Improving Physics Based Electron Neutrino Appearance Identification with a Long Short-Term Memory Network

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Andrew Vold

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Marvin Marshak

August, 2018

Acknowledgements

To begin, I would like to thank my family for their unwavering support throughout my life. They provided the resources and encouragement I needed during my graduate schooling. If it were not for all of them, I would not have been capable of pursuing my doctoral degree.

I would also like to thank my former adviser, Daniel Cronin-Hennessy for his knowledge and guidance during his time as my adviser. I thank Marvin Marshak for stepping in as my adviser when Dan needed to step down.

Finally, I am grateful for the assistance provided by my colleagues, Vladimir Bychkov and Jarrett Brown, the NOνA Reconstruction conveners, Evan Niner, Michael Baird, and Jianming Bian, and the ν_e Appearance Analysis conveners, Chris Backhouse and Alex Himmel. Performing an analysis such as this is no simple feat, and it would not have been feasible without their assistance.

Dedication

I dedicate this dissertation to my Lord, Jesus Christ. The fact that you are reading this thesis is testament to His grace. I can't count the number of times when God came through for me when I was in great need. My enrollment at the University of Minnesota for both my undergraduate and graduate education were nothing short of a miracle. During my time in graduate school, I experienced heartbreaks, crippling depression, and no foreseeable progress in my research. Under my own strength, I can promise you that I would not be writing this right now. There were times when I loathed being alive and barely had the strength to make it through the day. There was no way I could have made progress in my research and gain an education in artificial intelligence without the strength and guidance from God. Jesus gave me purpose for my life when I could not find any, and He carried me through the struggles when I lacked the strength to walk.

I don't know all of the details of your life: the hurts, disappointments, and struggles. However, I can tell you that from my experience, that God loves you so much, and that giving your life to Christ is the best decision you can possibly make. This won't make your problems instantly disappear, but I can promise you that Jesus will strengthen you and give you purpose in all of this. With God, there is hope for better circumstances. They may not arrive tomorrow, but rest assured, they will come.

Abstract

The NOνA experiment is a long baseline neutrino oscillation experiment with the objective to measure the oscillation probability of muon type neutrinos (ν_μ) into electron type neutrinos (ν_e). NOνA measures the interactions of neutrinos from the NuMI beam in two functionally identical liquid scintillator detectors. The far detector detects the appearance of electron neutrinos, leading to measurement of the oscillation parameters under study. Using an off axis beam with an 810 km baseline length, NOνA is sensitive to measuring the neutrino mass hierarchy, the CP violating parameter, and the octant of the mixing angle, θ_{23}. The data presented in this thesis has been collected from October 2013, until May 2018.

The first NOνA ν_e charged current identifier utilized an artificial neural network with the physical features of the highest energy reconstructed shower as inputs. The ν_e charged current identifier in this thesis utilizes a Long Short-Term Memory network with the physical features of every reconstructed shower in a particular interaction. In addition to the Long Short-Term Memory network, there are two Boosted Decision Trees to assist in event level selection. In the analysis of the data, 54 ν_e candidate events were detected with an expected background of 15 events. The results of this analysis prefer the normal mass hierarchy with maximal mixing and $\frac{\delta_{CP}}{\pi} = 1.92^{+0.08}_{-1.19}$. Results from this analysis differ from the published NOνA analysis, due to the differences of electron identification techniques.

Contents

List of Tables

List of Figures

Chapter 1

Neutrino Physics

1.1 History of Neutrinos

The story of neutrinos begins in 1930 when an experiment conducted by James Chadwick found that the electrons emitted from the beta decay of ^{214}Pb had a continuous energy spectrum [1].

This was puzzling at the time because the kinematics of a two body decay with a massive target demands a mono-energetic energy spectrum of the outgoing particle. There were two explanations which came out of this observation. Niels Bohr proposed that energy is conserved on a statistical basis [2]. Wolfgang Pauli proposed, however, that in these decays, there was a light, neutral particle, which led to a three body phase space, thus explaining the continuous electron energy spectrum [3].

In 1956, the neutrino was observed by a group of physicists by conducting a beta capture experiment. In particular, the group first observed the electron flavor in the capture of electron anti-electron neutrinos by the following process:

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

This measurement led to winning the Nobel Prize in physics in 1995 [4]

The other neutrino flavors were discovered by measuring the charged current interactions of the neutrinos from the other charged lepton decays. As a result of this, there are known to be three neutrino flavors which are identified by the charged lepton involved in a charged current interaction with the given neutrino [5, 6]

1.2 The Standard Model

The Weinberg-Glashow-Salam Model, also known as the Standard Model is the current description of elementary particles and their interactions. It is a $SU(3) \otimes SU_L(2) \otimes U(1)$ theory. The model is consists of interactions between fundamental half integer spin particles known as fermions and integer spin particles known as bosons.

The interactions in the Standard Model are mediated by Gauge Bosons. The simplest interaction is electromagnetism. Electromagnetism is a $U(1)$ gauge theory which describes the interactions of matter with the $U(1)$ gauge boson, known as the photon. The photon is massless, and it has spin 1. With this U(1) gauge symmetry, electromagnetic interactions must preserve electromagnetic charge. The next simplest interaction is known as the Weak Interaction. The Weak Interaction is a $SU_L(2)$ Yang-Mills Theory which describes the interactions of matter with the $SU_L(2)$ gauge bosons, known as the $W^{\pm}$ and the Z bosons. The gauge bosons of the Weak Interaction are very massive, so the interactions occur at small length scales. The most complex interaction of the Standard Model is the Strong Interaction. The Strong Interaction is a $SU(3)$ Yang-Mills Theory which describes the interactions of quarks with the $SU(3)$ gauge bosons, known as gluons. Gluons are also massless and have spin 1. With this SU(3) gauge symmetry, the strong charge, also known as color, must be conserved in strong interactions.

Matter in the Standard Model is comprised of two families of fermions: Quarks and Leptons. Quarks are the building blocks of hadrons. They are held in close proximity, as a result of the strength of the strong interaction at increasing distance. Since the strong interaction increases as the separation between quarks increases, it impossible to have free quarks. Since Quarks have color, electromagnetic charge, weak charge, and mass, they participate in all known Standard Model interactions. Leptons, on the other hand, contain no color. Therefore, the charged leptons participate in all interactions, besides the strong interaction. Neutrinos are neutral leptons, so they only participate in the weak interaction. There are three charged leptons, and there are three neutrinos with the same flavors as the charged leptons. Under the weak interaction, the leptons form weak doublets because the weak interaction is a$SU_L(2)$ gauge theory. The charged

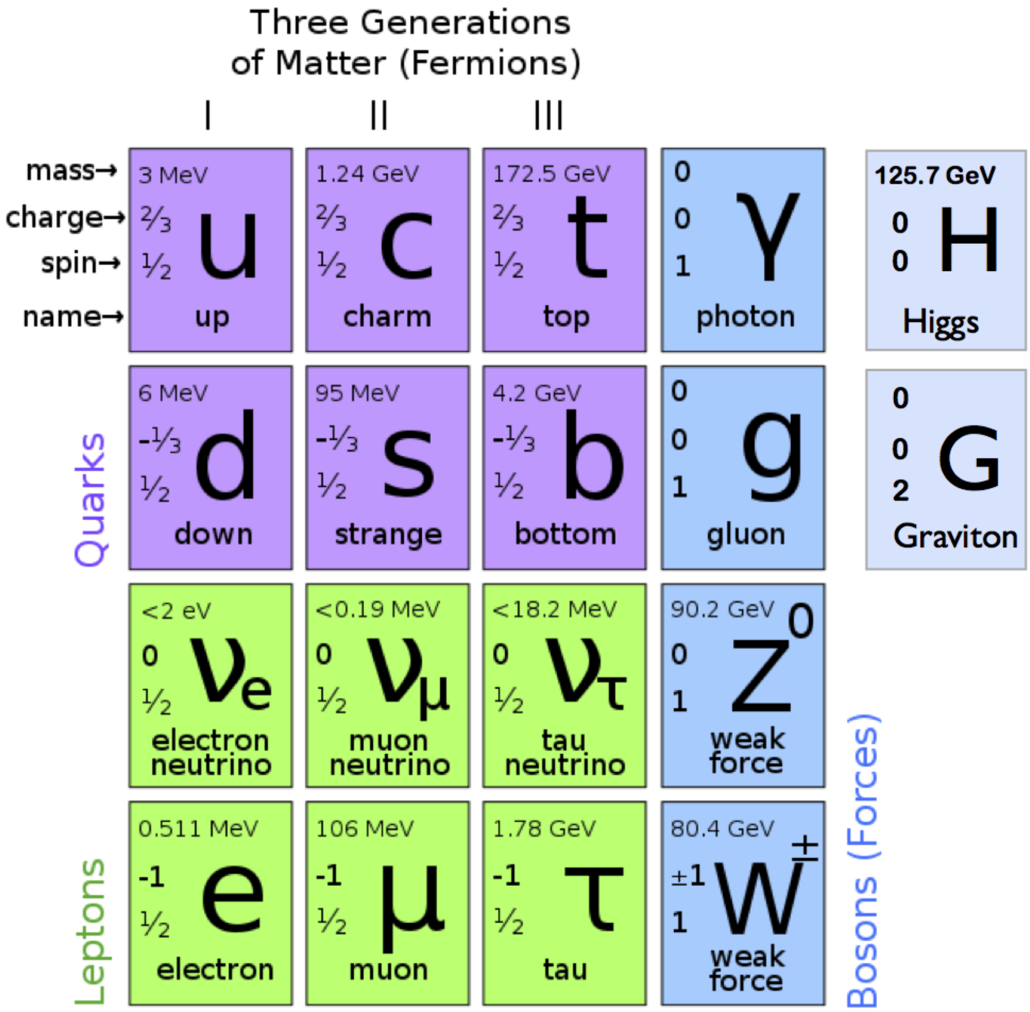


Figure 1.1: Known elementary particles in the Standard Model

lepton and the neutrino with the same flavor form a weak doublet with the charged lepton having $I_3 = +1/2$, and the neutrino having $I_3 = -1/2$.

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}$$

Figure 1.2: $SU_L(2)$ weak doublets for the three lepton generations.

1.3 Neutrino Fields

From experiments, neutrinos were found to be spin $\frac{1}{2}$ particles. As a result of this, neutrinos obey the Pauli Exclusion Principle, and thus must be described by anti-symmetric objects. The fields describing neutrinos must also be Lorentz invariant. The Lorentz Group is SO(1,3), which can be decomposed as:

$$SO(1,3) \simeq SU_L(2) \oplus SU_R(2) \tag{1.2}$$

As a result of this, there are two spin 1/2 irreducible representations of the Lorentz Group known as the Spinor Representations. The two spinor representations form right handed and left handed Weyl Spinors. A more compact way of using these spinors is to introduce the Dirac Spinor. The Dirac Spinor is a 4 vector, which consists of two Weyl Spinor components, as follows:

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} \bar{\psi} = \begin{pmatrix} \psi_R^\dagger & \psi_L^\dagger \end{pmatrix} \tag{1.3}$$

Due to its compact nature, the Dirac Representation is convenient to use. The Lagrangian density describing free fermions is the following:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \tag{1.4}$$

It is very interesting to note that to maintain Lorentz Invariance, the mass term only couples left handed chiral states to right handed chiral states, unlike the kinetic term which couples the chiral states. By finding the Euler-Lagrange equations, one finds the Dirac Equation:

$$(i\gamma^{\mu}\partial_{\mu} - m)\psi = 0 \tag{1.5}$$

A special case of equation 1.5 is when the particle is massless. In this case, m = 0, and we get the Weyl Lagrangian:

$$\mathcal{L} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi \tag{1.6}$$

One of the questions which remains in neutrino physics is whether or not they are Majorana Fermions. Since a Majorana Fermion state is equivalent to its charge conjugated state, it is impossible for the charged leptons to be Majorana Fermions. Since neutrinos are electrically neutral, it is possible for them to be Majorana Fermions. The Lagrangian for Majorana Fermions is as follows:

$$\mathcal{L} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - \frac{i}{2}(m_L\psi_L^{\dagger}\sigma_2\psi_L^{*} + m_R\psi_R^{\dagger}\sigma_2\psi_R^{*} + h.c.) \tag{1.7}$$

Since charge conjugation has the following form for Weyl Spinors: $\psi^c = -i\sigma_2\psi^*$, one can easily see from the Lagrangian that the mass terms couple the Weyl Spinors to their charge conjugated partners, meaning that Majorana Fermions are their own charge conjugates, which would lead to a violation in lepton number conservation for neutrinos. Observation of such a violation would occur in neutrinoless double beta decay, which is one of the great searches in the future of high energy physics experiments.

Combining the Dirac mass terms with the Majorana mass terms and using the notation $\psi^c = -i\sigma_2\psi^*$ one gets the following in the Weyl basis:

$$\mathcal{L}_m = -\frac{1}{2}\begin{pmatrix}\psi_L^{\dagger} & \psi_R^{\dagger c}\end{pmatrix}\begin{pmatrix}m_L & m_D \\ m_D & m_R\end{pmatrix}\begin{pmatrix}\psi_L^c \\ \psi_R\end{pmatrix} + h.c. \tag{1.8}$$

Since the above matrix is not diagonal, the chiral neutrino flavor states do not have a definite mass. When this matrix is diagonalized, one obtains the neutrino mass eigenvalues:

$$m_{\pm} = \frac{1}{2}\Big(m_L + m_R \pm \sqrt{(m_L - m_R)^2 - 4m_D^2}\Big) \tag{1.9}$$

It is important to note that right handed chiral states do not participate in the weak interaction, which will be explained later. Therefore, m_L is necessarily 0 and m_R need not be 0. The reason for this is because the effect of the m_L term is to create a right handed anti-neutrino and destroy a left handed neutrino. This would mean that the Higgs Field has a weak isospin of 1. The m_R term creates a left handed anti-neutrino and destroys a right handed neutrino. Since these fields have no weak isospin, the Higgs Field has 0 weak isospin, If we assume that $m_R >> m_D$, we obtain the following approximation:

$$\begin{aligned} m_- &\approx \frac{m_D^2}{m_R} \\ m_+ &\approx m_R(1 + \frac{m_D^2}{m_R^2}) \approx m_R \end{aligned} \tag{1.10}$$

It is worth noting that a very large m_R yields a small m_- and a large m_+. This is known as the See-Saw Mechanism, and this is a possible explanation for the tiny mass of the neutrino.

1.4 Weak Interactions

It has been experimentally observed that weak interactions violate Parity Conservation. In 1950, Chien-Shiung Wu observed that only left handed chiral particles participate in the weak interaction [7]. In order to have parity violation, weak currents have a vector-axial form (V-A). The weak charged current has the form:

$$J^\mu = \bar{\psi}_l \gamma^\mu (1 - \gamma_5) \psi_\nu + h.c. \tag{1.11}$$

and the neutral current has the form:

$$J^\mu = \bar{\psi}_l \gamma^\mu (g_V - g_A \gamma_5) \psi_\nu + h.c. \tag{1.12}$$

Here, ψ_l is the Dirac Spinor for the charged lepton, ψ_ν is the Dirac Spinor for the neutrino, and g_v and g_A are the couplings. Using properties of γ matrices, equation 1.11 may be rewritten as:

$$J^\mu = \frac{1}{4} \bar{\psi}_l \gamma^\mu (1 - \gamma_5)(1 + \gamma_5) \psi_\nu + h.c. = \bar{\psi}_{l,L} \gamma^\mu \psi_{\nu,L} \tag{1.13}$$

From this, it is evident that the weak interaction only occurs between left handed chiral fermions, where chiral state fermions are given by:

$$\psi_L = \tfrac{(1-\gamma_5)}{2}\psi,\ \psi_R = \tfrac{(1+\gamma_5)}{2}\psi$$

Here, ψ_L is the left handed chiral state, and ψ_R is the right handed chiral state. It is important to note that anti-particles have the opposite chiral states:

$$\bar{\psi_L} = \tfrac{(1+\gamma_5)}{2}\psi,\ \bar{\psi_R} = \tfrac{(1-\gamma_5)}{2}\psi$$

This shows that the left handed charged leptons and their corresponding neutrino form a weak isospin doublet, and that the right handed particles are weak isospin singlets. Right handed charged leptons, however, can participate in neutral current interactions because their electric charge gives them a weak hypercharge, given by:

$$Y_W = 2(Q - I_3) \tag{1.14}$$

Right handed charged leptons couple to the Z^0 boson as $-Qsin^2\theta_W$, where θ_W is the Weinberg Angle.

1.5 Neutrino Oscillations

Neutrinos were first thought to be massless particles. If this were the case, then neutrinos would travel at the speed of light and have infinite time dilation. Therefore, any oscillation in the neutrino frame of reference cannot be observed in any inertial frame of reference. As a result of this, there exists a one-to-one correspondence between eigenstates of the free particle propagation Hamiltonina, mass eigenstates, and eigenstates of the Weak interaction, flavor eigenstates.

The phenomenon of neutrino oscillations was first hypothesized by B. Pontecorvo. He predicted that neutrinos oscillated between particle and antiparticle states [8]. Evidence for neutrino oscillations first appeared when a deficit in the solar neutrino flux in the Solar Neutrino Search Experiment at the Homestake Mine was discovered [9]. The experiment expected that the total flux of neutrinos from the sun would be electron type neutrinos. $\nu_e \to \nu_\mu$ oscillations were suggested by Pontecorvo and Gribov to explain the deficit in the solar ν_e flux [8]. This hypothesis, however, was later shown to be incorrect when the MSW Effect in the sun was developed.

1.5.1 Oscillations in the Vacuum

The mathematical formalization of neutrino oscillations begins with the expansion of a particular neutrino flavor state $|\nu_\alpha\rangle$ in terms of the neutrino mass basis $|\nu_i\rangle$:

$$|\nu_\alpha\rangle = |\nu_\alpha\rangle = \sum_{i=1}^{3} |\nu_i\rangle \langle \nu_i | \nu_\alpha\rangle = \sum_{i=1}^{3} U_{i\alpha} |\nu_i(t)\rangle \tag{1.15}$$

Here, U_{if} are components of a 3×3 unitary, mixing matrix. Using Euler Angles and a complex phase, this matrix is conveniently parameterized as the Pontecorvo, Maki, Nakagawa, Sakata (PMNS) Matrix.

$$U = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{CP}} \\ -s_{12}c_{13} - c_{12}s_{23}s_{13}e^{i\delta_{CP}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{CP}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{CP}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{CP}} & c_{23}c_{13} \end{pmatrix} \tag{1.16}$$

Here, $s_{ij} = \sin(\theta_{ij})$, $c_{ij} = \cos(\theta_{ij})$, and δ is the CP violating phase, which gives non-identical oscillation probabilities for neutrinos and anti-neutrinos, if it is nonzero.

The dynamics of neutrino propagation through free space are realized by working in the mass eigenbasis. The time evolution of the mass eigenstate $|\nu_i(t)\rangle$ is governed by the Schrodinger Equation:

$$H |\nu_i(t)\rangle = i\partial_t |\nu_i(t)\rangle = E |\nu_i(t)\rangle \tag{1.17}$$

Where H is the Hamiltonian for free particle propagation. The solution of this equation is the familiar plane wave solution:

$$|\nu_i(t)\rangle = |\nu_i(0)\rangle e^{i(\mathbf{p_i}\cdot\mathbf{x_i} - E_i t)} \tag{1.18}$$

It is then convenient to align the momentum with the direction of propagation and to use the fact that the neutrinos are almost always ultra-relativistic: $t \approx L$. Here, L is the length of propagation, and natural units are used: $\hbar = c = 1$. This leads to:

$$|\nu_i(t)\rangle = |\nu_i(0)\rangle \, e^{iL(p-E)} \tag{1.19}$$

For a relativistic particle, $E = \sqrt{p^2 + m^2}$. Since neutrinos have mass energy much smaller than the kinetic energy, it is acceptable to simplify E by means of a Taylor Expansion:

$$E = p\Big(1 + \frac{m^2}{p^2}\Big)^{1/2} \approx p + \frac{m^2}{2p} \tag{1.20}$$

In the ultra-relativistic limit, $p \approx E$. This result may be inserted into equation 1.19 to yield:

$$|\nu_i(t)\rangle = |\nu_i(0)\rangle \, e^{\frac{-im^2L}{2E}} \tag{1.21}$$

Plugging this expression into equation 1.15 gives the following expression for the neutrino flavor state:

$$|\nu_\alpha(L)\rangle = \sum_{i=1}^{3} U_{i\alpha} \, |\nu_i(0)\rangle \, e^{\frac{-im_i^2L}{2E_i}} \tag{1.22}$$

From elementary quantum mechanics, the probability for a flavor state to oscillate to another flavor state is given by the projection squared of the new flavor state from the original:

$$P_{\alpha\to\beta}(L) = |\,\langle\nu_\beta|\nu_\alpha(L)\rangle\,|^2 \tag{1.23}$$

Now the form of equation 1.22 may be used to yield:

$$= \left|\left(\sum_{j=1}^{3} \langle\nu_j|\, U^*_{j\beta}\right)\left(\sum_{i=1}^{3} U_{i\alpha}\, |\nu_i(0)\rangle \, e^{\frac{-im_i^2L}{2E_i}}\right)\right|^2 \tag{1.24}$$

This expression simplifies by orthonormality, yielding:

$$
\begin{aligned}
&= \Big| \sum_{i=1}^{3} U_{i\beta}^{*} U_{i\alpha} e^{-i\frac{m_i^2 L}{2E_i}} \Big|^2 \\
&= \sum_{i=1}^{3} \sum_{j=1}^{3} U_{i\beta} U_{i\alpha}^{*} U_{j\beta}^{*} U_{j\alpha} e^{i\frac{\Delta m_{ij}^2 L}{2E}}
\end{aligned}
\tag{1.25}
$$

Where $\Delta m_{ij}^2 = m_i^2 - m_j^2$. This expression may be rewritten as:

$$
\begin{aligned}
P_{\alpha\rightarrow\beta}(L) = \delta_{\alpha\beta} - 4\sum_{i>j} \mathrm{Re}(U_{i\beta} U_{i\alpha}^{*} U_{j\beta}^{*} U_{j\alpha}) \sin^2\Big(\frac{\Delta m_{ij}^2 L}{4E}\Big) \\
+2\sum_{i>j} \mathrm{Im}(U_{i\beta} U_{i\alpha}^{*} U_{j\beta}^{*} U_{j\alpha}) \sin\Big(\frac{\Delta m_{ij}^2 L}{2E}\Big)
\end{aligned}
\tag{1.26}
$$

It is important to note that this probability is dependent on the mass squared differences between the mass states, the energy of the neutrino, and the length of the neutrino propagation. There are two mass squared differences, which have different orders of magnitude. As a result of this, studying the larger mass squared difference, which is done in NOνA, will cause small corrections to the probability from the smaller mass squared difference term. In order to obtain the maximum oscillation probability, the argument of the first sin in equation 1.30 must be set to π. This yields the oscillation length: $L_{osc} = \frac{4\pi E}{\Delta m_{ij}^2}$.

In the ν_e appearance analysis, the probability for a muon type neutrino to oscillate into an electron type neutrino is the measurement of interest. This is given by:

$$
P_{\nu_\mu\rightarrow\nu_e} = P_{atm} + P_{sol} + P_{int} \tag{1.27}
$$

P_{atm} is the leading order term, which is governed by the oscillation length involved in atmospheric neutrino experiments:

$$
P_{atm} = sin^2\theta_{23} sin^2 2\theta_{13} sin^2\Delta_{31} \tag{1.28}
$$

Where $\Delta_{ij} = 1.27\frac{\Delta m_{ij}^2(eV^2)L(m)}{E(GeV)}$. The numerical factor arises from converting natural units into units used in the experiment. P_{sol} is the subdominant term, which is governed by the solar length:

$$P_{sol} = cos^2\theta_{13}cos^2\theta_{23}sin^2 2\theta_{12}sin^2\Delta_{21} \tag{1.29}$$

This term is small because it has been observed that Δ_{21} is a small value. P_{int} is the interference term between the atmospheric and solar length probabilities, given by:

$$P_{int} = cos\theta_{13}sin\theta_{13}sin\theta_{23}[cos\delta cos\Delta_{32}sin\Delta_{31}sin\Delta_{21} \pm sin\delta sin\Delta_{32}sin\Delta_{31}sin\Delta_{21}] \tag{1.30}$$

The positive sign is for neutrinos, and the negative sign is for anti-neutrinos.

1.5.2 Oscillations in Matter

In the previous section, the calculation of neutrino flavor oscillation assumed the neutrinos propagate in the vacuum. However, in many oscillation experiments the neutrinos travel through the earth's crust. When neutrinos propagate through matter, there is a nonzero probability of interaction between the neutrinos and the matter through which they propagate.

Neutrinos may interact with the electrons in the earth. As mentioned earlier, the neutrinos may interact by an exchange of a $W^{\pm}$ boson in a charged current interaction or by an exchange of a Z boson in a neutral current interaction. These interactions are represented by the Feynman diagrams in figure 1.3

It is important to note that since electrons are the only leptons in the surrounding material, only electron neutrinos may undergo coherent charged current interactions. Neutral current cross-sections, on the other hand, are equivalent for all neutrino flavors.

Matter gives an effective mass to the neutrinos, which changes the oscillation probabilities [9]. In NOνA, a highly pure beam, initially composed of muon type neutrinos propagate through the earth. During propagation, non-zero amplitudes develop for ν_e and ν_τ. Since electron type neutrinos differ from the other flavors, as a result of the nonzero cross-section for coherent charged current scattering, they experience an effective potential, $V_{CC} = \sqrt{2}G_F n_e$. G_F is the Weak Fermi Coupling Constant, and n_e is the electron number density in the medium. This effect modifies the oscillation parameters as follows:

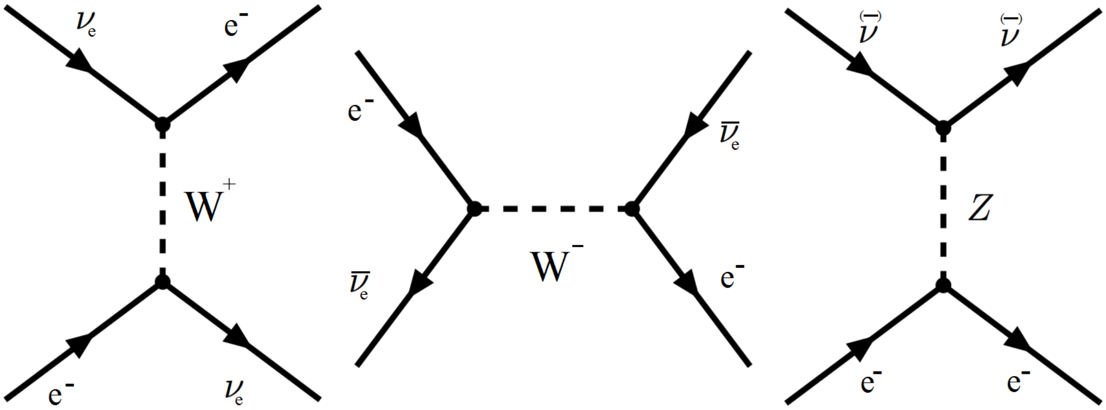


Figure 1.3: Feynman diagrams for coherent neutrino scattering. In the neutral current case, all neutrino flavors can participate. Whereas, in the charged current case, only electron type (anti)neutrinos can participate.

$$\begin{aligned}\Delta m^*_{32} &= \sqrt{\Delta m^2_{32}\sin^2 2\theta_{13} + (2EV_{CC} \mp \Delta m_{32}\cos 2\theta_{13})^2} \\ \sin^2 2\theta^*_{13} &= \frac{\Delta m^2_{32}\sin^2 2\theta_{13}}{(\Delta m^*_{32})^2}\end{aligned} \tag{1.31}$$

In the above equation, the negative term is for neutrinos, and the positive term is for anti-neutrinos . The phenomenon of the matter effect creating effective oscillation parameters is known as the Mikheyev-Smirnov-Wolfenstein (MSW) Effect [10]. At baselines on the order of 1000 km and energy on the order of 1 GeV, a ν_e resonance occurs, which greatly enhances or suppresses the probability for $\nu_\mu \rightarrow \nu_e$. This is given by:

$$P^{matter}_{\mu\rightarrow e} \approx \Big(1 \pm \frac{2EV_{CC}}{\Delta m^2_{32}}\Big) P^{vacuum}_{\mu\rightarrow e} \tag{1.32}$$

In the NOνA experiment, the experimental topic of this paper, this corresponds to approximately a 30% enhancement or suppression in $\nu_\mu \rightarrow \nu_e$ transition probability, depending on the sign of Δm^2_{32}.

