Improving the NOvA 3-Flavour Neutrino Oscillation Analysis

Veera Mikola

Doctoral Thesis

University College London

July 2023

Declaration

I, **Veera Mikola**, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Veera Mikola
July 20, 2023

Abstract

NOvA is a long-baseline neutrino experiment studying neutrino oscillations, a quantum mechanical interference phenomena where the observed neutrino flavour differs from that measured earlier, stemming from neutrino mass and flavour states mixing. NOvA consists of two functionally identical tracking calorimeter detectors deployed in the Fermilab NuMI beam. Both detectors are placed 14.6 mrad off the beam axis to achieve a narrow energy peak at 1.8 GeV at an oscillation maximum.

The NOvA 3-flavour oscillation analysis measures the neutrino oscillation parameters $sin^2\theta_{23}$ and Δm^2_{32} as well as sets limit to δ_{CP}, the octant of θ_{23} and the sign of Δm_{32}. The event selection for 3-flavour neutrino oscillation analysis ensures the maximum quantity of signal is made available and a minimal amount of background is present. The current disappearance analysis selection has an efficiency of 80% for selecting ν_μ CC events, and with some improvements additional events could be recovered into the analysis to improve the sensitivity to the aforementioned oscillation parameters. In an effort to recover the currently rejected signal event to the analysis, these events were trained in a classification neural network. The aim of the network was to divide the data into signal (ν_μ CC events) and background (NC and ν_e CC events). The highest performing network gave an additional figure of merit gain of 2.34 increasing the sensitivity by 4.3% in effective POT equal to 33 days of additional data taking. This was compared to changing the current particle identification event selection cuts, the best result out of the tested cut combinations gave an additional FOM of 3.36 equal to 3.7% increase in effective POT equivalent to 28 days of data taking.

Impact Statement

The work presented in this thesis was completed with the intention of improving the understanding the phenomena of neutrino oscillations in the NOvA neutrino oscillation experiment. Despite being one of the most abundant particles in nature, there are still a lot that is not known about the nature of neutrinos. Neutrino oscillations are a widely researched area in particle physics and have the potential to profoundly change our understanding of the Universe. Neutrino oscillations may provide answers to one of the most fundamental questions in the Universe: why is the Universe made of matter instead of antimatter? Neutrino oscillations can also provide insight into the nature of neutrino mass, which cannot be explained by the Standard model, currently the most comprehensive understanding of particle physics theory. This thesis uses neural networks in its analysis, which are widely used in many applications in particle physics as well as other areas of scientific research.

Outside academia, neutrino physics research has the potential to provide important technological advancements. Particle physics has already provided technology for particle detectors used in medical applications. Particle detectors produce a large amount of data which has forced researchers to find new techniques for storing and handling large data sets. These advancement have been adopted by private industries as well as governments outside of research environments. Especially the use of machine learning is widespread in a lot of industries, such as the tech industry. In the era of big data and the surge of artificial intelligence, understanding these technologies is crucial.

Acknowledgements

Firstly, I want to thank my supervisor Prof. Ryan Nichol for guidance and support throughout my PhD. I would like to thank my fellow PhD students and postdocs at UCL and all of my colleagues at NOvA for all the help and great memories we have experienced together. Specifically, I would like to thank Mike Wallbank and Kevin Mulder for answering all of my coding questions, and both Flavia Cicala and Simeon Bash for help and support when I needed it the most. Thank you to all the lovely people that I was fortunate enough to share my time at Fermilab with, especially thanks to Henry Lay for being a great friend.

Most of all, I want to thank my parents and my amazing sisters. Thank you for all the love, support and laughs. And finally, a big thank you to my partner Frazer, this would have never been possible without your support.

Contents

List of Figures

List of Tables

Chapter 1

Neutrinos

1.1 Introduction

Neutrinos are amongst the most abundant particles in the Universe. First postulated by Pauli in the 1930s [1], they were introduced as a way to solve the continuous energy of the final state electron in beta decays. If only one particle would be emitted in the decay, the energy of the final state electron should not be continuous but discrete to follow all principles of energy conservation. Pauli introduced the idea that as well as an electron, a chargeless fermion is emitted in the decay. Due to the lack of electric charge, Pauli thought neutrinos could never be detected and described his theory as a desperate remedy to help solve the problem.

Despite Pauli's doubts, neutrinos were experimentally discovered in 1956 [2]. This first discovery was that of an electron antineutrino, the first of the three neutrinos. The experiment was looking to record a signal of neutrinos reacting via an inverse beta decay.

$$\mathrm{p} + \overline{\nu}_e \longrightarrow \mathrm{n} + \mathrm{e}^+. \tag{1.1.1}$$

The experiment consisted of a large interaction volume filled with water and cadmium working as a liquid scintillator close to a nuclear reactor. The signal recorded consisted of two parts: rapidly annihilating positron signal and a capture of a neutron in the cadmium, which caused a photon signature from the nuclear de-excitation.

As with other leptons in the Standard Model (SM), neutrinos come in three flavours: electron, muon and tau. The muon neutrino was discovered in 1962 by the Brookhaven Experiment [3] and the tau neutrino in 2001 by the DONUT experiment [4]. The Brookhaven experiment used a 10-ton aluminium spark chamber to measure pions decaying to neutrinos and muons, and DONUT experiment used emulsion cloud chambers to measure tau neutrinos from leptonic decays of mesons, (mostly from D_S mesons).

The original description in the SM did not make a statement about the mass of the neutrinos so they were conventionally assumed to be massless in the SM. However due to the phenomenon of neutrino oscillations, neutrinos are now proven to have mass. Neutrino oscillations are a quantum mechanical interference phenomenon where the observed flavour of a neutrino differs from that measured earlier due to the mass (ν_1, ν_2, ν_3) and flavour states (ν_e, ν_μ, ν_τ) mixing. They will be explained more in section (1.3).

1.2 Neutrino Interactions

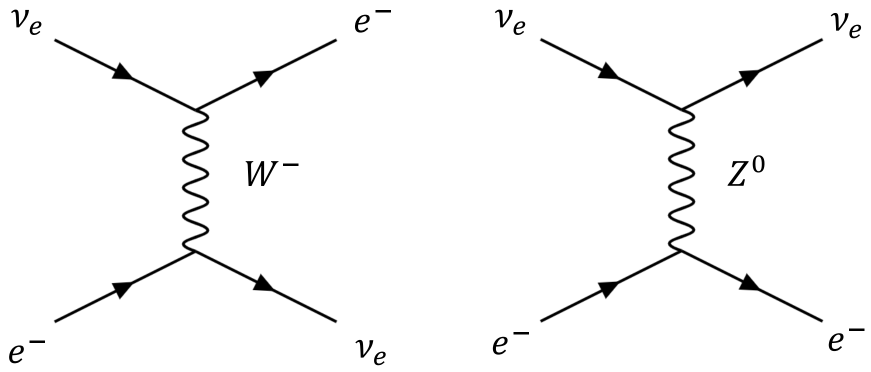

Figure 1.1: Feynman diagram for charged current (left) and neutral current (right) electron neutrino interaction.

Neutrinos are never directly observed, but only detected indirectly in particle interactions. Neutrino interactions are described in detail in the SM. Neutrinos only interact via weak force, which is a force mediated by three bosons: W^+, W^-, and a neutral Z^0 boson.

In weak interactions, $W^\pm$ mediate a charged current (CC) interactions, in which a neutrino changes into a corresponding lepton (ν_e with e etc.) with W boson sign chosen so that electric charge is conserved in the interaction. An example of this type of decay in neutrinos is seen on the left in figure 1.1 [5].

There exists three different types of CC neutrino interactions describing how the neutrino interacts with the target nucleus in the interaction. These are quasi-elastic scattering (QE), resonant pion production (RES) and deep inelastic scattering (DIS). In QE, neutrinos scatter quasi-elastically off an entire nucleon and liberate a nucleon from the target. It is the dominant type of interaction in low energies (energy less than 1 GeV). When moving to higher energies, the neutrino has enough energy

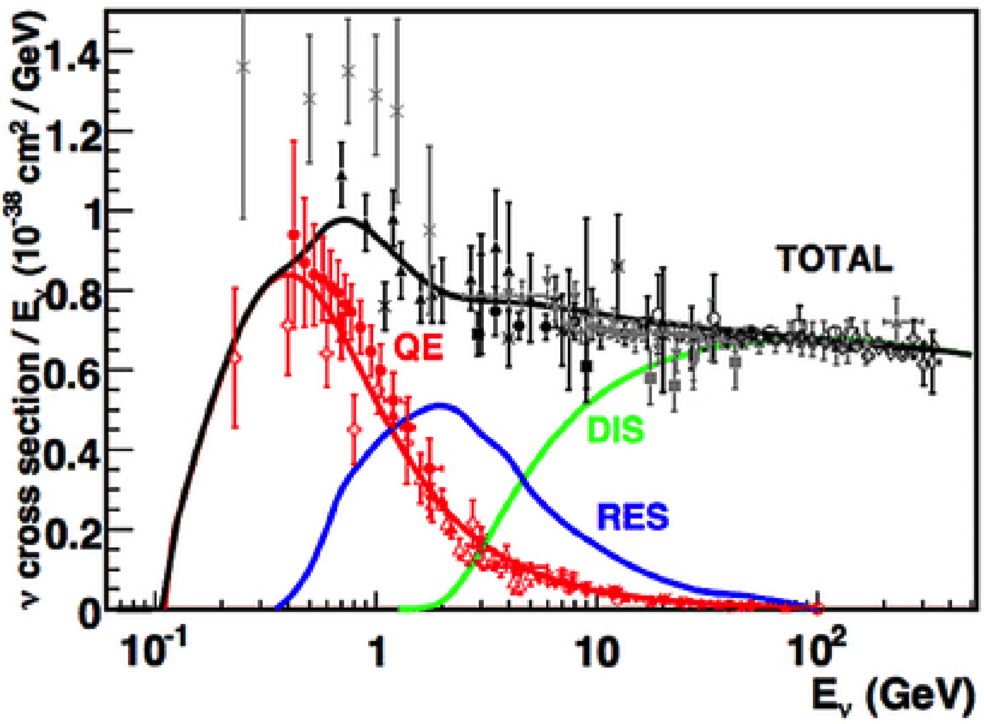

Figure 1.2: Types of CC interactions for different energies. QE (red) is for quasi-elastic scattering, dominating in low energies. RES (blue) is for resonant pion production and DIS (green) for deep inelastic scattering, the most prominent in high energies. The total cross-section taking into account all interaction types is also marked in black. Retrieved from [5].

to excite a nucleon to a resonance state, such as Δ^{++}. When the nucleon decays back to its normal state additional mesons (such as pions) are emitted [5].

With enough momentum transferred by the boson in an interaction, a neutrino can resolve individual quark constituents of the target nucleus resulting in a break up of the nucleon leading to a hadronic shower. This DIS interaction is the most dominant in the highest neutrino energies [5].

All of these interactions in terms of energy and cross-section are seen in figure 1.2. This illustrates what type of interaction dominates at which energy and what is the total cross-section [5].

Another type of weak interaction is a neutral current interaction mediated by a Z^0 boson. A Feynman diagram example of this interaction is seen on the right of figure 1.1. Since neutrinos are not detectable, all that is known about the interaction is what is detected from the interaction nucleus. That is why distinguishing between neutrino flavours in NC interactions is not possible [5].

1.3 Neutrino Oscillations

Neutrino oscillations were proposed first by Pontecorvo who created the first theory of neutrino oscillations in the 1950s [6]. Since at the time only one flavour of neutrinos was known, his first paper was not complete to describe neutrino oscillation theory as it is known now. Oscillations between 3-flavours were first considered by Maki, Nakagawa and Sakata in 1962 [7]. In the 70s, the oscillation theory was then shaped to a theory more like what it is now understood to be [8] [9].

1.3.1 Neutrino Mixing

The basic principle behind neutrino oscillations is mixing between neutrino flavour states (ν_e, ν_μ, ν_τ) and mass states (ν_1, ν_2, ν_3). Mathematically one can write [8]:

$$|\nu_\alpha\rangle = \sum_k U^*_{\alpha k} |\nu_k\rangle . \tag{1.3.1}$$

With α describing the three neutrino flavour states $\alpha = (e, \mu, \tau)$ and k the mass states $k = (1, 2, 3)$. U is the weight describing the mixing between flavours and masses.

In matrix form, the oscillations are described in Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [8]:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} . \tag{1.3.2}$$

Usually in neutrino physics, the quantities that are used to describe the oscillations include three mixing angles (θ_{12}, θ_{23}, θ_{13}), mass splittings (Δm_{12}, Δm_{23}) and a CP violation phase (δ_{CP}). These are derived from the PMNS matrix such that [10]:

$$sin^2\theta_{12} \equiv \frac{|U_{e2}^2|}{1 - |U_{e3}^2|} \tag{1.3.3}$$

$$sin^2\theta_{23} \equiv \frac{|U_{\mu 2}^2|}{1 - |U_{e3}^2|} \tag{1.3.4}$$

$$sin^2\theta_{13} \equiv |U_{e3}^2| \tag{1.3.5}$$

$$\delta_{CP} \equiv -\arg(U_{e3}). \tag{1.3.6}$$

It is then possible to write the weight U in terms of these oscillation parameters as [8] [9]:

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{CP}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.3.7}$$

where $s_{ab} = sin(\theta_{ab})$ and $c_{ab} = cos(\theta_{ab})$ with $a, b \in [1, 3], a \neq b$. δ_{CP} is a CP-violation term, which is discussed in more detail in section 1.5 [8] [9].

1.3.2 Two-Flavour Oscillations

This section and its equations were written using [8] [9] throughout.

For simplicity, only oscillations between two neutrino states will be considered in this section. Starting with an electron neutrino, considering a mix of the two mass eigenstates one gets

$$|\nu_e\rangle = U_{e1} |\nu_1\rangle + U_{e2} |\nu_2\rangle . \tag{1.3.8}$$

The mass eigenstates evolve in time according to free-particle solutions to the wave equation

$$\begin{aligned} |\nu_1\rangle (t) &= |\nu_1\rangle \exp\{i(\underline{P}_1 - E_1 t)\}, \\ |\nu_2\rangle (t) &= |\nu_1\rangle \exp\{i(\underline{P}_2 r - E_2 t)\}. \end{aligned} \tag{1.3.9}$$

Relating the mass and flavour eigenstates in a 2x2 unitary matrix

$$\begin{pmatrix} |\nu_e\rangle \\ |\nu_\mu\rangle \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} |\nu_1\rangle \\ |\nu_2\rangle \end{pmatrix} . \tag{1.3.10}$$

Thus the expression for the electron neutrino can be written as

$$|\nu_e\rangle (t) = \cos\theta |\nu_1\rangle \exp\{i(\underline{P}_1 r - E_1 t)\} + \sin\theta |\nu_2\rangle \exp\{i(\underline{P}_1 - E_1 t)\}. \tag{1.3.11}$$

At time T along positive z-axis at distance L the expression changes to

$$|\nu_e\rangle (L, T) = \cos\theta |\nu_1\rangle \exp\{i\phi_1\} + \sin\theta |\nu_2\rangle \exp\{i\phi_2\}, \tag{1.3.12}$$

where

$$\phi_i(L, T) = E_i T - |P_i| L. \tag{1.3.13}$$

Then one can express the mass eigenstates in terms of the weak eigenstates

$$\begin{aligned}|\psi(L,T)\rangle = \cos\theta(\cos\theta\,|\nu_e\rangle - \sin\theta\,|\nu_\mu\rangle)\exp\{-i\phi_1\} \\ +\sin\theta(\sin\theta\,|\nu_e\rangle + \cos\theta\,|\nu_\mu\rangle)\exp\{-i\phi_2\}.\end{aligned} \tag{1.3.14}$$

Rearranging the equation gives:

$$\begin{aligned}|\psi(L,T)\rangle = |\nu_e\rangle\,(\cos^2\theta\exp\{-i\phi_1\} + \sin^2\theta\exp\{-i\phi_2\}) \\ +|\nu_\mu\rangle\,(-\sin\theta\cos\theta\exp\{-i\phi_1\} + \sin\theta\cos\theta\exp\{-i\phi_2\}).\end{aligned} \tag{1.3.15}$$

Next, making an assumption that $|P_1|= |P_e|= P$ [11] and using an expression for kinematics $E^2 = m^2 + p^2$, $\phi_1 - \phi_2$ can be written as

$$\begin{aligned}\phi_1 - \phi_2 &= (E_1 - E_2)T \\ &= ((p^2 + m_1^2)^{1/2} - (p^2 + m_2^2)^{1/2}) \\ &= \frac{(m_1^2 - m_2^2)L}{2E},\end{aligned} \tag{1.3.16}$$

where in the last step an assumption that $p \cong E$ was made. For oscillations, the important quantity to find is a probability of $|\nu_e\rangle$ changing to $|\nu_\mu\rangle$.

$$\begin{aligned}P(\nu_e \to \nu_\mu) &= |\langle\nu_\mu|\Psi(L,T)\rangle\,|^2 \\ &= \sin^2 2\theta \sin^2\frac{\phi_1 - \phi_2}{2}\end{aligned} \tag{1.3.17}$$

Finally, an expression for the two flavour transformation probability is obtained

$$P_{\nu_e->\nu_\nu} = \sin^2(2\theta_{12})\sin^2(\frac{1.27\Delta m_{21}^2 L}{E}) \tag{1.3.18}$$

where $\Delta m_{21}^2 = m_2^2 - m_1^2$. A survivor probability for an electron neutrino can then be expressed as:

$$P_{\nu_e->\nu_e} = 1 - P_{\nu_e->\nu_\nu} = 1 - \sin^2(2\theta_{12})\sin^2(\frac{1.27\Delta m_{21}^2 L}{E}). \tag{1.3.19}$$

The prefix of 1.27 is added so in these equations L is measured in kilometers and E in GeV, which is easier for experimental consideration. When planning experiments, it is important that the variables that one has control over are chosen such that the oscillation probability is at a maximum. Looking at a version of an equation such that

$$\frac{1.27\Delta m_{21}^2 L}{E} = \frac{\pi}{2} \tag{1.3.20}$$

one can perform calculations to optimise the length of the baseline (L) as well as the energy (E) of the neutrino beam to get the maximum amount of oscillations for best results.

1.3.3 Three-Flavour Oscillations

This section and its equations were written using [8] [9] throughout.

The full 3-flavour oscillations are analogues to the 2-flavour case, however using the full PMNS matrix (equation 1.3.2). Considering an electron neutrino created in a CC interaction:

$$|\psi(0)\rangle = |\nu_e\rangle = U^*_{e1}|\nu_1\rangle + U^*_{e2}|\nu_2\rangle + U^*_{e3}|\nu_3\rangle . \tag{1.3.21}$$

The time evolution of the mass eigenstates is as follows:

$$|\psi(\mathbf{L},T)\rangle = U^*_{e1}|\nu_1\rangle e^{-i\psi_1} + U^*_{e2}|\nu_2\rangle e^{-i\psi_2} + U^*_{e3}|\nu_3\rangle e^{-i\psi_3}. \tag{1.3.22}$$

As before using equation 1.3.13 and writing in terms of weak eigenstates

$$\begin{aligned} |\psi(\mathbf{L},T)\rangle = U^*_{e1}&(U_{e1}|\nu_e\rangle + U_{\mu1}|\nu_\mu\rangle + U_{\tau1}|\nu_\tau\rangle)e^{-i\psi_1} \\ +U^*_{e2}&(U_{e2}|\nu_e\rangle + U_{\mu2}|\nu_\mu\rangle + U_{\tau2}|\nu_\tau\rangle)e^{-i\psi_2} \\ +U^*_{e3}&(U_{e3}|\nu_e\rangle + U_{\mu3}|\nu_\mu\rangle + U_{\tau3}|\nu_\tau\rangle)e^{-i\psi_3}. \end{aligned} \tag{1.3.23}$$

Using the unitarity condition of the PMNS, i.e. $UU^\dagger = 1$,

$$\begin{aligned} |\psi(\mathbf{L},T)\rangle = &(U^*_{e1}U_{e1}e^{-i\psi_1} + U^*_{e2}U_{e2}e^{-i\psi_2} + U^*_{e3}U_{e3}e^{-i\psi_3})|\nu_e\rangle \\ &+(U^*_{\mu1}U_{\mu1}e^{-i\psi_1} + U^*_{\mu2}U_{\mu2}e^{-i\psi_2} + U^*_{\mu3}U_{\mu3}e^{-i\psi_3})|\nu_\mu\rangle \\ &+(U^*_{\tau1}U_{\tau1}e^{-i\psi_1} + U^*_{\tau2}U_{\tau2}e^{-i\psi_2} + U^*_{\tau3}U_{\tau3}e^{-i\psi_3})|\nu_\tau\rangle . \end{aligned} \tag{1.3.24}$$

Thus the 3-flavour oscillation probability for ν_e to ν_μ is:

$$\begin{aligned} P(\nu_e \to \nu_\mu) = |\langle\nu_\mu|\Psi(L,T)\rangle|^2 \\ = |U^*_{e1}U_{\mu1}e^{-i\psi_1} + U^*_{e2}U_{\mu2}e^{-i\psi_2} + U^*_{e3}U_{\mu3}e^{-i\psi_3}|. \end{aligned} \tag{1.3.25}$$

Using complex number identities one gets:

$$\begin{aligned} P(\nu_e \to \nu_\mu) = 2\,\mathrm{Re}\Big\{U^*_{e1}U_{\mu1}U_{e2}U^*_{\mu2}[e^{i(\psi_2-\psi_1)} - 1]\Big\} \\ 2\,\mathrm{Re}\Big\{U^*_{e1}U_{\mu1}U_{e3}U^*_{\mu3}[e^{i(\psi_3-\psi_2)} - 1]\Big\} \\ 2\,\mathrm{Re}\Big\{U^*_{e2}U_{\mu2}U_{e3}U^*_{\mu3}[e^{i(\psi_3-\psi_2)} - 1]\Big\}. \end{aligned} \tag{1.3.26}$$

Using identities:

$$\begin{aligned}\mathrm{Re}\Big\{e^{i(\psi_j-\psi_1)}-1\Big\} &= \cos(\psi_j-\psi_i)-1\\ &= -2\sin^2\frac{\psi_j-\psi_i}{2} = -2s\sin^2(\Delta_{ji})\end{aligned} \tag{1.3.27}$$

$$\Delta_{ji} = \frac{\psi_j-\psi_i}{2} = \frac{(m_j^2-m_i^2)}{4E_\nu}. \tag{1.3.28}$$

The final 3-flavour probability can thus be written as:

$$\begin{aligned}P(\nu_e\rightarrow\nu_\mu) = 1 - 4&|U_{e1}|^2|U_{e2}|^2\mathrm{sin}^2(\Delta_{21})\\ -4&|U_{e1}|^2|U_{e3}|^2\mathrm{sin}^2(\Delta_{31})\\ -4&|U_{e2}|^2|U_{e3}|^2\mathrm{sin}^2(\Delta_{32}).\end{aligned} \tag{1.3.29}$$

1.4 Neutrino Oscillation Experiments

Multiple neutrino experiments over the years have managed to provide measurements of all three mixing angles, both mass differences as well as the sign of Δm_{12}. The least constrained neutrino oscillation parameter currently is δ_{CP}, which has not been measured to 5σ precision. The sign of Δm_{23} is also unknown, leaving the question of mass ordering open [12]. The last two are discussed in more detail in section 1.5.

The earliest neutrino oscillation experiments measured natural neutrino sources such as neutrinos from the Sun and the cosmic rays decaying to neutrinos in the atmosphere. Today, a lot of experiments focus on man-made neutrino sources such as neutrino beams and nuclear reactors, all of which will be discussed in this section.

1.4.1 Solar Neutrino Experiments

The earliest neutrino experiments all measured neutrinos from the Sun. From the beginning, these experiments showed a deficit in the number of observed solar neutrinos compared to what was predicted by the solar models describing the nuclear reactions in the Sun. The first experiment of this kind was conducted by Ray Davis, who discovered solar neutrinos via chlorine capture reaction in the Homestake mine [8]:

$${}^{37}\mathrm{Cl} + \nu_e \longrightarrow \mathrm{e}^- + {}^{37}\mathrm{Ar}. \tag{1.4.1}$$

The experiment discovered about one third of the expected rate of interactions [13]. Similar results were then recorded at other solar neutrinos experiments such as SAGE [14], GALLEX [15] and Kamiokande

[16].

The first measurements to confirm oscillations were from an experiment Super-Kamiokande (SK) [17] and the Sudbury Neutrino Observatory (SNO) [18]. Super-Kamiokande is a successor experiment to the Kamiokande experiment. It is a water Cherenkov detector, which measure light created when particles travel faster than the speed of light in the medium (water in this case). It is located at the Kamioka mine in Japan and it identifies neutrino flavours by the lepton created in the interaction that produces the Cherenkov light measured. Both its measurements from the Sun [19] and atmospheric neutrinos [17] (see section 1.4.2) showed evidence for oscillations. SK is not, however, sensitive to NC interactions since there are no leptons emitted in the final interaction state.

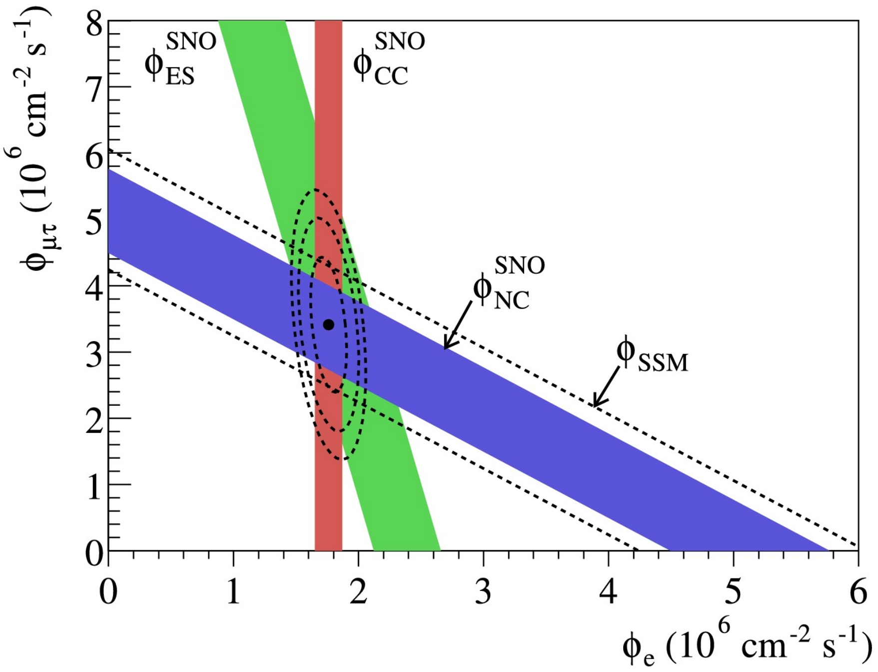

Figure 1.3: μ and τ solar neutrino flux vs e flux from the three types of neutrino interactions (NC, CC and elastic scattering, ES) measured by the SNO detector. Dashed lines show the flux predicted by a solar model and solid lines what is measured at the SNO detector. The bands intercept at the best fit value indicating that the results are consistent with neutrino oscillations and no distortion in the solar model. The limit of the bands represent $\pm 1\sigma$ errors. Retrieved from [20].

SNO was a heavy-water (D_2O) Cherenkov detector at Sudbury in Ontario, Canada. It had an advantage over Super-Kamiokande by being able to measure neutral current events as well as charged current events due to the use of heavy water. Thus, SNO measured both the total amount of neutrino interactions for all flavours via NC as well as the amount of ν_e CC interactions. The results are seen in figure 1.3, which show that the measured flux of all the flavours was consistent with measurements from the Sun. This then proves that the lack of ν_e was due to oscillations instead of an inaccurate solar model [8] [21].

The combination of results from SK and SNO proved the existence of neutrino oscillations and were awarded a Nobel Prize in physics in 2015 [22]. In terms of oscillation parameters, solar neutrino experiments measure θ_{12} and Δm^2_{12}. The current best values for these parameters are $sin^2\theta_{12} = 0.307 \pm 0.013$ and $\Delta m^2_{12} = (7.53 \pm 0.18) \times 10^{-5}\ eV^2$ [10].

1.4.2 Atmospheric Neutrino Experiments

Atmospheric neutrino experiments measure neutrinos from decays of cosmic rays that collide with nuclei in the Earth's atmosphere. The most common decays of cosmic rays producing neutrinos are [21]:

$$\begin{aligned} \pi^+ &\rightarrow \mu^+ + \nu_\mu \\ \mu^+ &\rightarrow e^+ + \nu_e + \overline{\nu_\mu}. \end{aligned} \tag{1.4.2}$$

The relative fluxes of ν_μ and ν_e can be predicted to 5% accuracy. Early atmospheric neutrinos saw a deviation from the predicted fluxes, including Kamiokande [23] and IMB [24]. This was a similar effect seen in solar neutrinos and explained similarly by oscillations. Since, atmospheric experiments have provided measurements of oscillation parameters θ_{23} and Δm^2_{32} as well as limits on the mass ordering [25]. The current best values for these parameters are $sin^2\theta_{23} = 0.534 \pm 0.021$ and $\Delta m^2_{32} = (-2.519 \pm 0.033) \times 10^{-3}\ eV^2$ assuming inverted mass ordering and $sin^2\theta_{23} = 0.547 + 0.021 - 0.024$ and $\Delta m^2_{32} = (2.437 + 0.018 - 0.024) \times 10^{-3}$ eV^2 assuming normal mass ordering (normal and inverted ordering are explained in detail in 1.5.2) [10].

Examples of experiments measuring atmospheric neutrino parameters include Soudan 2 [26], MACRO [27] and Super-Kamiokande, which has provided by far the most statistics from atmospheric measurements: [17], [28] and [29]. IceCube experiment has also measured atmospheric neutrino oscillations. Primarily built to detect PeV-scale astrophysical neutrinos, the detector also has neutrino oscillations measurement capabilities [30].

1.4.3 Long-Baseline Experiments

The basic idea behind long-baseline experiments is simple: compare the neutrino spectrum measurements at a detector near the source of the neutrinos to the spectrum measured at a far detector. Equation 1.3.20

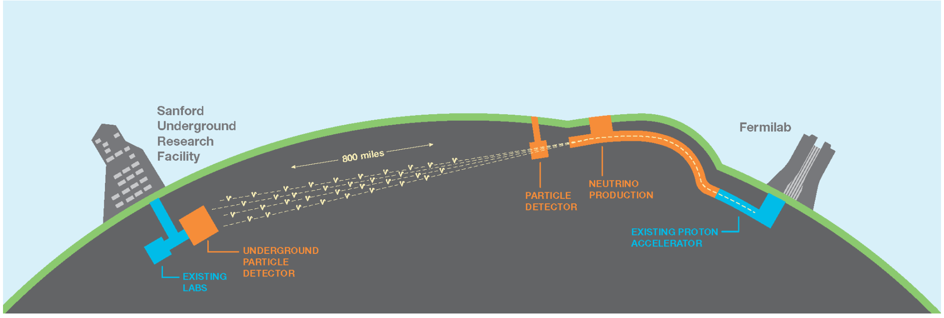

Figure 1.4: The proposed location of the DUNE detectors, retrieved from [32].

shows the relationship between the length of the baseline and the energy of the neutrinos, which is optimised in experiments so that the oscillations are maximal. Two currently running long-baseline experiments are introduced here: T2K and NOvA.

T2K is a long-baseline neutrino experiment with near detector located at JPARC (Japan Proton Accelerator Research Complex) and SK, which was mentioned earlier in section 1.4.1, used as a far detector. NOvA is a long-baseline experiment with a near detector located at Fermilab in Illinois and far detector at Ash River in Minnesota. Both of these experiments are sensitive to oscillation parameters Δm^2_{32} and θ_{23}. Both experiments can also set limits to δ_{CP} and the mass hierarchy. Especially NOvA's long baseline is an advantage when looking at the mass hierarchy using matter effects. More on NOvA and the long-baseline technology will be discussed in chapter 2 [31].

The next generation of long-baseline experiments are aiming to address the unknowns of neutrino physics, namely the δ_{CP} parameter and the mass hierarchy. The Super-Kamiokande experiment is upgrading to Hyper-Kamiokande, which is going to be a larger version of the SK with improved technology. Also, the T2K near detector is being upgraded alongside HyperK to get better oscillation results. Hyper-K is looking to start taking data in 2027, and its key goal is to measure δ_{CP}. The estimates on how long it is going to take to get an accurate measurement depend on the improvement on systematics but the estimates approximate that the non-zero value of δ_{CP} will be discovered in about 5 years [33].

Deep Underground Neutrino Experiment (DUNE), is another future long-baseline neutrino oscillation experiment currently under construction in the USA. The near detector will be located at Fermilab and the far detector at Sanford Underground Research Facility (SURF) in South Dakota. Both of the detector locations are seen in figure 1.4. Based on simulations, DUNE will be able to establish the neutrino mass ordering to a 5σ level in two to three years for all values of δ_{CP}. Determining δ_{CP} is a harder process requiring more data taking, with current estimates stating that to measure it to a level of 5σ can take anything between seven and fifteen years depending on what the exact value is [32] [33].

1.4.4 Reactor Experiments

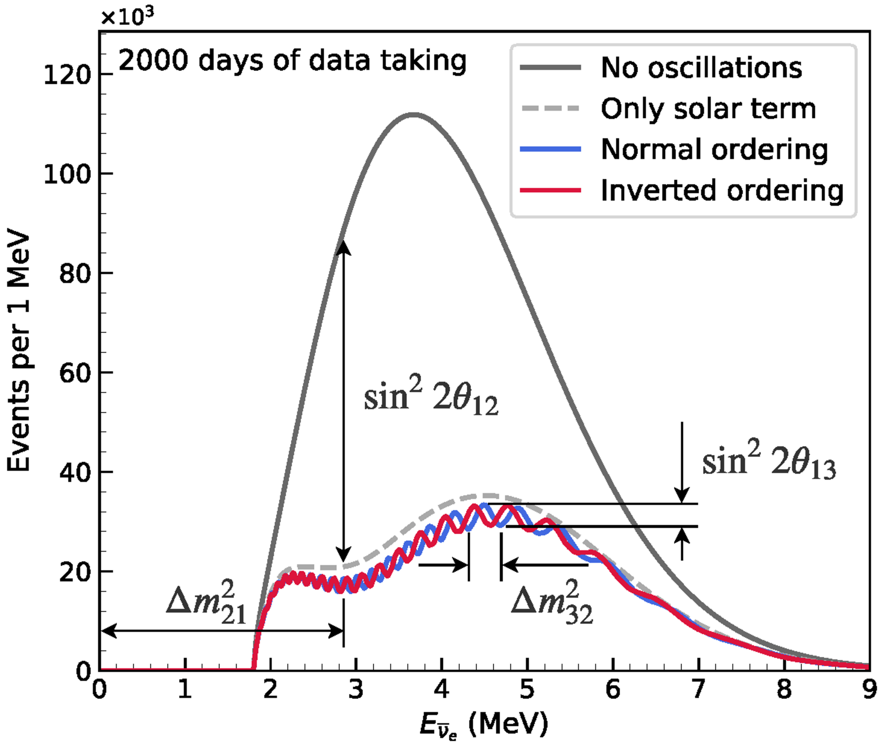

Figure 1.5: Expected oscillated spectra of JUNO experiment for both mass orderings: normal (blue) and inverted (red) as well as unoscillated spectra in black with assumptions of an average baseline of 52.5 km and a data-taking time of 6 years. Retrieved from [34].

Reactor neutrino experiments measure neutrinos coming from nuclear reactors. Nuclear reactors produce $\overline{\nu}_e$s isotropically from β-decays, which can then be measured via inverse β-decay in a detector some distance away from the source and compared to the expected number of $\overline{\nu}_e$s hence measuring the disappearance of $\overline{\nu}_e$. Reactor experiments are sensitive to oscillation angle θ_{13}, Δm^2_{13}, Δm^2_{23} and the mass hierarchy [35].

Reactor neutrinos experiments include RENO [36], DoubleCHOOZ [37], Daya Bay [38] and KamLAND [39]. KamLAND is a liquid scintillator detector at the Kamioka mine in Japan measuring reactor neutrinos from various sources coming from an average of 180 km away from the detector. Due to the fairly long baseline, KamLAND is also sensitive to θ_{12} oscillation angle as well as θ_{13}, Δm^2_{13} and Δm^2_{23}. RENO, Double-

CHOOZ and Daya Bay were all similar experiment with similar baseline lengths all using gadolinium doped liquid scintillator as a detector material measuring θ_{13} and Δm^2_{12}. Double Chooz stopped taking data in 2018, Daya Bay in 2020 and RENO in 2021 [25] [35] [40].

Jiangmen Underground Neutrino Observatory (JUNO) is going to be a liquid scintillator reactor neutrino experiment in China near Jiangmen city. It is bigger and more sensitive to neutrino oscillation parameters than previous reactor experiments. JUNO is going to start data taking in 2024 and it is estimated that JUNO will be able to determine the neutrino mass ordering before the two future long-baseline begin operations [33]. JUNO's expected oscillation spectra is seen in figure 1.5 showing all the neutrino oscillation parameters JUNO will be measuring.

1.5 CP Violation and Neutrino Mass

1.5.1 CP Violation

CP symmetry is a combined expression for two symmetries in the Universe: charge conjugation and parity. Charge conjugation switches particles for antiparticles and parity mirrors the spatial coordinates of a particle when changing to its antiparticle. Thus, these two symmetries together ensure that the laws of physics are the same when interchanging a particle to its antiparticle. When this symmetry is broken the process is called CP violation (CPV) [41].

The violation of the charge conjugation and parity symmetry was discovered in kaon decays in the 1960s [42]. Since the discovery, CPV has been well established in the quark sector and is described in the CKM matrix. One can thus make an assumption that a similar effect is present in the lepton sector, however no evidence for CPV in the lepton sector has been found. In the PMNS matrix (equation 1.3.2), one of the variables is δ_{CP}. If $\delta_{CP} \neq 0, \pi$ (and assuming no U parameter is zero) that means PMNS matrix has a complex phase and the CP-invariance is present in neutrino oscillations. In practice, this means that the oscillation probabilities between two flavours are not the same for particles and anti-particles, i.e. $P(\nu_e \rightarrow \nu_\mu \neq P(\overline{\nu}_e \rightarrow \overline{\nu}_\mu))$ [43].

Defining δ_{CP} is central for providing an explanation to the matter - antimatter imbalance in the Universe through leptogenesis. Leptogenesis, if proven, would have resulted in the imbalance between leptons and antileptons in the early universe. All evidence points to this imbalance since no significant source of antimatter has been found. When coming in contact with ordinary matter, antimatter and matter annihilate but no substantial signal of this happening has been found [43] [44] [45].

Eventually, leptogenesis results in an imbalance in baryon/antibaryon number via partial conversion of leptons to baroyns on the electroweak scale. Baryogenesis in the Universe is one of the Sakharov conditions for creating a matter-antimatter imbalance in the Universe [46]. Thus the combination of CPV in neutrinos leading to leptogenesis and out of thermal equilibrium interactions in the early Universe could provide an answer to the matter/antimatter imbalance [43] [44] [45].

1.5.2 Neutrino Mass

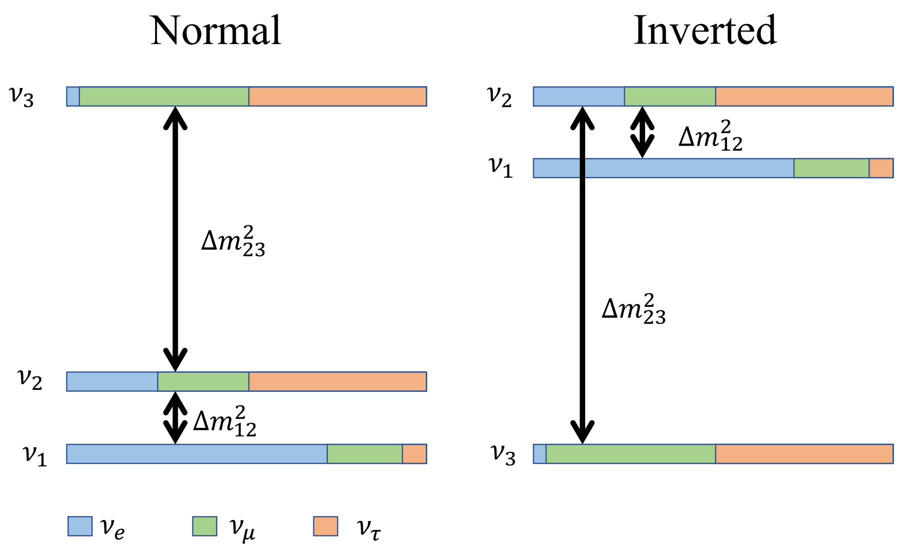

Figure 1.6: Illustration of the neutrino mass hierarchy.

As well as the existence of CP violation, there is another unanswered question in neutrino physics: the nature of neutrino mass. Since the existence of neutrino oscillations is proven, at least two of the mass states have to be non-zero. Neutrino oscillation experiments only measure the differences in neutrino masses but not the absolute mass. There are experimental measurements for both of the differences (Δm_{12} and Δm_{23}), but the sign is only known for Δm_{12} [47]. This means that the order of the masses of neutrino states v_1, v_2 and v_3 is not known, and this is known as the mass hierarchy problem, illustrated in figure 1.6. A model where v_3 is the biggest mass is called Normal Hierarchy (NH) (left side of figure 1.6) and other case where v_3 is the smallest is named Inverted Hierarchy (IH) (right side of figure 1.6). No experiment so far has managed to measure to prove it either way, but a lot of ongoing research is taking place [48] [49].

The question whether particles are their own anti-particles is relevant for neutrinos, since they are the only known fermions which are neutral with respect to both electric and colour charge. All current matter particles (fermions) described in the SM are Dirac particles, meaning they are not their own anti-particles and acquire their mass through interactions with the Higgs boson. The mass term in the SM is of form $m\overline{\psi}_L\psi_R + h.c$ with L and R annotating the left and right handed components (and h.c the hermitian conjugate of the term). This "handness", also referred to as chirality, is an intrinsic quantum mechanical property of a particle describing the mirror symmetry of the particles. In the SM, only left-handed particles (and right-handed anti-particles) take part in CC weak interactions. Since those interactions are the only way neutrinos are created, all neutrinos are left-handed and all anti-neutrinos are right-handed. It is hence not possible for a neutrino to acquire mass in the same way other particles due to the lack of right-handed neutrino component [9] [45] [48] [49] [50].

In contrast, according to Majorana's theory particles and antiparticles are indistinguishable from each other. One of the most researched theory for neutrino mass is the seesaw mechanism, which requires neutrinos to be Majorana in nature. The theory combines Majorana and Dirac mass terms, and requires there to exist a heavy partner to a neutrino, which is a Majorana particle so massive that it cannot have been seen in experiments. It gives neutrinos non-zero mass and also provides an explanation to the smallness of neutrino masses compared to other fermions in the SM [9] [48] [50].

1.5.3 Matter Effects

When neutrinos travel through matter, their oscillations are modified by the surrounding material. The effect was first considered by L. Wolfestein [51] and named matter effects, sometimes also called Mikheyev-Smirnov-Wolfestein (MSW) effect. This is due to the excess of electrons compared to muons or taus in matter (such as the Earth or the Sun), which causes coherent forward scattering of electrons. The Feynman diagrams for this scattering for both CC and NC are seen on figure 1.7. This effect is crucial in measuring the signs of the mass differences, and the matter effects in the Sun was central when measuring the sign of Δm_{12} [47].

The matter effect increases the oscillation probability for NH and decreases it for IH. The matter effect is also opposite for ν_e and $\overline{\nu}_e$ and can thus be confused with the CP-violating effect [49] [54]. Mathematically, the matter electrons add a potential term [52]:

$$V_e = \pm\sqrt{2}G_F N_e, \tag{1.5.1}$$

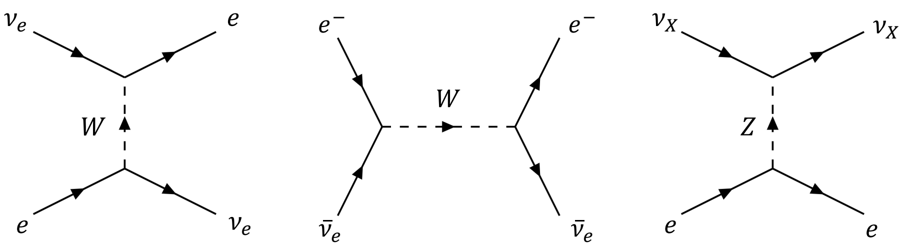

Figure 1.7: Coherent scattering of neutrinos for CC (left and middle) and NC (right) interactions. X in NC plots indicates that the interactions happens equally often for all three neutrino flavours. Reproduced from [52].

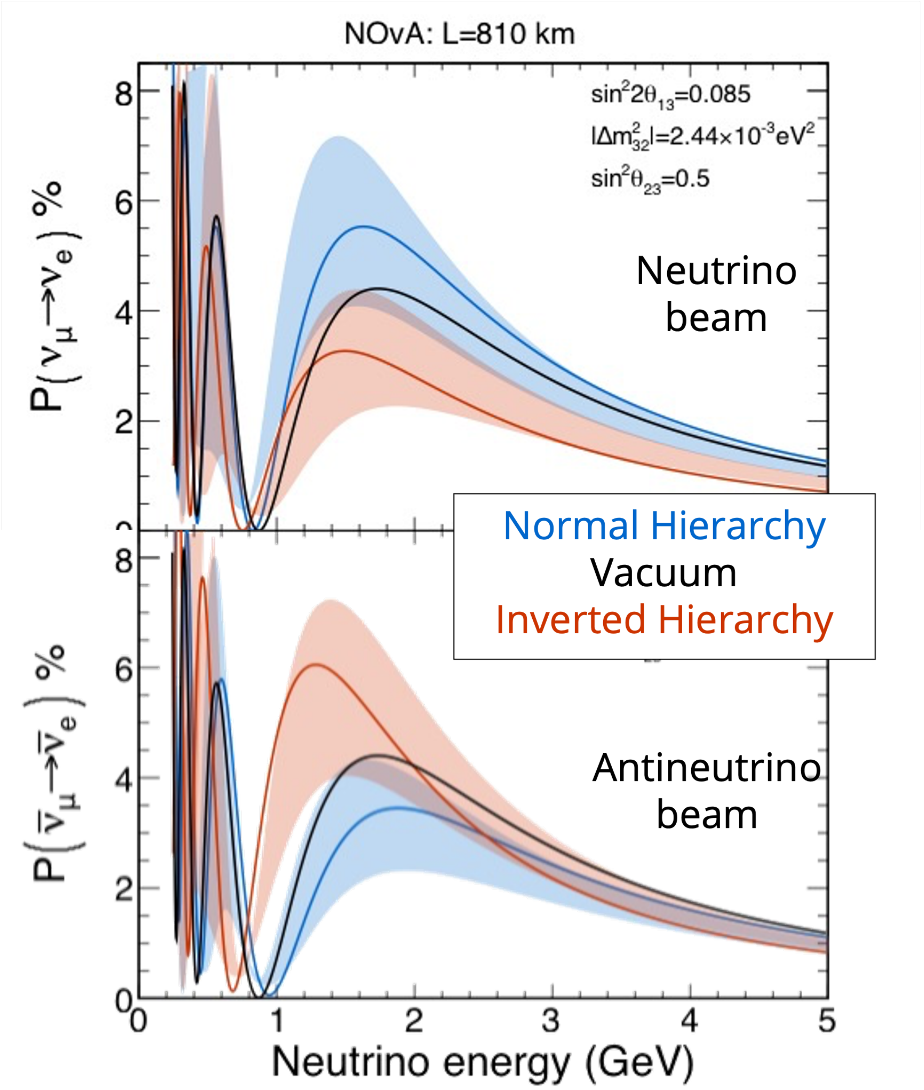

Figure 1.8: Illustration of matter effects for oscillation probability in the NOvA experiment. Black line describes the probability for oscillations in vacuum, blue line in matter for normal hierarchy and orange for inverted hierarchy assuming $\delta_{CP} = 0$. Upper plot describes the probability for neutrinos and lower plot for anti-neutrinos. The error bans in blue and orange illustrate the allowed range for δ_{CP} values. Retrieved from [53].

where G_F is the Fermi constant and N_e the electron number density. The sign depends on whether neutrinos or anti-neutrinos are considered [52] [55].

This then leads to a modification for the oscillation probability as:

$$\begin{aligned} P(\nu_\mu \to \nu_e) = \sin^2(\theta_{23})\sin^2(2\theta_{13})&\frac{\sin^2(\Delta(1-A)}{(1-A)^2)} \\ +\alpha\tilde{J}\cos(\Delta \pm \delta_{CP})&\frac{\sin(\Delta A)}{A}\frac{\sin(\Delta(1-A))}{1-A} \\ +\alpha^2\cos^2(\theta_{23})\sin^2(2\theta_{12})&\frac{\sin^2(\Delta A)}{a^2}, \end{aligned} \tag{1.5.2}$$

where $A = \dfrac{2\sqrt{2}G_F N_e E}{\Delta m^2_{31}}$, $\tilde{J} = \cos(\theta_{13})\sin(2\theta_{13})\sin(2\theta_{12})\sin(2\theta_{23})$, $\Delta = \frac{\Delta m^2_{31} L}{4E}$ and $\alpha = \frac{\Delta m^2_{21}}{\Delta m^2_{32}}$. The sign is set as positive for neutrinos and negative for antineutrinos [55] [56].

Current efforts to measure the sign of Δm_{23} and to solve the mass hierarchy problem are focusing on measuring this difference in matter effects in the Earth in long-baseline and reactor neutrino experiments. An example of how matter effects oscillation probabilities in the NOvA experiment is seen in figure 1.8 for both normal and inverted hierarchy for both neutrinos (first plot) and anti-neutrinos (second plot).

Chapter 2

NOvA

2.1 Introduction

NuMI Off-Axis ν_e Appearance Experiment, NOvA, is a long-baseline neutrino oscillation experiment devoted to studying neutrino oscillations. It consists of two functionally identical, finely granulated tracking calorimeter detectors. The NuMI facility at Fermilab in Illinois creates a neutrino beam, which is aimed at the near detector (ND) located 1 km from the beam target approximately 100 m underground. The neutrinos then travel 810 km to the far detector (FD) placed on the surface at Ash River, Minnesota near International Falls under a rock bed with a background cosmic ray flux of about 130 kHz. Both NOvA detectors are placed 14.6 $mrad$ off the beam axis to help to obtain a narrow energy peak at 1.8 GeV, where the oscillations are maximal. The detectors are functionally identical with the same design but differ only in size to reduce systematic errors and achieve more accurate results [54] [57].

2.2 NuMI Beam

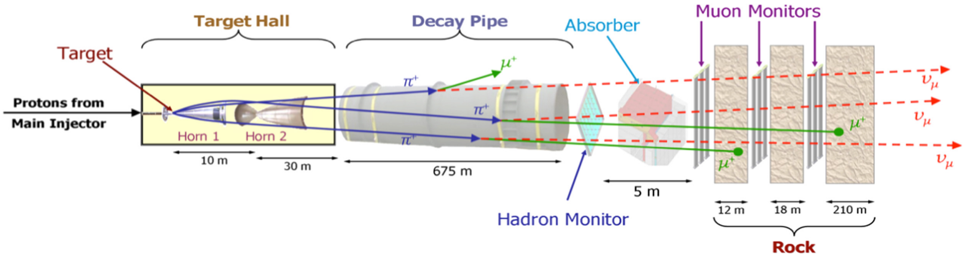

Figure 2.1: Schematic of the NuMI beamline components, not to scale. Protons from the main injector strike a target, then the created particles are focused onto a decay pipe with two magnetic horns. At the end of the pipe, the remaining particles (mostly muons) are absorbed leaving a beam of neutrinos or anti-neutrinos. Retrieved from [57].

