated via GENIE would be regenerated with NEUT in a process dubbed "NEUTRegen". The outputs from either method can then be plugged back into the NOvA simulation chain as if the events were generated by GENIE. NEUT version 5.4.1 was graciously provided by Hayato-san and Luke Pickering, both members of the T2K collaboration.

It was decided to pursue the NEUTRegen method since it allows us to utilize the existing G4NuMI-to-GENIE interface in order to produce NEUT events, as opposed to creating a brand new interface. Essentially, GENIE has complete knowledge of the beam flux and detector geometry from which it can make the decision on when and where a neutrino interacts and what target nucleus it interacts with. The neutrino momentum, direction, energy, and flavor are fed into NEUT along with the target nucleus and interaction current (CC or NC). NEUT then decides which interaction mode to use and generates the outgoing particles and event kinematics. The output from NEUT is then reformatted into the expected output that the NOvA framework expects and can be propagated like normal. The primary flaw with this method as compared to fully plugging NEUT into the NOvA framework is that it relies on GENIE to decide whether the neutrino interacts with a given target nucleus. This is unfavorable because the overall cross sections calculated by NEUT and GENIE may differ. To rectify this, we calculate the cross section ratio between the two interaction generators and apply that as an event weight. In this way we produce the event distributions that NEUT would create if it were implemented as a standalone generator.

T2K typically uses NEUT to simulate interactions on molecules like CH_2 or H_2O while in the NEUTRegen implementation I will be feeding it single nuclei. For the most part, there is no issue with using single nuclei except when the target is hydrogen, because NEUT was not designed to handle single free protons. In general, the interaction cross sections for NEUT and GENIE should be quite close to unit,m with some discrepancies arising from the use of different datasets and the particular implementations. However, while studying the cross section ratios, I found a significant discrepancy that only appeared for NC events on a hydrogen target, as can be seen in Figure 7.1. This discrepancy was skewed much more toward the high-energy DIS region, which is of much less concern to T2K since they operate at a beam energy around 0.6 GeV putting them nicely in the QE region.

Resolving the hydrogen issue ended up being relatively straightforward (after understanding the FORTRAN code and digging up old papers). In multi-pion production DIS interactions, NEUT calculates the neutral current cross section by first using a pre-calculated table to determine the charged current cross section then applying a NC/CC ratio pulled from Kim *et al.* [131]. The original NEUT code applies the NC/CC ratios from DIS on isoscalar targets, and the given ratio depends on the energy of the interacting neutrino and whether you are looking at neutrinos or antineutrinos. Reference [131] also reviews measurements of neutrino interactions with free protons and neutrons, which yield different ratios than those for isoscalar targets. From their results I extracted the ratios shown in

neutrino energy	isoscalar	free nucleon
< 3 GeV	0.26 (0.39)	0.43 (0.44)
> 6 GeV	0.30 (0.37)	0.50 (0.42)

Table 7.1: NC/CC ratio for neutrinos (antineutrinos) interacting with isoscalar or free nucleon targets at low and high incoming neutrino energy. Data drawn from Ref. [131].

Table 7.1 and use linear interpolation for the region $3 < E_\nu < 6$ GeV. Implementing this fix almost entirely removes the NC hydrogen DIS discrepancy between NEUT and GENIE, with any residual differences being similar to all other targets.

NEUT samples were generated using the NEUTRegen method, with these events being passed through the same standard NOvA reconstruction chain as the GENIE samples used in analysis from chapter 5. It was unfeasible to produce the same number of NEUT events as the main GENIE production, since their use case is, thus far, limited to this thesis and I would be the sole person tasked with producing these files. Instead, a smaller production

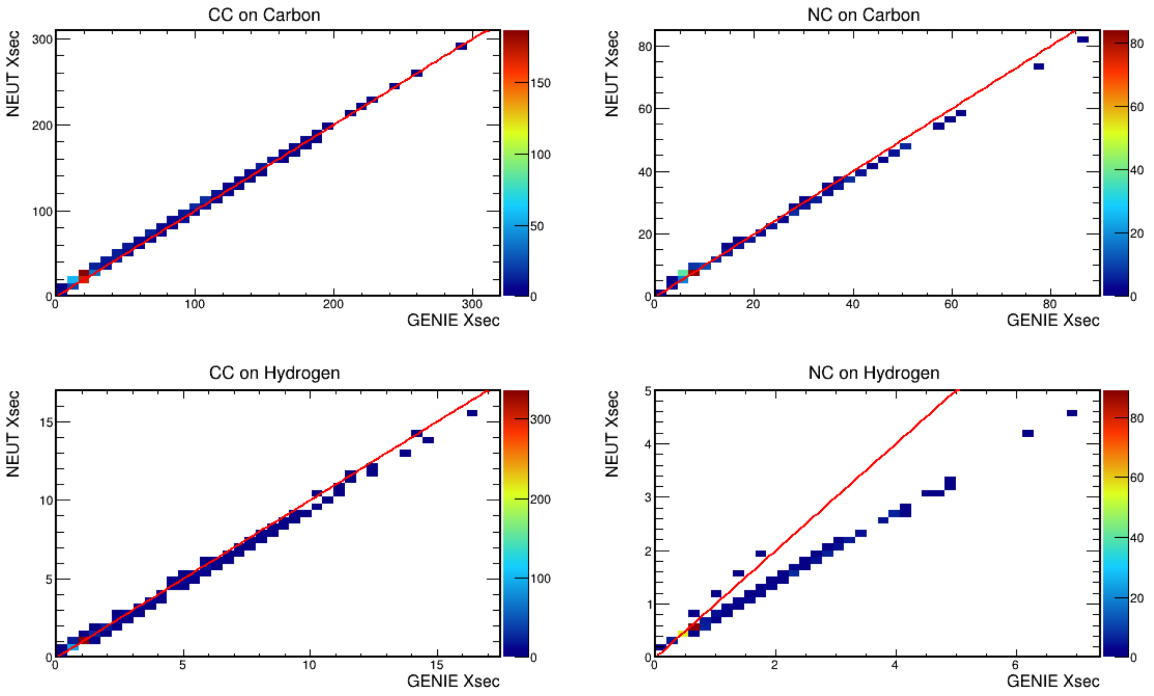


Figure 7.1: Comparisons of cross sections, in units of 10^{-38} cm^2, from NEUT and GENIE calculated for the same events. Left plots show CC interactions right plots show NC. Top plots are for interactions on Carbon and bottom are for Hydrogen. The red lines are $y = x$ showing the ideal line if the cross section from NEUT and GENIE matched exactly.

campaign was pursued, similar in size to campaigns performed by NOvA-at-large, which were used to test various modifications and train algorithms. The resulting datasets consist of about 1×10^6 events each for FD neutrino and antineutrino beam modes and about 2×10^5 events for each beam mode at the ND. A 30%:70%, testing:training split was used for all samples. The simulated events are distributed proportionally across each of the beam running conditions. Similarly distributed samples were pulled from the nominal NOvA GENIE simulation so as to have approximately equal sample sizes for both domains.

7.2.1 NEUT versus GENIE

With NEUT patched to handle NC interactions on hydrogen and NEUTRegen implemented, we can now investigate differences between NEUT and GENIE. Both interaction generators are, of course, attempting to represent the same underlying physics and indeed employ the same models in some cases.

Table 7.2 lists the models used by each neutrino interaction generator. From that table we see that for many interaction modes both NEUT and GENIE utilize the same underlying models with some small implementation differences, although in a few cases NEUT opts for different models or custom implementations built from multiple sources.

Table 7.2: A summary of the various models used by the GENIE and NEUT neutrino interaction generators.

Interaction	GENIE	NEUT
QE:	València using a local Fermi gas (LFG) [62, 63].	Benhar *et al.* using spectral functions (SF) [132] for certain nuclei (^{1}2C, ^{1}6O, and ^{5}6Fe) otherwise a relativistic Fermi gas is used.
2p2h:	València LFG [62, 63].	València LFG [62, 63].
RES:	Berger-Sehgal [64, 65] with axial mass of 1.12 GeV from Kuzmin *et al.* [133].	Berger-Sehgal [64, 65] with axial mass of 0.95 GeV from Graczyk-Sobczyk [134].

COH:	Berger-Sehgal [64, 65] with axial mass of 1.0 GeV.	Berger-Sehgal [64, 65] with axial mass of 1.0 GeV.
DIS:	AGKY model [56] for $W < 3$ GeV and Pythia v6.4 [135] for $W > 3$ GeV both using Bodek-Yang corrected parton distribution functions from GRV98 [66, 136].	Custom implementation [137] for $W < 2$ GeV and Pythia v5.7 [138] for $W > 2$ GeV both using Bodek-Yang corrected parton distribution functions from GRV98 [66, 136].
FSI:	hN semi-classical intranuclear cascade [68] with pion interactions from Salcedo *et al.* [139].	Custom intranuclear cascade model [130] with pion interactions from Salcedo *et al.* [139].

One could consult each of the references in Table 7.2 to compare each of the models used by the two interaction generators. However, I find it more instructive to simply use each generator and compare the results of certain interesting quantities. From Figure 7.2 we see that all of the cross section ratios are "reasonable"; that is, all cross section ratios are nearly equal to one. In both CC and NC interactions, the higher ratios come from events with lower interaction energies, while at high energies the ratios converge to about 1 for CC interactions and 0.95 for NC.

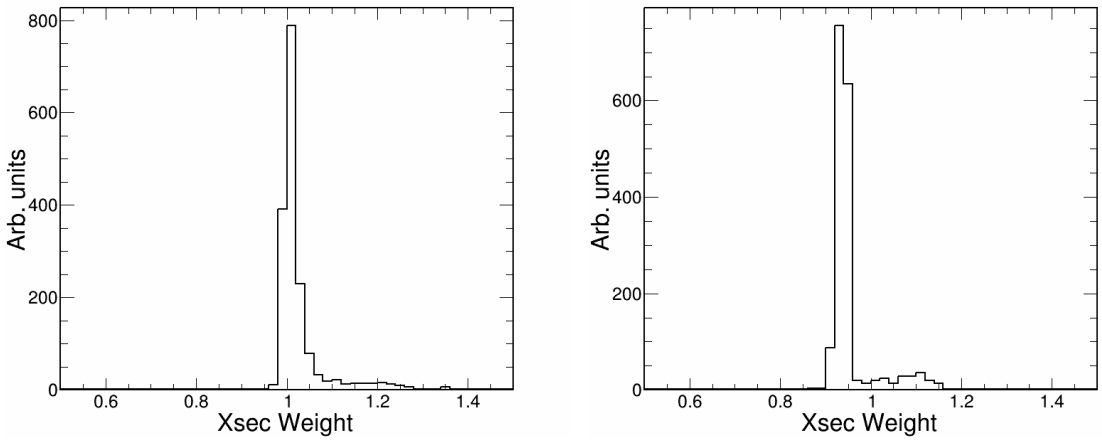


Figure 7.2: NEUT/GENIE ν_μ cross section ratios for all possible targets and incident neutrino energies between 0–30 GeV for (a) CC interactions and (b) NC interactions.

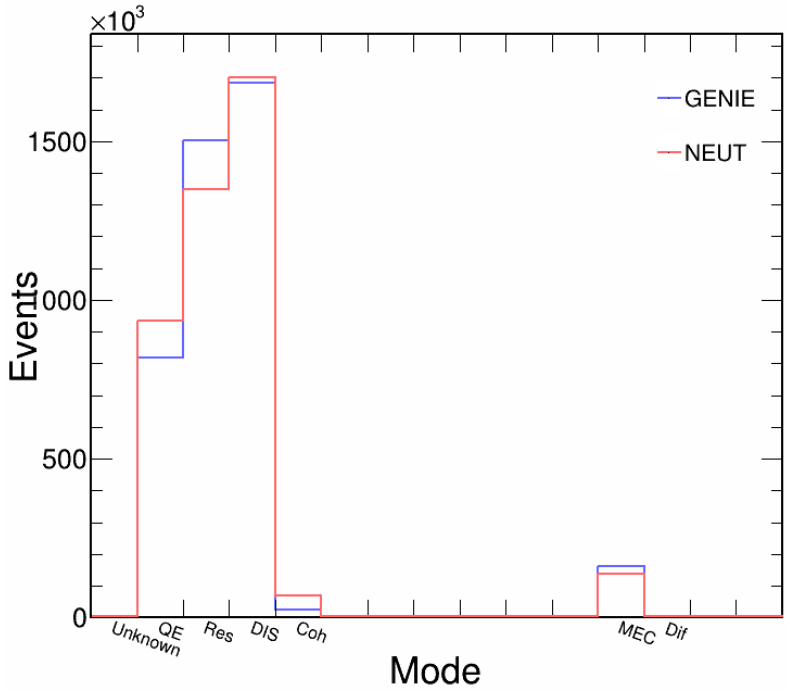


Figure 7.3: Unweighted comparison of interaction modes produced by (blue) GENIE and (red) NEUT at the far detector in the neutrino beam mode.

All of the following plots will show NEUT versus GENIE comparisons without the cross section ratio weights applied so as to have the event rates normalized between the two generators. From Figure 7.3 we see that NEUT has a preference for QE interactions at the expense of both resonance and coherent pion production as compared to GENIE. The rates of DIS and 2p2h interactions are comparable between both generators. This, of course, has knock-on effects when it comes to the kinematics of the generated events. Changes in these kinematic variables will then go on to alter some of the relevant inputs used by the LSTM EE.

Owing to the preference of QE interactions, NEUT produces fewer final state particles on average than GENIE (Figure 7.4a). Digging further into the production process, I found that GENIE also has a preference of generating large numbers of intermediate state particles, typically of low energy, while NEUT produces a comparatively small number. We also see, from Figure 7.4b, that NEUT prefers to give more energy to the outgoing lepton. This observation is borne out by the lower observed momentum transfer and hadronic invariant mass (Figures 7.4c and 7.4d respectively). NEUT also produces leptons that are more

forward-going and the fact that NEUT gives more energy to the outgoing lepton also means that the number of reconstructed hadronic hits will be reduced.

We have seen that even though NEUT and GENIE utilize the same or similar interaction models, they still produce some appreciable differences in the distributions of important kinematic variables. Now we can move on to comparisons of the various reconstructed quantities to see how these truth differences translate into the utilized inputs for the LSTM EE. From the plots in Figure 7.5, we can see some differences between NEUT and GENIE in the "important" input variables.

None of the reconstructed differences are nearly as large as those seen in the kinematic variables of Figure 7.4, however, there are some features of note. As expected, NEUT events have slightly lower reconstructed `calE` and `remPngCalE` (Figures 7.5a and 7.5b). While we see more event hits overall (Figure 7.5c), these can be attributed to an increase of lepton hits that is greater than the decrease seen in hadronic hits. A similar effect is seen in the 3D prong `calE` (Figure 7.5d), where we see fewer events at the middle energy values and an increase in events at both high energies, for the leptons, and low energies, for the hadrons. Finally, as expected we see that the 3D prongs are more forward going (Figure 7.5e) and in general the lengths of the 3D prongs are slightly larger (Figure 7.5f), owing to the increase in energy given to the outgoing leptons. The 2D prong inputs follow the same pattern as the 3D prongs. Even with the differences in inputs between the NEUT and GENIE domains being somewhat small, in combination across the many inputs used for each event, there may be some ability to discriminate between domains. If this is the case, then adversarially training out this ability could lead to a more domain-generalized neural network.

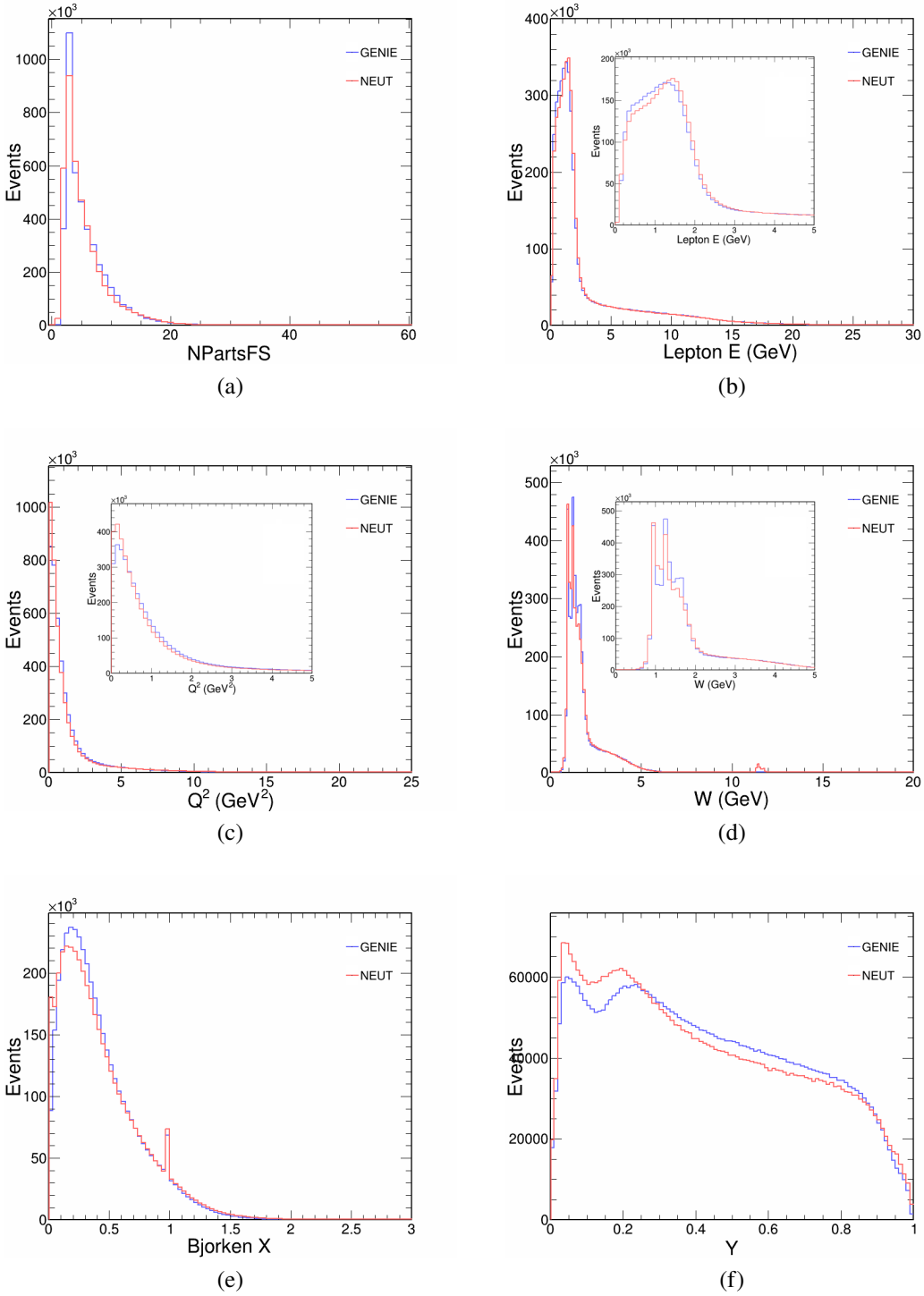


Figure 7.4: Unweighted comparison of truth variables for FD neutrino mode events between GENIE in blue and NEUT in red. (a) Number of final state particles, (b) outgoing lepton energy, (c) invariant four-momentum transfer Q^2, (d) invariant hadronic mass W, (e) Bjorken x, and (f) neutrino fractional energy loss y.