One must understand that this asymmetry between ν_e and $\bar{\nu}_e$ is not due to CP violation. This asymmetry may either enhance or decrease the effect of CP violation in the appearance probability. Therefore, the matter effect may resemble CP violation in the data. However, by measuring both neutrinos and anti-neutrinos, it may be possible to disentangle CP violation from the matter effect. NOνA, in conjunction with other experiments, seeks to measure or limit CP violation and resolve the sign of Δm^2_{32}. NOνA is the longest baseline experiment, which means that it is the most sensitive to measuring Δm^2_{32}, as seen from equation 1.32.

1.6 Neutrino Mass Hierarchy

An important parameter being measured in the NOνA Experiment is the hierarchy of the neutrino masses. From previous experiments which measured neutrinos from the atmosphere, it was discovered that the absolute value of the mass squared difference between m_1 and m_3, Δm^2_{31}, is equal to $2.44 \times 10^{-3} eV^2$ [11]. From experiments which measured neutrinos from the sun, it was discovered that the mass squared difference between m_1 and m_2, Δm^2_{21}, is equal to $7.53 \times 10^{-5} eV^2$ [12]. Different oscillation experiments are sensitive to different magnitudes of mass squared difference for a given energy and baseline length. This is summarized in table 1.1.

Source	E_ν in MeV	L in km	Δm^2 in eV^2
Solar	~ 1	1.5×10^8	$\sim 10^{11}$
Atmospheric	$\sim 10^3$	$\sim 10^4$	$\sim 10^{-4}$
Reactor	~ 1	~ 1	$\sim 10^{-3}$
Reactor	~ 1	~ 1	$\sim 10^{-5}$
Accelerator	$\sim 10^3$	~ 1	~ 1
Accelerator	$\sim 10^3$	$\sim 10^3$	$\sim 10^{-3}$

Table 1.1: Mass Hierarchy Sensitivities

It is important to note that $\Delta m^2_{21} << \Delta m^2_{31}$. As a result of this, $\Delta m^2_{32} \approx \Delta m^2_{31}$. This information, however, is not sufficient to describe the hierarchy of the neutrino masses. The sign of Δm^2_{32} is still not known. As shown in figure 1.4, $m_3 > m_2, m_1$ or $m_3 < m_2, m_1$. NOνA seeks to measure the sign of the mass squared difference between m_3 and m_2.

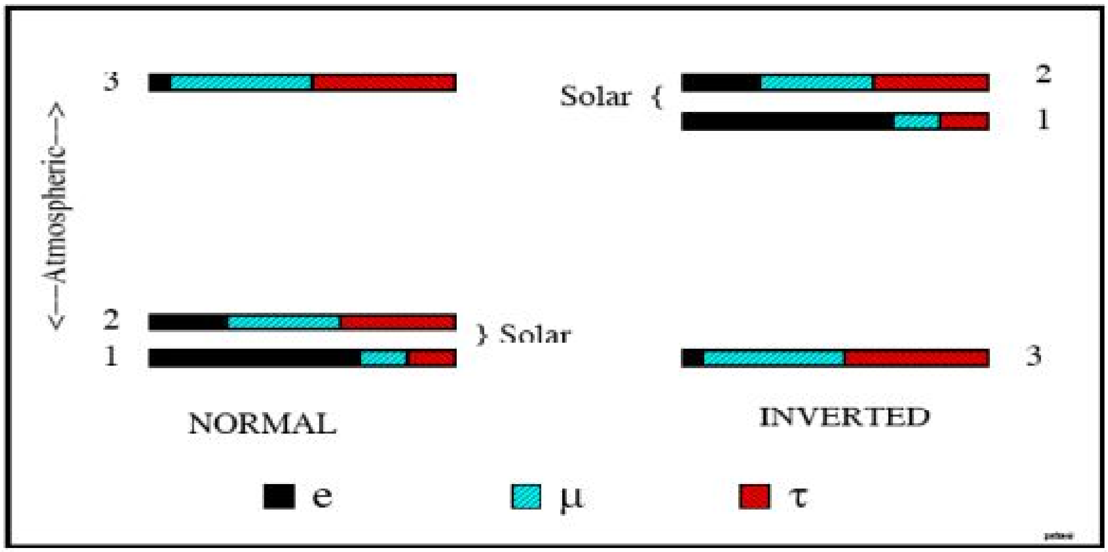


Figure 1.4: The left corresponds to the normal hierarchy, while the right corresponds to the inverted hierarchy. The flavor composition of the mass states are given by the colors.

1.7 CP Violation in Neutrino Oscillations

One of the most important physics targets in the NOνA experiment is the magnitude of CP violation in neutrino oscillations. This parameter can be accessed by measuring neutrino oscillations and anti-neutrino oscillations, under the same conditions. Measurement of this parameter will quantify the difference in oscillation probability for neutrinos and anti-neutrinos. If nature permits it, a measurement of δ_{CP} can be measured with modest statistical significance.

This parameter comes to be as a result of the complex nature of the PMNS Matrix. For CP symmetry to be broken, T symmetry must also be broken to maintain CPT symmetry, which is a conserved symmetry of nature. As a result of the parameterization of the PMNS Matrix, the CP violating phase, δ_{CP}, is the parameter which quantifies the magnitude of CP breaking in neutrino oscillations.

Chapter 2

Neutrino Oscillation Experiments

The goal of all neutrino oscillation experiments is to either measure the probability of appearance or disappearance of a neutrino flavor from an initial flavor. Currently, there are four types of neutrino oscillation experiments: solar, reactor, atmospheric, and accelerator. Each of these experiments have certain conditions, which make them suitable for measuring certain parameters of neutrino oscillation. This is due to the ratio of the energy of the produced neutrinos and the length of propagation.

2.1 Solar Neutrino Experiments

The first type of neutrino experiment conducted was a solar neutrino experiment. In the late 1960's, Raymond Davis conducted a search for neutrinos from thermonuclear reactions in the sun. The experiment consisted of a 380 m^3 detector located 1478 meters underground in the Homestake Gold Mine in South Dakota. The detector was filled with Perchloroethylene, a common chemical used in dry cleaning. Perchloroethylene was used for its rich Chlorine content. Neutrinos with energies from fusion processes in the sun undergo capture in Chlorine by the capture process:

$$\nu_e +^{37} Cl \rightarrow^{37} Ar + e^- \tag{2.1}$$

The threshold energy of the neutrino in this process is 0.814 MeV. By "bubbling" the detector with Helium, Davis was able to collect the Argon which resulted from the capture process and count the Argon nuclear decays [9] There are several processes in solar fusion reactions which produce neutrinos. The detector in the Homestake Experiment was sensitive to neutrino energies at the MeV scale. Theorist John Bahcall calculated the expected neutrino energy spectrum, which was encapsulated in the Standard Solar Model (SSM). SSM predicts that 91% of the neutrinos produced in the sun come from the proton proton fusion process:

$$p + p \to d + e^+ + \nu_e \tag{2.2}$$

The maximum energy of the neutrino from this process has a maximum energy of 0.4 MeV, making it short of the threshold energy of the ν_e capture process in Chlorine. Therefore, neutrinos from higher processes in the fusion chain reaction were needed. The following chain reaction produced neutrinos with energies above threshold for the experiment:

$$\begin{aligned} d + p &\to^3 He + \gamma \\ ^3He +^3 He &\to^4 He + 2p \\ ^3He +^4 He &\to^7 Be + \gamma \\ ^7Be + p &\to^8 B + \gamma \\ ^8B &\to^8 Be + e^+ + \nu_e \end{aligned} \tag{2.3}$$

The neutrinos from the ^8B decay have a maximum energy of 14 MeV, and they represent 1% of the total flux of neutrinos from the sun [11]. The experiment measured roughly one third of the neutrinos that were predicted in SSM. Further solar neutrino oscillation experiments measured this deficit as well. It was for this reason that Bruno Pontecorvo hypothesized that there was a flavor oscillation occurring in the solar neutrinos.

Neutrinos which are produced in the center of the sun traverse an exponentially decaying density of electrons to reach the surface. At the neutrino energies of interest for solar experiments, the electron density close to the center of the sun is near the resonant density, as previously discussed. The combination of these phenomena lead to an adiabatic evolution of the ν_e flavor state into the ν_2 mass state. Using the two flavor

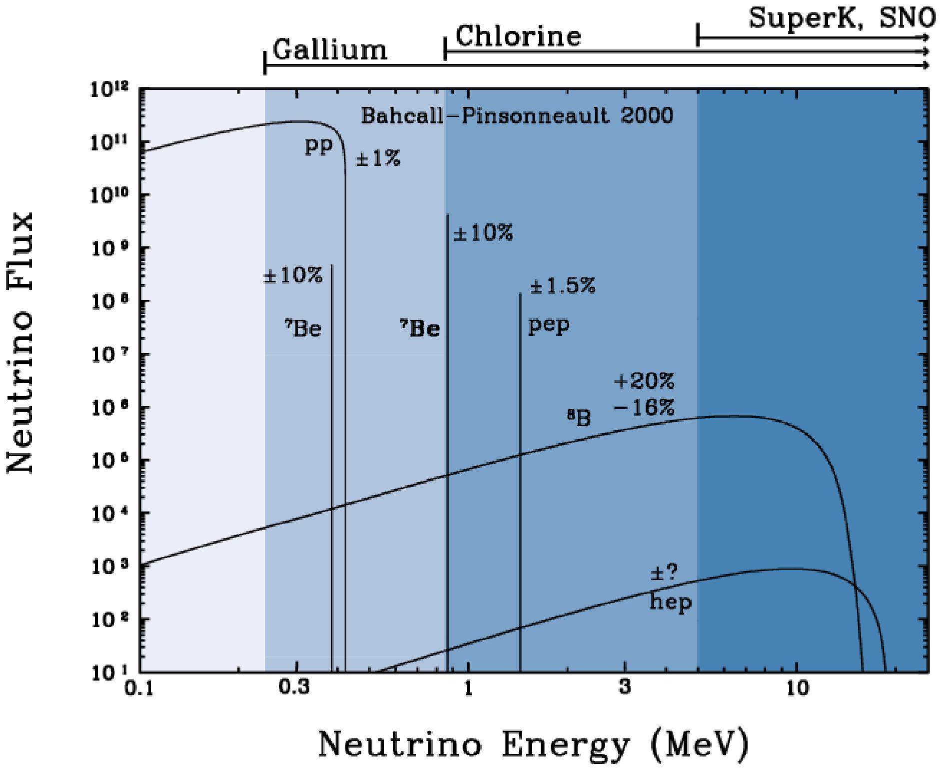


Figure 2.1: Spectra of the neutrinos from the reactions in the fusion chain. Neutrinos from the ^8B decay were observed in the Homestake Mine Experiment

oscillation model, we have:

$$\begin{aligned} |\nu_1\rangle &= \cos\theta\,|\nu_e\rangle + \sin\theta\,|\nu_\mu\rangle \\ |\nu_2\rangle &= -\sin\theta\,|\nu_e\rangle + \cos\theta\,|\nu_\mu\rangle \end{aligned} \tag{2.4}$$

Therefore, the probability of electron flavor survival is given by:

$$P_{\nu_e \to \nu_e} = |\,\langle \nu_e|\nu_2\rangle\,|^2 = \sin^2\theta \tag{2.5}$$

Neutrinos of lower energies do not undergo the adiabatic evolution into a pure mass state. Instead, to good approximation, they are produced in the ν_e flavor state and oscillate during propagation to earth. Once again, using the two flavor model, the probability for electron flavor survival is given by:

$$P_{\nu_e \to \nu_e} \approx 1 - \sin^2 2\theta \sin^2(\frac{\Delta m^2_{21} L}{4E}) \tag{2.6}$$

By combining the results from experiments which used neutrinos from Boron decays, such as the Homestake Experiment with experiments which used lower energy neutrinos from the proton-proton fusion process, such as Super-K, the oscillation parameters θ_{12} and Δm^2 were determined to be [12]

$$\begin{aligned} \sin^2(2\theta_{12}) &= 0.846 \pm 0.021 \\ \Delta m^2_{21} &= (7.53 \pm 0.18) \times 10^5\ eV^2 \end{aligned}$$

In 2001, the Sudbury Neutrino Observatory (SNO) Experiment, carried out by Arthur McDonald performed a flavor blind measurement of solar neutrinos and verified the solar model predictions [13].

2.2 Atmospheric Neutrino Experiments

Another large source of neutrinos which reach the surface of the earth are neutrinos which result from the collisions of cosmic ray protons with the nuclei in the atmosphere. The collisions produce a large number of pions. Neutrinos are produced from the charged pions by their decays into muons and muon type neutrinos. Muons which decay in the atmosphere also decay into electrons, muon type neutrinos, and electron type neutrinos, as seen in figure 2.2.

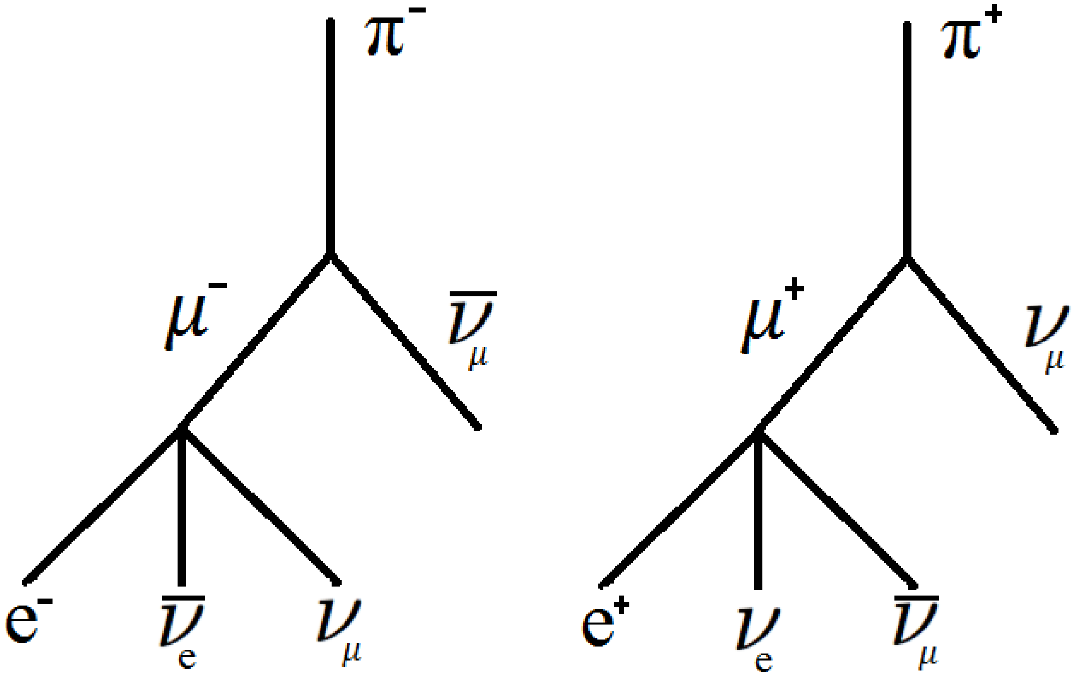


Figure 2.2: Mesons resulting from cosmic ray proton collisions undergo leptonic decays. Without oscillations, there should be twice as many muon type neutrinos as electron type neutrinos.

As seen in figure 2.2, one should expect to observe twice as many muon type neutrinos as electron type neutrinos. The Super Kamiokande Experiment, however, found that there was a far lower ν_μ CC interaction rate from neutrinos which passed through the earth versus neutrinos which entered the detector from above. The reason for this is because the matter effect from the earth enhanced the probability of oscillation from ν_μ to ν_τ. With neutrino energies ranging from the GeV to the TeV scale and baseline lengths used in atmospheric experiments, atmospheric neutrino oscillation experiments are sensitive to measuring $|\Delta m^2_{32}|$ and θ_{23}.

2.3 Reactor Neutrino Experiments

Another type of experiment which studies neutrino oscillations is a reactor experiment. Reactor experiments use $\bar{\nu}_e$'s from the Beta Decays of nuclei in fission reactors. These neutrinos are produced at the MeV scale. As a result of this, the detectors of these experiments are usually placed in close proximity to the nuclear reactors. This minimizes the matter effect and allows for greater sensitivity to measure the oscillation parameter, θ_{13} and $|\Delta m^2_{32}|$. This increased sensitivity to these parameters is the result of the $\frac{L}{E}$ minimizing terms in the oscillation probability with the parameter Δm^2_{21}. Therefore the two flavor model can be used to good approximation, yielding the survival probability:

$$P_{\bar{\nu}_e \to \bar{\nu}_e} \approx 1 - \sin^2 2\theta_{13} \sin^2 \left(\frac{\Delta m^2_{32} L}{4E} \right) \tag{2.7}$$

Reactor experiments determine the survival probability of $\bar{\nu}_e$ by measuring the rate of the absorption process: $\bar{\nu}_e + p \to e^+ + n$. Experiments such as Daya Bay, RENO, and Double CHOOZ, used Gadolinium doped liquid scintillator, which produces a flash of light from the annihilation of the positron with an electron in the scintillator. The neutron then thermalizes and is absorbed by the Gadolinium nucleus, which causes a delayed signal.

Daya Bay was the first experiment to measure that $\theta_{13} \neq 0$, and this measurement was made with 5.2σ significance [14]. The RENO and Double Chooz Experiments confirmed the results of Daya Bay. Using a joint fit from the reactor experiments, the following result was obtained: $\sin^2(2\theta_{13}) = 0.093 \pm 0.008$ [12].

2.4 Accelerator Experiments

The final class of neutrino oscillation experiments is the accelerator experiment. Accelerator experiments have the best control of $\frac{L}{E}$, which allows them to measure certain oscillation parameters with high sensitivity. Accelerator experiments function by smashing protons into a fixed target. These collisions produce a shower of hadrons and mesons. Charged pions and kaons can be focused by their charge, which controls the polarity of the neutrino beam. These neutrinos are aimed at a massive detector in order to measure the charged current interactions of the neutrinos.

An example of an accelerator experiment is the MINOS Experiment. MINOS used the NuMI beam from Fermilab and measured neutrino interactions in the Soudan Mine in northern Minnesota. With a baseline length of 734 km, and beam energies at the GeV scale, MINOS was optimized to measure θ_{23} and $|\Delta m^2_{32}|$. MINOS obtained the following results: $|\Delta m^2_{32}| = 2.32^{+0.12}_{0.08} \times 10^{-3} eV^2$ and $\sin^2(2\theta_{23}) > 0.90$ (90% confidence limit) [15].

NOνA, the experiment discussed in this thesis, is also an accelerator experiment, and it will be discussed in detail in the following chapter.

Chapter 3

The NOνA Experiment

The NuMI Off-Axis ν_e Appearance (NOνA) Experiment is a long baseline accelerator experiment which is designed to measure the appearance probability of electron neutrinos from a muon neutrino beam. NOνA will have greater sensitivity than previous long baseline experiments, as a result of its large beam power and baseline length. The NOνA Experiment consists of the Neutrinos at the Main Injector (NuMI) Beam and two functionally identical detectors. There is a near detector placed in the vicinity of the neutrino source, and there is a far detector placed 810 km, 14 mrad off-axis, from the neutrino source in Ash River, Minnesota. The long baseline length allows for maximal matter effect for neutrinos at energies optimal for oscillation. The off-axis beam creates a more narrow beam energy spectrum, which yields a larger fraction of oscillated neutrinos and smaller background.

3.1 The NuMI Beam

The NuMI Beam is created by colliding 120 GeV protons at a power of nearly 700 kW from the Main Injector ring at Fermilab with a fixed graphite target. The result of the collision is a combination of pions and kaons.

These particles then travel toward magnetic horns, which are tuned in such a way to focus the particles of a certain charge and to defocus particles of the opposite charge. Horns which are tuned to focus positively charged particles are called to be in the forward horn current (FHC) mode. This mode produces a beam of ν_μ, due to the decay: $\pi^+ \rightarrow \mu^+ + \nu_\mu$ with a 99.98% branching fraction, and $K^+ \rightarrow \mu^+ + \nu_\mu$ with a 63.55% branching fraction. The reverse horn current (RHC) is the opposite mode, and the beam will be mostly composed of $\bar{\nu_\mu}$. Near the beam energy peak, the contamination from anti-particle neutrinos is small. For the forward horn current, there is less than 2% $\bar{\nu}_\mu$ contamination, and for the reverse horn current, there is approximately 10% ν_μ contamination. These values were obtained from simulation. Understanding beam contamination from the data is currently undergoing exploration. Since μ^+ and μ^- look identical while traversing the detector, one must look at activity at the end of the muon's track. Negatively charged muons have the ability to be captuered in an atomic orbit and decay into electrons, whereas the positively charged muons collide with the target nuclei and produce hadronic showers. The outgoing electrons and neutrons from nuclear collisions on average occur much later in time than the neutrino interaction, which means these are the most probable source of out of time energy deposition and the end of the track and can thus be used to better understand the flavor composition of the NuMI beam.

The hadrons and muons which do not decay are then absorbed by steel and earth at the end of the decay pipe. This leaves a beam of mostly muon type neutrinos. There exists, however, some electron neutrino background due to the following processes: $K^+ \rightarrow \pi^0 + e^+ + \nu_e$ with a branching fraction of 5.07% and $\mu^+ \rightarrow e^+ + \nu_e + \nu_\mu$ with a branching fraction of nearly 100%. The same holds true for the charge conjugated processes. These beam ν_e's are removed by considering a signal region of 1-3 GeV for ν_e events. This is the result of the neutrino energy being peaked at 2 GeV. Around this energy, there is a larger probability for $\nu_\mu \rightarrow \nu_e$ oscillation. The majority of the beam ν_e's will fall outside of this energy window due to the larger phase space of the kaon decays and the smaller phase space of the muon decays.

In order to control the beam energy and flux of the NuMI Beam, the relative distance between the target and the first horn and the distance between the two horns can be adjusted [16].

The energy spectrum and flux of the neutrinos can be calculated, using relativistic kinematics. The hadrons produced are spinless, which means that the their rest frame decays are isotropic in order to conserve angular momentum. In the lab frame, however, the effect of boosting leads to the energy and flux to be a function of the angle of the

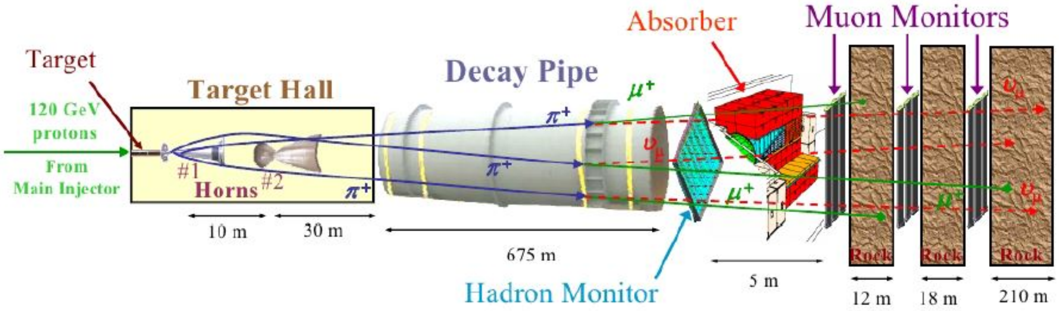


Figure 3.1: Schematic showing the components of the NuMI Beamline.

beam center to the target. Using the data from the mesonic decays and the knowledge gained from target hall simulations, the NuMI spectra can be understood.

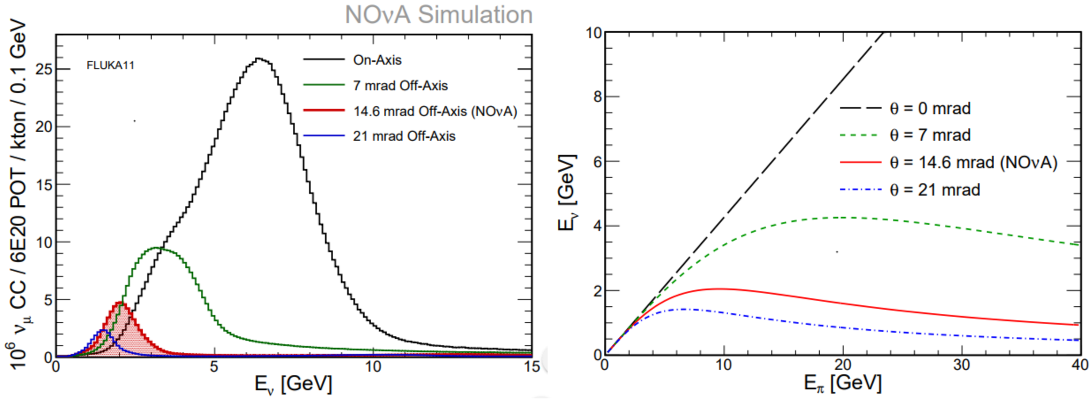


Figure 3.2: Left: Neutrino energy flux vs. neutrino energy for various off-axis angles. Right: Neutrino energy vs. meson energy for various off-axis angles.

The NOνA experiment utilizes the NuMI beam 14 mrad off-axis from the centerline. This configuration was chosen to produce a narrow neutrino energy peak around 2 GeV, as seen in Figure 3.3. A narrow energy beam with this peak location is optimal for measuring ν_e appearance with much lower background.

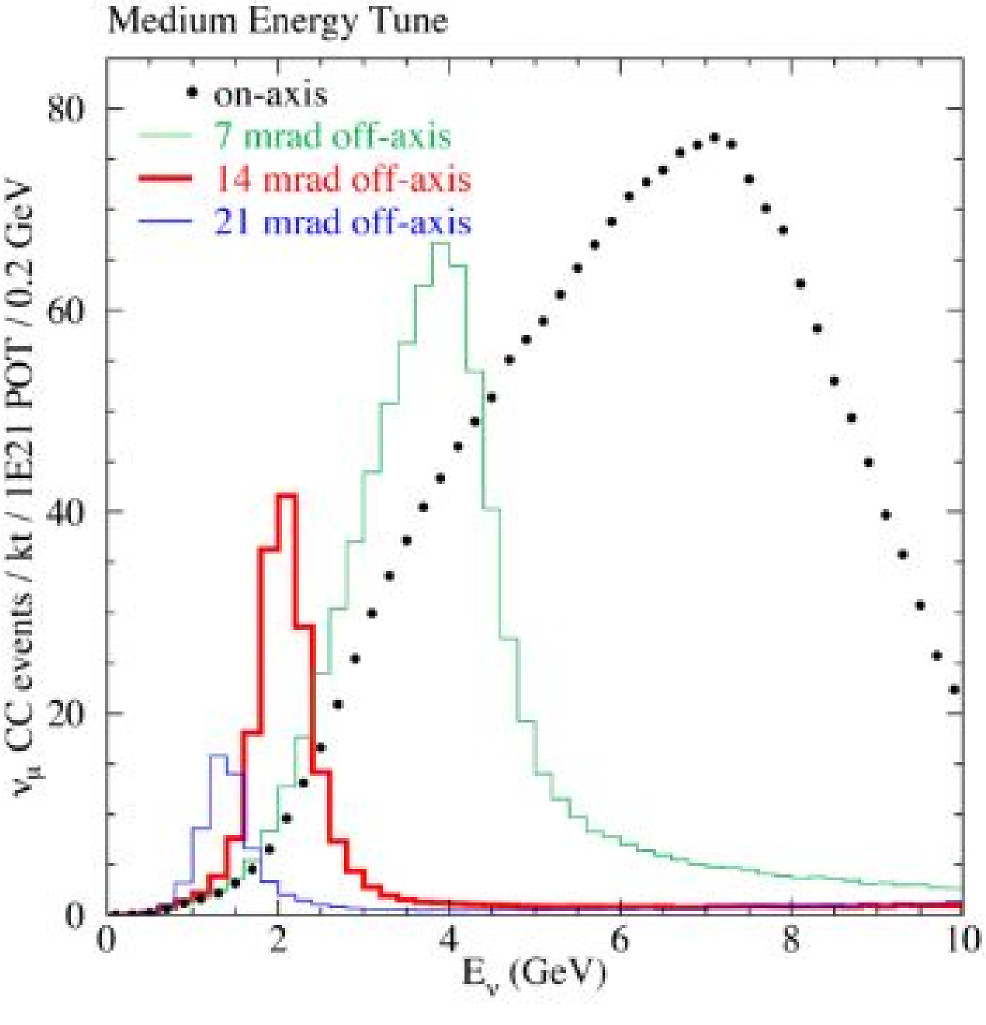


Figure 3.3: NOνA uses a 14 mrad off-axis beam, resulting in a narrow peak around 2 GeV which is shown in red.

3.1.1 The Fermilab Accelerator Complex

The process for producing the 120 GeV proton beam at Fermilab begins with protons from a linear accelerator being fed into a booster ring which accelerates the protons to an energy of 8 GeV. These protons were originally fed into the Main Injector Ring which accelerated them to an energy of 120 GeV. This was the process which was used by the MINOS Experiment, and it resulted in a beam power of roughly 350 kW.