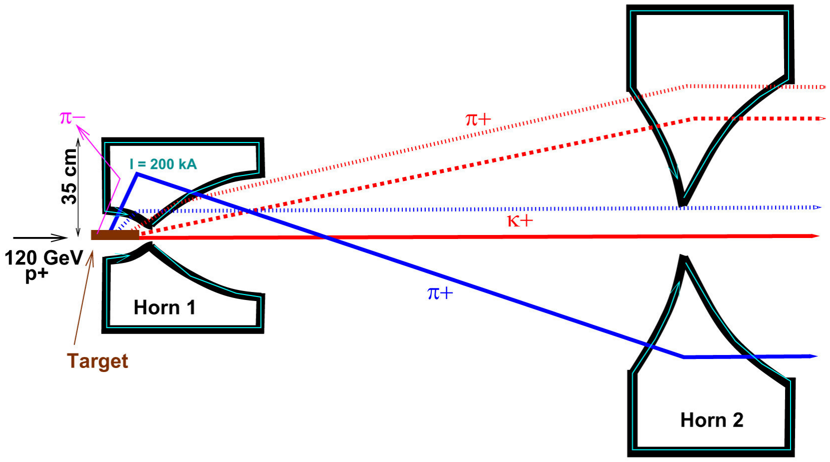

Figure 2.2: NuMI horn focusing effect for hadrons arriving in different angles. The solid red line is for a particle passing straight through. The red dashed line is where horn 2 focus the particles. The red dotted line describes particles that are underfocused by horn 1 and corrected by 2. Solid blue line is particles that are overfocused by both first horn and corrected by the second one, and blue dotted line is for particles focused by horn 1. Retrieved from [57].

The Neutrinos at the Main Injector (NuMI) beam is located at Fermilab and provides neutrinos for multiple research uses. A schematic of the beamline components is shown in figure 2.1. The beam is formed by first striking 120 GeV protons from the Fermilab Main Injector onto a graphite target of approximately 1 m in length. The interactions in the target produce pions and kaons, which are in turn focused into a 675 m decay pipe with two magnetic horns where they decay into muons and neutrinos [57]:

$$\begin{aligned} \pi^+ &\to \mu^+ + \nu_\mu \\ \pi^- &\to \mu^- + \overline{\nu_\mu}. \end{aligned} \tag{2.2.1}$$

$$\begin{aligned} K^+ &\to \mu^+ + \nu_\mu \\ K^- &\to \mu^- + \overline{\nu_\mu}. \end{aligned} \tag{2.2.2}$$

The two 3-metre-long magnetic horns keep the charged hadrons in the pipe and focuses them into a beam. This focusing effect is illustrated in figure 2.2. All the hadrons are produced in differing angles, which is shown by different line styles in the figure. Those along the beam axis pass the horns unaffected. At some angles, hadrons are well focused by the first horn and sometimes the first horn over- or underfocuses, in which case the second horn corrects the issue. The horns can also select the sign of the neutrino beam by reversing the power supply current. When neutrinos are selected, the horns are said to be in forward horn current (FHC) and when anti-neutrinos are selected, the horns are in

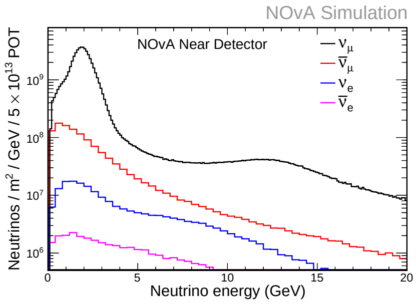

Figure 2.3: The composition of the NuMI beam at the NOvA near detector for neutrino (FHC) mode, produced using NOvA simulation. Retrieved from [59].

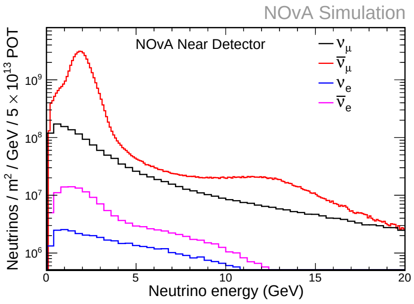

Figure 2.4: The composition of the NuMI beam at the NOvA near detector for anti-neutrino (RHC) mode, produced using NOvA simulation. Retrieved from [59].

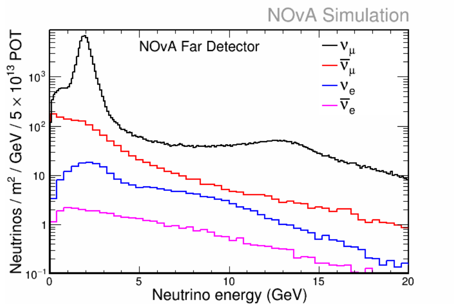

Figure 2.5: The composition of the NuMI beam at the NOvA far detector for neutrino (FHC) mode, produced using NOvA simulation. Retrieved from [59].

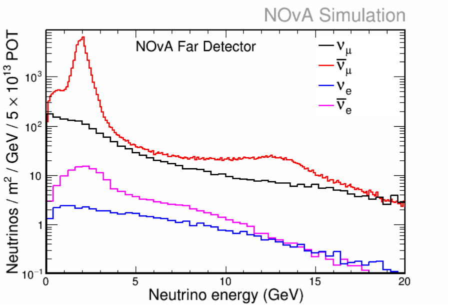

Figure 2.6: The composition of the NuMI beam at the NOvA far detector for anti-neutrino (RHC) mode, produced using NOvA simulation. Retrieved from [59].

reverse horn current (RHC).

Most of the selected hadrons decay according to equation 2.2.1 and equation 2.2.2 in the decay pipe, but there is also a hadron monitor and an absorber to remove of any residual hadrons after the decay pipe [57] [58].

At the end of the decay pipe, the beam made of muons and neutrinos is lead to a series of muon monitors in between large sections of rock to absorb the muons leaving only neutrinos (plus a small percentage of other particle contaminants). The created neutrino beam is highly pure with 96 percent ν_μ for FHC mode and 83 percent $\overline{\nu}_\mu$ for RHC mode. The difference between the purity and the width of the beam stems from the difference in cross-sections for ν_μ and $\overline{\nu}_\mu$. The simulated composition of the beam at the NOvA ND is shown in figures 2.3 and 2.4 for FHC and RHC respectively. Similarly, for FD the compositions are shown in figures 2.5 and 2.6. The narrow energy peak for the neutrinos is due to NOvA detectors being placed off-axis of the NuMI beam, explained more in the next section 2.2.1. It is evident that the majority of the contamination is from wrong sign neutrinos ($\overline{\nu}_\mu$ for FHC and ν_μ for RHC). There is also a small background contribution from electron neutrinos and anti-neutrinos, which result from decays of K^0 produced in small numbers, as well as decays of tertiary muons. Charged kaons also have an electron decay mode but it is highly suppressed compared to the muon decay mode. It is, however, responsible for some of the electron neutrino background contribution [12] [54] [57] [60].

2.2.1 Off-Axis Concept

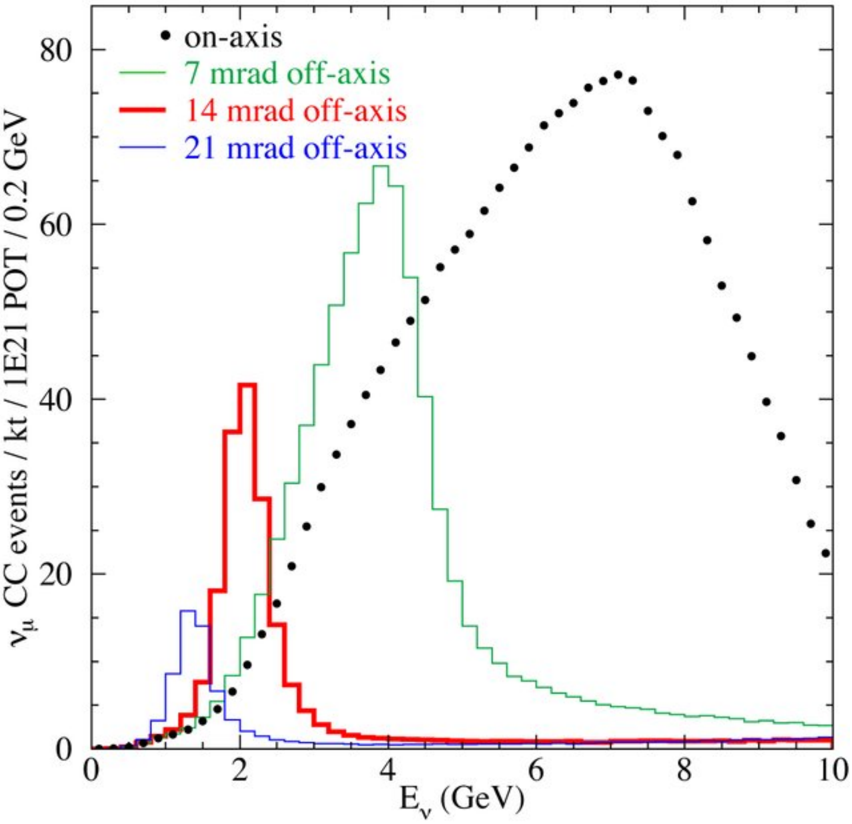

Figure 2.7: The off-axis effect in the NuMI beam for different axis angles with neutrino energy on the x-axis and a (scaled) number of CC muon neutrino events on the y-axis. Retrieved from [58].

The NuMI beam can be configured to produce a beam ranging from peaking at 3 GeV to 15 GeV. However, The NOvA experiment was designed to be off-axis leading to a narrow energy peak at 1.8 GeV, where the oscillations are maximal. As well as maximal neutrino mixing, this narrow energy spectrum helps with background rejection. Some NC events, which are treated as background in the NOvA analysis, produce showers that look very similar to electron showers and they can be mistaken for ν_e CC events. In NC events there is no outgoing charged lepton so the outgoing neutrino tends to carry away a large portion of the energy, hence there are fewer NC events at lower energies [12] [54].

The concept for off-axis can be explained by considering pions in their rest frame producing mono-energetic and isotropic neutrinos. The 4-momentums for a pion, a muon and a neutrino in the interaction can be written as follows [56]:

$$\begin{aligned} p_{\pi^+} &= (E_{\pi^+}, \overline{0}) \\ p_{\mu^+} &= (E_{\mu^+}, \overline{P}_{\mu^+}) \\ p_{\nu_\mu} &= (E_{\nu_\mu}, \overline{P}_{\nu_\mu}) \end{aligned} \tag{2.2.3}$$

One can then write the momentum of the muon as:

$$p_{\mu^+}^2 = p_{\pi^+}^2 + p_{\nu_\mu}^2 - 2p_{\pi^+}^2 p_{\nu_\mu}^2. \tag{2.2.4}$$

Using identities $p_{\mu^+}^2 = p_{\pi^+}^2 - p_{\nu_\mu}^2$ and $E^2 = m^2 + p^2$:

$$m_{\mu^+}^2 = m_{\pi^+}^2 - 2E_{\pi^+}E_{\nu_{\mu^+}} \tag{2.2.5}$$

and so the energy of the neutrino in the rest frame is:

$$E_{\nu_\mu}^{Rest} = \frac{m_{\pi^+}{}^2 - m_{\mu^+}{}^2}{2E_{\pi^+}} \tag{2.2.6}$$

Boosting from the rest frame to lab frame changes the energy such that:

$$E_{\nu_\mu}^{Rest} = \gamma E_{\nu_\mu}^{Lab}(1 - \beta). \tag{2.2.7}$$

Rearranging:

$$E_{\nu_\mu} = E_{\pi^+}(1 - \frac{m_{\mu^\nu}}{m_{\pi^+}})\frac{1}{1 + \theta^2\gamma^2} \tag{2.2.8}$$

Here θ is the angle between pion and neutrino directions and γ the Lorentz factor, i.e. $\gamma = \frac{E_\pi}{m_\pi}$. Since the masses of the particles are unchangeable, the energy of neutrinos as well as their flux decreases as the θ increases. The width of the neutrino energy spectrum also decreases as it is less dependant on the energy of the pion [52] [56].

2.3 NOvA Detectors

2.3.1 The NOvA Cell

The basic constituent unit of both NOvA detectors is a rectangular PVC plastic cell that contains liquid scintillator and a wavelength shifting fiber looping around the cell, seen on the right in figure 2.8. Charged particles

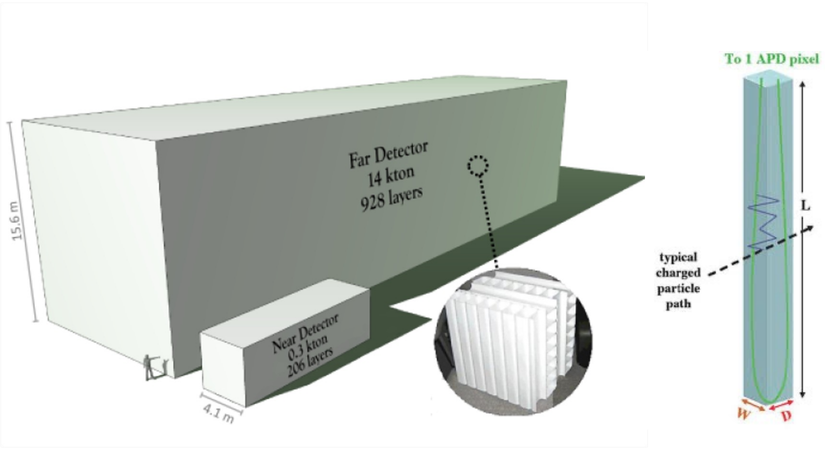

Figure 2.8: A schematic showing the scale between the two NOvA detectors. The zoomed-in part illustrates the placement of the two alternating orientations of the cell planes. On the right is pictured NOvA's basic building block, cell. Each cell has a wave length sifting fibre looping around and is connected to an APD to record the signal from the fibre. Retrieved from [61].

produce scintillator light in the cell, which is light emitted when material is exited by high energy particles. The light bounces around the cell until captured by the fibre or absorbed by the plastic or the scintillator. Each cell has an internal width of 3.8 *cm* in the beam direction (W), internal depth of 5.9 *cm* along beam direction (D) and an internal length of 15.5 *m* (L) in the far detector and 3.9 *m* in the near detector. The PVC walls of the cell are about 2.0 to 4.5 *mm* thick (see figure 2.8) [52] [54].

A set of 16 of cells are joined together to form an extrusion, two of which are then connected together to form a module. These planes of cells are placed alternating vertical and horizontal directions (see the zoomed in insert of figure 2.8) to get two orthogonal views of particle interactions in the detector to enable reconstruction of full 3D particle tracks [62] [63].

2.3.2 Liquid Scintillator

The components of the liquid scintillator are summarised in table 2.1. Most of the scintillator consists of mineral oil and pseudocumene [1,2,4 - Trimethybenzene] with trace concentrations of wavelength shifting fluors (PPO [2,5 - diphenyloxazone] and bis-MSB [1,4 - de(methylstyryl) benzene]). Pseudocumene is a scintillator that produces light which peaks at 360-390 *nm* and with the wavelength shifting fluors, the peak lighting wavelength is brought to 400-450 *nm* to match with the absorption spectrum of the fiber used. The resulting composition of the NOvA detectors

Component	**Purpose**	**Mass Fraction**
Mineral oil	solvent	95%
Pseudocumene	scintillant	4.1%
PPO	waveshifter	0.091%
bis-MSB	waveshifter	0.0031%
topocherol	antioxidant	0.0010%
Stadis-425	antistatic agent	0.0003%

Table 2.1: The composition of the NOvA liquid scintillator, values from [54].

is about 70% PVC and 30% scintillator [54].

2.3.3 Wavelength Shifting Fibre

Each cell in NOvA contains about 33.5 meters of 0.77 *mm* diameter wavelength shifting fiber with a core of polystyrene mixed with R27 dye to attain the wavelength shifting properties. The fibre captures the blue light produced by the scintillator (400-450 *nm*) and shifts it to a range of 490-550 *nm*. When the light attenuates in the fibre, the red light at about 520-550 *nm* survives and is recorded by the avalanche photo diodes [54].

2.3.4 Avalanche Photo Diodes (APDs)

At the top of the cell the fibre is connected to an avalanche photodiode (APD), seen in figure 2.9, temperature monitoring and a thermoelectric cooler that keep the temperature of the APD at about -15 degrees celsius to reduce noise. The APD is also connected to a dry gas system to remove moisture. Each of these APDs has a quantum efficiency of 85% at the 520-550 *nm* light that it is collecting [54].

In the APD, the light that is collected generates electron-hole pairs which propagate at the p-n junction in a 425 V reverse bias electric field. The electrons drift past the junction, where an avalanche effect multiplying electrons takes place. At each APD, the temperature is set and monitored carefully to keep a consistent amplification across all of the detector [54] [64].

2.3.5 NOvA Data Acquisition System

The main task for the data acquisition system (DAQ) is to record the data by concentrating it from the huge number of APDs into a single stream for analysis and storage.

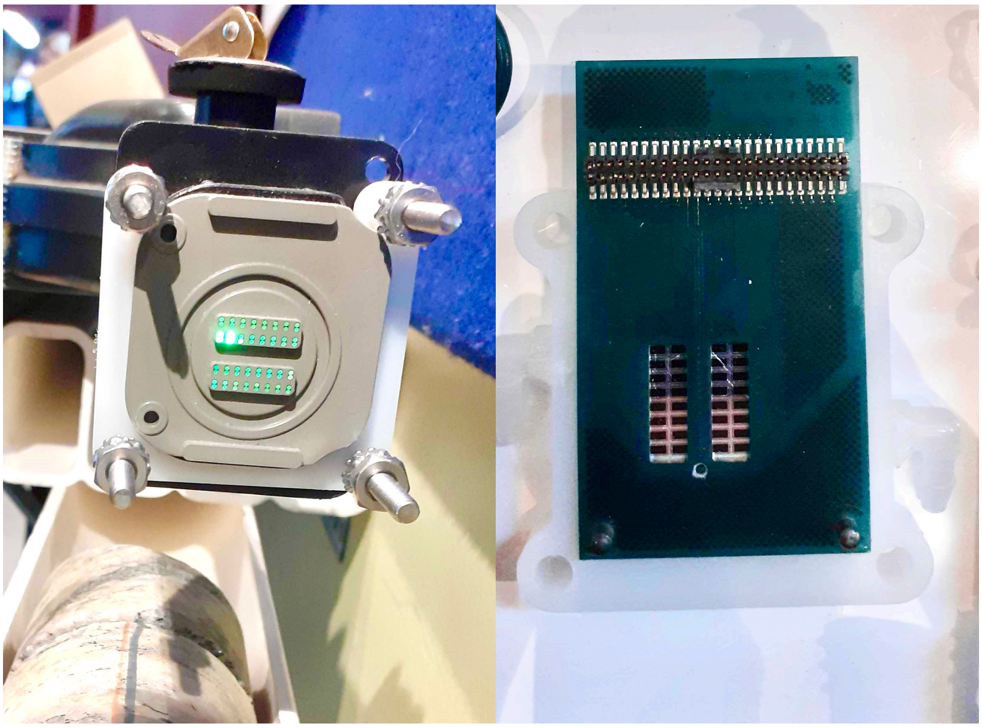

Figure 2.9: NOvA APD: the end of the scintillator cells where the fibres mount to the APD (left) and the front of the APD (right). Photo courtesy of Kevin Mulder.

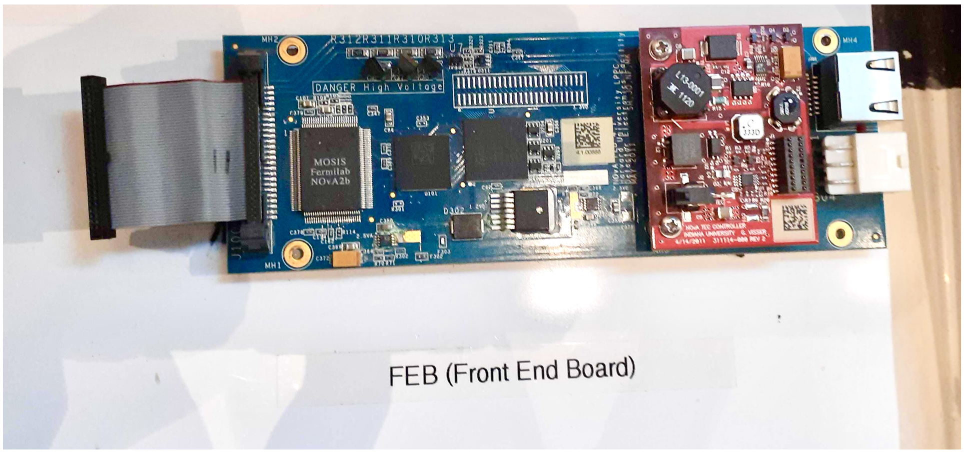

Figure 2.10: NOvA FEB up close. Photo courtesy of Kevin Mulder.

After the signal is collected at the APD, each APD is connected to a front-end board (FEB), pictured in figure 2.10, which digitises the signal from light to an electronic signal as well as separates and time tags the signal. The data is delivered to data concentrator modules (DCMs) and organised into 50 μs time blocks (micro-slices). In the DCM, the data is then organised into 5 ms intervals (milli-slices) which are routed to buffer nodes. All the DCMs send the data from across all the regions of the detector in the same pattern to the same DCM in each milli-slice to give each buffer node a full picture of the detector. In the buffer nodes, the events are built based on time tags and archived for storage. They

Figure 2.11: View from the top of the far detector, showing FEBs connecting to DCMs (grey boxes with yellow lights in the middle).

are also capable of performing real time analysis using a data driven trigger. A photo showing the FEBs connecting to DCMs on top of the far detector is seen in figure 2.11 [54] [64] [65].

The trigger in NOvA can be divided into two categories: clock based and data driven. The clock based ones get information about the NuMI beam spill, and select a time window when a beam spill occurs in Fermilab. This is primarily used to get rid of cosmic rays in the far detectors. NOvA also has data driven triggers, which looks at topologies in the data stream and make selections based on it. An example trigger is the supernova trigger that is looking for intense spills of neutrinos not from the NuMI beam [52] [65].

2.3.6 Near Detector

NOvA near detector, seen in figure 2.12, is located 105 meters underground 1,015 meters away from the NuMI beam target hall. The detector is 14.5 metres long, 4.1 metres tall and 2.8 metres wide weighing approximately 300 tons. The detector consists of 20,192 cells grouped in planes alternating in horizontal and vertical directions. The timing resolution of the near detector is 5 ns. An example event display with a beam spill is seen in figure 2.13, showing multiple potential neutrino events [40] [52] [67].

The first 12.5 meters are the actual detector volume and the last 2 meters form a muon catcher. The purpose of the catcher is to tag muons from ν_μ CC events and increase the quality of the muon energy reconstruction. It is made of two PVC module planes (one in both vertical

Figure 2.12: NOvA near detector at Fermilab, Illinois. Retrieved from [66].

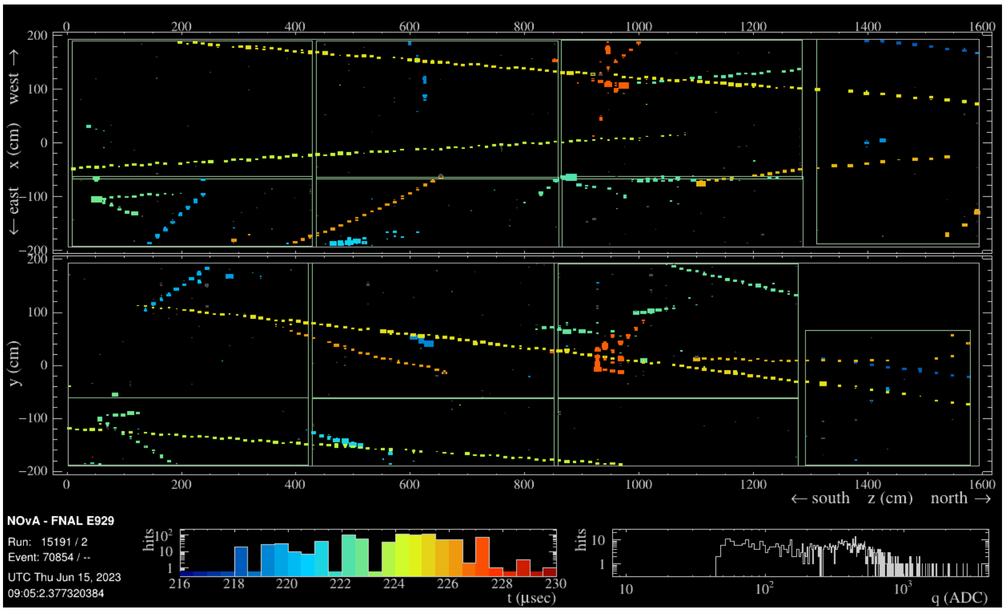

Figure 2.13: An event display of the ND during a NuMI beam spill.

and horizontal direction) interspersed with 10 *cm* thick steel planes [54] [63].

2.3.7 Far Detector

Figure 2.14: NOvA far detector at Ash River, Minnesota. Retrieved from [68].

The far detector, seen in figure 2.14, is located at Ash River approximately 810 *km* away from the near detector. It has the same design as the near detector albeit much bigger, weighing around 14 kilotons. The far detector has a width of 15.6 *m*, height of 15.6 *m* and length of 60 *m* in the beam direction [54] [63] [69].

The far detector is located on the surface but has a 122 *cm* concrete and 15 *cm* barite shielding on top to reduce the amount of recorded cosmic rays in the detector. Also, an accurate timing for the readout electronics is essential to decrease the amount of cosmogenic background, since the NuMI beam operates in 10 μs spills. The timing resolution of the FD is slightly higher than that of the near detector at 15 *ns*. This is due to FEBs sampling the APDs at four times the rate at the ND compared to FD. This is done because ND gets a much higher flux of neutrinos near the source [52] [54] [63] [69].

An event display including one 550 μs time window in the FD including all cosmic events is seen in 2.15. Figure 2.16 shows a neutrino event

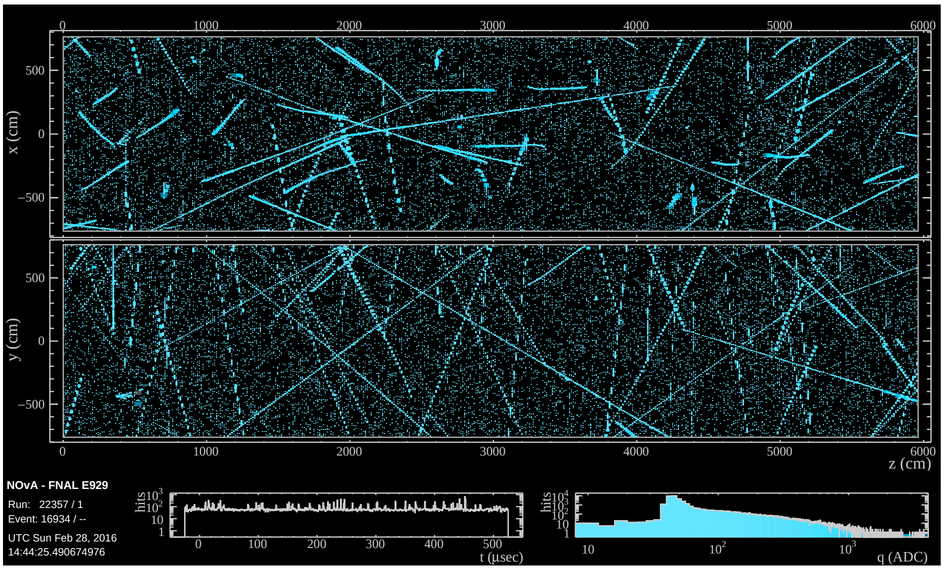

Figure 2.15: An event display of the FD with all cosmic events included in a 550 μs time window. Retrieved from [70].

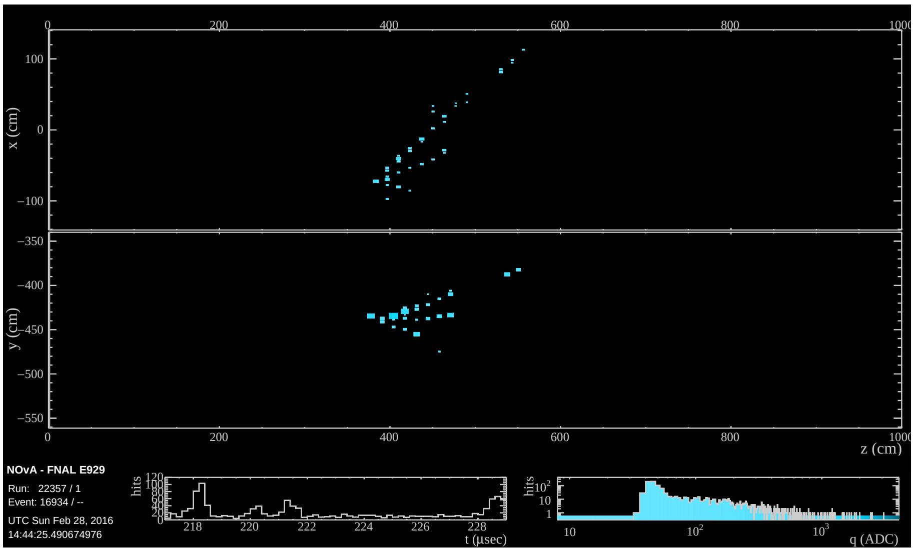

Figure 2.16: A neutrino candidate event from the previous event display in figure 2.15 after cosmic filtering has been applied. Retrieved from [70].

candidate after removing cosmics, ready for reconstruction (explained in the next chapter in section 3.2).

2.3.8 Neutrino Interactions in NOvA Detectors

The different interactions can be distinguished from each other in NOvA depending on the signatures the final state particles leave in the detectors. Examples of the three types of interactions measured in NOvA are shown in figure 2.17. ν_μ CC interactions (top of the figure 2.17) are distinguished by the long track created by a muon alongside a shorter track created by the final state nucleon, which deposits its energy much quicker compared to the electron. ν_e interactions (middle of the figure 2.17) are

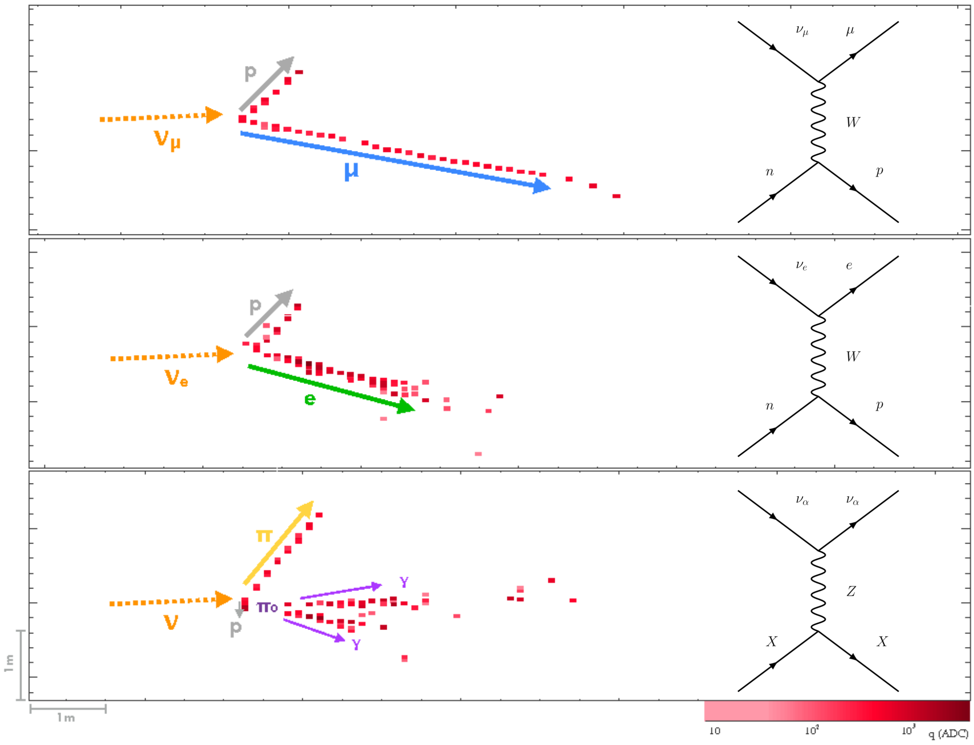

Figure 2.17: Examples of the three neutrino interactions in NOvA: ν_μ CC, ν_e CC and NC, showing the different signals left in the detectors. Retrieved from [71].

differentiated from ν_μ by the less defined and shorter electron track. This is due to the fact that electrons interact in the detector electromagnetically more than muons because of their lower mass. The neutral current interactions (bottom of the figure 2.17) cannot distinguish between neutrino flavours, and they are differentiated from CC interactions by the characteristic hadronic shower and a single pion track instead of two long tracks created by ν_μ CC event.

NOvA also has few ν_τ events, although the threshold for the events is 3.4 GeV which is on the upper end of the NOvA energy scale. τ is very short lived so only its decay products like pions, electrons and muons are visible in the detector [72]. For most NOvA analysis, including this thesis, ν_τ events are not relevant.

2.4 NOvA Oscillation Channels

There are two main oscillation channels in NOvA: electron neutrino appearance and muon neutrino disappearance. First, focusing on the disappearance, one can write down a simplified version for the oscillation probability for this channel. θ_{13} is known to be small and Δm_{12} is much smaller than Δm_{23}, so a 2-flavour approximation for this oscillation channel can thus be used (analogous to equation 1.3.18) [52]:

$$P_{\nu_\mu -> \nu_\mu} = 1 - \sin^2(2\theta_{23}) \sin^2(\frac{1.27\Delta m^2_{32} L}{E}). \tag{2.4.1}$$

Hence, NOvA can provide measurements of the oscillation parameters $\sin^2(2\theta_{23})$ and Δm^2_{32} using disappearance analysis alone. However, NOvA's sensitivity to ν_e appearance gives opportunities to also measure the mass hierarchy via matter effects, described in section 1.5.3, and CPV in neutrino oscillations, as described in section 1.5.1. The oscillation probability for the appearance can be written as [73]:

$$P_{\nu_\mu -> \nu_e} = P_{atm} + P_{sol} + \sqrt{P_{atm}}\sqrt{P_{sol}} cos(\Delta_{23} + \delta_{CP}), \tag{2.4.2}$$

where

$$\Delta_{ij} = \frac{1.27\Delta m_{ij} L}{E} \tag{2.4.3}$$

with $i, j = (1, 2, 3)$. The subscripts *atm* and *sol* correspond to atmospheric and solar, which refer to the experiments types that these probabilities are the dominant components in. The expressions for these probabilities are [73]:

$$P_{atm} = \sin(\theta_{23}) \sin(2\theta_{13}) \sin\left(\frac{\sin(\Delta_{31} - aL)}{\Delta_{13} - aL}\right) \tag{2.4.4}$$

and

$$P_{sol} = \cos(\theta_{23}) \sin(2\theta_{12}) \sin\left(\frac{\sin(aL)}{aL}\right)\Delta_{12}, \tag{2.4.5}$$

where $a = G_F N_e \sqrt{2}$, analogues to equation 1.5.1 and stems from matter effects. δ_{CP} is the CP-violation phase, which in NOvA can be measured comparing the oscillations for ν_e and $\overline{\nu}_e$ in this oscillation channel [52] [69].

NOvA can also measure the octant of θ_{23}, and it is considered that $\theta_{23} < \frac{\pi}{4}$ is the lower octant and $\theta_{23} > \frac{\pi}{4}$ the upper octant. Unfortunately, the disappearance analysis is invariant under $\theta_{23} \to \frac{2}{\pi} - \theta_{23}$ so it is not sensitivity to the octant of the θ_{23} but only the value of $\sin(\theta_{23})$. Hence, the constraint measurements for the octant is made for the appearance channel [52].

Chapter 3

NOvA 3-Flavour Analysis

The latest NOvA 3-flavour neutrino oscillation results were produced in 2020 and presented in [12]. The data used included a 50% beam exposure increase with improved analysis techniques and better simulation compared to earlier oscillation results from NOvA. The data included data taken with a neutrino beam from February 6th 2014 to March 20th 2020 with exposure of 13.6×10^{20} protons on target (POT), combined with antineutrino beam data taken from June 29th 2016 to February 26th 2019 with exposure 12.5×10^{20} POT. The average power of the proton source was 650 kW with a peak hourly-averaged power of 756 kW. The analysis included all NOvA neutrino channels producing a joint fit to neutrino oscillation parameters of $\nu_\mu \to \nu_e$, $\nu_\mu \to \nu_\mu$ as well as the corresponding anti-neutrino channels [12] [74]. This chapter explains the analysis methods used as well as highlights the latest results.

3.1 Simulation

NOvA uses a Monte-Carlo (MC) simulation to describe the NuMI beam, neutrino interactions and NOvA detectors. The simulation is analogues to the data extracted from the detectors. Accurate simulation is essential in the analysis, it is used for the development and testing of analysis methods and tools, estimating systematics and predicting the distribution of events in the detectors [52] [62].

3.1.1 Beam Simulation

The neutrino flux from the NuMI beam is predicted using a GEANT4 [75] based simulation named G4NuMI. The full production of hadrons from protons in the graphite target, decays of these hadrons to muons and neutrinos and muon absorption in the rock are run through the simulation. An accurate description of targets, horns and the decay pipe are all included in the simulation [52] [67].

The results are re-weighted using PPFX (package to predict the flux) to predict and estimate the uncertainties from the beam flux. This was first developed by the MINERvA collaboration [76]. The corrections are based on the hadronic spectrum attained from hadron production data that includes details of all the different interactions leading to neutrino production [52] [62] [67].

3.1.2 Event Generation

Neutrino interactions are simulated using a custom version of GENIE software [77] and tuned using NOvA ND data. The simulation uses nuclear physics models to predict the neutrino interactions and cross-sections in the detectors and surrounding material. It predicts the type of interaction, kinematics and outgoing particles and a full topology for each simulated event [12].

3.1.3 Detector Simulation

The detector simulation is using both GEANT4 [75] and NOvA specific software modules. Particle propagation and energy deposition in the materials is simulated using GEANT4, similar to beam simulation. Scintillator light production and transport in the cell, and conversion to electronic signal is done using NOvA specific software modules. The algorithm takes into account factors such as scintillator response in NOvA detectors, PVC reflectivity and absorption of the fibre to calculate how the light reaches the photodiode. The electronic readout simulation models how the signal is transported from raw signal to digitised data. It takes into account the response in the APDs by calculating noise simulation and a triggering firmware effect on data. Cosmic rays are included alongside beam neutrinos in the simulation [12] [52] [67].

3.2 Reconstruction

Reconstruction is a term for algorithms used to construct objects such as events, tracks, and showers using the hits from individual cells in the detector. In NOvA, the reconstruction chain is in three parts: event reconstruction, particle identification and energy estimation.

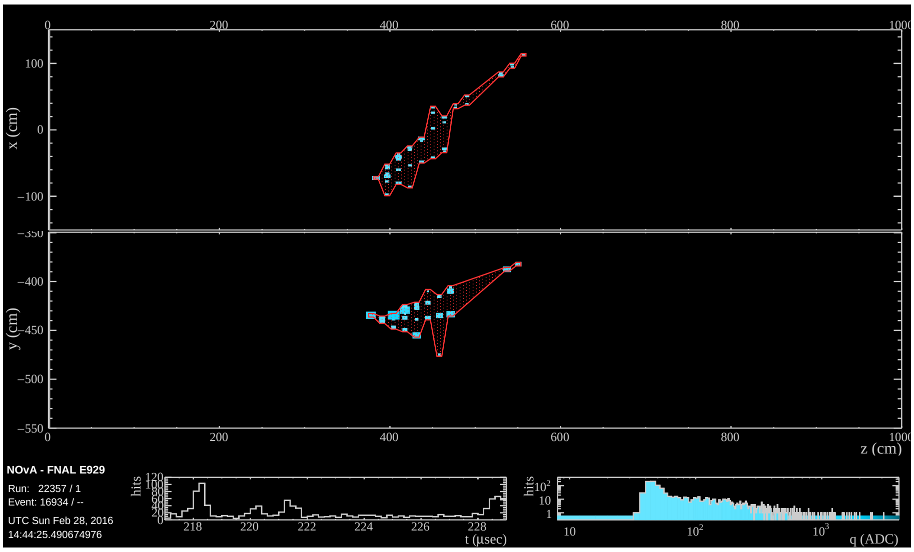

Figure 3.1: An example event from FD after hit finding and slicing. The clustered area of hits found by the algorithm is marked in red. Retrieved from [70].

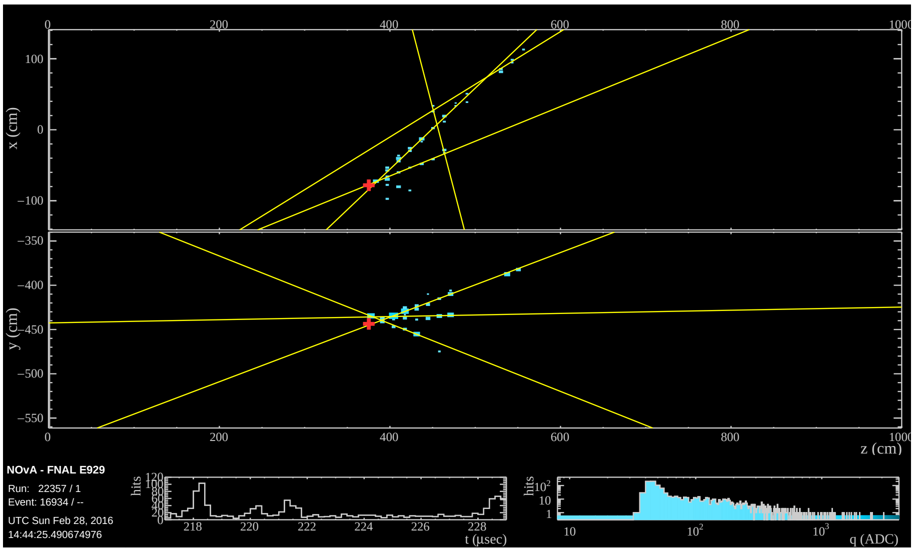

Figure 3.2: The same example event as seen in figure 3.1 after Hough transform has been performed to find straight lines, represented here in gold. The red cross indicates the elastic arms vertex. Retrieved from [70].

3.2.1 Event Reconstruction

Event reconstruction starts by hit finding. If light produced in a cell is above a set threshold, then it is recorded as a cell hit. The production of cell hits is followed by slicing, in which hits coincident in time and space are grouped to form a slice. Hits not part of any slice are considered as noise. An example event after slicing has been applied is seen in figure 3.1, where the red area indicates where the cluster of hits has been found [52] [62].

After slicing, a modified Hough transform [79] is applied to search

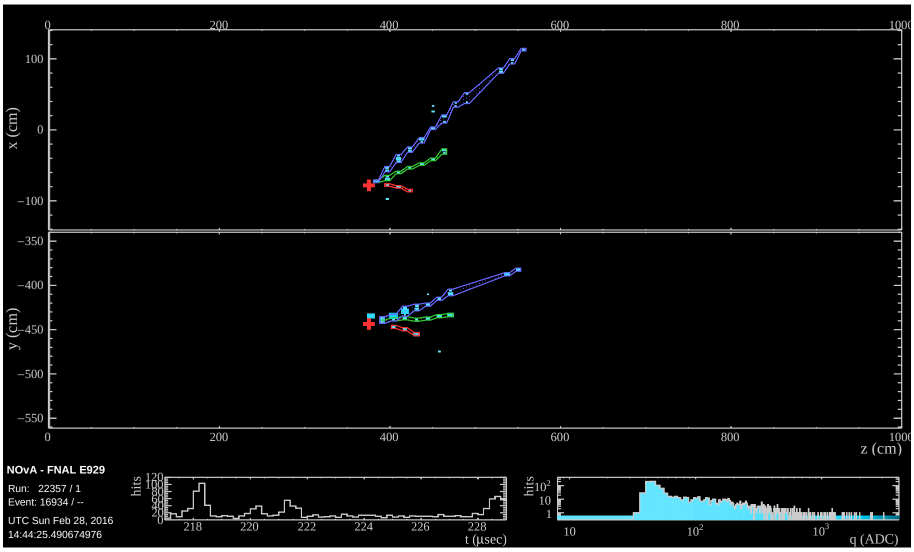

Figure 3.3: The same event as figures 3.1 and 3.2 after fuzzy-K algorithm has been performed. The blue, green and red regions represent the reconstructed prongs and the red cross is the elastic arms vertex. Retrieved from [70].

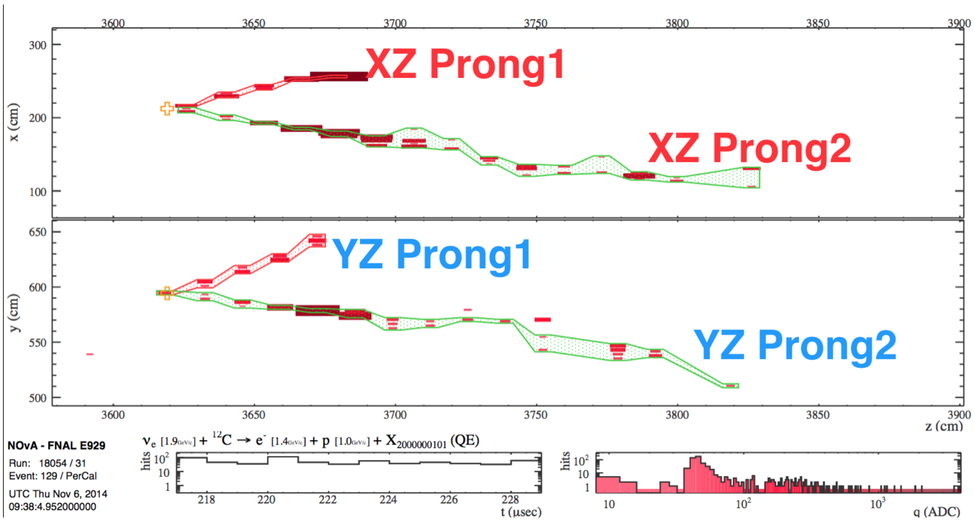

Figure 3.4: An example of 3D prong reconstruction using fuzzy-K algorithm for a simulated ν_e CC quasi-elastic interaction in the FD. One can see the same track in both XZ and YZ views and the labels demonstrate how the 2D prongs are matched in 3D. Retrieved from [78].

for prominent straight line features. This is a commonly used technique in image processing to identify straight lines. Then, to find primary neutrino interaction intersections, the elastic arms [80] algorithm is used. The results from Hough transform are used as a seed for the algorithm and the output of this algorithm is a 3D vertex point. The same example event as seen in figure 3.1 after Hough transform and elastic arms have been applied is seen in figure 3.2. The golden lines represent the found Hough lines and red cross the vertex point [52] [72].

The next step in the reconstruction chain is to form prongs, which are hit clusters with a start point and a direction. Prongs are found using a fuzzy-k algorithm, which is used to separate noise hits from actual tracks. The "fuzzy" part of the algorithm allows hits to belong to more than one prong, since prongs are separated in XZ and YZ views due to detector geometry. Using cell hits in a slice and starting from a 2D view, two views are matched into a full 3D picture. Prongs are then formed into tracks using Kalman and break point fitter, which are discussed in the next section. The same example event as was seen in figures 3.1 and 3.2, is presented in figure 3.3 after prong finding. To further demonstrate the matching of 2D to 3D, figure 3.4 shows a simulated ν_e CC event in two views with the prongs named and marked to demonstrate the view matching [52] [72].

Break Point Fitter

The break point fitter takes 3D prongs created by fuzzy-k and elasticated arms vertex (as described in the previous section), and constructs the particle's trajectory using multiple Coulomb scattering. The Coulomb scattering is determined from a straight line fit, and the energy loss along the prong path is also calculated. This energy loss if summed from the end of the track to the vertex [81].

Kalman Tracks

Kalman tracks, based on Kalman filter routine [82], are used to find muons and anti-muons, which helps with identifying ν_μ (and $\overline{\nu}_\mu$) CC events that are mostly characterised by long muon (and anti-muon) tracks, as seen in figure 2.17. The Kalman filter uses hit clusters from slices as inputs and forms track in two views: XZ and YZ. Each 2D view is matched to a full 3D track. To form 2D tracks, a seed is formed from a segment of a track that is from a pair of hits separated by less than 4 cells. Hits, which describe the energy deposit per cell, are added iteratively based on the seed's position and gradient. A new hit is included if the χ^2 value of the track is changed by an acceptable value, and it is repeated until no new hits are available [62] [67] [72].

Cosmic Tracks

The cosmic tracks are separated by directions of the tracks of incoming particles. Cosmics arrive to the detectors from above whereas beam particles are arriving from the direction of the beam. This algorithm is called a window tracker, and it takes into account the fact that cosmic muons undergo multiple Coulomb scattering in the detector that can change their direction. However, these tracks can still be assumed to follow a straight line trajectory over small portions of the track in Z-direction [67].

The algorithm works by first choosing a set of n planes in the Z-direction covering a chosen narrow range at the most upstream side of a slice. A straight line fit to that window is used to evaluate which hits belong to that specific 2D track. Following the window in the Z plane, hits that are consistent with the line corresponding to the first window are accepted to be part of the track. A new fit is then done with n-1 number of planes extrapolating the fit into the next window along Z-direction. Again, it is evaluated whether these hits belong to the line. The process is repeated until all planes in one slice have been considered. Like in Kalman tracks, the 2D tracks are matched to form full 3D tracks [52] [67].

3.2.2 Energy Estimation

Calorimetric Energy Calibration

To be able to meaningfully compare recorded deposited energy in the detectors, one must perform calorimetric energy calibration. This allows different recorded activities in the detector be comparable on both a relative and on an absolute scale. The calibration is divided into two parts: relative and absolute calibration [52] [62] [67].

In the FD, cosmic muons can be used for calibration as their energy loss by ionisation is well described by the Bethe-Bloch formula that can be expressed as (neglecting small corrections for highly relativistic particles) [84]:

$$-\frac{dE}{dx} = \frac{Dq^2 n_e}{\beta^2}\left[ln(\frac{2m_e c^2\beta^2\gamma^2}{I}) - \beta^2\right], \tag{3.2.1}$$

where E is the energy, x is the distance travelled through a medium, q charge, $D = \frac{4\pi\alpha^2\hbar}{m_e} = 5.1 \times 10^{-25} \quad MeVcm^2$, m_e the electron mass, $\beta = v/c$, $\gamma = (1-\beta^2)^{-1/2}$, n_e the electron density, and I mean ionisation potential of the atoms averaged over all electrons. To demonstrate, the mean energy loss rate described by Bethe-Bloch formula for three different particles in different media is seen in figure 3.5.