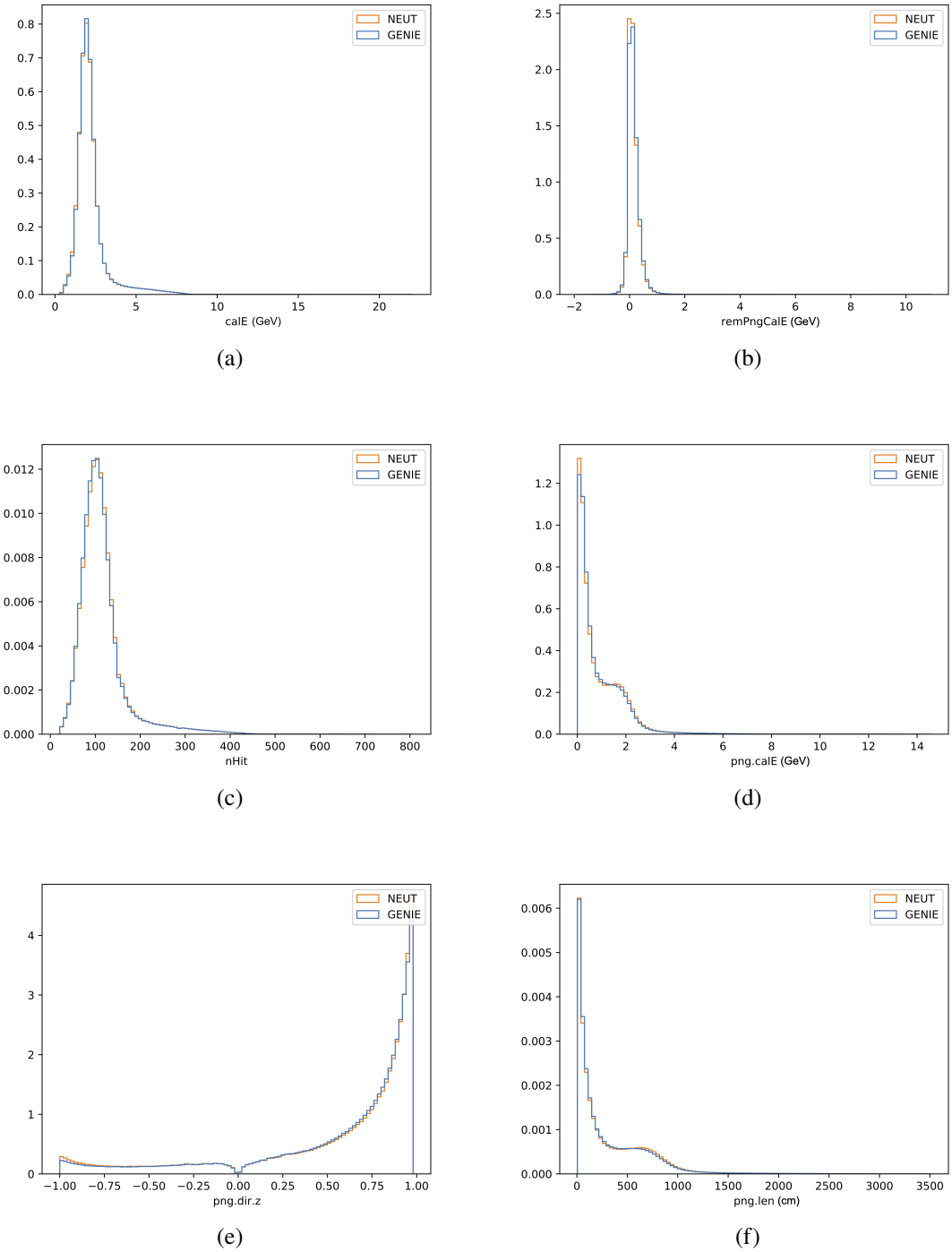


Figure 7.5: Area normalized comparison of reconstructed variables for FD neutrino mode events between GENIE in blue and NEUT in orange. (a) event `calE`, (b) event `remPngCalE`, (c) number of hits in the event, (d) 3D prong `calE`, (e) cosine of the 3D prong angle with respect to the z-direction, and (f) length of 3D prongs.

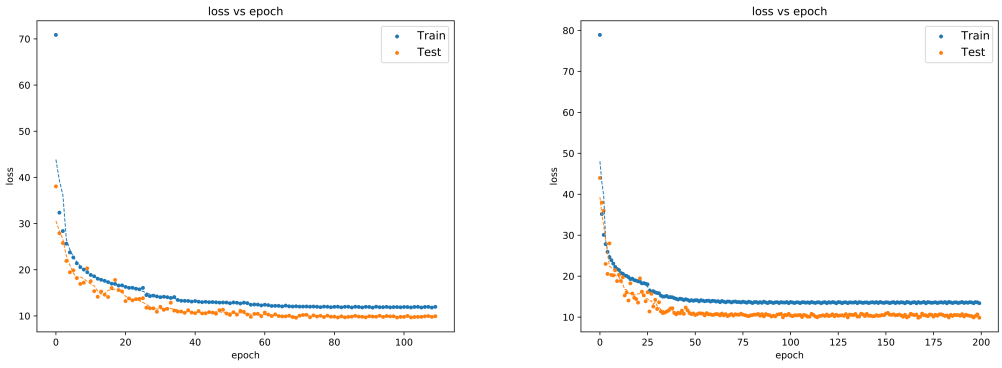
7.2.2 Results

Recall that the gradient reversal layer has an associated hyperparameter, λ, which controls the influence of the adversarial training. Higher values of λ will generally result in a more generalized network at the expense of performance, while negative values can be used in order to essentially de-generalize the network. Multiple values of λ (-1, 1, 2, 5, 10, 20, 50, 100) were tested in order to determine which level achieves sufficient domain generalization without an excessive hit to the performance. The value of λ was hand-tuned by choosing the the lowest value for which the AUC was close to 0.5. Higher values of λ result in decreased energy estimation performance without an increase in generalization.

In addition to the domain adversarial (DA) LSTM EE, I also trained the base network on GENIE and NEUT events separately as well as a combined dataset using both interaction generators. As an initial proof that the adversarial training proceeded as expected, we can look at the training loss versus epoch[4] which converges similarly for the DA-LSTM as it does for the standard LSTM trained on GENIE events as shown in Figure 7.6.

Table 7.3 shows the overall performance of each of these networks along with two DANN λ settings. Immediately we see that all trainings produce relatively similar performances, with the λ = 100 DA-LSTM performing only slightly better than the base GENIE network.

We can further interrogate the behavior of the adversarial training by looking at the receiver operating characteristic curve (ROC) and the area under the curve (AUC). The



(a) (b)

Figure 7.6: Train and test loss as a function of epoch for (a) the LSTM EE trained on only GENIE events and (b) the DA-LSTM EE trained with λ = 100.

[4]Epochs are essentially training time. For all of the networks shown here, 200 epochs were used and the training process was allowed to exit early once the improvement from epoch-to-epoch became negligible.

Training	Energy Resolution (%)	Bias ($\times 10^{-3}$)
GENIE	9.37	-13.74
NEUT	9.49	-4.33
Both	9.51	-6.01
DANN $\lambda = -1$	9.43	-10.14
DANN $\lambda = 100$	9.33	-4.54

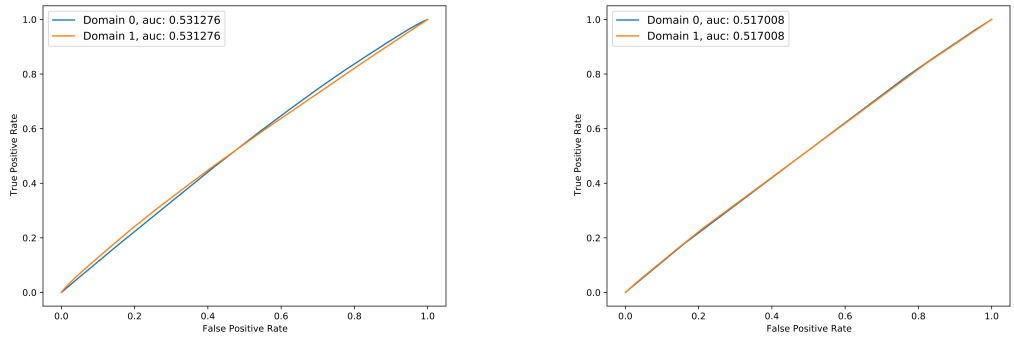
Table 7.3: Energy resolution and bias (RMS and mean, respectively, of the energy resolution distributions) for various LSTM EE versions.

ROC curve illustrates the performance of a classifier by comparing the rate of true versus false positives [140]. To produce these curves, the classification acceptance threshold is varied from zero to one and the classifier is tested against a sample to determine the rate of true positives and false positives for each threshold. An AUC of 1 corresponds to a perfect network which can always distinguish between the domains, while an AUC of 0.5 means that each decision is essentially just the flip of a coin.

As a baseline we use the DA-LSTM with $\lambda = -1$, which we take as the ability of the standard LSTM EE to distinguish between the two domains since, with a minus sign, it will de-generalize but with $\lambda = 1$ it will not contribute much to alter the weights of the feature extraction. Comparing the ROC and AUC of the $\lambda = -1$ DA-LSTM to the other λ settings allows us to see how domain invariant we can make the network and determine which λ yields the most generalized network. The ROC curves of $\lambda = -1$ and $\lambda = 100$ are shown in Figure 7.7. From Figure 7.7a, we see that the base LSTM EE has some small ability to distinguish between NEUT and GENIE events, but this is relatively minimal with an AUC of 0.53. Increasing λ to 100, Figure 7.7b, the DA-LSTM is able to almost completely train this ability out of the network and drive the AUC slightly closer to 0.5. However, the gain here is very small because the NEUT-GENIE differences that were evident with truth level information were only marginally reflected in the relevant reconstructed quantities.

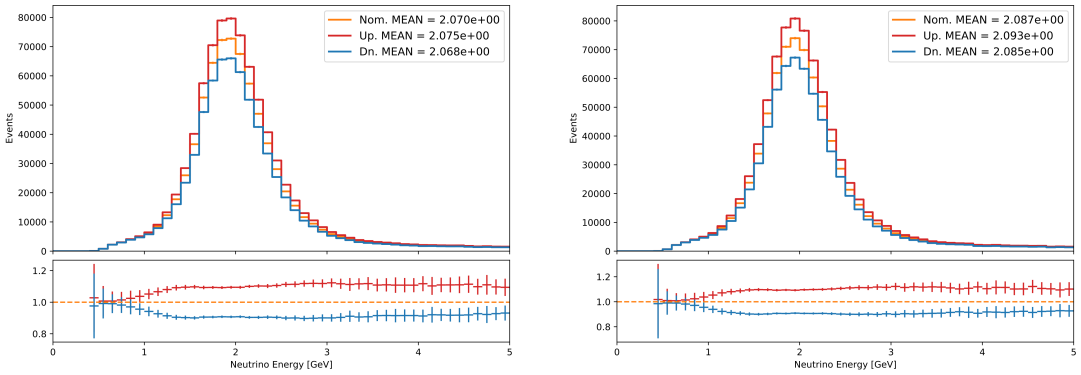
As a final comparison, we can look at the impact of the cross section systematic which we were attempting to constrain with the DA-LSTM training. Figure 7.8 shows the impact of the plus and minus 1 σ shifts in the uncertainty on the axial mass of RES interactions. We see essentially no difference in the impact of this systematic uncertainty between the normally trained, GENIE-only LSTM EE and the DANN version with $\lambda = 100$. Thus, with an initial AUC of 0.53, the NEUT and GENIE domains are not distinct enough to obtain any significant gains through adversarial domain generalization. We can conclude that in order to generalize the LSTM EE against the various domains associated with uncertainties in the

cross section tunes would require a more individually targeted approach particularly on those uncertainties which produce significant changes in the relevant reconstructed quantities.



(a) (b)

Figure 7.7: Comparison of the ROC curve and AUC for the (a) $\lambda = -1$ and (b) $\lambda = 100$ adversarially trained LSTM EEs on FD neutrino-mode samples. Domain 0 is GENIE and domain 1 is NEUT.



(a) (b)

Figure 7.8: Impact of systematic shifts to the axial mass of RES interactions and ratios to the nominal simulation for the (a) basic GENIE only trained LSTM EE and (b) the DANN version with $\lambda = 100$. The estimated energy of the nominal tune is shown in orange while the $\pm 1\,\sigma$ shifted distributions are shown in red and blue respectively. Plots are for unoscillated FD neutrino-mode samples.

7.3 Calibration Systematic Uncertainty Samples

In the last chapter we discussed the injection of noise as an method to generalize the LSTM EE against calibration domains. As a recap, random noise is injected into calibration-related inputs so that the neural network learns that these components are less reliable and thus we produce an energy estimator that is robust to shifts in the calibration. We know, however, that changes in calibration result in wide ranging effects that can impact many reconstructed quantities. These impacts are the reason why we produce separate calibration up and down samples in order to properly assess the related systematic uncertainties. What's more, changes will occur in nearly all reconstructed quantities, with more basic variables impacting higher level ones. This makes it impossible to fully capture the effects by simply reweighting these quantities after the fact.

We can also pursue domain generalization using the adversarial approach, which will allow us to fully capture all of the correlated changes that occur when the calibration is shifted. Here we have three training domains: the nominal calibration tune, the calibration up sample, and the calibration down sample. Since the adversarial domain classifier is implemented via the softmax activation function, the DA-LSTM will easily be able to handle more than two source domains.

As we did with the neutrino generator domains, let us now compare quantities from the three calibration domains. Each calibration sample uses the same underlying neutrino models, so all of the truth distributions will be the same; therefore, we will only compare interesting reconstructed quantities. We can immediately see the impact on the calibration-related quantities in Figure 7.9. These variables all shift as expected for the calibration up and down samples as compared to the nominal calibration. Figure 7.10 displays how the calibration can have a knock-on effect to certain basic reconstructed quantities. There we can see some minor variations in the reconstructed 3D prong directions and lengths which are likely due to the use of calibrated energy in the view matching algorithm for the fuzzyk prongs (the 2D prongs show almost no variation in direction or length as the calibration is shifted). Finally, in Figure 7.11 we can see how the scores from CNN_{png} change. The larges of these changes are at the low end of each particle score, though the log-scale allows us to see the variations that occur all along the CNN_{png} range.

Aside from the direct calibration-related quantities, most of the other reconstructed values have relatively small variations as the calibration is shifted. Deep neural networks in general, and the LSTM EE in particular, are complex multi-layered algorithms which can

produce non-linear responses that are able to pick up on these subtle changes. Therefore, properly handling all of the correlated changes during the domain generalization process can become relevant even if the variations are relatively minor. Samples from all three calibration domains were combined in equal proportion, with about 30% of the events being used as a testing sample and the remainder being the training sample. The FD neutrino and antineutrino beam modes each have training sample sizes of about 1×10^6 events while for the ND the training samples consist of 1×10^6 events for neutrino mode and 7×10^5 events for antineutrino mode.

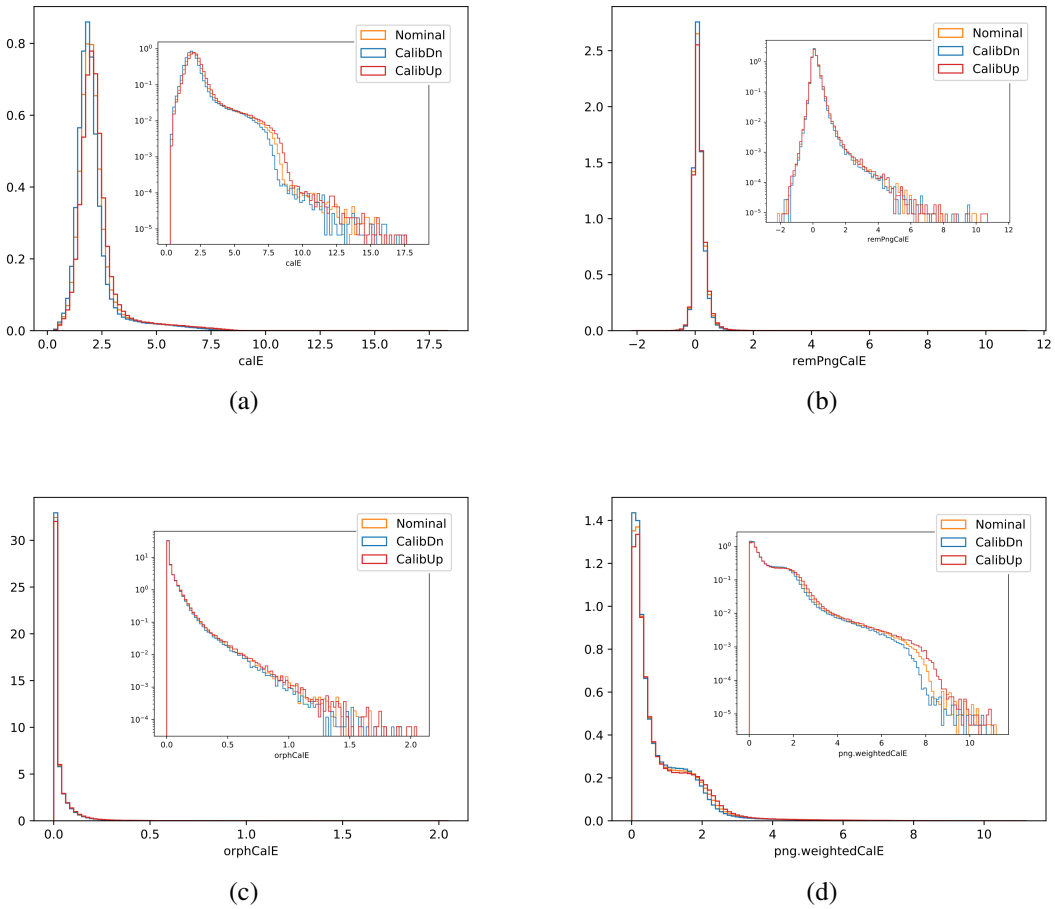


Figure 7.9: Area normalized comparison of calibration-related reconstructed variables for FD neutrino mode events between the nominal calibration in orange, calibration up in blue, and calibration down in red. (a) event `calE`, (b) event `remPngCalE`, (c) event `orphCalE`, and (d) 3D prong `weightedCalE`. Inset plots have a log scale along the y-axis to highlight the small variations associated with each domain.