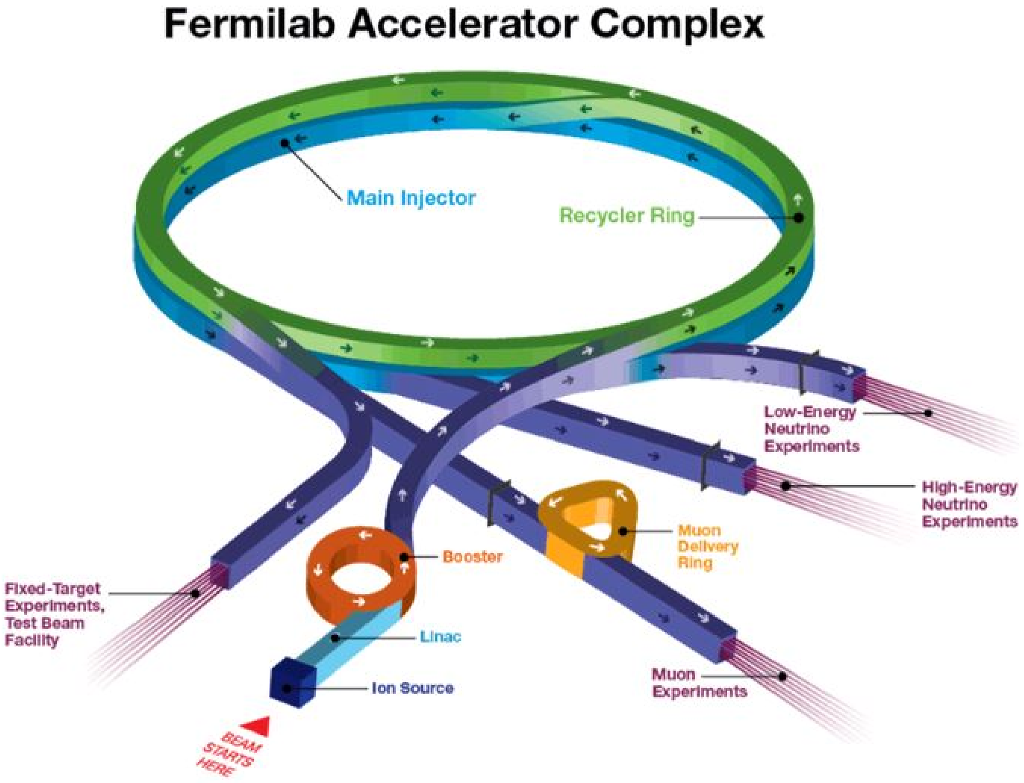


Figure 3.4: The accelerator complex at Fermilab which is used to produce the NuMI Beam.

In order to reach the beam power requirements of NOνA, several improvements were needed. First, the cycle time for the Main Injector was reduced from 2.2 to 1.3 seconds. The next improvement was to use the anti-proton storage ring from the Tevatron as a recycler ring for the protons from the booster. With the recycler ring, the protons may be fed into the Main Injector by a process called slip-stacking. Slip-stacking is the process by which protons from the recycler ring are fed into the main injector with a particular timing and momentum configuration.

3.2 The NOνA Detectors

Cross-sections for neutrino interactions are very small. It is for this reason that a highly intense NuMI beam is used. In addition to an intense beam, it is necessary that the detectors of neutrino experiments be sufficiently massive. In the NOνA experiment, there are two detectors: the near and far detector. Each of these detectors uses orthogonally rotated, stacked planes, but they have different purposes.

The near detector is roughly a 300 ton detector, which is placed in a cavern 1 km away from the NuMI source. At its location, the near detector receives roughly 1 million times larger flux than the far detector. For this reason, the mass of the near detector need not be so large. The purpose of the near detector is to measure the energy distribution, which is the product of the flux and the interaction cross-section, before the neutrinos oscillate. This allows for a prediction of the beam composition at the far detector, thus decreasing reliance on simulation. The near detector must also have a long and narrow design to allow enough length for charged particles to be contained in the detector. The near detector has an iron muon ranger at the end to better contain muons, which are in the minimum ionizing region around the 1 GeV scale.

The far detector is a 14 kton detector placed 810 km from Fermilab at Ash River, Minnesota. This distance was chosen to maximize the probability for $\nu_\mu \rightarrow \nu_e$ appearance for the off-axis NuMI beam energy while remaining in the United States. The far detector will be used to observe the ν_e appearance signal. Since the far detector is so massive, it would have been a costly ordeal to place it underground. Strategic timing cuts for the NuMI beam must be implemented, and topological cuts in the data must be applied to reject most of the cosmic background. By using a 10 μs NuMI spill window and applying the topological cuts, cosmic background poses a minimal hindrance to the ν_e appearance analysis.

3.2.1 The Detector Cell

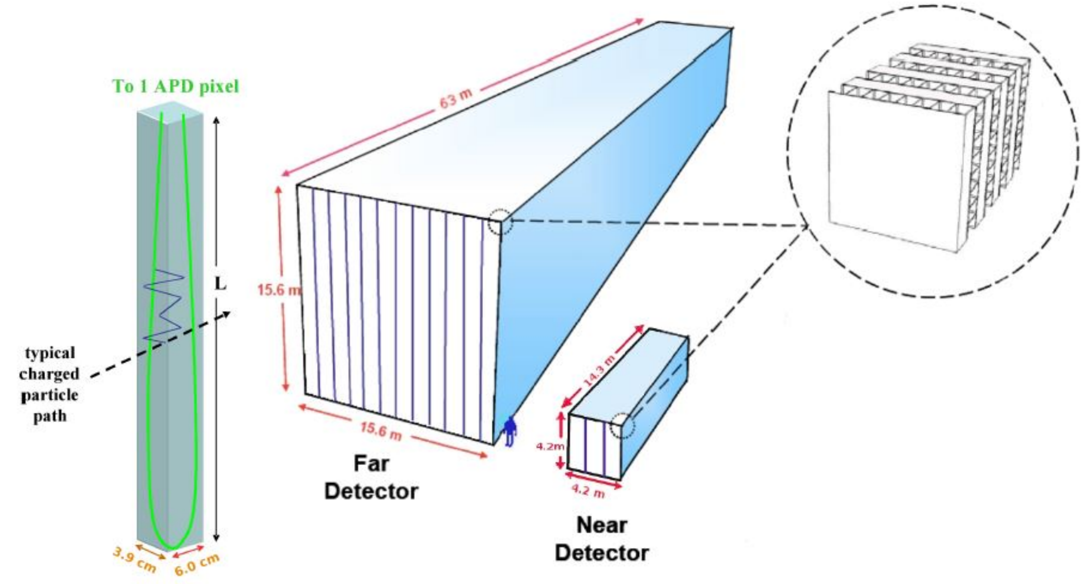
A NOνA detector cell consists of a hollow, 4 cm x 6 cm, rectangular cross-section tube made out of PVC. The ND cells are roughly 4 m. long, and the FD cells are roughly 15 m. long. These dimensions were chosen to allow for efficient assembly, transportation, and installation, given the detectors' fiducial mass requirements. Each tube consists of a wavelength shifting fiber (WLS), immersed in liquid scintillator. Nearly all of the liquid scintillator is composed of mineral oil, but 4.1% of it consists of pseudocumene [1,2,4-Trimethybenzene] as the scintillant. Pseudocumene produces photons in the ultra-violet spectrum when charged particles excite it. PPO [2,5-diphenyloxazole] and bis-MSB [1,4-di(methylstyryl)benzene] are used to shift the ultra-violet light to the visible light spectrum in the blue-light region. This is the optimal absorption region for the WLS fiber. The photons are absorbed by the WLS fiber, where many of them undergo total internal reflection and are shifted to green light. This is the region where the light is most efficiently absorbed by Avalanche Photo Diode (APD), which is located at the end of the cell. The quantum efficiency of the APD in this region is roughly 85%. This signal is then amplified and sent into a system of electronics, which can then extract timing and energy deposition information from the light.

3.2.2 Tracking

In order to obtain a useful spatial reconstruction of charged particles' trajectories, the detector cells needed to be strategically arranged. First, 32 cells were aligned side by side to form a plane. The planes were then stacked orthogonally and aligned with the direction of the beam. With this configuration, it is possible to describe the cartesian coordinates of a particle's trajectory. Planes which face vertically give the x coordinate, and planes which face horizontally give the y coordinate. The planes were chosen to face the beam direction to allow the beam direction provide the z coordinate. This can all be seen in Figure 3.5

3.2.3 Electronics and Triggering

As stated earlier, detector cells produce light as charged particles traverse them. In order to gain any useful information from this, there must exist a method to convert



(a) (b)

Figure 3.5: (a) A single cell of the detector. (b) The near and far detectors with a zoomed image of the alternating plane design.

such signals into an electronic form and effectively handle the large volume of those electronic signals.

NOνA accomplishes the first task by sending all of the wavelength shifting fibers from a particular plane into an avalanche photo-diode (APD). Due to their high quantum efficiency for the amplification of the shifted light, APD's provide a cost efficient way to convert and amplify the light signal into a useful electric signal. APD's are more heavily affected by shot and dark electrical noise than other amplifying technologies such as a photomultiplier tubes. This shortcoming is overcome by applying a voltage gain of 100 or 150 to the APD's and having them operate at -15 ^oC [17].

After a particular APD converts the light signal into an amplified electronic signal, the resulting electronic signal is fed into a four component system called the front end board (FEB). Each FEB consists of an an application-specific integrated circuit (ASIC), an analog to digital converter (ADC), a multiplexer, and a field programmable gate array (FPGA). The ASIC is a CR-RC circuit which increases the signal to noise ratio by amplifying and widening the electronic pulses. The ASIC is CR-RC circuit which amplifies and widens the signals in time to better distinguish physics signals from noise. The ADC converts the analog signal by digitizing it with respect to a 64 MHz digital clock. The digital signals from each pixel in the APD are then multiplexed by an eight-fold multiplexer for the far detector and a two-fold multiplexer in the near detector. The output of the multiplexer is fed into the FPGA which uses dual correlated sampling (DCS) to alert the data acquisition system (DAQ) that it has just sampled a physics hit. The procedure for determining the validity of a hit consists of comparing the ADC quanitities of hits temporally separated by three clock ticks. If the ADC discrepancy exceeds a predetermined threshold, then the pulse is determined to be a hit, and it is sent to the DAQ.

The DAQ receives hits from many sources including electronic noise, cosmic rays, neutrinos, and potentially supernova neutrinos. In order to determine the source of a particular hit, a triggering system needs to be applied. The NuMI beam spill has a duration of 10 μs and is repeated every 500 μs. Therefore, hits which occur in the beam spill window will be collected by the NuMI beam spill trigger. Cosmic rays, however, continuously traverse the detector. Therefore, in order to have a pure cosmic hit collection, a 10 Hz minimum bias trigger also is applied. The data collected by the

DAQ system is finally stored on hard disks at both near and far detector sites. From there, the data can be permanently stored on tape at Fermilab.

Chapter 4

Simulations

Every high energy physics experience is unique, whether it be by calorimetry, tracking, detection strategy, instrumentation, and external environment. Therefore, there are no theories or prior data which can adequately model the phenomena of an experiment. It is for this reason, monte-carlo (MC) simulation is used. By using experiment specific parameters and results from phenomenological theories and previous experimentation, one can develop highly accurate predictions with monte-carlo simulations. NOνA simulates neutrino beam flux, neutrino interactions, particle propagation, and detector response. In this chapter, the phenomena and methods for simulating such phenomena will be discussed.

4.1 Simulating the NuMI Beam

The first step in the simulation chain is modeling the content and flux of the neutrino beam from the decays of the products of the proton on target collisions. It is important to be able to model the neutrino flux for both the near and far detectors, so that the systematic uncertainties from the beam flux can be minimized by performing a near/far detector extrapolation of the ratio of the simulated and actual neutrino bean fluxes. The source of this systematic uncertainty is the proximity to the beam source for both of the detectors. Since the near detector is much closer to the beam source, it occupies a much larger solid angle than the far detector.

NOνA carries out the simulation, using three pieces of software: GEANT4 to simulate the particles traversing the detector hall geometry, FLUKA [18] to model the particle decays, and FLUGG [19] to interface between the previous two pieces of software. With these pieces of software, it is possible to trace the back from the outgoing neutrinos to the protons which collided with the target.

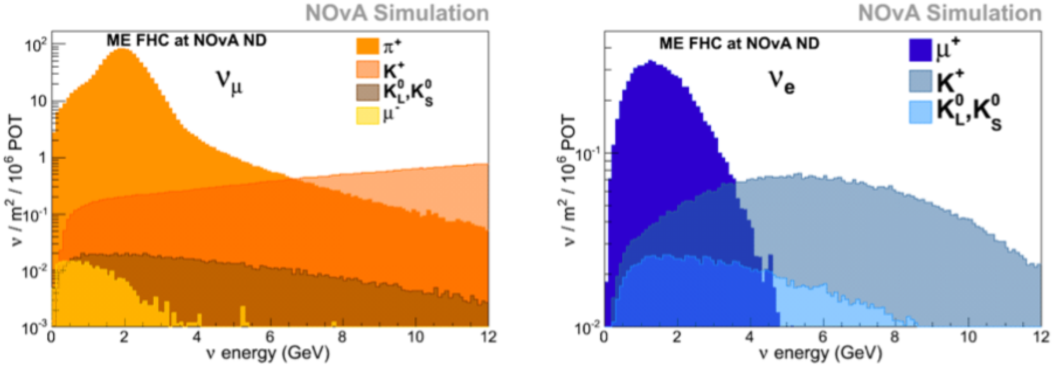


Figure 4.1: Lower energy neutrinos primarily come from pion decays, and high energy neutrinos primarily come from kaon decays.

4.2 Simulating Neutrino Interactions

Given the nature of the NuMI beam and the detector specifications, simulation software is needed to model the event rate of particular interactions and the kinematics of such interactions. NOνA uses the software GENIE to accomplish this. Genie is a simulation software which uses the composition of the detector and interaction cross section models to predict the event rate of various interactions as a function of interaction energy. In order to make an event rate estimation, GENIE uses the output of the beam simulation and convolves the neutrino beam flux with the interaction cross-sections. There are many interaction modes to be considered, but the most common ones include quasi-elastic scattering, resonant baryon production, deep inelastic scattering, coherent nuclear scattering, and meson exchange current scattering. These interactions have a strong energy dependence, and they tend to differ in their four momentum transfer, Q^2.

The quasi-elastic (QE) model was developed by Llewellyn Smith, and its name correctly describes the kinematics of this type of interaction. QE interactions occur in the low Q^2 region of phase space, which leads to the neutrino interacting with the whole nucleons, instead of their components. As a result of this, GENIE models the target of the interaction with a dipole form factor and solely gives the axial vector mass of the target as a free parameter [20]. Therefore, QE interactions are easier to classify due to the lower hadronic energy fraction in the interaction.

The GENIE implementation of resonant baryon production (RES) uses the Rein-Segal Model [21]. The Rein-Segal Model models the baryon resonances using the Feynman-Kislinger-Ravndal Model (FKR), which treats the baryon wavefunctions as excited states of three quarks bounded in a relativistic harmonic oscillator [22]. To simplify computations, GENIE neglects the interference between resonances and treats the outgoing leptons as massless. Just as with the QE case, GENIE leaves the axial vector mass as the only free parameter.

Deep inelastic scattering (DIS) is the phenomenon in which a neutrino interacts with just the quark inside of a nucleon, rather than the nucleon itself. DIS is modeled in two regions of the neutrino energy spectrum. At the lower end of the energy spectrum, there is a higher probability for smaller Q^2 transfer, so there must be a form factor correction. This is achieved by adding corrections from the Bodek and Yang model to the individual quark interaction [23]. At the higher end of the NuMI energy spectrum, DIS becomes the dominant interaction mode. In this region and interaction mode, the neutrinos transfer high amounts of Q^2 to the parton target. Therefore, it is in this region in which a neutrino interaction with a valence or sea quark becomes a good approximation. Unlike QE and RES, DIS has several free parameters which can be tuned to fit the spectra from the data. The weights of the reweighted model are determined by calculating the ratio of the modified cross-section to the original cross-section.

One of the more rare interactions to occur in NOνA is the coherent scattering of a neutrino off of the nucleus as a whole (COH). In this interaction mode, the neutrino does not exchange enough Q^2 to excite the nucleus into a higher energy state. Instead, the neutrino scattering causes the nucleus to recoil and thus collide with other nuclei. COH interactions, however, are most prominent in the MeV range, and thus their impact to the NOνA interaction simulation is negligible.

Up until now, all of the interactions described neglected nuclear interactions between the nucleons. These models treated the nuclei as a Fermi gas inside of the nuclear region. This was discovered to be a problem when the spectra of the simulation did not match that of the data. The data indicated that QE-like events were underrepresented. There was no method to tune the GENIE parameters to correct for this phenomenon without increasing the data/simulation discrepancy for the higher Q^2 interactions. As a result of this, nucleon-nucleon interaction in the nucleus was considered. More particularly, a 2 proton, 2 hole 2p2h model shows promising results. Under this model, there is a contribution to the neutrino/nuclear cross-sections by considering the meson exchange current between two of the nucleons (MEC). WIth MEC, there was an improved correspondence between the data and simulation for neutrino interactions.

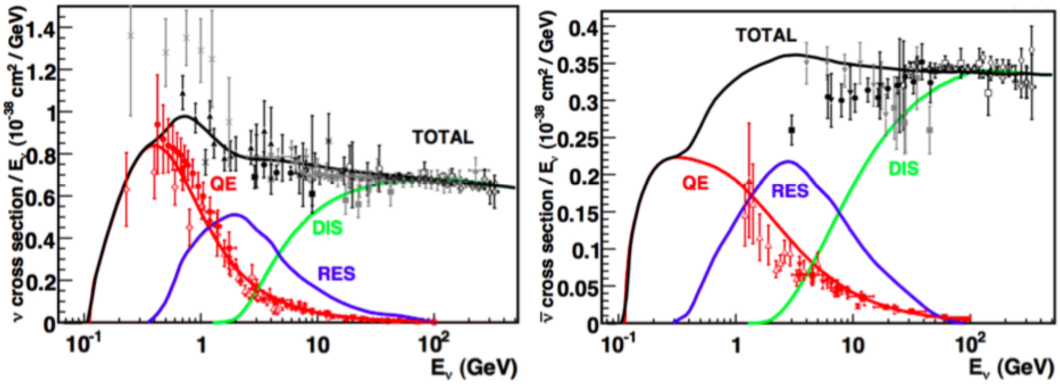


Figure 4.2: Neutrino interaction cross-sections for various interaction modes as a function of neutrino energy in the vicinity of NOνA's beam energy.

4.3 Simulating Particle Propagation

Once the interaction has been sufficiently modeled, it becomes necessary to model how the products of the interaction propagate through the detector. In order to do this, NOνA uses software called GEANT4, which inputs the detector geometry and models how particles interact and decay within the detector and its surroundings [16].

GEANT is highly tunable due to its vast physics lists. NOνA uses a particular physics list which best describes the particle trajectories for NOνA's particular detector geometry and kinematics. GEANT simulates particle trajectories step by step. At each step, the software selects the next action based on the probability drawn from the physics models. In order to avoid superfluous computational cost, an adaptive stepping procedure is employed which determines the step sized based upon the variance of the interactions in a given interaction length.

In each cell, the energy deposited into the scintillator and WLS is summed to form a simulated cell hit. An important consideration in forming this type of hit is a scintillating phenomenon known as Birk's Suppression [24]. Birk's Suppression occurs when a rapidly moving charged particle saturates the scintillator in its vicinity, which typically occurs toward the end of a particle's trajectory. Without this suppression, the light output per unit length, $\frac{dL}{dX}$ to be directly proportional to the energy output per unit length, $\frac{dE}{dX}$. Instead, the two quantities are related by:

$$\frac{dL}{dX} = L_0 \frac{\frac{dE}{dX}}{1 + k_B \frac{dE}{dX}} \tag{4.1}$$

Here, L_0 is a constant unique to the scintillator type and k_B is the Birks Suppression factor. To better match the data from the cells, NOνA adds an additional quadratic correction from the Chou Model [25]. The final model of the light response to a particle's energy deposition in a cell is given by:

$$\frac{dL}{dX} = L_0 \frac{\frac{dE}{dX}}{1 + k_B \frac{dE}{dX} + k_C \frac{dE}{dX}^2} \tag{4.2}$$

Using information from protons and muons in the ND data and MC, the suppression parameters were found by doing a simple χ^2 minimization technique.

4.4 Simulating Detector Response

The next step in the simulation chain is to take the quantities gathered from GEANT and model the way in which the detector cells and electronics respond. There needs to be a way to model the photon transport in the cell and the response of the electronic readout.

4.4.1 Simulating Photon Transport

Now that there is a sufficient model for light levels for the detector cells, there needs to be a method for converting those values into discrete photons to be read by the APD. Doing a direct measurement of the light collection and transport efficiency is impossible, due to the fact that the cells are closed. NOνA handles this dilemma by employing a simple ray-tracing algorithm which calculates the paths of wave fronts through media with varying refraction and boundaries with varying reflection. In this model, assumptions about the features of the scintillator, WLS, and cell boundaries need to be made. One of the features being exploited is the absorption and attenuation length of the photons in the WLS. This was found by using the attenuation curves from the cosmic ray muon studies which gave the best data/MC agreement. With the ray tracing model, an estimate of the photon collection rate as a function of energy deposition can be found along the surface of the fiber. From there, the propagation of the photons is simulated using GEANT4.

Since the fiber is in a loop, the photons which are collected are halved and sent in both directions of the fiber. The fiber attenuates the photons which travel through it. Therefore, the attenuation curve from the cosmic ray muon studies is used to estimate the mean number of photons as a function of position in the fiber. The number of photons to reach the APD is Poisson distributed, so a Poisson smearing factor is applied to the number prediction. Finally, the APDs have a quantum efficiency of 85%, causing a 15% reduction in the collected photons.

4.4.2 Simulating the Detector Response

Now that a prediction for the number of photo-electrons (PE) arriving at the APD have been made, the next task is to simulate the process by which the APDs convert the PE into cell hits. This is done by converting the analog signal into a digital signal and sending the signal to a dual correlated sampling node, which simulates the triggering method used for the data. Digitizing the analog signal is done by using an analog to digital conversion (ADC) with a temporal resolution of 62.5 ns. The magnitude of ADC values at each time step is scaled by the ratio of PE to the PE which saturates the APD. The ADC pulse is then shaped by a CR-RC circuit, as seen in figure 4.3.

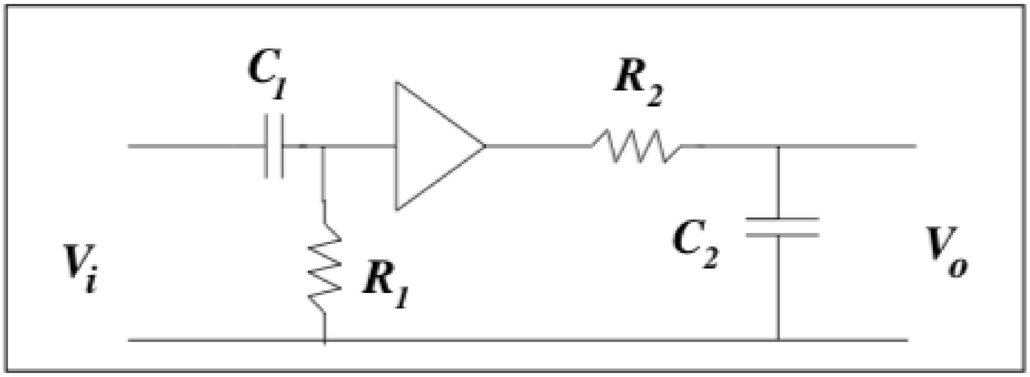


Figure 4.3: Circuit design of CR-RC circuit

The response of the ADC voltage is given by:

$$f(t) = f(0)\frac{\tau_1}{\tau_1 - \tau_2}(e^{-\frac{t-t_0}{\tau_1}} - e^{-\frac{t-t_0}{\tau_2}}) \tag{4.3}$$

Here, τ_1 is the time constant of the differentiator component of the circuit, τ_2 is the time constant of the integrator component of the circuit, and $t - t_0$ is the time difference between the PE response and input ADC pulse. Gaussian-Markov distributed noise is applied to the pulse to simulate the noise of the CR-RC circuit.

In the APD's, there is an effect in which charge deposited in one of the pixels reduces the charge of the other pixels by a factor of 1.86%. This effect is known as APD Sag. This is simulated by summing the ADC for each pixel, scaling the sum by 1.86%, and subtracting that amount for each pixel. Before this is done, each pixel's ADC is scaled up by 1.86% to handle the pixel's self sag. Since there is a non-linear relationship between a pixel's ADC count and the calorimatric energy of a cell, the effect of the sag is enhanced when considering calorimetric energy. Depending on the topology of a particular interaction, the APD sag's effect on calorimetric energy is in the range of 4-8%.

Finally after pulse shaping, the digitized ADC signal is sent to the DCS node, which considers the difference in pulse height for ADC counts separated by three time steps. If the difference in pulse height exceeds a predetermined threshold, the DCS algorithm triggers a cell hit. This fixed threshold is determined differently for data and MC.

Chapter 5

Reconstruction

In order to extract the physics out of an event in one of the detectors, the cell hits must be treated to exploit the features of the physics under investigation. This is known as reconstruction. The goal of reconstruction in the ν_e appearance analysis is to accentuate the features of ν_e charged current interactions while reducing the features from background, such as ν_μ charged current interactions, neutral current interactions, and cosmic rays. The particles from these interactions have different ionization signatures. Electrons, the target in the ν_e appearance analysis, begin as a thin column and widen throughout propagation. Muons, however, typically have a thin column profile throughout the entirety of propagation. This can be seen in Figure 5.1.

The reconstruction chain begins by calibrating the cell hits to convert light level readings in to energy measurements. After calibration, cell hits are clustered by space and time to find the hits which were involved in the neutrino interaction. Next, a Hough Transformation is applied to identify a line of hits which could belong to a particle trajectory. Following the Hough Transformation, the Elastic Arms vertex algorithm is applied to identify where the neutrino interaction occurred. Then, the Fuzzy-K Means clustering algorithm is applied to create three dimensional prong objects. Finally, the Reclustering Algorithm is applied to more appropriately encapsulate the showering nature of the electron in the NOνA detectors.

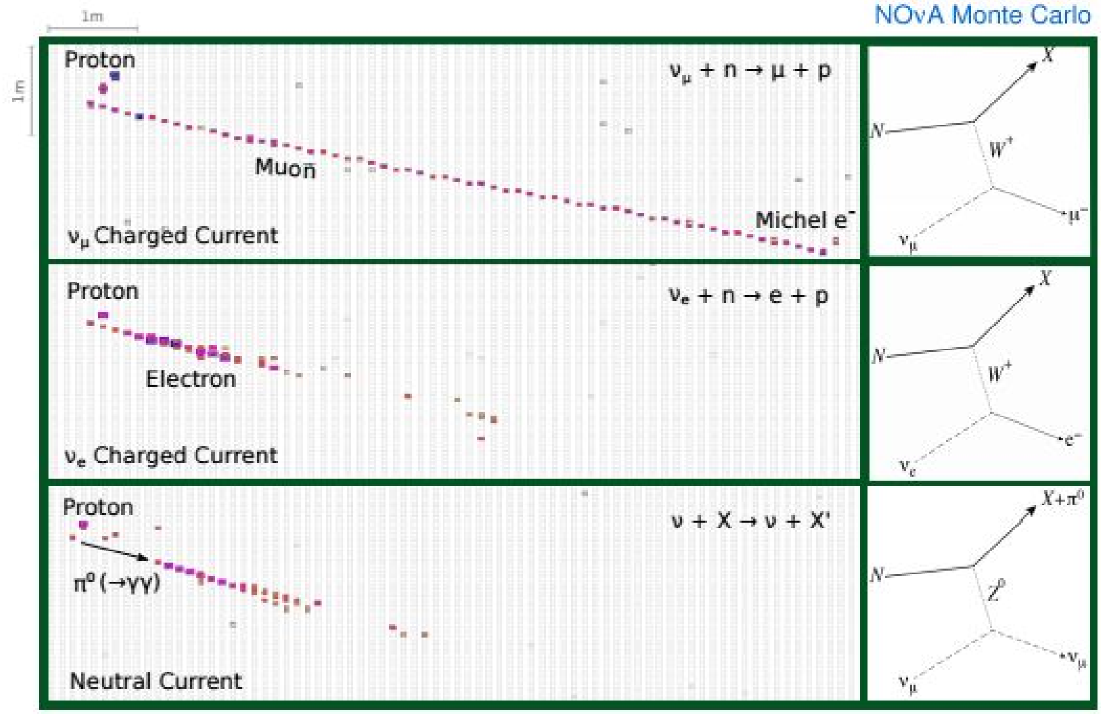


Figure 5.1: Signatures of various neutrino interactions. The shapes of the different outgoing particles have different features which can be exploited for a particular analysis.