There is also a selection applied to the cells to make sure the ones used for calibration have a well defined path length meaning a muon deposited energy by ionisation is withing the cell. The selected cells are called tricells, and for each of them a hit has to be between two cells which also have hits. This is because cells not surrounded by other cells

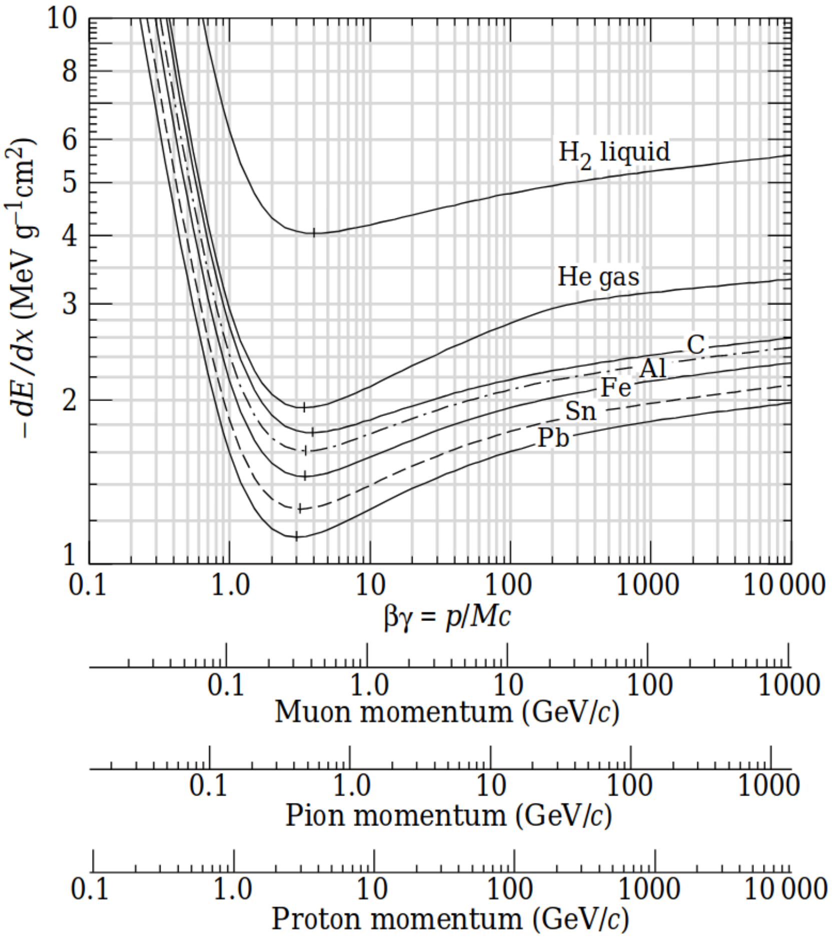

Figure 3.5: Mean energy loss rate calculated via Bethe Bloch formula in different materials for muons, pions and protons. Retrieved from [10].

.

with hits cannot be guaranteed to have been caused by a muon passing through both upper and lower horizontal boundaries of the cell. This requirement makes it possible for the track angle to be measured well as the enter and leave points are well known. A diagram showing an example of a tricell is seen in figure 3.6. Track angle combined with a trigonometric calculation gives an accurate path length that is shown in the figure [52] [67].

Relative calibration removes any dependence on a position in the detector response due to for example changes in light attenuation in the cell or distance from the readout electronics. Hits further away from the readout will be more attenuated when travelling compared to hits close to the readout. This appears as though hits close to the readout have higher energy compared to those further away, which relative calibration corrects for. This is done by performing a fit to the distribution of mean recorded number of photoelectrons per centimeter (PE/cm) from tricell hits relative to the distance from the readout. This is demonstrated in figure 3.7 on the right, which shows how these values change before and

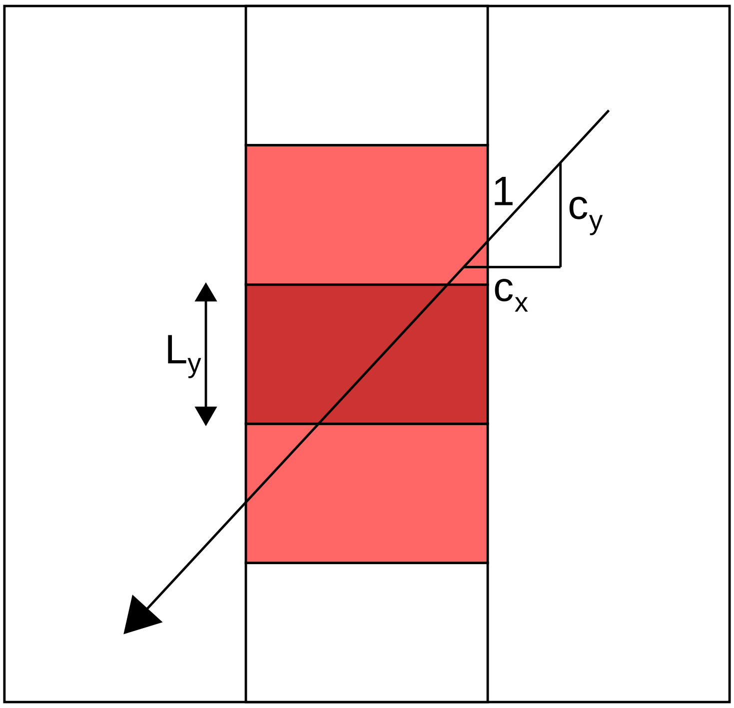

Figure 3.6: Diagram of a tricell hit (dark red). The cell is counted as a tricell hit when each of its neighbour cells (light red) is also hit. Using trigonometry and track angle, the path length is found to be L_y/c_y . C stands for cell, L for length, and x and y refer to a position in the detector. Retrieved from [83]

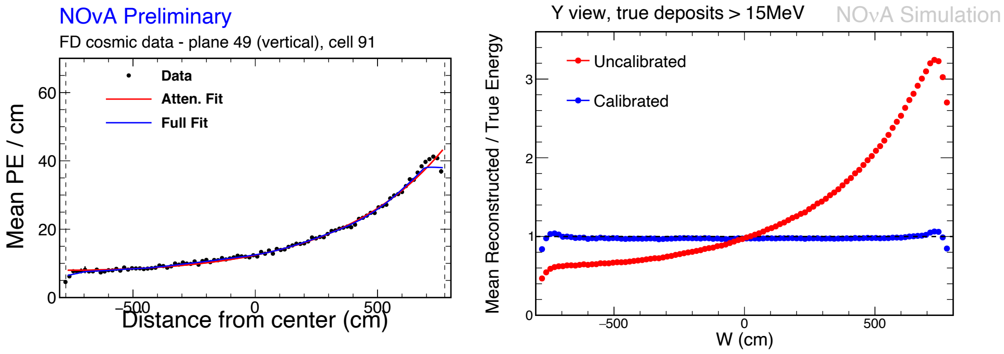

Figure 3.7: Relative calibration schematic. The left plot shows PE/cm for tricell hits in cosmic data relative to distance from the centre of the cell in a single horizontal cell in FD. The solid lines shows the best fit to the cosmic data. The right plot shows the hit position along the length of the cell, W, in the x-axis against ratio of reconstructed and true energy on the y-axis with uncalibrated (red) and calibrated (blue) cases. Retrieved from [83].

after calibration. The figure also shows how the number of PE/cm drops at near the detector edges (left plot). The mean values of the energy per *cm* are divided with the corrected number of photoelectrons per *cm* in the cell to obtain the calorimetric energy scale. The value is calculated for each detector for each data taking period separately. Separate numbers are calculated for reconstructed data and simulation [52] [62] [67].

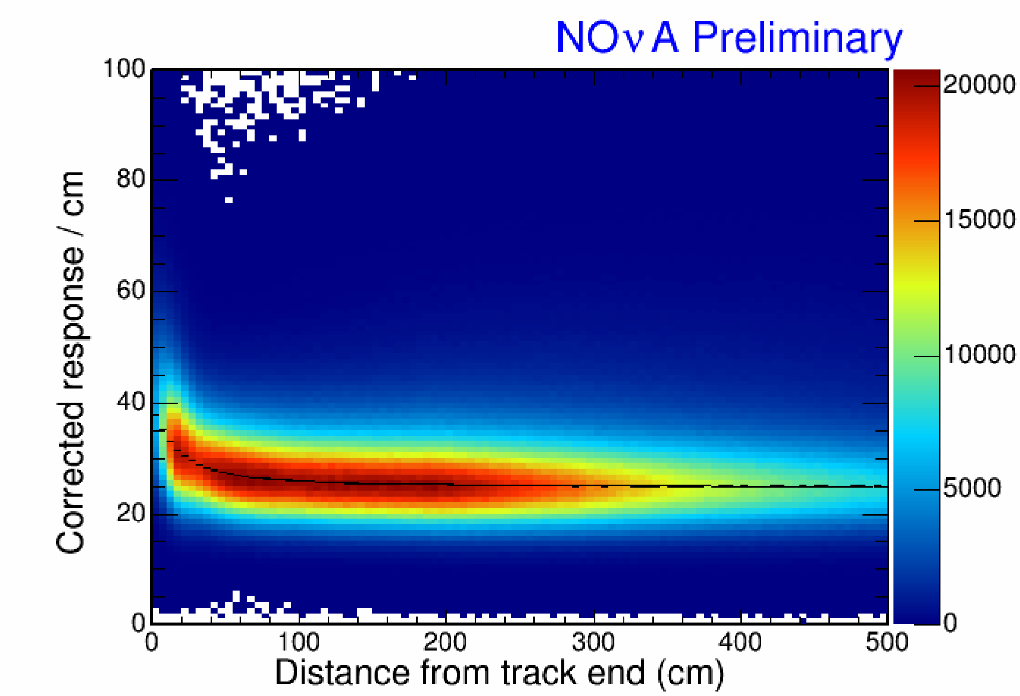

Figure 3.8: A diagram describing the absolute calibration procedure. Each histogram entry is from a tricell hit on a stopping muon track, with distance to the stopping point of the muon track shown on the x-axis and the path normalised attenuation corrected detector response shown on the y-axis. The black line shows the mean of a fit to the peak of the attenuation corrected detector response at particular distances from the track end. Hits between 100 *cm* and 200 *cm* form the track window that is used for absolute calibration. Retrieved from [85].

Absolute calibration gives an energy scale (in NOvAs case the most convenient is *GeV*) independent from the recorded number of photoelectrons (PE) to the hits from tricells. This is calculated using the Bethe-Bloch equation (equation 3.2.1) that describes an energy loss for a particle when travelling through a medium, which gives a scale factor for calculating an energy in *GeV* from the attenuation-calibrated detector output. Hits between 100 *cm* and 200 *cm* from the track end, a track window, are used because their detector response is known to be uniform within 1.8% [86]. Hits within that track window correspond to hits with minimum ionising region for muons. Figure 3.8 shows the distance from the track end and the corrected response per *cm* alongside the best fit to the peak of the attenuation corrected detector response [52] [62] [67].

Calibration also corrects for detector aging. It is known that both the overall light response and the number of hits observed decreases as a function of time but the underlying cause for detector aging is not known. Its effects are, however, taken into account by doing the calibration in small periods of time [52] [62] [67].

Even with the most careful consideration, some discrepancies exist between data and simulation in the calibration. This uncertainty is accounted as a calibration systematic uncertainty, described in section 3.6.1.

ν_μ Energy

The energy estimation is performed based on the charge deposited by the final state particles. The energy of ν_μ and $\overline{\nu}_\mu$ CC events can simply be defined as [52]:

$$E_{\nu_\mu(\overline{\nu}_\mu)} = E_{\mu(\overline{\mu})} + E_{had} \tag{3.2.2}$$

where $E_{\mu(\overline{\mu})}$ is the energy of the outgoing (anti)muon calculated from the length of the track. E_{had} is the hadronic energy estimated by adding up all the energy from visible hits not attributed to the muon. The visible energy is then corrected using a spline based fit made using simulation. In the end, the energy resolution of E_{had} is about 30% and 4% for the $E_{\mu(\overline{\mu})}$, which ends up giving the final neutrino energy a resolution of 9% for neutrinos and 8% for antineutrinos in the latest analysis. The slight difference is due to the fractional energy of the hadronic shower being higher for neutrinos compared to antineutrinos [52] [67].

3.2.3 Particle Identification

Convolutional Visual Network

For particle identification, NOvA uses a convolutional visual network (CVN), which is a convolutional neural network (CNN) deep learning method. The CVN is trained on the reconstructed tracks to identify the type of interaction: ν_μ CC, ν_e CC, ν_τ CC, NC and cosmogenic background [72].

One type of misidentification in NOvA is mistaking NC interactions as CC interactions. This is because NC interactions can produce pions that can look like leptonic activity. The only difference between these tracks is a spike in energy deposition at the end of the pion track. Also, a neutral pion decaying into a pair of photons producing electromagnetic showers can appear very similar to an electromagnetic shower produced by electrons other than the gap between the shower and the vertex in neutral pions. The CVN is trying to minimise these misidentifications.

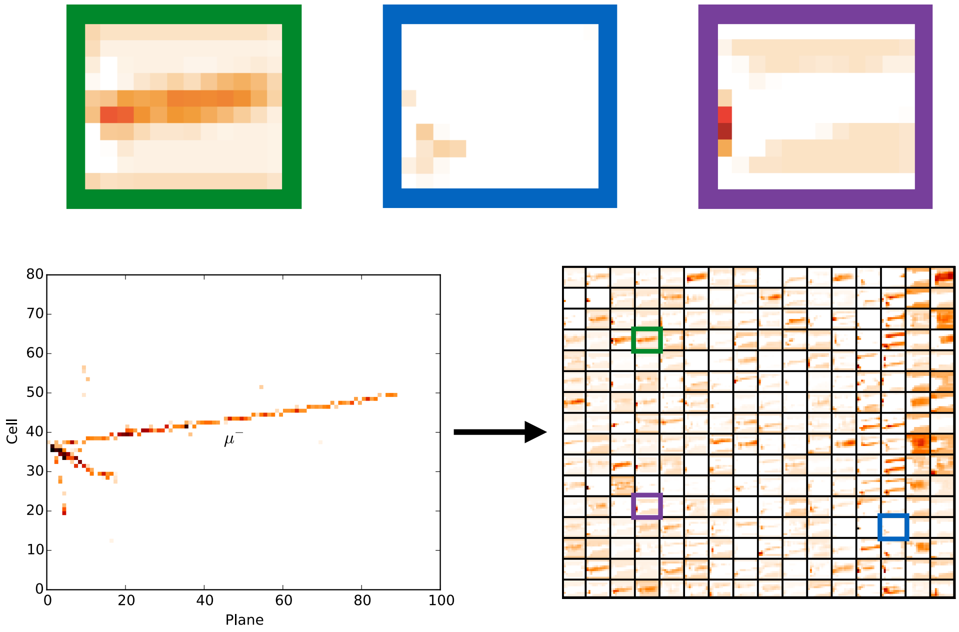

Figure 3.9: Figures describing ν_μ CC CVN. A true ν_μ CC interaction is seen on the bottom left figure. The bottom right figure includes the feature maps extracted from the bottom left event. Green, blue and purple highlighted plots above are feature maps that represent muon tracks, electron showers and hadronic activity respectively. Retrieved from [72].

A convolutional neural network is an image recognition technique using images as input. In NOvA's case these are reconstructed event slices in XZ and YZ views, which are called pixel maps. Each of these pixels is assigned an intensity proportional to deposited charge. The input is formed by first clustering energy deposits recorded together in space and time into slices. These grids of 100 planes deep and 20 cells wide contain the slice and they effectively separate neutrino interactions from cosmic rays and other noise. The size and placement was chosen so that the majority of neutrino events are fully contained. An example of a pixel map is seen in figure 3.9 on the bottom left [52] [72].

The architecture of the CVN model is complicated: the input layer is followed by 13 hidden layers of four types (convolution, pooling, local response normalisation (LRN) and inception model). The final output is formed by a softmax layer. These layers perform a series of linear operations on the original output images to extract key features and form them into feature maps. Each map represents a variable or a strength of response to a certain condition in the underlying image. An example of feature maps can be seen in figure 3.9 in the bottom right with examples of three feature maps highlighted above. These three images can be interpret as responses to muon tracks, electron showers and hadronic activity [52] [72].

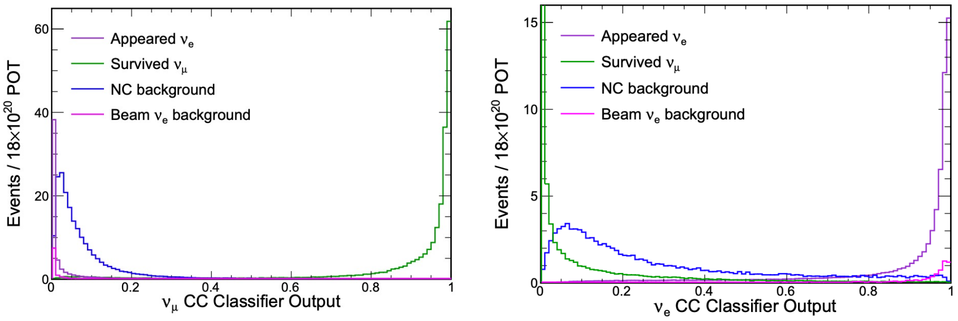

Figure 3.10: **Left**: ν_μ CC classifier output distribution for the CVN training using FD simulation. Appearing ν_e CC interactions, surviving ν_μ CC, NC interactions, and NuMI beam ν_e CC interactions are showed in purple, green, blue and magenta respectively. Retrieved from [72]. **Right**: ν_e CC classifier output distribution for the CVN training using FD simulation. Appearing ν_e CC interactions, surviving ν_μ CC, NC interactions, and NuMI beam ν_e CC interactions are showed in purple, green, blue and magenta respectively. Retrieved from [72].

The CVN was trained using NOvA simulation for both FHC and RHC separately. The outputs are variables between 0 and 1 that are interpreted as a probability of how ν_μ CC, ν_e CC, ν_τ CC, NC or cosmic-like the particle is. The results from this training using FD simulation are seen in figure 3.10 for ν_μ CC and ν_e CC outputs [12] [52] [72].

NOvA also has a another CVN training called ProngCVN for particle identification, whose goal is to classify individual reconstructed prongs within a slice. The architecture is similar to that of the original CVN with output of XZ and YZ pixel maps containing individual prongs. The output of the network describes how likely a prong is to originate from an electron, a proton, a photon, a pion or a muon [52].

Reconstructed Muon Identifier

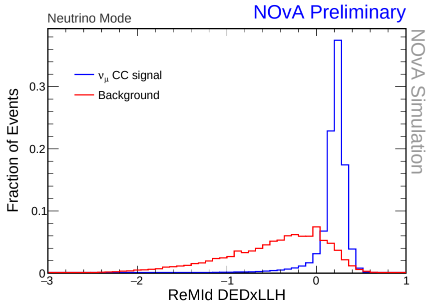

Figure 3.11: $\frac{dE}{dx}$ LLH for FD FHC simulation.

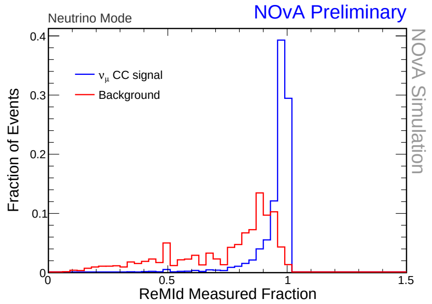

Figure 3.12: Fraction of planes used in the $\frac{dE}{dx}$ LLH calculation for FD FHC simulation.

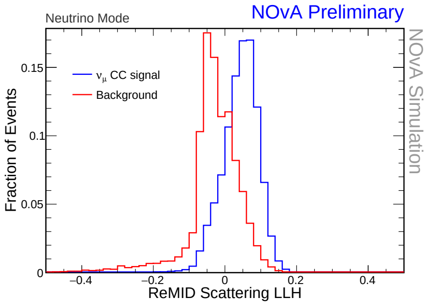

Figure 3.13: Scattering LLH for FD FHC simulation.

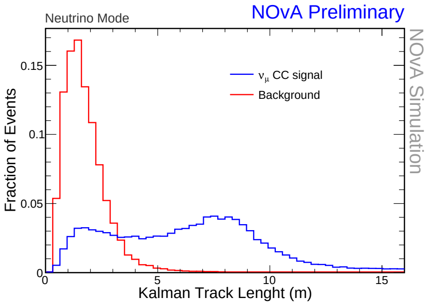

Figure 3.14: Track Length for FD FHC simulation.

Reconstructed Muon Identifier (ReMID) is a Boosted Decision Tree (BDT) based algorithm to identify muon-like tracks in a slice. A decision tree is a machine learning algorithm which loops over the chosen training variables performing a one dimensional split of the data by maximising the chosen figure of merit (FOM). Boosting, which combines a set of weak classifiers into a one final powerful classifier, is applied to help with misclassification. ReMID is specifically trying to combat the misidentification of pion and muon track similarities, mentioned earlier in this section [87] [88].

ReMID uses simulated ν_μ CC as signal and all other events (mostly NC) as background in its training. The input consists of 4 different variables from the Kalman track algorithm (section 3.2.1): $\frac{dE}{dx}$ log-likelihood (LLH), scattering LLH, track length and fraction of planes used in the $\frac{dE}{dx}$ LLH calculation (MeasFrac). All of these variables for simulated FD data are shown in figures 3.11, 3.12, 3.13 and 3.14. These plots show how the variables differ when ν_μ CC is considered as signal with all of them showing good separation between signal and background [89].

The simulation used includes the processes that describe the deposited energy ($\frac{dE}{dx}$), which is determined by the the Bethe-Bloch formula (see equation 3.2.1). Determining the LLH of $\frac{dE}{dx}$ is done by comparing $\frac{dE}{dx}$ values to distribution against the muon hypotheses. $\frac{dE}{dx}$ for muon tracks is different from pion track $\frac{dE}{dx}$ that usually loses energy via hadronic scattering as well as ionisation unlike muon that mainly lose energy via ionisation. Scattering LLH, in turn, describes how much each track deviates from a straight line. Muon tracks only negligibly differ from a straight line so they are good input variables to use. Hadronic showers usually contain more than one short track rather than a single long track, which is why variable track length is useful. The last variable, non-hadronic plane fraction, counts the planes that have energy outside of the track excluded from $\frac{dE}{dx}$ LLH calculations [52] [67] [89].

Cosmic Rejection BDT

Cosmic Rejection BDT is an additional algorithm to remove cosmogenic background. Both ν_μ and ν_e have separately trained BDTs, and the results are a value for each event that describes how likely the event is to be cosmic-like with score of 0 for a cosmic-like and 1 for a neutrino-like event. The inputs are a wide range of kinematic variables from simulated beam and cosmic events. [90].

For ν_e events, the BDT is trained separately for RHC and FHC as well as for core and peripheral samples. A peripheral sample is a special sample for ν_e events that contains highly electron neutrino like events according to the CVN, but these events do not pass containment cuts [52].

3.3 Event Selection

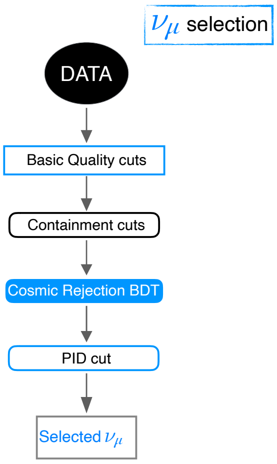

Figure 3.15: Current ν_μ selection flowchart, retrieved from [91].

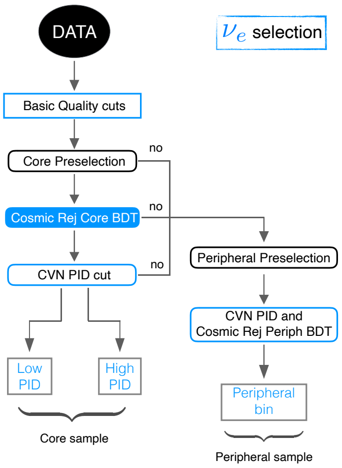

Figure 3.16: Current ν_e selection flowchart, retrieved from [91].

NOvA's event selection ensures that a minimum amount of background particle interactions are in the analysis. Event selection for ν_μ consists of 4 steps, illustrated in figure 3.15. The selection for ν_e is seen in figure 3.16, and it is slightly more complicated than the ν_μ selection as there are two different ν_e samples: core and peripheral. The focus in the next sections is on ν_μ selection for the purpose of the topic of this thesis, although some information of ν_e selection is included for clarity.

ν_μ **Basic Quality Cuts**
$0\ GeV < \nu_\mu$ Energy $< 5\ GeV$
ReMID PID Value > 0
Number of Hits > 20
Number of Continuous Planes > 4
Number of tracks found by a cosmic tracker > 0

Table 3.1: NOvA 2020 analysis basic quality cuts for ν_μ, values from [90]

.

ν_e **Basic Quality Cuts in ND**	**FD**
$20 \leq$ Number of hits ≤ 200	$1\ GeV < \nu_e$ Energy $< 4\ GeV$
$0\ GeV \leq \nu_e$ Energy $< 4.5\ GeV$	$30 <$ Number of Hits < 150
$100\ cm <$ Longest Prong $< 500\ cm$	$100\ cm <$ Longest Prong $< 500\ cm$

Peripheral
$0\ GeV \leq \nu_e$ Energy $< 4.5\ GeV$
ν_e cosmic hits per plane ≥ 8
Number of prongs $\neq 0$
Valid result from elastic arms

Table 3.2: NOvA 2020 analysis basic quality cuts for ν_e, values from [90].

3.3.1 Basic Quality Cut

The first stage of event selection is applying basic quality cuts to remove unambiguously poor quality events. For ν_μ this includes no zero energy events as well as an upper energy limit of 5 GeV (due to constraints of the detector), at least one Kalman track with assigned ReMID value (ReMID PID value > 0), at least one straight line track is present (track found by tracker called cosmic tracker), enough hits (> 20) are produced to register as an event, and enough continuous planes are present to make sure no vertical tracks are chosen. The basic quality cut is summarised in table 3.1 [90].

The basic quality cut for ν_e is seen in table 3.2. A different cut is applied for ND and FD, unlike in the case of ν_μ selection. In the ND, the number of hits is constrained between 20 and 200, and the energy is between 0 and 4.5 GeV. Also the length of the longest prong is set between 100 and 500 pixels. In the FD, the energy is set between 1 and 4 GeV, number of hits is set between 30 and 150, and the longest prong length is set between 100 and 500. The peripheral sample also has different cuts, energy is set between 0 and 4.5, there are enough hits per plane, and that the reconstruction is sufficient (no zero prong results, valid elastic arms result) [90].

3.3.2 Containment Cut

ν_μ **ND Containment Cuts**
Shower Min X,Y > -180 *cm*
Shower Max X,Y < 180 *cm*
Shower Min Z > 40 *cm*
Shower Max Z < 1525 *cm*
Only one track in the muon catcher
Slice first plane > 1
Slice last plane < 212
Kalman tack start Z < 1100 *cm*
Kalman tack stop Z < 1275 *cm* or Kalman Y position at mucat boundary < 55 *cm*
Kalman track projected forward cells > 5
Kalman track projected backward cells > 10

Table 3.3: NOvA 2020 analysis ν_μ containment cuts in the ND, values from [90].

ν_μ **FD Containment Cuts**	
Prong Cuts	Containment Cuts
Prong distance to top > 60 *cm*	Kalman track projected forward cells > 6
Prong distance to bottom > 12 *cm*	Kalman track projected backward cells > 6
Prong distance to east > 16 *cm*	Cosmic track projected forward cells > 5
Prong distance to west > 12 *cm*	Cosmic track projected backward cells > 7
Prong distance to front > 18 *cm*	Planes to front > 2
Prong distance to back > 18 *cm*	Planes to back > 3

Table 3.4: NOvA 2020 analysis ν_μ containment cuts in the FD, values from [90].

The next stage of the event selection is containment, which provides assurance that charged particles deposit their energy within the detector. This helps to choose events where the total energy can be accurately calculated and removes obvious backgrounds such as rock muons in the near detector and cosmic rays in the far detector [90].

The containment cuts for ν_μ in both ND and FD are seen in tables 3.3 and 3.4 respectively [90].

The containment cuts for ν_e are seen in tables 3.5 and 3.6 for ND and FD respectively. The cuts are very similar as for ν_μ, although slightly simpler.

ν_e **ND Containment Cuts**
Shower Min X,Y > -150 *cm*
Shower Max X,Y < 150 *cm*
Shower Min Z > 100 *cm*
Shower Max Z < 1225 *cm*
-100 *cm* < Vertex X < 160 *cm*
-160 *cm* < Vertex Y < 100 *cm*
150 *cm* < Vertex Z < 900 *cm*
Planes to front > 10

Table 3.5: NOvA 2020 analysis ν_e containment cuts at the ND, values from [90].

ν_e **FD Containment Cuts**
Prong distance to top > 63 *cm*
Prong distance to bottom < 12 *cm*
Prong distance to east > 12 *cm*
Prong distance to west > 12 *cm*
Prong distance to front > 18 *cm*
Prong distance to back > 18 *cm*

Table 3.6: NOvA 2020 analysis ν_e containment cuts at the FD, values from [90].

3.3.3 Cosmic Rejection

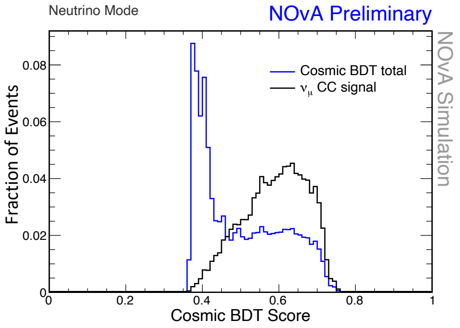

Figure 3.17: The cosmic BDT classification results (blue) and the share of ν_μ CC events for FD FHC simulation.

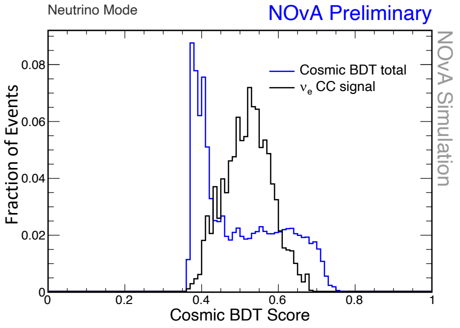

Figure 3.18: The cosmic BDT classification results (blue) and the share of ν_e CC events (not including peripheral) for FD FHC simulation.

Cosmic rejection BDT cut is applied next, using the algorithm that was described in section 3.2.3. For ν_μ selection, the cut was determined to be that Cosmic Rejection PID > 0.45. Results from the cosmic BDT are seen in figure 3.17 showing the total cosmic BDT score and the share of ν_μ signal. The y-axis was normalised to show the distributions of both on the same scale, even though there were more events in total compared to signal [90].

For ν_e events, the BDT is optimised separately for FHC and RHC modes and for core and peripheral sample. For the core sample, the cut for the FHC sample is set at PID score ≥ 0.49 and for RHC the PID score limit is ≥ 0.47. Figure 3.18 demonstrates the cosmic BDT scores for FHC ν_e simulation. Similar normalisation was also applied to these plots as the similar ν_μ plots. The cosmic BDT and PID cuts are linked to each other for the peripheral sample and will be explained in the next section [90].

3.3.4 PID Cut

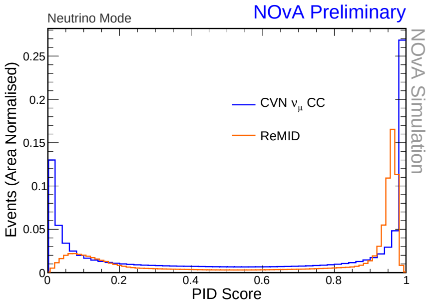

Figure 3.19: CVN and ReMID scores for ν_μ CC for FHC ND simulation to demonstrate the cuts applied for ν_μ sample.

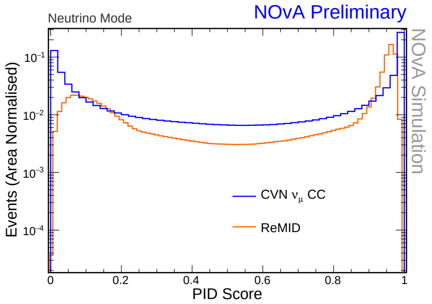

Figure 3.20: CVN and ReMID scores for ν_μ CC for FHC ND simulation to demonstrate the cuts applied for ν_μ sample, log scale plot.

The PID cut is the last cut applied before retrieving the final energy spectra and selected events for the NOvA analysis. The PID cuts were optimised by using an FOM of $\frac{S}{\sqrt{(S+B)}}$, where S is signal and B is background. This formula gives the most amount of signal whilst not compromising on the quality by giving as little background as possible.

The plot showing the scores for both ReMID and CVN are seen in figures 3.19 and 3.20 (the latter having a logarithmic y-axis). The final PID cut for ν_μ sample is set as ReMID > 0.3 and CVN score > 0.8 [90].

Similar analysis is performed for ν_e, separately for core and peripheral samples. The FOM is the same as for ν_μ, but there are different cuts for FHC and RHC samples. For FHC the ν_e CVN score is set to ≥ 0.84 and for RHC as ≥ 0.85 [90].

For the peripheral sample, the ν_e CVN cut is set as ≥ 0.97 if cosmic BDT is > 0.62 and 0.6 for FHC and RHC respectively, and ≥ 0.995 if the cosmic BDT score is > 0.57 for FHC and 0.56 for RHC.

Cut Level	ν_μ	$\overline{\nu_\mu}$	Total Beam Bkg	NC	Cosmics	Effic	Purity
NoCut	602	43.1	124	563	2.52×10^6		
Quality	569	41.0	121	490	4.72×10^5	0.946	1.20×10^{-3}
Containment	249	14.1	83.9	336	6.99×10^3	0.413	0.033
Cosmic Rej.	239	13.8	78.2	293	1.54×10^3	0.397	0.111
Full PID	201	12.5	0.717	2.51	6.02	0.334	0.956

Figure 3.21: Final expected sample numbers for ν_μ FHC after each cut for signal and various backgrounds using 2020 analysis simulation. Retrieved from [92].

Cut Level	$\overline{\nu_\mu}$	ν_μ	Total Beam Bkg	NC	Cosmics	Effic	Purity
NoCut	193	111	45.8	225	1.63×10^6		
Quality	186	106	45.0	195	2.99×10^5	0.964	6.21×10^{-4}
Containment	86.7	35.5	32.0	132	4.38×10^3	0.449	0.019
Cosmic Rej.	84.9	32.8	28.9	107	635	0.440	0.099
Full PID	79.5	25.7	0.336	0.790	1.65	0.412	0.966

Figure 3.22: Final expected sample numbers for ν_μ RHC after each cut for signal and various backgrounds using 2020 analysis simulation. Retrieved from [92].

Cut Level	ν_e	$\overline{\nu_e}$	Total Beam Bkg	NC	Cosmics	Effic	Purity
NoCut	96.8	1.84	1.58×10^3	820	2.95×10^6		
Data Qual.	93.2	1.78	1.5×10^3	680	5.12×10^5	0.964	1.81×10^{-4}
Containment	71.4	1.39	560	510	2.58×10^4	0.738	2.65×10^{-3}
Cosmic Rej.	63.5	1.02	91.5	169	353	0.656	0.094
Full PID	52.0	0.859	11.4	4.73	1.51	0.537	0.738
Basic Cuts	38.7	0.689	549	508	4.53e+05	0.400	8.53e-05
Full PID	8.63	0.194	3.84	0.253	1.63	0.0892	0.593

Figure 3.23: Final expected sample numbers for ν_e FHC after each cut for signal and various backgrounds using 2020 analysis simulation. The core sample is shown on the first five rows and the peripheral sample on the last two rows. Retrieved from [92].

Cut Level	$\overline{\nu_e}$	ν_e	Total Beam Bkg	NC	Cosmics	Effic	Purity
NoCut	27.0	5.14	723	328	1.92×10^6		
Data Qual.	26.4	4.92	693	266	3.25×10^5	0.979	8.12×10^{-5}
Containment	20.9	3.72	252	199	1.61×10^4	0.774	1.26×10^{-3}
Cosmic Rej.	19.3	2.57	27.4	57.8	296	0.715	0.048
Full PID	17.3	1.96	5.43	1.6	0.589	0.639	0.643
Basic Cuts	8.55	2.25	259	200	2.88×10^5	0.317	2.97×10^{-5}
Full PID	3.01	0.419	2.11	0.117	0.998	0.112	0.453

Figure 3.24: Final expected numbers for ν_e RHC after each cut for signal and various backgrounds using 2020 analysis simulation. The core sample is shown on the first five rows and the peripheral sample on the last two rows. Retrieved from [92].

3.3.5 Final Event Samples

The final expected number of events after all cuts for both FHC and RHC are seen in figures 3.21 and 3.22 respectively. The numbers are shown for both ν_μ and $\overline{\nu_\mu}$ and the numbers are shown for each cut individually. For ν_e the same numbers are seen in figures 3.23 and 3.24. All of these plots were made using simulation.

3.4 Binning

3.4.1 ν_μ Quartile Bins

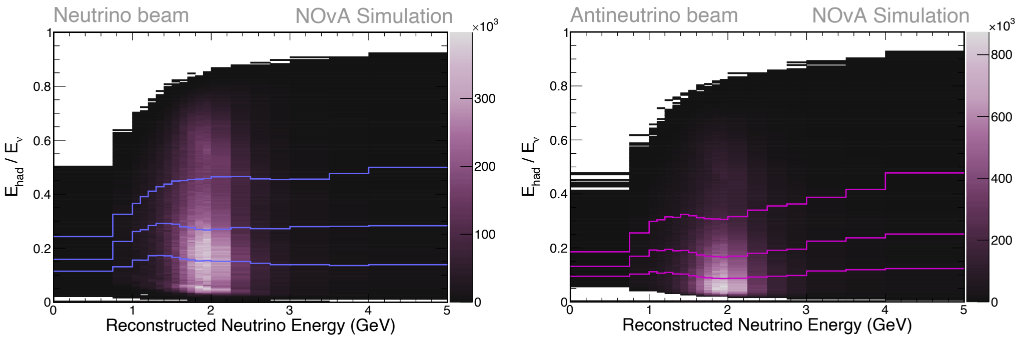

Figure 3.25: E_{had}/E_ν vs. reconstructed neutrino energy showing the quartile subsample limits for FHC (blue lines, left) and RHC (pink lines, right). Retrieved from [93].

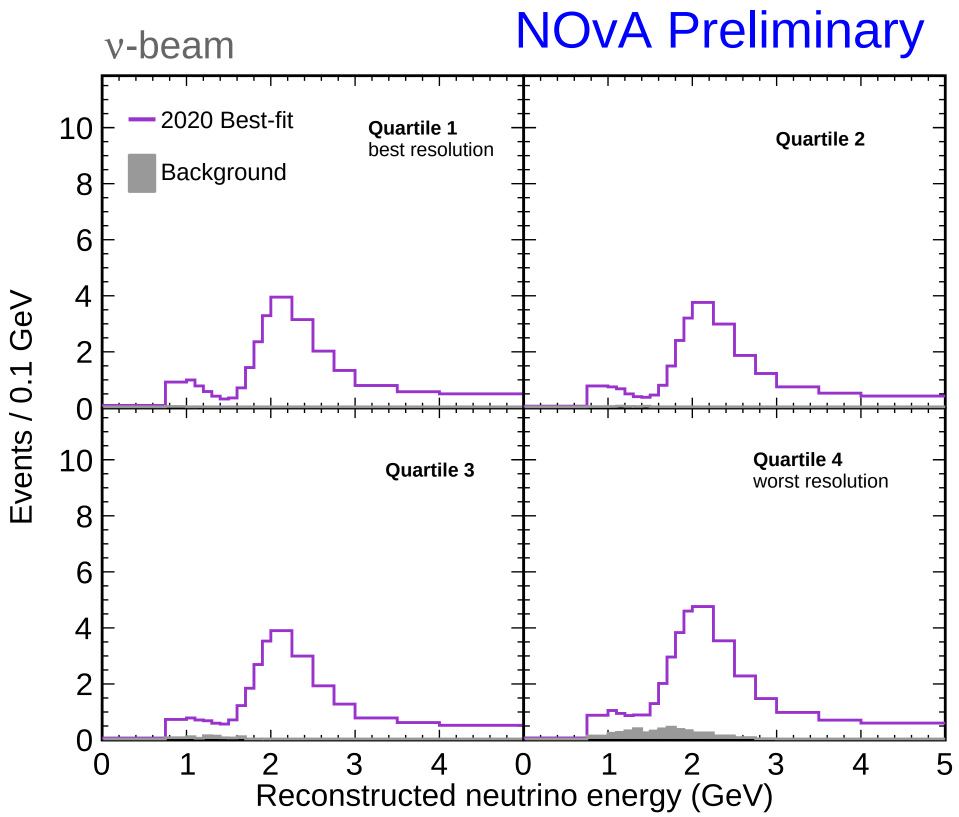

Figure 3.26: The ν_μ FD predicted reconstructed neutrino energy spectra based on simulation for the 2020 oscillation analysis split into quartiles based on hadronic energy fraction in FHC mode. Retrieved from [94].

As mentioned in section 3.2.2, the energy of ν_μ is the sum of the energy of the muon and hadronic components. The hadronic energy resolution is worse than that for the muon, hence ν_μ events with a smaller hadronic energy fraction have a better energy resolution compared to events with

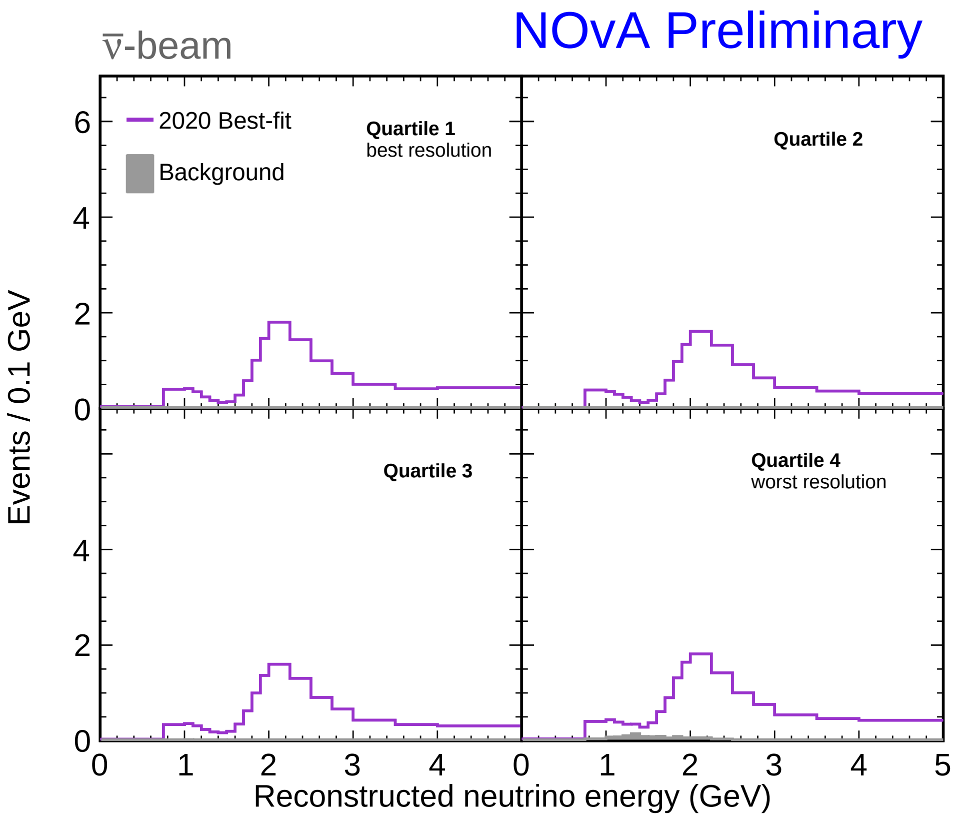

Figure 3.27: The ν_μ FD predicted reconstructed neutrino energy spectra based on simulation for the 2020 oscillation analysis split into quartiles based on hadronic energy fraction in RHC mode. Retrieved from [94].

a large hadronic component. This is why the ν_μ events are split into four different subsamples called quartiles all containing 25% of the total number of events. The split is based on the hadronic energy fraction: E_{had}/E_ν, where E_{had} is the hadronic energy and E_ν the total energy. Quartile 1 has the best resolution containing the events with the smallest hadronic energy fraction and the smallest background, while quartile 4 has the opposite. The limits are demonstrated in figure 3.25, and the reconstructed neutrino energy spectra for each quartile are seen in figures 3.26 and 3.27 for ν_μ FD simulation data for FHC and RHC respectively [67].

3.4.2 ν_e Binning

The ν_e sample is divided into core and peripheral samples as explained in section 3.2.1. Based on the CVN scores (also called PID values) given for the final selected events, the events are put into two different bins in the core sample: those with high PID values and those with low PID values. Thus there are three different ν_e bins: low and high PID as well as the peripheral bin. The spectra of reconstructed neutrino energy for each bin

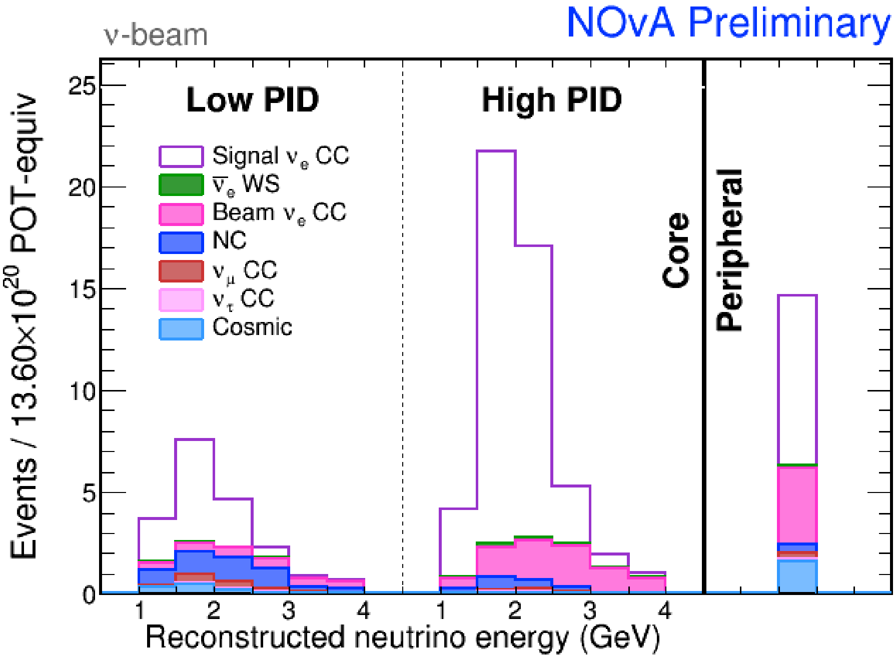

Figure 3.28: The ν_e FD predicted reconstructed neutrino energy spectra based on simulation used in the 2020 oscillation analysis split into two PID samples and a peripheral sample in FHC mode. WS is an abbreviation of wrong sign. Retrieved from [95].

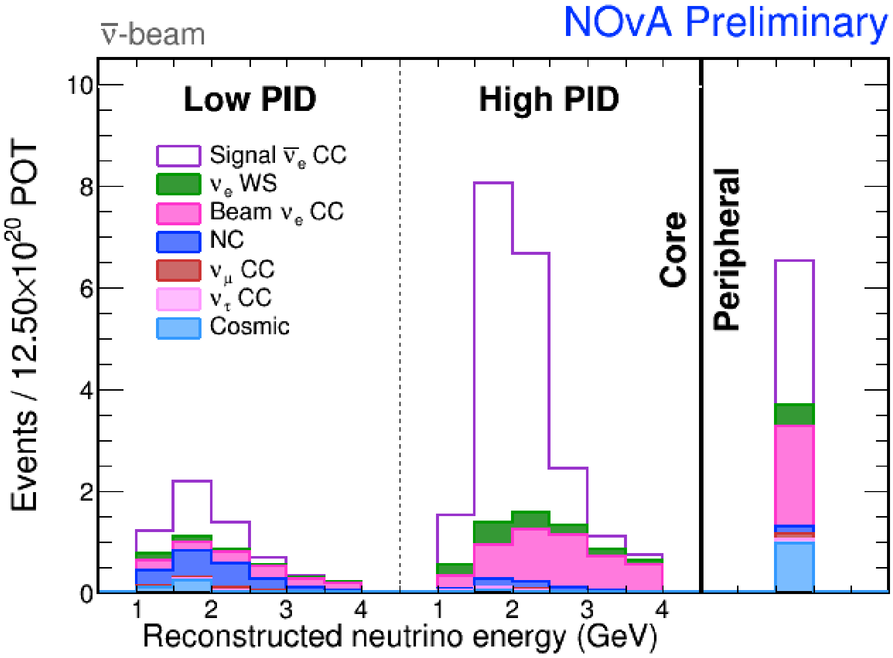

Figure 3.29: The ν_e FD predicted reconstructed neutrino energy spectra based on simulation used for the 2020 oscillation analysis split into two PID samples and a peripheral sample in RHC mode. WS is an abbreviation of wrong sign. Retrieved from [95].

in 2020 oscillation analysis are seen in figures 3.28 and 3.29 for FHC and RHC respectively. The plots show how the high PID bin has more signal compared to background than the low PID bin, and also that most of the background in the high PID bin is beam ν_es, which are irreducible. In turn, the low PID sample has more NC background compared to the high PID sample [52].

It is also notable that as the peripheral events failed the containment cut, their energy information is not used in the fit to the oscillation parameters as their full energy may have not been deposited in the detector. This sample has the most cosmic events as these events are mostly located near the detector edges [52].

3.5 Decomposition and Extrapolation

3.5.1 Decomposition

Decomposition is a technique used to predict the events seen at the FD using information from the ND. Thus, it is necessary to estimate the composition of each ND sample. For ν_μ ND sample, BEN (Beam Electron Neutrino) decomposition [96] is used to predict the number of ν_e beam events from decays of muon and kaons, which is a major background in the ν_e appearance analysis. ν_μ sample is also used to predict the number of survived ν_μ events as well as appeared ν_e events. Michel decomposition [97] uses Michel electrons, which are electrons produced in a muon decay. It is used for NC and ν_μ samples to calculate the number of Michel electrons in bins of reconstructed neutrino energy and CVN, which is used to rescale the ν_μ CC component in simulation to comply with the data [52] [62].

3.5.2 FD Extrapolation

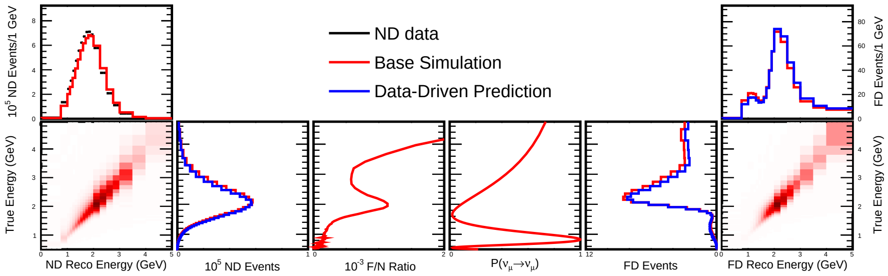

Figure 3.30: The extrapolation process for ν_μ data. First, the ND reconstructed energy is converted to true energy using a matrix. Using oscillation probabilities and ND/FD ratios a true far detector energy spectrum is obtained. Then, using both true energy spectras, the spectra to give final predictions at the FD are obtained. Retrieved from [98].

