

**Muon Neutrino Disappearance in NOvA with a Deep
Convolutional Neural Network Classifier**

**A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY**

Dominick Rosario Rocco

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy**

Dr. Gregory Pawloski

March, 2016

© Dominick Rosario Rocco 2016
ALL RIGHTS RESERVED

Acknowledgements

This work would not have been possible without my family. My parents, Dorothy and Bill, have always been a great support. They raised me to adulthood and ensured I could always pursue my dreams. My brother, Joseph, is my oldest friend. The lives we live today are an imprint of the experiences we have shared. I owe the utmost gratitude to Michelle, who has been my closest ally throughout recent years. There is no way I could have completed this work without her support, both material and emotional.

I will always remember the other graduate students who started this journey with me; in particular Terry Bretz-Sullivan, Barry Costanzi, Mark Pepin, Bern Youngblood, Tanner Prestergard, Michael Janas, and Tobias Gulden. Working through coursework with them was both enlightening and enjoyable. I have also enjoyed the friendship of Alexander Gude and Charles McEachern, who happened to start on the road before me, but were there all the same.

My fellow graduate students in the Minnesota neutrino group – Susan Lein, Nick Raddatz, Kanika Sachdev, and Jan Zirnstein – were a constant resource. The openness we encouraged created an environment for us all to succeed; the conversations we shared were instrumental to my development; the laughter we enjoyed was the ultimate diversion. I cannot imagine a better group with whom to have taken this journey.

This opportunity would not have existed without the Minnesota neutrino group as a whole, especially my advisor Gregory Pawloski. I would also like to thank Ken Heller for bringing me on in the first place, and Marvin Marshak for making the experience relatable. Navigating through this environment would have been considerably more difficult without the shrewd lessons and caring guidance of Jody Kaplan. Toward the end of my time as a student, I learned learned a great deal from conversations from Justin Hietala, who became a cherished friend despite the short time we worked together.

Of course, none of this work would have been possible without the NOvA collaboration; especially the leadership and vision of Gary Feldman, Mark Messier, and Ryan Patterson. I would like to thank Christopher Backhouse and Gavin Davies for the vast knowledge which they shared freely. I would also like to recognize the graduate students from other institutions with whom I have learned, particularly Evan Niner, Michael Baird, and Fernanda Psihas.

Finally, I owe the many thanks to Alexander Radovic and Adam Aurisano for breathing life into this project at just the right moment. The success demonstrated in this dissertation would not have been achieved without their effort.

Dedication

To the memory of my grandfather, Dominick A. Rocco.

His wisdom was a guiding light through my most formative years. The values he instilled were priceless; notably patience, determination, strength, and compassion. The caring he demonstrated was admirable; not just to those close to him, but to all people. The friendship he offered was fulfilling; hearts, minds, and souls were thoroughly enriched by his love, wisdom, and laughter. Along with the many others he touched, I would not be the person I am without his influence. In loving memory, I dedicate this work to him.

Abstract

The NuMI Off-axis Neutrino Appearance Experiment (NOvA) is designed to study neutrino oscillation in the NuMI (Neutrinos at the Main Injector) beam. NOvA observes neutrino oscillation using two detectors separated by a baseline of 810 km; a 14 kt Far Detector in Ash River, MN and a functionally identical 0.3 kt Near Detector at Fermilab. The experiment aims to provide new measurements of Δm_{32}^2 and θ_{23} and has potential to determine the neutrino mass hierarchy as well as observe CP violation in the neutrino sector. Essential to these analyses is the classification of neutrino interaction events in NOvA detectors. Raw detector output from NOvA is interpretable as a pair of images which provide orthogonal views of particle interactions. A recent advance in the field of computer vision is the advent of convolutional neural networks, which have delivered top results in the latest image recognition contests. This work presents a novel approach particle physics analysis in which a convolutional neural network is used for classification of particle interactions. The approach has been demonstrated to improve the signal efficiency and purity of the event selection, and thus physics sensitivity. Early NOvA data has been analyzed (2.74×10^{20} POT, 14 kt equivalent) to provide new best-fit measurements of $\sin^2(\theta_{23}) = 0.43$ (with a statistically-degenerate compliment near 0.60) and $|\Delta m_{32}^2| = 2.48 \times 10^{-3}$ eV².