5.1 Cell Hit Calibration

In order to be able to measure the energy of a neutrino interaction in the NOνA detectors, one must find a way to convert ADC signal into energy. To do this, NOνA uses muons from data, due to their abundant supply from cosmic rays in the far detector and neutrino interactions in the surrounding rock in the near detector.

The muons used for calibration must pass several quality cuts, and they must be sufficiently reconstructed using simple tracking algorithms. Due to the muon's predictable $\frac{dE}{dx}$ behavior, the muon's $\frac{dE}{dx}$ is used and the channel response is calibrated for each channel in the detectors. Tri-cells in a reconstructed track are used for calibration for two reasons: their use greatly reduces poor results from poor reconstruction, and it is known that the particle passed through both the top and the bottom of the cell, instead of traversing a corner of the cell. With this knowledge, one can obtain the equation for the path length of a particle traversing said tri-cell.

Another important quantity needed to properly calibrate the detectors is the mean position along the cell, also known as W. Using the track reconstruction and the neighboring planes, one may obtain an estimate for W. For convenience, the range of W is between -775 and 775 cm. Smaller values of W indicate that the particle's mean position is farther away from the readout, and the opposite holds true for larger values of W.

5.1.1 Cosmic Muon Bias Correction

Cosmic ray muons which enter the far detector do not have a uniform $\frac{dE}{dx}$ along their reconstructed tracks. Two phenomena cause such a bias in the visible energy which depends on the length along the track: shadowing and threshold effects. These must be corrected before a correction in the cell attenuation can be applied. Shadowing is an effect which is caused by the top of the detector shielding the bottom of the detector. Since cosmic ray muons primarily originate from the top of the detector, the hits toward the top of the detector are farther away from the end of the track.

The next effect which needs to be corrected is the threshold effect. For a cell to be considered a hit, the ADC pulse must be above a certain height. An estimation of the number of photo-electrons is made by multiplying the ADC by a scale factor. The number of photo-electrons (PE) produced from a hit follows a Poisson Distribution.

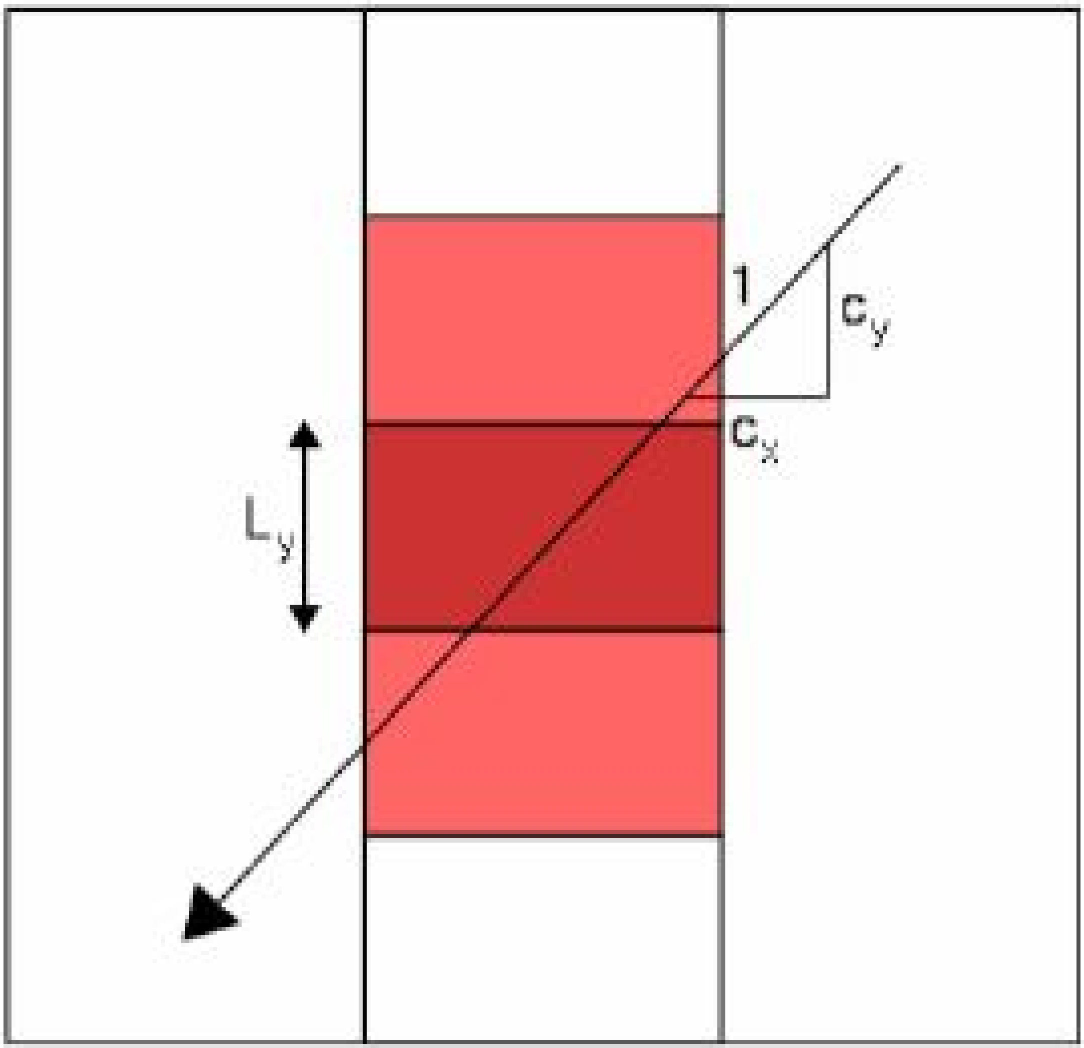


Figure 5.2: A tri-cell used for calibration. The dark red cell is the cell used for calibration. The path length is $\frac{L_y}{c_y}$.

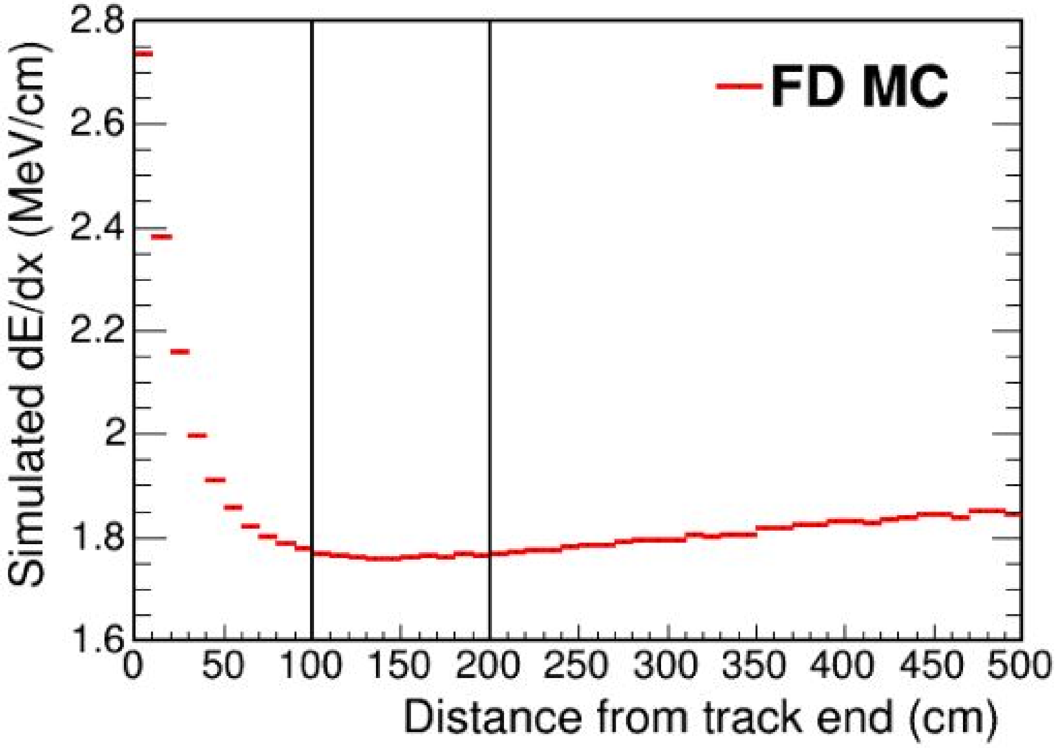


Figure 5.3: $\frac{dE}{dx}$ as a function of distance from the track end. Hits farthest away from the track end have a very non-uniform $\frac{dE}{dx}$ shape.

Therefore, short minimum ionizing hits are highly affected because only the upward fluctuations will produce enough photo-electrons above threshold. This is most commonly seen in the horizontal cells for cosmic ray muons because the steep angle of descent into the detector causes more of the energy deposition to occur in the vertical cells. Horizontal cells with a small W are even more greatly affected.

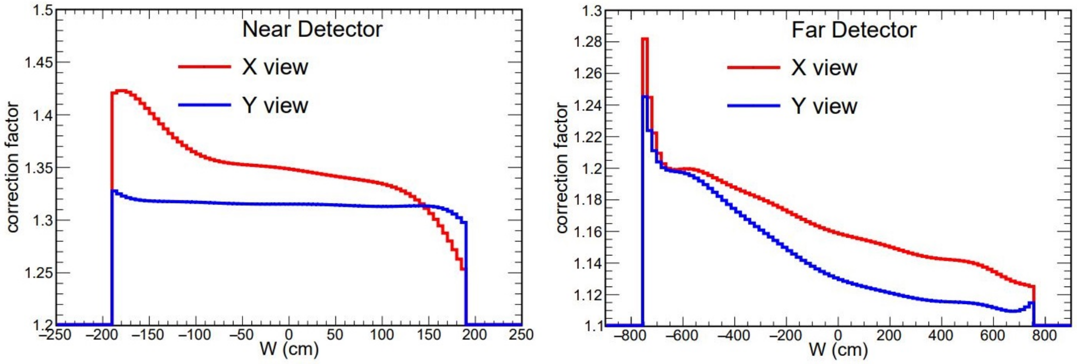


Figure 5.4: Correction factors used for relative attenuation correction in both detectors.

5.1.2 Attenuation Correction

Since the fibers used in the cells are wavelength shifting, the light which is captured is attenuated as it travels to the electronics. Therefore, hits that occur farther away from the electronics will on average a lower PE, and thus be more greatly affected by threshold effects. It is for this reason that a correction needs to be applied to the estimated PE. In order to do this, a two dimensional histogram of the PE over path length is plotted versus W for every channel in the detector. This histogram is then fit to the equation of the form:

$$y(W) = C + A\left(exp\Big(\frac{W}{l}\Big) + exp\Big(-\frac{W+L}{l}\Big)\right) \tag{5.1}$$

In the above equation, C and A are free variables, L is the length of the cell, W is the median value in each W bin, and l is the attenuation length of the fiber. Due to a phenomenon called ”roll-off” which is when light is absorbed in the black plastic at the cell ends, instead of the fiber. This light deficit is well modeled with an fourth order polynomial, so the following equation is multiplied by equation 5.1.

$$y = \begin{cases} 1 - \alpha_R(W - W_R)^4 & W > W_R \\ 1 - \alpha_L(W - W_L)^4 & W < -W_L \end{cases} \tag{5.2}$$

Due to variations in fiber performances, there are usually large residuals from the exponential form. In these circumstances, an interpolation algorithm called Locally Weighted Scatterplot Smoothing (LOWESS) is applied to smooth the cell response [26].

The result of the fit gives a prediction for the PE independent of the distance away from the readout which is proportional to the deposited energy in the cell.

5.1.3 Absolute Energy Calibration

Once the correction for the PE has been applied, it becomes necessary to multiply it by a scale factor to obtain a measurement for the deposited energy in the cell. To do this, cosmic ray muons which stop inside of the detector are used because the distance away from the muon end is well known. Cell hits which happen between 100 and 200 *cm* from the track end are used because they fall in the Bethe-Bloch flat region as seen in figure 5.6.

Using monte-carlo simulation for muons, the desired scale factor is found by taking the ratio of the true $\frac{dE}{dx}$ and the $\frac{PE_{Corr}}{cm}$. With this scale factor, PE_{Corr} from data can be converted into cell energy deposition.

5.2 Hit Clustering with Slicer

The first step in reconstructing the physics of the measured event is to cluster the cell hits in space and time. NOνA uses an algorithm called Slicer attempts to cluster all of the hits which were involved in a particular physics event. The Slicer Algorithm uses the Density-Based Spatial Clustering of Application with Noise (DBSCAN) Algorithm. DBSCAN performs clustering by finding cores, data points with the number of neighboring points greater than a pre-defined minimum and within a certain tolerance of distance in a parameter space. Once the cores are identified, DBSCAN finds weak neighbors, data points within the distance tolerance of a core without neighbors. Data

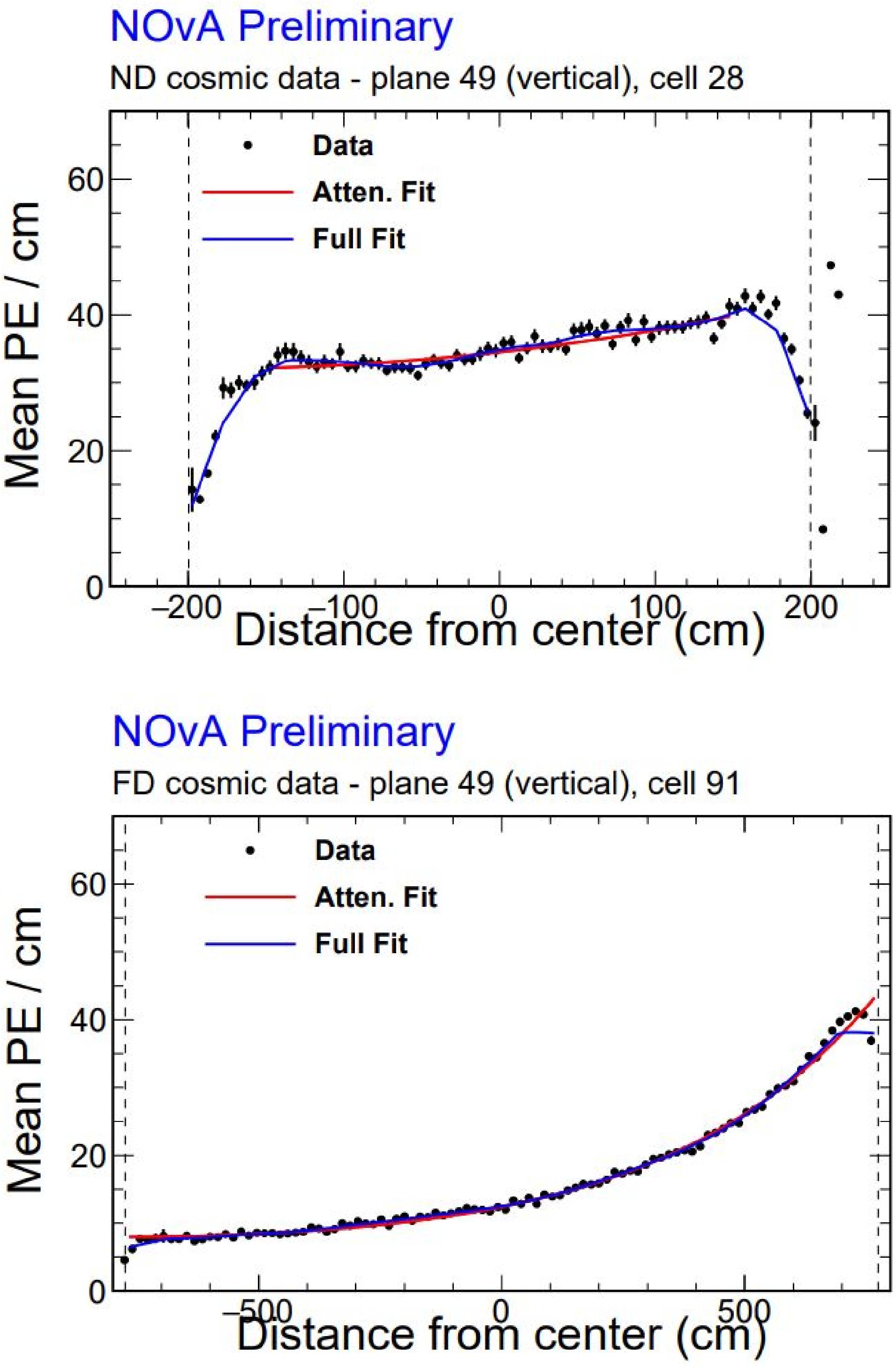


Figure 5.5: ND (top) and FD (bottom) fit for PE over path length versus W. The blue curve represents the fit after the LOWESS algorithm was applied. Notice how the curve better fits the data near the ends.

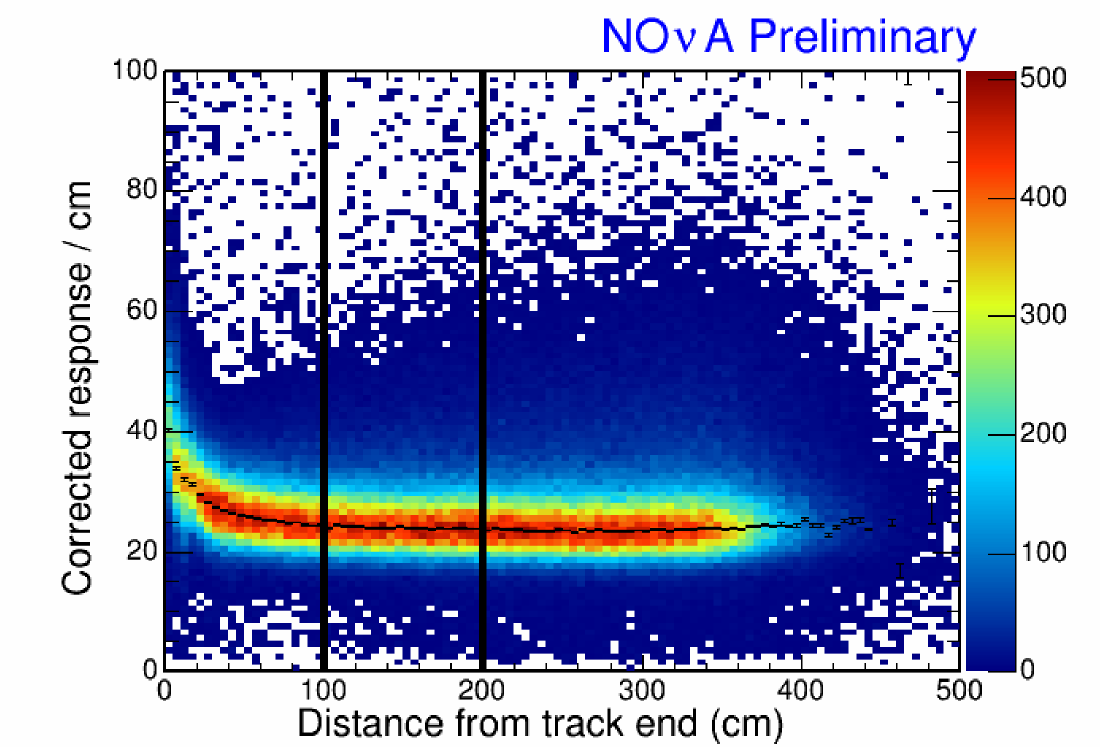


Figure 5.6: Cell response per unit length versus distance from track end. Curve between lines is approximately constant and is called flat region.

points which are neither cores or weak neighbors are not clustered and considered noise. If a cluster does not contain a minimum number of hits, defined by the user, the hits in the cluster are also considered noise [27].

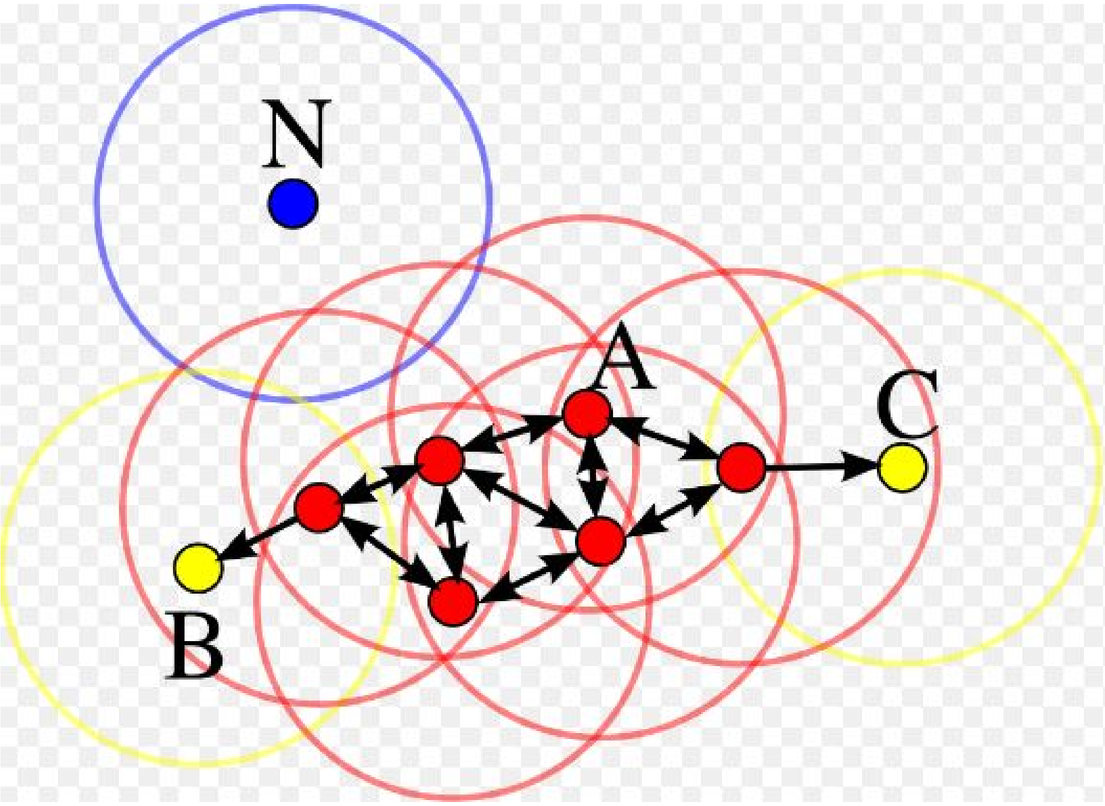


Figure 5.7: Cell response per unit length versus distance from track end. Curve between lines is approximately constant and is called flat region. `https://commons.wikimedia.org/w/index.php?title=File:DBSCAN-Illustration.svg&oldid=219879325`

In NOνA, the data points clustered by DBSCAN are cell hits, and the separation metric between two hits in the same view is given by the following equation:

$$L = \left(\frac{\Delta T - \Delta\vec{r}/c}{T_{res}}\right)^2 + \left(\frac{\Delta Z}{\alpha}\right)^2 + \left(\frac{\Delta X or Y}{\alpha}\right)^2 + \left(\frac{\beta}{PE}\right)^5 \tag{5.3}$$

Here, T_{res} is the timing resolution of the hits added in quadrature, $\Delta\vec{r}$ is the vector distance between the two points in their view, $\Delta X, Y, Z$ are the distances between the two points for the respective component, α and β are free parameters which are separately tuned for each detector, due to the vastly different event rates. PE in this equation is the sum of squares of the two hits in question. Since the electronic noise falls off as $PE^{-2.5}$, the PE term is raised to the fifth power. This term has the role is to suppress noise hits.

The measures of performance of the slicer are quantities called efficiency and purity. Efficiency is the amount of an interaction's energy is reconstructed in the slice. Purity is the fraction of the slice's energy which is deposited by the particles of the interaction. The parameters of slicer are tuned to maximize the efficiency and purity [28] [29].

5.3 The Hough Transform

Once all of the hits in the interaction are clustered, it may be possible to reconstruct the physics of the neutrino iteraction. The first step in doing this is to identify physically significant lines in the slice. In order to do this, a Hough Transform is applied [30]. The Hough Transform is an effective computer vision algorithm which identifies lines, shapes, and curves. For the purposeses of reconstructing particle trajectories, the Hough Transform aims to identify physically significant lines. The algorithm works by iterating through all possible pairs of points in each view of the slice and fitting a line through them. The lines are parameterized by polar coordinates to effectively handle vertical lines. Since a particular slice has an order of 100 hits, the number of potential lines is large. In order to efficiently handle all of the lines, a gaussian smear voting scheme is applied to remove lines which poorly fit the data. The gaussian smear voting score is defined as follows:

$$\begin{aligned} vote &= e^{-\frac{(\rho-\rho_0)^2}{2\sigma_\rho^2}} e^{-\frac{(\theta-\theta_0)^2}{2\sigma_\theta^2}} \\ \sigma_\rho &= \frac{3}{\sqrt{12}} \\ \sigma_\theta &= \frac{3}{d\sqrt{6}} \end{aligned} \tag{5.4}$$

Where d is the distance between the two points, and σ is the width of the standard top-hat distribution. When plotted in the Hough Space, pairs of hits which form a well fitting line produce peaks in Hough Space. Peaks which fall below the threshold value of the average bin height in Hough Space are removed. The parameteres of the most prominent peak are weighted by the vote and averaged. The cells from such lines are then removed, so the process can be repeated. This iterative process prevents the leading lines from interfering with subleading ones. This process is repeated until all lines above threshold are formed [30].

5.4 The Elastic Arms Vertexing

In order to reconstruct a neutrino interaction, the origin of the interaction must be located. It is from this interaction location where the outgoing particle trajectories originate. In order to do this, NOνA employs the Elastic Arms Vertexing Algorithm, also known as the "Method of Deformable Templates." Elastic arms considers the lines in the given picture and calculates a prediction for where they converge. In NOνA, the Hough lines are used as an initial guess, and the lines are tuned to yield the best possible event vertex prediction. With a vertex identified, the tuned Hough lines from the event are taken as "arms", which are vectors pointing away from the vertex.

5.5 The Fuzzy-K Vertex Algorithm

The elastic arms algorithm is a highly performing algorithm for finding the physical lines in a particular slice, but it under performs with clustering cell hits belonging to such lines. This issue is mitigated by supplementing the work done by elastic arms with Fuzzy-k Means Clustering Algorithm. Fuzzy-k means clustering is a variant of the well known K-means Clustering algorithm which is an unsupervised learning technique with the objective of finding the membership of a particular data point to a collection of data points in a particular space. Using Fuzzy-k clustering with elastic arms vertexing, NOνA is able to reconstruct the particle trajectories of a particular physics event.

The original k means clustering algorithm has the objective of placing n data points into k sets, where $k \leq n$, to minimize the sum of squares distance of each point to a cluster centroid. This objective is formulated as follows:

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \mu_{\mathbf{i}}\|^2 \tag{5.5}$$

In this algorithm, a predefined number of centroids is placed randomly in the space, and their coordinates are shifted in order to minimize the objective function. One of the shortcomings of k-means clustering for NOνA reconstruction purposes is handling datapoints which can appropriately have membership in multiple clusters. This is a common occurrence in particle trajectories with large energy contamination from other particles. NOνA handles this shortcoming by applying this algorithm with soft clustering, also known as fuzzy clustering. Fuzzy-k means clustering is nearly identical to k-means clustering, but the objective function is modified as follows:

$$\arg\min_{S} \sum_{i=1}^{n} \sum_{j=1}^{c} \|\mathbf{x} - \mu_{\mathbf{i}}\|^2 \tag{5.6}$$

Therefore, hits are allowed to have membership in multiple clusters. A useful extension of this algorithm is to use a probabilistic approach to cluster membership along with another extension which allows for an unknown number of clusters to be found [31][32]. By doing this, hits which have a probability under a particular threshold will be excluded from the clusters. This is especially useful for NOνA because noise hits far away from the particle trajectories are not clustered.

NOνA uses the hits from the XZ and YZ views of the detector for 2D clustering. The reconstructed vertex from the elastic arms algorithm is determined to be the origin, and the angular separation from the Z axis is used as the separation criteria. Therefore, the problem becomes two 1D problems.

5.6 Hit Reclustering

Though the Fuzzy-k clustering algorithm performs reasonably well for reconstructing particle trajectories, there is room for improvement for reconstructing electron trajectories. The reason for this is that Fuzzy-k was not exploited to handle the conical shape of the electrons during Bremmstralung Radiation. On average, electrons are minimum ionizing for the first 8 planes, so there is little reclustering need. However, electrons begin to radiate after an average of 8 planes. Therefore, cell hits which fall within 20 cells transverse to the propagation axis are reclustered.