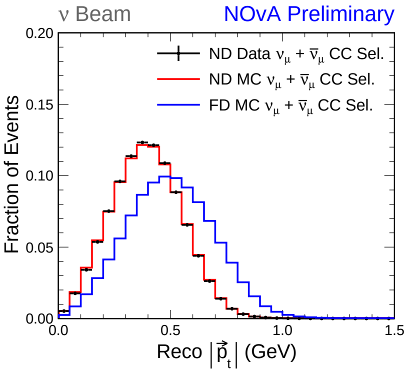

Figure 3.31: FHC Distribution of the fraction of selected events for each bin of reconstructed p_T for ND data (black), ND simulation (red) and FD simulation (blue). Retrieved from [99].

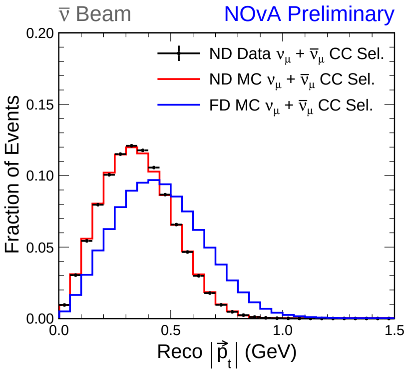

Figure 3.32: RHC Distribution of the fraction of selected events for each bin of reconstructed p_T for ND data (black), ND simulation (red) and FD simulation (blue). Retrieved from [99].

NOvA uses a method called extrapolation, where unoscillated spectrum at the ND is used to get predictions of the spectra at the FD, to reduce systematics that are correlated between the two detectors [12] [52] [67].

The extrapolation process starts with calculating predicted rates of neutrino events as well as background in the ND using simulation. The reconstructed neutrino energy spectrum is transformed to match with the true energy spectrum using a matrix method. The true ND spectrum is multiplied by FD / ND event ratio of the simulated samples in bins of true neutrino energy. This step takes into account differences in beam (ND closer to the source leading to wider range of beam axis angles), differences in selection efficiency and acceptance between ND and FD, and the differences between cross-sections of ν_μ and ν_e. Using oscillation probabilities, the predicted disappearance or appearance spectra can be obtained for FD. The oscillation probabilities include the matter effects with Earth's crust density assumed to be uniformly 2.84 g/cm^3. The predicted spectra can then be converted back to obtain reconstructed neutrino energy for FD. This process is illustrated in figure 3.30, and it is performed to all 4 quartile bins separately [12] [52] [67].

To further reduce systematics, extrapolation is performed separately for other variables than neutrino energy. Extrapolation was performed in bins of transverse momentum of the final-state lepton (p_T) in the latest NOvA 3-flavour analysis, which is especially useful for lowering the impact of cross-section uncertainties. In both ND and FD, the bin with the lowest p_T values will have events that are closer to each other kinematically compared to the bin with highest p_T values. In the FD,

p_T is divided into 3 bins and summed over and thus not used separately for the oscillation fit. The fraction of events for each value of p_T is seen in figures 3.31 and 3.32 for FHC and RHC respectively, showing how the values for p_T differ in ND and FD [12] [52] [67].

3.6 Systematic Uncertainties

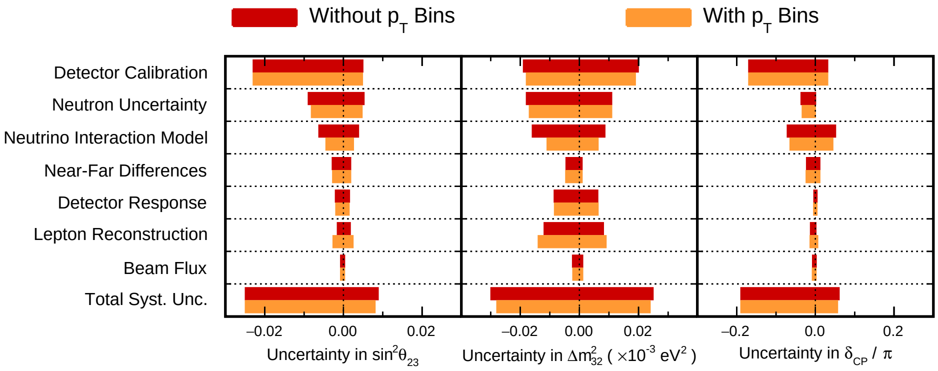

Figure 3.33: Summary of systematics at NOvA for the three measured oscillation variables. Red is the uncertainty level without and orange with p_T extrapolation. Retrieved from [12].

NOvA's 3-flavour oscillation analysis has over 60 individual systematic uncertainties. They can be grouped into 7 main groups: detector calibration, detector response, beam flux, lepton reconstruction, neutrino cross-section, neutron uncertainty and near-far differences. Systematics can also be grouped into correlated and uncorrelated. Correlated systematics apply to both detectors and uncorrelated are different for the two detectors. FD extrapolation, explained in section 3.5, helps to minimise detector correlated systematics like beam flux and neutrino cross-sections [67].

Systematic uncertainties are evaluated with different methods including varying the simulation via event re-weighting, repeating extrapolation as described in section 3.5, or simulating extra event samples. Reweighting is used in estimating beam flux and cross-sections. Systematic uncertainties dealing with neutron uncertainties are an example of systematics using recalculating via extrapolation. Simulating extra events is required for uncertainties such as calibration and light level (part of detector response) [12] [67].

3.6.1 Detector Calibration

The main source of systematic uncertainty in NOvA is detector calibration, the primary source being calorimetric energy scale uncertainty. It is determined using protons in the ND that describe the difference in data and MC. Whether this difference between data and simulation exists in the FD cannot be confirmed due to the lack of statistics leading to two energy scale uncertainties, absolute (the ND and the FD shifted in the same direction) and relative (shifted in opposite directions) [67].

The absolute energy scale can be improved by studying how the detector performs in a charged particle beam. Such a test beam programme was conducted with a scaled down NOvA near detector 1/5th of the size of the original detector. Results from this test run will improve the future analyses by decreasing the dominant NOvA systematics, and are currently being analysed [40].

Other detector calibration systematics are calibration shape and calibration drift. Uncertainty in calibration shape stems from differences in calibration in the middle and the end of the cell, which the calibration cannot account for perfectly. It highlights the differences between reconstructed and true energies for events close to the detector edges. The calibration drift systematic is due to the number of hits decreasing in the detectors, most likely due to the aging scintillator. This is corrected by reducing the number of light hits created in the simulation to match the decrease in number of hits [67].

3.6.2 Detector Response

The uncertainty in detector response includes light response in the cells in both detectors. When light is produced in the NOvA cell (see section 2.3.1), two types are produced: scintillation light and Cherenkov light (created when particles are moving faster than the speed of light in that particular medium). This is described in NOvA's light production model, which is tuned using ND simulation. The light model uncertainties describe the systematic uncertainties related to the production of both of these light types and how they are transmitted in the scintillator liquid and collected by the fibres [67] [74].

3.6.3 Beam Uncertainties

Flux uncertainties mostly arise from differences between data and simulation in the production of hadrons in the NuMI facility. Things like target position, beam spot size and position, and current of the horns are the causes of systematic uncertainties. As mentioned in section 3.1.1, a package called PPFX is used to constrain and predict uncertainties arising from beam flux [52] [67].

3.6.4 Neutron Uncertainties

Neutron uncertainties arise from the poor modelling of fast neutrons, (kinetic energy of 1 MeV or more) which especially affects $\overline{\nu}_\mu$ samples. This is due to $\overline{\nu}_\mu$ interactions frequently containing neutrons. Also, a deficit of low energy (less than 20 MeV) neutrons are observed in the data compared to the simulation in previous NOvA analyses. The cause for this discrepancy has not yet been found. This is modelled by scaling the contribution of these low energy neutrons to the hadronic energy to calculate the neutron systematic uncertainty [67] [74] [100].

3.6.5 Cross-section Uncertainties

Cross-section uncertainties are calculated from cross-section models, using GENIE event re-weighting framework in the calculations. The different models are described in greater detail in [12].

3.6.6 Lepton Reconstruction

Lepton reconstruction mostly consists of uncertainty in the muon energy scale. There are uncertainties arising from the use of GEANT and the Bethe-Bloch equation (see equation 3.2.1) relevant to both detectors' muon energy scale. There are also muon energy uncertainties that arise from differences between the detectors, such as from the ND muon catcher (described in section 2.3.6) that is not present in the FD [74].

3.6.7 Near-Far Differences

This uncertainty is due to differences in the detectors, the main one being the size difference. Even though these systematics are mostly cancelled between the detectors because of the use of extrapolation, especially the method of p_T extrapolation, some uncertainties remain. This uncertainty especially affects the ν_e signal since the extrapolation is done using the ν_μ ND sample. ν_μ and ν_e have different containment and PID performances resulting in kinematic differences between the samples.

This systematic effect is estimated by re-weighting the muon neutrino samples in ND data and MC to match the FD electron neutrino sample for four kinematic variables. These are true four momentum transfer squared, reconstructed angle of the primary lepton with respect to the beam and, the total reconstructed transverse momentum of the event. The re-weighted ND spectra are then passed through the extrapolation procedure and the resulting FD distributions are compared to the nominal spectra [52] [74].

3.7 Oscillation Fits

NOvA performs a fit to the oscillation parameters using frequentist statistical methods. The predicted energy is fit to the observed data from the FD assuming each analysis bin i is described by Poisson statistics. The probability P_i of observing O_i events with expectation E_i for an oscillation parameter θ_i can be written as [52] [101]:

$$P(O_i|E_i(\theta_i)) = \frac{E_i^{O_i} e^{E_i}}{O_i!}. \tag{3.7.1}$$

The Poisson likelihood is an expression of joint probability across all bins, which is expressed as [101]:

$$L(\bar{\theta}) = \prod_{i \in bins} P(O_i|E_i(\theta_i)) = \prod_{i \text{ in bins}} \frac{e^{-E_i} E^{O_i}}{E_i!} \tag{3.7.2}$$

The oscillation parameters are calculated as best fit values that maximise this likelihood ratio. Maximum likelihood is a common technique in statistics used to estimate values of parameters for a given data sample [102]. Likelihood ratio λ can be used as a goodness-of-fit check and is defined as [52]:

$$\lambda(\bar{\theta}) = \prod_{i \in bins} \frac{e^{-E_i} E^{O_i}}{E_i!} - \prod_{i \in bins} \frac{e^{-O_i} O^{O_i}}{O_i!} \tag{3.7.3}$$

The likelihood ratio can be written in terms of $-2ln\lambda$ which follows the chi-square distribution when data can be considered to have Gaussian errors [52]:

$$-2ln\lambda(\overrightarrow{\theta}) = 2 \sum_{i \in bins} \left[E_i(\overrightarrow{\theta}) - O_i + O_i ln(\frac{O_i}{E_i(\overrightarrow{\theta}})] \right] \tag{3.7.4}$$

This equation does not take into account an extra degree of freedom stemming from the systematic degrees of freedom $\overrightarrow{\delta}$. Adding it the equation gives [74]:

$$\chi^2 = -2ln\lambda(\overrightarrow{\theta}, \overrightarrow{\delta}) = 2 \sum_{i \in bins} \left[E_i(\overrightarrow{\theta}, \overrightarrow{\delta}) - O_i + O_i ln(\frac{O_i}{E_i(\overrightarrow{\theta}, \overrightarrow{\delta})}) \right] + \sum_{j=1}^{N} \frac{\delta_j^2}{\sigma_j^2} \tag{3.7.5}$$

where σ_j values describe 1σ ranges used to assign nuisance parameter penalty terms for each systematic [62]. The values of the parameters $\overrightarrow{\theta}$ and $\overrightarrow{\delta}$ that minimise the equation 3.7.5 gives the best fit values for the oscillation parameters [67]:

$$\Delta\chi^2 = \chi^2 - \chi^2_{min}. \tag{3.7.6}$$

The confidence intervals for each oscillation parameter are also defined using equation 3.7.6. For 2D oscillation fits, these are 1σ (68%) confidence contour defined as $\Delta\chi^2 < 2.3$, 2σ as $\Delta\chi^2 < 6.18$ and 3σ as $\Delta\chi^2 < 11.38$ for the number of parameters NOvA is fitting.

All of the oscillations fits use Feldman-Cousins (FC) unified technique [103] [104] to determine the confidence integrals. It is widely used in Poissonian processes to limit boundaries [12] [74].

3.8 Latest 3-Flavour Oscillation Results

As mentioned in the beginning of the chapter, the latest analysis included all NOvA neutrino channels producing a joint fit of $\nu_\mu \to \nu_e$, $\nu_\mu \to \nu_\mu$ and corresponding anti-neutrino channels. Improvements compared to earlier analyses included 50 % beam exposure increase and better simulation and analysis techniques [12].

Fit Parameter	**Value**
Baseline length L	$810\ km$
Average matter density ρ	$2.84\ g/cm^3$
Δm^2_{21}	$7.53 \times 10^{-5} eV^2$ [10]
$sin^2(\theta_{12})$	0.307 [10]
$sin^2(\theta_{13})$	0.0210 ± 0.0011 [10]

Table 3.7: Parameters used in the latest oscillation parameter fits, values from [12].

The constrained parameters that were used in the fit are listed in table 3.7. The parameters Δm^2_{32}, $sin^2(\theta_{23})$ and δ_{CP} were varied with no constraints, and all the systematic uncertainties were assigned penalty terms equal to the square of the number of standard deviations by which they vary from the assigned nominal value. As mentioned in section 2.4, NOvA is sensitive to $sin^2(\theta_{13})$ but it is more accurately measured by reactor experiments so it is constraint in the analysis [12] [52].

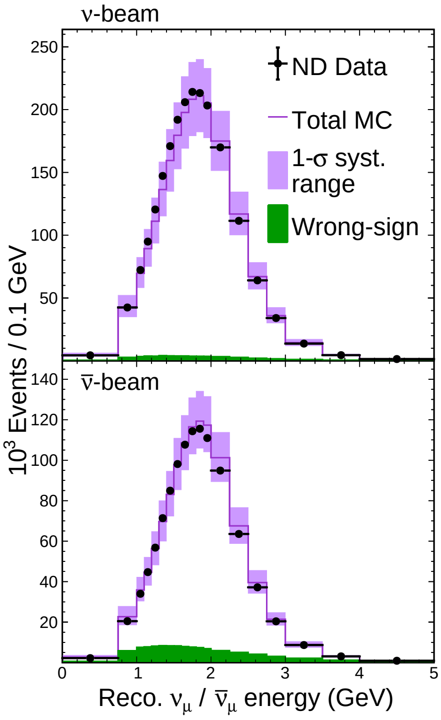

Figure 3.34: ν_μ ND spectra for the 2020 3-flavour oscillation analysis. Retrieved from [12].

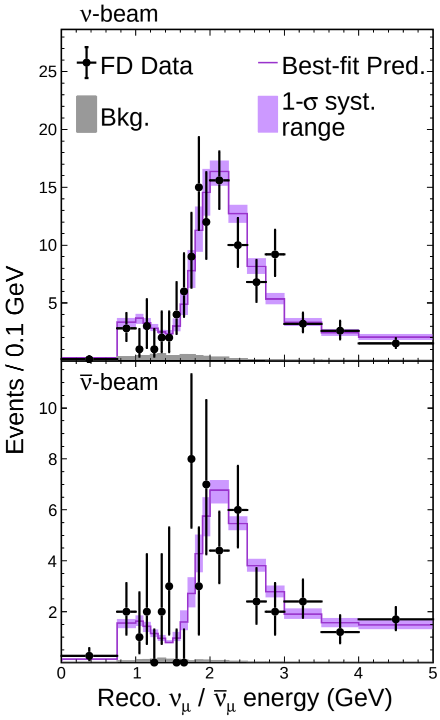

Figure 3.35: ν_μ FD spectra for the 2020 3-flavour oscillation analysis. Retrieved from [12].

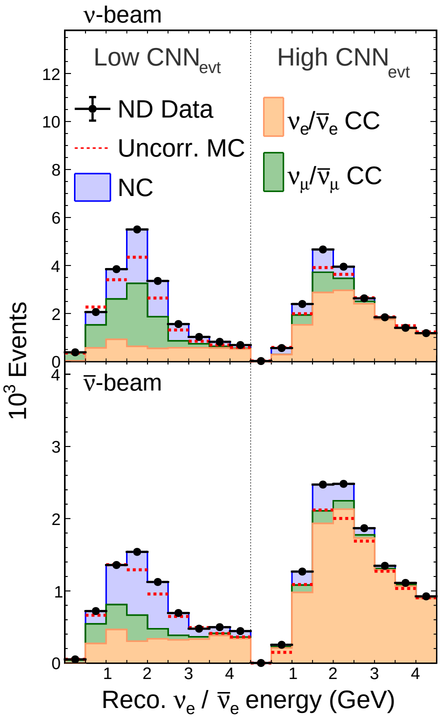

Figure 3.36: ν_e ND spectra for the 2020 3-flavour oscillation analysis. Retrieved from [12].

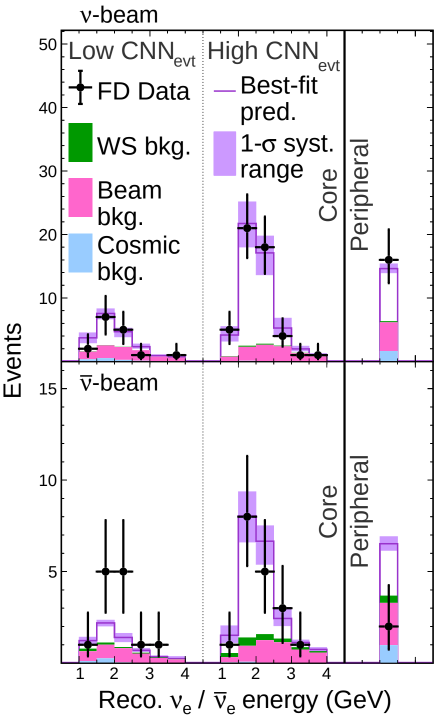

Figure 3.37: ν_e FD spectra for the 2020 3-flavour oscillation analysis. Retrieved from [12].

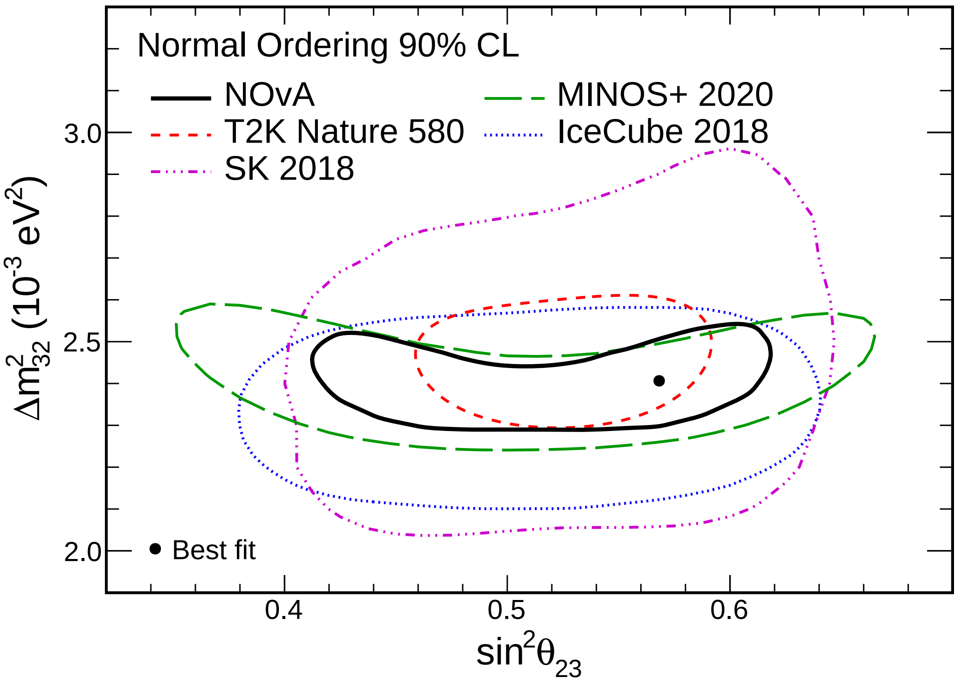

Figure 3.38: The best fit and 90% confidence level region for oscillation parameters Δm^2_{32} and $sin^2\theta_{23}$. Results from other experiments measuring the same oscillation parameters are included for comparison. The contours include a Feldman-Cousin correction. Retrieved from [12].

3.8.1 Final Data Samples

Figures 3.34 and 3.35 show the total reconstructed neutrino energy spectra for selected ν_μ event for ND and FD respectively. The same spectra for ν_e is seen in figures 3.36 and 3.37 for both ND and FD. For ν_e the energy is shown for all three bins separately: low and high PID as well as the peripheral bin. Even though the fit is done in four quantiles for ν_μ (as explained in 3.4.1), the final plots are put into one energy spectra.

3.8.2 Results

The final contours for oscillation parameters Δm^2_{23} and $sin^2\theta_{23}$ is seen in figure 3.38 (assuming normal mass ordering). The best fit point and confidence intervals were calculated using maximum likelihood technique described in section 3.7. Other experiments measuring the same oscillation parameters are also included for comparison. The final fit for δ_{CP} against $sin^2\theta_{23}$ for both normal and inverted mass ordering (see section 1.5.2) is seen in figure 3.39. Results from T2K are included for comparison.

All of the results from the analysis are summarised in table 3.8 including all versions of upper and lower octants for $sin^2(\theta_{23})$, and normal

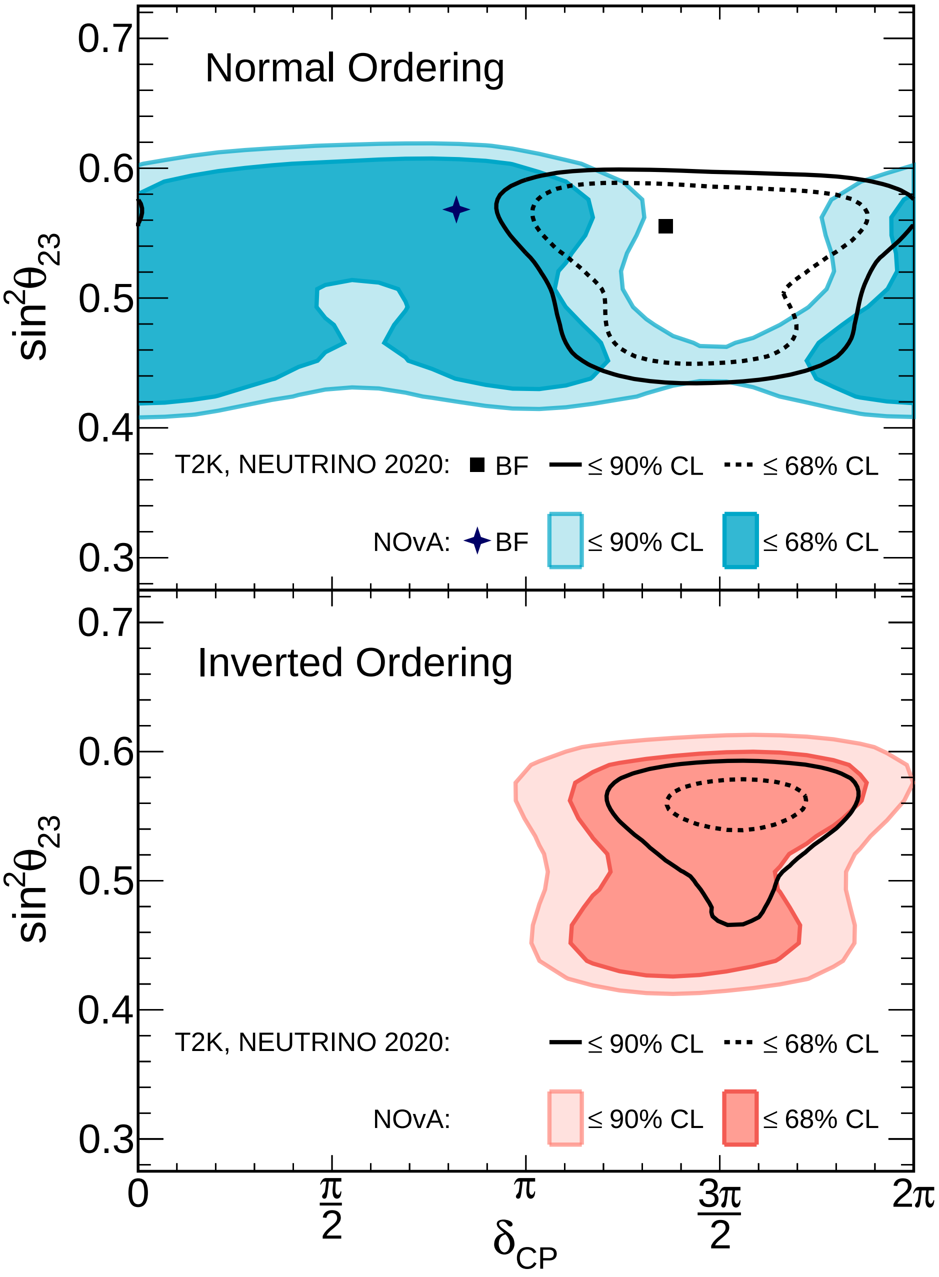

Figure 3.39: **The upper plot**: the best fit as well as 90% and 68% confidence level regions for δ_{CP} and $sin^2\theta_{23}$ for normal mass ordering for both NOvA as well as T2K. All confidence level regions include FC corrections. **The lower plot**: Inverted mass ordering, 90% and 68% confidence level regions, T2K results included for comparison. All confidence level regions include FC corrections. Retrieved from [12].

and inverted mass ordering. The results show a preference for normal ordering and upper octant for $sin^2(\theta_{23})$. Assuming these conditions, the neutrino oscillations parameters are measured as $(2.41 \pm 0.07) \times 10^{-3} eV^2$ and $0.57 + 0.03 - 0.04$ for Δm^2_{32} and $sin^2(\theta_{23})$ respectively.

The results disfavour combinations leading to strong asymmetry in ν_e vs. $\overline{\nu}_e$ appearance in inverted ordering with $\delta_{CP} = \pi/2$ excluded with

Parameter	**NO, UO**	**NO, LO**	**IO, UO**	**IO, LO**
$\Delta m^2_{32}\ 10^{-3} eV^2$	$+2.41 \pm 0.07$	$+2.39$	-2.45	-2.44
$sin^2\theta_{23}$	$0.57^{+0.03}_{-0.04}$	0.46	0.56	0.46
$\delta_{CP}(\pi)$	$0.82^{+0.27}_{-0.87}$	0.07	1.52	1.41
Rejection Significance	-	1.1σ	0.9σ	1.1σ

Table 3.8: NOvA 2020 oscillation analysis results, all values from [12]. NO stands for normal ordering, IO for inverted ordering, UO for upper octant and LO for lower octant. Values from [12].

confidence of more than 3σ and $\delta_{CP} = 3\pi/2$ being disfavoured at 4σ level. However, the 90% confidence level regions cover all δ_{CP} values for all combinations of octant and mass ordering. Hence, the data cannot prove whether neutrino oscillations exhibit CP-violation at 5σ level [12].

As it is seen in figure 3.39, both T2K and NOvA slightly prefer normal ordering, however together in the NO space T2K's best fit point is in the area that NOvA disfavours with only a very small region of overlap. This is mostly due to T2K having observed a more pronounced asymmetry in ν_e and $\overline{\nu}_e$ appearance. It does, however, seem that the experiments have overlapping contours in inverted ordering, meaning a joint fit between these experiments would prefer inverted ordering [12] although individually they prefer normal ordering. It has been suggested that these differences are due to additional nonstandard matter effects, explained more in [105] [106]. In these nonstandard interaction (NSI) models there are additional matter effects in neutrino oscillations that stem from additional interactions with e.g. sterile neutrinos or unknown heavy particles. It is also possible that the discrepancy in the results is due to statistical fluctuations or an unknown systematic error [105] [106].

Chapter 4

ν_μ Event Selection

This chapter describes in more detail the event selection used in the 2020 3-flavour analysis in NOvA, paying particular attention to where improvement could be made. The events that are rejected from the disappearance analysis are introduced, describing some of their properties in an effort to understand why they are not selected and how they could be recovered back into the analysis.

4.1 Improvements to the Current Analysis

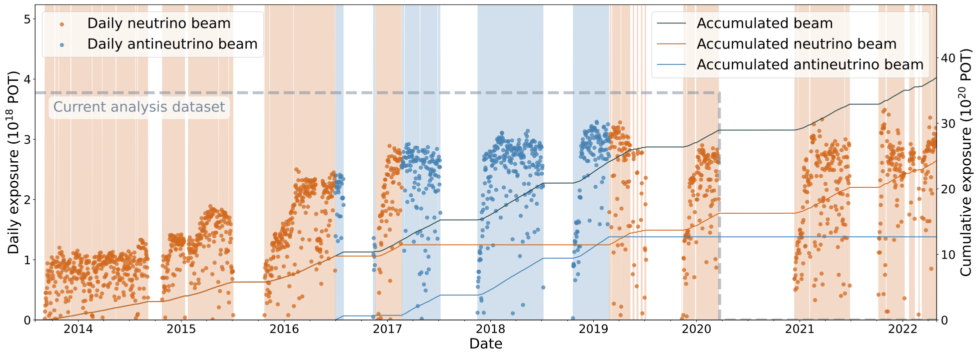

Figure 4.1: Daily POT (protons on target) recorded by NOvA from the start of data taking in 2014 to 2022 with neutrino beam (FHC, orange) and antineutrino beam (RHC, blue). Total accumulated POT is marked in grey. The plot also shows the point of up to which the current analysis uses data. Retrieved from [107].

Multiple improvements to the next 3-flavour analysis are underway in NOvA, covering many areas in the analysis chain. The main improvement to the analysis will come from the extra collected data. The total recorded protons on target (POT) for NOvA is seen in figure 4.1. The plot shows all the collected POT in both FHC and RHC modes since the start of data taking in 2014 and the point up to which the data has been analysed. This year, NOvA is predicted to collect enough data in neutrino mode to add 13.6×10^{20} POT to the analysis, doubling the current FHC data set. After this has been achieved, a switch to RHC mode is planned, but the timeline of this remains uncertain.

One of the ways the future results will be improved is using new fitting techniques or upgrading the existing ones for better ones. There will also be updates to how the energy is estimated for events, and systematics will be reviewed and potentially changed. Also, the selection for both ν_e and ν_μ will be re-optimised to account for new data.

NOvA has been taking data since 2014 and has published 6 iterations of 3-flavour oscillation analyses. Many of the steps in the analysis chain have been optimised to a very precise level. Large upgrades to the analysis are difficult to achieve at this stage, however many small improvements together with new recorded data can bring significant improvement to the results.

4.2 Current ν_μ Selection

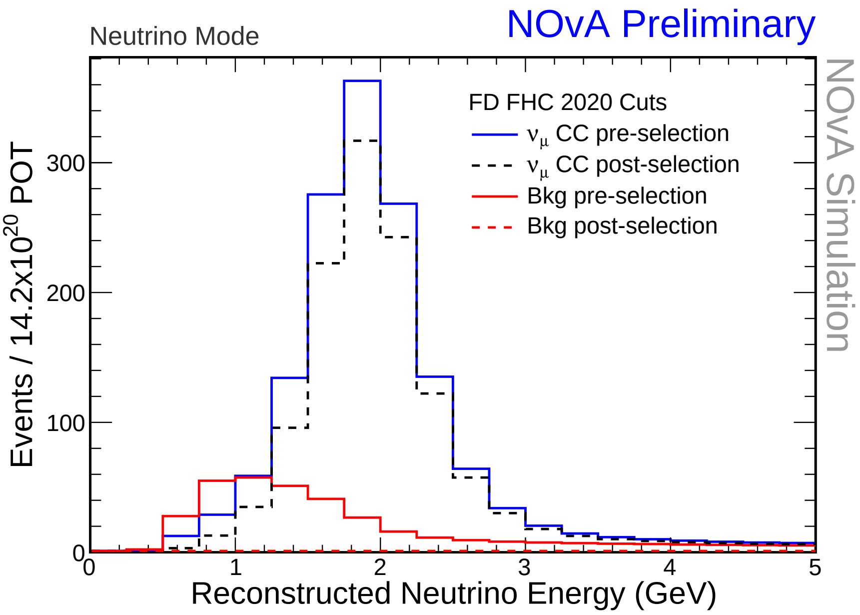

Figure 4.2: The current NOvA 3-flavour oscillation analysis selection for ν_μ CC disappearance using FD FHC decaf files (quality and containment cuts applied). In blue is the number of simulated events before selection and in black dashed line events after cosmic rejection and PID cuts. Red and red dashed line are the same for background. This plot is made with no oscillations applied.

The goal of this analysis is to investigate how many more ν_μ CC events could be added into the disappearance analysis from events that

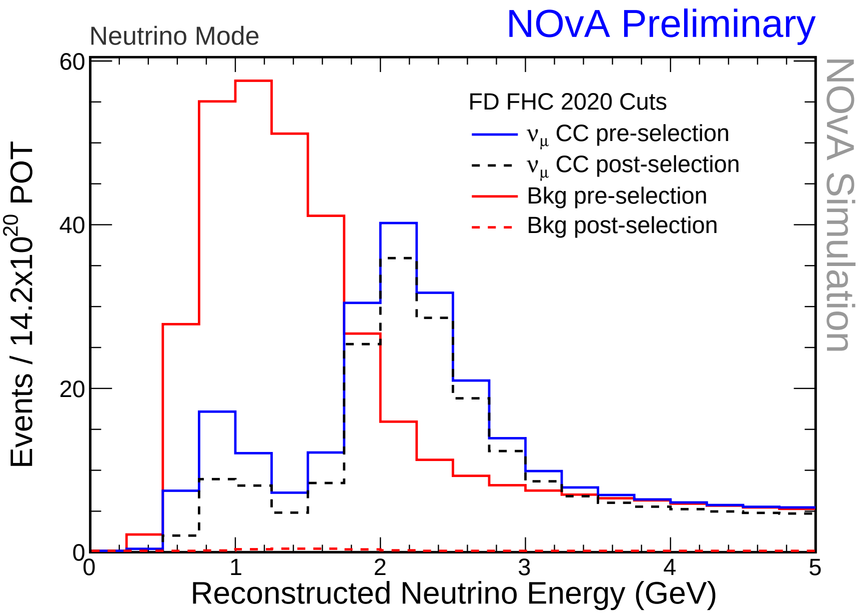

Figure 4.3: The same figure as 4.2 expect oscillations are applied.

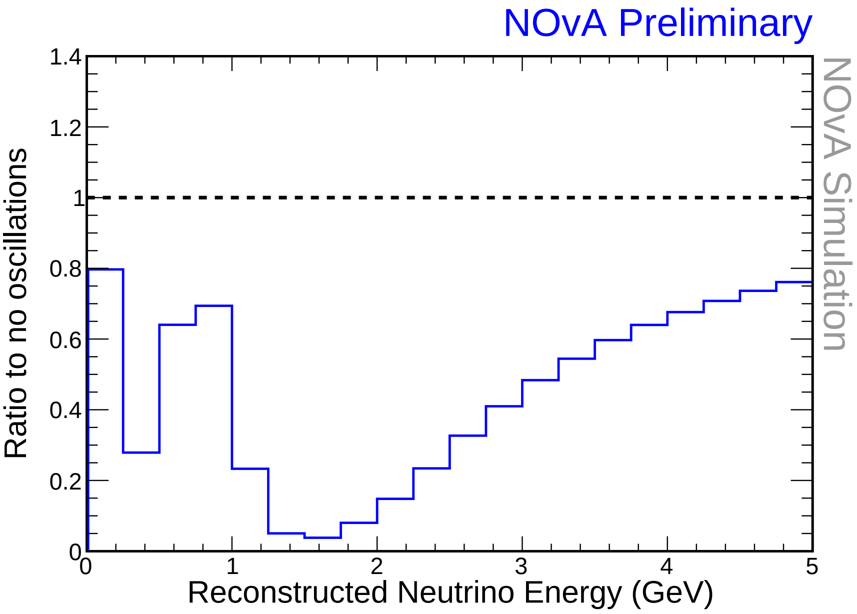

Figure 4.4: Ratio of oscillated to non-oscillated for ν_μ CC 2020 oscillation analysis events using FD FHC simulation. The plot clearly shows a dip at the oscillation maximum of 1.8 GeV.

are currently rejected by the analysis cuts. Hereafter, signal refers to ν_μ CC events and background to all other events unless otherwise specified.

This analysis focuses on the last two cuts and starts with simulation files that already have the first two cuts (basic quality and containment, see section 3.3) applied, as well as a loose cosmic rejection cut to remove some cosmic-like events. The loose cosmic cut is defined as Cosmic BDT score > 0.4 (see figures **??** and 3.17). These files are called decafs, and will be referred as such going forward.

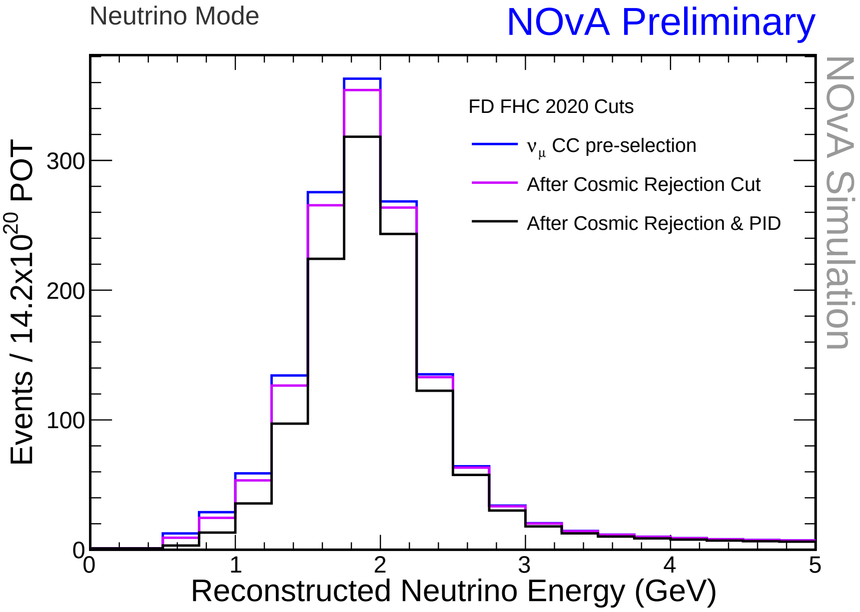

Figure 4.5: A plot showing how much signal is discarded by cosmic rejection and PID cuts separately using FD FHC simulation. No oscillations applied.

The use of decafs both speeds up the analysis due to their smaller size and makes sure the added events are of better quality. Events removed by the basic quality are of poor quality and have unwanted features such as having too few hits in a track or vertical tracks. For events that fail containment cuts, reconstructed neutrino energy cannot be defined. These events would have to have similar treatment to the ν_e peripheral sample, which is not the goal of this analysis.

Figures 4.2 and 4.3 show the reconstructed neutrino energy spectrum before and after the current 2020 3-flavour cuts are applied for FD FHC decafs. The oscillations were calculated using the same values as for the latest 3-flavour analysis (see table 3.7), and parameters θ_{23}, Δm^2_{32} and δ_{CP} were defined as $\frac{\pi}{4}$, 2.44×10^{-3} and 0 respectively.

A plot showing the ratio of oscillated to non-oscillated events is also included at figure 4.4 demonstrating clearly the oscillation dip at approximately 1.8 GeV. From these figures, it is clear that the cuts remove almost all the background but also some signal. For the oscillated case, there are 250 events in the decaf files, out of which 200 are chosen. The number of background is 356, out of which 3 are chosen. So it is found that 20% of signal events are removed from the analysis by the cuts.

To investigate the events further, figures 4.5 and 4.6 show how many signal events are rejected by both cosmic rejection and PID cuts separately. These plots clearly show that a larger number of signal is cut by the PID cuts. This can be interpreted that most of the signal that could

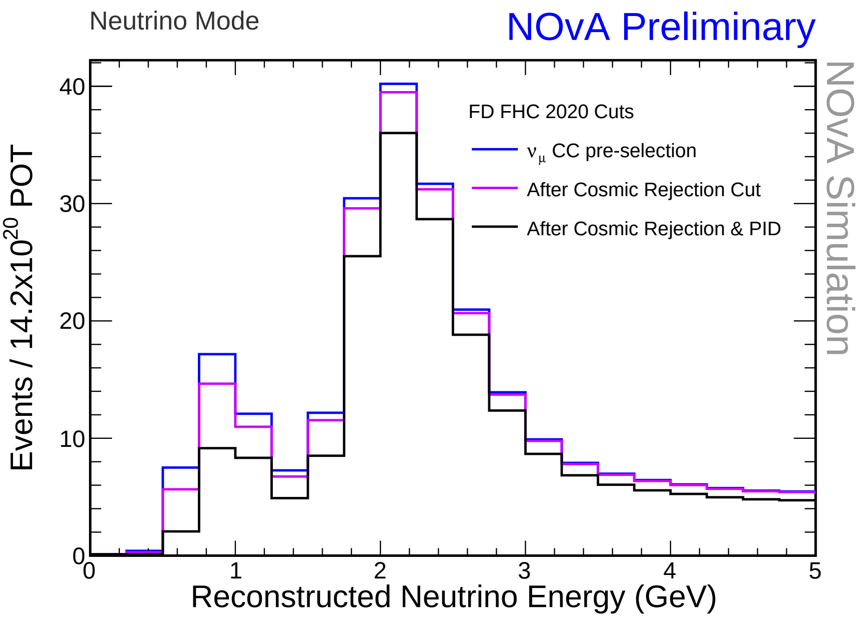

Figure 4.6: Same as figure 4.5 expect oscillations are applied.

be gained are events that are misidentified as something else than ν_μ CC signal instead of cosmic events.

It is also evident from these plots that more events are discarded at the lower energies. It would be particularly useful to try to recover some of these events back, since currently for events with energy less than 1 GeV more than 50% of signal is lost.

4.2.1 Efficiency and Purity

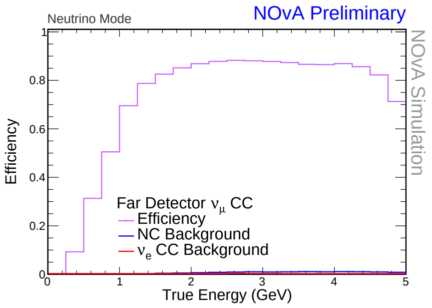

Figure 4.7: FD FHC signal efficiency as well as mis-identification efficiency for NC and ν_e CC events for 2020 3-flavour disappearance analysis.

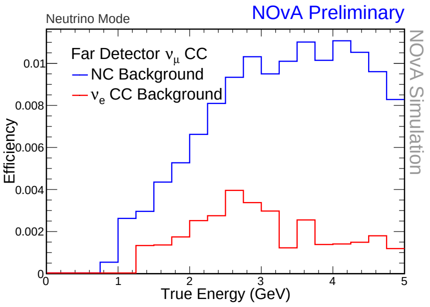

Figure 4.8: FD FHC mis-identification efficiency for NC and ν_e CC events for 2020 3-flavour disappearance analysis.

To further describe the selected 2020 sample, one can evaluate purity and efficiency. Purity is defined as the percentage of desired neutrino

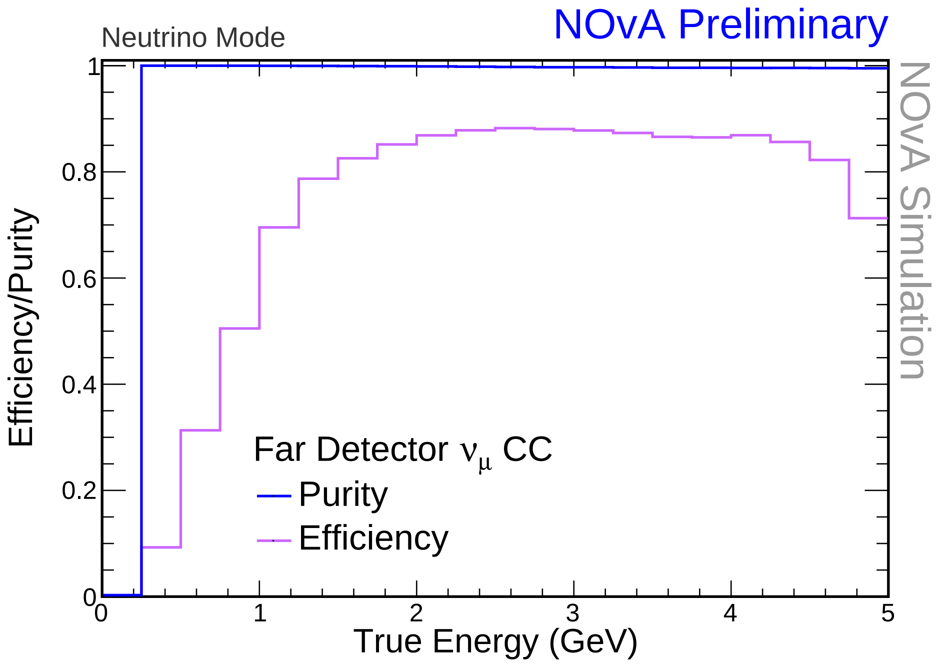

Figure 4.9: FD FHC Signal efficiency and purity for 2020 3-flavour disappearance analysis.

events in the full selected sample, i.e. for ν_μ CC events it is defined as:

$$\frac{\nu_\mu \text{ CC events that pass the cuts}}{\text{all events that pass the cuts}}. \tag{4.2.1}$$

Efficiency is a percentage of how well the selection chooses the wanted neutrino events, so for ν_μ CC events the definition is:

$$\frac{\text{selected } \nu_\mu \text{ CC events}}{\text{all } \nu_\mu \text{ CC events}}. \tag{4.2.2}$$

Figures 4.7 and 4.8 show the efficiencies for three samples: signal and two mis-identification efficiencies: ND and ν_e CC (number of selected NC and ν_e CC events out of all of those events). The efficiency for selecting signal is worse for low energies and increases to the highest value at around the oscillation maximum. There is also a small dip in efficiency just before 5 GeV, but this is much smaller than the difference in the lower energy region. One reason for this is that high energy muons are easier to detect, and thus lower energy events are missed disproportionately. Also, the dip at around 5 GeV is due to only the sample of events between 0 and 5 GeV being trained on the CVN, which causes a drop in efficiency in these high energy events.

Both the background mis-identification efficiencies are very low which is expected from figure 4.3 as it has almost no background after cuts. The mis-identification efficiency is higher for NC events, which is due to the

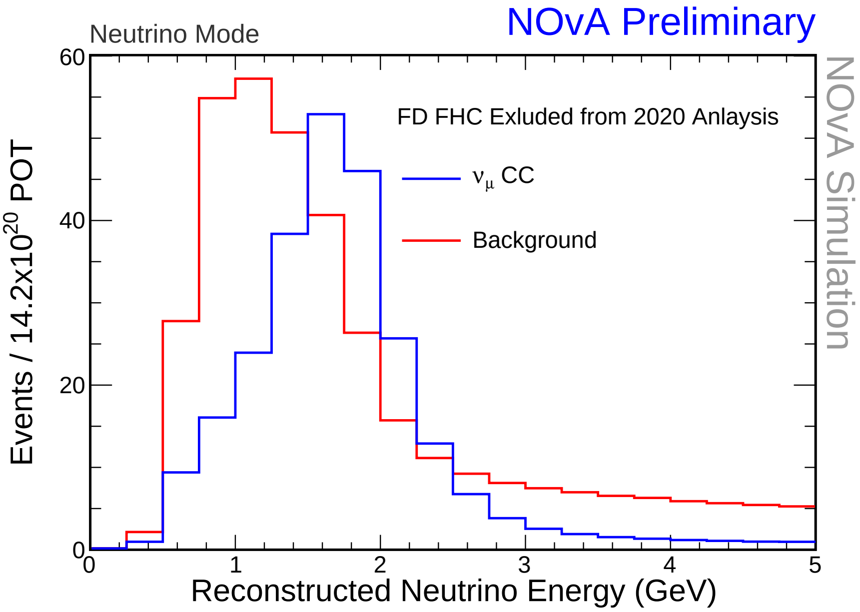

Figure 4.10: The reconstructed energy spectrum for events rejected by the current analysis cuts, non-oscillated.

majority of the background expected to be NC events. The difference is small, as can be seen in figure 4.8.

The purity along with efficiency is seen in figure 4.9. Purity is almost 100% for the whole sample, which is expected from having such a small background. The sample used for fitting to evaluate the oscillation parameters is thus effectively all signal.

4.3 Rejected Events

The reconstructed neutrino energy spectrum of only the events not selected by the current analysis is seen in figure 4.10 for the non-oscillated and in figure 4.11 for the oscillated case. This shows that a maximum of 50 events is the limit that can be recovered into the analysis (from the oscillated version), although it is expected to be lower since achieving 100% recovery rate is not possible. The number of background events in this sample is 353. Figure 4.11 also clearly shows that low energy events (energy less than 1.5 GeV) are cut out proportionally more than higher energy events as it was predicted by plots shown previously.

A plot showing a ratio of oscillated to non-oscillated is seen in figure 4.12 for the rejected events. If compared to the version of the plots for the 2020 analysis (see figure 4.4), the oscillation dip stays the same, however there are some differences in the lower energy events.

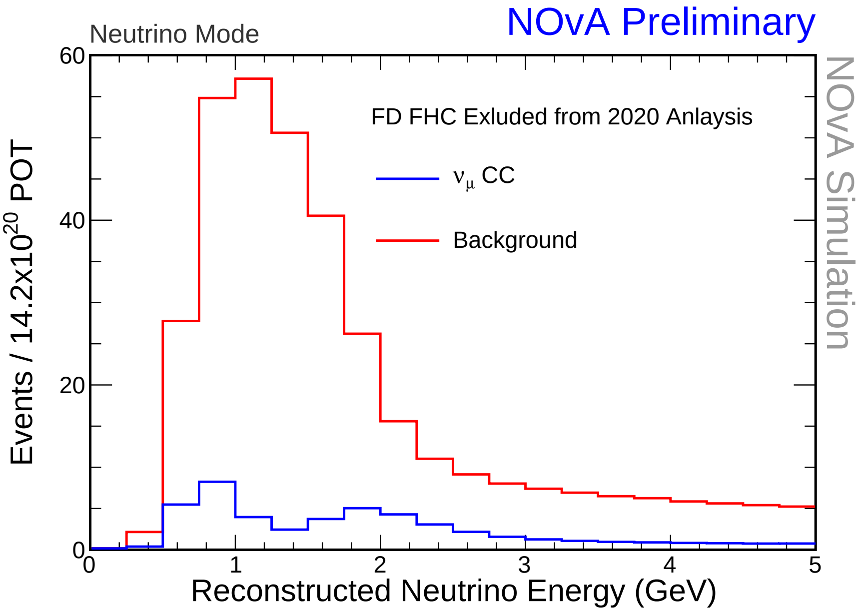

Figure 4.11: The reconstructed energy spectrum for events rejected by the current analysis cuts, oscillated.