Contents

Acknowledgements	i
Dedication	iii
Abstract	iv
List of Tables	v
List of Figures	vi
1 Introduction	1
1.1 The Standard Model	2
1.2 History of Neutrino Physics	4
1.2.1 The Neutrino Hypothesis	4
1.2.2 Neutrino Discovery – Methods and Observations	5
1.2.3 The Solar Neutrino Problem	9
2 Neutrino Physics	11
2.1 Neutrino Interactions	11
2.2 Neutrino Oscillation	13
2.2.1 A Straightforward Interpretation	13
2.2.2 Oscillation in Vacuum	14
2.2.3 Oscillation in Matter	16
2.2.4 Recent Experiments and Results	18

3 NOvA	20
3.1 NuMI and Off-axis Alignment	21
3.2 NOvA Detectors	24
3.2.1 Cell Structure	25
3.2.2 Light Collection	25
3.2.3 Data Acquisition	26
3.2.4 Event Topologies in NOvA	27
3.2.5 The NOvA Event Display	29
3.2.6 Exposure	30
4 Monte Carlo Simulation	31
4.1 Beam Simulation	31
4.2 GENIE– Neutrino Interaction Simulation	32
4.3 Geant4– Interaction Product Propagation	35
4.4 Photon Collection and Attenuation	37
4.5 Electronic Readout	38
4.6 Oscillated Spectra	40
5 Event Reconstruction	41
5.1 Slicing	42
5.2 Tracking	44
5.2.1 CosmicTrack – Least-squares Regression	44
5.2.2 KalmanTrack – Kalman Filtering	45
5.3 Calibration	46
5.3.1 Attenuation and Cell-to-cell Correction Procedure	47
5.3.2 Absolute Energy Calibration Procedure	49
5.3.3 Calibrated Energy Depositions for Hits	49
5.4 Muon Identification	52
5.5 Energy Estimation	52
6 Neural Networks	56
6.1 Artificial Neural Networks	56
6.2 Supervised Learning – Backpropagation	60

6.3	Deep Learning	63
6.3.1	Convolution Layers	64
6.3.2	Pooling	64
6.3.3	Network-in-Network and GooLeNet	65
6.3.4	Regularization	67
7	Convolutional Neural Network Event Classifier	69
7.1	NOvA Events as Images	70
7.2	Architecture	75
7.3	Regularization	77
7.4	Training	78
7.5	Training Results	81
7.6	Filter Visualization	81
8	Event Selection	90
8.1	Cosmic-Ray Preselection	91
8.2	Reconstruction Quality	93
8.3	Muon Track Designation	93
8.4	Containment	93
8.4.1	Far Detector Containment	93
8.4.2	Near Detector Containment	94
8.5	Background Rejection	94
8.6	FD Selected Sample	95
8.7	ND Selected Sample	106
9	Analysis	114
9.1	Prediction and Extrapolation	114
9.1.1	ν_μ CC Signal Extrapolation	116
9.1.2	Background Prediction	117
9.1.3	Detector Configurations	118
9.2	Fitting	118
9.2.1	Maximum-likelihood Fitting	118
9.2.2	The Feldman-Cousins Procedure	122

10 Systematic Uncertainties	124
10.1 Treatment of Systematic Uncertainties	124
10.1.1 Relative vs. Absolute Uncertainties	125
10.2 Flux Uncertainty	125
10.3 Cross Section Uncertainty	131
10.4 Particle Propagation Uncertainty	146
10.5 Scintillation Production Uncertainty	153
10.6 Calorimetric Energy Uncertainty	160
10.7 Detector Mass Uncertainty	170
10.8 Muon Range Uncertainty	173
10.9 Hadronic Energy Uncertainty	176
10.10 Negligible Uncertainties	179
10.11 Total Systematic Uncertainty	179
11 Results	181
11.1 FD Selected Sample	181
11.2 Selected Event Displays	198
11.3 Fit Results	205
11.4 Feldman-Cousins Confidence Interval	208
12 Conclusion	210
References	213
Appendix A. Glossary and Acronyms	224
A.1 Glossary	224
A.2 Acronyms	225