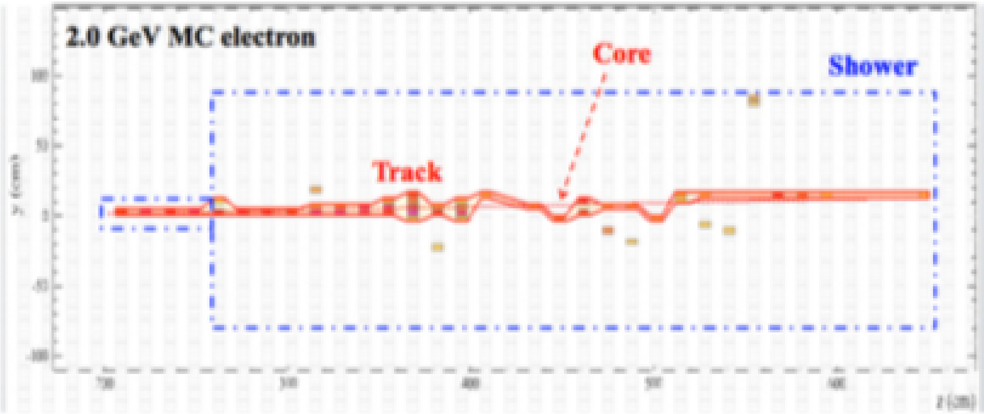


Figure 5.8: The Bremstrahlung Radiation of electrons leads to cell hits which fall outside of the Fuzzy-K clustered prong. Creating a reclutering region increases electron reconstruction efficiency with marginal purity cost.

5.7 Cell Energy Deconvolution

In neutrino interactions, there are instances in which daughter particle trajectories overlap in one or both of the views. Since Fuzzy-k clustering allows for membership of hits in multiple prongs, there exists cells in which several particles have deposited energy. As a result of this, a deconvolution algorithm is applied to the hits in overlapping showers, which predicts the percentage of a particular cell's energy corresponding to one of the showers in the overlap. The equation is given as follows:

$$E^i_{cell} = \frac{PE^{cell}}{a_i}\frac{(PE_i^{shower}P_i/a_i)}{\sum_{j=1}^{ncells} PE_j^{shower}P_j/a_j} \tag{5.7}$$

Where PE is the number of photo-electrons, P_i is an exponentially decaying function with respect to the distance of the ii^{th} cell to the shower core, and a_i is an attenuation factor to convert PE into E [33].

Chapter 6

Machine Learning Techniques for Particle Classification

In particle classification, there are many variables which can be exploited. Elementary classification techniques such as constant cuts and logistic regression struggle as the number of training variables increases. Not only that, but these algorithms also struggle when there exist nonlinear correlations between the variables. It is for this reason that the use of machine learning techniques has grown in scientific experiments for classification analyses. Several of the algorithms and their applications in the NOνA will be discussed in this chapter.

6.1 Artificial Neural Networks

One of the most common learning algorithms used in machine learning is the artificial neural network (ANN). ANN's were inspired by the structure of the brain and its configuration of neurons. The most common type of ANN is the feed-forward multi-layer percepetron (MLP).

In a MLP, all of the neurons in a particular layer connect to all of the neurons in the following layer. Each neuron in a particular layer contains an activation function which performs a nonlinear calculation. The combination of all of these activations allow for the modeling of nontrivial data trends. For binary classification tasks, the output of the

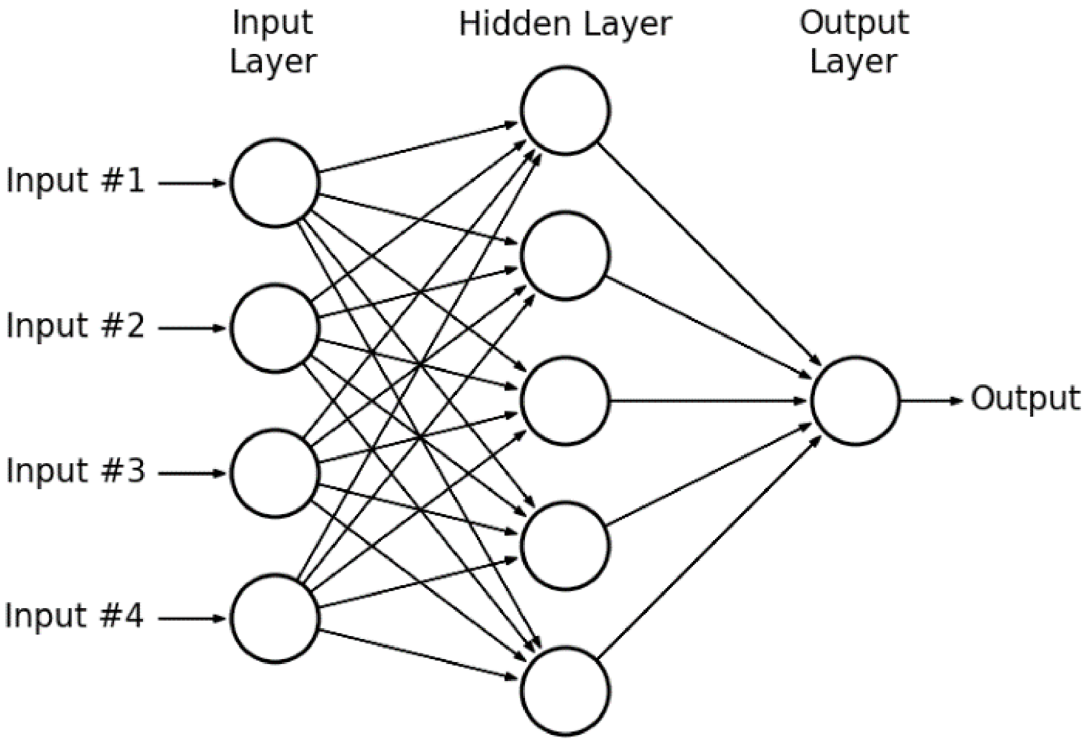


Figure 6.1: MLPs model nonlinear behavior by passing inputs through several layers of neurons with simple nonlinear activations.

ANN is the sigmoid function, given by:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{6.1}$$

This is an ideal final output because the domain spans the entire real line, and the range is between 0 and 1. Therefore, the output of the model give the probability of an unknown state to be binary type 1. ANN's are trained in an iterative manner by performing a gradient descent on a particular cost function. For classification tasks, the binary cross-entropy function is the cost function to be minimized. It is given by:

$$J(h) = -\frac{1}{m} \sum_{i}^{m} \Big(y_i ln(h_i) + (1 - y_i) ln(1 - h_i) \Big) \tag{6.2}$$

where m is the number of samples, h is the predicted value, and y is the ground truth value. The cross-entropy function is ideal for classification because the range spans the positive real line, and the domain spans 0 to 1. As a result of this, sufficiently large gradients exist near the region of interest for classification.

In gradient descent, the parameters of the model are tuned to yield a globally minimal cost function. Since, the gradient vector points in the direction of greatest rate of change, a step along the gradient vector is taken given as follows:

$$J := J - \alpha \partial_\theta J \tag{6.3}$$

where α is the learning rate, and ∂_θ is the gradient with respect to the parameter θ. This computation requires knowledge about the gradient of the prediction as a function of the parameters. This is calculated by means of back propagation. Back propagation is a numerical method of approximating the derivative of the prediction with respect to the parameters. As its name implies, back propagation feeds the difference of the prediction and the ground truth into the network in reverse order. This process leads to a shifting of the parameters which minimizes said difference.

6.2 Recurrent Neural Networks

There are many applications not well suited for MLPs. One important example is when the number of inputs are unknown. In order to effectively model such a scenario, the learning algorithm needs to be able to adapt to the varying number of inputs. Recurrent Neural Networks (RNN) are well suited to handle this scenario. RNNs handle these sequences of data by having a constant feedforward architecture with weights varying by the position in the sequence. Not only that, RNN modules have information from previous models, and in some applications, later modules, fed in, This can be seen in Figure 6.2.

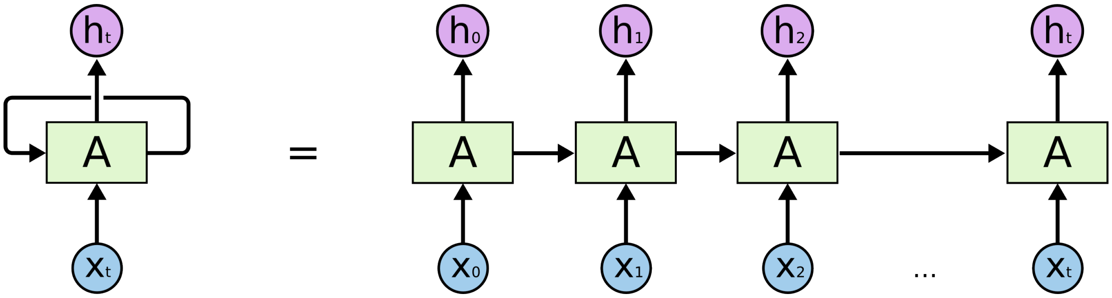


Figure 6.2: RNNs can be viewed as temporally connected feed forward networks.

RNNs are concatenated sequentially separated ANNs. By using temporally separated parameters, RNNs learn dynamic feature correlation. It is important to note that time is a term used loosely. In the context of RNNs, time is the domain on which the model samples sequentially.

6.2.1 Long Short-Term Memory Networks

Classical RNNs are highly performing models for sequential modeling tasks. However, when the sequences are long, traditional RNNs experience the issue of vanishing gradients. This is the phenomenon in which small gradient values decrease in magnitude after passing through each layer in back propagation. The Long Short-Term Memory Network (LSTM) is a recurrent architecture which mitigates this issue.

LSTMs employ two new neurons in each module with the task of prioritizing how important it is to remember or forget a particular sample of the sequence. These neurons

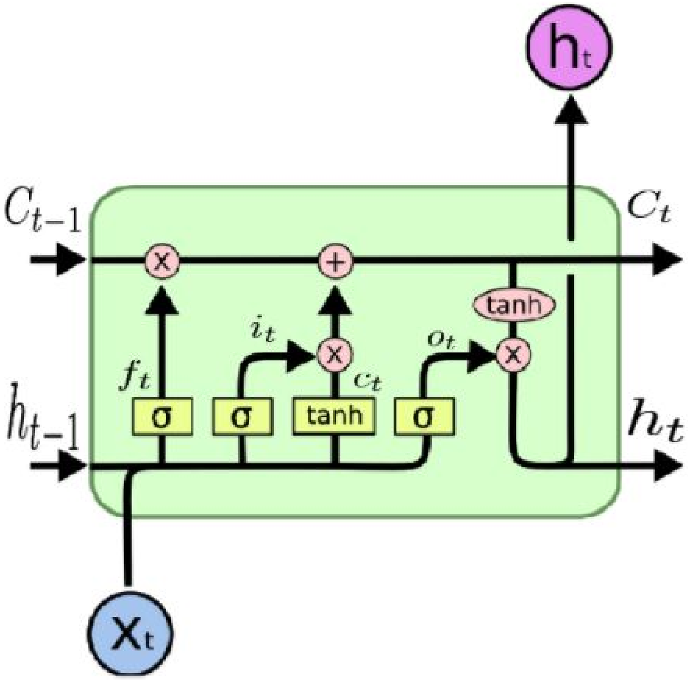


Figure 6.3: Single cell LSTM architecture

are known as the ”remember” and ”forget” gates, and they are trained along with all of the model’s other parameters by means of back propagation [34][35]. By employing this method, LSTMs give higher priority to the instances of a sequence in which there is large potential for learning. This is the machine learning architecture which was used to train the $\nu_e CC$ classifier in this analysis.

6.3 Decision Trees

One of the most popular machine learning models in science is the decision tree model. Decision tree classifiers and regressors both function by considering discriminating performance of a feature at a particular node and forming more decision branches from such node. Conditions which yield the highest separation power are used as the selection criteria at the nodes. This is illustrated very well in Figure 6.4.

6.3.1 Boosted Decision Trees

Quite often, it is found that standard decision trees are not robust against statistical fluctuations in the training dataset. As the tree depth increases, so does the risk of overfitting to said fluctuations. Simply using shallower trees is not a viable solution, for the discrimination power of a decision tree scales with the depth. An alternative solution is to use an ensemble of shallow trees. There are two common implementations of this: boosting and bagging. It was found that using adaptive boosted decision trees yielded the highest separation significance for identifying signal $\nu_e CC$ events over background.

Adaptive boosting is the procedure in which a shallow tree is trained on the training dataset. Since shallow trees are weak learners, the single tree models the data for a fraction of the parameter space. The adaptive boosting procedure then randomly selects several samples which fail the selection of the first tree and use them to train the second tree. This procedure happens until the entire parameter space is modeled. In its use, a boosted decision tree (BDT) uses a probability weighted superposition of the elements of the tree ensemble to model the event in question [37]. The implementation of BDTs in this analysis will be explained later in this thesis.

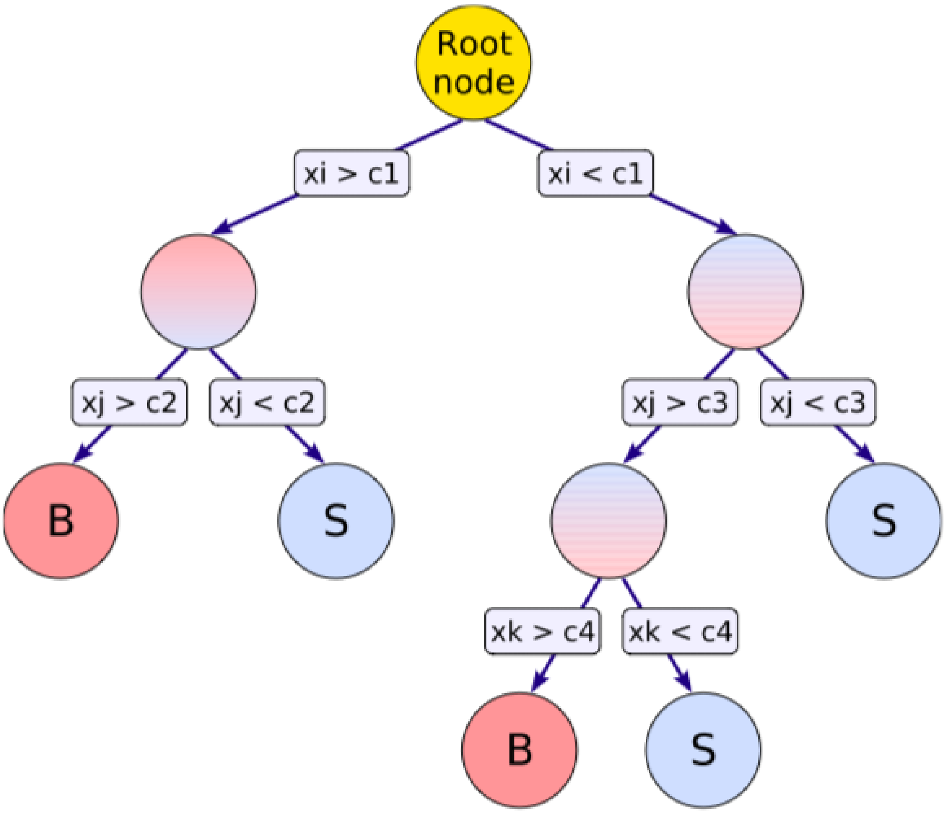


Figure 6.4: Decision trees define selection criteria for particular features at each node and produce more decision branches from the given node. Here, x represents a discriminating variable, S represents signal events, and B represents background events [36]

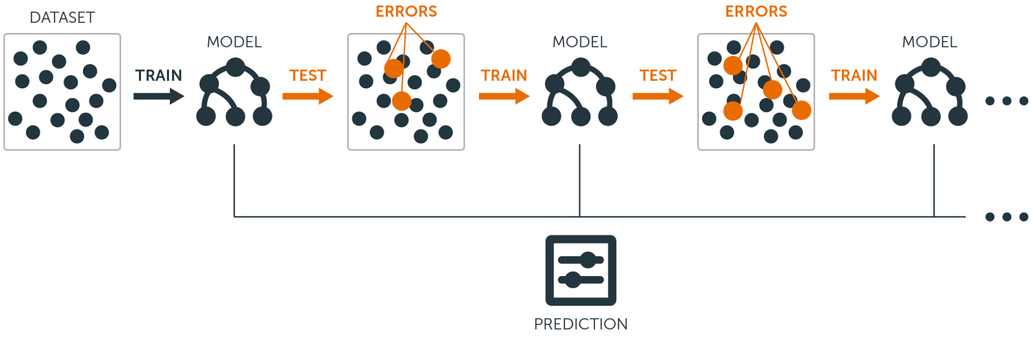


Figure 6.5: The BDT ensemble is formed by successively training weak learners on the events which fail selection by the previous learner.

Chapter 7

ν_e CC Identification Models

The NOνA reconstruction chain provides useful estimations of the kinematics and topologies of a particular physics event. Features extracted from the reconstructed information can be exploited to discriminate between the signal and background events. The primary $\nu_e CC$ identifier in NOνA neglects this information, and instead, uses the low level information from the cell hits. By using higher level features from reconstruction, it will be shown that a shallower model is capable of achieving similar results with far less training time and data. This classification algorithm goes by the uninspired name "SliceLID", and it capitalizes on the power of the LSTM architecture. In addition to the development of SliceLID, an empirical cutting scheme in a multivariate particle identification (PID) space will be explained.

7.1 Shower Based Classification

SliceLID is a composite model in which shower level features are fed into a shallow LSTM, and the output of the LSTM is concatenated with the calorimetric energy of the slice. Finally, the information is passed on to the output layer which consists of a single neuron with a sigmoid activation. Features which best describe the kinematics and topology of a particle in a reconstructed shower were chosen to maximally exploit the intra-slice, energy dependent, nonlinear correlations.

7.1.1 Log Likelihood Based Identification

One of the ways to distinguish particles in detectors is to observe the energy deposited per unit length ($\frac{dE}{dx}$). Due to mass differences of charged particles traversing the detector, the behavior of $\frac{dE}{dx}$ of the particles differ. This behavior is well simulated in the GEANT4 simulation framework. As a result of this and the large number of simulated neutrino interactions, meaningful information about the particle trajectories can be made. The reconstruction information from the simulations were used to model the energy deposition behavior for electrons, muons, protons, charged pions, neutral pions, and photons in both longitudinal and transverse directions. Histograms of $\frac{dE}{dx}$ were made for each particle hypothesis and binned by total calorimetric energy and position along the reconstructed prong. To form probability distribution functions, these histograms were area normalized.

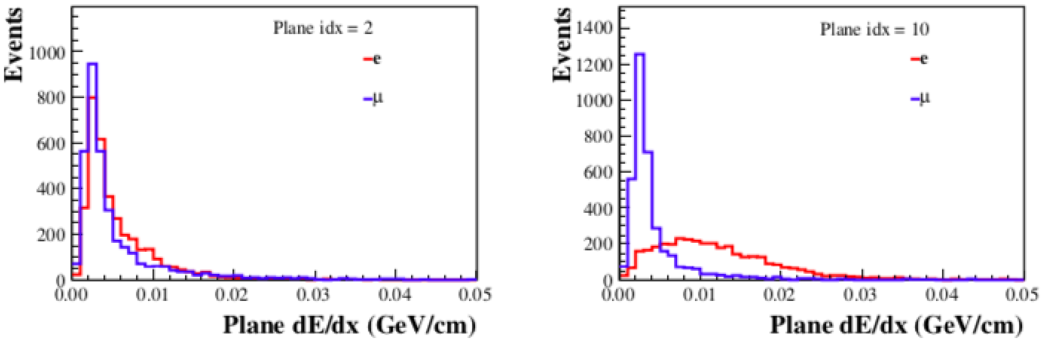


Figure 7.1: The mass differences between different charged particles leads to unique longitudinal and transverse $\frac{dE}{dx}$ behavior along the axis of propagation.

These distributions give the likelihood of an unknown particle being a hypothesis particle at a particular location along its propagation. It is possible to take the product of all of the likelihoods is taken over all of the planes to find the net likelihood of being a particle type. Doing this, however, does not yield much information since the true particle hypothesis tends toward 1, and all other hypotheses lie near 0. Therefore, the logarithm of the likelihood is taken at each plane and summed to yield a net log likelihood of being a hypothesis particle. Working on the logarithmic scale allows for much greater distinguishing power of the particle hypotheses. The differences of the log likelihoods for the electron hypothesis and all other hypotheses are used as inputs into the learning algorithm.

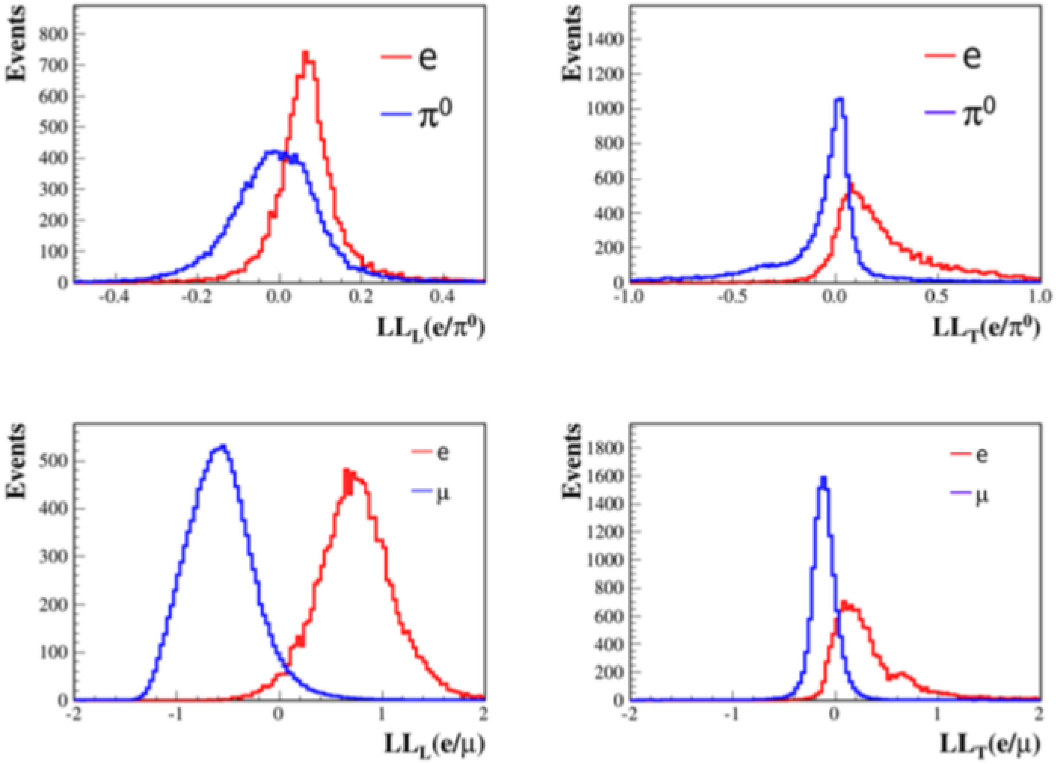


Figure 7.2: Differences in the log likelihood shapes for different charged particles provides strong discriminating power.

7.1.2 LID

The original NOνA $\nu_e CC$ identifier used log likelihood based $\nu_e CC$ based identification (LID). LID considered the log likelihood differences for the leading energy reconstructed shower. In addition to these log likelihoods, LID used slice level information to model the topology of the reconstructed neutrino interaction. These variables include the average $\frac{dE}{dx}$ of the leading shower, the reconstructed π^0 mass from the two leading energy prongs, the energy clustered in the vicinity of the vertex, and the angle of the leading prong with respect to the beam axis. LID was a high performing model for events with a low hadronic energy fraction, but its performance degraded when the hadronic energy fraction increased. In order to yield a gain in selection performance, it is necessary to consider information from the subleading prongs.

7.1.3 SliceLID

LID laid the groundwork for physics based $\nu_e CC$ identification. There is still much information to be gained from considering the hadronic activity of the interaction. This was achieved by employing a model which could handle a variable number of prongs in a robust manner. The LSTM was chosen to satisfy this requirement, and the additional physical features were reconsidered for more effective signal selection. This model was given the uninspired name SliceLID, and it is the selector used for this analysis.

7.1.4 SliceLID Architecture

After it was decided that a RNN would be used to handle the variable number of prongs for each event, the next task was to choose an optimal recurrent architecture. The LSTM was chosen to mitigate the effects of the vanishing gradient problem. Though the number of prongs for a particular neutrino interaction are small, there are many parameters in the models for each prong. Therefore, the problem of vanishing gradients was still a problem to handle.

In the SliceLID scheme, each reconstructed prong is used as an input for a particular model time step. A limit of 7 prongs was applied to eliminate obvious backgrounds. The inputs for each prong is as follows: longitudinal and transverse log likelihood differences of each particle listed and the electron hypothesis, the outputs of the convolutional neural network particle identifier for each particle hypothesis, the calorimetric energy of the prong, the length of the prong, the energy asymmetry between views for each prong, and the cosine angle of the prong with respect to the z axis. In addition to these inputs for each prong, the slice energy is used as an input. The vector of each prong, sorted by energy, was fed into the LSTM for each time step, and each were given a layer of 16 neurons with the ReLU activation. These time separated models were connected, and the output of the final model was the traditional sigmoid output.

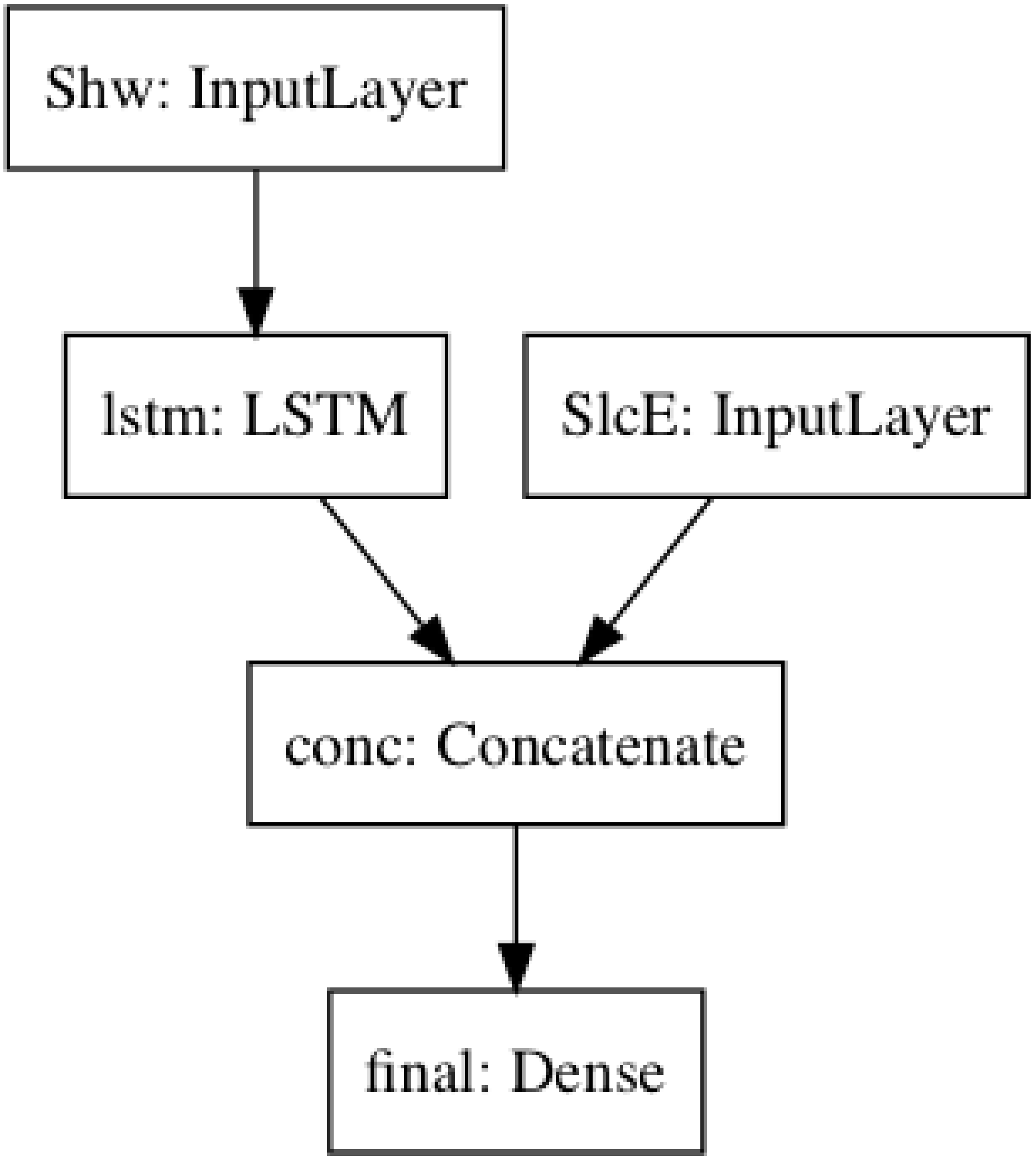


Figure 7.3: The SliceLID architecture feeds the shower level information into an LSTM, and the output of the LSTM is concatenated with the slice calorimetric energy. This concatenated vector is fed into a single neuron with a sigmoid activation.