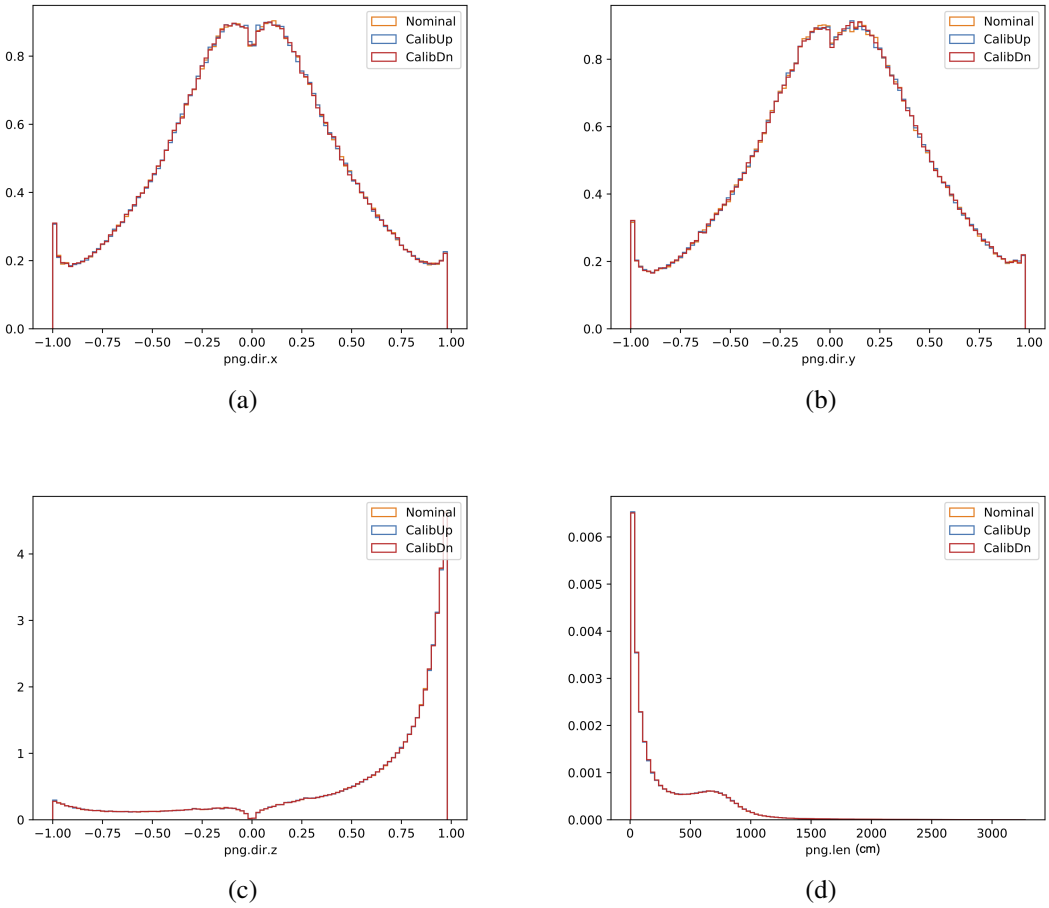


Figure 7.10: Area normalized comparison of reconstructed 3D prong variables for FD neutrino mode events between the nominal calibration in orange, calibration up in blue, and calibration down in red. Cosine of the 3D prong angles with respect to the (a) z-direction, (b) y-direction, and (c) z-direction, and (d) length of 3D prongs.

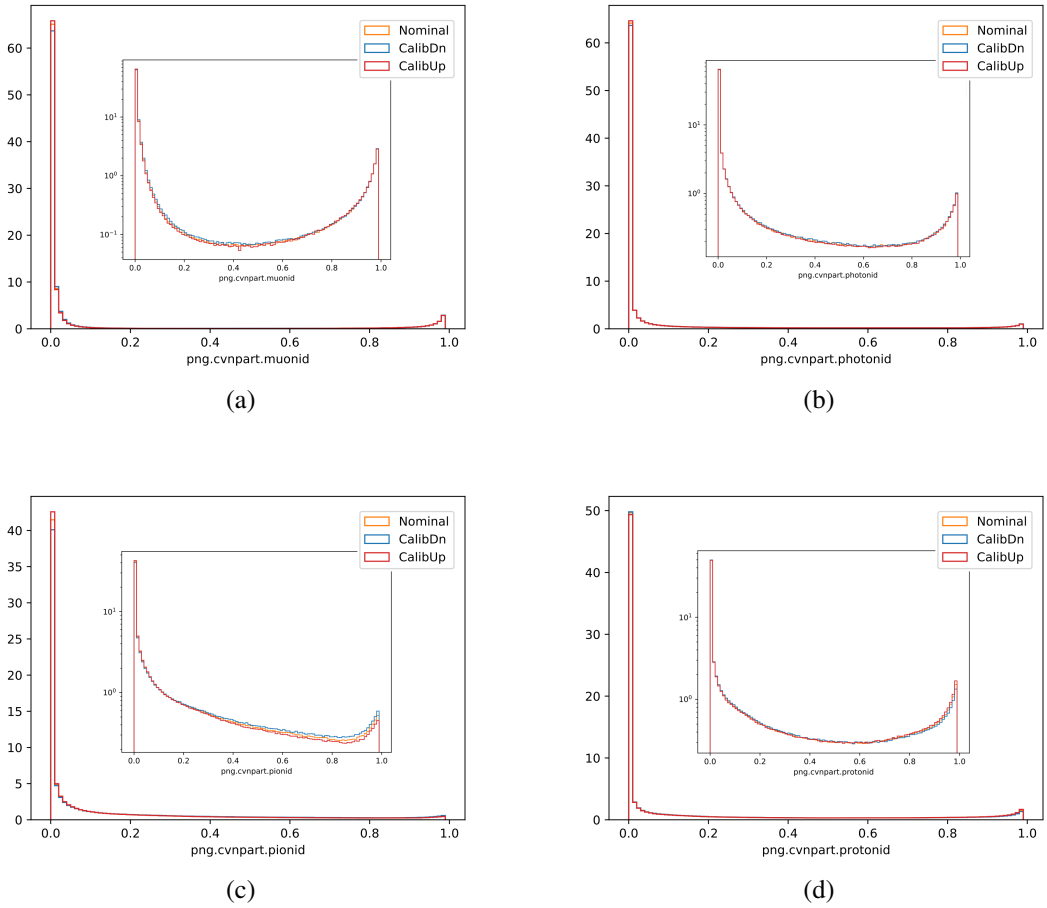
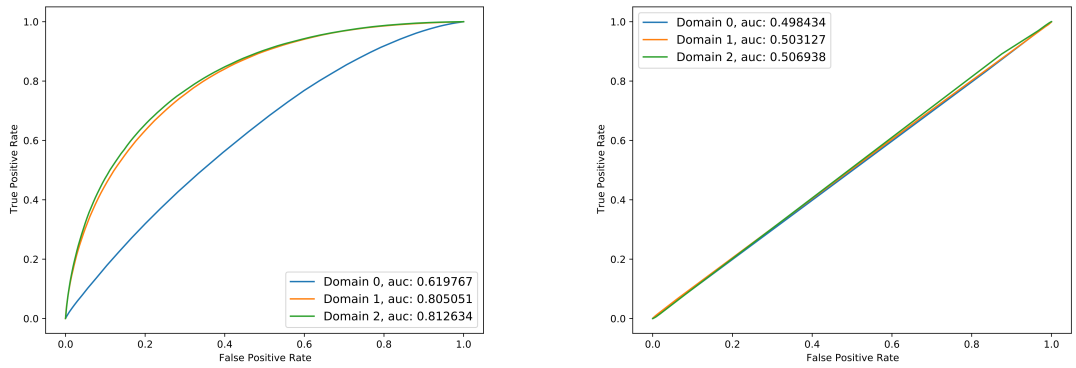


Figure 7.11: Area normalized comparison of CNN_{png} scores for FD neutrino mode events between the nominal calibration in orange, calibration up in blue, and calibration down in red. CNN_{png} scores for a (a) muon, (b) photon, (c) pion, and (d) proton. Inset plots have a log scale along the y-axis to highlight the small variations associated with each domain.

7.3.1 Results

Just as with the NEUT-GENIE DA-LSTM, the calibration DA-LSTM was trained with various λ settings (-1, 1, 2, 5, 10, 20, 50, 100) to determine the ideal value for this hyperparameter. A λ of 10 was found to provide almost complete domain generalization with the lowest impact on energy estimation performance. The ROC curves of the $\lambda = -1$ and $\lambda = 10$ DANNs are shown in Figure 7.12 and we see that the baseline network, represented by $\lambda = -1$ can distinguish between the shifted calibration domains and the nominal simulation. Interestingly, the network is better able to pick out events simulated with either the calibration shifted up or down than it is at identifying the nominal simulation. With $\lambda = 10$, the network is almost completely generalized between these three domains, yielding AUCs right around 0.5.

The performance of the DA-LSTM is shown in Table 7.4 along with the base LSTM trained on the nominal simulation with and without noise injection and a sample of events from all three domains trained without noise injection. The noise injected LSTM (NI-LSTM) trained on the nominal sample has the worst energy resolution while the DA-LSTM with $\lambda = 10$ has an energy resolution comparable to the other trainings and also yields the lowest energy bias. In all cases, the LSTM networks far outperform the standard spline-based EE. The other features of the DA-LSTM follows closely those discussed regarding the standard LSTM in Chapter 6, namely a flat resolution and bias across most of the energy



(a) (b)

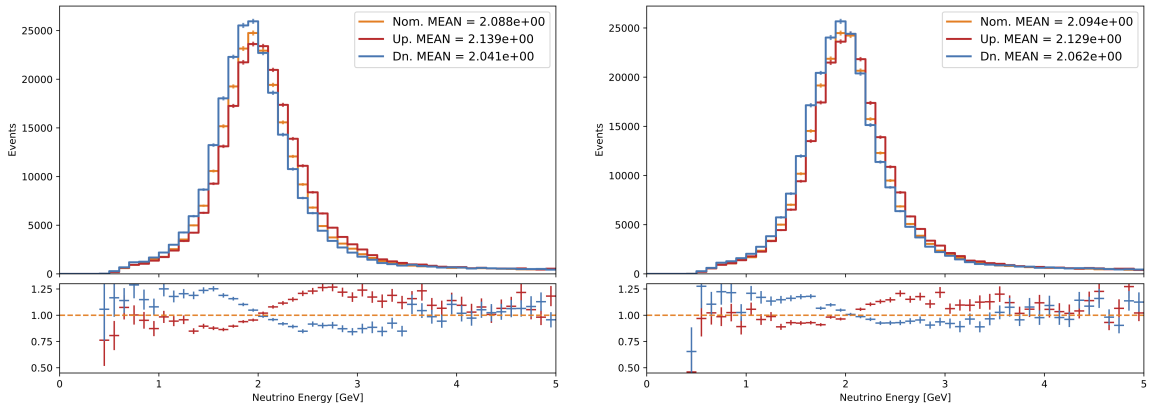
Figure 7.12: Comparison of the ROC curve and AUC for the (a) $\lambda = -1$ and (b) $\lambda = 10$ adversarially trained LSTM EEs on FD neutrino-mode samples. Domain 0 is the nominal simulation, domain 1 is calibration up, and domain 2 is calibration down.

Training	Energy Resolution (%)	Bias ($\times 10^{-3}$)
Spline EE	9.13	-14.96
Nominal Base LSTM	8.11	-4.33
Nominal NI-LSTM	8.83	-3.30
Combined Base LSTM	8.12	-6.39
DA-LSTM $\lambda = -1$	8.12	-6.78
DA-LSTM $\lambda = 10$	8.14	-1.02

Table 7.4: Energy resolution and bias (RMS and mean, respectively, of the energy resolution distributions) for the spline-based EE and the various LSTM EE versions on FD neutrino-mode samples. "Base" signifies that these LSTMs did not use noise injection on the calibration inputs. "Combined" is a sample containing events from all three calibration domains.

range under consideration.

Figure 7.13 compares the behavior of the nominal base LSTM and the DA-LSTM in response to the calibration systematic shifts. The DA-LSTM EE is more robust to these shifts, with the ratios of shifted to nominal being closer to unity than for the base LSTM EE. Comparing This method of domain generalization to the noise injection technique showed in Figure 6.5b reveals that the DA-LSTM is not as calibration-ignorant as the NI-LSTM. This is partially to be expected since the noise injection technique utilized a 20% discrete noise model while the calibration systematic samples only use a 5% shift in the calibration.



(a) (b)

Figure 7.13: Energy spectra and ratios of shifted to nominal calibration for the FD neutrino-mode (a) basic LSTM and (b) DA-LSTM evaluated on the (orange) nominal, (red) calibration up, and (blue) calibration down MC samples.

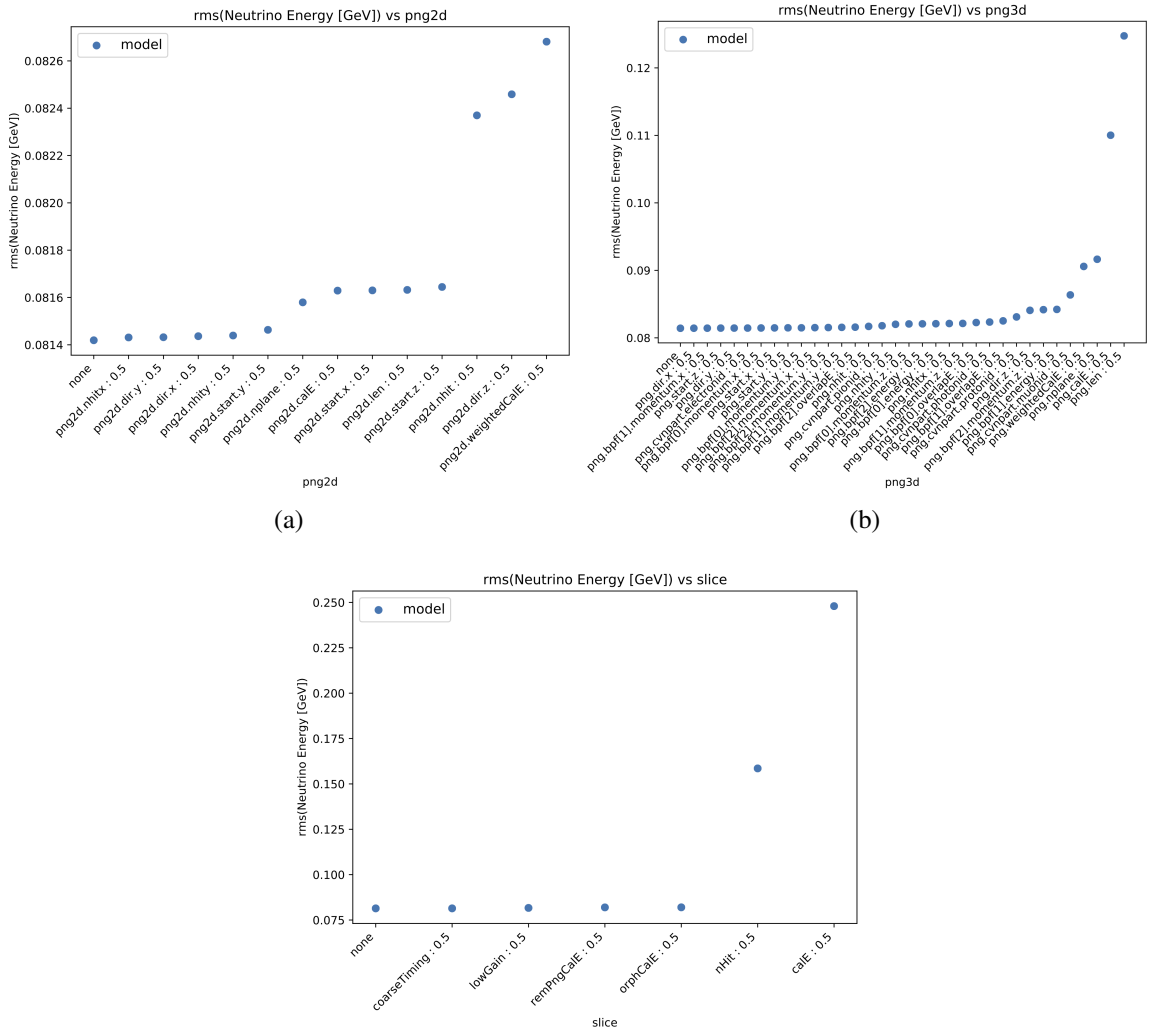


Figure 7.14: RMS of the energy resolution for the DA-LSTM EE trained and evaluated on far detector, neutrino-mode data as a function of perturbed input. Each input was modified by sampling a gaussian with a 50% spread. Prong 2D inputs are shown in (a), prong 3D inputs are in (b), and slice inputs are in (c).

The input importance ranking also provides an interesting story. Recall that we can determine the relative importance of the various inputs by perturbing each one individually and seeing how the performance (energy resolution) changes. Comparing Figure 7.14 to Figure 6.4 we see some shuffling in the order of the 2D and 3D prong inputs. For 3D prongs, `weightedCalE` drops from second to fourth most important and `nplane` moves up. For 2D prongs, `weightedCalE` is now the most important; however, the fourth through eighth

inputs now carry much more importance than they did with the baseline LSTM EE. The ranking of the slice inputs did not change much, but there are only six inputs anyway. With the slice inputs we do see a reduction in the impact of changes to `calE`; perturbations cause the energy resolution to increase to 25% with the DA-LSTM compared to 35% with the base architecture.

Table 7.5 shows the performance of the DA-LSTM EE compared to the spline-based EE for each combination of detector and neutrino beam mode. For all networks except ND neutrino-mode training, the DA-LSTM significantly outperforms the spline-based EE in both resolution and bias. The DA-LSTM also provides better energy resolution than the base NI-LSTM. Although the response to the calibration systematic is less robust for the DA-LSTM versus the NI-LSTM, adversarial domain generalization allows us to capture the changes that occur to all of the input variables as the calibration is shifted. Applying the adversarial technique with calibration up and down domains generated with shifts greater than 5% would likely yield robustness closer to the noise injection method while also properly handling all of the correlated variations across the reconstruction chain.