List of Tables

4.1	Oscillation parameters used to weight MC prediction	40
7.1	Class distribution for training dataset	80
10.2	All reweightable free parameters in GENIE	134
10.3	Redundant knobs in GENIE Reweighting	135
10.4	Shifts induced by alternative Geant4 physics lists	146
10.5	Shifts induced by alternative scintillation production properties	153
10.6	Shifts induced in alternative calibration samples	160
10.7	Uncertainties in density of NOvA cells	170
A.1	Acronyms	225

List of Figures

1.1	Particles in the Standard Model	3
1.2	Oscilloscope trace from Cowan and Reines' experiment	6
2.1	Neutrino interaction diagrams	12
2.2	Interactions which permit coherent scattering	17
3.1	A schematic of the NuMI neutrino source	22
3.2	Effects of off-axis design	23
3.3	Drawings of the NOvA detectors	24
3.4	Event topologies in NOvA	28
3.5	An example NOvA event display	29
4.1	Simulated spectrum of observed FD events	33
5.1	An example NOvA event display	42
5.2	Depiction of clustering with the DBSCAN algorithm	43
5.3	Illustration of the tracking concept in NOvA	45
5.4	Depiction of a tri-cell hit used for calibration	48
5.6	2D histogram depicting deposited energy per path length (dE/dX) . . .	51
5.7	Muon energy fit	53
5.8	Hadronic energy fit	54
6.1	Basic rendering of a biological neuron	57
6.2	Visualization of a simple neural network	59
6.3	Conceptual diagram of a convolutional neural network	65
6.4	Inception module architecture	66
6.5	Complete GooLeNet architecture	68
7.1	Image formed from a ν_μ CC interaction	71
7.2	Image formed from a ν_e CC interaction	72

7.3	Image formed from an NC interaction	73
7.4	Comparison of continuous hit intensity scale and discretized scale	74
7.5	Diagram of the convolutional neural network architecture used for event classification	76
7.6	Histogram depicting class distribution for training dataset	78
7.7	Histogram of the ν_e classifier output	82
7.8	Histogram of the ν_μ classifier output	83
7.9	7×7 Convolutional Filters	85
7.10	7×7 convolutional filter outputs for a ν_μ CC event	86
7.11	7×7 convolutional filter outputs for a ν_e CC event	87
7.12	7×7 convolutional filter outputs for an NC event	88
7.13	Inception module filter output	89
8.1	Comparison of track topologies for cosmic-ray and neutrino events . . .	91
8.2	Illustration of track distance projection	92
8.3	Distribution of softmax output from convolutional neural network for signal and backgrounds	95
8.4	Distribution of CosmicTrack start Y position for signal and backgrounds	96
8.5	Distribution of transverse momentum estimated from slice mean position for signal and backgrounds	97
8.6	Distribution of transverse momentum estimated from reconstructed prongs for signal and backgrounds	98
8.7	Final selected spectrum for ν_μ CC signal and backgrounds	100
8.8	Selected ν_μ CC signal event spectrum and efficiency	101
8.9	Selected ν background event spectrum and efficiency	102
8.10	Comparison of ν_μ CC signal event spectrum and efficiency to existing approach	103
8.11	Comparison of ν background event spectrum and efficiency to existing approach	104
8.12	Sensitivity comparison between existing analysis approach and CNN technique	105
8.13	Slice N_{Hit} distribution for selected ND sample	107
8.14	Muon ID distribution for selected ND sample	108