7.1.5 SliceLID Preselection

There are many background events in the training files which are easy to distinguish from the signal. It has been shown that a machine learning model's classification performance is sub-optimal when the training dataset is saturated with easy background. For this reason, a series of loose preselection cuts were made to eliminate the obvious background and retain nearly all of the signal. The cuts were as follows:

- Slice calorimetric energy < 6 GeV
- $40 <$ # Hits in the slice < 160
- Longest prong in the slice < 700 cm

With the simple background gone, the machine learning model was able to place larger effort on the difficult to distinguish cases.

7.1.6 SliceLID Training

The shallow architecture of SliceLID was chosen in order to optimize selection performance and training time. It was found that deeper architectures yielded marginal gain with significant increase in training time. With such a small architecture, the model was trained with a MacBook Air. The duration of model training on a simple cpu was on the order of hours. This is drastically shorter and less resource intensive than the training requirements for a deep convolutional neural network, like the one being used by the collaboration. In order to mot effectively utilize the computing resources, the model was trained categorically by the fraction of the leading prong calorimetric energy and the slice calorimetric energy. A requirement on the model validation loss improvement was imposed for each epoch, so separating the model allows for abbreviated training for the kinematic regions with more discriminating power. Plots for the final training epochs of the FHC models are as follows:

7.2 Event Based Classification

SliceLID is a powerful particle classification tool with a high accuracy. Accuracy, however, is not the performance metric needed for this analysis. Instead, the figure of

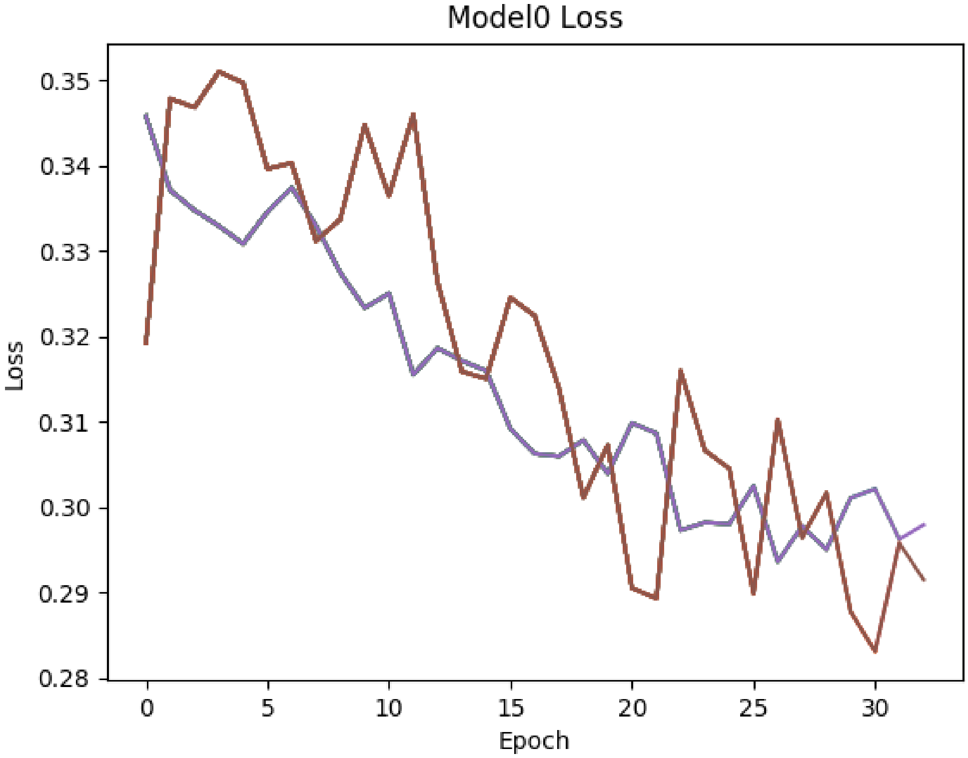


Figure 7.4: Cost function for SliceLID with energy fraction < 0.25

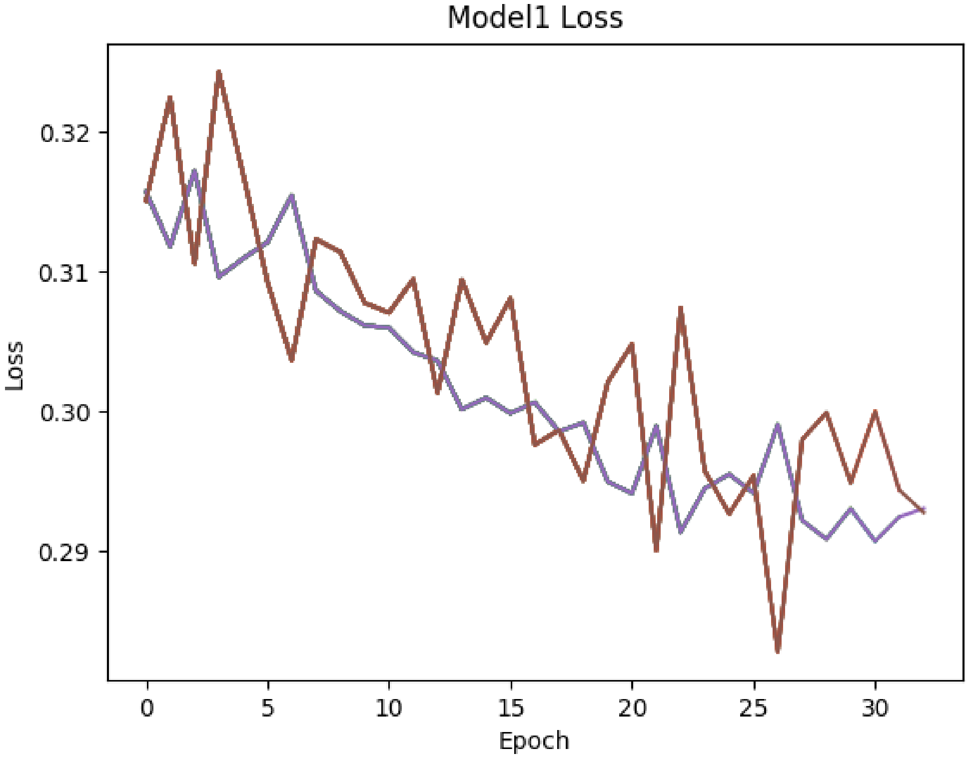


Figure 7.5: Cost function for SliceLID with 0.25 < energy fraction < 0.5

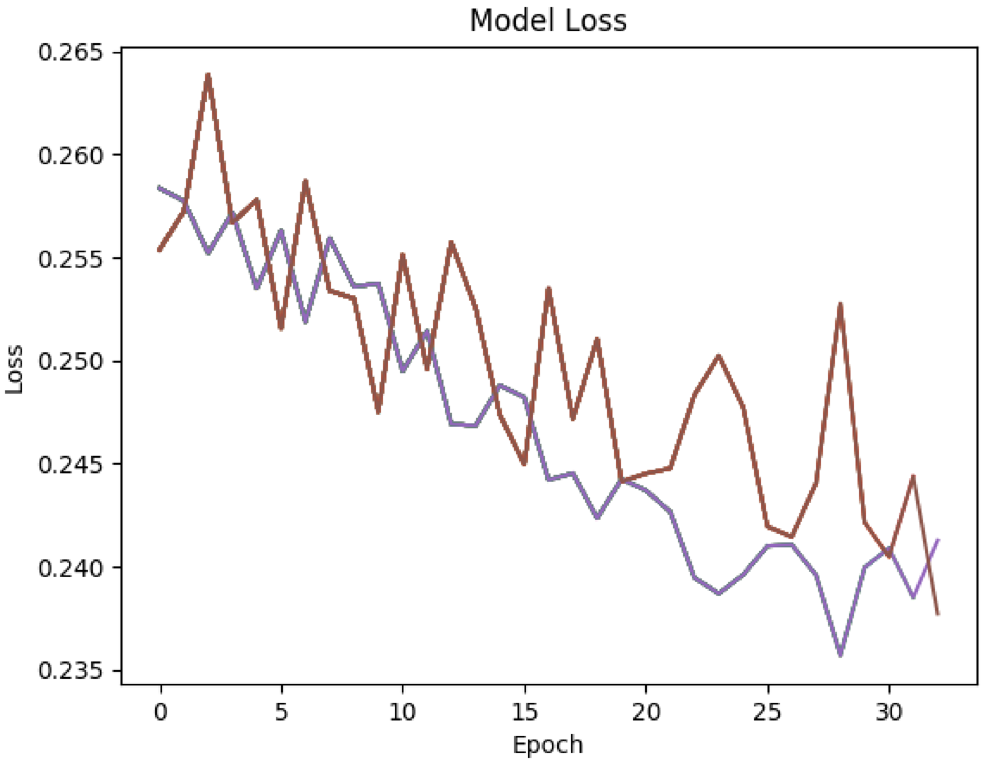


Figure 7.6: Cost function for SliceLID with 0.5 < energy fraction < 0.75

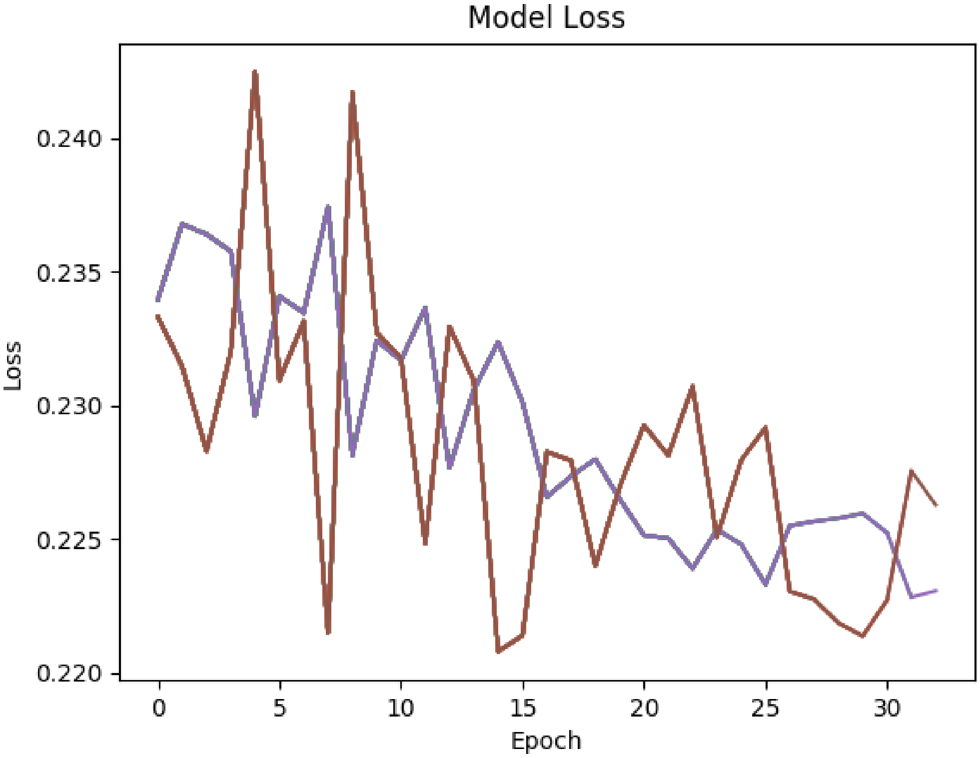


Figure 7.7: Cost function for SliceLID with energy fraction > 0.75

merit (FOM) is the optimization objective for the classification. FOM is a measure of statistical significance of separation, and it will be discussed in more detail in the next chapter. In order to have the ability to optimize for FOM, two additional classifiers are trained and used in combination with SliceLID. The two additional classifiers are BDTs with slice level inputs:

- SliceLID Score
- Reconstructed ν_e energy
- Reconstructed Hadronic energy
- Longest prong length
- Number of hits
- Cosine angle between the leading shower and the beam axis
- Ratio of transverse Momentum to total momentum
- x and y components of the momentum
- Hit sparseness asymmetry between the two views
- Coordinates of the reconstructed vertex
- Shortest distance to the top of the detector
- Shortest distance to an edge of the detector which is not the top

The first BDT serves as a tool for distinguishing between the ν_e CC signal and beam backgrounds. Whereas, the other BDT has the purpose of separating the signal from cosmic background. Most of the features of these two BDTs were chosen simply because many of them are used as rectangular preselection cuts in the official ν_e analysis. The use of BDTs over rectangular cuts has many advantages. Some of them include [36]

- Efficiently handle many input features
- Carve away high purity regions of the feature space

- Effectively handles weakly discriminating variables
- Robust against non-trivial correlations between input variables
- No data preparation needed before training
- Handle two BDT scores, instead of individually handling the input features

It is for this reason two BDTs were chosen over the rectangular cut method.

Chapter 8

Event Selection

Throughout the entire data collection in NOνA, there will be an incredible number of background events for every signal event in the far detector. Processing this many terabytes of data in a timely manner is not feasible. Not only that, but poorly reconstructed events can greatly degrade selection and energy estimation performance. To NOνA's benefit, many of these events are simple to identify and reject with little computational cost. The following cut scheme was applied to remove these background events: cosmic veto, data quality cuts, and a series of empirical cuts in a multidimensional strong discriminator space to maximize the figure of merit (FOM) for signal identification over background. In this chapter, each of the elements of the cutting chain will be discussed in further detail.

8.1 Quality Cuts

In order to ensure that the data being analyzed is of considerable quality for the analysis, several data quality cuts are applied before applying any analysis specific cuts. These include constraints on: number of active di-blocks, live time, hit rate, tracking fraction, slice number, and time stamp. Explanations for the cuts are as follows:

- Number of active di-blocks ≥ 4: Ensure the fiducial volume for the particular run is large enough to obtain large enough event rate and neutrino energy efficiency
- Live time: Live time must be large enough to allow for stable data taking for a

required continuous duration

- Hit rate: Ensures the data taking and noise levels fall within their expected values for the particular trigger

- Tracking fraction: Large number of 3D tracks gives verification of the stability of the time synchronization of the DCMs

- Slicing: large number of slices show sufficient readouts and stable timing resolution.

- Time stamp: Selects hits which fall in the beam spill trigger window

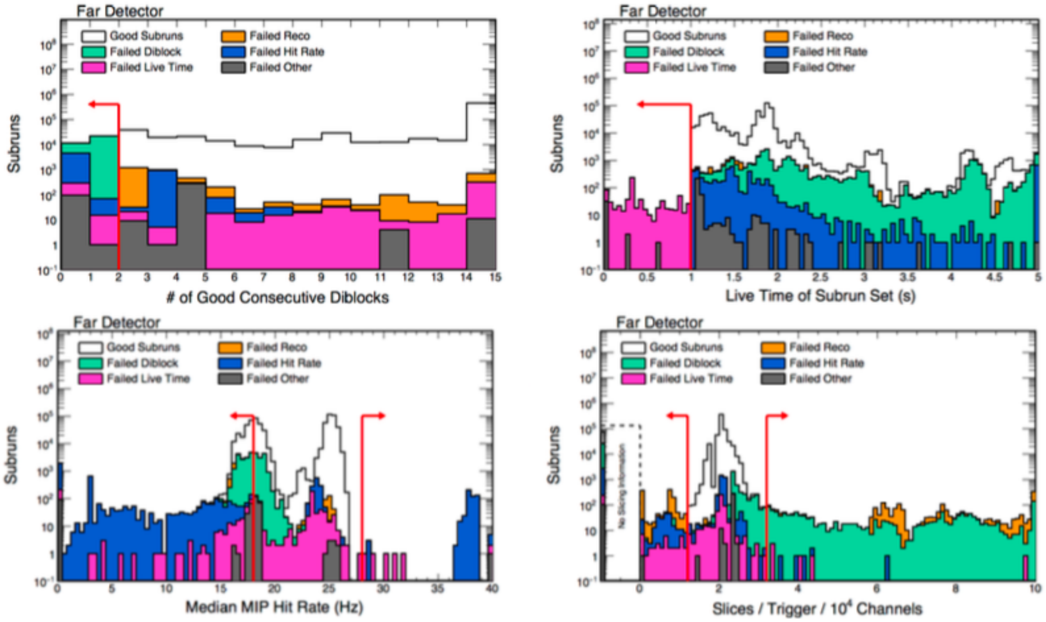


Figure 8.1: Events which failed the data quality selection have color given in the legend. The red arrows point in the direction of cut failure.

8.1.1 Cosmic Veto

The FD is located on the surface of the earth, which means that there is a large cosmic ray muon rate in the detector. Many of these are easy to reject, due to their topology. Therefore, a cosmic veto is applied to remove the obvious cosmic ray muon tracks: tracks which propagate perpendicular to the FD z-axis and tracks which completely pass through the detector.

8.1.2 Event Quality

When a charged particle deposits a lot of energy into a particular channel, there is an occasional effect in which a channel in the APD saturates and contaminates neighboring channels. This leads to a plane with several contiguous cell hits. This phenomenon can greatly hinder the reconstruction quality of the event. Therefore, any slice with a plane which has 8 or more contiguous cell hits is rejected.

8.1.3 Official ν_e CC Signal Selection

In the standard ν_e analysis, there exists rectangular preselection cuts in parameter space and a deep learning ν_e CC classifier cut. The rectangular cuts serve as a preselection and cosmic rejection to the main particle identifier (PID) known as the Convolutional Visual Network (CVN). The cutting scheme can be seen in the following figure and table [38]:

Cut	ν_e CC	$\bar{\nu}_e$	beam ν_e	NC	ν_μ CC	ν_τ CC	Cosmics
No Cuts	55.44	1.11	15.87	424.95	242.50	2.98	3.26×10^6
Veto	54.3	1.09	15.37	323.74	226.54	2.86	3.26×10^6
Analysis Mask	54.3	1.09	15.37	323.74	226.54	2.86	3.26×10^6
Event Quality	53.50	1.08	15.12	304.79	222.14	2.79	2.88×10^6
Containment	40.92	0.83	10.17	235.78	106.55	1.75	94172
Backward Photon	40.59	0.83	10.09	230.85	103.52	1.73	90340.2
p_T/p	40.02	0.82	9.98	215.97	101.49	1.69	71734.5
Preselection	38.28	0.71	7.47	108.43	43.75	1.07	15734.7
CVN	31.09	0.58	5.74	5.08	0.93	0.31	2.34
Nearest Slice	30.98	0.58	5.74	5.05	0.92	0.31	2.02

Table 8.1: Official ν_e Analysis Selection Flow

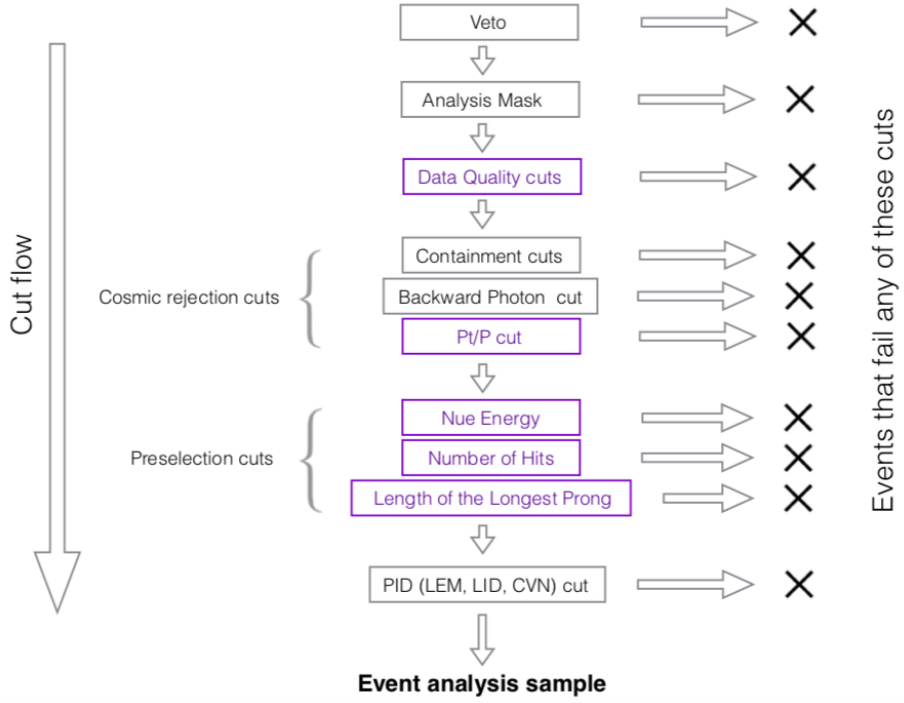


Figure 8.2: Survival Numbers for a given ν_e Cut

A direct comparison was made between CVN and SliceLID with the official preselection/cosmic rejection criteria. Since SliceLID had not been trained with cosmic ray data, its cosmic background rejection is weak. Therefore, cosmic ray background was ignored for the comparison. The results are as follows:

PID	Signal	Background	FOM
CVN	28.92	11.98	4.52
SliceLID	27.89	11.79	4.43

Table 8.2: SliceLID/CVN Comparison with Official Preselection

As seen from the table, with the standard ν_e CC preselection and cosmic rejection cuts, SliceLID's separation power falls just short of CVN's. It is important to note, that the cutting scheme used by the official analysis was tuned to optimize for CVN. It is for this reason that a new cut scheme is performed for this analysis.

8.2 Particle Identification Selection

Once all of the easy backgrounds and reconstruction failures have been removed, the more cumbersome task of selecting oscillated ν_e signal from the various backgrounds comes. The event rates for signal and the various backgrounds differ, and it is the goal of this analysis to achieve the highest sensitivity as possible for oscillation parameter measurement. Therefore, the statistical significance of Poisson distributed signal over background is chosen as the optimization objective, also known as the figure of merit (FOM):

$$FOM = \frac{Sig}{\sqrt{Sig + Bkg}} \tag{8.1}$$

It is for this reason that machine learning algorithms are trained and used as particle identifiers (PID). The implementation of signal selection with said algorithms will be explained in this section.

8.2.1 The SliceLID ν_e CC Identifier

SliceLID is chosen as one of the PIDs to be used because of its ability to use the outputs of reconstruction chain and perform high accuracy signal classification. There exists high discriminating power in the shower level information, and it's amplified with the addition of the slice energy. On its own, however, SliceLID is not optimal for classification because it was optimized to minimize the cross-entropy loss of a training set which had not been oscillated or POT normalized. Therefore, additional PIDs are used to enhance selection sensitivity.

8.2.2 Selection BDTs

As mentioned in the previous section, the standard ν_e appearance analysis cuts were chosen to yield the highest possible CVN performance. This analysis, however, uses many of the same features for preselection and cosmic rejection, but instead of employing rectangular cuts, the discriminating power of BDTs is used. As a result of this, there are only three independent parameters for performing the selection: the SliceLID score, the preselection BDT score, and the cosmic rejection BDT score. The optimal values for these are found by performing rectangular cuts in this lower dimensional PID space. Rectangular cuts are appropriate to use here because the features are low dimensional, have low intra-correlation, and are highly discriminating. After performing this procedure, the following selection cuts were determined:

- SliceLID > 0.73
- Beam BDT > 0.52
- Cosmic Rejection BDT > 0.56

For ease of viewing, the PID plots are shown in 2D instead of 3D to demonstrate the discriminating power and low intra-correlation.

Table 8.3 shows the content of the data through the selection process:

Using equation 8.1, the ratio of the FOM from the official analysis to the FOM from this analysis is 0.998.

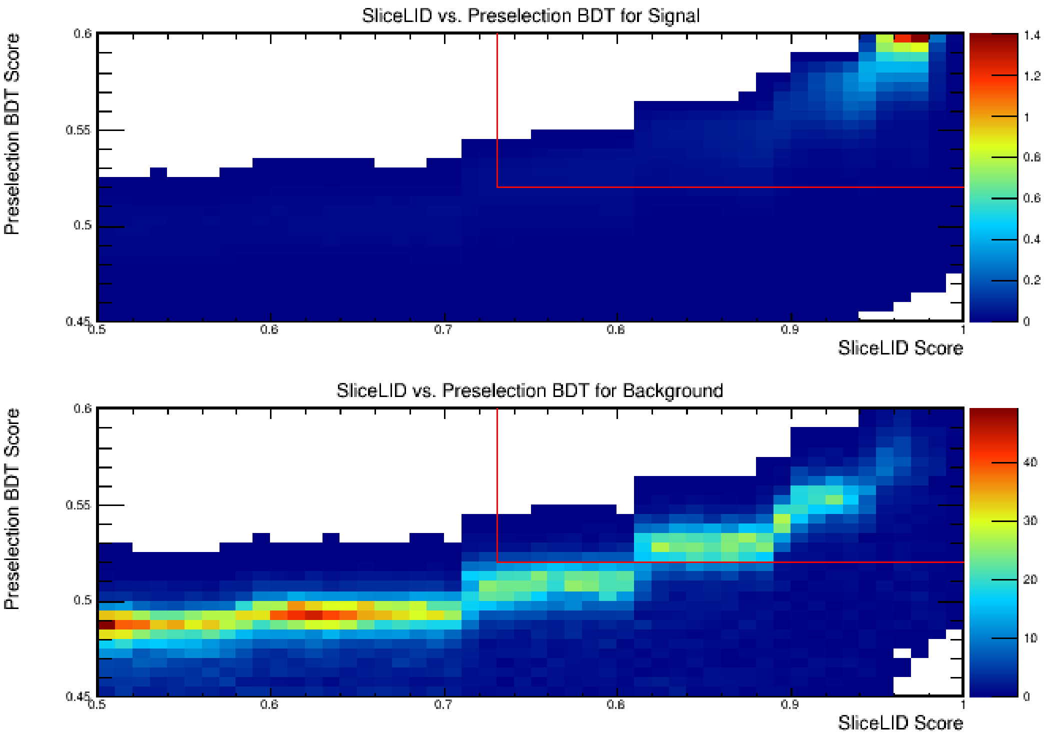
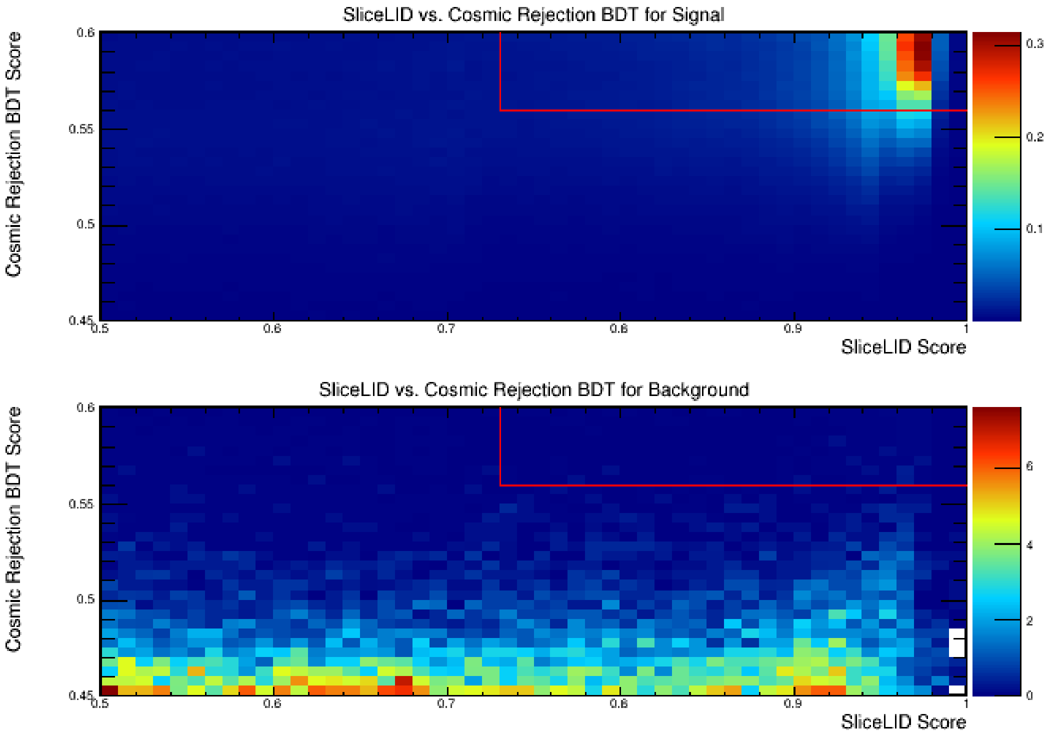


Figure 8.3: Preselection BDT/SliceLID Colormap

Cut	ν_e CC	beam ν_e	NC	ν_μ CC	ν_τ CC	Cosmics
Quality Cuts	41.25	12.82	20.75	95.23	0.91	6473.51
SliceLID	36.75	9.29	8.80	2.65	0.52	2593.22
Preselection BDT	34.58	6.23	6.15	1.45	0.35	1386.85
Cosmic Rejection BDT	30.00	5.22	4.40	0.84	0.29	2.16

Table 8.3: SliceLID Analysis Selection Flow for 8.85×10^{20} POT


SliceLID vs. Cosmic Rejection BDT for Signal

SliceLID vs. Cosmic Rejection BDT for Background

Figure 8.4: Cosmic Rejection BDT/SliceLID Colormap

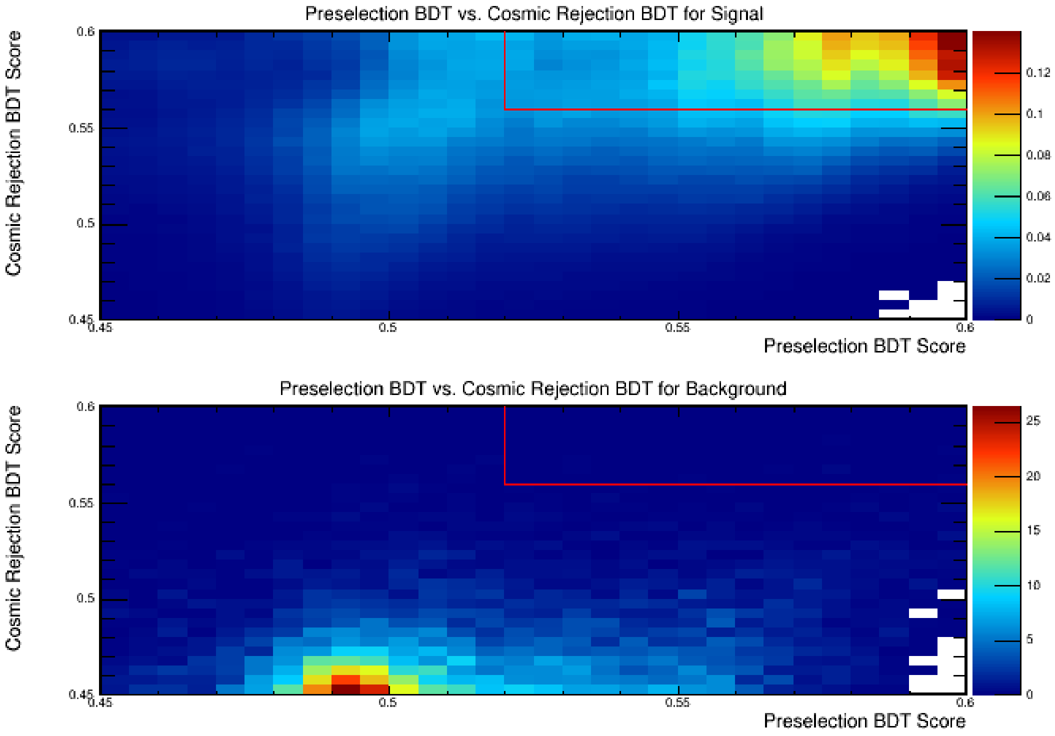


Figure 8.5: Cosmic Rejection BDT/Preselection BDT Colormap

8.3 Energy Estimation

A critical component to performing the ν_e appearance analysis is the accurate prediction of the neutrino energy. The calorimetric energy does not suffice for this task because of the amount of information being lost from the dead material. Therefore, a more sophisticated algortihm must be applied to regress the information from the slice to the true neutrino energy. The reconstructed energy is modeled by separating the energy of the interaction into an electromagnetic portion and a hadronic portion. The electromagnetic energy is found by summing the calorimetric energy of each cell hit in the prongs which pass the selection criteria of NOνA's electromagnetic particle identification algorithm. The hadronic energy is the calorimetric energy of every cell hit in the slice which does not belong to any of the electromagnetic prongs. A quadratic function of electromagnetic and hadronic energies is fit to give the final energy estimation [38].