Energy Estimator		Resolution (%)	Bias ($\times 10^{-3}$)
FD	FHC	8.14 (9.13)	-1.02 (-14.96)
	RHC	7.17 (8.20)	-4.51 (-11.68)
ND	FHC	10.83 (10.82)	-1.28 (-24.0)
	RHC	9.33 (9.41)	-2.82 (-21.9)

Table 7.5: Energy resolution and bias (RMS and mean, respectively, of the energy resolution distributions) of the DA-LSTM EE and spline-based EE, in parentheses, for the FD and ND in neutrino (FHC) and antineutrino (RHC) beam modes.

As we did with the NI-LSTM architecture in the last chapter, we can evaluate the Asimov sensitivities and impact of systematic uncertainties when using the DA-LSTM and compare it to the spline-based EE. Again, we only consider the impact on Δm^2_{32} and $\sin^2\theta_{23}$ since we are dealing with a ν_μ energy estimator. From Figure 7.15, we see that the DA-LSTM provides greater measurement sensitivity to both $\sin^2\theta_{23}$ and Δm^2_{32} as compared to the spline-base EE. However, looking back at Figure 6.7, we see that it is still outperformed by the NI-LSTM. This is confirmed when we consider the systematic impacts shown in Figure 7.16. The neutron uncertainty is constrained similarly by the DA-LSTM and the NI-LSTM, while the DA-LSTM is actually slightly more robust against the detector response and cross section uncertainties. The DA-LSTM also reduces the impact of the calibration uncertainties when compared to the spline-based EE.

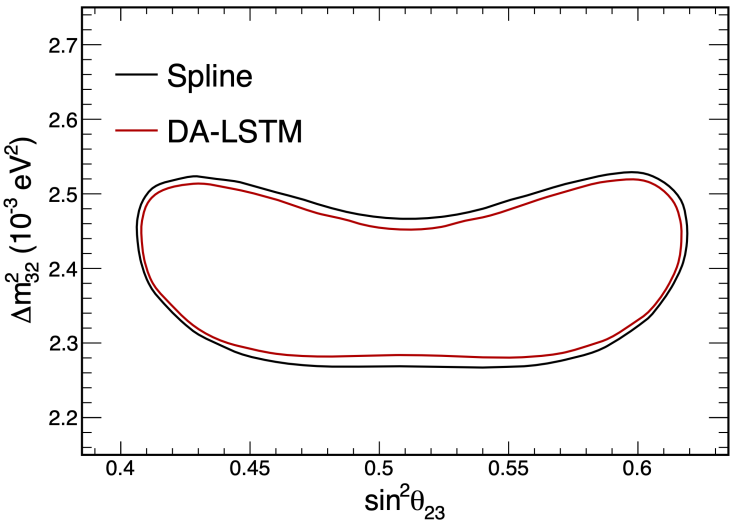


Figure 7.15: Asimov 90% sensitivity contours of Δm^2_{32} versus $\sin^2\theta_{23}$ with ν_μ-only data for the (black) spline-based and (red) DA-LSTM EEs.

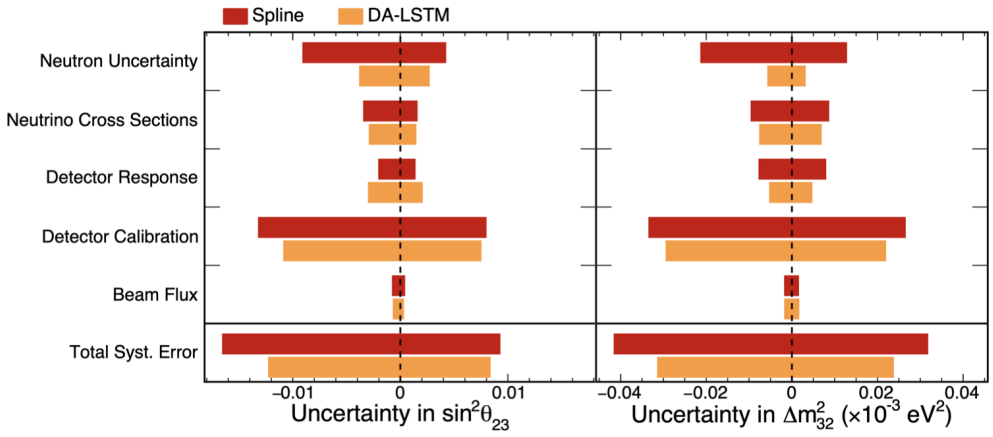


Figure 7.16: Impact of systematic uncertainties on the $\sin^2\theta_{23}$, Δm^2_{32} sensitivities with ν_μ-only data. Uncertainties for the (red) spline-based and (orange) DA-LSTM EEs are shown for comparison. Detector response includes the light-level and Cherenkov systematic uncertainties.

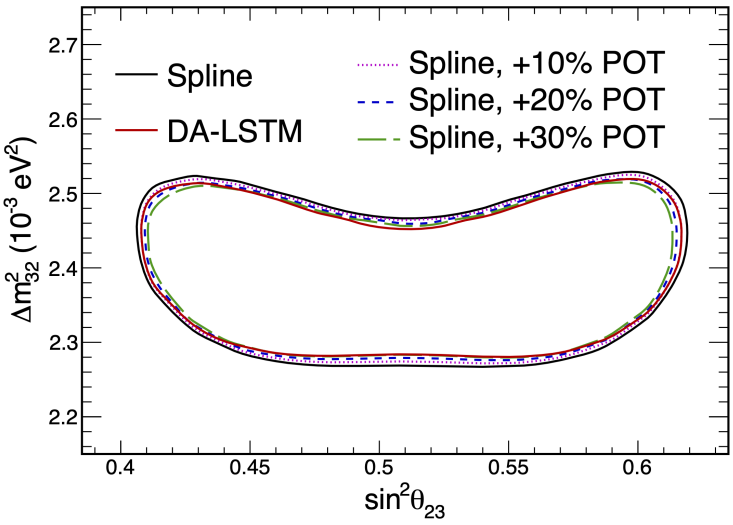


Figure 7.17: Asimov 90% sensitivity contours of Δm^2_{32} versus $\sin^2\theta_{23}$ with ν_μ-only data for the (black) spline-based and (red) DA-LSTM EEs. The spline EE contours are also shown with 10%, 20%, and 30% more data.

Finally, we can compare the impact of the DA-LSTM EE to the improvement we would obtain by increasing the size of our datasets while still employing the spline-based EE. From Figure 7.17,we see that the DA-LSTM EE provides the equivalent sensitivity of 10–20% more data regarding the measurement of $\sin^2\theta_{23}$, similar to the NI-LSTM. The sensitivity to Δm^2_{32} of the DA-LSTM is almost exactly equivalent to a 30% increase in the dataset while using the spline EE.

8. Conclusion

Adversarial domain adaptation and generalization techniques have been studied for LHC searches as well as in neutrino experiments; however, all of these cases have employed domain adversarial neural networks (DANNs) to classification problems [141, 142, 143, 144, 145]. I've explored the implementation of a DANN architecture added to an LSTM-based regression for ν_μ energy estimation. Two separate trainings were performed to generalize against neutrino interaction generators and the calibration systematic uncertainty.

The implementation of the NEUT neutrino interaction generator alongside GENIE opens up the possibility of creating new datasets that can be used to compare the generators used by the two leading long-baseline neutrino oscillation experiments. The particular GENIE and NEUT tunes explored here did not provide enough difference at the level of reconstructed quantities in order for the DANN to produce effective domain generalization. However, we now have the groundwork to employ not only additional GENIE and NEUT tunes as source domains, but this technique can be scaled to any number of various neutrino interaction generators. Implementing NEUT with the NOvA framework also opens up more possibilities for comparing the results of NOvA and T2K and can help to resolve the measurement tensions between the two experiments.

The second DANN was trained using the calibration systematic samples as the source domains and compared to domain generalization achieved through input noise injection. While the Δm^2_{32} sensitivity for the DA-LSTM generalized against the calibration systematic samples is not quite on par with the NI-LSTM, we did see some added benefits. First, because the DA-LSTM is trained on completely reprocessed samples, it is able to fully capture all of the effects associated with a change in the calibration. These changes propagate to modifications in prong directions and particle identification, which are difficult to replicate using the noise injection method. Second, the DA-LSTM is slightly more robust to the light model uncertainties. Finally, the comparison of the DA-LSTM and NI-LSTM is not quite fair since the noise injection method employed a discrete noise model with 20% variations, while the calibration systematic samples used for the adversarial training only have a 5% discrete variation from the nominal sample.

The choice of source models becomes particularly important in order to achieve a desirable level of generalization while maintaining the performance of the neural network.

As we saw with both the neutrino interaction generators and the calibration samples, DNNs have the capacity to pick up on very small, correlated changes to input variables and use that information to determine the source domain. Through the use of a DANN, we can remove the ability to effectively extract domain-dependent features, rendering the network more model-independent. However, care must be taken when using this method to reduce the impact of systematic uncertainties. In order to be most effective, the source domains should cover a wide range of model variations. Utilizing domain variations that are larger than the systematic uncertainty will better generalize toward a network that is ignorant to the comparatively small changes associated with that uncertainty. Domains with large variations also have the benefit of covering a larger swath of model space to keep us from falsely reducing the impact of a systematic uncertainty to a point that underestimates the true effect [146].

The adversarial domain generalization technique investigated here can easily be expanded to handle any and all models that bracket the central-value prediction. The majority of the literature focuses on only two domains; however, this can be expanded to any number of domains in various ways. One can use the simple method employed here by using a softmax activation for the domain classifier along with categorical crossentropy for the loss function, which is typical for multi-class classification problems [147]. More sophisticated loss functions can also be used which combine individual comparisons of each systematic domain against the nominal domain [128]. Further, techniques can also be used wherein independent networks are trained separately on each domain and then combined in a pair-wise manner [148].

We were restricted to domain generalization rather than adaptation in the cases we explored here. This is due to the fact that we do not have a large sample of target domain data that is independent of the neutrino events we are trying to measure. However, NOvA has been undertaking a Test Beam effort (Appendix F) which is seeking to collect an independent dataset of well-understood, tagged particles which could serve as a target domain. The addition of real data in the training procedure could help drive the LSTM EE to more strongly consider those event features which are common between data and simulation and ignore those that are mis-modeled.

With the prevalence of machine learning and DNNs in particle physics, it can be difficult to evaluate the amount of model dependence we have in training our algorithms. In many cases it is not feasible to collect large samples of target domain data, so we must rely on simulation to provide training samples. Adversarial domain generalization or adaptation

is a compelling tool which can help us to reduce the dependence on a particular model selection by training our neural networks away from model-specific features and toward those that are present in data or are common across many models.

A. Notation

Throughout this text I will follow the conventions and notations passed down from the venerable Peskin and Schroeder. Therefore, I use natural units:

$$\hbar = c = 1$$

and will typically hold off on writing out either $\hbar$ or c, save in instances where it may be informative to do so.

I also follow the typical particle physics convention and take the relativistic metric tensor to be mostly negative:

$$g_{\mu\nu} = g^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

with Greek indices denoting the space-time dimensions (0, 1, 2, 3 or t, x, y, z respectively) and Roman indices (eg. i or j) denoting only the spatial dimensions. Rather confusingly Greek subscripts can also denote particle types but hopefully it will be clear when that's the case. Additionally a four-vector will be represented as italic and a spatial three-vector as bold:

$$r^\mu = (r^0, \boldsymbol{r}) \quad \text{and} \quad r_\mu = g_{\mu\nu} r^\mu = (r^0, -\boldsymbol{r})$$

where repeated indices are to be summed over following Einstein's summation notation.

Commutation is denoted by square brackets and anticommutation with curly brackets:

$$[A, B] \equiv AB - BA \quad \text{and} \quad \{A, B\} \equiv AB + BA$$

Pauli's sigma matrices are:

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

which obey the following commutation and anticommutation relations:

$$\left[\sigma^i, \sigma^j\right] = 2i\epsilon^{ijk}\sigma^k \quad \text{and} \quad \left\{\sigma^i, \sigma^j\right\} = 2\delta^{ij} I$$

where ϵ^{ijk} is the totally antisymmetric Levi-Civita symbol which is +1 for $(i, j, k) = (1, 2, 3)$ and any even permutations (ie. $(2, 3, 1)$, is -1 for any odd permutations (ie. $(3, 2, 1)$, and 0 for any repeated indices. The Kroenecker delta is δ^{ij} (equal to 1 if $i = j$ and 0 otherwise) and I is the identity matrix.

We also have the Dirac matrices, γ^μ, in n dimensions which follow the anticommutation relation:

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu} \times I_{n\times n}$$

and therefore the form of the Dirac matrices depends on the particular basis we choose. Further, we can define a fifth gamma matrix, γ^5, such that:

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = -\frac{i}{4!}\epsilon^{\mu\nu\rho\sigma}\gamma_\mu\gamma_\nu\gamma_\rho\gamma_\sigma$$

with the following properties:

$$\begin{aligned}
(\gamma^5)^\dagger &= \gamma^5 && \text{(it is hermitian)} \\
(\gamma^5)^2 &= 1 \times I && \text{(it has eigenvalues } \pm 1\text{)} \\
\left\{\gamma^5, \gamma^\mu\right\} &= 0 && \text{(it anticommutes with the other gamma matrices)}
\end{aligned}$$

In particular, using the Dirac basis we have:

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix}, \quad \gamma^5 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

with 1 denoting the appropriately sized identity matrix. In the Weyl/chiral basis they are:

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix}, \quad \gamma^5 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

which has the nice property that left and right-handed spinors can be "chosen" using γ^5:

$$\psi_L = \frac{1}{2}(1 - \gamma^5)\psi \quad \text{and} \quad \psi_R = \frac{1}{2}(1 + \gamma^5)\psi$$

with:

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}$$

Finally, I will also make use of the "slash notation":

$$\not{p} \equiv \gamma^\mu p_\mu$$

B. Electroweak Unification

The majority of this appendix follows a combination of Halzen & Martin [149] with Peskin & Schroeder [43]. It is included here more for my own benefit as a coherent blending of multiple explanations.

B.1 The Higgs Mechanism

We start off with a theory obeying an $SU(2) \times U(1)$ symmetry. This group has four generators, three $T^a = \frac{1}{2}\sigma^a$ from $SU(2)$ and a single generator for the $U(1)$ rotations. These generators are then coupled to the boson fields, A^a_μ and B_μ yielding the covariant derivative:

$$D_\mu = \partial_\mu - igA^a_\mu T^a - ig'\frac{Y}{2}B_\mu \tag{B.1}$$

where g and g' are coupling constants of the $SU(2)$ and $U(1)$ groups respectively and Y is the $U(1)$ charge of whatever particle this derivative is applied to. Next, we introduce a complex scalar field, ϕ, with $Y = +1$ which transforms under the $SU(2) \times U(1)$ group as:

$$\phi \rightarrow e^{i\alpha^a T^a} e^{i\beta/2}\phi \tag{B.2}$$

In analogy to spin, which is also part of an $SU(2)$ group, we use a basis where the T^3 operator defines the eigenstates of "isospin" and can create raising and lowering operators:

$$T^\pm = T^1 \pm iT^2 = \frac{1}{2}\left(\sigma^1 \pm i\sigma^2\right) \tag{B.3}$$

In this basis the field ϕ is a spinor:

$$\phi = \begin{pmatrix}\phi_a \\ \phi_b\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1 + i\phi_2 \\ \phi_3 + i\phi_4\end{pmatrix} \tag{B.4}$$

where ϕ_a and ϕ_b have T^3 eigenvalues of $+1/2$ and $-1/2$ respectively. Now, we can formulate a Lagrangian for ϕ with a potential $V(\phi)$:

$$\mathcal{L} = |D_\mu\phi|^2 - \mathcal{L}_{KE} - V(\phi) + \text{h.c.} \tag{B.5}$$

where the kinetic energy term of the gauge fields is:

$$\mathcal{L}_{KE} = \frac{1}{4}(A^i_{\mu\nu})^2 + \frac{1}{4}(B_{\mu\nu})^2 + \text{h.c.} \tag{B.6}$$

with:

$$A^i_{\mu\nu} = \partial_\mu A^i_\nu - \partial_\nu A i_\mu + g\epsilon^{ijk} A^j_\mu A^k_\nu$$
$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$$

Next, we take the potential to be $V(\phi) = -\mu^2\phi^\dagger\phi + \lambda\left(\phi^\dagger\phi\right)^2$. If $\mu^2 < 0$, then the scalar field ϕ would have some mass, μ and the minimum of the potential is located at $|\phi| = 0$. However, taking $\mu^2 > 0$ makes ϕ massless and moves the minimum of the potential away from 0. We usually take the minimum of the potential to define the "vacuum" state and a vacuum condition of $|\phi| = 0$ is typical. With $\mu^2 > 0$, the vacuum state is found to be where:

$$\phi^\dagger\phi = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{\mu^2}{2\lambda} \equiv v^2 \tag{B.7}$$

where v is called the "vacuum expectation value" or vev. We can then choose a particular minimum to expand ϕ about. As we'll see later a "nice" choice is at:

$$\phi_1 = \phi_2 = \phi_4 = 0 \quad \text{and} \quad \phi_3 = v$$

This is equivalent to:

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix}$$

By choosing a particular direction (here ϕ_3) we are inherently breaking the $SU(2)$ symmetry. Because the vacuum field is now at some nonzero value the Lagrangian is no longer symmetric under $SU(2)$ transformations. Expanding around this point with some perturbation, $h(x)$, and some rotations, α^a, gives:

$$\phi(x) \to e^{iT^a\alpha^a/v}\begin{pmatrix}0\\ \frac{v+h(x)}{\sqrt{2}}\end{pmatrix}$$
$$\sim \begin{pmatrix}1+i\frac{\alpha^3}{v} & i\frac{\alpha^1-i\alpha^2}{v}\\ i\frac{\alpha^1+i\alpha^2}{v} & 1-i\frac{\alpha^3}{v}\end{pmatrix}\begin{pmatrix}0\\ \frac{v+h(x)}{\sqrt{2}}\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}i(v+h(x))\frac{\alpha^1-i\alpha^2}{v}\\ (v+h(x))\frac{1-i\alpha^3}{v}\end{pmatrix}$$

and keeping terms up to $O(1)$ in perturbations $h(x)$ and α^i gives:

$$\phi \sim \frac{1}{\sqrt{2}} \begin{pmatrix} i\alpha^1 + \alpha^2 v \\ v + h(x) - i\alpha^3 \end{pmatrix}$$

which, for small α^i, is:

$$\phi \sim \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{B.8}$$

this new field, $h(x)$, is the Higgs boson and is all that remains with our choice of vacuum state. We can replace all the ϕs in the Lagrangian with this value.