8.15	Muon track length distribution for selected ND sample	109
8.16	CNN classifier output distribution for selected ND sample	110
8.17	Hadronic cluster N_{Hit} distribution for selected ND sample	111
8.18	Hadronic energy distribution for selected ND sample	112
8.19	Reconstructed energy distribution for selected ND sample	113
9.1	MC Prediction for ND and FD	115
9.2	Schematic view of the extrapolation procedure	117
9.3	ND and FD reco-true reco-true matrices	119
9.4	ND spectra: reconstructed, data, true and reweighted	120
9.5	FD prediction: reconstructed, true and reweighted	121
10.1	FD hadron production uncertainty band	126
10.2	ND hadron production uncertainty band	127
10.3	Hadron production uncertainty in FD/ND ratio	128
10.4	Systematic effect of combined flux uncertainty	130
10.5	Systematic effect of MaCCQE uncertainty	137
10.6	Systematic effect of MaCCRES uncertainty	138
10.7	Systematic effect of MvCCRES uncertainty	139
10.8	Systematic effect of MaNCEL uncertainty	140
10.9	Systematic effect of MaNCRES uncertainty	141
10.10	Systematic effect of MvNCRES uncertainty	142
10.11	Summed GENIE uncertainty band for FD	143
10.12	Summed GENIE uncertainty band for ND	144
10.13	Systematic effect of combined GENIE uncertainty	145
10.14	Systematically shifted spectrum alternative Geant4 physics list: FTF_BIC	147
10.15	Systematically shifted spectrum alternative Geant4 physics list: FTFP_BERT	148
10.16	Systematically shifted spectrum alternative Geant4 physics list: QGSC_BERT	149
10.17	Systematically shifted spectrum alternative Geant4 physics list: QGSP_BIC_HP	150
10.18	Systematic effect of particle propagation energy scale uncertainty . . .	151
10.19	Systematic effect of particle propagation energy normalization uncertainty	152
10.20	Systematically shifted spectrum with $k_B = 0.02$ cm/MeV	154
10.21	Systematically shifted spectrum with $k_B = 0.01$ cm/MeV	155

10.22 Systematic effect of absolute scintillation production normalization uncertainty	156
10.23 Systematic effect of relative scintillation production normalization uncertainty	157
10.24 Systematic effect of absolute scintillation production energy scale uncertainty	158
10.25 Systematic effect of relative scintillation production energy scale uncertainty	159
10.26 Systematically shifted spectrum with -5% absolute energy scale	162
10.27 Systematically shifted spectrum with -5% absolute energy scale	163
10.28 Systematically shifted spectrum with random cell-to-cell variations	164
10.29 Systematically shifted spectrum with artificial attenuation slopes	165
10.30 Systematic effect of absolute calorimetric energy normalization uncertainty	166
10.31 Systematic effect of relative calorimetric energy normalization uncertainty	167
10.32 Systematic effect of absolute calorimetric energy scale uncertainty	168
10.33 Systematic effect of relative calorimetric energy scale uncertainty	169
10.34 Systematic effect of absolute detector mass uncertainty	171
10.35 Systematic effect of relative detector mass uncertainty	172
10.36 Systematic effect of FD muon track range uncertainty	174
10.37 Systematic effect of ND muon track range uncertainty	175
10.38 Systematic effect of absolute hadronic energy scale uncertainty	177
10.39 Systematic effect of relative hadronic energy scale uncertainty	178
10.40 Effect of total systematic uncertainty	180
11.1 FD observed spectrum with prediction in the absence of neutrino oscillation	182
11.2 CNN softmax output distribution for selected FD events with MC prediction	184
11.3 Slice N_{Hit} distribution for selected FD events with MC prediction	185
11.4 Number of tracks formed by KalmanTrack distribution for selected FD events with MC prediction	186
11.5 Muon track length distribution for selected FD events with MC prediction	187
11.6 Muon track start X distribution for selected FD events with MC prediction	188