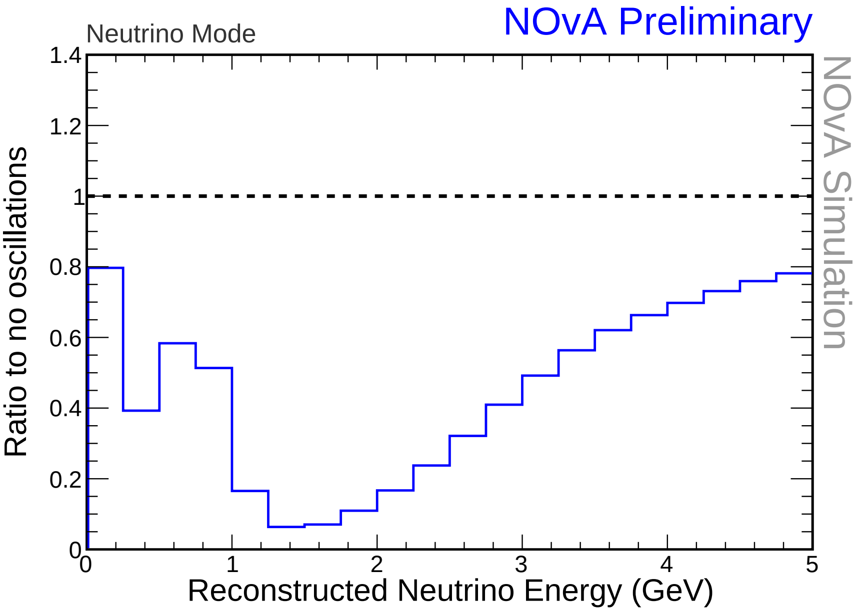

Figure 4.12: Ratio of oscillated to non-oscillated for ν_μ CC for the signal rejected by the current cuts. The plot shows a dip at the oscillation maximum of 1.8 GeV.

True Energy

The true neutrino energy spectrum of the events rejected by the current cuts is seen in figure 4.13 for non-oscillated and in figure 4.14 for the oscillated case. Reconstructed neutrino energy is measured by the detectors, whereas neutrino oscillations happen in the true neutrino energy space. The difference between the true and reconstructed energy comes from the energy resolution of NOvA detectors. The main difference is in

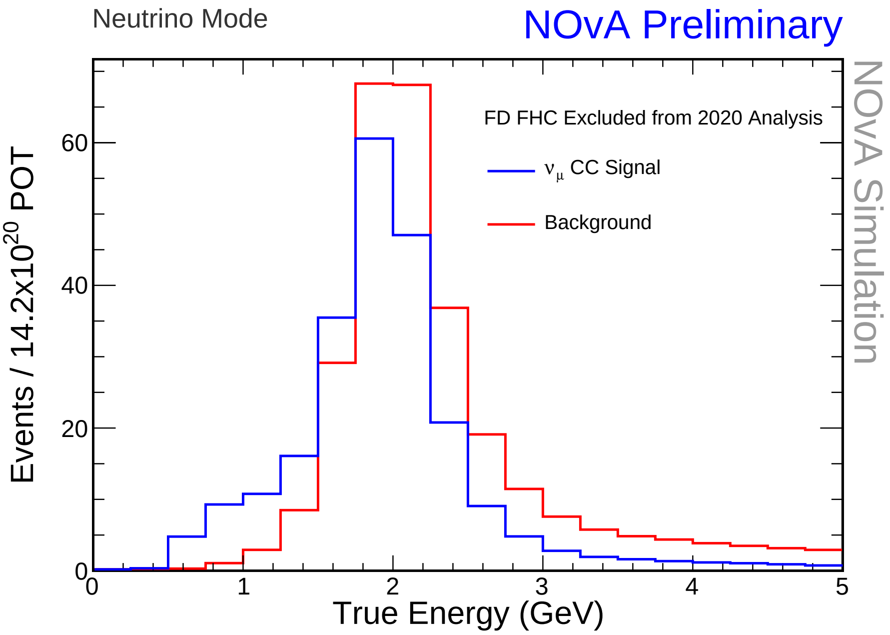

Figure 4.13: The true energy spectrum for events rejected by the current analysis cuts, non-oscillated.

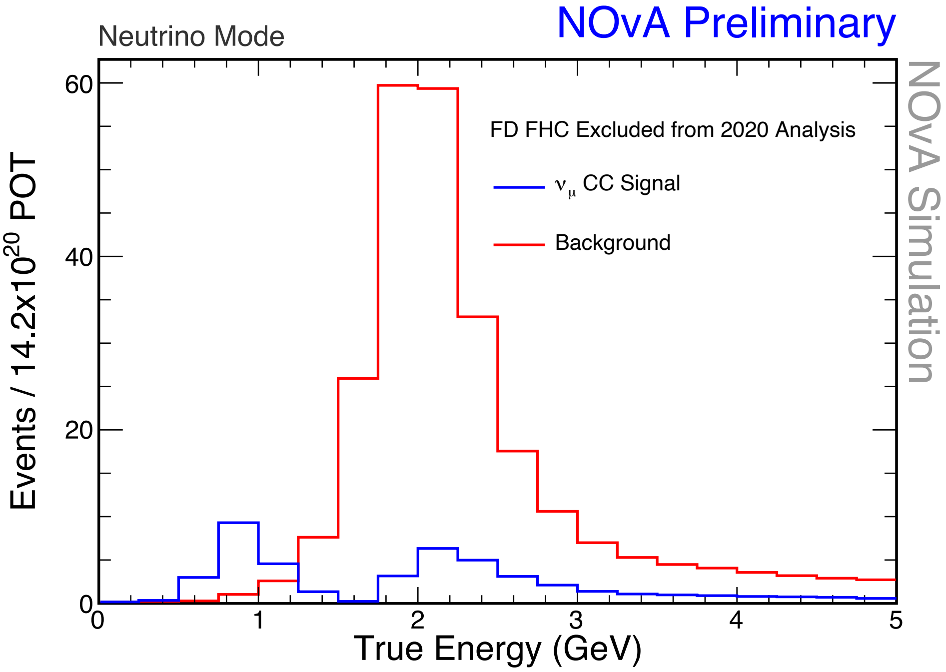

Figure 4.14: The true energy spectrum for events rejected by the current analysis cuts, oscillated.

the background, which stems mostly from under-estimation of the energy of NC events.

4.3.1 Event Properties

Looking more closely into the types of events that are rejected from the analysis, there are aspects that can be learned and possibly used to

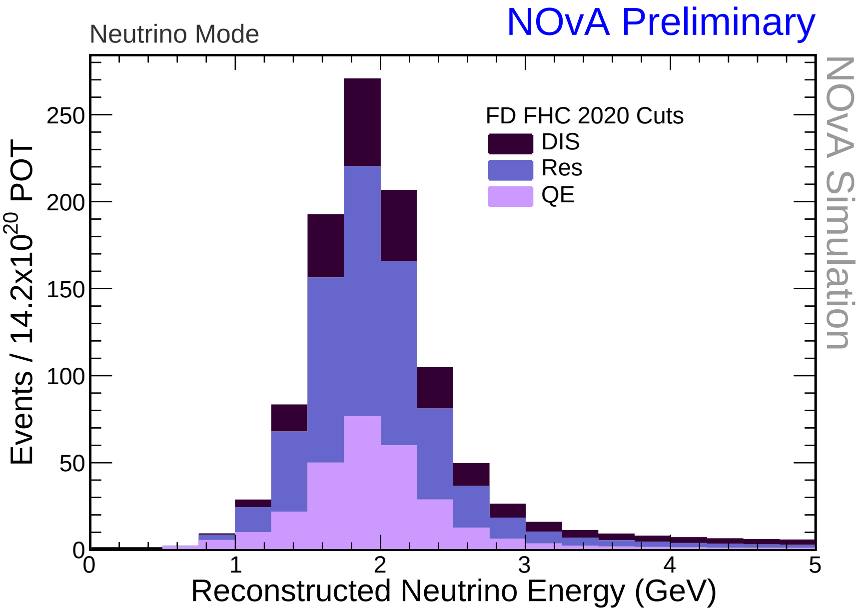

Figure 4.15: 2020 selected FD FHC signal broken down by the type of neutrino interactions: deep inelastic scattering (DIS), resonant pion production (RES) and quasi-elastic scattering (QE), non-oscillated.

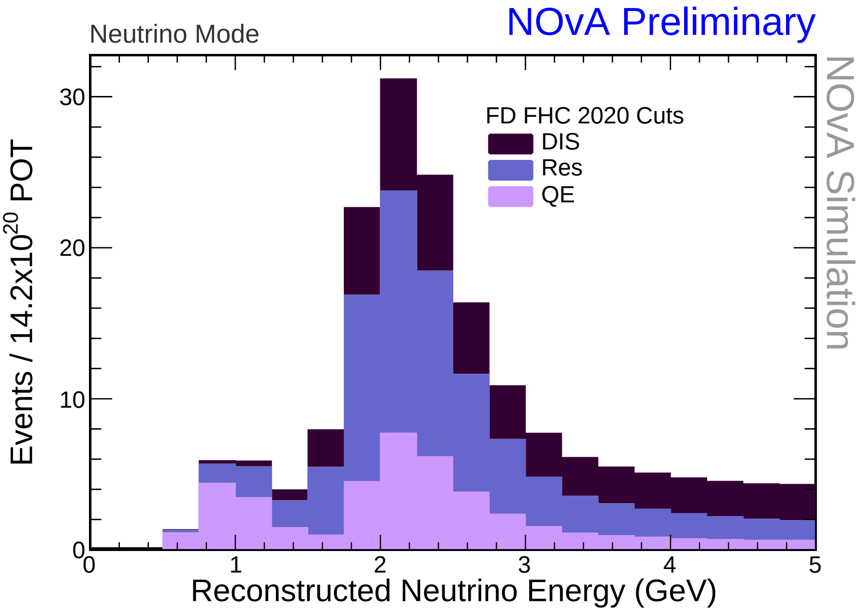

Figure 4.16: 2020 selected FD FHC signal broken down by the type of neutrino interactions: deep inelastic scattering (DIS), resonant pion production (RES) and quasi-elastic scattering (QE), oscillations applied.

aid with recovering the events. Figures 4.15 and 4.16 show the events broken into three types of neutrino events: quasi-elastic scattering (QE), resonant pion production (RES) and deep inelastic scattering (DIS) (see section 1.2) for 2020 selection (both oscillated and not oscillated). For the rejected events, the same figures are seen in figures 4.17 and 4.18 for non-oscillated and oscillated respectively. From the oscillated plots, the 2020 sample has 43 QE, 79 RES and 52 DIS, plus 26 events not belonging

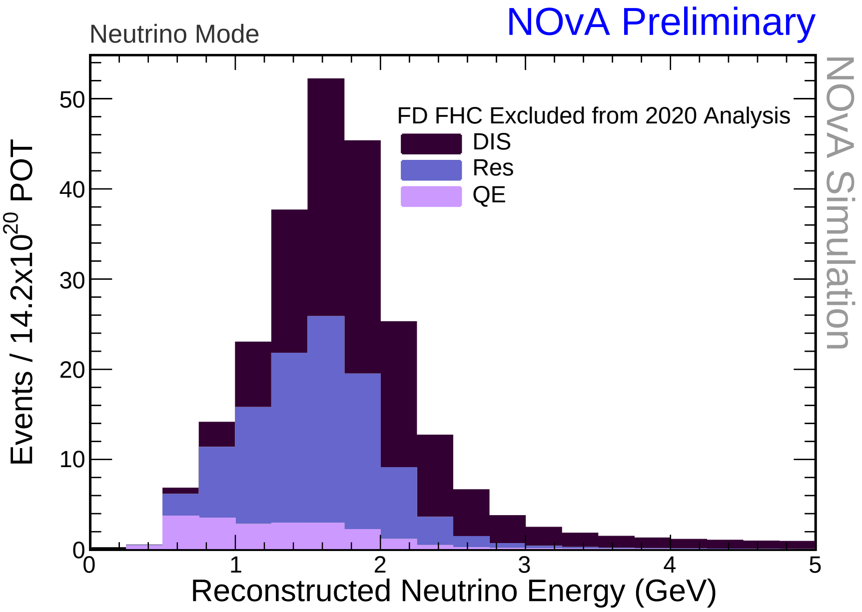

Figure 4.17: Events rejected from the current disappearance analysis for FD FHC, divided into three types of neutrino interactions: deep inelastic scattering (DIS), resonant pion production (RES) and quasi-elastic scattering (QE), non-oscillated.

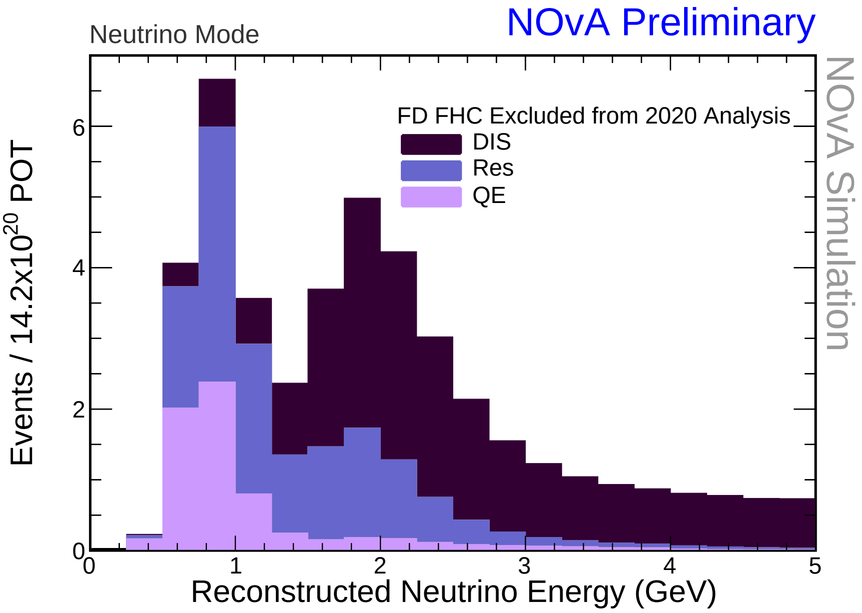

Figure 4.18: Events rejected from the current disappearance analysis for FD FHC, divided into three types of neutrino interactions: deep inelastic scattering (DIS), resonant pion production (RES) and quasi-elastic scattering (QE), oscillations applied.

to any of these categories. After the cuts, there are 7 QE, 14 RES, 23 DIS and 6 other (non-categorised) events remaining.

The non-categorised events can either be QE, RES or DIS that were poorly reconstructed and not correctly identified, or they can be so called meson exchange current (MEC) events. When neutrinos interact with a

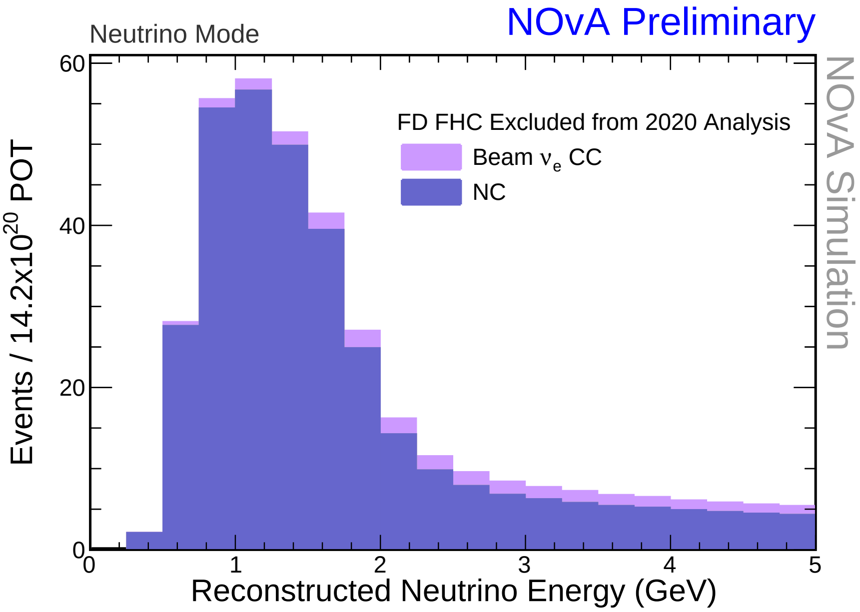

Figure 4.19: Background in the rejected sample by component: ν_e CC and NC events. All other events only contribute negligible amount.

nucleus, the nucleons also interact with each other according to the underlying substructure. In MEC, pairs of nucleons exchange a virtual meson, such as π^0, and the neutrino interacts with one of the nucleons. Both nucleons are ejected from the nucleus leaving two holes, hence why the interaction is also sometimes called the 2p2h (2 particle 2 hole) interaction. The modelling of MEC is difficult and thus is not being included as its own category in the simulation. It is also not included in figure 1.2 [56] [52].

From these plots, it seems that most quasi-elastic scattering events are selected, whereas deep inelastic scattering is more prominent in the event spectrum for rejected events. There are also more RES events chosen compared to DIS. This is due to quasi-elastic scattering events having the least energy in the hadronic system, which is harder to estimate accurately. This makes reconstruction easier and hence more of them are chosen for the analysis. RES events have energy in the hadronic system between DIS and QE, so they are chosen more than DIS but less than QE.

The background sample broken down by neutrino event type is shown in figure 4.19. These show that most non-ν_μ CC events are from neutral current events with a small contribution from ν_e beam neutrinos (see section 2.2). In total, there are 330 NC and 23 ν_e events.

As well as looking at the events in the reconstructed and true neutrino energy space, other variables were reviewed with the goal of finding features that connect the rejected events. These include variables describing

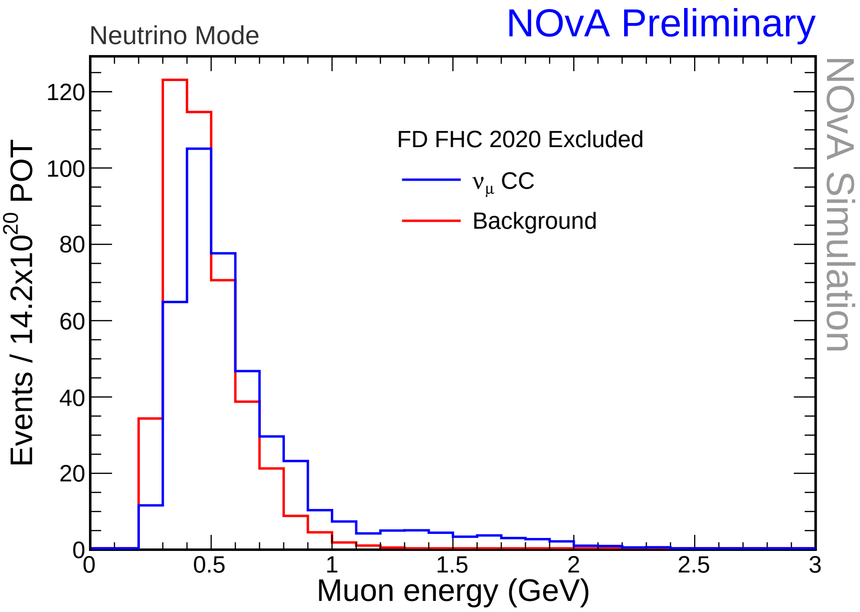

Figure 4.20: Muon energy for signal and background for the events rejected from the 2020 analysis, non-oscillated.

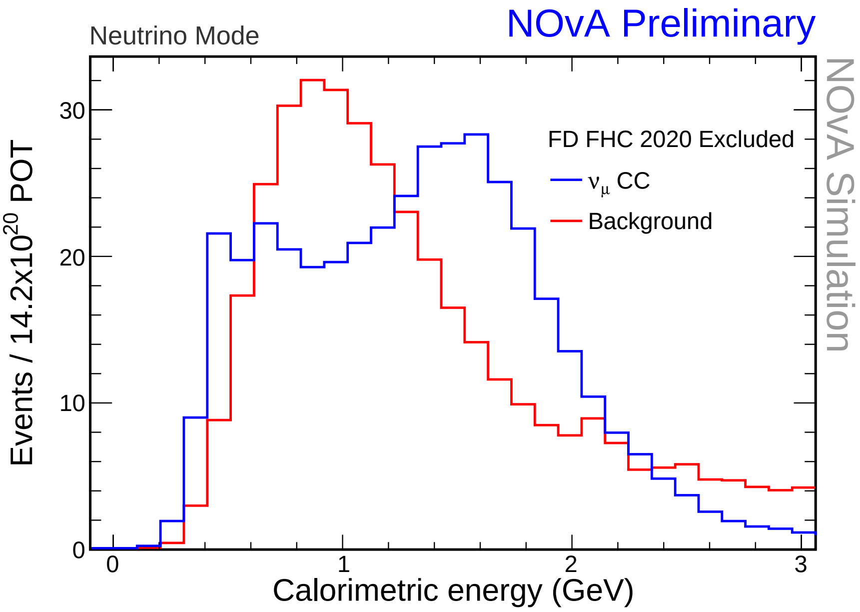

Figure 4.21: Calorimetric energy of the slice for signal and background for the events rejected from the 2020 analysis, non-oscillated.

properties such as event energies and the number of hits, lengths, etc. of tracks.

Examples of two different energies, muon energy and calorimetric energy of the slice are seen in figures 4.20 and 4.21 respectively. Muon energy does not appear to differ much between signal and background whereas calorimetric one does. For events where no muon is present, the muon energy is defined as the energy of the longest prong in the event. Since most of the background is NC, for these events the muon energy can be assumed to be the energy of the pion in most cases instead of

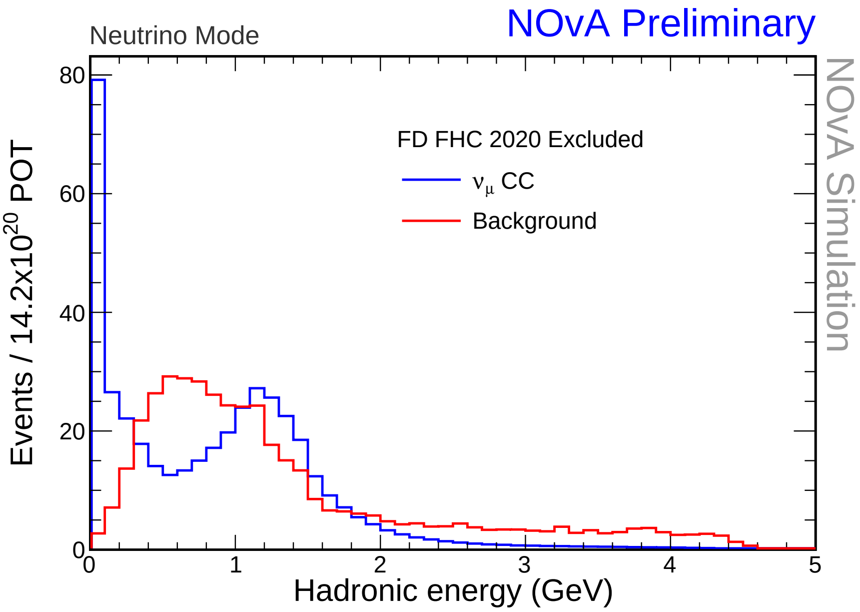

Figure 4.22: Hadronic energy for signal and background for the events rejected from the 2020 analysis, non-oscillated.

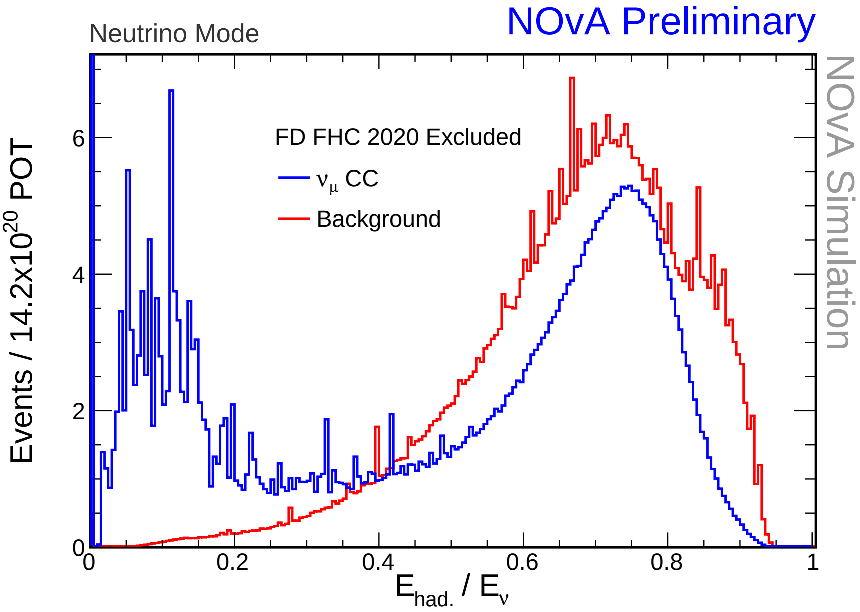

Figure 4.23: Hadronic energy fraction for signal and background for the events rejected from the 2020 analysis, non-oscillated.

zero, which is why the distributions are very similar. For calorimetric energy, there are two signal peak in the distribution. This is because in NC events most energy is released via hadronic shower whereas for QE CC events most of the energy is carried out by the muon and equal to less total energy in the system. There are, however, also ν_μ CC events with a higher portion of hadronic energy corresponding to RES and DIS events.

Another variables that shows a difference between signal and background is the hadronic energy, shown in figure 4.22, and the fraction of hadronic energy of the total energy: E_{had}/E_ν (see section 3.4.1), shown in figure 4.23. The hadronic energy is generally smaller for signal than background, but the signal also displays a second peak at 1.1 GeV. The hadronic energy fraction also has two signal peaks, at about 0.1 and at 0.7, whereas the background only peaks at 0.7. The shape is similar to the shape of the calorimetric energy and is due to the distribution of energy in different event types (QE vs. RES and DIS).

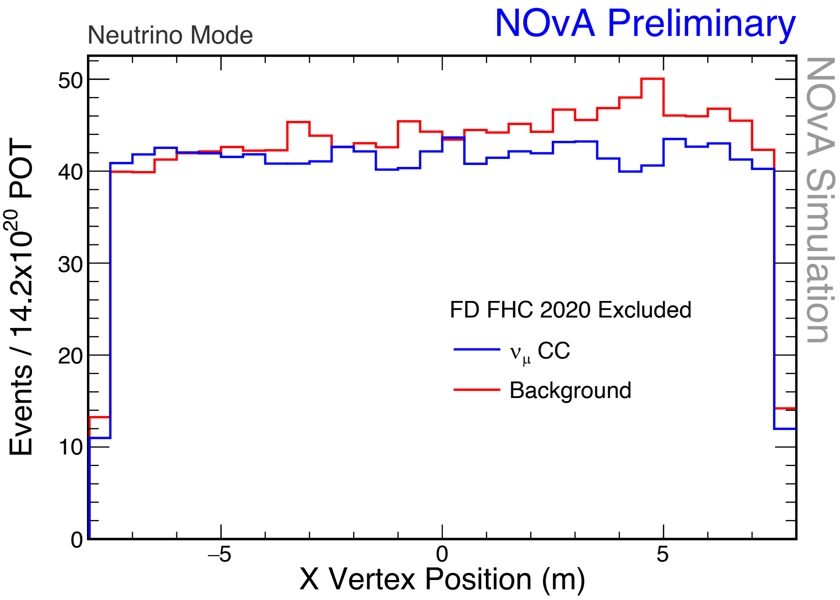

Figure 4.24: X vertex position (width) for signal and background for the events rejected from the 2020 analysis. The zero is defined in the middle of the detector in terms of width, non-oscillated.

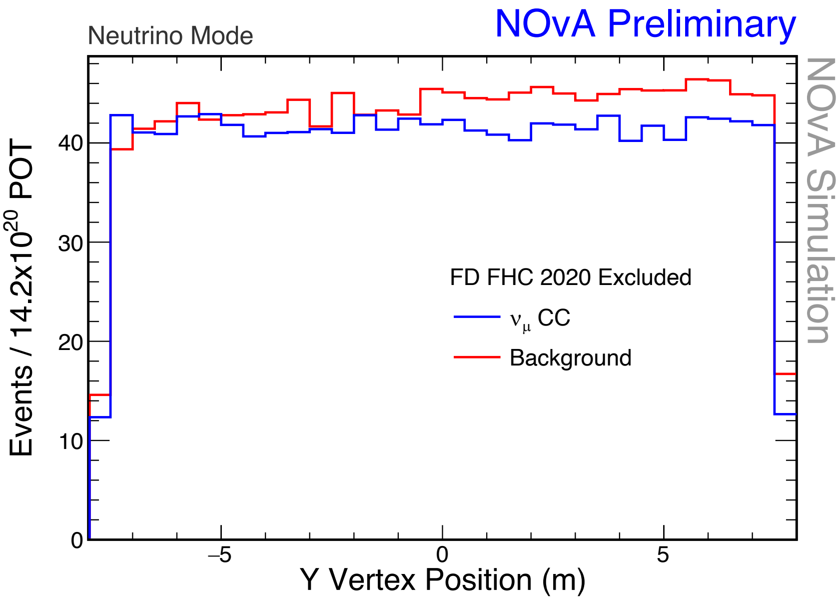

Figure 4.25: Y vertex position (height) for signal and background for the events rejected from the 2020 analysis. The zero is defined in the middle of the detector in terms of height, non-oscillated.

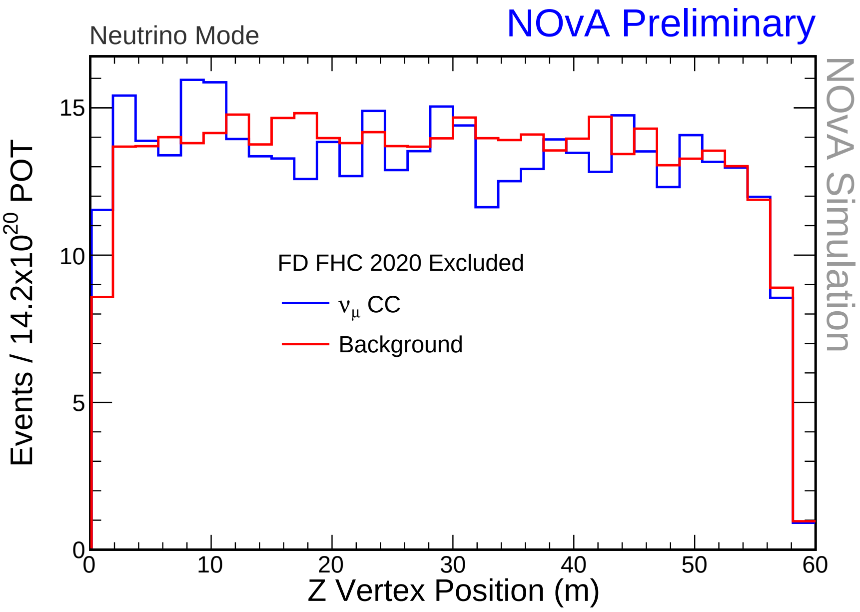

Figure 4.26: Z vertex position (depth) for signal and background for the events rejected from the 2020 analysis. The zero position is the starting point of the detector, non-oscillated.

Figures 4.24, 4.25 and 4.26 show the distribution of events in the far detector by vertex position in X, Y and Z views. The distribution of events is highly uniform in all views. For example, there is no excess of signal or background events at the back of the detector.

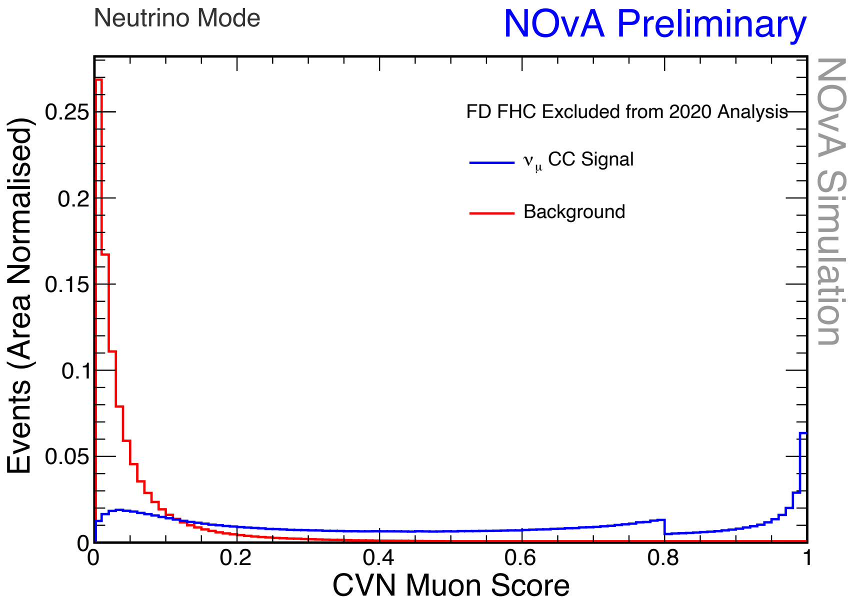

Figure 4.27: The CVN muon score for events rejected from the 2020 disappearance analysis, non-oscillated. The y-axis has been normalised by area to show probability.

Finally, variables that are useful to investigate for rejected events are their CVN muon and RemID scores. These two are the variables that form the final PID cut for the last analysis (see section 3.3.4). These

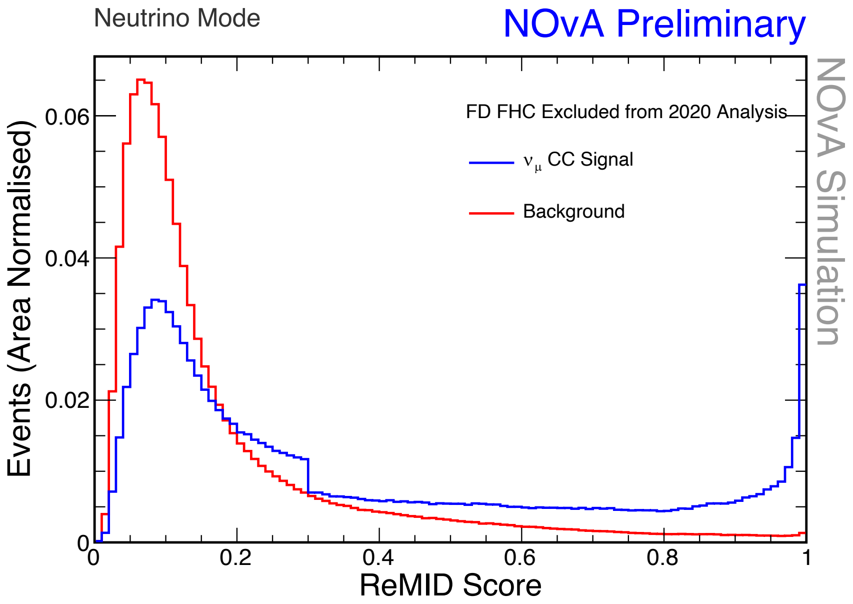

Figure 4.28: The ReMID score for events rejected from the 2020 disappearance analysis, non-oscillated. The y-axis has been normalised by area to show probability.

two variables for the rejected events are seen in figures 4.27 and 4.28 for CVN and RemID respectively. The CVN muon score has a drop at 0.8 and ReMID a similar one at 0.3, since those are the two cut limits used in the last analysis. Many signal events in this sample have a ReMID and CVN score close to 0 as well as 1. Thus it is clear that for both of them that there are a lot of signal events very similar to background due to the overlap with background close to the score value of 0.

Several other variables not highlighted here were also studied. Most of them having similar distributions to the vertex positions or muon energy, where there was no clear distinction between signal and background. It thus seems that most variables do not show clear distinction between signal and background, and including multiple variables to form a cut to recover signal would be the best approach. Some variables highlighted here, like hadronic energy, show a difference and could potential be used to form a cut to recover events. However, there is more potential to include multiple variables to form a cut rather than rely on one variable.

This suggests suggests that there is no clear type of event that is dismissed by the analysis that could be used to recover these events back. Obviously some variables, like hadronic energy, show a difference between signal and background and could potential be used to form a cut to recover events. However, there is more potential if multiple variables are included to form a cut rather than rely on one variable.

4.4 Maximum Improvement

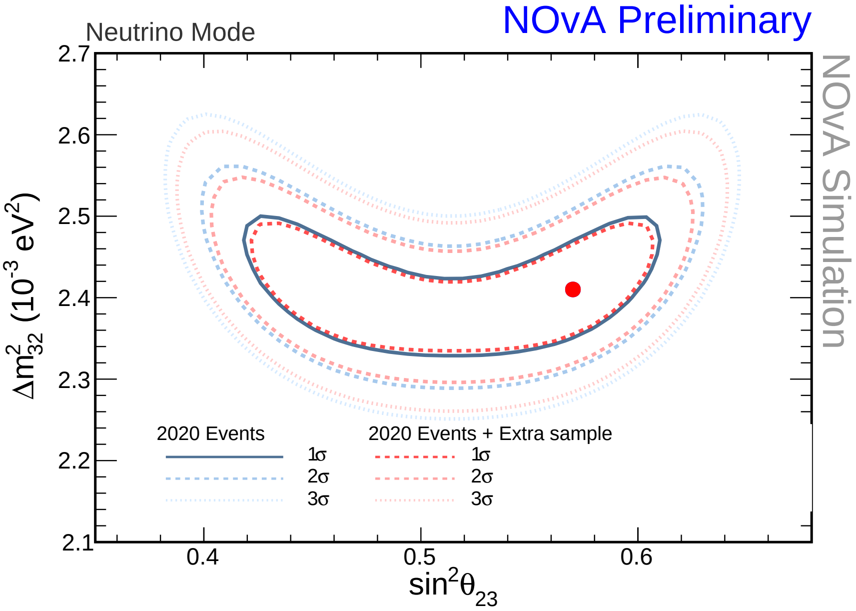

Figure 4.29: Theoretical upper bound improvement in sensitivity to oscillation parameters if all the currently rejected signal events could be added back into the analysis. These events are equivalent to an improvement of 14% in effective POT. Blue contour is the original fit and red the fit when the cut events are added back to the analysis as an additional sample. For both cases contours of 1σ, 2σ and 3σ are drawn. The best fit points appears equal in the plot but do differ to a degree too small to be seen in the figure.

As mentioned earlier (section 4.2), 20% of all signal events are rejected by the current cuts. If all of these extra signal events (and none of the background) could be added, the maximum amount of extra sensitivity to oscillation parameters gained is shown in figure 4.29.

From this plot, it can calculated that the maximum increase in effective POT is approximately 14% which is equivalent to running the detector for an additional 6 months. This calculation was conducted assuming that the detector is running 24 hours a day every day. This is a theoretical maximum value and in the end the real world improvement will be less. However, this demonstrates that a meaningful improvement is possible and can be made if these events are added to the analysis.

Chapter 5

Improvements to Event Selection

This chapter describes the methods used to attempt to recover the ν_μ events currently rejected from the NOvA disappearance oscillation analysis. The analysis uses neural networks, which are introduced first to outline some key concepts. Then, the process of choosing the right training is discussed, and results are shown describing the extent to which signal can be added back into the analysis and how much it increases the sensitivity to neutrino oscillation parameters. Finally, a section on how changing the current cuts could improve the results is included for comparison to the use of neural network training.

5.1 Neural Networks

Artificial neural networks (ANN or simply NN) model the behaviour of real biological neural network systems, such as the human brain. Actual biological systems are very effective at processing large amounts of information simultaneously, i.e. they are excellent at parallelism, which is beneficial for information processing in particle physics. Many particle detectors, including NOvA, gather large amounts data and need ways to process this information in an efficient way, hence neural networks are regularly used. No artificial system is close to the effectiveness and complexity of biological systems but they are constantly improving [108] [109] [110].

Figure 5.1 shows a schematic of a neuron (or a node), which is a basic element of ANNs. A neuron's inputs (x) are multiplied by weights (w) described by the output (z) of the neuron:

$$z_i = \sum_i w_{ij} x_j - s_i. \tag{5.1.1}$$

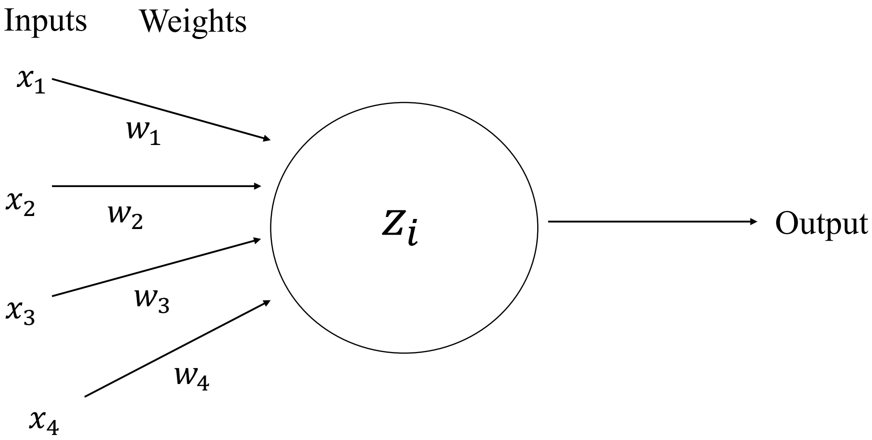

Figure 5.1: A schematic of an artificial neuron with inputs x_i and weights w_i giving the output (z) z_i

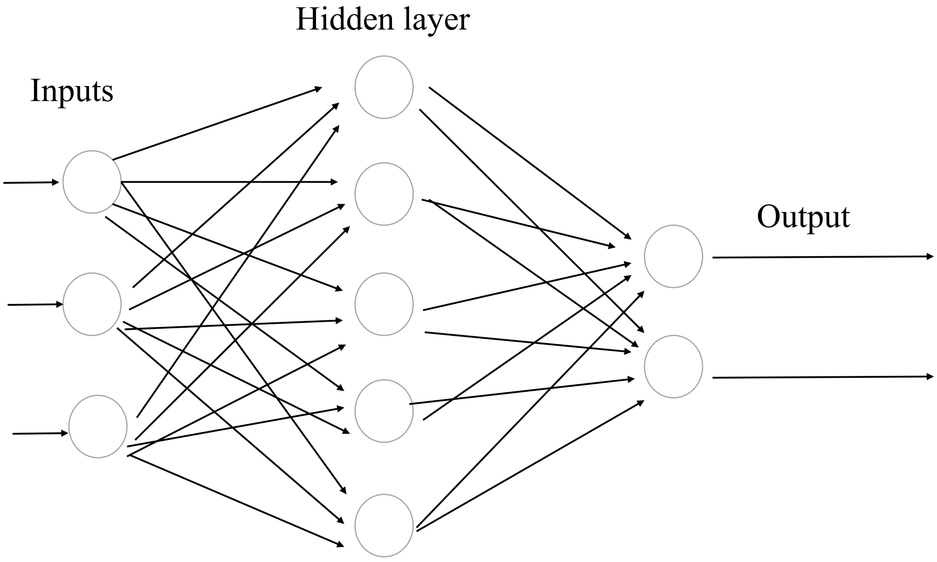

Figure 5.2: Example of a feed forward ANN with one hidden layer.

In equation (5.1.1) w_{ij} are the weights of a neuron i, x_j the inputs and s_i the bias (sometimes also called threshold or offset) set for the network training. The output signal is calculated using an activation function, dependant on the value of the net signal and the bias. Many options for activation functions are available depending on the application, including linear, non-linear, step, sigmoid and gaussian functions [108] [109].

Almost all neural networks consist of multiple interconnected neurons, which are organised into layers resulting in multi-layer ANNs. The input layer consists of a number of neurons which maps on to an output layer with a number of hidden layers in between. The architecture of an ANN with one hidden layer is seen in figure 5.2.

If the signal in an ANN is always sent to the layer above the next one, a feed forward network is formed. This is a common type of neural network used in particle physics, especially in classification problems which are relevant to the work in this thesis [108] [109] [110].

5.1.1 Neural Network Learning Processes

The learning process of an ANN determines the optimal weights used for the network. The learning types can be classified into three main categories: supervised learning, unsupervised learning, and reinforcement learning. In unsupervised learning, the network is given only an input and the network is trained to respond to patterns that are statistically significant in the training set. Similarly, in reinforcement learning only the input is feed to the network and the weights are adjusted according to a criteria to reward certain parts whilst others are penalised. This analysis uses supervised learning, and it is explained below in more detail [109] [110].

In supervised learning, the network is given a training data set as well as a target, which is the desired output. The weights are adjusted to minimise the difference between the actual output of the ANN and the target. A common usage for supervised learning is a classification network. In these networks, all data are assigned to a class found by the network by training with events of known classification. The event is given a value between 0 and 1, 1 describing the event belonging to a class and 0 if not. For example, in the context of events classified as signal or background, 1 would represent a perfect signal-like event and 0 a perfect background-like event. This can be interpreted as a probability distribution of how likely a particular event belongs to a desired category [108] [109] [110].

5.1.2 Network Performance

The performance of ANNs can be determined in multiple ways. One common approach is to calculate the error for another data set not used in the training, in which case the error is called the generalisation error. For the best performance of the network, the generalisation error needs to be as low as possible, which means the accuracy is highest possible. In classification problems, the accuracy is typically the percentage of data being correctly classified to the correct group [109].

Initialising the network and choosing the correct number of variables like epochs, hidden layers and the right activation function is complex

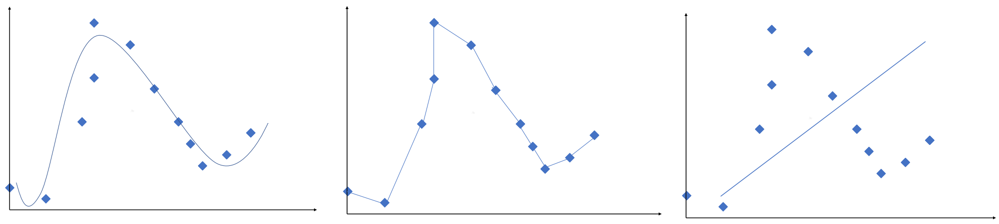

Figure 5.3: Schematic demonstrating under- and overtraining. The first plot describes good training, second one overtraining and the last one undertraining.

and must be uniquely optimised for each problem. For example, having many hidden neurons will ensure good learning but may compromise the ability of the ANN to generalise to other data samples. With too few hidden layers, the network might not learn the patterns in the data and will not be able to classify events correctly [109] [110].

Another problem that can occur if an initial network is not prepared correctly right is over-training. Illustrated in figure 5.3, an overtrained network gets too familiar with the training sample and starts to memorise patterns in it. This kind of network will not be capable of classifying samples outside the training input set. In this case, the generalisation error is higher than the error of the data set. The opposite can also occur, in which case a network is undertrained. In this situation the network does not learn the patterns in the data. This is also illustrated in figure 5.3, and can often be resolved with additional training time [109] [110].

5.1.3 Neural Networks in Particle Physics

In particle physics, neural networks can be used for both offline data analysis as well as online triggering, to select the best events to record when taking data. A review of different ways neural networks are used in particle physics reserach can for example be found in [111].

5.2 Neural Network Training

In this study, events currently rejected by the NOvA 3-flavour disappearance analysis cuts are trained in a neural network. The spectra of the events is seen in figure 4.10 for the non-oscillated case. This figure was scaled to show how many events are expected in real data, however the MC sample used was much larger.

The MC sample used was an upgraded version of that used in the last 3-flavour analysis in 2020. The previous analysis used data from production5 (which refers to the number of productions of simulation files), which was also the starting point of the analysis. Later, it was moved to the next generation of MC files called prod5.1. This included a larger sample and some bug fixes as well as changes to systematics. The sample included approximately 1.4 million signal and 2.3 million background (2.15 million NC and 150 000 ν_e CC) events that did not pass the current cuts.

These events were compressed into a csv file from NOvA datafiles containing multiple parameters describing event energies, momentum and track information that is provided by the NOvA reconstruction. This network uses supervised learning (see section 5.1.1) so the file also included a variable that gives a value of 0 to background and 1 to ν_μ CC signal to be used as a target for the learning.

The training was done using TensorFlow (TF) [112], a free and open-source software that can be used in machine learning. Because of its wide use, good documentation and support is available making it relatively easy to use. It has also been used in NOvA before in reconstruction (section 3.2). Version 1.12.0 of TF was used as it is compatible with NOvA software.

5.2.1 Training Variable Selection

The first trainings included all the variables in the files, which were then removed based on their performance. These trainings were done with the old prod5 data but not when moving to prod5.1 data, since these were expected to yield the same results. The main goal of the training was to demonstrate general proof of concept and learn about the behaviour of the network.

A statistical test called the Kolmogorov-Smirnov (K-S) test was used on initial selection of training variables. It is a test to check for goodness-of-fit by determining how much two distributions differ. The test checks whether a single sample could have been sampled from another given reference distribution, a low test score indicating that the two distributions are similar and the sample follows the chosen distribution. In a case of choosing variables for network training, the more different the distributions for signal and background are, the better it can be expected to perform in the training, and thus the higher the K-S test value should be. Variables that have very similar distributions for signal and background will not help the network to learn the differences between them [113].

Another way to test variables is to draw spectra for each variable (as mentioned in section 4.3.1), and just from a spectrum it can be investigated if some variables have similar distributions for signal and background, and thus should not be used in the training. The downside of this method is the amount of time it takes, especially if a lot of variables are tested.

The process of picking the best variables was started by choosing the highest K-S test scores for 18 variables (all shown on table 5.1). For all of these variables, a spectrum showing the distribution between signal and background was drawn, choosing the ones with the best separation power. Based on these tests, 10 variables were selected. Those not chosen are seen in appendix A and in the last chapter (figures 4.20, 4.24, 4.25, and 4.26), demonstrating that they have similar distributions for signal and background.

Variable	**K-S test score**
RemID:dE/dx LLH	0.409
Reconstructed Muon Transverse Momentum: Pt	0.319
Reconstructed Invariant Mass (W)	0.317
Number Hits in the Hadronic Track	0.310
Hadronic Energy Fraction	0.291
RemID: Non-hadronic Plane Fraction	0.288
Hadronic Energy	0.251
Muon Energy	0.211
Track Calorimetric Energy	0.174
Reconstructed 4-Momentum Transfer	0.171
Reconstructed Muon Pt/P	0.141
Calorimetric Energy	0.131
RemID: Scattering Log-likelihood	0.111
Vertex Y position	0.0504
Slice Duration	0.0337
Slice Time	0.0254
Vertex X position	0.0252
Vertex Z position	0.0215

Table 5.1: Variables and their KS test values tested for neural network training.

After the 10 variables were chosen, all of their distributions for ND data and simulation were drawn. This is to check if there are any disagreements between data and simulation. There were three variables that do not match between data and MC, and so to avoid biased results in data those were cut out. Examples of those variables are seen in figures 5.4, 5.5 and 5.6. These are three of the variables used in the RemID training (see section 3.2.3), and it is clear that they do not agree between MC and data even within the uncertainty limits (purple shading). Thus, those were not used in training.

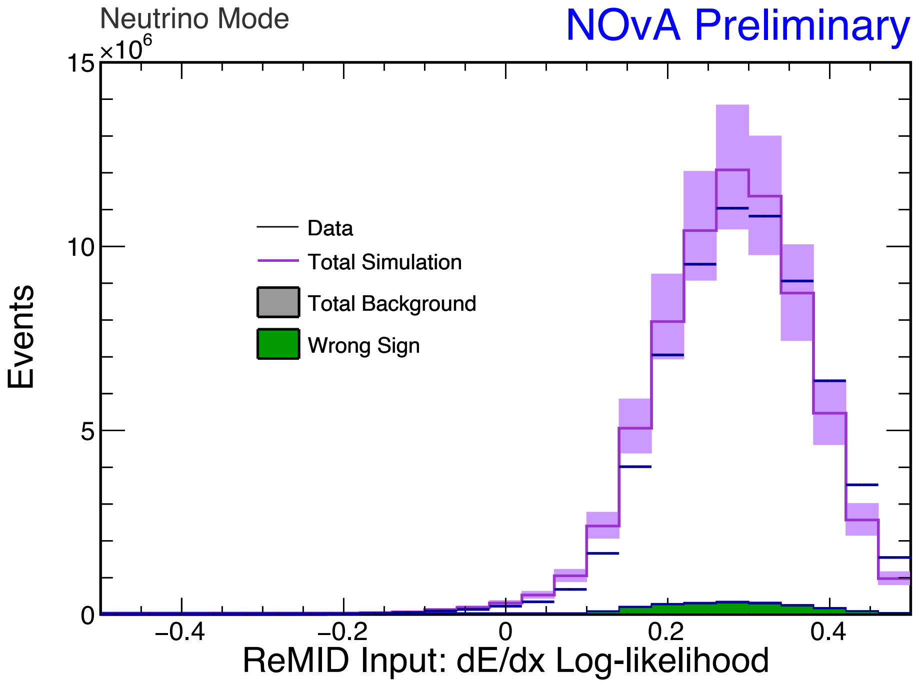

Figure 5.4: RemID variable $\frac{dE}{dx}$ LLH (see section 3.2.3) comparison plot for ND MC and data including background.