Inspecting the coupling of this scalar field to the gauge fields (the second and third terms of covariant derivative) we have:

$$\left| \left(-igA_\mu^a T^a - ig' \frac{Y}{2} B_\mu \right) \phi \right|^2$$
$$= \frac{1}{2} \begin{pmatrix} 0 & v + h(x) \end{pmatrix} \left(gA_\mu^a T^a + \frac{1}{2} g' B_\mu \right) \left(gA^{b\mu} T^b + \frac{1}{2} g' B^\mu \right) \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}$$

where we used the fact that the $U(1)$ charge of ϕ is $Y = +1$. If we ignore the terms where $h(x)$ couples to the gauge fields and perform the matrix multiplications we are left with:

$$\frac{1}{2} \frac{v^2}{4} \left[g^2 (A_\mu^1 + iA_\mu^2)(A^{1\mu} - iA^{2\mu}) + (-gA_\mu^3 + g' B_\mu)(-gA^{3\mu} + g' B^\mu) \right]$$

which can be rewritten as:

$$\frac{1}{2} \frac{v^2}{4} \left[g^2 (A_\mu^1)^2 + g^2 (A_\mu^2)^2 + (-gA_\mu^3 + g' B_\mu)^2 \right] \tag{B.9}$$

We initially started with four massless vector bosons, but now can identify four new bosons that are linear combinations of the originals. Three of these have obtained masses:

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (A_\mu^1 \mp iA_\mu^2) \quad \text{with mass} \quad m_W = g \frac{v}{2} \tag{B.10}$$

$$Z_\mu = \frac{1}{\sqrt{g^2 + g'^2}} (gA_\mu^3 - g' B_\mu) \quad \text{with mass} \quad m_Z = \sqrt{g^2 + g'^2} \frac{v}{2} \tag{B.11}$$

We can also formulate another vector field, orthogonal to Z_μ, which did not gain a mass:

$$A_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(gA^3_\mu + g'B_\mu) \tag{B.12}$$

This new massless field is being labeled as A_μ since it will be identified as the photon field which commonly uses that notation, it is not to be confused with the A^a_μ fields that are part of the original $SU(2)$ group. It is relatively easy to see that:

$$W^+_\mu T^+ + W^-_\mu T^- = A^1_\mu T^1 + A^2_\mu T^2$$

which is why we labeled the new W fields as plus and minus. It is a bit harder to see that we can make the replacement:

$$gA^3_\mu T^3 + g'YB_\mu = \frac{1}{\sqrt{g^2 + g'^2}} Z_\mu (g^2 T^3 - g'^2 \frac{Y}{2}) + \frac{gg'}{\sqrt{g^2 + g'^2}} A_\mu (T^3 + \frac{Y}{2})$$

Using these substitutions to rewrite the covariant derivative in terms of the massive fields gives:

$$\begin{aligned} D_\mu = \partial_\mu - i\frac{g}{\sqrt{2}} W^+_\mu T^+ - i\frac{g}{\sqrt{2}} W^-_\mu T^- - i\frac{1}{\sqrt{g^2 + g'^2}} Z_\mu (g^2 T^3 - g'^2 \frac{Y}{2}) \\ -i\frac{gg'}{\sqrt{g^2 + g'^2}} A_\mu (T^3 + \frac{Y}{2}) \end{aligned} \tag{B.13}$$

We can see that the scalar field ϕ along with the potential ended up breaking the original symmetry group and in the processes gave mass to the gauge bosons. What's more, from the covariant derivative we can identify $(T^3 + Y/2)$ as the electric charge operator, Q since these values define the coupling strength to the emergent photon field. Going back to our choice of ϕ for the vacuum state, since the scalar field had a $U(1)$ charge of $Y = +1$ by choosing the isospin down component, ϕ_b, as the minimum the total electric charge of the vacuum state is $Q = -1/2 + 1/2 = 0$. Therefore, our choice of vacuum state breaks the $SU(2) \times U(1)_Y$ symmetry but leaves the $U(1)_Q$ symmetry intact. That is why this procedure leaves the photon massless and we end up with a theory where electric charge is conserved.

There is another "simplifying" change we can make by representing the change of basis

from (A^3_μ, B_μ) to (Z_μ, A_μ) as a matrix:

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_w & -\sin\theta_w \\ \sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} A^3_\mu \\ B_\mu \end{pmatrix} \tag{B.14}$$

then,

$$\cos\theta_w = \frac{g}{\sqrt{g^2 + g'^2}} \quad \text{and} \quad \sin\theta_w = \frac{g'}{\sqrt{g^2 + g'^2}} \tag{B.15}$$

and we can identify the electron charge as:

$$e = g\sin\theta_w$$

therefore, we can recast Eq.B.13 as:

$$D_\mu = \partial_\mu - i\frac{g}{\sqrt{2}}(W^+_\mu T^+ + W^-_\mu T^-) - i\frac{g}{\cos\theta_w} Z_\mu (T^3 - \sin^2\theta_w Q) - ieA_\mu Q \tag{B.16}$$

B.2 Fermions

By coupling this theory to the fermions we can also generate their masses in a similar manner. As we stated in 2.1, because the charged weak interactions only couple to left-handed fermions we are unable to insert "standard" mass terms for the fermions. Because we must separate the left and right handed states for the charged weak interactions we arrange the left-handed states into doublets and the right-handed ones as singlets for the leptons and quarks.

$$E_{iL} = \begin{pmatrix} \nu_{iL} \\ e_{iL} \end{pmatrix} \quad \text{and} \quad Q_{iL} = \begin{pmatrix} u_{iL} \\ d_{iL} \end{pmatrix} \tag{B.17}$$

where the index i denotes the lepton and quark generations:

$$\nu_i = (\nu_e, \nu_\mu, \nu_\tau) \qquad e_i = (e, \mu, \tau)$$
$$u_i = (u, c, t) \qquad d_i = (d, s, b)$$

The T^3 for the left-handed doublets are naturally $\pm 1/2$ while the eigenvalues are 0 for the right-handed singlets. Then we can assign the proper Y eigenvalues for each component such that we obtain the observed electric charges for each particle. Thus, the left-handed

lepton fields have $YE_{iL} = -1$ and the quark fields have $YQ_{iL} = +1/3$ while the right-handed fields have $Ye_{iR} = -2, Y\nu_{iR} = 0, Yu_{iR} = +4/3, Yd_{iR} = -2/3$

With these representations we can easily define the kinetic terms for the fermions:

$$\mathcal{L} = \bar{E}_{iL}(i\not{D})E_{iL} + \bar{e}_{iR}(i\not{D})e_{iR} + \bar{Q}_{iL}(i\not{\partial})E_{iL} + \bar{u}_{iR}(i\not{D})u_{iR} + \bar{d}_{iR}(i\not{D})d_{iR} + \text{h.c.}$$

Expanding this by plugging in Eq. B.16 gives the following interaction currents:

$$J_W^{\mu+} = \frac{1}{\sqrt{2}}(\bar{\nu}_{iL}\gamma^\mu e_{iL} + \bar{u}_{iL}\gamma^\mu d_{iL}); \tag{B.18}$$

$$J_W^{\mu-} = \frac{1}{\sqrt{2}}(\bar{e}_{iL}\gamma^\mu \nu_{iL} + \bar{d}_{iL}\gamma^\mu u_{iL}); \tag{B.19}$$

$$\begin{aligned} J_Z^\mu = \frac{1}{\cos\theta_w}\Bigg[& \bar{\nu}_{iL}\gamma^\mu\left(\frac{1}{2}\right)\nu_{iL} + \bar{e}_{iL}\gamma^\mu\left(-\frac{1}{2} + \sin^2\theta_w\right)e_{iL} + \bar{e}_{iR}\gamma^\mu\left(\sin^2\theta_w\right)e_{iR} \\ & + \bar{u}_{iL}\gamma^\mu\left(\frac{1}{2} - \frac{2}{3}\sin^2\theta_w\right)u_{iL} + \bar{u}_{iR}\gamma^\mu\left(-\frac{2}{3}\sin^2\theta_w\right)u_{iR} \\ & + \bar{d}_{iL}\gamma^\mu\left(-\frac{1}{2} + \frac{1}{3}\sin^2\theta_w\right)d_{iL} + \bar{d}_{iR}\gamma^\mu\left(\frac{1}{3}\sin^2\theta_w\right)d_{iR}\Bigg]; \end{aligned} \tag{B.20}$$

$$J_{EM}^\mu = \bar{e}\gamma^\mu(-1)e + \bar{u}\gamma^\mu\left(+\frac{2}{3}\right)u + \bar{d}\gamma^\mu\left(-\frac{1}{3}\right)d \tag{B.21}$$

where, for the EM current I've recombined the left and right-handed spinors back into one.

Since we have the scalar ϕ, which is an $SU(2)$ spinor, we can use it to combine the left and right-handed field into gauge invariant terms for the leptons and quarks:

$$\mathcal{L}_{lm} = -\lambda_{ei}\bar{E}_{iL}\cdot\phi e_{iR} - \lambda_{\nu i}\epsilon^{ab}\bar{E}_{iLa}\phi^\dagger\nu_{iR} + \text{h.c.} \tag{B.22}$$

$$\mathcal{L}_{qm} = -\lambda_{di}\bar{Q}_{iL}\cdot\phi d_{iR} - \lambda_{ui}\epsilon^{ab}\bar{Q}_{iLa}\phi_b^\dagger u_{iR} + \text{h.c.} \tag{B.23}$$

where I have left in the right-handed neutrino field for symmetry and because this is a natural extension to the SM to provide a neutrino mass[1]. All of these terms have a total Y that add up to zero and are invariant under $SU(2)$ rotations since we combine two doublets with a singlet. The next step is to plug in the vacuum state for ϕ and see the emergent mass terms:

[1] As I stated in a footnote of ch. 2.1, this is a Dirac mass term but we may also introduce Majorana terms. These would provide a "natural" way to describe the small neutrino mass.

$$\mathcal{L}_{lm} = -\frac{1}{\sqrt{2}}\lambda_{ei} v \bar{e}_{iL} e_{iR} - \frac{1}{\sqrt{2}}\lambda_{\nu i} v \bar{\nu}_{iL} \nu_{iR} + \text{h.c.} \tag{B.24}$$

$$\mathcal{L}_{qm} = -\frac{1}{\sqrt{2}}\lambda_{di} v \bar{d}_{iL} d_{iR} - \frac{1}{\sqrt{2}}\lambda_{ui} v \bar{u}_{iL} u_{iR} + \text{h.c.} \tag{B.25}$$

These are all of the single generation terms, however, terms with mixed generations are also gauge invariant, and therefore, we should add those in as well. This would entail changing the i indices on the right-handed fermions into a j and turning all λ_{fi} into matrices, λ_{fij} to represent the generational mixing. There is no a priori reason that any of these lambda matrices should be diagonal but we can diagonalize them to find the mass eigenstates (up to now we've been working in the interaction eigenstates of the fermions). This diagonalization takes the form of unitary matrix transformations:

$$e_{iL} = (U_e)_{ij} e^M_{jL}, \quad \nu_{iL} = (U_\nu)_{ij} \nu^M_{jL}, \quad u_{iL} = (U_u)_{ij} u^M_{jL}, \quad d_{iL} = (U_d)_{ij} d^M_{jL}, \tag{B.26}$$

where f_{iL} represents the interaction eigenstates (also called flavor states) and f^M_{iL} are the mass eigenstates. We can then recast the weak interaction currents into the mass eigenstates:

$$J^{\mu-}_W = \frac{1}{\sqrt{2}}(\bar{e}_{iL}\gamma^\mu \nu_{iL} + \bar{d}_{iL}\gamma^\mu u_{iL}) = \frac{1}{\sqrt{2}}\left(\bar{e}^M_{iL}\gamma^\mu (U_e^\dagger U_\nu)_{ij} \nu^M_{jL} + \bar{d}^M_{iL}\gamma^\mu (U_d^\dagger U_u)_{ij} u^M_{jL}\right) \tag{B.27}$$

It is customary to redefine $(U_d^\dagger U_u)_{ij}$ as a new matrix, V^{CKM}_{ij}, called the Cabbibo-Kobayashi-Maskawa (CKM) matrix. Further, we have not observed any mixing between the charged leptons, therefore, U_e is either identity or very close to it and $e_{iL} \sim e^M_{iL}$. Finally, the neutrino mixing matrix is commonly referred to at the PMNS matrix, named after Pontecorvo, Maki, Nakagawa, and Sakata, four physicists who formulated the initial neutrino mixing theory.

While we could have introduced the PMNS matrix by hand, allowing for right-handed neutrino states gives us a natural way to construct this matrix and also produces neutrino masses. What's more, even if we were to put in the PMNS matrix by hand, the resulting neutrino oscillations would necessitate massive neutrinos. Neutrino mixing means that the neutrino flavor states are made from linear superpositions of some other states. As shown in Appendix C, the probability of one neutrino flavor turning into another relies on the mass difference of these other states. Therefore, neutrino mixing requires that there be some underlying mass eigenstates. It is possible, of course, to formulate neutrino mass eigenstates without the inclusion of right-handed neutrinos in a number of ways. These

include extensions to the Higgs sector with more Higgs fields [150, 151] or producing Majorana masses by allowing for non-renormalizable terms [152].

C. 3 Flavor Oscillations with Wave Packets

As discussed in Chapter 2.3, the "standard" derivation of neutrino oscillations is strictly incorrect. Here I present the "better" wave packet derivation, mostly for my own benefit and following closely with Ref. [153].

Take the neutrino wave function to be some wave packet such that the neutrino mass states generated at the source are described by:

$$\left|\nu_i^S(\boldsymbol{x},t)\right\rangle = \Psi_i^S(\boldsymbol{x},t)\left|\nu_i\right\rangle = \int \frac{d^3p}{(2\pi)^{3/2}} f_i^S(\boldsymbol{p}-\boldsymbol{p}_i) e^{i\boldsymbol{p}\cdot\boldsymbol{x}-iE_i(\boldsymbol{p})t}\left|\nu_i\right\rangle \tag{C.1}$$

where $f_i^S(\boldsymbol{p}-\boldsymbol{p}_i)$ is a sharply peaked momentum distribution for the packet of mass state i with a mean momentum of $\boldsymbol{p}_i$ and energy $E_i(\boldsymbol{p}) = \sqrt{|\boldsymbol{p}|^2+m_i^2}$. The S denotes that these neutrino states were produced at the source. We then Taylor expand the energy about the mean momentum $\boldsymbol{p}_i$:

$$E_i(\boldsymbol{p}) = \sqrt{|\boldsymbol{p}|^2+m_i^2} \sim E_i(\boldsymbol{p}_i) + \left.\frac{\partial E_i(\boldsymbol{p})}{\partial p^j}\right|_{p_i^j}(p^j - p_i^j) \tag{C.2}$$

where p^j are the momenta in the three spatial dimensions and j is not an index for the neutrino mass states. Performing the derivative on the second term in the energy approximation gives:

$$E_i(\boldsymbol{p}) \sim E_i(\boldsymbol{p}_i) + \frac{\boldsymbol{p}_i\cdot(\boldsymbol{p}-\boldsymbol{p}_i)}{E_i(\boldsymbol{p}_i)} = E_i(\boldsymbol{p}_i) + \boldsymbol{v}_i\cdot(\boldsymbol{p}-\boldsymbol{p}_i) \tag{C.3}$$

where $\boldsymbol{v}_i$ is the group velocity of the i^{th} mass state's wave packet. We can then rearrange the wave packet from Eq. C.1 to get:

$$\Psi_i^S(\boldsymbol{x},t) \sim e^{i\boldsymbol{p}_i\cdot\boldsymbol{x}-iE_i(\boldsymbol{p}_i)t} g_i^S(\boldsymbol{x}-\boldsymbol{v}_i t) \tag{C.4}$$

with

$$\begin{aligned} g_i^S(\boldsymbol{x}-\boldsymbol{v}_i t) &= \int \frac{d^3p}{(2\pi)^{3/2}} f_i^S(\boldsymbol{p}-\boldsymbol{p}_i) e^{i(\boldsymbol{p}-\boldsymbol{p}_i)\cdot\boldsymbol{x} - i\boldsymbol{v}_i\cdot(\boldsymbol{p}-\boldsymbol{p}_i)t} \\ &= \int \frac{d^3p}{(2\pi)^{3/2}} f_i^S(\boldsymbol{p}) e^{i\boldsymbol{p}\cdot(\boldsymbol{x}-\boldsymbol{v}_i t)} \end{aligned} \tag{C.5}$$

where in the final term we've just made a change of integration variable. This integral term, $g_i^S(\boldsymbol{x}-\boldsymbol{v}_i t)$, is a shape factor for the i^{th} wave packet.

In a similar manner, We can write the wave function for the mass states at the detection location which is a distance $\boldsymbol{L}$ from the source. Here we don't have to worry about the time component since these states make up the detected neutrino and we have:

$$\Psi_i^D(\boldsymbol{x}-\boldsymbol{L}) \sim e^{i\boldsymbol{p}_i\cdot(\boldsymbol{x}-\boldsymbol{L})} g_i^D(\boldsymbol{x}-\boldsymbol{L}) \tag{C.6}$$

where $g_i^D(\boldsymbol{x}-\boldsymbol{L})$ is defined in the same manner as source shape factor but we've swapped $\boldsymbol{x}$ for $(\boldsymbol{x}-\boldsymbol{L})$ and set $t=0$.