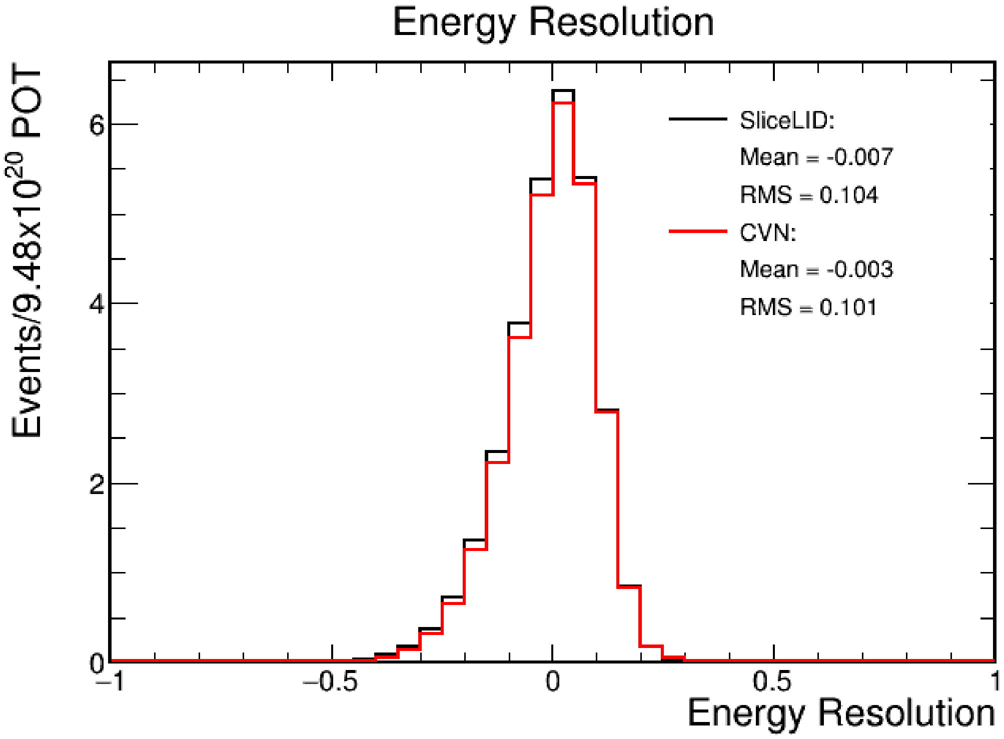


Figure 8.6: Energy Resolutions for this analysis and the standard ν_e preselection/cosmic rejection with optimal CVN cut

Chapter 9

Analysis

The main objective of NOνA is to quantify the sign of Δm^2_{32}, θ_{23}, and δ_{CP}. In order to accomplish this task, this analysis performs a joint fit, using the results of the 2018 FHC ν_μ disappearance result and the results of the multidimensional PID selection procedure with the 2018 FHC NuMI data. In order to obtain such a result, there must exist an unbiased prediction of the signal spectrum at the FD. With this prediction and the FD NuMI data, a maximum likelihood fit is made with both statistical and systematic uncertainties.

9.1 Predicting the FD Data

It is impossible for pure monte-carlo simulation to make high accuracy predictions on its own. There are too many experimentally unique phenomena to model, so fitting to data can greatly improve the simulation's predictive power. It is for this reason that NOνA uses the ND. The ND simulation is tuned to the ND data, so that simulation at the FD can be much more representative of the FD data. Creating a prediction at the FD requires several steps. The NuMI spectra as well as the background spectra must be corrected and then extrapolated to the FD. After this, FD specific backgrounds to the analysis such as neutrino interactions in the rock and cosmic rays must also be understood.

9.1.1 ND Decomposition

Due to its proximity to the NuMI Beam source, the ND contains a rich sample of many interactions. With these high statistics datasets and simulated samples, the relationship between truth quantities from the simulation and the reconstructed information is better understood.

The first decomposition is ν_μ CC decomposition. This technique selects events which pass the ν_μ CC selection requirements and compares the true energy from the simulation to the reconstructed energy from the data. Since the ND and FD are functionally identical, this information will be crucial for estimating the true energy of the signal from its reconstructed energy at the FD.

The next decomposition technique to be considered is called proportional decomposition. Proportional decomposition assumes that the beam content of the simulation and data are the same, so only the ratios of the spectra are used for the extrapolation procedure. Since the content of the FHC tuned NuMI beam is well understood, this decomposition technique is not used for this analysis.

The final decomposition procedure to be considered is the combination of Michel decomposition and beam electron neutrino (BEN) decomposition. Michel decomposition functions by selecting $\pi^\pm$ in the ν_μ CC sample. This selection procedure identifies the charged pions by the Michel electrons from the following decay:

$$\begin{aligned} \pi^+ &\to e^+ + \nu_e \\ \pi^- &\to e^- + \bar{\nu}_e \end{aligned} \tag{9.1}$$

By understanding the number of $\pi^\pm$ in the ν_μ ND sample, a better understanding of the NC contribution to the total background. BEN decomposition functions by selecting ν_μ CC events in the ND and recording the parents. The number of Kaon parents is an important quantity because the large Kaon mass offers a much greater phase space for the ν_e daughters. Therefore, BEN decomposition gives a greater handle on the number of intrinsic beam ν_e events.

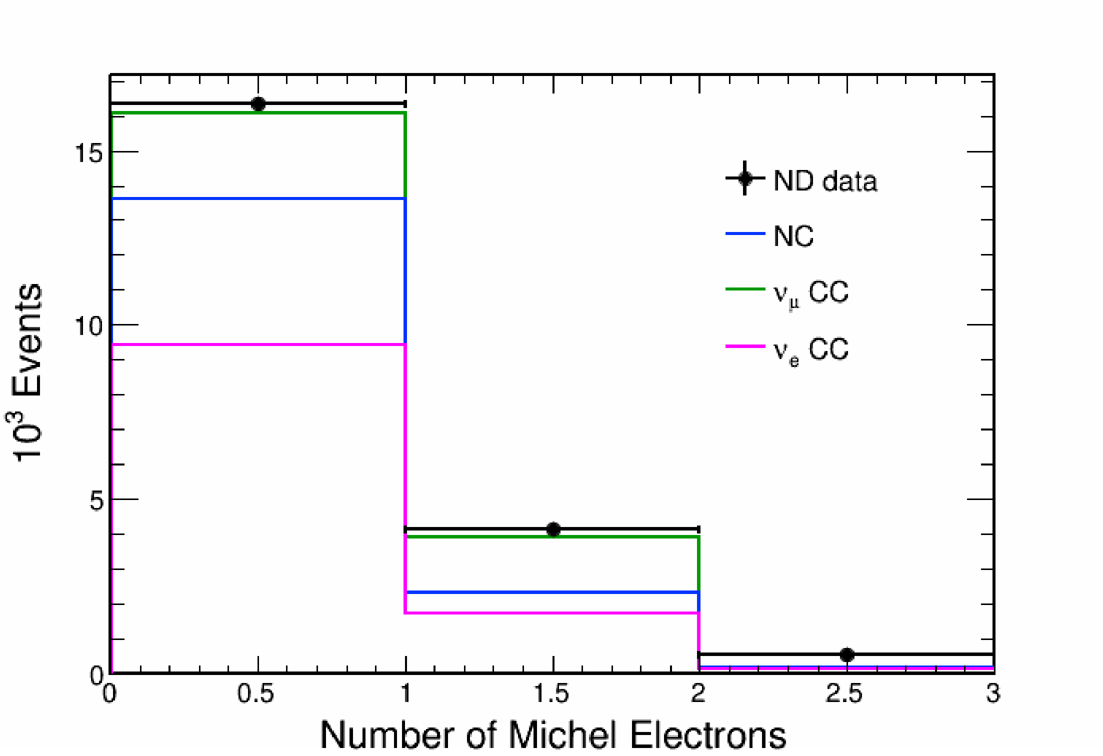


Figure 9.1: Comparison of the number of selected Michel electrons from the data and the corrected MC

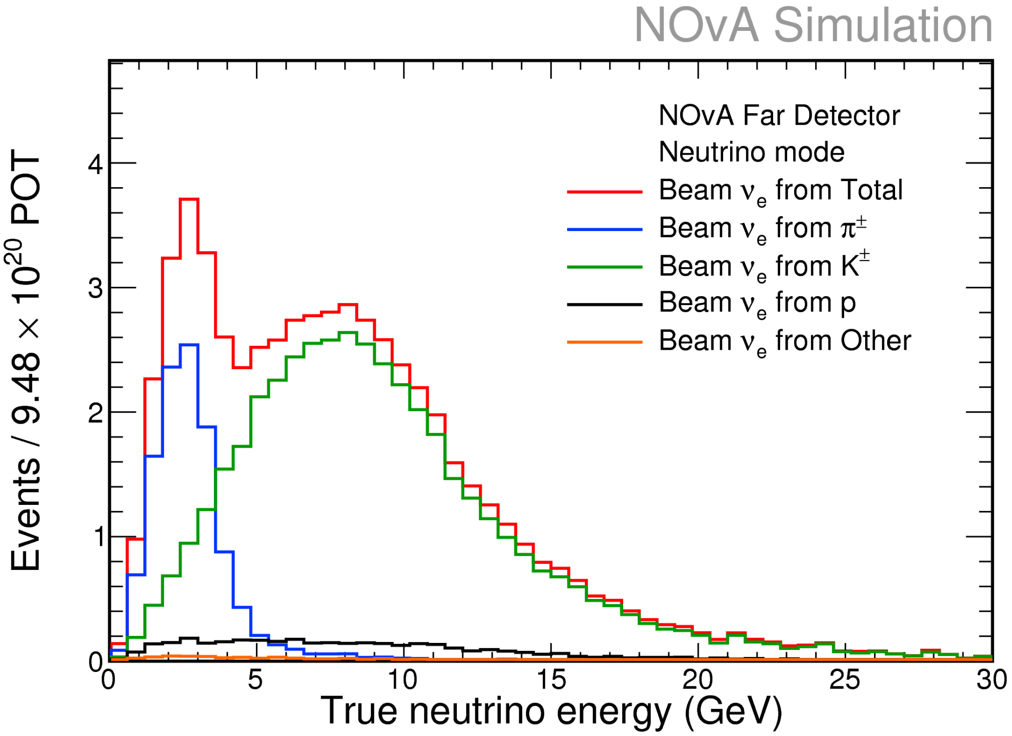


Figure 9.2: Beam ν_e sources at the FD

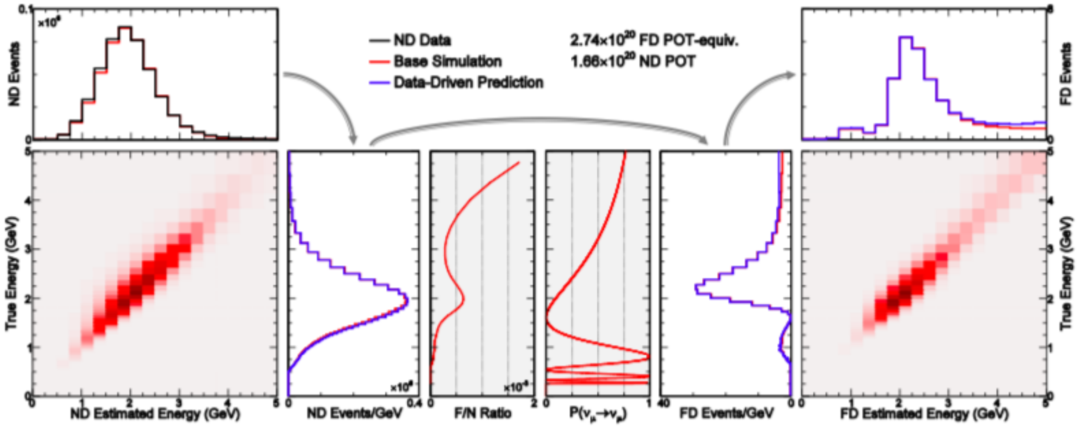


Figure 9.3: Steps implemented to for signal extrapolation at the FD

9.1.2 Near to Far Extrapolation

By now, the ND sample has been decomposed into its respective components. With this, a series of steps are made to relate FD reconstructed information to truth level information for each component. First, a matrix is formed which relates the truth information and reconstructed information in the ND. Next, a far/near ratio is applied to incorporate the flux change from having the FD 810 km away. Then simple oscillations are applied which make no assumptions about the oscillation parameters to be measured. Finally, the reconstructed data is transformed by the extrapolated matrix to yield truth level spectra. All of these steps can be seen in figure 9.1.2.

9.2 Treatment of Systematic Uncertainty

Uncertainties in the modeling of experimental operations lead to a decrease in sensitivity of the measurement of the oscillation parameters in question. This is because the systematic uncertainties add in quadrature to the statistical uncertainty. Therefore, in the large statistics limit, the measurement uncertainty is dominated by systematic uncertainty. It is of vital importance that the systematic uncertainties of any analysis be quantified, so that the parameter fit values can be given at a known confidence interval. NOνA has the advantage of having functionally identical near and far detectors, so many of the detector specific systematic uncertainties are either significantly reduced or completely eliminated. Nevertheless, there still exist sources of systematic uncertainty which cannot be neglected. They are broken up into the following categories:

- Calibration
- Light levels
- Cherenkov light
- Beam transport
- Hadron production
- Neutrino interaction modeling
- Extrapolation
- Normalization

  Ordinarily, there is a category for the modeling of neutrino interactions in the detectors' surroundings. However, this category will not be discussed because it is not incorporated into this analysis. The systematic uncertainties in this analysis are quantified by shifting the nominal sample positively and negatively by the quantity's quoted valued, and the uncertainty is propagated to the oscillation fit.

9.2.1 Calibration

One of the greatest sources of systematic uncertainty for the ν_e appearance analysis is calibration uncertainty. Though the ND and FD are functionally identical, they are calibrated differently. As mentioned in chapter 5, both detectors are calibrated with cosmic ray muons. Since the FD is located on the surface of the earth, it is exposed to a large statistics sample of cosmic rays, whereas the ND has a much lower cosmic ray interaction rate. In addition to this, since the ND is 100 meters underground, many of the cosmic ray muons range out of the minimum ionizing region of the Bethe-Bloch curve and cause a higher fraction of Bremsstrahlung Radiation.

In order to handle these uncertainties, deliberate miscalibrations are applied to the nominal simulated sample. These include a 5% shift to the absolute energy calibration scale, changes to the response slope shape, and random shifts in the MC. The effect of these shifts are determined by comparing the event rate difference to the nominal sample.

9.2.2 Light Levels and Cherenkov Light

The equation for the amount of light yielded in the scintillator is given in equation 4.2. It was discussed in chapter 4 that the scintillator saturates when a charged particle has a large $\frac{dE}{dx}$. This phenomenon is modeled by adding the Birks and Chou terms to the light yield equation. The systematic uncertainty of this is evaluated by shifting the thresholds for saturation and shifting the light level amount by 10% while compensating for that with a compensating shift to the absolute energy calibration. Further, the light yield from the Cherenkov model used by NOνA is shifted, yielding a 2.6% decline in proton response.

9.2.3 Beam Transport and Hadronic Production

Differences in the working conditions of the beam and its simulation is a source of systematic uncertainty. This is quantified by shifting the locations of working components and the beam spot size. Uncertainty in hadronic production is handled by generating a statistical ensemble of cases in which the hadron production models are varied within their uncertainties. Each case is propagated through the simulation chain, and their predicted flux is determined.

9.2.4 Neutrino Interaction Modeling

The neutrino interaction software, GENIE, sets default values for the parameters of the neutrino interactions. These values can be tuned to better match the data in a particular experiment. As a result of this, the parameters which are used for the interaction cross-sections, hadronization, and final state interactions are tuned to best match the ND data. The uncertainties from this are determined by the quality of the fit to the data.

9.2.5 Normalization and Extrapolation

The normalization systematic is simply a way for considering uncertainties in the exposure, fiducial mass, and ND reconstruction efficiency. The extrapolation systematics are the uncertainties which arise from predicting the FD oscillated ν_e signal from the ν_μ CC spectrum at the ND. Not only that, but the differences in the detector sizes lead to different levels of containment, and thus differences in the reconstructable phase space.

The effect of the quadrature summed systematic shifts on the selection can be seen in figures 9.4,9.5.

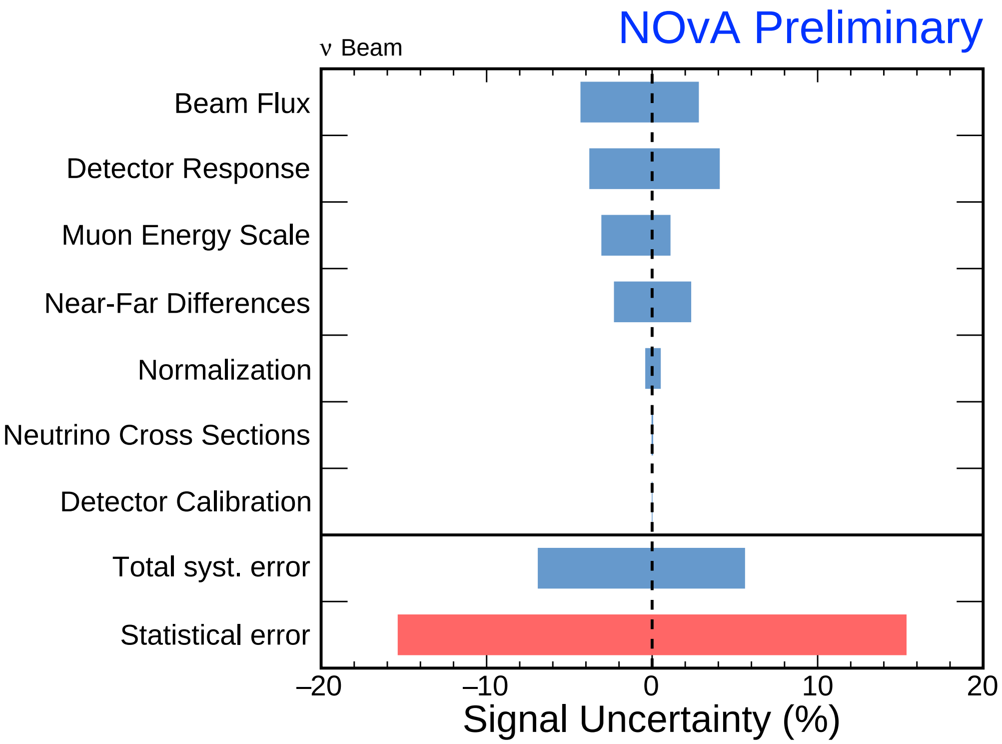


Figure 9.4: Net effect of all the systematics on the total uncertainty of the signal measurement.

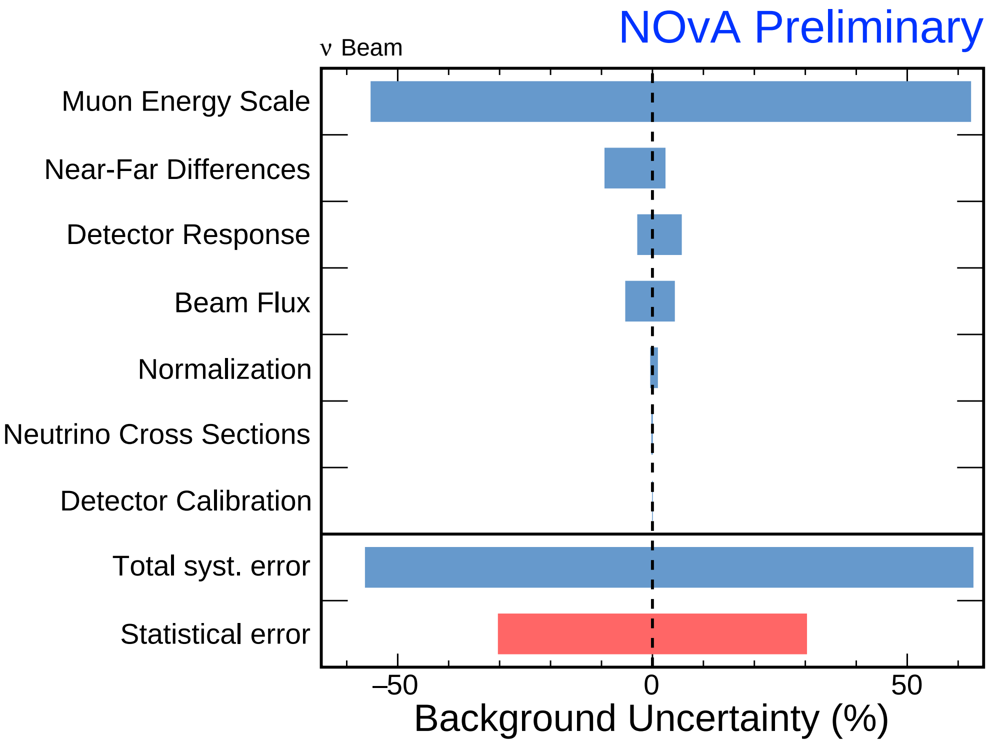


Figure 9.5: Net effect of all the systematics on the total uncertainty of the background measurement.

Chapter 10

Results

Now that the prediction objects are made and the FD NuMI data is collected, a fit to the ν_e appearance probability(equation 1.30) can be made. The discussion of parameter fitting in this chapter follows closely the section of maximum likelihood fitting in the statistics section of the PDG [12]. The oscillation parameters are tuned such that the maximum likelihood of the fit is found [39]. The data and predictions are binned by the reconstructed energy with a bin width of 0.25 GeV. Since the binned data is Poisson distributed, the likelihood function can be written as follows:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{\mu_i^{n_i}(\boldsymbol{\theta}) e^{-\mu_i(\boldsymbol{\theta})}}{n_i!} \tag{10.1}$$

where N is the number of bins, n_i is the number of entries in the i^{th} bin which are Poisson distributed, $\boldsymbol{\theta}$ is the vector of tunable oscillation parameters, and $\mu_i(\boldsymbol{\theta})$ is the average number of entries per bin. With binned data, one is able to obtain a goodness of fit statistic and maximum likelihood estimators by maximizing the likelihood function. Maximizing the likelihood accomplishes the same task as maximizing the likelihood ratio:

$$\lambda(\boldsymbol{\theta}) = \frac{L(\boldsymbol{n}; \boldsymbol{\theta})}{L(\boldsymbol{n}; \boldsymbol{\mu})} \tag{10.2}$$

The denominator is the likelihood function with tunable means which have maximum likelihoods at $\boldsymbol{\mu} = \boldsymbol{n}$. It is more convenient to work with the logarithm of the likelihood because the differentiation from optimization handles sums over products, and the parameters which maximize the log-likelihood also maximize the likelihood. This yields:

$$ln(\lambda(\boldsymbol{\theta})) = \sum_{i=1}^{N} \Bigg[n_i - \mu_i(\boldsymbol{\theta}) - n_i ln\Big(\frac{n_i}{\mu_i(\boldsymbol{\theta})}\Big) \Bigg] \tag{10.3}$$

Most often, the maximum likelihood function is found numerically. The most common method being gradient descent. Therefore, a -2 factor is added, which is convergent with gradient descent.

$$-2ln(\lambda(\boldsymbol{\theta})) = 2\sum_{i=1}^{N} \Bigg[\mu_i(\boldsymbol{\theta}) - n_i + n_i ln\Big(\frac{n_i}{\mu_i(\boldsymbol{\theta})}\Big) \Bigg] \tag{10.4}$$

If μ_i is sufficiently large, $-2ln(\lambda(\hat{\boldsymbol{\theta}}))$ is the χ^2 distribution for Poisson distributed data.

10.1 Systematic Nuisance Parameters

No experimental operation is perfect. There always exists sources of systematic uncertainty which hinder the performance of fitting. Therefore, additional degrees of freedom are added to the fit by means of nuisance parameters $\boldsymbol{\nu}$. The addition of nuisance parameters decreases systematic uncertainty at the cost of increasing statistical uncertainty. This is due to the fact that finding maximum likelihood values for all the parameters leads to correlations between $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$. To reduce the effect of overfitting the model to the nuisance parameters, a penalty term in the form of the variance weighted Frobenius Norm is added to the log-likelihood ratio giving:

$$-2ln(\lambda(\boldsymbol{\theta},\boldsymbol{\nu})) = 2\sum_{i=1}^{N} \Bigg[\mu_i(\boldsymbol{\theta},\boldsymbol{\nu}) - n_i + n_i ln\Big(\frac{n_i}{\mu_i(\boldsymbol{\theta},\boldsymbol{\nu})}\Big) \Bigg] + \sum_{i=1}^{M} \frac{\nu_i^2}{\sigma_i^2} \tag{10.5}$$

With this, a map of the $\Delta\chi^2$ can be made in the parameter space with which confidence intervals can be obtained.