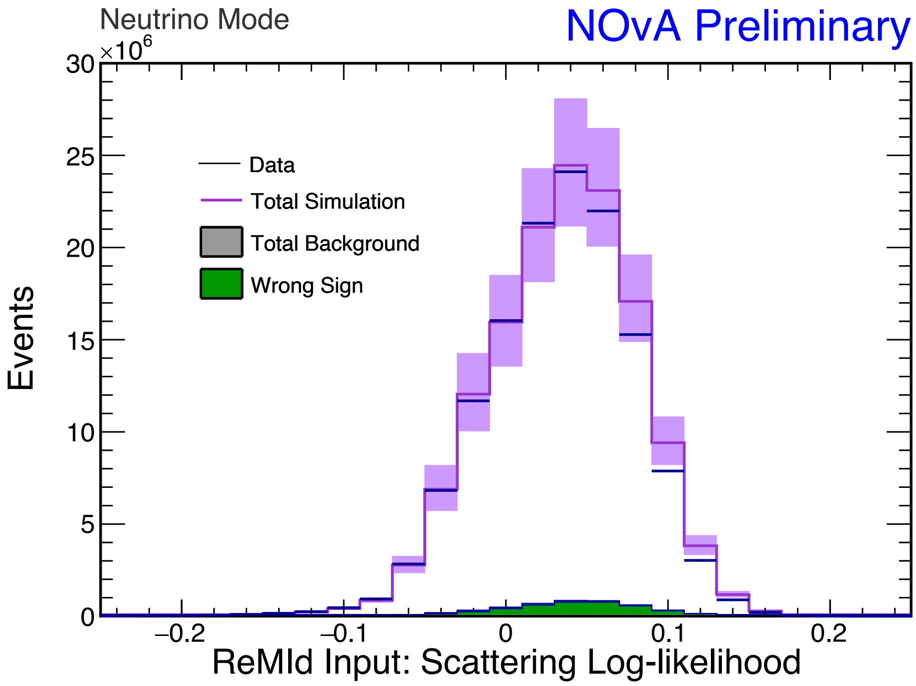

Figure 5.5: RemID variable scattering LLH (see section 3.2.3) comparison plot for ND MC and data including background.

5.2.2 Network Structure

TF offers many options for network structure, although most of those are only useful for image recognition and require 3D images to be used. However, for the uses of this study, simple connected layers should be sufficient. Also, the easier and faster the training is, the better the re-

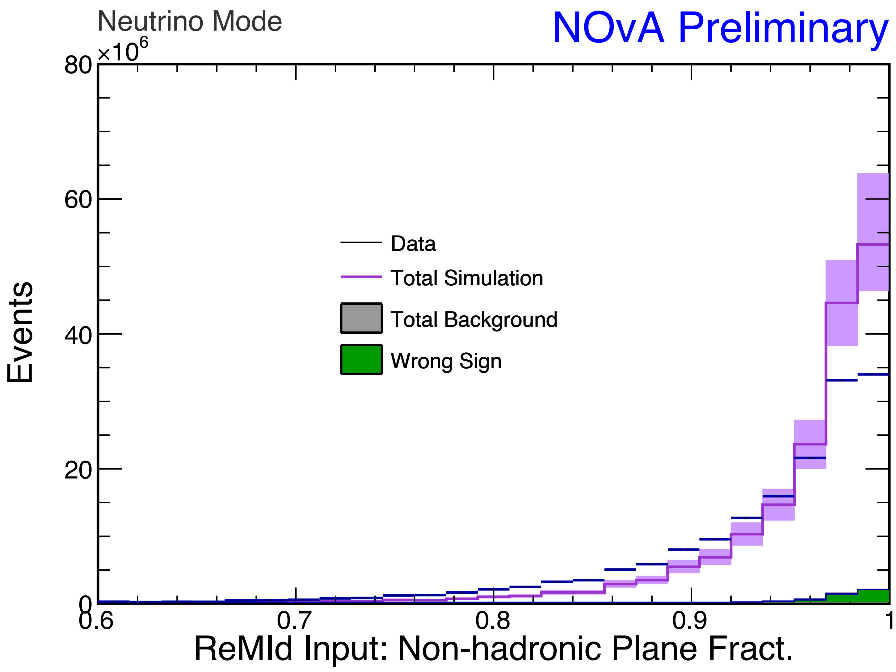

Figure 5.6: RemID variable of fraction of planes used in $\frac{dE}{dx}$ LLH calculation (see section 3.2.3) comparison plot for ND MC and data including background.

producability and fit with other parts of the NOvA 3-flavour analysis will be. Although, the expected limiting factor in this study will likely be how similar the signal and background events are to each other, not necessarily the computing power, considering the events that are trained.

The activation function used in the network was exponential linear unit (ELU). ELU is defined as [114]:

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha(e^x - 1) & \text{if } x \leq 0, \end{cases} \tag{5.2.1}$$

with $0 < \alpha$, which is the hyperparameter that controls how ELU saturates negative net inputs. This is an improvement compared to RELU (rectified linear units) activation function, which gives negative arguments an activation of zero. This can sometimes lead to bias in training, but ELU does not have this problem [114]. When tested, ELU performed the best with RELU performing very similarly. Other tested activation functions included sigmoid, tanh and exponential but they did not give results as good as ELU.

For the output layer, a softmax function was used. Softmax is a probability that describes whether the input value belongs to a desired category. The outputs are values between 0 and 1, which sums to 1. This can be interpreted as a probability distribution for the input values.

Mathematically, softmax calculates the value for a vector x as [115]:

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \tag{5.2.2}$$

where $i = 1, 2, .., N$, x_N describes the input value of the network and $f(x_i)$ is the probability of input belonging to a ith category. Softmax is easy to use when visualising the classification probability of a network as it can be interpret as how effective a network is at separating between signal and background [115] [116].

The optimiser used was Adam, which implements the Adam (adaptive moment estimation) algorithm. The Adam algorithm is a gradient-based optimisation of stochastic objective functions. It aims to adapt the learning rate to each parameters in the network from estimates of first and second moments of the gradient based on the statistics of the gradient. It is efficient and uses little memory, and is also powerful for training that includes large data sets or many parameters. In this study, the data sample is large and many parameters were tested so Adam was a well-suited option. Other options for optimisers are available in TF and they were also tested briefly, but Adam performed the most successfully or to a very similar level as the others [117] [118].

A loss function is used in training to keep track of the goodness of the classification as well as define the gradients that are used in training the network [119]. The loss function in this training was calculated by sparse categorical cross-entropy. This is commonly used to calculate loss functions for networks where the output has two label classes, like signal and background. Another variable to describe the performance in the network is accuracy. In TF, accuracy is a variable that calculates how often predictions equal labels, i.e. how often an event is correctly matched with its true value. This describes the network's ability to correctly identify events [120].

The data used in the training was divided to train, test and validation data sets, 80% is allocated to train, 10% to test, and 10% to validate. The network is first fully trained on the train set, then validated on the validation set to observe how well the training generalises to data outside the training test set. Finally, a test set is used to quantify how the network performs and what the final distribution provided by the softmax corresponds to.

The biggest issue in setting an ideal number of layers and their size was the networks tendency to categorise all events into one category (usually as 0, i.e. background). This was prevalent when using many layers, and consequently the number of layers was decreased whilst their size was increased.

5.2.3 Tuning

To correct the imbalance of different amounts of signal and background in the training data, class weights were used. These describe the percentage of each category in the data set, which are assigned as weights in the training. This is especially helpful if one of the data sets is very small, since there are fewer of those events for the network to learn from.

Over-training is a common problem in classification networks (as explained in section 5.1.2). One way that over-training can be mitigated in TF is dropout layers. These layers limit the amount of events trained in the next level, preventing the network from getting too used to the training set. These were tested and were ready to use in the event of over-training, but ultimately those were not needed in the training.

In addition, an early stopping was implemented to the training. This callback is used to stop the training when a chosen variable has stopped improving. This variable can be loss or accuracy for either train or validation set, and the number of epoch after which the improvement has stopped can also be chosen.

A learning rate optimiser was also set to further test the training in an order to evaluate if any improvement could be achieved from changing the learning rate, despite using the Adam algorithm. Another tested feature on the training was the use of a scaler, which standardises the input variables based on their mean values to improve training.

5.2.4 7 Variable Training

A version of a training that uses seven variables and assigns two values to each event is first introduced. The two values were how signal-like (pure signal 1) and how background-like (pure background 0) the event is.

Variables and Model Architecture

For this training seven variables were chosen (method explained in section 5.2.1), all of them listed in table 5.2. Three of them were introduced in the last chapter, and their spectra are seen in figures 4.21, 4.22 and 4.23. The rest are seen in figures 5.7, 5.8, 5.9 and 5.10. Appendix B shows the data-simulation comparison plots for all the listed variables. They show

Training variables
Hadronic Energy
Reconstructed Invariant Mass
Hadronic Energy Fraction
Transverse Muon Momentum
Number of Hits in a Hadronic Track
Calorimetric Energy
Reconstructed Muon Pt/P

Table 5.2: 7 variables used in the first network.

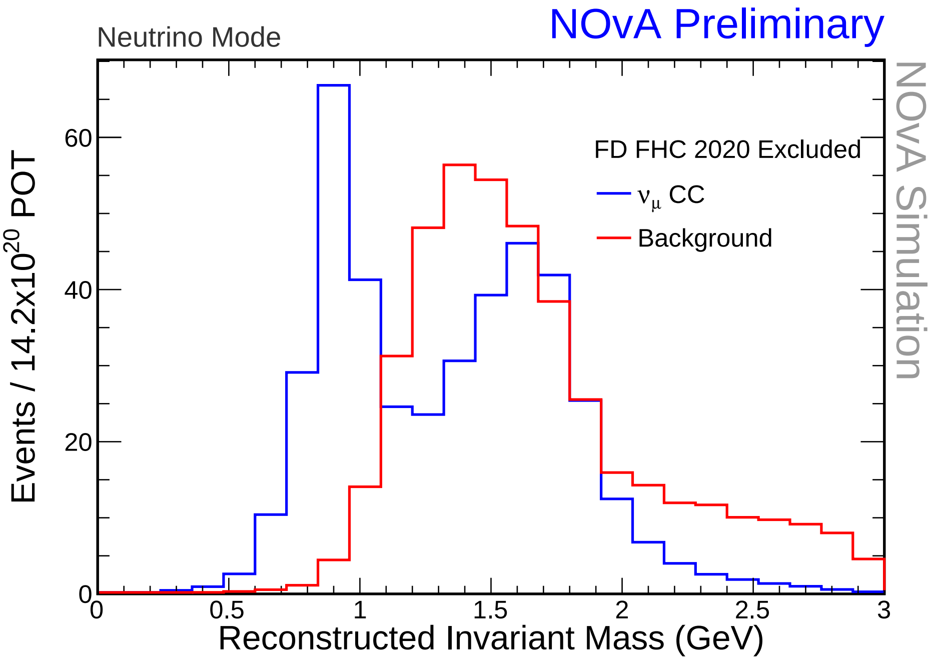

Figure 5.7: FD FHC simulation: reconstructed invariant mass for events rejected from 2020 disappearance analysis, non-oscillated.

that all the variables have an agreement between simulation and data within the uncertainty limits.

The model architecture used for this training is seen in figure 5.11. The input layer consist of 7 input variables connected to a dense layer of size 700. That is connected to 3 dense layers of size 1500 and finally to the softmax output layer. The architecture was chosen based on trial and error, tests changing the number and the size of the layers were done to investigate different options. This model ensured that enough complexity is introduced by adding 4 hidden layers but also their size is big enough for the data set. No more complexity was introduced, since when testing adding more layers or changing their size lead to the network not converging but rather identifying all events as background.

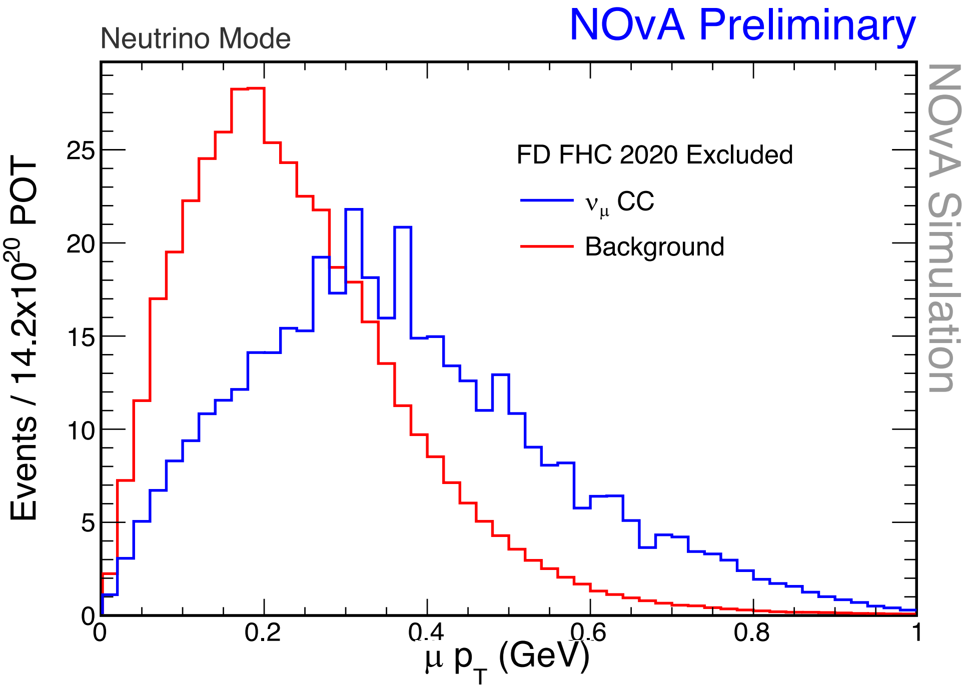

Figure 5.8: FD FHC simulation: reconstructed muon transverse momentum for events rejected from 2020 disappearance analysis non-oscillated.

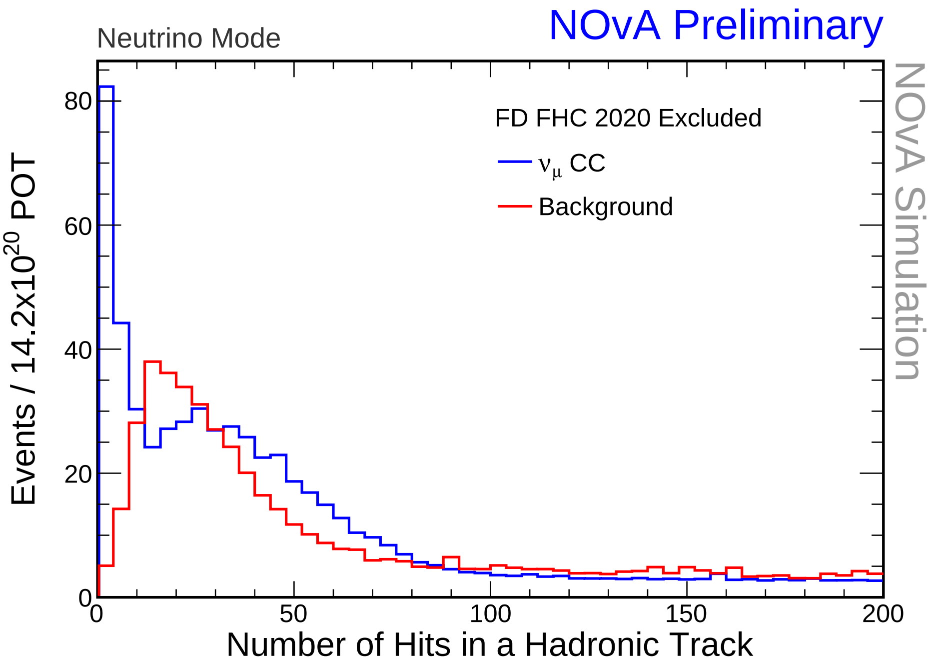

Figure 5.9: FD FHC simulation: number of hits in a hadronic kalman track for events rejected from 2020 disappearance analysis non-oscillated.

Results

Figure 5.12 shows the classification score (also called the PID) which are the results from the two output variables from the final softmax layer using test data. A figure showing the accuracy and loss per epoch for the training is seen in figure 5.13. As it is expected, the loss is decreasing

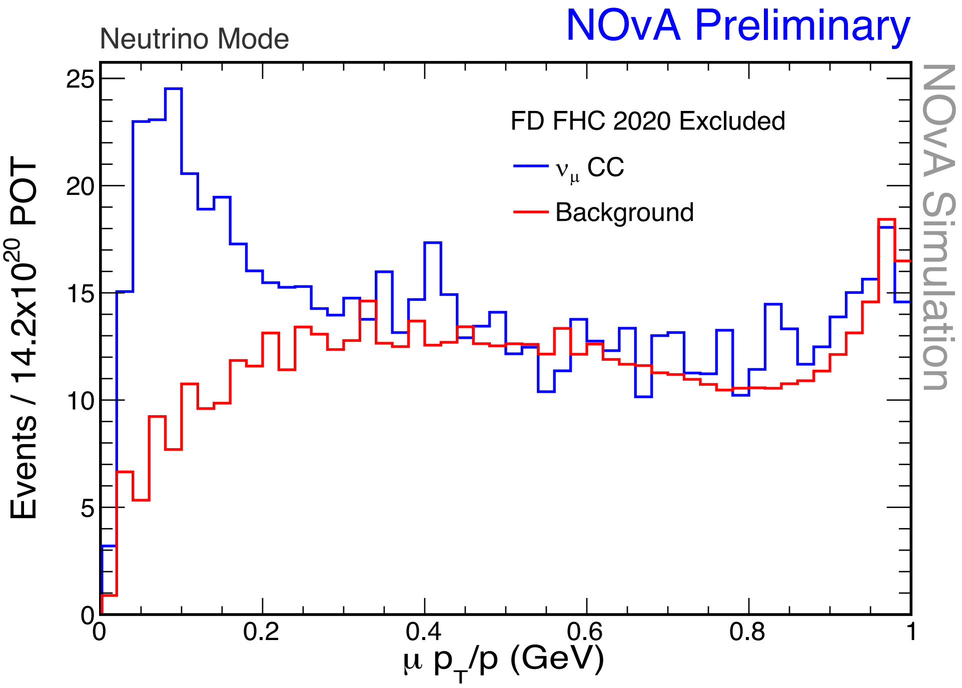

Figure 5.10: FD FHC simulation: reconstructed transverse muon momentum divided by the total momentum (Muon Pt/P) for events rejected from 2020 disappearance analysis non-oscillated.

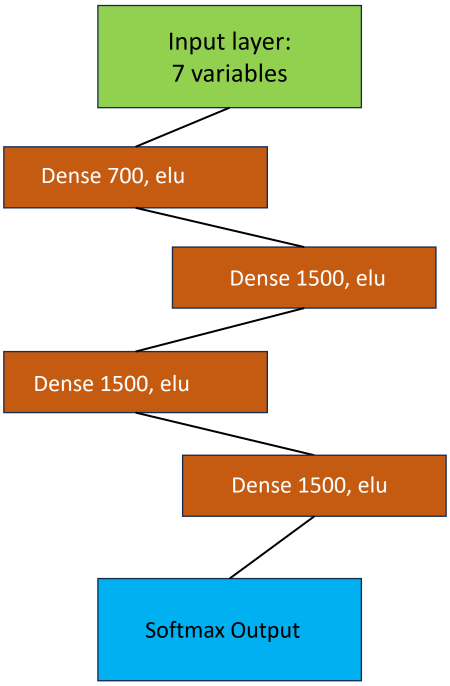

Figure 5.11: Neural network model architecture for the network with 7 input variables. The input layer is connected to a dense layer of size 700, which is connected to 3 dense layers of size 1500. The final layers is using a softmax function to turn the results into two numbers between 0 and 1 describing how signal and background like an event is.

and the accuracy increasing, although it seems that both improve only for the first few epochs and only marginally change for the rest of the

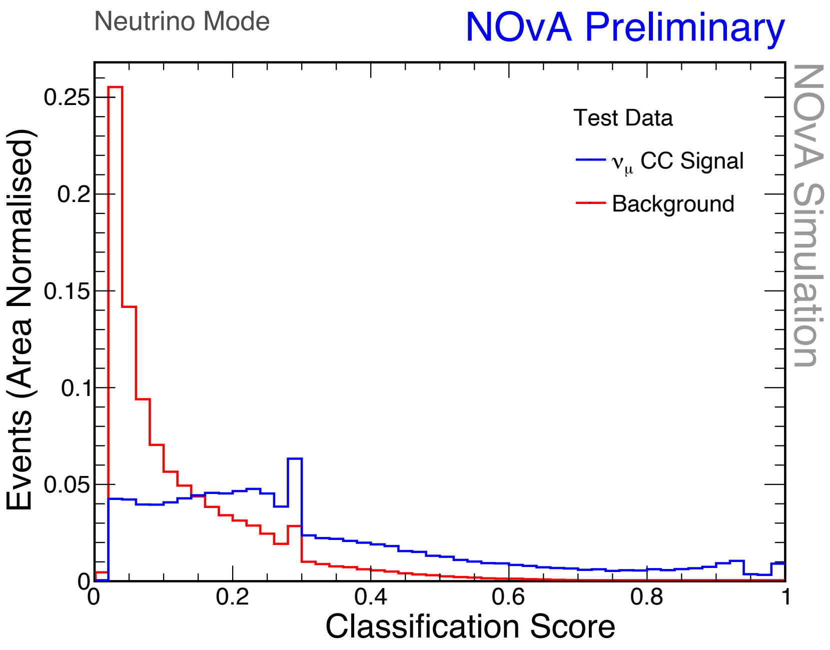

Figure 5.12: Classification score plot from training classifying events into signal and background. 7 variables and 20 epochs were used in the training. The y-axis has been area normalised by area to show a probability instead of absolute event numbers.

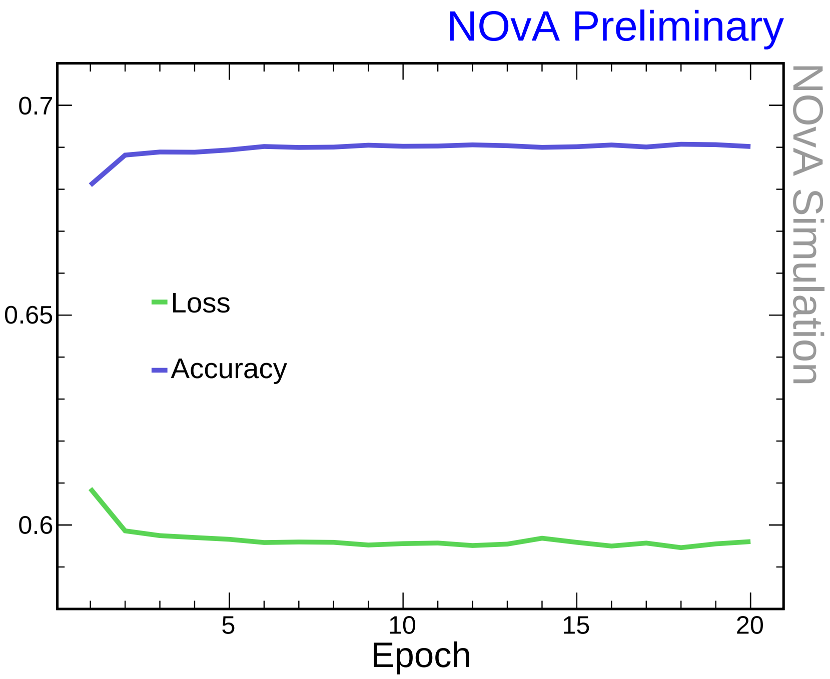

Figure 5.13: Plot showing loss and accuracy from training classifying events into signal and background. 7 variables and 20 epochs were used in the training

training.

In the PID distribution (figure 5.12) there are a few values where signal and background both have peaks, which most likely resulted from a certain type of event or misidentification of events. As it is evident

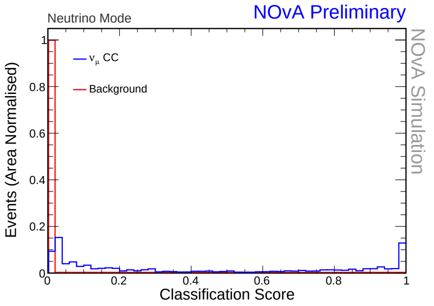

Figure 5.14: PID score for events above 1.5 *GeV* using NOvA FD FHC simulation for the network with 7 variables and 20 epochs. The y-axis has been normalised by area to show a probability instead of absolute event numbers

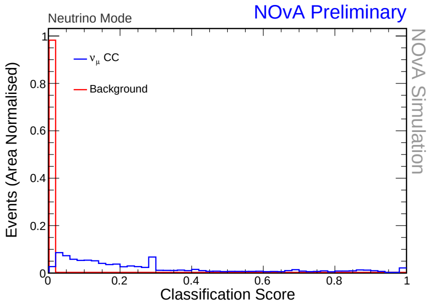

Figure 5.15: PID score for events below 1.5 *GeV* using NOvA FD FHC simulation for the network with 7 variables and 20 epochs. The y-axis has been normalised by area to show a probability instead of absolute event numbers

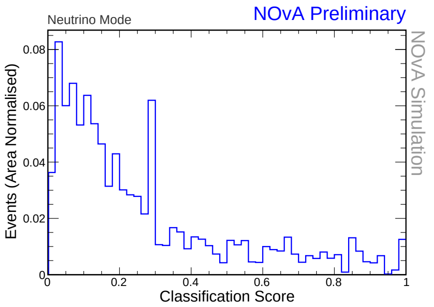

Figure 5.16: PID score for QE signal events only using NOvA FD FHC simulation for the network with 7 variables and 20 epochs. The y-axis has been normalised by area to show a probability instead of absolute event numbers

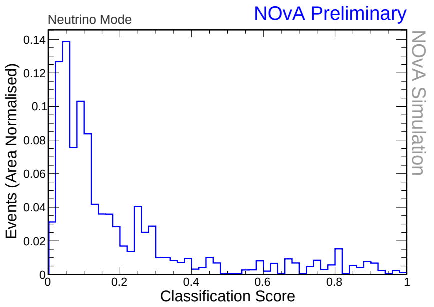

Figure 5.17: PID score for RES signal events only using NOvA FD FHC simulation for the network with 7 variables and 20 epochs. The y-axis has been normalised by area to show a probability instead of absolute event numbers

from the distrubution plots of the training variables, a lot of them have a structure in the signal events. For example, figures 4.21 and 4.22 both show two signal peaks. As it was discussed in section 4.3.1, these arise from the differences in QE and DIS/RES events. In QE events most of the energy is carried out by the muon emitted and no energy is carried by hadrons. In QE/RES the energy is distributed between hadronic system and the muon. Variables such as hadronic energy fraction (figure 4.23) and number of hits in a hadronic track (figure 5.9) show the two peaks as well as they depend on the hadronic energy. The muon momentum is more uniform (figure 5.8), but its fraction of the total momentum (5.10) shows this difference once again as in QE events all momentum is carried out by the muon. It looks like that the network is learning these patterns and it is seen in the final PID plots.

To better understand the shape of the PID classification plot for this

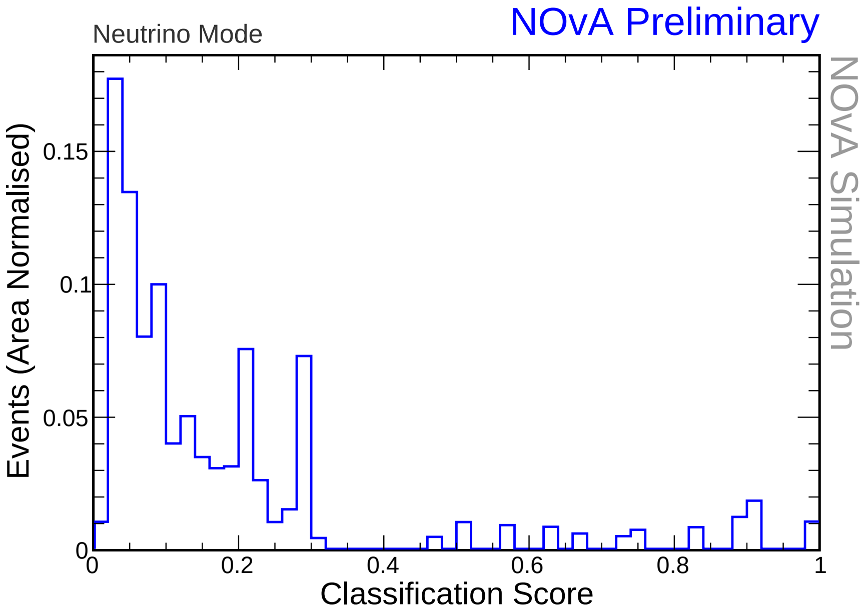

Figure 5.18: PID score for DIS signal events only using NOvA FD FHC simulation for the network with 7 variables and 20 epochs. The y-axis has been normalised by area to show a probability instead of absolute event numbers

network, some additional plots were created. Separate PID plots for high energy (more than 1.5 GeV) and for low energy (less than 1.5 GeV) are seen in figures 5.14 and 5.15. These figures were created after the network results were downloaded to NOvA software, so they were not produced using the same test data. Based on these figures, most of the signal that is assigned a PID score of 1 is high energy events, and the peak on signal at 0.3 PID consists of low energy events. The background is focused more on 1 in both of these plots, possibly because these plots were run for the whole data set whereas the original PID plot is only shown for the test data.

Furthermore, separating the signal score for three different event types, QE, RES and DIS, was performed. The figures are seen in 5.16, 5.17 and 5.18 for QE, RES and DIS respectively. The events with PID score of 1 consist mostly of QE and DIS events, whereas at lower PID scores a mix of all events is observed. The peak at 0.3 appears in all of the plots, but it is proportionally the biggest for the low energy events and the QE spectrum.

As mentioned in section 5.2.3, the differences between the amount of signal and background were corrected by class weights in the network. This was not, however, done for the different signal categories (QE, RES, DIS) discussed here. As it can be seen from plot 4.17 there are a lot less QE events compared to DIS and RES. A future improvement to this analysis would be to test class weights to these event categories, and whether it would improve the results.

5.2.5 5 Variable Training

Missing Variables	**FOM** $(\frac{S}{\sqrt{(S+B)}})$
All variables	2.47
Hadronic Energy	2.45
Reconctructed Invariant Mass (W)	2.46
Hadronic Energy Fraction	2.41
Muon Pt/P	2.48
Calorimetric Energy	2.48
Number of Hits in Hadronic Track	-
Reconstructed Transverse Momentum	-

Table 5.3: Network tests performed with excluding each variable and their FOM values with 0.8 cut on classification score.

To further optimise the variables used in the training, the network was run separately without each variable to investigate whether leaving them out changes the training. For each one of these trainings, a FOM was calculated using the same formula as was used for the 2020 cut optimisation: $S/\sqrt{(S+B)}$ (see section 3.3.4) using a cut of 0.8 in classification score. The results can be seen in table 5.3. The two variables with no FOM result caused the network to converge into a situation where all events were assigned as background, i.e. 0, so there are no results for those two networks. Thus it can be concluded that those two variables are essential for the network to converge and generate results.

Based on these results two variables were left out of the training: calorimetric energy and Muon Pt / P, because without them the training improved. One can assume that without these variables the training should improve. The new training variables are listed in table 5.4 and the model architecture was kept the same as for the 7 variable training (see figure 5.11).

Training variables
Hadronic Energy
Reconstructed Invariant Mass
Hadronic Energy Fraction
Transverse Muon Momentum
Number of Hits in a Hadronic Track

Table 5.4: 5 variables for the network after testing the performance of each of the original 7 variables.

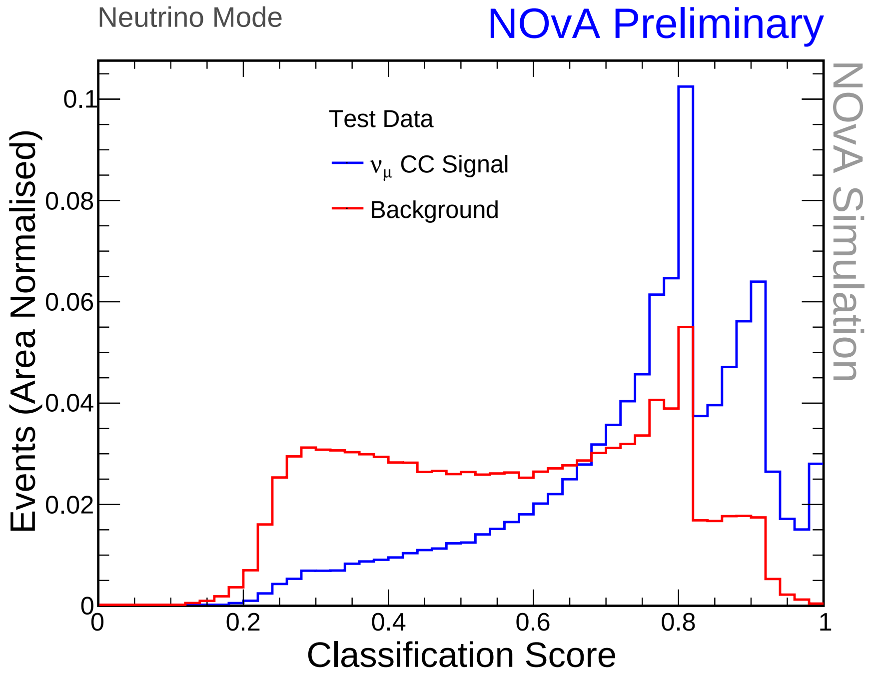

Figure 5.19: Classification score plot from the training classifying events into signal and background using 5 variables and 20 epochs. The y-axis has been normalised by area to show a probability instead of absolute event numbers.

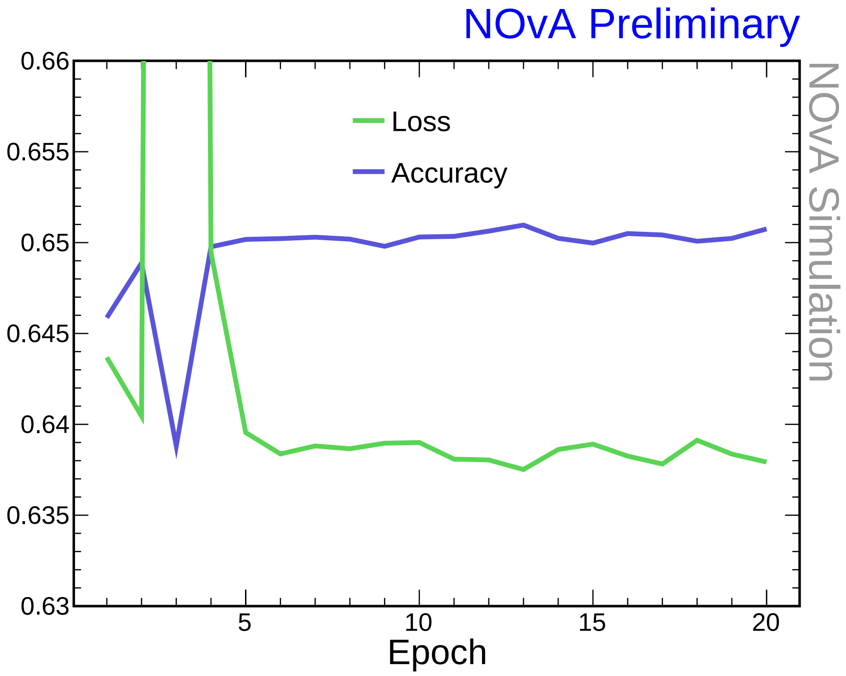

Figure 5.20: Plot showing loss and accuracy from the training classifying events into signal and background using 5 variables and 20 epochs.

The results from this new training are seen in figures 5.19 and 5.20 for PID as well as loss and accuracy respectively. The accuracy decreases and the loss increases for the third epoch, but after that the training falls back into the expected behaviour, so that particular epoch is assumed to be an anomaly and no further action was taken. Interpreting the PID, the results appear slightly improved that for the 7 variable training as expected.

The shape of the PID distribution for this network is very different than that for the 7 variables. The background does not peak at 0 but rather at 0.7, and there are three signal peaks at 0.7, 0.9 and 1. It is once again most likely that this plot is showing the internal structure of some of the variables, as discussed previously. The 0.7 peak has both signal and background events, so these events would then be RES and DIS events, and the two other signal peaks mostly consists of QE events as less background is present in those peaks.

When producing PID plots for only high or low energy events, as was done for the 7 variable training, no such shape was found. The same was discovered for dividing to QE, RES and DIS events, or dividing background into NC and ν_e categories. This could be due to the test sample (that the PID plots are produced with) having fewer events than included on the separate PID plots using simulation. It is possible that only a small subset of events is the cause of this behaviour, which cannot be reproduced when all data is included. As this behaviour cannot be fully explained, using this network to obtain final results should be done with caution. Some conclusions can be draws from the structure of the variables in regards the shape of the PID plot, but not all features are satisfactorily understood. To further investigate this behaviour, the input of this network with different cuts on PID could be drawn for different variables to learn more.

As discussed earlier, two variables, number of hits in a hadronic track and reconstructed transverse momentum, clearly demonstrated that they are essential for the network to converge (see table 5.3). Given more time, just training these two variables could be tested to see what kind of network would be obtained. The results could provide more insight into the behaviour of the network and explain better the shapes present in the PID plots.

Training Optimasation

To further improve the results and optimise the network training, two callbacks were added to the training: early stopping and learning rate scheduler (as discussed in section 5.2.3). Also, the use of a standard scaler was tested to standardise the training variables.

The learning rate in Adam optimiser assigns each individual parameter a learning rate separately (as explained in section 5.2.3), but it was investigated if manually setting an learning rate optimiser as:

$$Learning\ rate = 10^{-3} * 10^{\frac{epoch}{30}} \tag{5.2.3}$$

would change the results. In this equation the learning rate starts at approximately 0.01, which is the default learning rate for Adam in TF. The training is run for 50 epochs, resulting in the last training having a learning rate of approximately 0.05.

The early stopping callback stops training when a certain variable does not improve any further. In this case, loss was set as an early stopping variable and set to stop if it does not improve for 1 epoch. Others variables were also tested, but the valuation loss and accuracy fluctuated slightly causing the training to stop very quickly, and were thus not used. Accuracy also stopped improving with loss in all tests, so the use of accuracy would have provided identical results.

In an attempt to improve the shape of the PID distribution, a standard scaler [121] from scikit-learn library [122] was added to the training. This standardises the input variables in the training by removing the mean value and scaling each variable to be between 0 and 1. This should remove any proportional dependency on variables with higher mean values or different distributions of values.

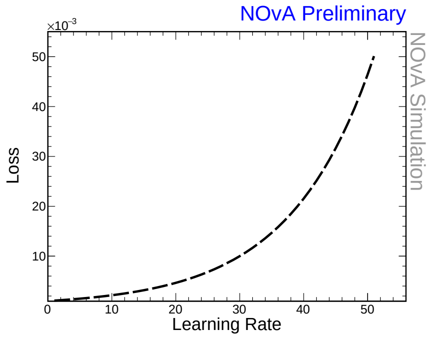

Figure 5.21: Loss vs. learning rate for the 5 variables model when learning rate is varied according to equation 5.2.3

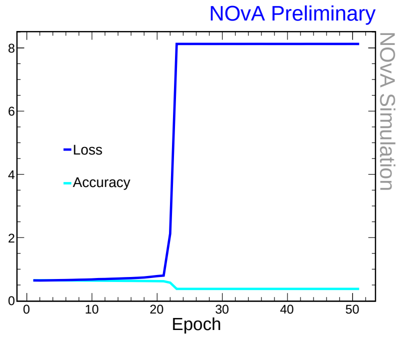

Figure 5.22: Loss and accuracy for the 5 variables model when learning rate is varied according to equation 5.2.3

The results from the learning rate test are seen in figures 5.21 and 5.22. The first plot has loss plotted against the learning rate. The plot is clearly increasing logarithmically, indicating that increasing the learning rate is detrimental to the training performance. This is also evident from the second plot, showing the accuracy and loss, especially at approximately 20 epochs when the loss increases dramatically to above 8. As discussed in section 5.2.2, Adam changes the learning rate for different parameters so it is possible that changing it manually removes its ability for effective optimasation. Therefore, it was decided that the learning rate was kept as the default.

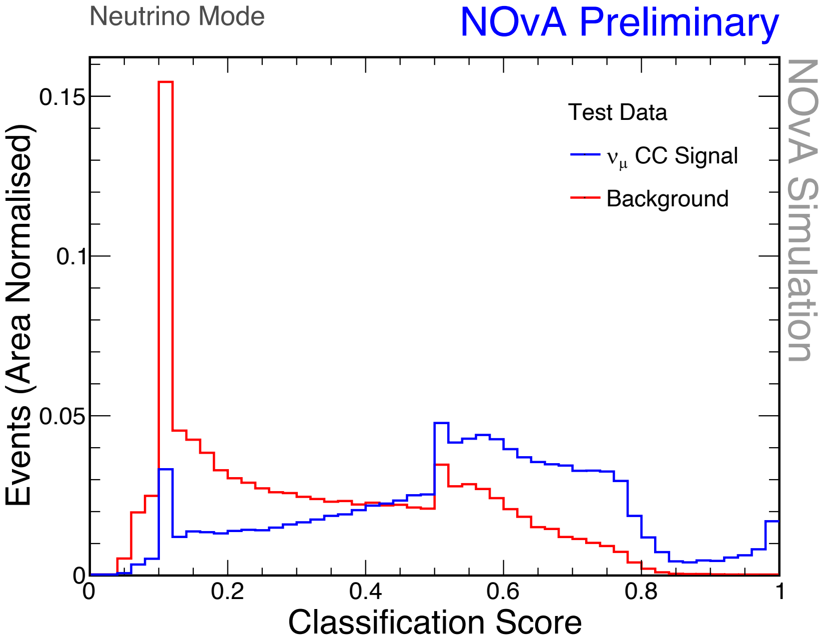

Figure 5.23: Classification score plot from training with 5 variables using early stopping. The y-axis has been normalised by area to show a probability instead of absolute event numbers.

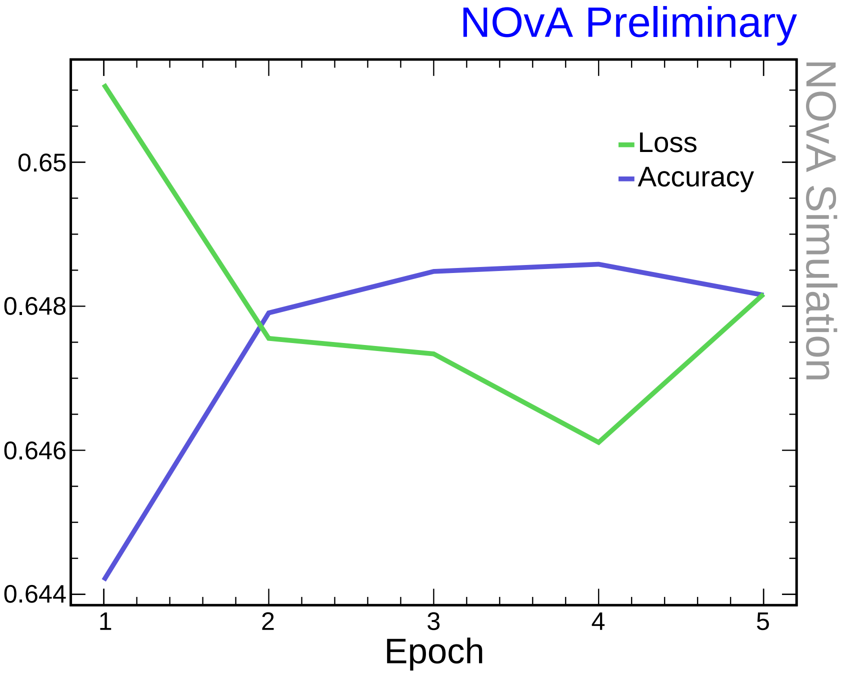

Figure 5.24: Accuracy and loss for the last 5 variable training with early stopping based on loss.

After optimising the learning rate, the early stopping was applied to the training to identify the ideal number of epochs for the best training results. The result are seen in figures 5.23 and 5.24. It is difficult to say confidently how much impact applying early stopping has on the number of events that can be recovered. Again, the shape of the PID plot has unexplained features that were not seen when running on NOvA software, consistent with the previous training with these variables. As discussed, the training variables show underlying structures that can be seen in the results, but it does not explain all the features.

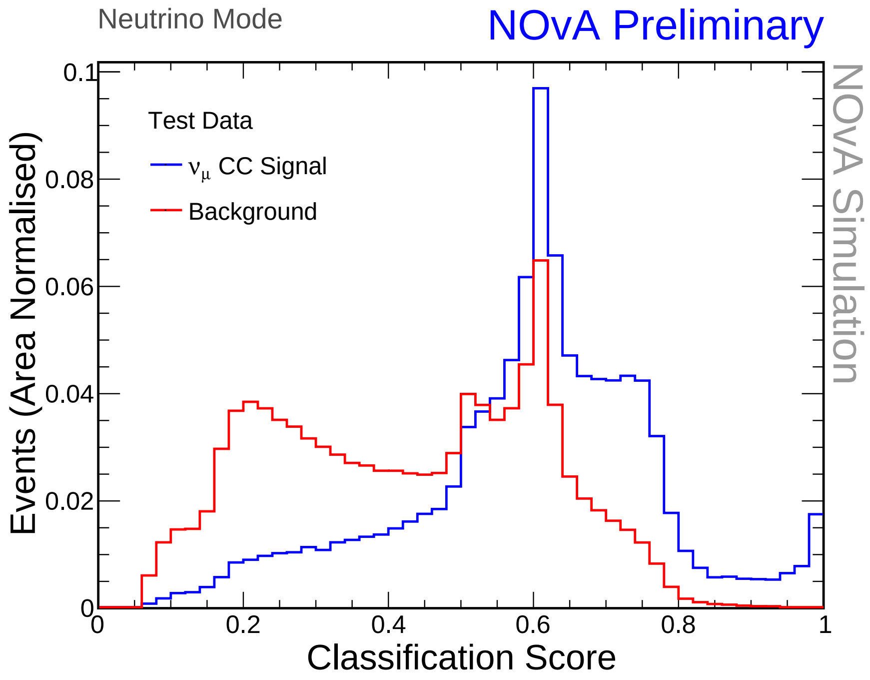

Figure 5.25: Classification score plot from training with 5 variables using early stopping and standard scaling for variables. The y-axis has been normalised by area to show a probability instead of absolute event numbers.

The next improvement implemented to the training was used to standardises the input variables in the training. This should remove any proportional dependency on variables with higher mean values, and could potentially explain the PID shape, if one of the variables with higher mean skews the results. The PID score for this training is seen in figure 5.25, and the loss and accuracy in figure 5.26. The shape of the distribution is notably different than before applying the scaler, however the signal and background still do not peak at 1 and 0 respectively, and it cannot be determined whether the scaler provided any additional improvement to the network.

5.2.6 Network Including Events Above 5 GeV

The events above 5 *GeV* are not used in the NOvA analysis, since these events are far from the oscillation maximum. However, it was tested if including these in the training would improve the results. The variables

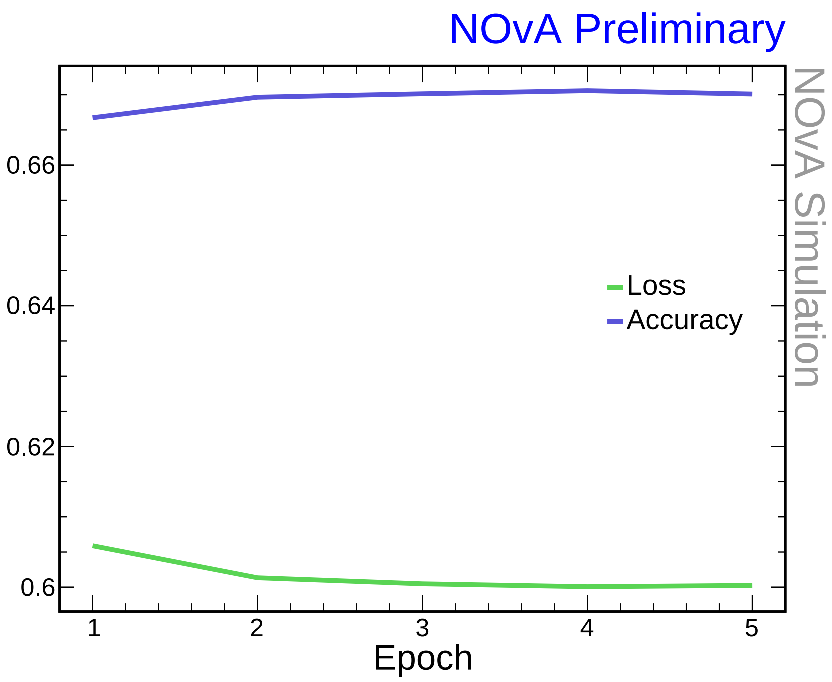

Figure 5.26: Accuracy and loss for the last 5 variable training with early stopping and a standard scaler applied.

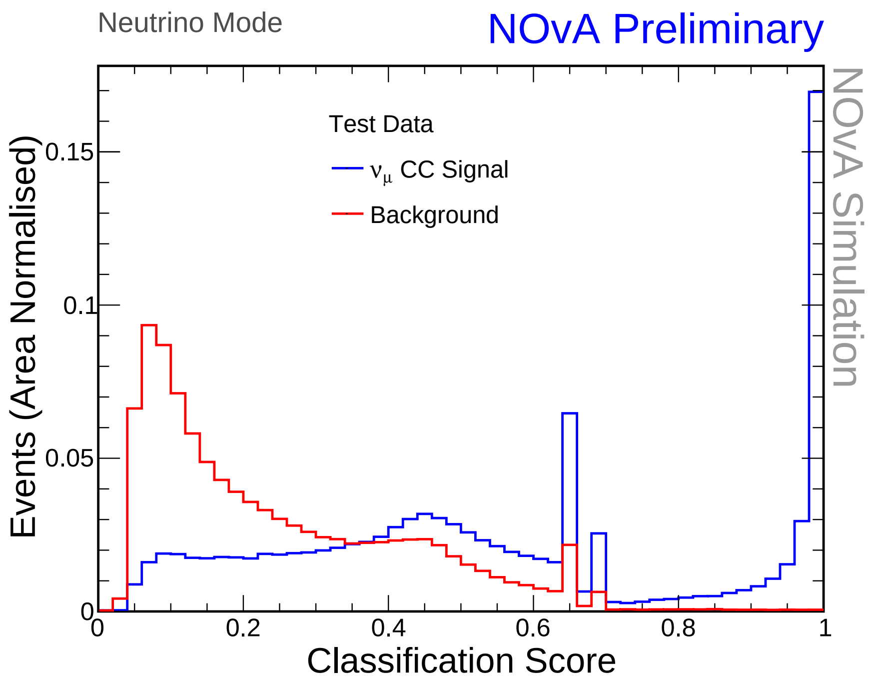

Figure 5.27: Classification score plot from training that includes events with energy above 5 GeV. The y-axis has been normalised by area to show a probability instead of absolute event numbers. 5 variables and 5 epochs were used in the training.

and the network architecture was kept the same as for the previous network. The results are seen in figures 5.27 and 5.28 showing the PID plot as well as loss and accuracy respectively.