We can now formulate the oscillation probability by first taking the transition amplitude:

$$\begin{aligned} A(\nu_\alpha \to \nu_\beta; \boldsymbol{L}) &= \sum_i U_{\alpha i}^* U_{\beta i} e^{i\boldsymbol{p}_i\cdot\boldsymbol{L} - iE_i(\boldsymbol{p}_i)t} \int d^3x g_i^{D*}(\boldsymbol{x}-\boldsymbol{L}) g_i^S(\boldsymbol{x}-\boldsymbol{v}_i t) \\ &= \sum_i U_{\alpha i}^* U_{\beta i} e^{i\boldsymbol{p}_i\cdot\boldsymbol{L} - iE_i(\boldsymbol{p}_i)t} G_i(\boldsymbol{L}-\boldsymbol{v}_i t) \end{aligned} \tag{C.7}$$

and the oscillation probability is:

$$\begin{aligned} P(\nu_\alpha \to \nu_\beta; L) &= \int dt |A(\nu_\alpha \to \nu_\beta; L)|^2 \\ &= \sum_{i,j} U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^* \int dt G_i(L-v_i t) G_j^*(L-v_j t) e^{-i\Delta\phi_{ij}(L,t)} \end{aligned} \tag{C.8}$$

where, for simplicity, we are aligning the propagation direction with the line connecting the source and detector. We've also defined:

$$\Delta\phi_{ij}(L,t) = (E_i(p_i) - E_j(p_j))t - (p_i - p_k)L \equiv \Delta E_{ij} t - \Delta p_{ij} L \tag{C.9}$$

We are getting close since Eq. C.8 already resembles Eq. 2.16 only with an ugly integral term. Taking relativistic neutrinos we can expand the energy difference about the

average energy of the two mass states, E:

$$\begin{aligned}\Delta E_{ij} &= \left.\frac{\partial E_i}{\partial p_i}\right|_{avg} p_i - \left.\frac{\partial E_j}{\partial p_j}\right|_{avg} p_j + \left.\frac{\partial E_i}{\partial m_i^2}\right|_{avg} m_i^2 - \left.\frac{\partial E_j}{\partial m_j^2}\right|_{avg} m_j^2 \\ &= v\Delta p_{ij} + \frac{\Delta m_{ij}^2}{2E}\end{aligned} \tag{C.10}$$

where v is the average group velocity of the wave packets of the two mass states. Then we plug this back into Eq. C.9 to get:

$$\Delta\phi_{ij}(L,t) = \frac{\Delta m_{ij}^2}{2E}t - (L - vt)\Delta p_{ij} \tag{C.11}$$

where the first term in this phase difference is the same phase factor from the "standard" oscillation derivation which results in Eq. 2.16. Plugging this phase into Eq. C.8 we have:

$$P(\nu_\alpha \to \nu_\beta; L) = \sum_{i,j} U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^* e^{-i\frac{\Delta m_{ij}^2 L}{2E}} \int dt G_i(L - v_i t) G_j^*(L - v_j t) e^{-i(L-vt)\Delta p_{ij}} \tag{C.12}$$

We can now work through the integral term. Written in full this term is:

$$\begin{aligned}I_{ij} = \iiint dt dp_1 dp_2 f_i^S(p_1) f_i^{D*}(p_1) f_j^{S*}(p_2) f_j^D(p_2) \\ \times e^{ip_1(L-v_i t) - ip_2(L-v_j t)} e^{-i(L-vt)\Delta p_{ij}}\end{aligned} \tag{C.13}$$

I've not been explicit but all of these integrals are from $-\infty$ to ∞ and so we can group the time exponentials and perform the time integral to yield a Dirac delta function:

$$\begin{aligned}I_{ij} = \iint dp_1 dp_2 f_i^S(p_1) f_i^{D*}(p_1) f_j^{S*}(p_2) f_j^D(p_2) \\ \times \delta(v_i p_1 - v_j p_2 - v\Delta p_{ij}) e^{i(p_1 - p_2 - \Delta p_{ij})L}\end{aligned} \tag{C.14}$$

Then we perform one of the momentum integrals using the delta function:

$$\begin{aligned}I_{ij} = \frac{1}{v_j}\int dp f_i^S(p) f_i^{D*}(p) f_j^{S*}(\frac{1}{v_j}(v_i p - v\Delta p_{ij})) f_j^D(\frac{1}{v_j}(v_i p - v\Delta p_{ij})) \\ \times\ e^{i((1-v_i/v_j)p - (1-v/v_j)\Delta p_{ij})L}\end{aligned} \tag{C.15}$$

We can now make some reasoned arguments to see the impact of this integral term. The

exponential term is dependent on the group velocities of the mass state wave packets. We can see that if the velocities are close to each other then the integral loses its dependence on oscillation distance, L. We can write this as:

$$\left|1 - \frac{v_i}{v_j}\right| pL << 1 \tag{C.16}$$

Further, since the momentum distribution functions, $f(p)$, are sharply peaked at an argument of zero, their contributions to the integral are largest when p is smaller than the momentum spread, $|p| \lesssim \sigma_p$. Thus the condition becomes:

$$\begin{gathered} \left|1 - \frac{v_i}{v_j}\right| \sigma_p L << 1 \\ \text{or} \\ L << \sigma_x \frac{v}{v_i - v_j} = l_{coh} \end{gathered} \tag{C.17}$$

where $\sigma_x = 1/\sigma_p$ and l_{coh} is the coherence length. What this tells us is that the integral is independent of the oscillation distance, L, if L is smaller than the distance the neutrino states would have to travel to start separating due to non-equal group velocities. In other words, oscillations can occur as long as the neutrino states remain coherent. If, on the other hand, the mass states decohere before they are detected then the integral term will work to suppress the oscillation probability due to rapid oscillations in the exponential.

Next, we look at the momentum distribution terms in the integral. Again, will only consider parts of momentum space where $|p| \lesssim \sigma_p$ since this region is where the first two terms will be nonzero. Then we have a condition on the second pair of distribution terms:

$$\begin{gathered} v_i \sigma_p - v \Delta p_{ij} >> 0 \\ \text{or} \\ \sigma_x v \Delta p_{ij} << 1 \end{gathered} \tag{C.18}$$

From the relativistic condition we obtained Eq. C.10 and we also know that the energy difference between the mass states will also be small. Thus, we have $v\Delta p_{ij} \sim \frac{\Delta m^2_{ij}}{2E}$ which is the length that sets the scale for oscillations to occur. Then the second condition can be rewritten as:

$$\sigma_x << l_{osc} \tag{C.19}$$

Here, the spatial variance is set by the extent of the source and detector. What this condition tells us is that oscillations will be suppressed if the extent of either the source or detector is large compared to the characteristic oscillation length, $l_{osc} = 2E/\Delta m^2_{ij}$. This is referred to as the localization condition and if it is not met then the integration will average out any oscillations

Now we can complete our analysis. If both the coherence and localization conditions are met then the integral term just provides a normalization constant and we obtain the standard oscillation probability:

$$P(\nu_\alpha \rightarrow \nu_\beta; L) = \sum_{i,j} U^*_{\alpha i} U_{\beta i} U_{\alpha j} U^*_{\beta j} e^{-i\frac{\Delta m^2_{ij} L}{2E}} \tag{C.20}$$

If the coherence condition is not met then the integral term will suppress oscillations while if the localization condition is violated then oscillations will be averaged out. On reflection, both conditions are sensible since the mass state wave packets must travel close together in order to interfere and produce oscillations and if the source and detector are not localized then spatial extents of the wave packets will be large and the interference will average out. With both conditions satisfied we recover the standard oscillation probability which we can derive by making expolicity incorrect assumptions. Both of these conditions are upheld in the majority of oscillation experiments.

D. Neutrinos From Pions

Charged pions primarily produce neutrinos through via $\pi^{\pm} \rightarrow \mu^{\pm} + \nu_{\mu}$ (99.988% branching ratio). Neutrino energy and flux can be determined as a function of the angle from the direction of the traveling pions. Here, we will consider two reference frames, the pion center of mass frame and the laboratory frame as shown in Figure D.1.

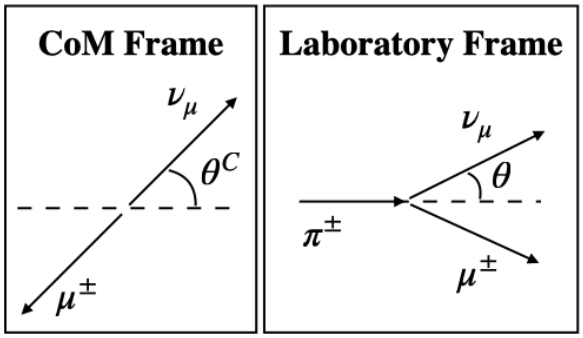


Figure D.1: Two reference frames for $\pi^{\pm} \rightarrow \mu^{\pm} + \nu_{\mu}$ decay. The left side shows the center-of-mass (CoM) frame and the right side shows the laboratory frame.

We can immediately write the four-momenta for all particles in both reference frames, taking the dotted line in Figure D.1 as the z-axis and up as the y-axis:

$$p_{\pi}^{C} = \begin{pmatrix} m_{\pi} & 0 & 0 & 0 \end{pmatrix} \tag{D.1}$$

$$p_{\nu}^{C} = \begin{pmatrix} E_{\nu}^{C} & 0 & E_{\nu}^{C} \sin\theta^{C} & E_{\nu}^{C} \cos\theta^{C} \end{pmatrix} \tag{D.2}$$

$$p_{\mu}^{C} = \begin{pmatrix} E_{\mu}^{C} & 0 & -E_{\nu}^{C} \sin\theta^{C} & -E_{\nu}^{C} \cos\theta^{C} \end{pmatrix} \tag{D.3}$$

$$p_{\pi} = \begin{pmatrix} E_{\pi} & 0 & 0 & |\boldsymbol{p}_{\pi}| \end{pmatrix} \tag{D.4}$$

$$p_{\nu} = \begin{pmatrix} E_{\nu} & 0 & E_{\nu} \sin\theta & E_{\nu} \cos\theta \end{pmatrix} \tag{D.5}$$

$$p_{\mu} = \begin{pmatrix} E_{\mu} & 0 & -E_{\nu} \sin\theta & |\boldsymbol{p}_{\pi}| - E_{\nu} \cos\theta \end{pmatrix} \tag{D.6}$$

Here, Eqs. D.4–D.6 are the four-momenta in the lab frame and Eqs. D.1–D.3 are in the CoM frame. Conservation of momentum has been used to express the muon momentum components in terms of the neutrino momentum. The neutrino was treated as massless to get $E_\nu^2 = |\boldsymbol{p}_\nu|^2$.

We can also apply a boost along the z-axis to the CoM equations to get another set of four-momenta in the lab frame.

$$p_\pi = \begin{pmatrix} \gamma m_\pi & 0 & 0 & \gamma\beta m_\pi \end{pmatrix} \tag{D.7}$$

$$p_\nu = \begin{pmatrix} \gamma E_\nu^C(1+\beta\cos\theta^C) & 0 & E_\nu^C\sin\theta^C & \gamma E_\nu^C(\cos\theta^C+\beta) \end{pmatrix} \tag{D.8}$$

$$p_\mu = \begin{pmatrix} \gamma(E_\mu^C - \beta E_\nu^C\cos\theta^C) & 0 & -E_\nu\sin\theta & \gamma(-E_\nu^C\cos\theta^C + \beta E_\mu^C) \end{pmatrix} \tag{D.9}$$

with $\gamma = (1-\beta^2)^{-1/2} = E_\pi/m_\pi$.

We can now use conservation of four-momentum squared to find the neutrino energy in both reference frames. In both frames we will use:

$$p_\pi = p_\nu + p_\mu \xrightarrow{\text{rearrange}} p_\mu = p_\pi - p_\nu \xrightarrow{\text{square}} (p_\mu)^2 = (p_\pi - p_\nu)^2$$

$$\Rightarrow m_\mu^2 = m_\pi^2 - 2p_\pi^\alpha p_{\nu\,\alpha}$$

where α denotes the four-momentum components and we have used the fact that neutrinos are nearly massless to set $m_\nu^2 = 0$. Doing the four-momentum multiplication and rearranging, we get:

$$E_\nu^C = \frac{1}{2}(1 - \frac{m_\mu^2}{m_\pi^2})m_\pi \tag{D.10}$$

and

$$E_\nu = \frac{(1-\frac{m_\mu^2}{m_\pi^2})m_\pi^2}{2(E_\pi - |\boldsymbol{p}_\pi|\cos\theta)} \tag{D.11}$$

To determine the neutrino flux as a function of angle θ in the lab frame we first note that in the CoM frame the angular probability distribution for the neutrino is uniform:

$$\frac{dP}{d\Omega^C} = \frac{1}{4\pi}$$

where the differential solid angle, $d\Omega$, is equal to $\sin\theta d\theta d\phi$. Therefore, in we find the probability distribution in the lab frame to be:

$$\frac{dP}{d\Omega} = \frac{dP}{d\Omega^C}\frac{d\Omega^C}{d\Omega} = \frac{1}{4\pi}\frac{\sin\theta^C}{\sin\theta}\frac{d\theta^C}{d\theta} \tag{D.12}$$

Equating the energy and y-components of Eqs. D.5 and D.8 we find:

$$E_\nu^C \sin\theta^C = E_\nu \sin\theta$$

and

$$\gamma E_\nu^C (1 + \beta\cos\theta^C) = E_\nu$$

then plugging in the results of E_ν^C and E_ν gives:

$$\frac{\sin\theta^C}{\sin\theta} = \frac{E_\nu}{E_\nu^C} = \frac{m_\pi}{E_\pi - |\boldsymbol{p}_\pi|\cos\theta} \tag{D.13}$$

and

$$\cos\theta^C = \frac{1}{\beta}\left(\frac{1}{\gamma(E_\pi - |\boldsymbol{p}_\pi|\cos\theta)} - 1\right) \tag{D.14}$$

Next, noting that the derivative of $\cos^{-1} x$ is $1/\sqrt{1+x^2}$, we find that:

$$\frac{d\theta^C}{d\theta} = \left[1 - \frac{1}{\beta^2}\left(\frac{m_\pi}{\gamma(E_\pi - |\boldsymbol{p}_\pi|\cos\theta)} - 1\right)^2\right]^{-1/2} \frac{m_\pi p_\pi \sin\theta}{\beta\gamma(E_\pi - |\boldsymbol{p}_\pi|\cos\theta)^2} \tag{D.15}$$

Eqs. D.13 and D.15 give us all we need to determine $dP/d\Omega$, however, the resulting formula is messy and not particularly informative. We can make some simplifying assumptions for a high energy pion beam. At high energy, $\beta \approx 1$ and we can Taylor expand $|\boldsymbol{p}_\pi| = \sqrt{E_\pi - m_\pi}$ to get $|\boldsymbol{p}_\pi| \approx E_\pi - m_\pi^2/(2E_\pi)$. Further, since in the lab frame the majority of decay products will be forward going, we can use the small angle approximations:

$$\cos\theta \approx 1 - \frac{1}{2}\theta^2$$
$$\sin\theta \approx \theta$$

These approximations in turn give:

$$\gamma(E_\pi - |\boldsymbol{p}_\pi|\cos\theta) \approx \gamma E_\pi \left[1 - \left(1 - \frac{m_\pi^2}{2E_\pi^2}\right)\left(1 - \frac{1}{2}\theta^2\right)\right] \approx \frac{m_\pi}{2}(1 + \gamma^2\theta^2) \tag{D.16}$$

Plugging Eqs. D.13 and D.14 into D.12 then applying the small angle and high energy approximations ($\beta \approx 1$) along with Eq. D.16 gives us:

$$\frac{dP}{d\Omega} \approx \frac{1}{4\pi}\frac{m_\pi}{\frac{m_\pi}{2\gamma}(1+\gamma^2\theta^2)}\left[1 - \left(\frac{m_\pi}{\gamma\frac{m_\pi}{2\gamma}(1+\gamma^2\theta^2)} - 1\right)^2\right]^{-1/2} \frac{m_\pi E_\pi \theta}{\gamma\frac{m_\pi^2}{4\gamma^2}(1+\gamma^2\theta^2)^2}$$

which, with a lot of cancellations, simplifies to:

$$\frac{dP}{d\Omega} \approx \frac{1}{4\pi}\left(\frac{2\gamma}{1+\gamma^2\theta^2}\right)^2$$

meaning that the neutrino flux over some area, A, at some distance, z, is:

$$\Phi_\nu \approx \frac{A}{4\pi z^2}\left(\frac{2\gamma}{1+\gamma^2\theta^2}\right)^2 \tag{D.17}$$

We can also apply the same approximations to Eq. D.11 and find a simplified form of the neutrino energy distribution in the lab frame:

$$E_\nu \approx \frac{(1 - m_\mu^2/m_\pi^2)E_\pi}{1+\gamma^2\theta^2} \tag{D.18}$$

as a reminder $\gamma = E_\pi/m_\pi$. Similar relations hold for muon neutrinos originating from kaon decay with all π subscripts being replaced with K. However, the branching ratio for $K^\pm \rightarrow \mu^\pm + \nu_\mu$ is only 63.6%.

E. Gradient Reversal Layer Implementation

This follows closely the code developed by Ref. [129] with some minor modifications.

```
from keras.engine import Layer
import keras.backend as K
from tensorflow.python.framework import ops
import tensorflow as tf

def reverse_gradient(X, hp_lambda):
    """Flips the sign of the incoming gradient during training."""
    try:
        reverse_gradient.num_calls += 1
    except AttributeError:
        reverse_gradient.num_calls = 1

    grad_name = "GradientReversal%d" % reverse_gradient.num_calls

    @ops.RegisterGradient(grad_name)
    def _flip_gradients(unused_op,grad):
        return [tf.negative(grad) * hp_lambda]

    g = K.get_session().graph
    with g.gradient_override_map({'Identity': grad_name}):
        y = tf.identity(X)

    return y
```

```
class GradientReversal(Layer):
    """
    Layer that flips the sign of gradient during backpropagation.

    Parameters
    ----------
    hp_lambda : float
        Gradient reversal loss multiplier.
    """

    def __init__(self, hp_lambda, **kwargs):
        self.supports_masking = True
        self.hp_lambda = hp_lambda
        super(GradientReversal, self).__init__(**kwargs)

    @staticmethod
    def get_output_shape_for(input_shape):
        return input_shape

    def build(self, input_shape):
        self.trainable_weights = []

    def call(self, x, mask=None):
        return reverse_gradient(x, self.hp_lambda)

    def get_config(self):
        config = {'hp_lambda': self.hp_lambda}
        base_config = super(GradientReversal, self).get_config()
        return dict(list(base_config.items()) + list(config.items()))
```

F. The Test Beam

The success of the NOνA experiment depends, in part, on the ability to understand the limits of the collected data. As the experiment continues to take data the statistical uncertainties will naturally decrease making it more important to understand and possibly address the causes of systematic errors. Among the leading systematic uncertainties on the are calibration and the detector response. These all impact NOvA's ability to accurately measure the energy deposited by neutrinos and understand the activity of final state particles emerging from those neutrinos. To address these issues, a subset of the NOvA collaboration began installing a scaled-down version of the NOvA detector design in 2018 at the Fermilab Test Beam Facility (FTBF). This Test Beam (TB) detector has been exposed to a beam of primarily pions, protons, muons, and electrons with momenta between 0.2–2 GeV in order to characterize the detector response and validate the calibration procedure. Additionally, the test beam effort will result in a library of particles at known energies that can be used in the development and training of new simulation and particle identification techniques.