10.2 Feldman-Cousins Corrections

In cases where the statistics are not large, such as this one, $\Delta\chi^2$ alone cannot be used for creating confidence intervals. Not only that, the parameters being used in the fit do not span the real number line. Instead, they have physical boundaries. It is for this reason, that a method for handling small signal statistics with physical boundaries was created by Feldman-Cousins [40]. This approach is a Frequentist approach in which many experiments are run with simulation to determine probabilities for fluctuations of a particular parameter near physical boundaries while fixing all other parameters. Due to the time and computation cost of the Feldman-Cousins procedure, the corrections will not be applied to the fitting procedure in this analysis.

10.3 Fitting the Data

After performing the selection procedure over the entire FHC dataset, 54 events in total were observed, compared to the prediction of 42.8 signal events and 15.3 background events, resulting in an observation of ν_e appearance at 5.27σ significance. The spectrum can be seen in figure 10.3. By using the maximum likelihood fitting procedure previously discussed with both ν_e and ν_μ selected data, the analysis preferred maximal mixing with the normal mass hierarchy and $\frac{\delta_{CP}}{\pi} = 1.92^{+0.08}_{1.19}$.

The contours and uncorrected significance slices for the for the parameter fits can be seen below.

10.4 Event Displays

In this section, several interesting event displays will be shown. These will show a glimpse of the similarities and differences of the selection of this analysis and the official analysis. Plots will be shown for both selected NuMI triggered events and out of time cosmic events.

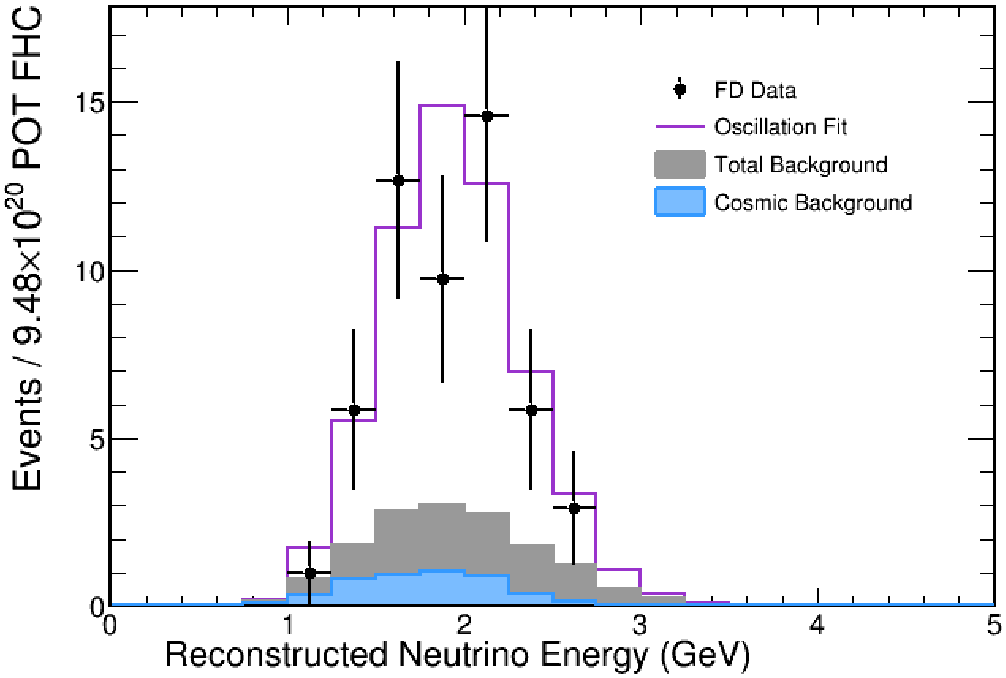


Figure 10.1: Comparison of the FD prediction with the selected NuMI data

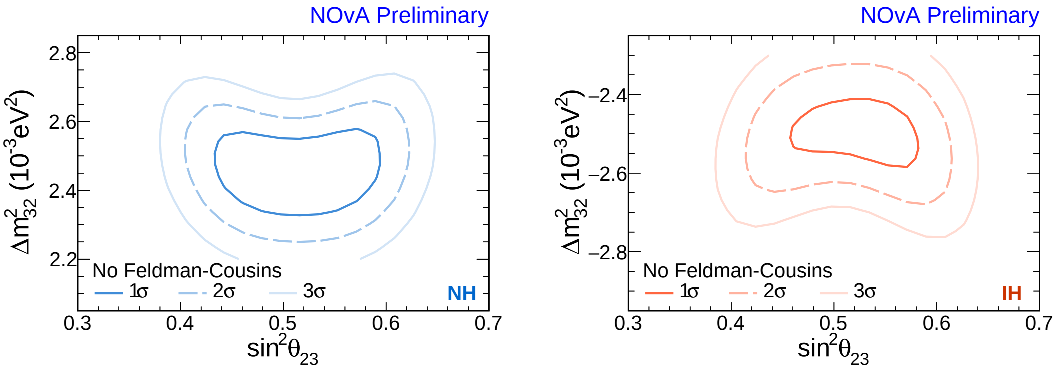


Figure 10.2: Contours showing the uncorrected confidence intervals for Δm^2_{32} vs θ_{23}

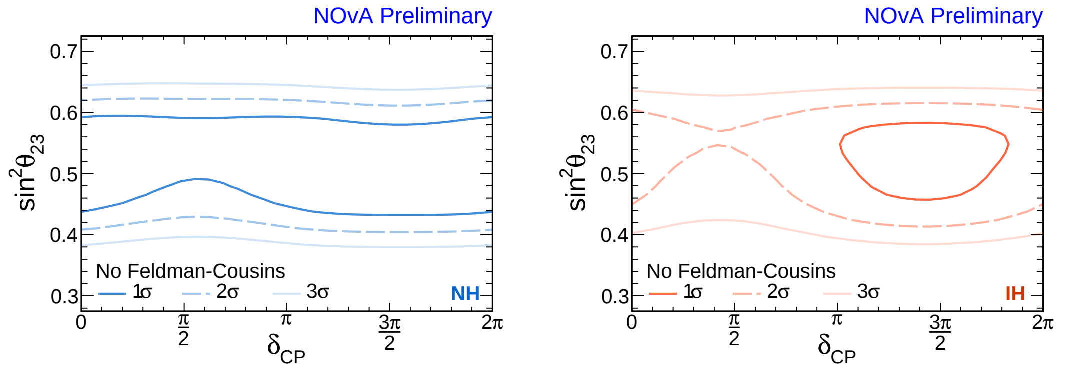


Figure 10.3: Contours showing the uncorrected confidence intervals for Δm^2_{32} vs θ_{23}

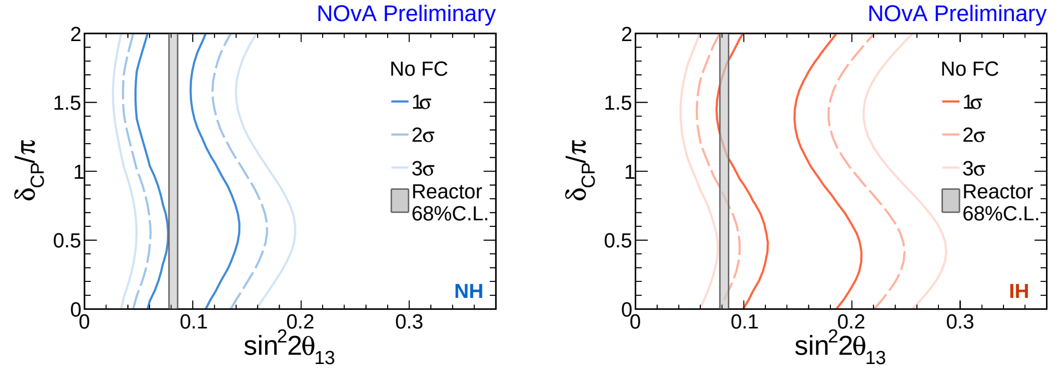


Figure 10.4: Contours showing the uncorrected confidence intervals for δ_{CP} vs θ_{13}

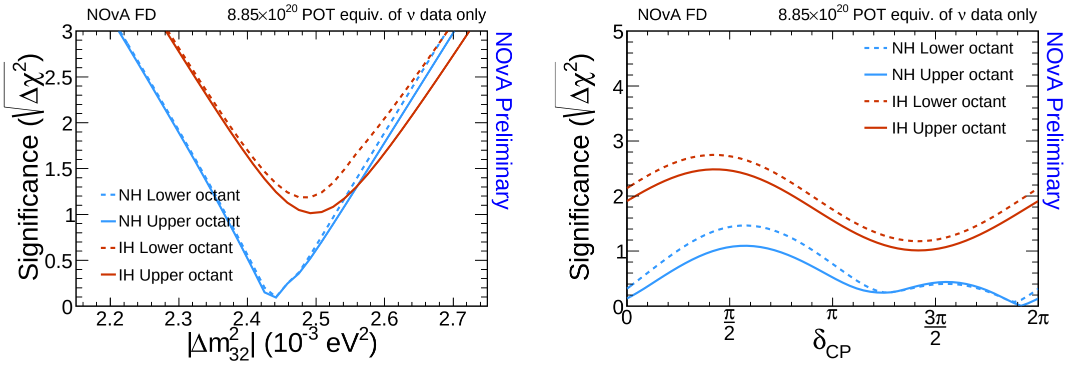


Figure 10.5: Uncorrected significance slices for Δm^2_{32} and δ_{CP}

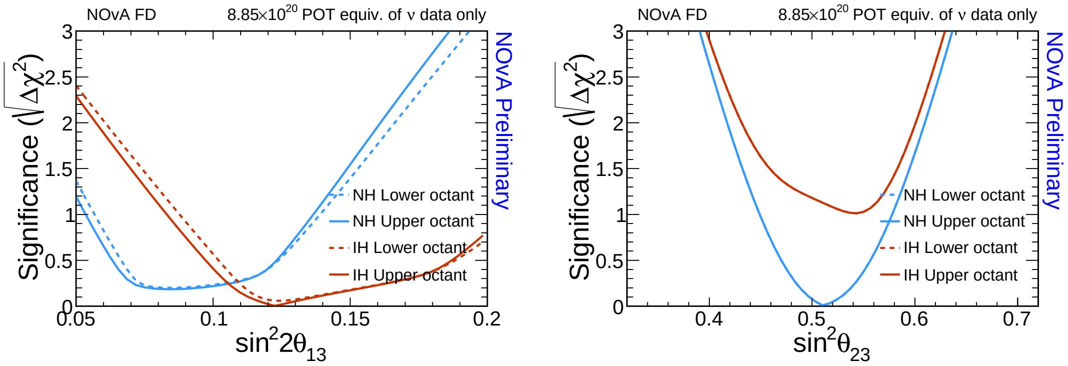


Figure 10.6: Uncorrected significance slices for $sin^2 2\theta_{13}$ and $sin^2\theta_{23}$

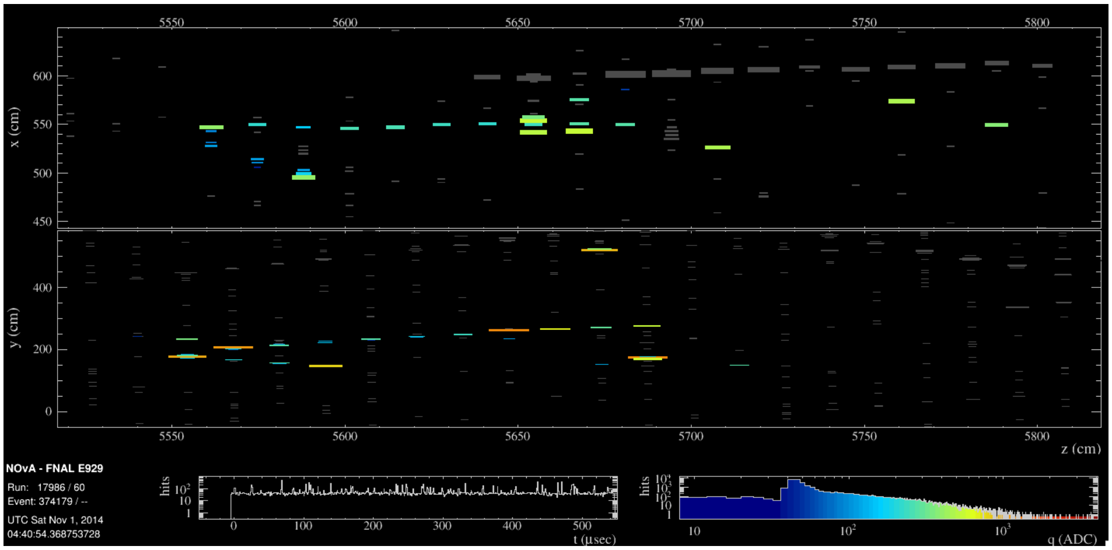


Figure 10.7: An event which passes all of the selection for this analysis but has a CVN score ¡ 0.2

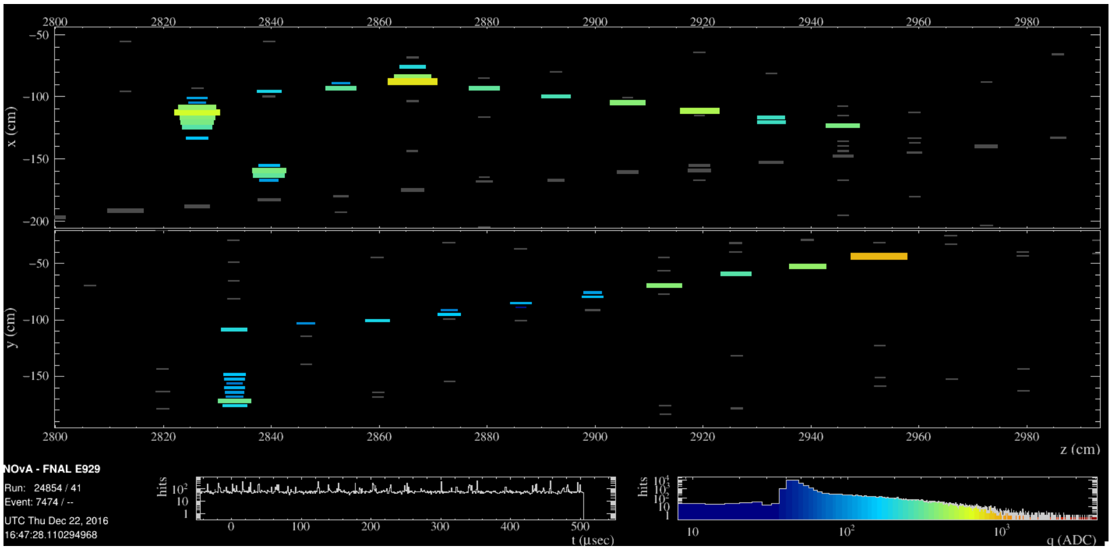


Figure 10.8: An event which fails the selection of this analysis but passes CVN selection

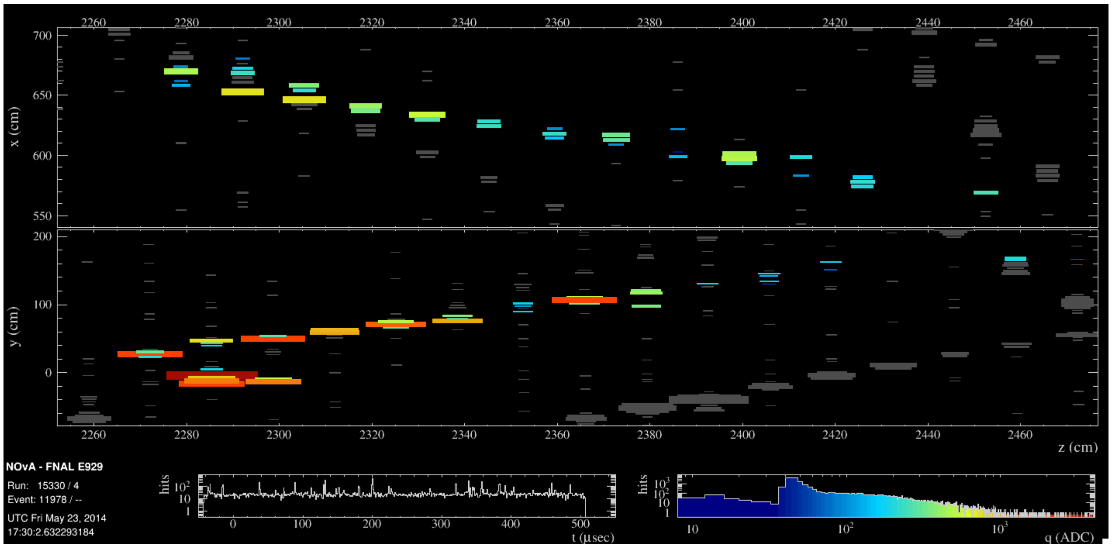


Figure 10.9: An event which passes both SliceLID and CVN selection

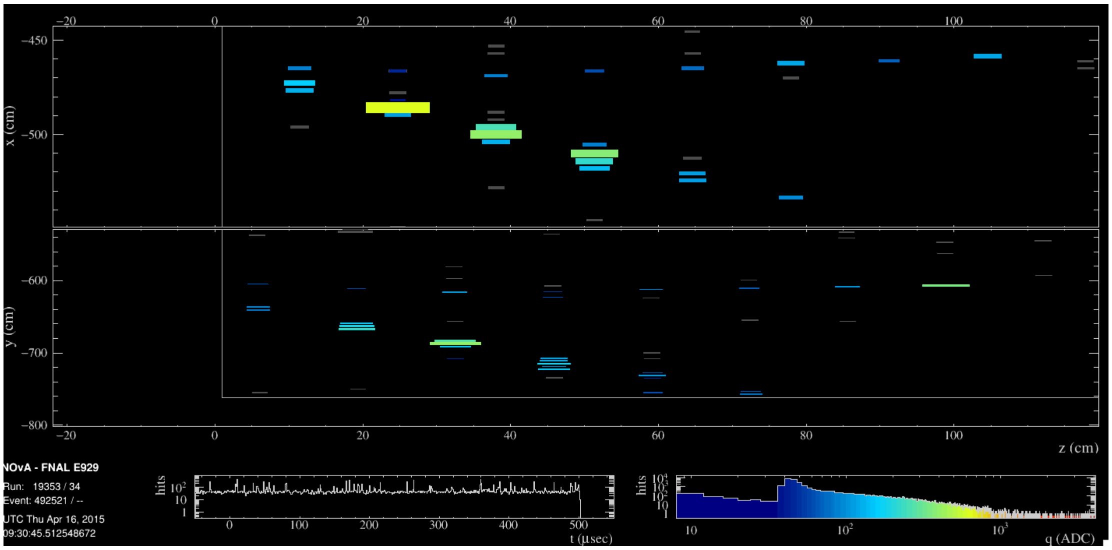


Figure 10.10: An event in the timing sideband which fails the selection of this analysis but passes CVN selection for its peripheral sample

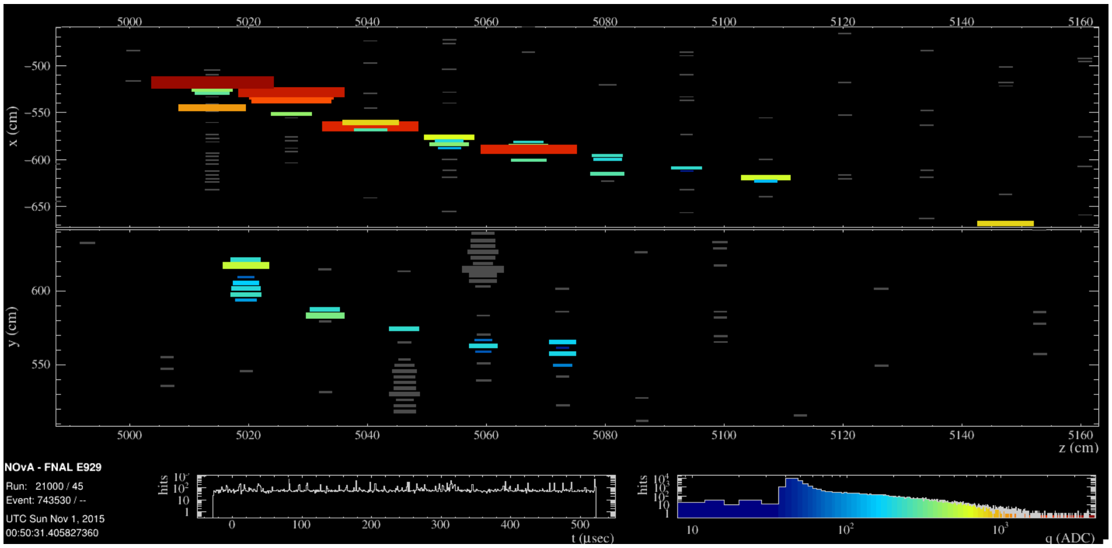


Figure 10.11: An event which passes both PID selection criteria in the timing sideband

Chapter 11

Conclusion

In this thesis, it was shown that a physics based recurrent neural network ν_e CC identifier achieved selection performance comparable to that of a deep convolutional neural network ν_e CC identifier. Using high level physics inputs to the model yielded smaller systematic uncertainties than the low level cell energy information. In addition to the recurrent neural network identifier with shower level inputs, preselection and cosmic rejection boosted decision trees were trained, and rectangular cuts to the resulting three dimensional particle identifier space were made to optimize the figure of merit for parameter measurement. With this selection criteria, 54 events were selected in the trigger spill window at the FD. 43 signal events with 15 background were predicted. With the predicted and data spectra of the reconstructed ν_e CC energy, the best fit of this analysis was determined to be normal hierarchy, maximal mixing, and $\delta_{CP} = 1.9\pi$ with a yet to be determined statistical significance.

References

[1] J Chadwick. Intensittsverteilung im magnetischen Spectrum der β-Strahlen von radium B + C. *Verhandl. Dtsc. Phys. Ges.*, 16:383, 1914.

[2] Niels Bohr. Faraday lecture. chemistry and the quantum theory of atomic constitution. *J. Chem. Soc.*, pages 349–384, 1932.

[3] Wolfgang Ernst Friederich Pauli. Open letter to radioactive persons. Technical report, 1930.

[4] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire. Detection of the free neutrino: a confirmation. *Science*, 124(3212):103–104, 1956, http://science.sciencemag.org/content/124/3212/103.full.pdf.

[5] I. V. Anicin. The Neutrino - Its Past, Present and Future. *ArXiv Physics e-prints*, March 2005, physics/0503172.

[6] K. Kodama et al. Observation of tau neutrino interactions. *Physics Letters B*, 504(3):218 – 224, 2001.

[7] Chien-Shiung Wu, Ernest Ambler, RW Hayward, DD Hoppes, and Ralph Percy Hudson. Experimental test of parity conservation in beta decay. *Physical review*, 105(4):1413, 1957.

[8] Bruno Pontecorvo. Neutrino experiments and the problem of conservation of leptonic charge. *Sov. Phys. JETP*, 26(984-988):165, 1968.

[9] Bruce T. Cleveland, Timothy Daily, Jr. Raymond Davis, James R. Distel, Kenneth Lande, C. K. Lee, Paul S. Wildenhain, and Jack Ullman. Measurement of the solar

electron neutrino flux with the homestake chlorine detector. *The Astrophysical Journal*, 496(1):505, 1998.

[10] Esteban Roulet. Mikheyev-smirnov-wolfenstein effect with flavor-changing neutrino interactions. *Physical Review D*, 44(4):R935, 1991.

[11] Raymond Davis Jr, Don S Harmer, and Kenneth C Hoffman. Search for neutrinos from the sun. *Physical Review Letters*, 20(21):1205, 1968.

[12] C Patrignani, Particle Data Group, et al. Review of particle physics. *Chinese physics C*, 40(10):100001, 2016.

[13] Q Retal Ahmad, RC Allen, TC Andersen, JD Anglin, G Bühler, JC Barton, EW Beier, M Bercovitch, J Bigu, S Biller, et al. Measurement of the rate of ν e+ d p+ p+ e- interactions produced by b 8 solar neutrinos at the sudbury neutrino observatory. *Physical Review Letters*, 87(7):071301, 2001.

[14] FP An, JZ Bai, AB Balantekin, HR Band, D Beavis, W Beriguete, M Bishai, S Blyth, K Boddy, RL Brown, et al. Observation of electron-antineutrino disappearance at daya bay. *Physical Review Letters*, 108(17):171803, 2012.

[15] P Adamson, C Andreopoulos, R Armstrong, DJ Auty, DS Ayres, C Backhouse, G Barr, M Bishai, A Blake, GJ Bock, et al. Measurement of the neutrino mass splitting and flavor mixing by minos. *Physical Review Letters*, 106(18):181801, 2011.

[16] T Eberl. Nuclear instruments and methods in physics research, section a: Accelerators, spectrometers, detectors and associated. *Nucl. Instrum. Methods Phys. Res., A*, 502:306–33, 2003.

[17] Susan M. Kasahara. Nova data acquisition system. *Physics Procedia*, 37:1876 – 1883, 2012. Proceedings of the 2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011).

[18] A. Fasso’, A. Ferrari, S. Roesler, J. Ranft, P. R. Sala, G. Battistoni, M. Campanella, F. Cerutti, L. De Biaggi, E. Gadioli, M. V. Garzelli, F. Ballarini, A. Ottolenghi, D. Scannicchio, M. Carboni, M. Pelliccioni, R. Villari, V. Andersen,

A. Empl, K. Lee, L. Pinsky, T. N. Wilson, and N. Zapp. The FLUKA code: present applications and future developments. *ArXiv Physics e-prints*, June 2003, physics/0306162.

[19] M Campanella, PR Sala, S Vanini, and A Ferrari. First calorimeter simulation with the flugg prototype. Technical report, ATL-COM-SOFT-99-004, 1999.

[20] CH Llewellyn Smith. Neutrino reactions at accelerator energies. *Phys. Rept*, 3(5):261–379, 1972.

[21] Dieter Rein and Lalit M Sehgal. Neutrino-excitation of baryon resonances and single pion production. *Annals of Physics*, 133(1):79–153, 1981.

[22] R Ph Feynman, M Kislinger, and F Ravndal. Current matrix elements from a relativistic quark model. *Physical Review D*, 3(11):2706, 1971.

[23] Arie Bodek and UK Yang. Higher twist, ξw scaling, and effective lo pdfs for lepton scattering in the few gev region. *Journal of Physics G: Nuclear and Particle Physics*, 29(8):1899, 2003.

[24] John Betteley Birks. Scintillations from organic crystals: specific fluorescence and relative response to different radiations. *Proceedings of the Physical Society. Section A*, 64(10):874, 1951.

[25] CN Chou. The nature of the saturation effect of fluorescent scintillators. *Physical Review*, 87(5):904, 1952.

[26] Cleveland WS LOWESS. A program for smoothing scatterplots by robust locally weighted regression [tekst]. *The American Statistician*, 35:54, 1981.

[27] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.

[28] Michael David Baird. An analysis of muon neutrino disappearance from the numi beam using an optimal track fitter. 2015.

[29] Evan Niner. Observation of electron neutrino appearance in the numi beam with the nova experiment. 2015.

[30] Leandro AF Fernandes and Manuel M Oliveira. Real-time line detection through an improved hough transform voting scheme. *Pattern recognition*, 41(1):299–314, 2008.

[31] Joseph C Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. 1973.

[32] James C Bezdek. Objective function clustering. In *Pattern recognition with fuzzy objective function algorithms*, pages 43–93. Springer, 1981.

[33] Kanika Sachdev. Muon neutrino to electron neutrino oscillation in nova. 2015.

[34] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[35] Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with lstm. 1999.

[36] A Hcker, J Stelzer, F Tegenfeldt, H Voss, K Voss, A Christov, S Henrot-Versill, M Jachowski, A Krasznahorkay, Y Mahalalel, X Prudent, and P Speckmayer. TMVA - Toolkit for Multivariate Data Analysis with ROOT: Users guide. Technical Report CERN-OPEN-2007-007, CERN, Geneva, Mar 2007.

[37] mariajesusbigml. Introduction to Boosted Trees, Mar 2017.

[38] Fernanda Psihas. *Measurement of Long Baseline Neutrino Oscillations and Improvements from Deep Learning.* PhD thesis, Indiana University, 2018.

[39] John Aldrich et al. Ra fisher and the making of maximum likelihood 1912-1922. *Statistical science*, 12(3):162–176, 1997.

[40] Gary J Feldman and Robert D Cousins. Unified approach to the classical statistical analysis of small signals. *Physical Review D*, 57(7):3873, 1998.