The separation between signal and background improves significantly

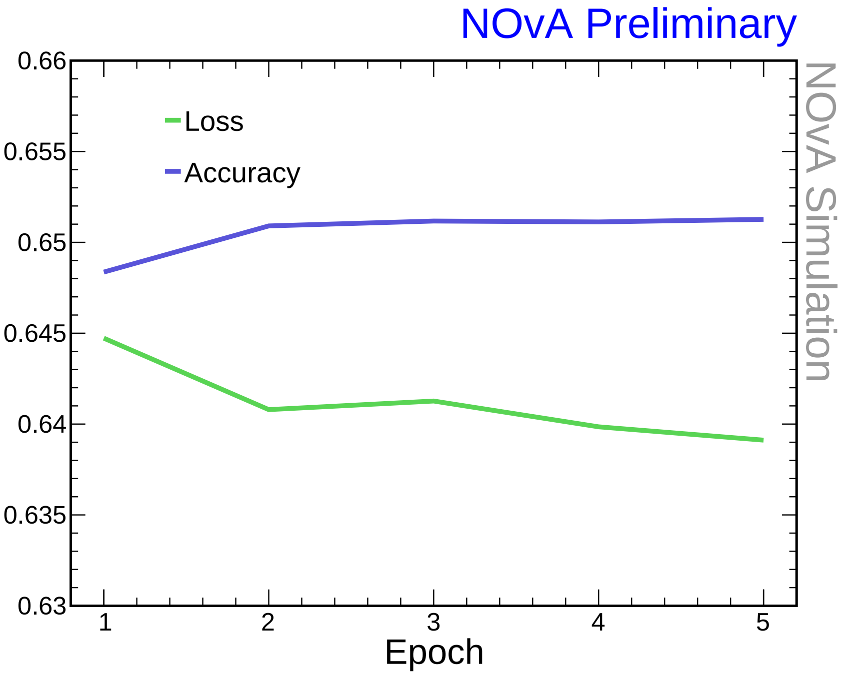

Figure 5.28: Plot showing loss and accuracy for training splitting the final results into signal and background when events with neutrino energy higher than GeV are included. 5 variables and 5 epochs were used in the training.

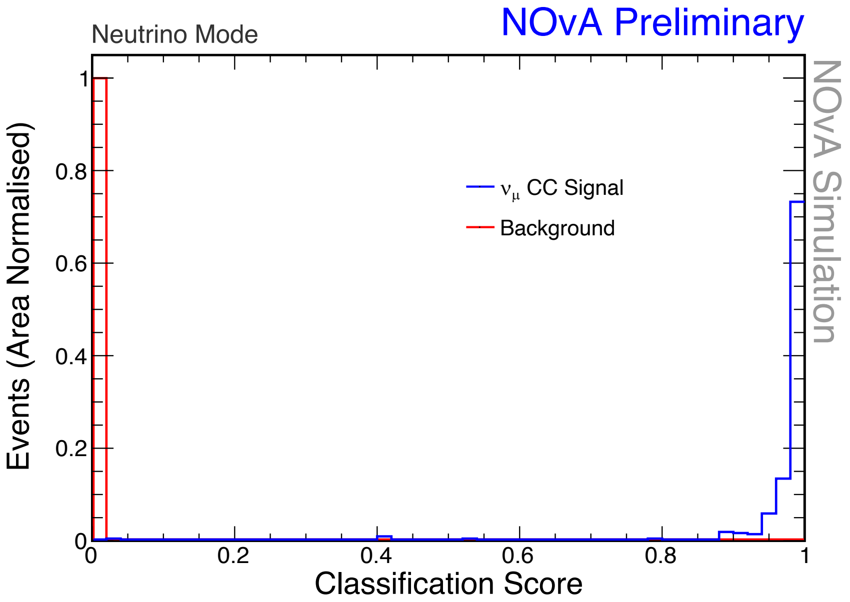

Figure 5.29: PID score for only events above 5 GeV. The y-axis has been normalised by area to show a probability instead of absolute event numbers. 5 variables 5 epochs were used in training.

compared to other trainings. To check which events this additional separation affects, the PID score was only plot for events above 5 GeV, seen in figure 5.29. Unfortunately, the improvement appears to only be for events above 5 GeV and this training is not expected to bring any additional events that could be recovered below 5 GeV. The two peaks at

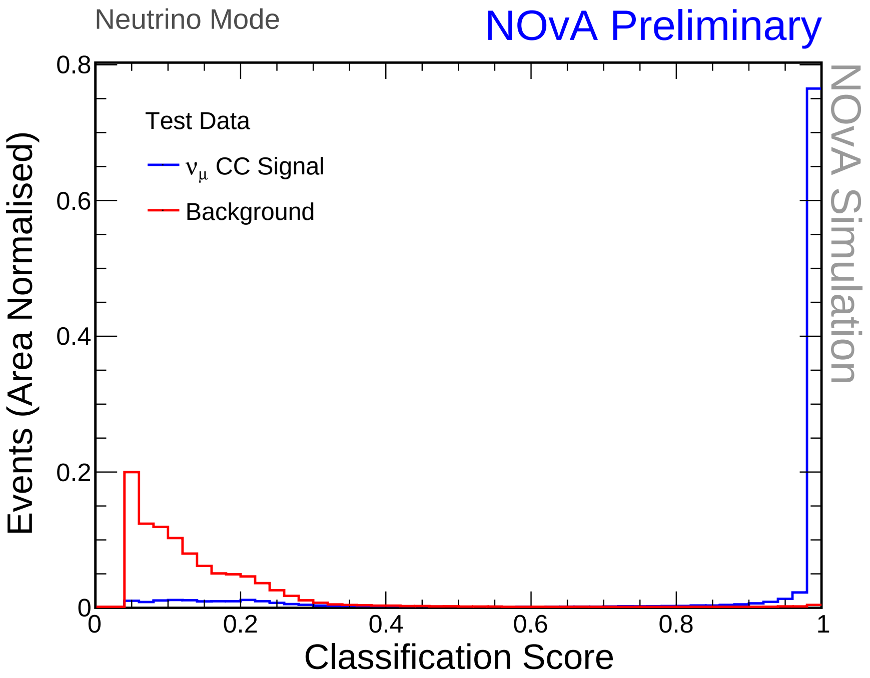

Figure 5.30: Classification score plot training only events that have neutrino energy above 5 GeV. The y-axis has been normalised by area to show a probability instead of absolute event numbers. 5 variables 5 epochs were used in the training.

about PID score 0.7 seem to account for all the events less than 5 GeV, whereas the good separation is only for the higher energy events.

This was further tested on training only the events above 5 GeV, and the PID score plot is seen in figure 5.30. This plot has a similar separation between signal and background as in figure 5.29 and has no peaks at PID score of 0.7. The only notable difference is that the background peak is less pronounced at 0, but that could be due to low number of epochs. This further proves all actual gains in separation was for events above 5 GeV, so this network was not used for any further testing.

As have been discussed previously, there is a structure to the variables stemming from the differences between QE and DIS/RES events. From this, a working hypothesis can be constructed that assumes that dividing the events to two categories, those events with reconstructed neutrino energy of less than 1.5. GeV and those events above, could give interesting results. Events with low energy have disproportionately lot of QE events (as well as some RES and DIS events) whereas the higher energy events are almost all DIS and RES events. These networks could get rid some of the structures present in the previous PID plots and to provide more insight to the results of the networks described. This test was done by using the 5 variables and the same architecture as used in the previous models. However, these results did not give any improvement but rather produced worse results. There are multiple explanations for this, one possible is that the variables and the architecture do not work for these networks but rather would have to be specifically optimised which there was no time for. Also, especially the sample of events below 1.5 GeV was much smaller. Much more time would have had to be spent on optimising this network if better results were wanted.

5.2.7 3 Variable Network

Training variables
Hadronic Energy
Transverse Muon Momentum
Number of Hits in a Hadronic Track

Table 5.5: The 3 variables trained in the network.

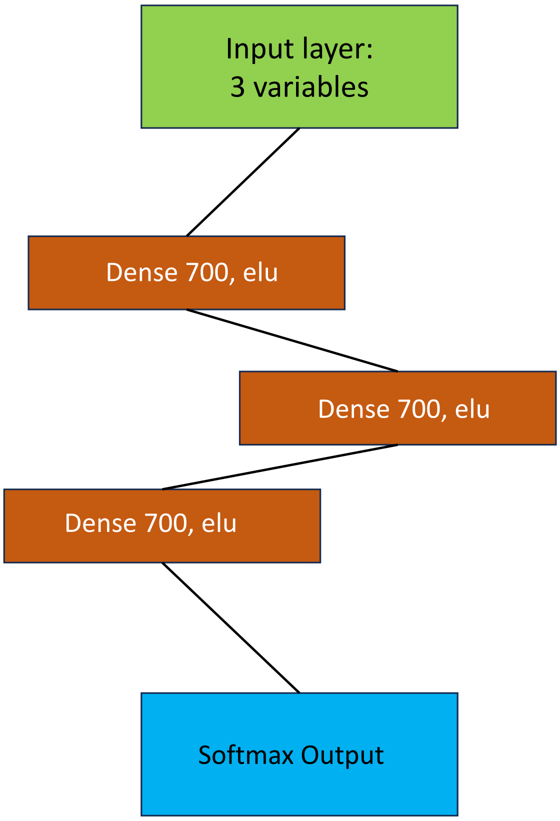

Figure 5.31: Model architecture for the 3 variable network. The input layer is connected to a dense layer of size 700, which is connected to 2 additional dense layers of size 700. The final layer uses a softmax activation function.

A training with only three variables was tested to compare to the earlier 5 variable model. The three chosen variables were hadronic energy, transverse muon momentum and number of hits in a hadronic track. These are all also used in the earlier trainings, their spectra seen in figures 4.22, 5.8 and 5.9, and the MC-data comparison plots in appendix B as before. The variables were chosen based on testing combinations of the earlier 5 variables, and the chosen three were found to produce the most separation. The model architecture is slightly different than for the 5 variable model, and is seen in figure 5.32. A similar procedure of trial and error was used to find the best option. The main difference is that the size of the hidden layers was reduced, as the number of variables was reduced.

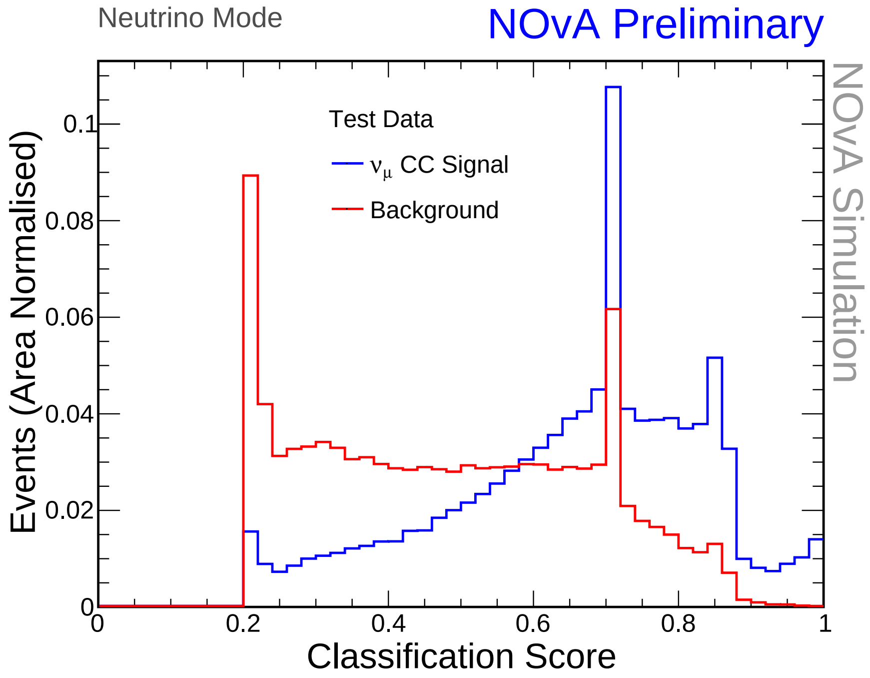

Figure 5.32: Classification score plot from training using 3 variables for 20 epochs. The y-axis has been normalised by area to show a probability instead of absolute event numbers.

Results from the training with 3 variables for 20 epochs are seen in figures 5.32 and 5.33. The results look promising, so they were tested further to see how much signal could be gained alongside the earlier training with 5 variables. The reason for the better performance of the 3 variable network could be due to multiple factors, for example the removal of highly correlated variables. Correlations between the training variables were not considered in this study, but it is known that some of the variables are dependent on each other. For example the hadronic energy fraction is dependent on the total hadronic energy, which can impact the behaviour of the network. Also, the architecture of the network was chosen on trial and error basis, and thus it is possible that the architecture found for 3 variables was better than that used for 5 variables.

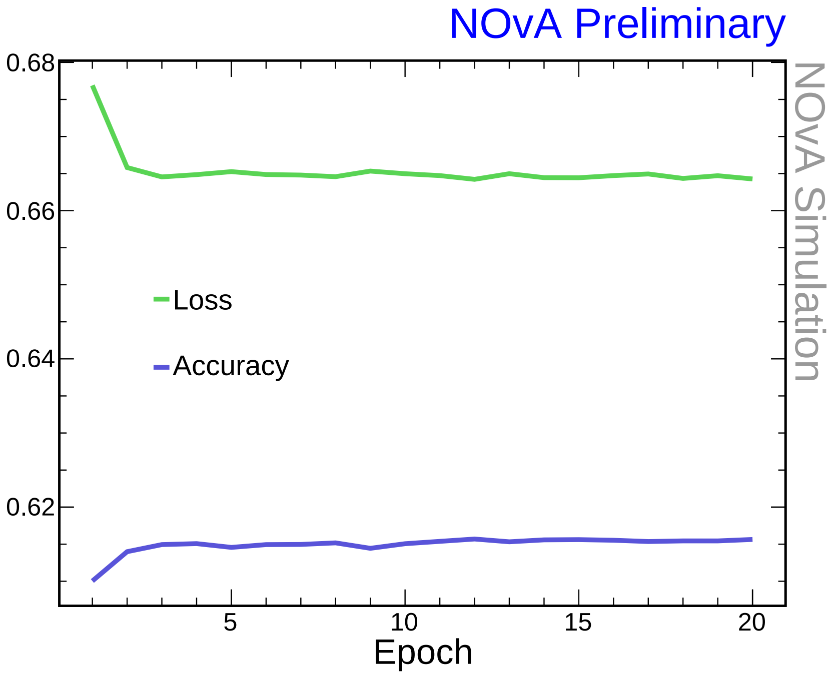

Figure 5.33: Plot showing loss and accuracy, using 3 variables and training for 20 epochs

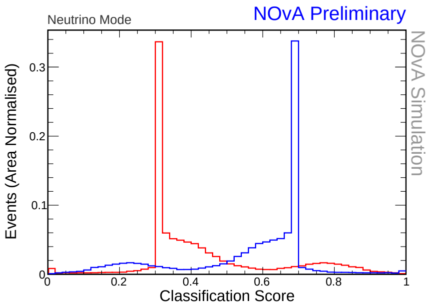

Figure 5.34: PID score for events above 1.5 *GeV* using NOvA FD FHC simulation using 3 variables and training for 20 epochs. Blue line indicates signal and red background. The y-axis has been normalised by area to show a probability instead of absolute event numbers.

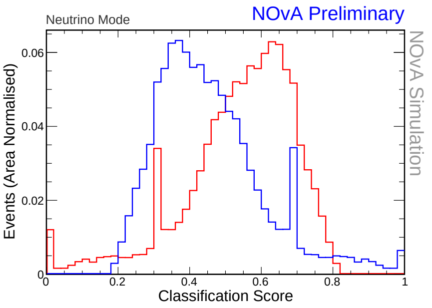

Figure 5.35: PID score for events below 1.5 *GeV* using NOvA FD FHC simulation using 3 variables and training for 20 epochs. Blue line indicates signal and red background. The y-axis has been normalised by area to show a probability instead of absolute event numbers.

To further understand the shape of the classification score seen in figure 5.32, two additional PID plots were created. Figures 5.34 and 5.35 show the PID distribution for events above 1.5 *GeV* and below 1.5 *GeV* respectively. As previously discussed, the low energy events have most of the QE events and the higher energy events are mostly DIS and RES events. Based on these results, it appears that the background peak at a PID score of 0.7 is mostly a contribution from low energy events and the signal peak at the same value are high energy events. The purest signal at 1 seems to be a contribution from both low and high energy events. The peak at 0.2 is more complicated, since these plots peak closer to 0.3 than 0.2. If this is assumed to be the effects of calculating the PID in test data, then it can be assumed that the peak is mostly high energy background events.

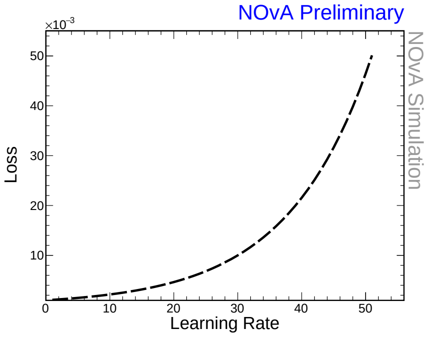

Figure 5.36: Loss vs. learning rate for the 3 variables model when learning rate is varied according to 5.2.3

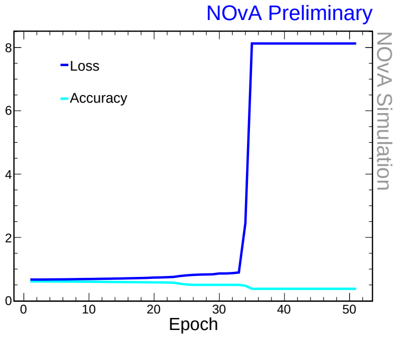

Figure 5.37: Loss and accuracy for the 3 variables model when learning rate is varied according to 5.2.3

The same training optimasation tests were performed for this training as for the previous 5 variable network (see section 5.2.5). The learning rate test was performed first, and the results are seen in figures 5.36 and 5.37. It was also clear that there was no need to change the learning rate for this model as the standard learning rate produces the lowest loss during the test.

Results from training after early stopping is applied are seen in figures 5.38 and 5.39. Applying early stopping appears to bring improvement, but this has to be tested against the training without early stopping to verify. The difference in PID plot compared to the training with 20 epoch (figure 5.32) is that a higher peak at PID score of 0.7 is seen, which makes the background peak at 0.2 and the signal peak above 0.8 appear smaller.

Adding a standard scaler to the 3 variable model produces a network described in figures 5.40 for PID and 5.41 for loss and accuracy. The only difference for this network compared to the network with no scaling (figure 5.38) is that the peak at 0.7 shifts to lower classification score to approximately 0.58. Also, the less prominent background peak moves from 0.3 closer to PID score of 0.2. This is not expected to bring any change compared to the version of the training with no standard scaler, so it was not used in any further analysis.

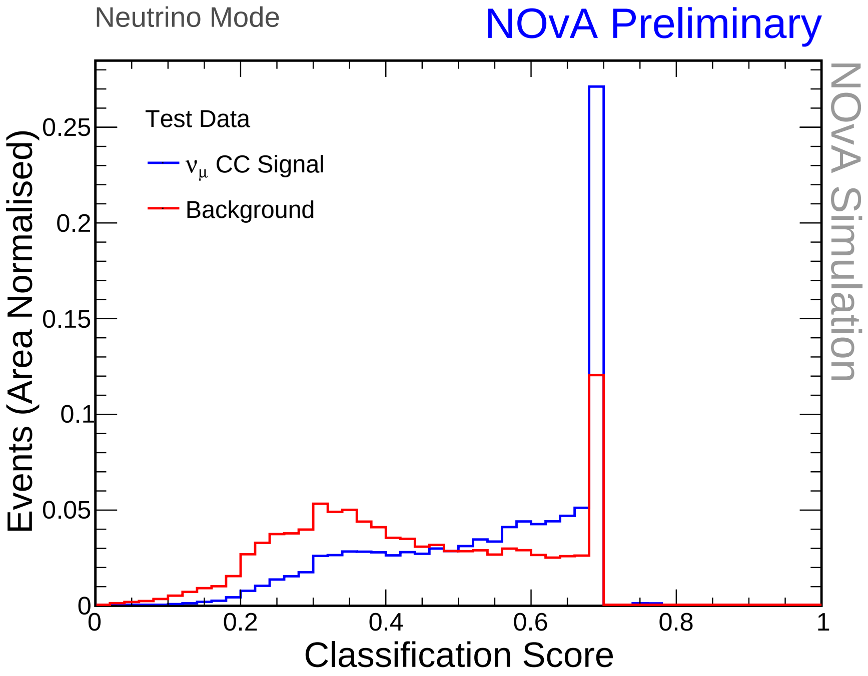

Figure 5.38: Classification score plot from training with 3 variables using early stopping. The y-axis has been normalised by area to show a probability instead of absolute event numbers.

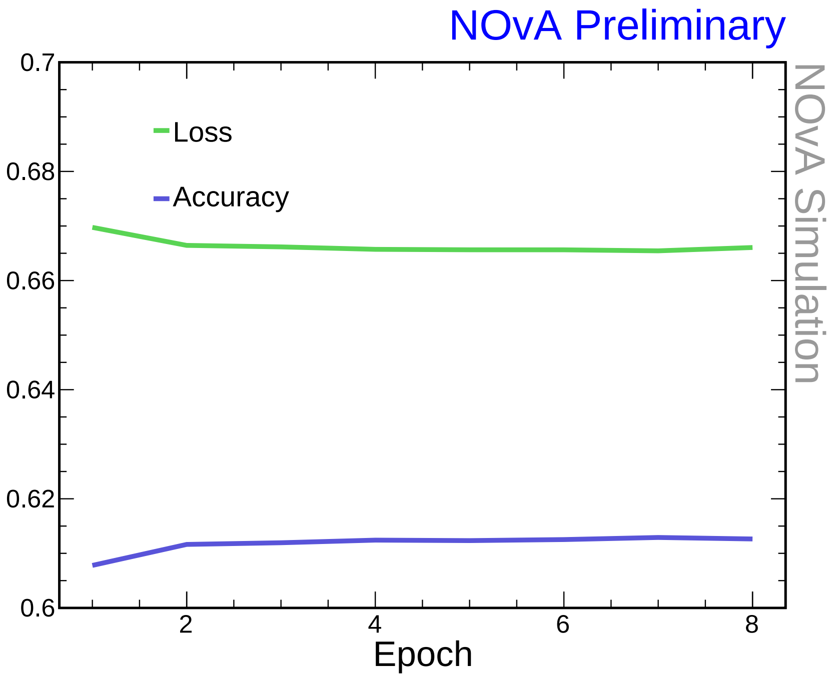

Figure 5.39: Accuracy and loss for the last 3 variable training with early stopping based on loss applied.

5.3 Sensitivity Improvement

After the trainings were completed and the networks saved to a h5 file format, they were converted to protobuf (pb) file format to be integrated into NOvA software. Prod5.1 FD FHC simulation data is used and run

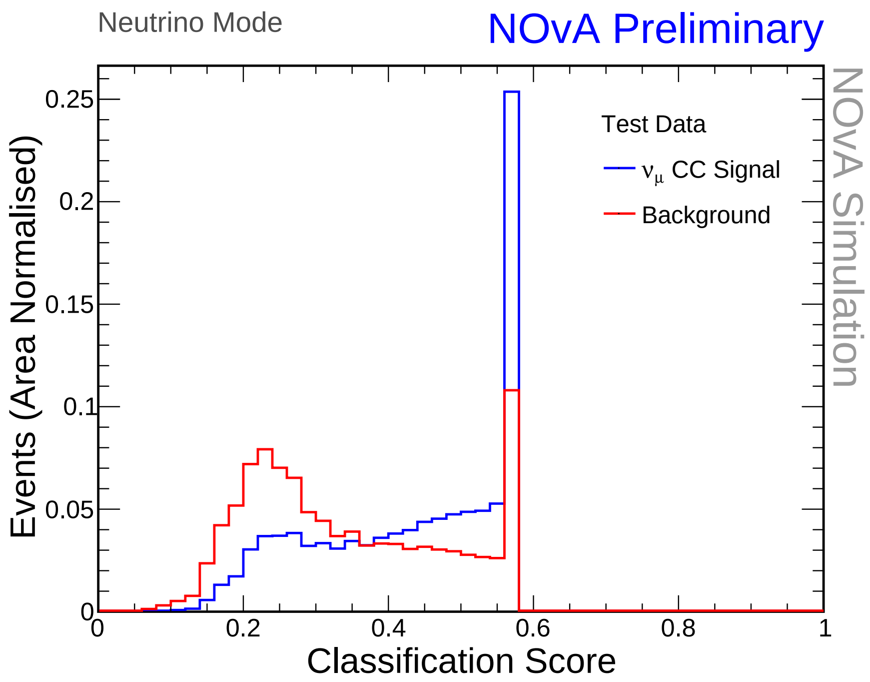

Figure 5.40: Classification score plot from training with 3 variables using early stopping and a standard scaler. The y-axis has been normalised by area to show a probability instead of absolute event numbers.

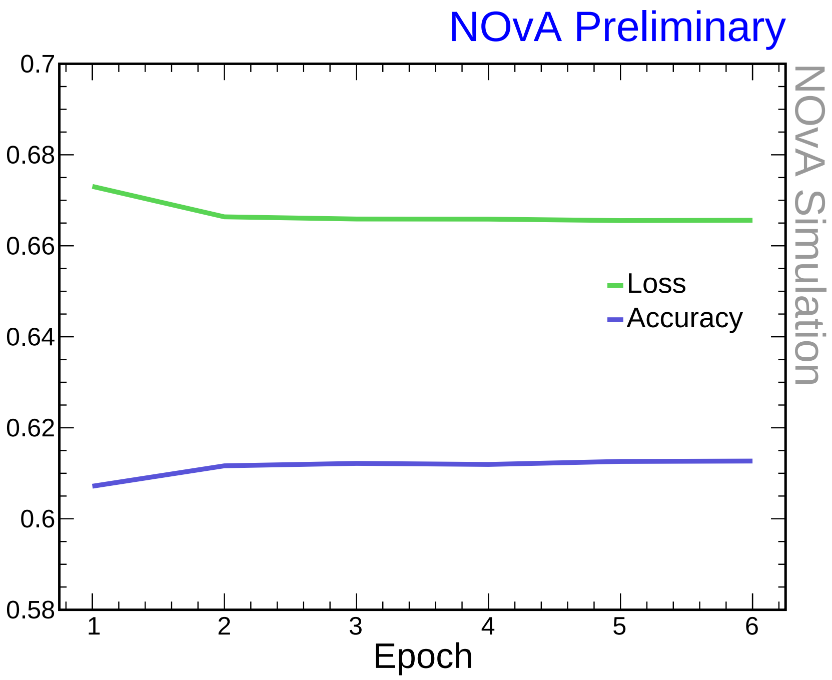

Figure 5.41: Accuracy and loss for the last 3 variable training with early stopping and standard scaling.

through the training. Using the results from this run, a cut based on the classification score is applied to investigate how much signal (and background) can be recovered to the oscillation analysis. These events can then be used to test how much sensitivity to the oscillation parameters is gained by adding the events as a separate sample to the fitting.

As was discussed in section 5.2.5, it was clear that the behaviour of the 5 variable network cannot be fully explained. It was still tested whether the results could be used to extract the number of signal events that could be added to the analysis using this network. However, the PID distributions (figures 5.19, 5.24 and 5.25) did not correspond to what the spectrum looks like when running the network through NOvA software. Integrals showing the number of events using two cuts on PID distribution (0.6 and 0.8) were also calculated from both the original PID distribution as well as those produced from NOvA software, and they did not agree. Possible explanations include something changing in the behaviour of the file when converting from h5 to pb, or some other incompatibilities with the NOvA software with this network. Even after long and rigorous testing no feasible explanation was found, and thus it was decided only the results from the 3 variable network were used in this final part of the analysis.

5.3.1 Reconstructed Neutrino Energy

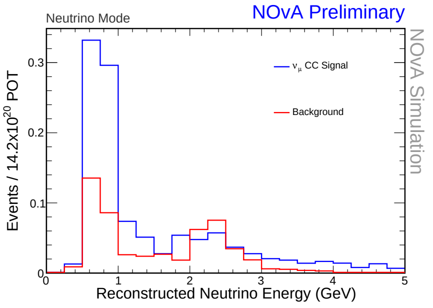

Figure 5.42: Final energy spectrum after 0.8 cut on classification score using 3 variable model trained for 20 epochs.

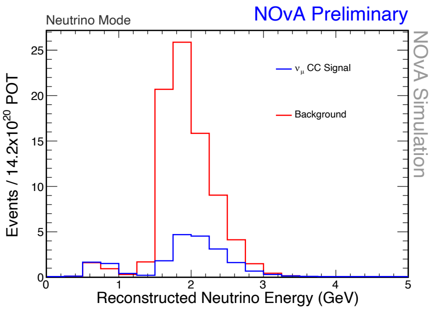

Figure 5.43: Final energy spectrum after 0.5 cut on classification score using 3 variable model trained for 20 epochs.

The reconstructed neutrino energy spectra for the 3 variable network for PID cuts 0.8 and 0.5 are seen in figures 5.42 and 5.43. All of these plots are for the networks before learning rate test, early stopping or standard scaler. All of the plots in this section include the effect of oscillations, since those figures give the best estimate on how much signal is gained when data is used. The same plots using early stopping are seen in figures for 3 variable model in figures 5.44 and 5.45 for 0.8 and 0.5 cuts respectively.

From these plot, it can be investigated that there is a clear cut between 0.5 and 0.8, where most signal (and also more background) is gained with the smaller cut value. Also, applying early stopping improved the distribution for this network by introducing more signal.

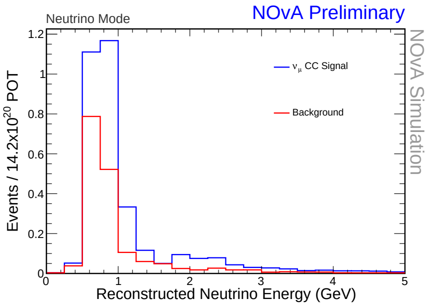

Figure 5.44: Final energy spectrum after 0.8 cut on classification score using 3 variable model with early stopping applied.

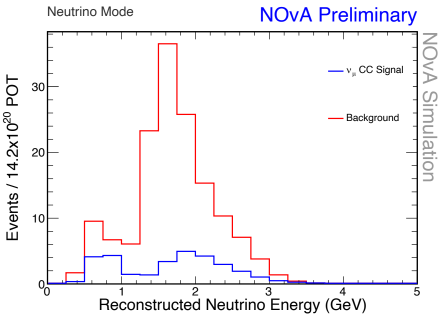

Figure 5.45: Final energy spectrum after 0.5 cut on classification score using 3 variable model with early stopping applied.

5.3.2 Figure of Merit

To properly evaluate the effect of adding these events into the analysis, a figure of merit (FOM) was used to compare the results. The FOM used in these calculations was defined as $FOM = \frac{S}{\sqrt{S+B}}$, where S is the number of signal events and B the number of background events between reconstructed neutrino energy of 0 and 5 GeV. This is the same that was used in the optimisation of cuts in the last analysis 3.3.4, and earlier in this chapter in section 5.2.4.

Model	PID Cut	Signal	Background	FOM
3 var model 20 epoch	0.8	1.13	0.535	0.875
3 var model 20 epoch	0.7	4.41	4.04	1.52
3 var model 20 epoch	0.65	16.7	57.02	1.94
3 var model 20 epoch	0.6	21.1	82.4	2.07
3 var model 20 epoch	0.5	28.3	136	2.21
3 var model 20 epoch	0.4	34.9	204	2.26
3 var model early stop	0.8	3.28	1.71	1.27
3 var model early stop	0.7	6.74	9.39	1.86
3 var model early stop	0.65	21.5	68.4	2.26
3 var model early stop	0.6	24.5	90.9	2.28
3 var model early stop	**0.5**	**31.4**	**148**	**2.34**
3 var model early stop	0.4 cut	35.9	212	2.28

Table 5.6: Number of signal and background as well as FOM for various 3 variable training models with different cuts on classification score.

The FOM calculated for 3 variable models are seen in table 5.6. The model with standard scaler applied was not included since the results are expected to be identical to those with only early stopping applied as discussed. The best performing network out of these was with a cut of 0.5 giving an FOM of 2.34. The model with 20 epochs appears to give lower values for all the cuts included here, and it can be expected that cuts lower than 0.35 will not improve the results as primarily background will be added. The clear cut between the peak at 0.7 and 0.65 is apparent from these results, which was also seen in the PID distributions.

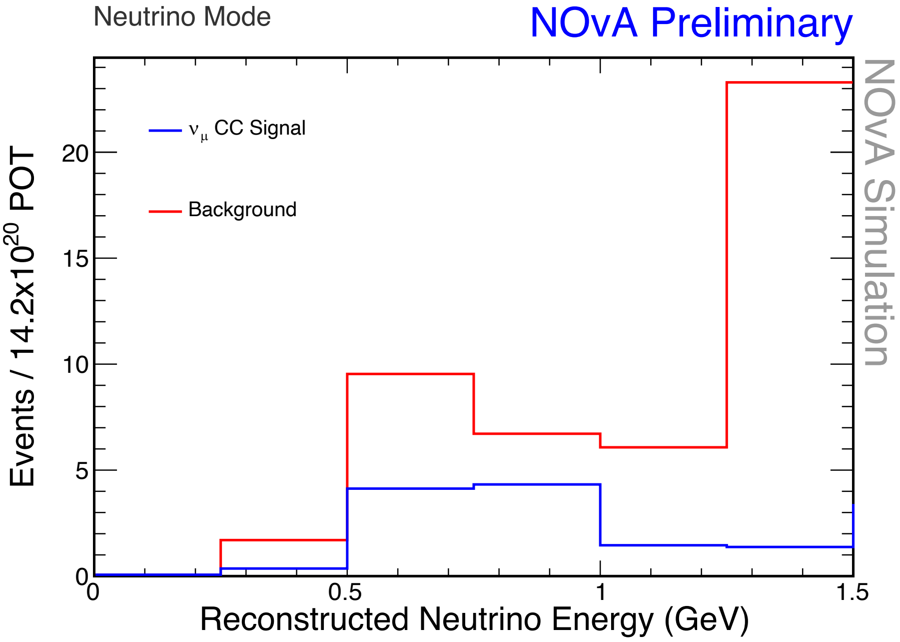

Figure 5.46: Reconstructed neutrino energy spectrum after applying a PID cut of 0.5 using 3 variable network for events between energies of 1 and 5 GeV.

One can also calculate FOM for a certain limit for a chosen model. For example, including only events between 0 and 1.5 GeV for 3 variable model with early stopping and a cut of 0.5, spectrum seen in figure 5.46, gives an FOM of:

$$\frac{11.6}{\sqrt{11.6 + 45.8}} = 1.22. \tag{5.3.1}$$

This is obviously lower in terms of FOM but gains back the low energy events with less background so those could be a beneficial addition to the analysis.

5.3.3 Sensitivity Contours

The improvement to the sensitivity to oscillation parameters is evaluated by adding the recovered events as a separate sample using the cut on the PID score to the fit. The theoretical upper maximum improvement to sensitivity is seen in figure 4.29, but much lower improvement is expected from the results from the network. All of the fits were made without FC corrections (see section 3.7) in the interest of time and simplicity. Table 5.7 shows the starting values for oscillation parameters used in the fit as well as parameters kept constant in the fit. All of these were the values

Fitting Parameters
$\Delta m^2_{23} = 2.41 \times 10^{-3} eV^2$
$\theta_{23} = 0.568$
$\delta_{CP} = 0.82\ \times\pi$
$\Delta m^2_{21} = 7.53 \times 10^{-5} eV^2$ [10]
$sin^2(\theta_{12}) = 0.307$ [10]
$sin^2(\theta_{13}) = 0.0210 \pm 0.0011$ [10]

Table 5.7: Parameters used in the oscillation fits.

also used in the 2020 oscillation fits and the same as used when producing figure 4.29.

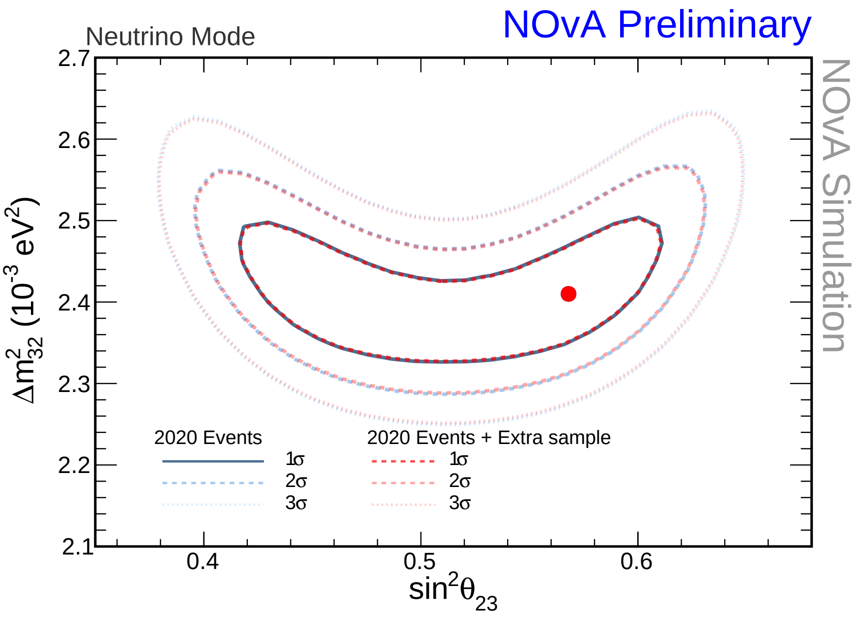

Figure 5.47: The sensitivity contour for 3-flavour oscillation parameters for the 2020 3-flavour analysis sample as well as the new sensitivity after the extra sample has been added to the fit. 3 variables with early stopping applied, 0.5 cut on classification score.

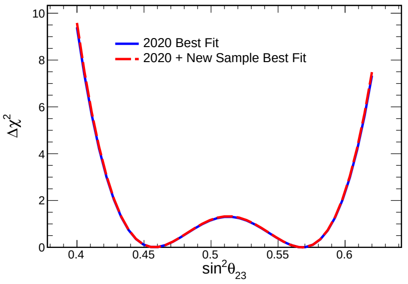

Figure 5.48: 1D fit to $sin^2(\theta_{23})$ before and after adding events from 3 variable early stopping model with 0.5 cut on classification score.

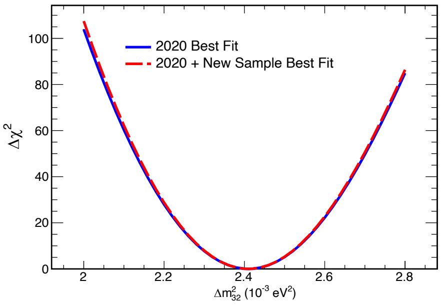

Figure 5.49: 1D fit to Δm^2_{32} before and after adding events from 3 variable early stopping model with 0.5 cut on classification score.

The improved sensitivity contour when events from the 3 variable early stopping model with a cut of 0.5 on PID are added as a separate sample, is seen in figure 5.47. There is a minor difference in the contours shrinking at 2 and 3 σ levels when the new sample is added to the fit. The increase in sensitivity is equivalent to 4.3% increase in effective POT which is equivalent to approximately 33 days of extra data taking (assuming that NOvA is collecting data 24 hours per day). Similarly, plots showing a 1D fit to the two oscillations parameters, $sin^2(\theta_{23})$ and Δm^2_{32}, were produced and are presented in figures 5.48 and 5.49 respectively. From these plots, it appears that the best fit point is approximately the same and the improvement occurs at higher values of $\Delta\chi^2$.

The results from these plots show that most of the sensitivity is gained at the highest σ level and the change is minor. However, there are other ongoing improvements to the analysis, which together with adding this extra sample could bring more significant improvement.

5.4 Changing Current Cuts

As seen in figure 4.6, most of the rejected signal used in the trained are those that failed the PID cuts. The effect of relaxing these cuts was investigated, to study what would be the gain in signal and how much background would be introduced to the analysis. This provides useful information and a point of comparison to the improvement from the network training.

Figure of Merit

The FOM used in these calculations was defined as $FOM = \frac{S}{\sqrt{S+B}}$ as before. This ensures these results are comparable to the results from the network. For comparison, the current PID cuts are set as RemID > 0.3 and NumuID > 0.8 (see section 3.3.4), and the cosmic cut as > 0.45 (see section 3.3.3).

All the signal and background numbers as well as FOM values for different changes to current cuts are seen on table 5.8. The best score of the tested combinations is changing the cuts to RemID > 0.05 and NumuID > 0.5. Figures 5.50 and 5.51 show the total spectrum after adding the events on top of the current selected events and the spectrum for only the added events respectively.

The only value to change for cosmic cuts is to lower from > 0.45 to > 0.4 to be equivalent to the loose cosmic rejection cut applied to decaf files. The effect of adding these cosmic cuts gives an FOM of 0.094, which

Cut Change	**Signal**	**Background**	**FOM** ($\frac{S}{\sqrt{(S+B)}}$)
RemID > 0.2	1.47	0.19	1.14
RemID > 0.1	3.05	0.41	1.64
RemID > 0.05	3.66	0.5	1.80

Cut Change	**Signal**	**Background**	**FOM** ($\frac{S}{\sqrt{(S+B)}}$)
NumuID > 0.7	3.87	0.84	0.39
NumuID > 0.6	6.29	0.41	2.22
NumuID > 0.5	8.08	2.65	2.47
NumuID > 0.4	9.57	3.80	2.62

Cut Change	**Signal**	**Background**	**FOM** ($\frac{S}{\sqrt{(S+B)}}$)
RemID > 0.2, NumuID > 0.7	5.73	1.18	2.18
RemID > 0.1, NumuID > 0.6	11.1	2.95	2.96
RemID > 0.05, NumuID > 0.5	**15.12**	**5.17**	**3.36**
CosmicID > 0.4	3.704	1.19	0.094

Table 5.8: Number of signal and background as well as FOM after changing the current analysis cuts. The first section describes changing the RemID cut but keeping NumuID the same, second section changing NumuID but not RemID and the last section changing both RemID and NumuID as well as changing CosmicID.

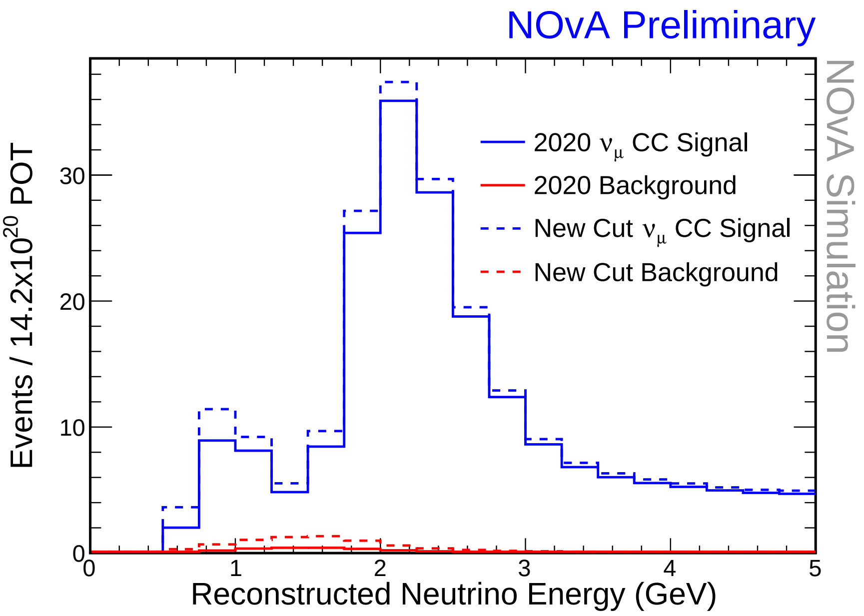

Figure 5.50: Reconstructed neutrino energy spectrum before and after changing the current PID cuts to RemID > 0.05 and NumuID > 0.5.

is much lower than that for PID cut, so no further testing was done.

Sensitivity Contours

The sensitivity plots after adding the additional events from changing the PID cuts as a separate sample are seen in figures 5.52 (2D fit), 5.53 (1D Δm^2_{23} fit) and 5.54 (1D $sin\theta_{23}$ fit). Based on these improved sen-

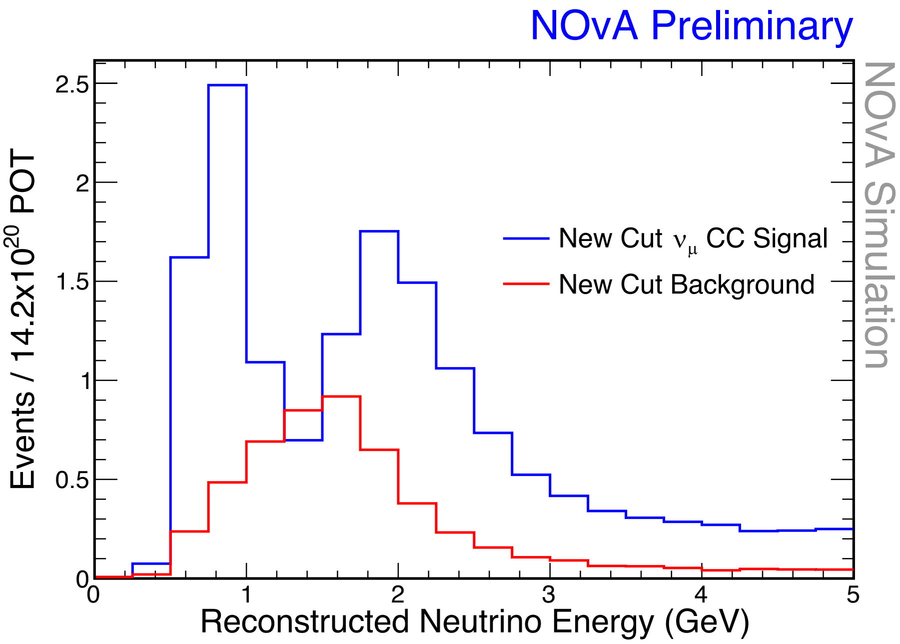

Figure 5.51: Reconstructed energy spectrum of events added to the analysis after changing the current PID cuts to RemID > 0.05 and NumuID > 0.5.

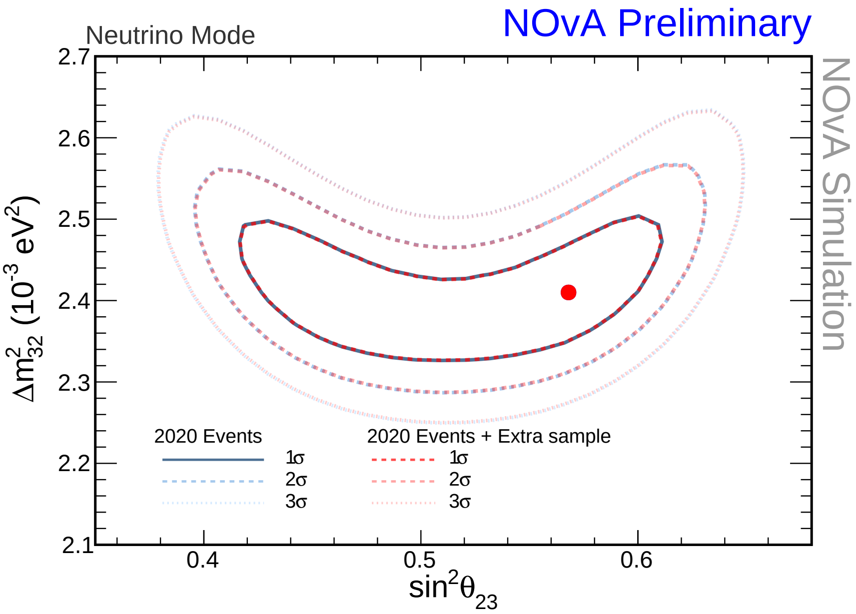

Figure 5.52: The sensitivity contours for 3-flavour oscillation parameters for the 2020 analysis sample as well as the new sensitivity after adding a sample of variables given by changing current PID cuts to RemID > 0.05 and NumuID > 0.5

sitivities the increase in effective POT is 3.7% which is equivalent to approximately 28 days of data taking. These values are very close to results from the sensitivity improvement using the neural network, although slightly smaller. This is probably due to this sample having less signal, even though the number of background is smaller and the FOM is higher.

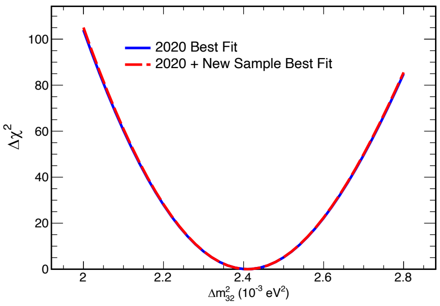

Figure 5.53: 1D fit to Δm^2_{32} before and after adding events by changing the PID cut to RemID > 0.05 and NumuID > 0.5

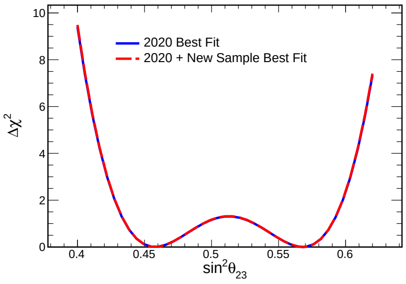

Figure 5.54: 1D fit to $sin^2(\theta_{23})$ before and after adding events by changing the PID cut to RemID > 0.05 and NumuID > 0.5

5.5 Summary and Further Work

Multiple neural network trainings were tested to find the most efficient way to recover events that are currently rejected from the NOvA 3-flavour disappearance analysis back into the oscillation analysis to gain some additional FOM and increase the sensitivity to oscillation parameters. The best results were for a network with 3 variables after applying early stopping based on loss with a cut of 0.5 on classification score. This gave an additional FOM of 2.34 for all energies and FOM of 1.22 for events between 0 and 1.5 GeV. The change in sensitivity is very small (4.3% gain in effective POT, about 33 days of extra data taking), but together with other improvements applied to the current oscillation analysis, adding these events as en extra sample could be beneficial.

These values can be compared to results from changing the current analysis PID cuts. Unfortunately, the FOM value was smaller than changing the current PID cut to RemID > 0.05 and NumuID > 0.5, which gave an additional FOM of 3.36. However, the number of signal events is much higher using networks than changing the current cuts, and the network brings more events at the lower energy range that is is rejected most. Also, the change in sensitivity is slightly smaller after these cut changes, equating to a 3.7% increase in effective POT and about 28 days of extra data taking.

There was no time to run this analysis using NOvA data, since the release of new NOvA oscillation analysis was pushed back to 2024 due to issues with gathering enough data to double the collected FHC POT that was planned for the new analysis. There was a shut down due to Covid-19 pandemic and hardware issues in the beamline that slowed down the data taking. Otherwise, the results could have been checked with data and possibly be added into the new oscillation analysis as separate sample to the fit to bring a small improvement to the sensitivity.