11.7 Muon track start Y distribution for selected FD events with MC prediction	189
11.8 Muon track start Z distribution for selected FD events with MC prediction	190
11.9 Muon track stop X distribution for selected FD events with MC prediction	191
11.10 Muon track stop Y distribution for selected FD events with MC prediction	192
11.11 Muon track stop Z distribution for selected FD events with MC prediction	193
11.12 Muon track Z -direction cosine distribution for selected FD events with MC prediction	194
11.13 Hadronic cluster N_{Hit} distribution for selected FD events with MC prediction	195
11.14 Hadronic energy distribution for selected FD events with MC prediction	196
11.15 Muon ID distribution for selected FD events with MC prediction	197
11.16 Selected FD ν_μ CC QE candidate	199
11.17 Selected FD ν_μ CC QE candidate	200
11.18 Selected FD ν_μ CC QE candidate	201
11.19 Selected FD ν_μ CC QE candidate	202
11.20 Selected FD ν_μ CC QE candidate	203
11.21 Selected FD ν_μ CC QE candidate	204
11.22 FD observed spectrum with best-fit and systematic uncertainties	205
11.23 Measured 90% confidence interval with and without systematic uncertainties	206
11.24 Comparison of 90% confidence interval to existing results	207
11.25 Feldman-Cousins 90% confidence interval	209
12.1 Predicted sensitivity of this analysis assuming NOvA design exposure	211

Chapter 1

Introduction

This work aims to further the understanding of a type of particle called the neutrino, which is arguably the most mysterious of all known particles. Neutrinos are neutral particles of relatively small mass which are only known to interact through the weak and gravitational forces. Since they only couple to the subdominant forces, neutrino interactions are astonishingly rare; a neutrino could travel through a light year of lead and have less than a 50% chance of interacting [1]. From the perspective of particle physicists, the low probability of observing neutrino interactions has made their characterization an extremely challenging task.

Neutrinos undergo a curious phenomenon called oscillation, or the ability for members of the neutrino family to transform into other members. The NOvA (NuMI Off-axis Neutrino Appearance) experiment has built two detectors separated by a distance of nearly 1,000 km in order to observe neutrino oscillation. This work aims to better measure the frequency and amplitude of one neutrino oscillation mode, particularly by improving the method by which neutrinos are identified within NOvA detectors.

The data output from NOvA detectors can be interpreted roughly as top and side view images of neutrino interactions. In that sense, identification of neutrino interactions in NOvA detectors can be seen as an image classification task. The image classification community has made great strides through convolutional neural network classifiers. Implementing that technique for interaction classification in NOvA will be the central thrust of this dissertation.

1.1 The Standard Model

High energy particle physics seeks to address the age-old topic of describing the constituents of matter throughout the universe. The Greek atomist theories are among the earliest attempts at answering this question. Democritus is well known for putting forth the theory that all matter was made up of indivisible, or fundamental, particles called atoms; Plato furthered this idea by proposing that the all matter was comprised of four basic elements: earth, air, fire and water [2]. In the 19th century, Dimitri Mendeleev made huge strides in understanding the structure of matter through the organization of the periodic table [3]. The elemental atoms represented in the periodic table, however, are not fundamental. The Standard Model of particle physics represents our latest attempt to describe matter in terms of fundamental particles. Of course, there is no guarantee that the particles of the Standard Model are truly fundamental, but no experiment has been performed to suggest otherwise. A graphic which displays all of the particles in the Standard Model can be seen in Figure 1.1.

In the Standard Model, a particle is an entity with intrinsic charge, mass, and spin¹. Particles in the Standard Model form two distinct categories: fermions and bosons. In units of \hbar , fermions have a half-integer spin, e.g. $\frac{1}{2}, \frac{3}{2}$, while bosons have integer spin. Matter is primarily made up of the fermions in the Standard Model, all of which have a spin of $\frac{1}{2}$. The bosons in the Standard Model are force carriers, or in other words, they mediate the interactions of the electromagnetic, strong and weak forces. All of the bosons in the Standard Model have a spin of 1 or 0. Each charged particle in the Standard Model is complimented by an antiparticle which has the same mass and spin, but opposite charge [3].