F.1 Secondary and Tertiary Particle Beams

FTBF provides a beam consisting of primarily pions with an energy that can be tuned between 8–80 GeV [154]. This secondary beam collides with a copper target and passes through a collimator to produce a tertiary beam. The tertiary beam is momentum selected by a magnet with a variable field to deliver particles with momenta from 0.2–2 GeV/c. Each beam spill lasts 4.2 s and produces that trigger the beamline data collection and detector readout.

F.2 The Main Detector

The TB detector uses the same detector technology as the near and far detectors. The test beam detector is $2.6 \times 2.6 \times 4.1$ m^3 allowing for full containment of muons up to 0.9 GeV while containment for 2 GeV pions is greater than 95% longitudinally and greater than 98% transversely. It consists of two blocks, the first with 32 planes and the second with 31. The detector is outfitted mostly by FD-type front end boards with four ND-type boards

placed in the center of each block.

F.3 Beamline technology

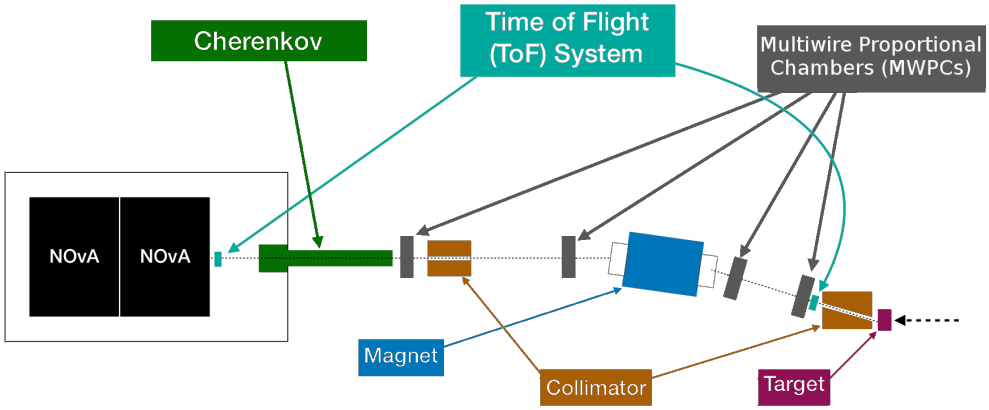


Figure F.1: Layout of the tertiary beamline components

As shown in Figure F.1, the tertiary beamline consists of a combination of detectors for momentum measurement and particle identification along with a single sweeping dipole magnet with a maximum field of 1.8 T for momentum selection. Two time-of-flight (TOF) scintillators are placed on either end of the tertiary beamline. With a separation of 13.2 m, these provide particle identification with good separation between slow protons and the other particles. A Cherenkov counter filled with CO_2 at 1 atm enables further separation of electrons due to a threshold energy deposition of 20 MeV. A set of four multi-wire proportional chambers (MWPCs) in conjunction with the sweeping magnet provide the ability for momentum tagging of tertiary beam particles.

F.4 Preliminary Results

The particle identification and momentum reconstruction capabilities are shown in Figure F.2 for data recorded with magnets currents of 500 A, 750 A, and 1000 A. We can clearly see two bands appearing along with a small band between them. The horizontal band corresponds to fast-moving, light particles which have an approximately constant

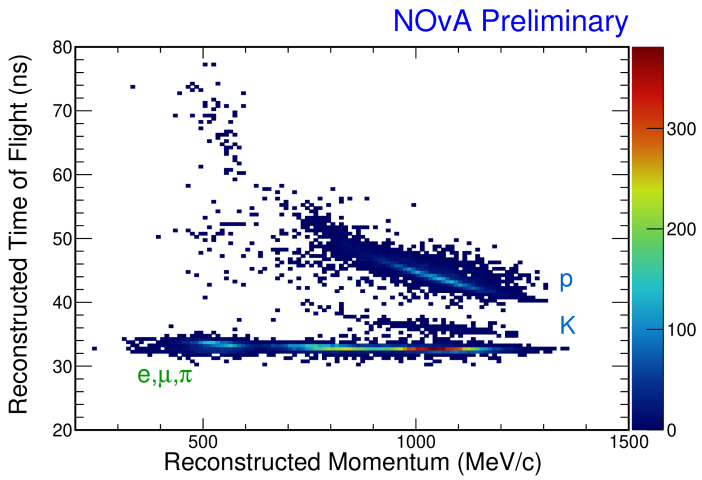
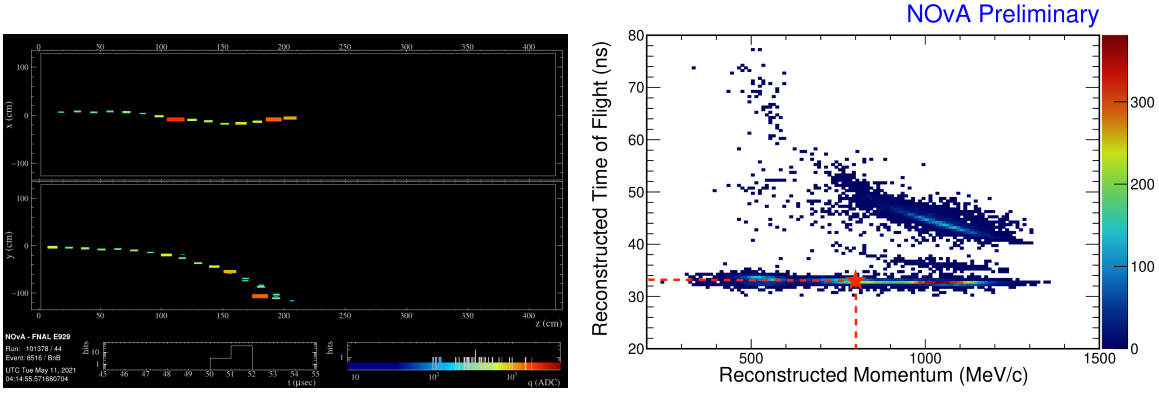


Figure F.2: Reconstructed time-of-flight vs momentum for TB data, labels showing where each particle type resides. The color gradient indicates data accumulation in the parameter space.

time-of-flight across all selection magnet settings. The curved band of events to the top right of the plot are associated with the heavy protons and the smaller curved band is from kaons, which have a mass between protons and muons.

When an event triggers the beamline detectors a signal is also sent to the DAQ of the main detector to record in-time information. The beamline and detector events are then associated with each other by time-matching them during post-processing. Figure F.3a



(a) (b)

Figure F.3: (a) Candidate pion event with a reconstructed time-of-flight of 38.2 s and momentum of 0.8 GeV/c. (b) Indication where the particle lies in the TOF vs momentum plot.

shows a pion candidate event and the location of this event in the TOF vs momentum plot. Due to the reconstructed TOF and momentum, this particle was likely due to an electron, muon, or pion. The Cherenkov detector was not triggered on this event, reducing the likelihood that this was an electron. Finally, looking at the particle path in the detector shows it underwent a significant amount of scattering indicating that it is likely a pion rather than a muon, which would have a much straighter path through the detector.

F.5 Plans

The NOvA TB detector is fully installed, outfitted, and has been collecting data since May 2019. Various beam conditions and effects have been investigated to optimize the run-time and limit the impacts of background events. The TB effort is currently in its fourth period of data taking which is scheduled to end at the beginning of summer 2022. Thus far the TB has collected $O(10^4)$ "analysis quality" events and continues to collect particles at various energies to inform our understanding of the detector response and calibration procedure. The library of collected events will also provide an independent sample of particles in the same range of energies as those produced by neutrino interactions in the main NOvA experiment. These events can then be used to analyze our current algorithms and develop new methods to more accurately simulate and reconstruct the neutrino interactions we observe.

Bibliography

[1] Henri Becquerel. On the Invisible Radiations Emitted by Phosphorescent Substance. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, 122: 501–503, 1896.

[2] Henri Becquerel. On the Invisible Radiations Emitted by the Salts of Uranium. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, 122: 559–564, 1896.

[3] H. Becquerel. Emission of New Radiations by Metallic Uranium. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, 122:1086–1088, 1896.

[4] E. Rutherford. Uranium Radiation and the Electrical conduction Produced by it. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 47:109–163, Sep. 1899.

[5] J. Chadwick. Intensitätsverteilung im magnetischen Spectrum der β-Strahlen von radium B + C. *Verhandlungen der Deutschen Physikalischen Gesellschaft*, 16: 383–391, 1914.

[6] G. Ecker. James Chadwick: ahead of his time. *arXiv e-prints*, 2020.

[7] C. D. Ellis and W. A Wooster. The Average Energy of Disintegration of Radium E. *Proc. Roy. Soc. Lond. A*, 117:109–123, Dec. 1927. doi: 10.1098/rspa.1927.0168.

[8] W. Pauli. Open letter to the group of radioactive people at the Gauverein meeting in Tübingen, Dec. 1930.

[9] Fred L. Wilson. Fermi's Theory of Beta Decay. *Am. J. Phys.*, 36:1150–1160, 1968. doi: 10.1119/1.1974382.

[10] H. Bethe and R. Peierls. The "Neutrino". *Nature*, 133(3362):532–532, Apr. 1934.

[11] Luis W. Alvarez. A Proposed Experimental Test of the Neutrino Theory, Apr. 1949.

[12] B. Pontecorvo. Inverse beta process. *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.*, 1:25–31, 1991.

[13] F. Reines, C. L. Cowan, F. B. Harrison, A. D. McGuire, and H. W. Kruse. Detection of the Free Antineutrino. *Phys. Rev.*, 117:159–173, Jan. 1960. doi: 10.1103/PhysRev.117.159.

[14] F. Reines and C. L. Cowan. A Proposed experiment to detect the free neutrino. *Phys. Rev.*, 90:492–493, May 1953. doi: 10.1103/PhysRev.90.492.

[15] C. L. Cowan, F. Reines, F. B. Harrison, E. C. Anderson, and F. N. Hayes. Large liquid scintillation detectors. *Phys. Rev.*, 90:493–494, May 1953. doi: 10.1103/PhysRev.90.493.

[16] F. Reines and C. L. Cowan. Detection of the Free Neutrino. *Phys. Rev.*, 92:830–831, Nov. 1953. doi: 10.1103/PhysRev.92.830.

[17] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire. Detection of the Free Neutrino: A Confirmation. *Science*, 124:103–104, Jul. 1956. doi: 10.1126/science.124.3212.103.

[18] Raymond Davis. Attempt to Detect the Antineutrinos from a Nuclear Reactor by the $Cl^{37}(\overline{\nu}, e^{-})A^{37}$ Reaction. *Phys. Rev.*, 97:766–769, Feb. 1955. doi: 10.1103/PhysRev.97.766.

[19] John N. Bahcall. Solar Neutrinos. I. Theoretical. *Phys. Rev. Lett.*, 12:300–302, Mar. 1964. doi: 10.1103/PhysRevLett.12.300.

[20] Raymond Davis. Solar Neutrinos. II. Experimental. *Phys. Rev. Lett.*, 12:303–305, Mar. 1964. doi: 10.1103/PhysRevLett.12.303.

[21] 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.

[22] 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.

[23] T. D. Lee and C. N. Yang. Question of Parity Conservation in Weak Interactions. *Phys. Rev.*, 104:254–258, Oct. 1956. doi: 10.1103/PhysRev.104.254.

[24] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental Test of Parity Conservation in Beta Decay. *Phys. Rev.*, 105:1413–1415, Feb. 1957. doi: 10.1103/PhysRev.105.1413.

[25] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov. 1967. doi: 10.1103/PhysRevLett.19.1264.

[26] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519: 367–377, May 1968. doi: 10.1142/9789812795915_0034.

[27] F. J. Hasert et al. Observation of Neutrino Like Interactions without Muon or Electron in the Gargamelle Neutrino Experiment. *Nucl. Phys. B*, 73:1–22, Sep. 1974. doi: 10.1016/0550-3213(74)90038-8.

[28] G. Danby, J-M. Gaillard, K. Goulianos, L. M. Lederman, N. Mistry, M. Schwartz, and J. Steinberger. Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos. *Phys. Rev. Lett.*, 9:36–44, Jul. 1962. doi: 10.1103/PhysRevLett.9.36.

[29] V. Alles-Borelli et al. Limits on the Electromagnetic Production of Heavy Leptons. *Lett. Nuovo Cim.*, 4S1:1156–1159, Nov. 1970. doi: 10.1007/BF02753664.

[30] Yung-Su Tsai. Decay Correlations of Heavy Leptons in $e^+ + e^- \rightarrow l^+ + l^-$. *Phys. Rev. D*, 4:2821–2837, Nov. 1971. doi: 10.1103/PhysRevD.4.2821.

[31] M. L. Perl et al. Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation. *Phys. Rev. Lett.*, 35:1489–1492, Dec. 1975. doi: 10.1103/PhysRevLett.35.1489.

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

[33] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. Remarks on the unified model of elementary particles. *Prog. Theor. Phys.*, 28:870–880, Jun. 1962. doi: 10.1143/PTP.28.870.

[34] V. N. Gribov and B. Pontecorvo. Neutrino astronomy and lepton charge. *Phys. Lett. B*, 28:493, Dec. 1969. doi: 10.1016/0370-2693(69)90525-5.

[35] M. Z. Akrawy et al. Measurement of the Z^0 Mass and Width with the OPAL Detector at LEP. *Phys. Lett. B*, 231:530–538, 1989. doi: 10.1016/0370-2693(89)90705-3.

[36] D. Decamp et al. Determination of the Number of Light Neutrino Species. *Phys. Lett. B*, 231:519–529, 1989. doi: 10.1016/0370-2693(89)90704-1.

[37] L. Wolfenstein. Neutrino Oscillations in Matter. *Phys. Rev. D*, 17:2369–2374, 1978. doi: 10.1103/PhysRevD.17.2369.

[38] Y. Fukuda et al. Evidence for Oscillation of Atmospheric Neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, Aug. 1998. doi: 10.1103/PhysRevLett.81.1562.

[39] 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.

[40] W. Buchmuller, R. D. Peccei, and T. Yanagida. Leptogenesis as the origin of matter. *Ann. Rev. Nucl. Part. Sci.*, 55:311–355, 2005. doi: 10.1146/annurev.nucl.55.090704.151558.

[41] J. W. F. Valle. Neutrino masses: evidences and implications. *J. Phys. Conf. Ser.*, 485:012005, 2014. doi: 10.1088/1742-6596/485/1/012005.

[42] Alexandre Deur, Stanley J. Brodsky, and Guy F. de Teramond. The QCD Running Coupling. *Nucl. Phys.*, 90:1, 2016. doi: 10.1016/j.ppnp.2016.04.003.

[43] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995. ISBN 978-0-201-50397-5.

[44] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961. doi: 10.1016/0029-5582(61)90469-2.

[45] P. A. Zyla et al. (Particle Data Group). Review of Particle Physics. *Prog. Theor. Exp. Phys.*, 2020(8):083C01, 2020. doi: 10.1093/ptep/ptaa104.

[46] S. Raby and R. Slansky. Neutrino masses: How to add them to the standard model. *Los Alamos Sci.*, 25:64–71, 1997.

[47] C. H. Llewellyn Smith. Neutrino Reactions at Accelerator Energies. *Phys. Rept.*, 3: 261–379, 1972. doi: 10.1016/0370-1573(72)90010-5.

[48] J. L. Hewett et al. Fundamental Physics at the Intensity Frontier. May 2012. doi: 10.2172/1042577.

[49] E. Kh. Akhmedov and A. Yu. Smirnov. Neutrino oscillations: Entanglement, energy-momentum conservation and QFT. *Found. Phys.*, 41:1279–1306, 2011. doi: 10.1007/s10701-011-9545-4.

[50] S. P. Mikheev and A. Yu. Smirnov. Resonant amplification of neutrino oscillations in matter and solar neutrino spectroscopy. *Nuovo Cim. C*, 9:17–26, 1986. doi: 10.1007/BF02508049.

[51] Hiroshi Nunokawa, Stephen J. Parke, and Jose W. F. Valle. CP Violation and Neutrino Oscillations. *Prog. Part. Nucl. Phys.*, 60:338–402, 2008. doi: 10.1016/j.ppnp.2007.10.001.

[52] Accelerator Division. *Concepts Rookie Book*, pages 29–46. FNAL Operations Department, Batavia, IL, Aug. 2020.

[53] Sacha E. Kopp. The NuMI beam at FNAL and its use for neutrino cross section measurements. *AIP Conf. Proc.*, 967(1):49–52, 2007. doi: 10.1063/1.2834509.

[54] S. Mufson et al. Liquid scintillator production for the NOvA experiment. *Nucl. Instrum. Meth. A*, 799:1–9, 2015. doi: 10.1016/j.nima.2015.07.026.

[55] C. Andreopoulos et al. The GENIE Neutrino Monte Carlo Generator. *Nucl. Instrum. Meth. A*, 614:87–104, 2010. doi: 10.1016/j.nima.2009.12.009.

[56] Costas Andreopoulos et al. The GENIE Neutrino Monte Carlo Generator: Physics and User Manual. 10 2015.

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

[58] Geant4 Collaboration. Geant4 10.4 release notes. *geant4-data.web.cern.ch, https://geant4-data.web.cern.ch/ ReleaseNotes/ReleaseNotes4.10.4.html*, 2017.

[59] V. V. Uzhinsky. The Fritiof (FTF) Model in Geant4. In *International Conference on Calorimetry for the High Energy Frontier*, pages 260–264, 2013.

[60] D. H. Wright and M. H. Kelsey. The Geant4 Bertini Cascade. *Nucl. Instrum. Meth. A*, 804:175–188, 2015. doi: 10.1016/j.nima.2015.09.058.

[61] L. Aliaga et al. Neutrino Flux Predictions for the NuMI Beam. *Phys. Rev. D*, 94(9): 092005, 2016. doi: 10.1103/PhysRevD.94.092005. [Addendum: Phys.Rev.D 95, 039903 (2017)].

[62] J. Nieves, Jose Enrique Amaro, and M. Valverde. Inclusive quasi-elastic neutrino reactions. *Phys. Rev. C*, 70:055503, 2004. doi: 10.1103/PhysRevC.70.055503. [Erratum: Phys.Rev.C 72, 019902 (2005)].