Given more time, statistical uncertainties could have been taken into account and tested how the different variables used in the training are susceptible to the effects of systematics. Another improvement would be to check for correlations between the training variables. As it was discussed in section 5.2.7, this is possibly the explanation for the improvement on the training with reduced variables, and it could further the understanding of the behaviour of the network.

Also, additional testing on the behaviour of the 5 variable model could bring some insight into how it could be improved and why it behaves so differently on the small test sample when calculating the classification score. Some explanations, like the shape of the training variables or errors when converting the h5 network file to pb file, were discussed but a more in depth analysis would further improve the understanding. Further improvements could include plotting the output for various variables using different PID cuts and better optimasation of the network architecture. This might allow for this network to be used in further analysis to see its impact on the sensitivity, as in this thesis it was not included on the sensitivity testing.

In section 5.2.5, it was suggested that training with just the two variables, number of hits in hadronic track and reconstructed transverse momentum, could be constructed to further the understanding of the training variables and the unexplained behaviour of the constructed networks. Later, in section 5.2.6, it was briefly discussed that dividing the events into two categories: events with reconstructed neutrino energy more than 1.5 GeV and less than 1.5 GeV, could be trained separately. As mentioned, this was briefly tested using previously used variables and network architecture, and it did not improve results. With more, this training could have been optimised better in terms of variables and architecture, possibly given a better improvement in gained events into the analysis. Also, using class weights to balance the number of QE, RES and DIS events in the training sample could be tested to see if it benefits teh results.

Another study similar to this analysis would be to construct a training using a completely different machine learning method. BDTs are used in NOvA in the construction of the cosmic rejection cut (see section 3.3.3), so they seem like the most interesting method to test. This could bring meaningful insight and a point of comparison to these results. It can especially improve the integration to NOvA software, since these methods are already used in the analysis.

Chapter 6

Conclusions

6.1 Conclusions

In this thesis a new way to improve the sensitivity to oscillation parameters Δm^2_{23} and $sin\theta_{23}$ in NOvA ν_μ disappearance analysis was introduced. Using neural networks, signal rejected from the previous analysis was trained to look at effects of recovering these events.

The current NOvA oscillation analysis selection has an efficiency of 80% for selecting ν_μ CC events for the disappearance analysis. Most of the rejected events are those that failed the particle identification cut, and the events are not all one type but have different properties. It was found that slightly more QE events are selected compared to RES and DIS ν_μ CC events, and a lot more events with low energy (less than 1.5 GeV) are not selected.

Multiple different networks using different architectures and variables were tested. In the end, the impact on the sensitivity was studied using a network that trained 3 variables (hadronic energy, transverse muon momentum and hadronic energy per hits) using all events that have reconstructed neutrino energy between 1 and 5 GeV. These variables were chosen based on their K-S test scores as well as their distributions between signal and background trying to find variables with the most separation power. The final training included early stopping applied to loss for optimal number of epochs. The use of a standard scaler and changing the learning rate were also tested, but those did not bring additional improvements to the network.

When run through NOvA software testing different cuts on the classification (PID) score, the highest value of added FOM ($\frac{S}{\sqrt{(S+B)}}$) (2.34) was achieved with a cut of 0.5 on the classification score. The sensitivity improvement after adding this sample was also tested, although changes to the sensitivity were marginal. The increase in effective POT was 4.3 % which is equivalent to approximately 33 days of extra data taking. Comparing this to changing the current PID analysis cuts, the FOM using the neural networks was smaller. The sensitivity was tested on changing the current PID cut to RemID > 0.05 and NumuID > 0.5 (FOM 3.36) which gave an increase in effective POT of 3.7% equal to 28 days of additional data taking, which is smaller than the increase in sensitivity using neural networks most likely due to a smaller number of signal events.

Some improvement for the analysis were discussed, including taking into account correlations and effects of systematic uncertainties in the chosen training variables, additional training on just two most important variables, using class weights to balance different types of events (QE, RES and DIS) in the signal sample, and having two training samples based on the value of the reconstructed neutrino energy. Constructing a similar study using BDTs and the possible improvement compared to neural network was also discussed briefly.

Bibliography

[1] E. Fermi, "An attempt of a theory of beta radiation. I," *Z. Phys*, vol. 88, pp. 161–177, 1934. DOI: `10.1007/BF01351864`.

[2] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, "Detection of the Free Neutrino: a Confirmation," *Science*, vol. 124, pp. 103–104, 1956. DOI: `10.1126/science.124.3212.103`.

[3] G. Danby *et al.*, "Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos," *Phys. Rev. Lett.*, vol. 9, pp. 36–44, 1962. DOI: `10.1103/PhysRevLett.9.36`.

[4] T. Patzak, "First direct observation of the tau neutrino," *Europhys. News*, vol. 32, pp. 56–57, DOI: `10.1051/epn:2001205`.

[5] J. A. Formaggio and G. P. Zeller, "From eV to EeV: Neutrino cross sections across energy scales," *Rev. Mod. Phys.*, vol. 84, pp. 1307–1341, 2012. DOI: `10.1103/RevModPhys.84.1307`.

[6] B. Pontecorvo, "Mesonium and antimesonium," *Sov. Phys. JETP*, vol. 33, pp. 549–551, 1957.

[7] Z. Maki, M. Nakagawa, and S. Sakata, "Remarks on the unified model of elementary particles," *Prog. Theor. Phys*, vol. 28, pp. 870–880, 1962. DOI: `10.1143/PTP.28.870`.

[8] C. Giunti and C. W. Kim, *Fundamentals of Neutrino Physics and Astrophysics.* New York: Oxford University Press, 2007, ISBN: 978-0-19-850871-7.

[9] M. Thomson, *Modern particle physics.* New York: Cambridge University Press, 2013, ISBN: 978-1-107-03426-6. DOI: `10.1017/CBO9781139525367`.

[10] M. Tanabashi *et al.*, "Review of Particle Physics," *Phys. Rev. D*, vol. 98, no. 3, p. 030 001, 2018. DOI: `10.1103/PhysRevD.98.030001`.

[11] C. Giunti, “Quantum mechanics of neutrino oscillations,” in *11th International School on Particles and Cosmology*, Kabardino Balkaria, Russian Federation, May 2001. arXiv: `hep-ph/0105319`.

[12] M. A. Acero *et al.*, “Improved measurement of neutrino oscillation parameters by the NOvA experiment,” *Phys. Rev. D*, vol. 106, no. 3, p. 032 004, 2022. DOI: `10.1103/PhysRevD.106.032004`.

[13] R. Davis, “A review of the Homestake solar neutrino experiment,” *Prog. Part. Nucl. Phys*, vol. 32, pp. 13–32, 1994. DOI: `10.1016/0146-6410(94)90004-3`.

[14] J. Abdurashitov *et al.*, “Results from SAGE (The Russian-American gallium solar neutrino experiment),” *Phys. Lett. B*, vol. 28, pp. 234–248, 1994. DOI: `10.1016/0370-2693(94)90454-5`.

[15] M. Cribier, “Results of the whole GALLEX experiment,” *Nucl. Phys. B Proc. Suppl.*, vol. 70, A. Bottino, A. Di Credico, and P. Monacelli, Eds., pp. 284–291, 1999. DOI: `10.1016/S0920-5632(98)00438-1`.

[16] Y. Suzuki, “Kamiokande solar neutrino results,” *Nucl. Phys. B (Proc. Suppl.)*, vol. 38, no. 1, pp. 54–59, 1995, Neutrino 94. DOI: `10.1016/0920-5632(94)00733-C`.

[17] Y. Fukuda *et al.*, “Evidence for oscillation of atmospheric neutrinos,” *Phys. Rev. Lett.*, vol. 81, pp. 1562–1567, 1998. DOI: `10.1103/PhysRevLett.81.1562`.

[18] Q. R. Ahmad *et al.*, “Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^{8}B solar neutrinos at the Sudbury Neutrino Observatory,” *Phys. Rev. Lett.*, vol. 87, p. 071 301, 2001. DOI: `10.1103/PhysRevLett.87.071301`.

[19] Y. Fukuda *et al.*, “Measurements of the solar neutrino flux from Super-Kamiokande’s first 300 days,” *Phys. Rev. Lett.*, vol. 81, pp. 1158–1162, 1998. DOI: `10.1103/PhysRevLett.81.1158`.

[20] A. Bellerive, J. R. Klein, A. B. McDonald, A. J. Noble, and A. W. P. Poon, “The Sudbury Neutrino Observatory,” *Nucl. Phys. B*, vol. 908, pp. 30–51, 2016. DOI: `10.1016/j.nuclphysb.2016.04.035`.

[21] M. D. Messier, “Review of neutrino oscllations experiments,” *eConf*, vol. C060409, p. 018, 2006. arXiv: `hep-ex/0606013`.

[22] T. N. Prize. “the nobel prize in physics 2015.” (2015), [Online]. Available: `https://www.nobelprize.org/prizes/physics/2015/summary/` (visited on 03/27/2023).

[23] K. S. Hirata *et al.*, “Observation of a small atmospheric muon-neutrino / electron-neutrino ratio in Kamiokande,” *Phys. Lett. B*, vol. 280, pp. 146–152, 1992. DOI: `10.1016/0370-2693(92)90788-6`.

[24] R. Becker-Szendy *et al.*, “Neutrino measurements with the IMB detector,” *Nucl. Phys. B Proc. Suppl.*, vol. 38, pp. 331–336, 1995. DOI: `10.1016/0920-5632(94)00765-N`.

[25] M. Sajjad Athar *et al.*, “Status and perspectives of neutrino physics,” *Prog. Part. Nucl. Phys.*, vol. 124, p. 103 947, 2022. DOI: `10.1016/j.ppnp.2022.103947`.

[26] M. C. Sanchez *et al.*, “Measurement of the L/E distributions of atmospheric neutrinos in Soudan 2 and their interpretation as neutrino oscillations,” *Phys. Rev. D*, vol. 68, p. 113 004, 2003. DOI: `10.1103/PhysRevD.68.113004`.

[27] M. Ambrosio *et al.*, “Measurements of atmospheric muon neutrino oscillations, global analysis of the data collected with MACRO detector,” *Eur. Phys. J. C*, vol. 36, pp. 323–339, 2004. DOI: `10.1140/epjc/s2004-01951-9`.

[28] Y. Fukuda *et al.*, “Study of the atmospheric neutrino flux in the multi-GeV energy range,” *Phys. Lett. B*, vol. 436, pp. 33–41, 1998. DOI: `10.1016/S0370-2693(98)00876-4`.

[29] Y. Fukuda *et al.*, “Measurement of a small atmospheric muon-neutrino electron-neutrino ratio,” *Phys. Lett. B*, vol. 433, pp. 9–18, 1998. DOI: `10.1016/S0370-2693(98)00476-6`.

[30] M. G. Aartsen *et al.*, “Measurement of Atmospheric Neutrino Oscillations at 656 GeV with IceCube DeepCore,” *Phys. Rev. Lett.*, vol. 120, no. 7, p. 071 801, 2018. DOI: `10.1103/PhysRevLett.120.071801`.

[31] M. V. Diwan, V. Galymov, X. Qian, and A. Rubbia, “Long-Baseline Neutrino Experiments,” *Ann. Rev. Nucl. Part. Sci.*, vol. 66, pp. 47–71, 2016. DOI: `10.1146/annurev-nucl-102014-021939`.

[32] B. Abi *et al.*, “Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume II: DUNE Physics,” Tech. Rep., Feb. 2020. DOI: `10.48550/arXiv.2002.03005`.

[33] B. Jamieson, “Future Neutrino Experiments,” in *20th Conference on Flavor Physics and CP Violation*, Oxford, Mississippi, USA, Jul. 2022. DOI: `10.48550/arXiv.2207.05044`.

[34] A. Abusleme *et al.*, “Sub-percent precision measurement of neutrino oscillation parameters with JUNO,” *Chin. Phys. C*, vol. 46, no. 12, p. 123 001, 2022. DOI: `10.1088/1674-1137/ac8bc9`.

[35] C. Bemporad, G. Gratta, and P. Vogel, “Reactor Based Neutrino Oscillation Experiments,” *Rev. Mod. Phys.*, vol. 74, p. 297, 2002. DOI: `10.1103/RevModPhys.74.297`.

[36] J. K. Ahn *et al.*, “RENO: An Experiment for Neutrino Oscillation Parameter θ_{13} Using Reactor Neutrinos at Yonggwang,” Tech. Rep., Mar. 2010. DOI: `10.48550/arXiv.1003.1391`.

[37] Y. Abe *et al.*, “Improved measurements of the neutrino mixing angle θ_{13} with the Double Chooz detector,” *JHEP*, vol. 10, p. 086, 2014. DOI: `10.1007/JHEP02(2015)074`.

[38] F. P. An *et al.*, “A side-by-side comparison of Daya Bay antineutrino detectors,” *Nucl. Instrum. Meth. A*, vol. 685, pp. 78–97, 2012. DOI: `10.1016/j.nima.2012.05.030`.

[39] T. Mitsui, “Kamland experiment,” *Nucl, Phys. B - Proc. Supp.*, vol. 117, pp. 13–17, 2003. DOI: `10.1016/S0920-5632(03)90478-6`.

[40] T. M. Lackey, “Proton Scattering in NOvA Test Beam,” Ph.D. dissertation, Indiana University, Jul. 2022.

[41] V. Barger, D. Marfatia, and K. Whisnant, *The Physics of Neutrinos.* Princeton, USA: Princeton University Press, 2012, ISBN: 978-0-691-12853-5.

[42] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, “Evidence for the 2π Decay of the K_2^0 Meson,” *Phys. Rev. Lett.*, vol. 13, pp. 138–140, 1964. DOI: `10.1103/PhysRevLett.13.138`.

[43] G. C. Branco, R. G. Felipe, and F. R. Joaquim, “Leptonic CP violation,” *Rev. Mod. Phys.*, vol. 84, pp. 515–565, 2012. DOI: `10.1103/RevModPhys.84.515`.

[44] N. Cosme, “Leptogenesis, neutrino masses and gauge unification,” *JHEP*, vol. 08, p. 027, 2004. DOI: `10.1088/1126-6708/2004/08/027`.

[45] W. Buchmuller, R. D. Peccei, and T. Yanagida, “Leptogenesis as the origin of matter,” *Ann. Rev. Nucl. Part. Sci.*, vol. 55, pp. 311–355, 2005. DOI: `10.1146/annurev.nucl.55.090704.151558`.

[46] A. Sakharov, “Violation of cp invariance, c asymmetry, and baryon asymmetry of the universe,” *Soviet Physics Uspekhi*, vol. 34, no. 5, p. 392, May 1991. DOI: `10.1070/PU1991v034n05ABEH002497`.

[47] S. Fukuda *et al.*, "Determination of solar neutrino oscillation parameters using 1496 days of Super-Kamiokande I data," *Phys. Lett. B*, vol. 539, pp. 179–187, 2002. DOI: `10.1016/S0370-2693(02)02090-7`.

[48] X. Qian and P. Vogel, "Neutrino Mass Hierarchy," *Prog. Part. Nucl. Phys.*, vol. 83, pp. 1–30, 2015. DOI: `10.1016/j.ppnp.2015.05.002`.

[49] A. Ghosh, T. Thakore, and S. Choubey, "Determining the Neutrino Mass Hierarchy with INO, T2K, NOvA and Reactor Experiments," *JHEP*, vol. 04, p. 009, 2013. DOI: `10.1007/JHEP04(2013)009`.

[50] E. K. Akhmedov, G. C. Branco, and M. N. Rebelo, "Seesaw mechanism and structure of neutrino mass matrix," *Phys. Lett. B*, vol. 478, pp. 215–223, 2000. DOI: `10.1016/S0370-2693(00)00282-3`.

[51] L. Wolfenstein, "Neutrino Oscillations in Matter," *Phys. Rev. D*, vol. 17, pp. 2369–2374, 1978. DOI: `10.1103/PhysRevD.17.2369`.

[52] A. C. Booth, "Electron neutrino appearance at the NOvA experiment," Ph.D. dissertation, University of Sussex, Oct. 2021.

[53] E. Catano-Mur, "Dpf2017 - physics reach of electron neutrino measurements," NOvA internal note 21764, Tech. Rep., Aug. 2017.

[54] D. S. Ayres *et al.*, "The NOvA Technical Design Report," Tech. Rep., Oct. 2007. DOI: `10.2172/935497`.

[55] L. H. Whitehead, "Neutrino Oscillations with MINOS and MINOS+," *Nucl. Phys. B*, vol. 908, pp. 130–150, 2016. DOI: `10.1016/j.nuclphysb.2016.03.004`.

[56] S. Bash, "The Chips Prototype Water Cherenkov Detector," Ph.D. dissertation, University College London, Feb. 2023.

[57] P. Adamson *et al.*, "The NuMI Neutrino Beam," *Nucl. Instrum. Meth. A*, vol. 806, pp. 279–306, 2016. DOI: `10.1016/j.nima.2015.08.063`.

[58] S. E. Kopp, "The NuMI neutrino beam and potential for an off axis experiment," *J. Phys. G*, vol. 29, K. Long and R. Edgecock, Eds., pp. 1757–1762, 2003. DOI: `10.1088/0954-3899/29/8/343`.

[59] L. Aliaga, "2017-2018 beam plots," NOvA internal note 20843, Tech. Rep., Jul. 2018.

[60] L. Kolupaeva, "Recent three-flavor neutrino oscillation results from the NOvA experiment," *J. Phys. Conf. Ser.*, vol. 1690, no. 1, P. Teterin, Ed., p. 012 172, 2020. DOI: `10.1088/1742-6596/1690/1/012172`.

[61] R. B. Patterson, "The NOvA Experiment: Status and Outlook," *Nucl. Phys. B Proc. Suppl.*, vol. 235-236, pp. 151–157, 2013. DOI: `10.1016/j.nuclphysbps.2013.04.005`.

[62] S. Bending, "Measurement of 3-Flavour Neutrino Oscillation Parameters in the NOvA Experiment," Ph.D. dissertation, University College London, Oct. 2020.

[63] T. G. Olson, "Measurement of $d^2\sigma/d|\vec{q}|dE_{avail}$ and 2p2h contribution using charged current ν_μ interactions in the NOvA Near Detector," Ph.D. dissertation, Tufts University, Sep. 2021.

[64] B. Tapia Oregui, "Towards a measurement of neutral pion production in neutral current interactions with the NOνA near detector," Ph.D. dissertation, University of Texas at Austin, Aug. 2022.

[65] A. Norman, "The NOvA Data Acquisition System," *J. Phys. Conf. Ser.*, vol. 396, p. 012 035, 2012. DOI: `10.1088/1742-6596/396/1/012035`.

[66] ""fermilab creative services"." (2023), [Online]. Available: `https://vms.fnal.gov/index/vms-home` (visited on 01/12/2023).

[67] Y. Zhang, "Measurement of neutrino oscillations using neutrino and antineutrino beams in the NOvA experiment," Ph.D. dissertation, University of Sussex, Mar. 2023.

[68] G. Davies, "Images of the far detector," NOvA internal note 20792, Tech. Rep., Jun. 2017.

[69] A. T. C. Sutton, "Domain Generalization with Machine Learning in the NOvA Experiment," Ph.D. dissertation, University of Virginia, Feb. 2022.

[70] G. Davies, "Reconstruction event displays," NOvA internal note 21152, Tech. Rep., Jun. 2017.

[71] F. Psihar, "Event displays for nue selected events," NOvA internal note 15647, Tech. Rep., Aug. 2016.

[72] A. Aurisano *et al.*, "A Convolutional Neural Network Neutrino Event Classifier," *JINST*, vol. 11, no. 09, P09001, 2016. DOI: `10.1088/1748-0221/11/09/P09001`.

[73] H. Nunokawa, S. J. Parke, and J. W. F. Valle, "CP Violation and Neutrino Oscillations," *Prog. Part. Nucl. Phys.*, vol. 60, pp. 338–402, 2008. DOI: `10.1016/j.ppnp.2007.10.001`.

[74] M. Bair, L. Suter, and J. Wolcott, "Summary of the 2020 FHC+RHC 3-Flavor oscillation analysis," NOvA internal note 44422, Tech. Rep., Apr. 2020.

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

[76] L. Aliaga *et al.*, "Neutrino Flux Predictions for the NuMI Beam," *Phys. Rev. D*, vol. 94, no. 9, p. 092 005, 2016. DOI: `10.1103/PhysRevD.94.092005`.

[77] C. Andreopoulos *et al.*, "The GENIE Neutrino Monte Carlo Generator," *Nucl. Instrum. Meth. A*, vol. 614, pp. 87–104, 2010. DOI: `10.1016/j.nima.2009.12.009`.

[78] S. Yu, "Blessing package: Fuzzykprong view matching with a nue qe event," NOvA internal note 28490, Tech. Rep., May 2018.

[79] P. V. Hough, "Method and means for recognizing complex patterns," U.S. Patent 3 069 654 A, Dec. 18, 1962.

[80] M. Ohlsson, "Extensions and explorations of the elastic arms algorithm," *Comput. Phys. Commun.*, vol. 77, pp. 19–32, 1993. DOI: `10.1016/0010-4655(93)90033-9`.

[81] B. Behera, G. Davies, and F. Psihas, "Event Reconstruction in the NOvA Experiment," in *Meeting of the APS Division of Particles and Fields*, Batavia, Illinois, USA, Oct. 2017. DOI: `10.48550/arXiv.1710.03772`.

[82] R. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng*, vol. 82, pp. 35–45, 1960. DOI: `10.1115/1.3662552`.

[83] C. Backhouse, A. Radovic, P. Singh, and M. M. Campbell, "The attenuation and threshold calibration of the nova detectors," NOvA internal note 13579, Tech. Rep., Aug. 2017.

[84] B. R. Martin, *Nuclear and Particle Physics*. Chichester: John Wiley and Sons, 2006, ISBN: 9780470019993.

[85] L. Vinton, "Far detector stopping muon energy scale calibration," NOvA internal note 13475, Tech. Rep., May 2015.

[86] L. Vinton, "Calorimetric energy scale calibration of the nova detectors," NOvA internal note 13579, Tech. Rep., Jul. 2015.

[87] V. V. Gligorov and M. Williams, “Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree,” *JINST*, vol. 8, P02013, 2013. DOI: `10.1088/1748-0221/8/02/P02013`.

[88] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, “Boosted decision trees, an alternative to artificial neural networks,” *Nucl. Instrum. Meth. A*, vol. 543, no. 2-3, pp. 577–584, 2005. DOI: `10.1016/j.nima.2004.12.018`.

[89] N. L. Raddatz, “Reconstructed muon identification,” NOvA internal note 11206, Tech. Rep., May 2014.

[90] E. Catano-Mur, N. Nayak, A. Sutton, and K. Warlburton, “Event selection for the 2020 3-flavor analysis,” NOvA internal note 44040, Tech. Rep., Mar. 2020.

[91] F. Psihas, “3flavor 2020 cut flow chart,” NOvA internal note 44349, Tech. Rep., May 2015.

[92] T. Warburton, “3-flavour cut flow tables,” NOvA internal note 45880, Tech. Rep., Aug. 2020.

[93] T. Warburton, “Ana 20 numu quantile plots,” NOvA internal note 47242, Tech. Rep., Sep. 2020.

[94] A. T. C. Sutton, “Ana2020 numu reconstructed energy blessing package,” NOvA internal note 45809, Tech. Rep., May 2020.

[95] A. M. Hall, “Ana2020 far detector mc breakdown blessing package,” NOvA internal note 45892, Tech. Rep., Jul. 2020.

[96] S. Yu and Z. Djurcic, “Bendecomp ana2018 technote,” NOvA internal note 27833, Tech. Rep., Apr. 2018.

[97] D. Pershey, “Technote for micheldecomp2017,” NOvA internal note 22523, Tech. Rep., Sep. 2017.

[98] S. You, “Blessing package: Extrapolation cartoon for ana2018,” NOvA internal note 29612, Tech. Rep., Jun. 2018.

[99] A. Mislivec, “Pt extrapolation blessed plots for the 2020 analysis,” NOvA internal note 45980, Tech. Rep., Jun. 2020.

[100] R. B. Patterson and M. J. Wetstein, “2020 neutron systematic tech note,” NOvA internal note 43811, Tech. Rep., Feb. 2020.

[101] S. Baker and R. D. Cousins, “Clarification of the Use of Chi Square and Likelihood Functions in Fits to Histograms,” *Nucl. Instrum. Meth.*, vol. 221, pp. 437–442, 1984. DOI: `10.1016/0167-5087(84)90016-4`.

[102] G. Cowan, *Statistical data analysis.* New York: Oxford University Press, 1998, ISBN: 978-0-19-850156-5.

[103] F. Gary and C. Robert, "Unified approach to the classical statistical analysis of small signals," *Phys. Rev. D*, vol. 57, pp. 3873–3889, 1998. DOI: `10.1103/PhysRevD.57.3873`.

[104] M. A. Acero *et al.*, "The Profiled Feldman-Cousins technique for confidence interval construction in the presence of nuisance parameters," Jul. 2022. arXiv: `2207.14353`.

[105] P. B. Denton, J. Gehrlein, and R. Pestes, "CP-Violating Neutrino Nonstandard Interactions in Long-Baseline-Accelerator Data," *Phys. Rev. Lett.*, vol. 126, no. 5, p. 051 801, 2021. DOI: `10.1103/PhysRevLett.126.051801`.

[106] S. S. Chatterjee and A. Palazzo, "Nonstandard neutrino interactions as a solution to the NOνA and t2k discrepancy," *Phys. Rev. Lett.*, vol. 126, p. 051 802, 5 2021. DOI: `10.1103/PhysRevLett.126.051802`.

[107] A. Back, "Blessing Package for All-time FD POT Exposure - updated for May 2022," NOvA internal note 54987, Tech. Rep., May 2022.

[108] H. Kolanoski, "Application of artificial neural networks in particle physics," *Nucl. Instrum. Meth. A*, vol. 367, W. Bartl, M. Krammer, G. Neuhofer, M. Regler, and A. Taurok, Eds., pp. 14–20, 1995. DOI: `10.1016/0168-9002(95)00743-1`.

[109] L. Teodorescu, "Artificial neural networks in high-energy physics," *Inverted CERN School of Computing : 2005 and 2006 edition*, pp. 13–22, DOI: `10.5170/CERN-2008-002.13`.

[110] A. Abraham, "Artificial neural networks," in *Handbook of Measuring System Design.* John Wiley and Sons, Ltd, 2005, ISBN: 0-470-02143-8. DOI: `10.1002/0471497398.mm421`.

[111] M. D. Schwartz, "Modern Machine Learning and Particle Physics," *Harvard Data Science Review*, 2021. DOI: `10.1162/99608f92.beeb1183`.

[112] M. Abadi *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015. [Online]. Available: `http://tensorflow.org/`.

[113] V. W. Berger and Y. Zhou, “Kolmogorov-Smirnov Test: Overview,” in *Wiley StatsRef: Statistics Reference Online*. John Wiley and Sons, Ltd, 2014, ISBN: 9781118445112. DOI: `10.1002/9781118445112.stat06558`.

[114] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs),” in *ICLR 2016*, San Juan, Puerto Rico, USA, 2016. DOI: `10.48550/arXiv.1511.07289`.

[115] M. Wang, S. Lu, D. Zhu, J. Lin, and Z. Wang, “A high-speed and low-complexity architecture for softmax function in deep learning,” in *2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, Chengdu, China, 2018, pp. 223–226. DOI: `10.1109/APCCAS.2018.8605654`.

[116] “"tf.keras.activations.softmax".” (2023), [Online]. Available: `https://www.tensorflow.org/api_docs/python/tf/keras/activations/softmax` (visited on 05/20/2023).

[117] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *the 3rd International Conference for Learning Representations*, San Diego, California, USA, 2015. arXiv: `1412.6980`.

[118] Z. Zhang, “Improved adam optimizer for deep neural networks,” in *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, 2018, pp. 1–2. DOI: `10.1109/IWQoS.2018.8624183`.

[119] L. Li, M. Doroslovaki, and M. Loew, “Approximating the gradient of cross-entropy loss function,” *IEEE Access*, vol. 8, pp. 111 626–111 635, 2020. DOI: `10.1109/ACCESS.2020.3001531`.

[120] “"tf.keras.metrics.accuracy".” (2023), [Online]. Available: `https://www.tensorflow.org/api_docs/python/tf/keras/metrics/Accuracy` (visited on 05/20/2023).

[121] “"sklearn.preprocessing.standardscaler".” (2023), [Online]. Available: `https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html` (visited on 12/07/2023).

[122] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

Appendix A

Variables Not Used in Training

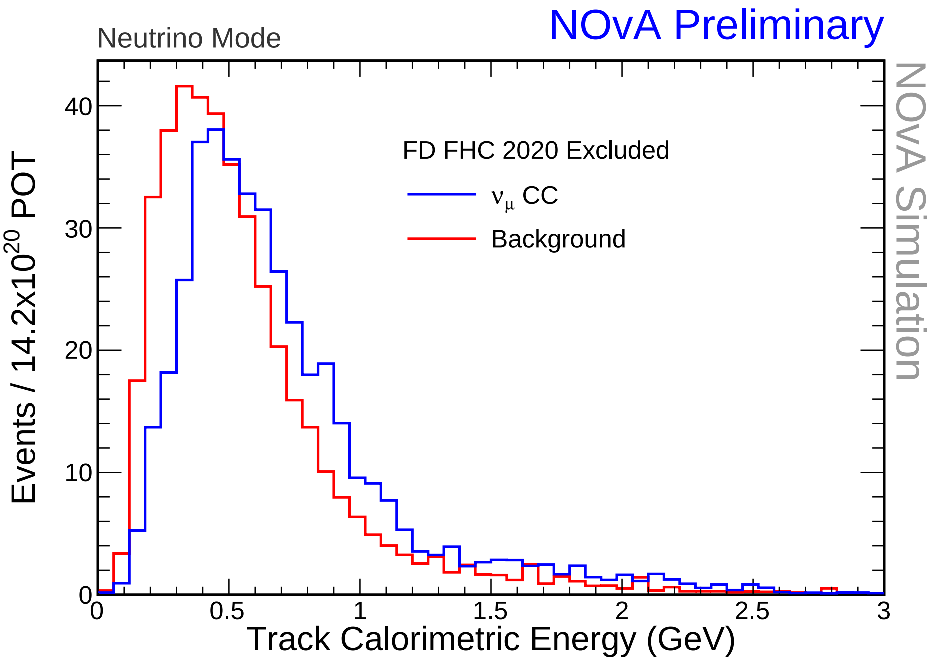

Figure A.1: FD FHC simulation track calorimetric energy for events rejected from disappearance analysis.

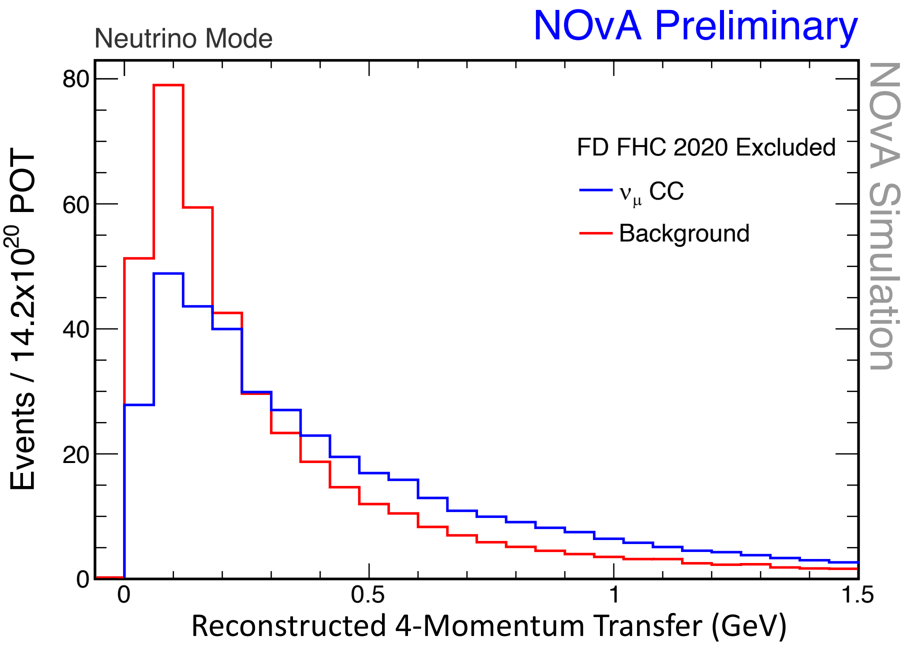

Figure A.2: FD FHC simulation reconstructed 4-momentum transfer for events rejected from 2020 disappearance analysis.

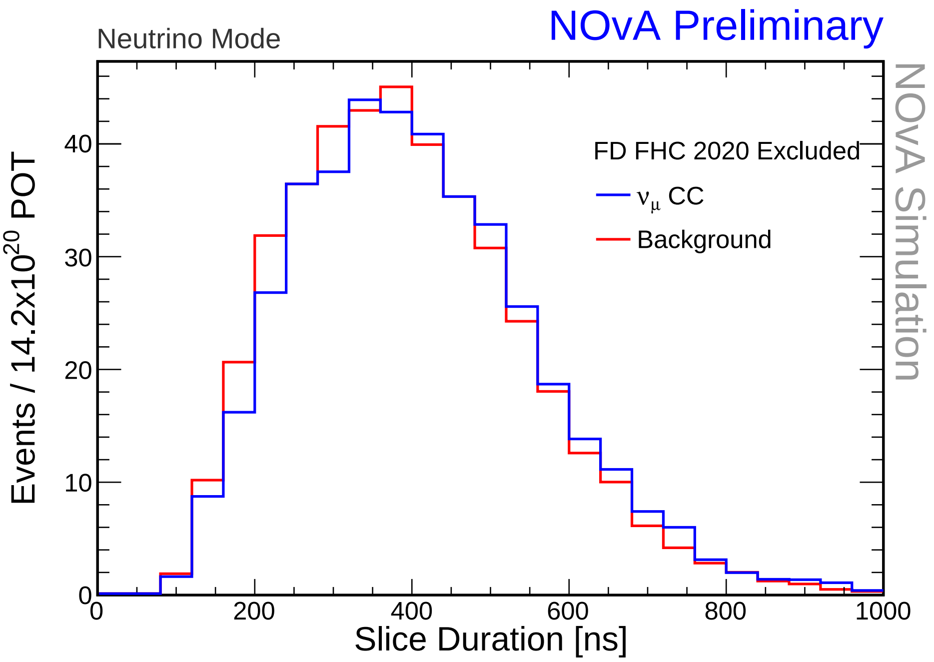

Figure A.3: FD FHC simulation slice duration for events rejected from 2020 disappearance analysis.

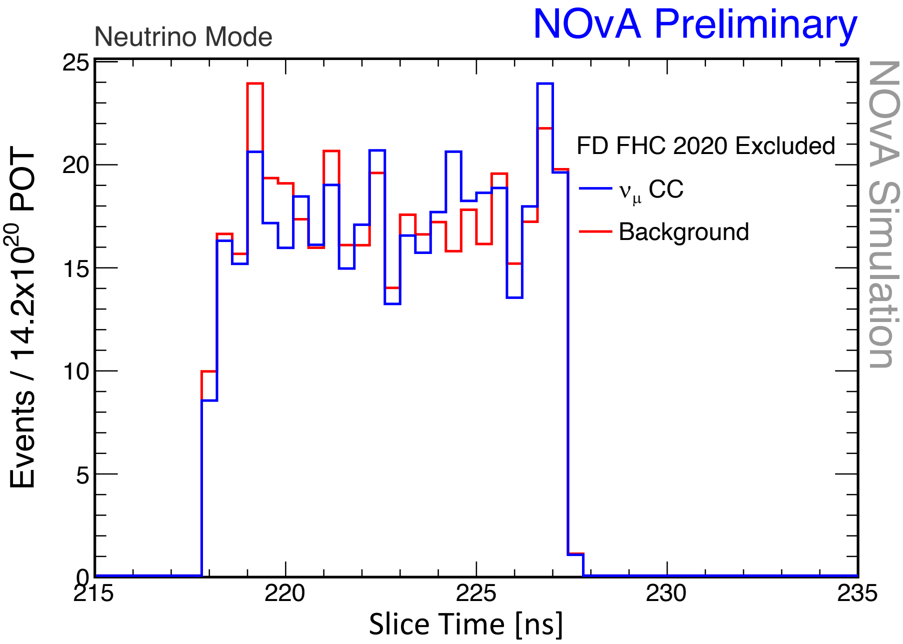

Figure A.4: FD FHC simulation slice time for events not rejected from 2020 disappearance analysis.

Appendix B

ND MC-Data Comparison Plots for Training Variables

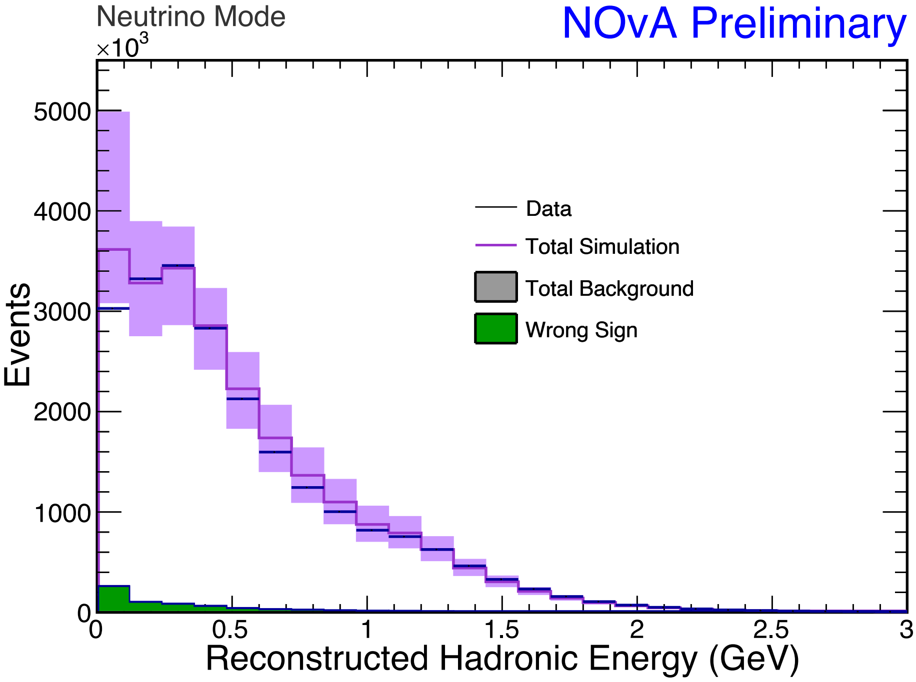

Figure B.1: MC-data comparison for ND FHC data, reconstructed hadronic energy.

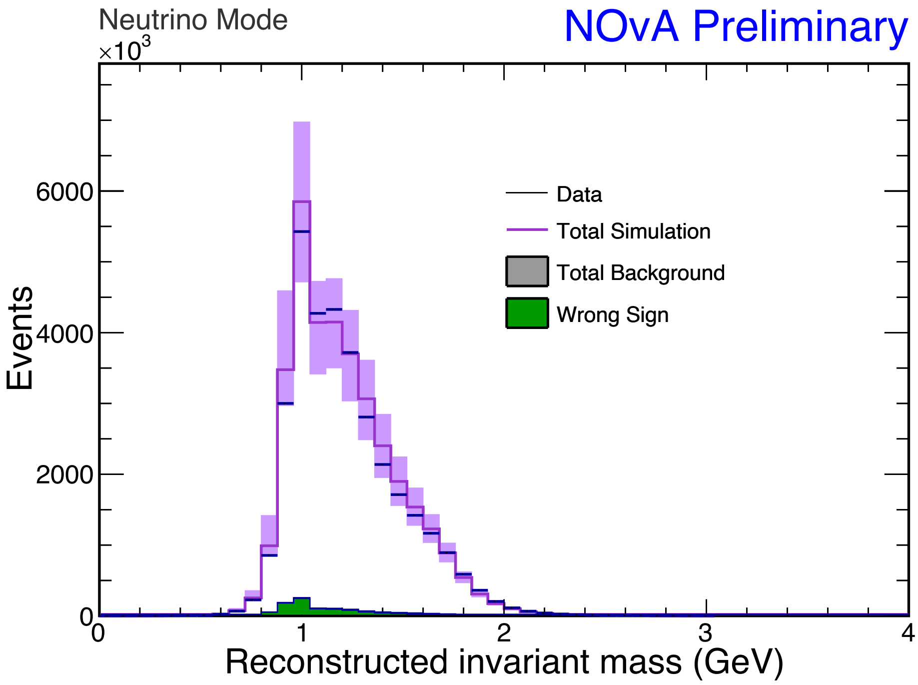

Figure B.2: MC-data comparison for ND FHC data, reconstructed invariant mass.

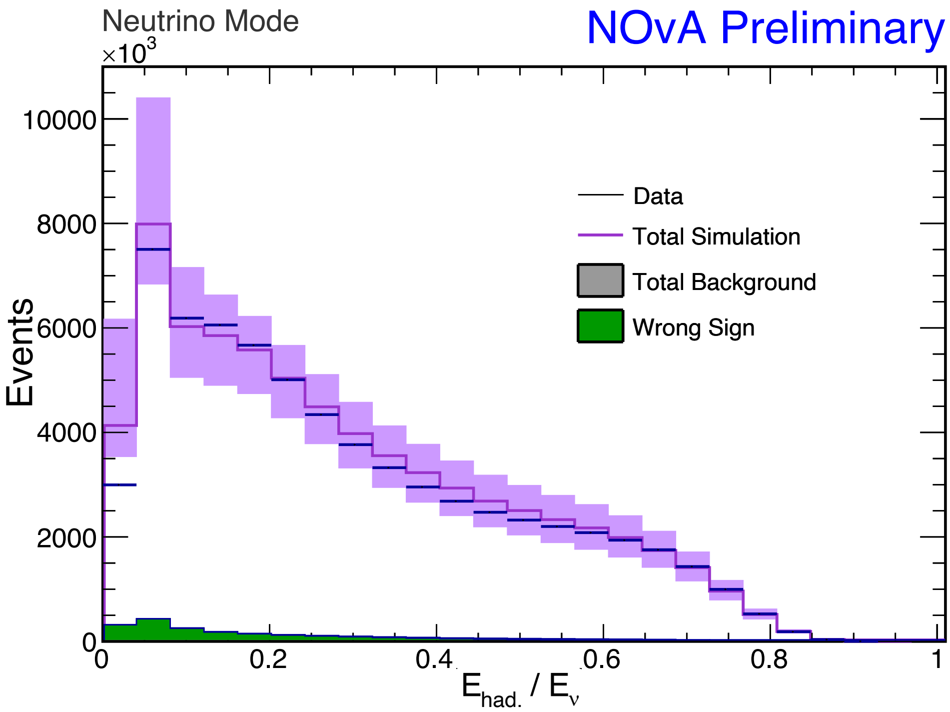

Figure B.3: MC-data comparison for ND FHC data, hadronic energy fraction.

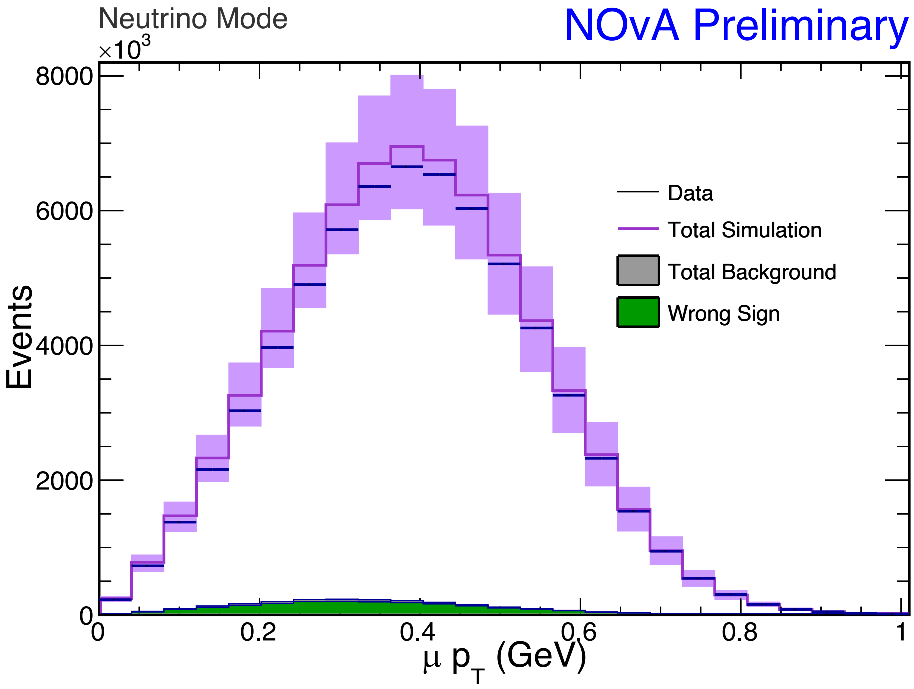

Figure B.4: MC-data comparison for ND FHC data, reconstructed muon transverse momentum.

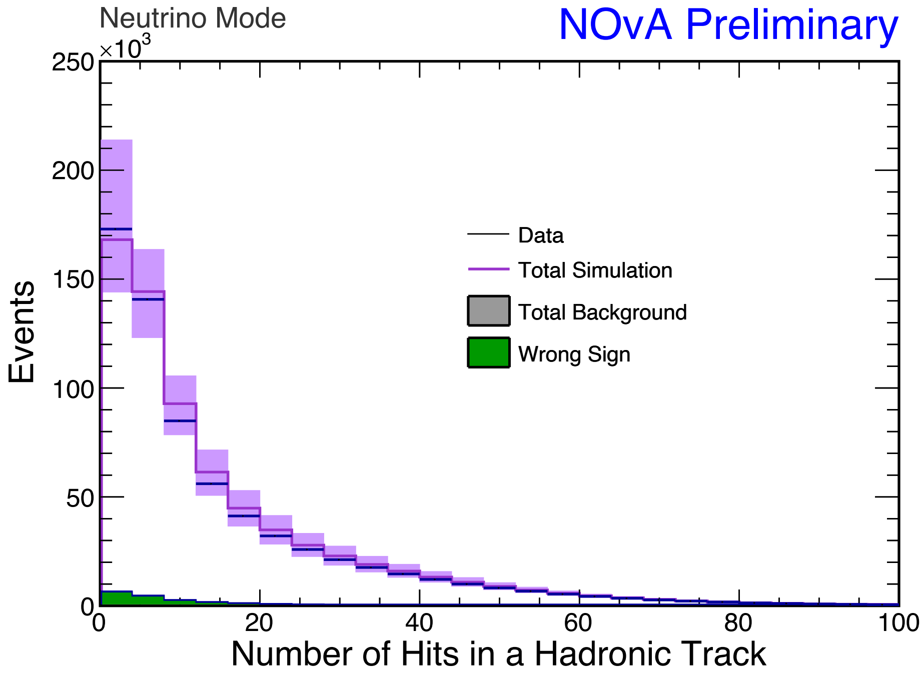

Figure B.5: MC-data comparison for ND FHC data, number of hits in the hadronic kalman Track.

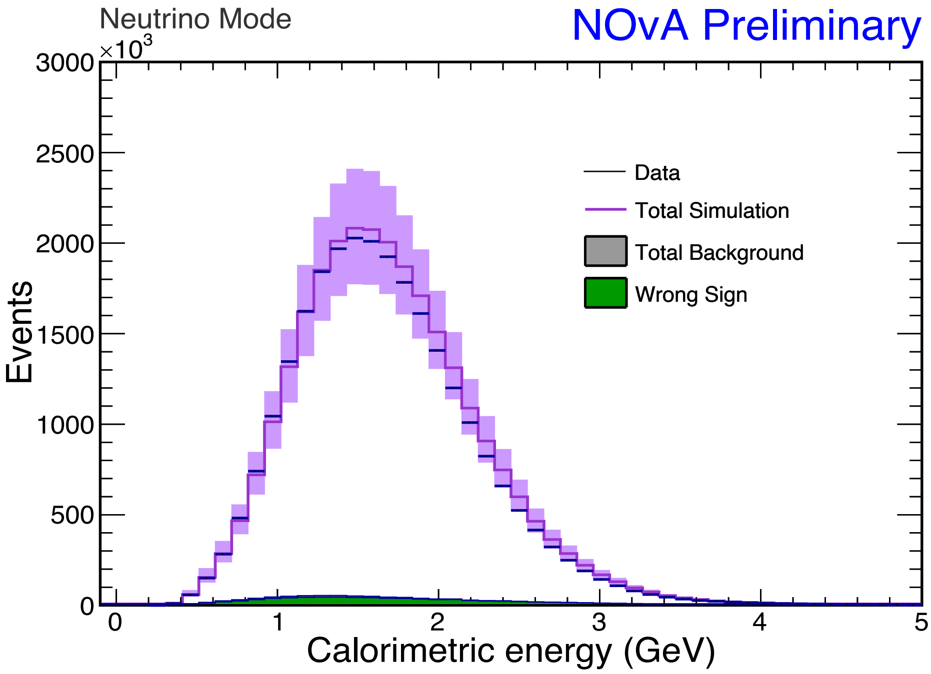

Figure B.6: MC-data comparison for ND FHC data, calorimetric energy.

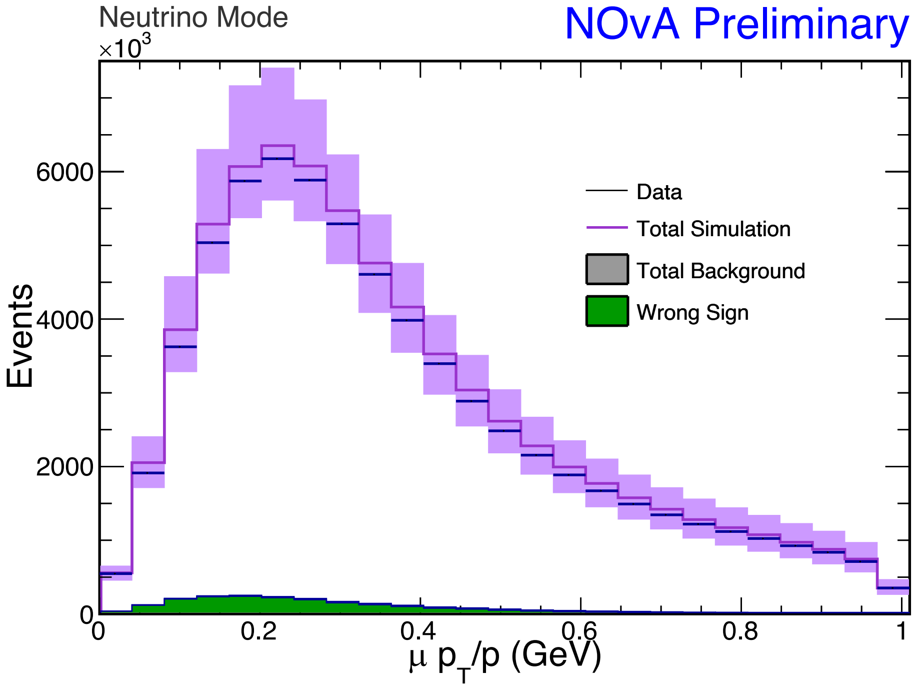

Figure B.7: MC-data comparison for ND FHC data, reconstructed transverse muon momentum divided by the total momentum.