[63] R. Gran, J. Nieves, F. Sanchez, and M. J. Vicente Vacas. Neutrino-nucleus quasi-elastic and 2p2h interactions up to 10 GeV. *Phys. Rev. D*, 88:113007, Dec 2013. doi: 10.1103/PhysRevD.88.113007.

[64] Ch. Berger and L. M. Sehgal. Lepton mass effects in single pion production by neutrinos. *Phys. Rev. D*, 76:113004, Dec 2007. doi: 10.1103/PhysRevD.76.113004.

[65] Ch. Berger and L. M. Sehgal. Partially conserved axial vector current and coherent pion production by low energy neutrinos. *Phys. Rev. D*, 79:053003, Mar 2009. doi: 10.1103/PhysRevD.79.053003.

[66] A Bodek and U. K. Yang. Higher twist, xi(omega) scaling, and effective LO PDFs for lepton scattering in the few GeV region. *J. Phys. G*, 29:1899–1906, 2003. doi: 10.1088/0954-3899/29/8/369.

[67] T. Yang, C. Andreopoulos, H. Gallagher, K. Hoffmann, and P. Kehayias. A Hadronization Model for Few-GeV Neutrino Interactions. *Eur. Phys. J. C*, 63:1–10, 2009. doi: 10.1140/epjc/s10052-009-1094-z.

[68] Steven Dytman. Final state interactions in neutrino-nucleus experiments. *Acta Phys. Polon. B*, 40:2445–2460, 2009.

[69] R. Gran et al. Antineutrino Charged-Current Reactions on Hydrocarbon with Low Momentum Transfer. *Phys. Rev. Lett.*, 120:221805, Jun 2018. doi: 10.1103/PhysRevLett.120.221805.

[70] R. M. Sternheimer. The density effect for ionization loss in materials. *Phys. Rev.*, 88:851–859, 1952. doi: 10.1103/PhysRev.88.851.

[71] J. B. Birks. Scintillations from Organic Crystals: Specific Fluorescence and Relative Response to Different Radiations. *Proc. Phys. Soc. A*, 64:874–877, 1951. doi: 10.1088/0370-1298/64/10/303.

[72] Chris Hagmann et al. Cosmic-ray shower generator (CRY) for Monte Carlo transport codes. *2007 IEEE Nuclear Science Symposium Conference Record*, 2:1143–1146, 2007. doi: 10.1109/NSSMIC.2007.4437209.

[73] P. Balaban et al. The probability distribution of gains in avalanche photodiodes. *IEEE Transactions on Electron Devices*, 23(10):1189–1190, 1976. doi: 10.1109/T-ED.1976.18570.

[74] A. Rodriguez and A. Laio. Clustering by fast search-and-find of density peaks. *Science, American Association for the Advancement of Science*, 344(6191):1492–1496, 2014. doi: 10.1126/science.1242072.

[75] 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.

[76] Leandro A.F. Fernandes and Manuel M. Oliveira. Real-time line detection through an improved Hough transform voting scheme. *Pattern Recognition*, 41(1):299–314, 2008. doi: https://doi.org/10.1016/j.patcog.2007.04.003.

[77] Mattias Ohlsson, Carsten Peterson, and Alan L. Yuille. Track finding with deformable templates the elastic arms approach. *Computer Physics Communications*, 71(1):77–98, 1992. doi: https://doi.org/10.1016/0010-4655(92)90074-9.

[78] M. Baird, J. Bian, M. Messier, E. Niner, D. Rocco, and K. Sachdev. Event Reconstruction Techniques in NOvA. *J. Phys. Conf. Ser.*, 664(7):072035, 2015. doi: 10.1088/1742-6596/664/7/072035.

[79] Miin-Shen Yang and Kuo-Lung Wu. Unsupervised Possibilistic Clustering. *Pattern Recogn.*, 39(1):5–21, Jan. 2006. doi: 10.1016/j.patcog.2005.07.005.

[80] J. Dunn. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 3(3):32–57, Sep. 1973. doi: 10.1080/01969727308546046.

[81] N.H. Kuiper. Tests concerning random points on a circle. *Proceedings of the Koinklijke Nederlandse Akademie van Wetenschappen*, Series A. 63:38–47, 1962.

[82] R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth. A*, 262:444–450, Jun. 1987. doi: 10.1016/0168-9002(87)90887-4.

[83] Pierre Billoir. Progressive track recognition with a Kalman like fitting procedure. *Comput. Phys. Commun.*, 57:390–394, Dec. 1989. doi: 10.1016/0010-4655(89)90249-X.

[84] Pierre Billoir and S. Qian. Simultaneous pattern recognition and track fitting by the Kalman filtering method. *Nucl. Instrum. Meth. A*, 294:219–228, Sep. 1990. doi: 10.1016/0168-9002(90)91835-Y.

[85] G. Lutz. Optimum Track Fitting in the Presence of Multiple Scattering. *Nucl. Instrum. Meth. A*, 273:349, Dec. 1988. doi: 10.1016/0168-9002(88)90836-4.

[86] Yann Coadou. Boosted Decision Trees and Applications. *EPJ Web Conf.*, 55:02004, 2013. doi: 10.1051/epjconf/20135502004.

[87] A. Aurisano et al. A Convolutional Neural Network Neutrino Event Classifier. *JINST*, 11(09):P09001, Apr. 2016. doi: 10.1088/1748-0221/11/09/P09001.

[88] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. doi: 10.1109/CVPR.2016.90.

[89] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-Excitation Networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018. doi: 10.1109/CVPR.2018.00745.

[90] Mark Sandler et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *arXiv e-prints*, 2019. doi: 1801.04381.

[91] Micah C. Groh. *Constraints on Neutrino Oscillation Parameters from Neutrinos and Antineutrinos with Machine Learning*. PhD thesis, Indiana U., 2021.

[92] Laurens van der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(86):2579–2605, Nov. 2008.

[93] Fernanda Psihas. *Measurement of Long Baseline Neutrino Oscillations and Improvements from Deep Learning*. PhD thesis, Indiana U., Feb. 2018.

[94] L. Michel. Interaction between four half spin particles and the decay of the μ meson. *Proc. Phys. Soc. A*, 63:514–531, 1950. doi: 10.1088/0370-1298/63/5/311.

[95] Aaron S. Meyer, Minerba Betancourt, Richard Gran, and Richard J. Hill. Deuterium target data for precision neutrino-nucleus cross sections. *Phys. Rev. D*, 93(11): 113015, 2016. doi: 10.1103/PhysRevD.93.113015.

[96] P. Adamson et al. Study of quasielastic scattering using charged-current $\nu\mu$ -iron interactions in the MINOS near detector. *Phys. Rev. D*, 91(1):012005, 2015. doi: 10.1103/PhysRevD.91.012005.

[97] B. W. Allardyce et al. Pion reaction cross-sections and nuclear sizes. *Nucl. Phys. A*, 209:1–51, 1973. doi: 10.1016/0375-9474(73)90049-3.

[98] A. Saunders et al. Reaction and total cross-sections for low-energy pi+ and pi- on isospin zero nuclei. *Phys. Rev. C*, 53:1745–1752, 1996. doi: 10.1103/PhysRevC.53.1745.

[99] O. Meirav, E. Friedman, R. R. Johnson, R. Olszewski, and P. Weber. Low-energy Pion - Nucleus Potentials From Differential and Integral Data. *Phys. Rev. C*, 40: 843–849, 1989. doi: 10.1103/PhysRevC.40.843.

[100] S. M. Levenson et al. Inclusive pion scattering in the delta (1232) region. *Phys. Rev. C*, 28:326–332, 1983. doi: 10.1103/PhysRevC.28.326.

[101] D. Ashery, I. Navon, G. Azuelos, H. K. Walter, H. J. Pfeiffer, and F. W. Schleputz. True Absorption and Scattering of Pions on Nuclei. *Phys. Rev. C*, 23:2173–2185, 1981. doi: 10.1103/PhysRevC.23.2173.

[102] D. Ashery et al. Inclusive pion single charge exchange reactions. *Phys. Rev. C*, 30: 946–951, 1984. doi: 10.1103/PhysRevC.30.946.

[103] E. S. Pinzon Guerra et al. Measurement of σ_{ABS} and σ_{CX} of π^+ on carbon by the Dual Use Experiment at TRIUMF (DUET). *Phys. Rev. C*, 95(4):045203, 2017. doi: 10.1103/PhysRevC.95.045203.

[104] Melanie Day and Kevin S. McFarland. Differences in Quasi-Elastic Cross-Sections of Muon and Electron Neutrinos. *Phys. Rev. D*, 86:053003, 2012. doi: 10.1103/PhysRevD.86.053003.

[105] K. Abe et al. Measurement of neutrino and antineutrino oscillations by the T2K experiment including a new additional sample of ν_e interactions at the far detector. *Phys. Rev. D*, 96(9):092006, 2017. doi: 10.1103/PhysRevD.96.092006. [Erratum: Phys.Rev.D 98, 019902 (2018)].

[106] Jonathon Shlens. A Tutorial on Principal Component Analysis. *ArXiv*, abs/1404.1100, 2014.

[107] N. Agafonova et al. Final Results of the OPERA Experiment on ν_τ Appearance in the CNGS Neutrino Beam. *Phys. Rev. Lett.*, 120(21):211801, 2018. doi: 10.1103/PhysRevLett.120.211801. [Erratum: Phys.Rev.Lett. 121, 139901 (2018)].

[108] Gary J. Feldman and Robert D. Cousins. A Unified approach to the classical statistical analysis of small signals. *Phys. Rev. D*, 57:3873–3889, 1998. doi: 10.1103/PhysRevD.57.3873.

[109] J. Neyman. Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. *Phil. Trans. Roy. Soc. Lond. A*, 236(767):333–380, 1937. doi: 10.1098/rsta.1937.0005.

[110] K. Abe et al. Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations. *Nature*, 580(7803):339–344, 2020. doi: 10.1038/s41586-020-2177-0. [Erratum: Nature 583, E16 (2020)].

[111] K. Abe et al. Atmospheric neutrino oscillation analysis with external constraints in Super-Kamiokande I-IV. *Phys. Rev. D*, 97(7):072001, 2018. doi: 10.1103/PhysRevD.97.072001.

[112] P. Adamson et al. Precision Constraints for Three-Flavor Neutrino Oscillations from the Full MINOS+ and MINOS Dataset. *Phys. Rev. Lett.*, 125(13):131802, 2020. doi: 10.1103/PhysRevLett.125.131802.

[113] M. G. Aartsen et al. Measurement of Atmospheric Neutrino Oscillations at 6–56 GeV with IceCube DeepCore. *Phys. Rev. Lett.*, 120(7):071801, 2018. doi: 10.1103/PhysRevLett.120.071801.

[114] Mehreen Saeed. An Introduction To Recurrent Neural Networks And The Math That Powers Them, Sept. 2021. URL https://machinelearningmastery.com/an-introduction-to-recurrent-neural-networks-and-the-math-that-powers-them/.

[115] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*, chapter 6.5 Back-Propagation and Other Dierentiation Algorithms. MIT Press, Cambridge, massachusetts, USA, 2016.

[116] Peltarion. Categorical crossentropy, 2022. URL https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/loss-functions/categorical-crossentropy.

[117] Michael Phi. Illustrated Guide to LSTMs and GRUs: A step by step explanation, Sept. 2018. URL https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21.

[118] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997. doi: 10.1162/neco.1997.9.8.1735.

[119] Wikimedia Commons. Schematic of the long-short term memory cell, a component of recurrent neural networks, 2018. URL https://commons.wikimedia.org/wiki/File:LSTM_Cell.svg. File: `LSTM_Cell.svg`.

[120] Dmitrii Torbunov. *Improving Energy Estimation at NOvA with Recurrent Neural Networks*. PhD thesis, University of Minnesota, May 2021.

[121] Sergey Ioffe and Christian Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, page 448456, 2015.

[122] Muriel Gevrey, Ioannis Dimopoulos, and Sovan Lek. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, 160(3):249–264, 2003. doi: https://doi.org/10.1016/S0304-3800(02)00257-0.

[123] Eilam Gross and Ofer Vitells. Wald's theorem and the "Asimov" data set. URL https://indico.cern.ch/event/74940/contributions/2088584/attachments/1047729/1493442/Wald-Asimov.pdf. ATLAS Statistics Forum, Dec. 2009.

[124] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. doi: 10.1140/epjc/s10052-011-1554-0. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[125] Kaiyang Zhou et al. Domain Generalization in Vision: A Survey. 2021.

[126] Jindong Wang and. Generalizing to Unseen Domains: A Survey on Domain Generalization. In *International Joint Conferences on Artificial Intelligence Organization*, 2021.

[127] Yaroslav Ganin and Victor Lempitsky. Unsupervised Domain Adaptation by Backpropagation. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, page 11801189. JMLR.org, 2015.

[128] Han Zhao et al. Multiple Source Domain Adaptation with Adversarial Training of Neural Networks. 05 2017.

[129] Michele Tonutti et al. Robust and Subject-Independent Driving Manoeuvre Anticipation through Domain-Adversarial Recurrent Neural Networks. *Robotics Auton. Syst.*, 115:162–173, 2019.

[130] Yoshinari Hayato and Luke Pickering. The NEUT Neutrino Interaction Simulation. 6 2021. doi: 10.1140/epjs/s11734-021-00287-7.

[131] Jihn E. Kim, Paul Langacker, M. Levine, and H. H. Williams. A theoretical and experimental review of the weak neutral current: a determination of its structure and limits on deviations from the minimal $SU(2)_L \times U(1)$ electroweak theory. *Rev. Mod. Phys.*, 53:211–252, Apr 1981. doi: 10.1103/RevModPhys.53.211. URL https://link.aps.org/doi/10.1103/RevModPhys.53.211.

[132] O. Benhar, A. Fabrocini, S. Fantoni, and I. Sick. Spectral function of finite nuclei and scattering of GeV electrons. *Nucl. Phys. A*, 579:493–517, 1994. doi: 10.1016/0375-9474(94)90920-2.

[133] Konstantin S. Kuzmin, Vladimir V. Lyubushkin, and Vadim A. Naumov. Axial masses in quasielastic neutrino scattering and single-pion neutrino production on nucleons and nuclei. *Acta Phys. Polon. B*, 37:2337–2348, 2006.

[134] Krzysztof M. Graczyk and Jan T. Sobczyk. Lepton mass effects in weak charged current single pion production. *Phys. Rev. D*, 77:053003, 2008. doi: 10.1103/PhysRevD.77.053003.

[135] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006. doi: 10.1088/1126-6708/2006/05/026.

[136] M. Glück, E. Reya, and A. Vogt. Dynamical parton distributions revisited. *Eur. Phys. J. C*, 5:461–470, 1998. doi: 10.1007/s100520050289.

[137] Yoshinari Hayato. A neutrino interaction simulation program library NEUT. *Acta Phys. Polon. B*, 40:2477–2489, 2009.

[138] Torbjorn Sjostrand. High-energy-physics event generation with PYTHIA 5.7 and JETSET 7.4. *Computer Physics Communications*, 82:74–89, 1994.

[139] L. L. Salcedo, E. Oset, M. J. Vicente-Vacas, and C. Garcia-Recio. Computer Simulation of Inclusive Pion Nuclear Reactions. *Nucl. Phys. A*, 484:557–592, 1988. doi: 10.1016/0375-9474(88)90310-7.

[140] F. Pedregosa et al. Receiver operating characteristic (roc), 2021. URL https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html.

[141] Jose M. Clavijo, Paul Glaysher, Jenia Jitsev, and Judith M. Katzy. Adversarial domain adaptation to reduce sample bias of a high energy physics event classifier *. *Mach. Learn. Sci. Tech.*, 3(1):015014, 2022. doi: 10.1088/2632-2153/ac3dde.

[142] Christoph Englert, Peter Galler, Philip Harris, and Michael Spannowsky. Machine Learning Uncertainties with Adversarial Neural Networks. *Eur. Phys. J. C*, 79(1):4, 2019. doi: 10.1140/epjc/s10052-018-6511-8.

[143] Aishik Ghosh, Benjamin Nachman, and Daniel Whiteson. Uncertainty-aware machine learning for high energy physics. *Phys. Rev. D*, 104(5):056026, 2021. doi: 10.1103/PhysRevD.104.056026.

[144] G. N. Perdue et al. Reducing model bias in a deep learning classifier using domain adversarial neural networks in the MINERvA experiment. *JINST*, 13(11):P11020, 2018. doi: 10.1088/1748-0221/13/11/P11020.

[145] Marta Babicz, Saul Alonso-Monsalve, and Stephen Dolan. Neutrino interaction event filtering at liquid argon time projection chambers using neural networks with minimal input model bias. *PoS*, ICRC2021:1075, 2021. doi: 10.22323/1.395.1075.

[146] Aishik Ghosh and Benjamin Nachman. A cautionary tale of decorrelating theory uncertainties. *Eur. Phys. J. C*, 82(1):46, 2022. doi: 10.1140/epjc/s10052-022-10012-w.

[147] Zhongyi Pei, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Multi-Adversarial Domain Adaptation. In *AAAI*, 2018.

[148] Farzaneh Khoshnevisan and Min Chi. Unifying Domain Adaptation and Domain Generalization for Robust Prediction Across Minority Racial Groups. In *Machine Learning and Knowledge Discovery in Databases. Research Track*, pages 521–537, Cham, 2021. Springer International Publishing. ISBN 978-3-030-86486-6.

[149] F. Halzen and Alan D. Martin. *QUARKS AND LEPTONS: AN INTRODUCTORY COURSE IN MODERN PARTICLE PHYSICS*. John Wiley & Sons, 1984. ISBN 978-0-471-88741-6.

[150] Sahar Bahrami. *The Higgs Triplet Model: mixing in the neutral sector, vector-like fermions, and dark matter*. PhD thesis, Concordia U., Montreal, 7 2016.

[151] Daniel A. Camargo, Alex G. Dias, Téssio B. de Melo, and Farinaldo S. Queiroz. Neutrino Masses in a Two Higgs Doublet Model with a U(1) Gauge Symmetry. *JHEP*, 04:129, 2019. doi: 10.1007/JHEP04(2019)129.

[152] Sacha Davidson. Neutrino Physics. *CERN Yellow Rep. School Proc.*, 3:167–198, 2018. doi: 10.23730/CYRSP-2018-003.167.

[153] Evgeny Kh. Akhmedov and Alexei Yu. Smirnov. Paradoxes of neutrino oscillations. *Phys. Atom. Nucl.*, 72:1363–1381, 2009. doi: 10.1134/S1063778809080122.

[154] Fermilab Test Beam Facility. Beam Structure and Delivery Path. URL https://ftbf.fnal.gov/beam-delivery-path/.