Fermions are further grouped into two categories: quarks and leptons. Atomic nuclei are built up from quarks; a proton is comprised of two up quarks and one down quarks, a neutron is comprised of two down quarks and one up quark. Combinations of three quarks are known as baryons, while those of two quarks are called mesons. Recent experiments have revealed particles which are likely to contain four or five quarks [5, 6]. Distinct baryons and mesons are not merely formed by different combinations of quarks, but also from excited states of those combinations. Quarks in baryons and mesons are

¹Spin is the property of a particle which represents its intrinsic angular momentum. The particles are not actually spinning in a physical sense, although they behave as if they were [3].

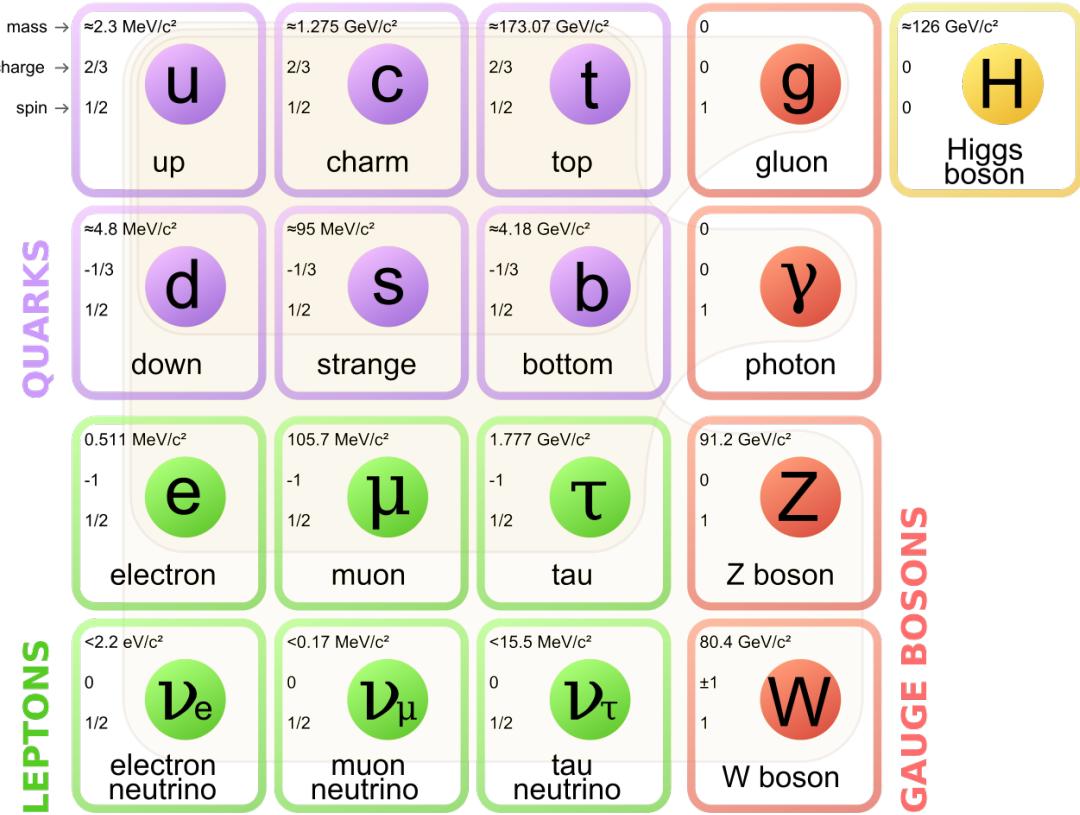


Figure 1.1: Particles in the Standard Model

There are six quarks: up, down, charm, strange, top, and bottom. The three charged leptons – e , μ , τ – each have a neutral partner in the form of a neutrino – ν_e , ν_μ , ν_τ , respectively. Interactions in the Standard Model are mediated by force-carrying bosons. The photon (γ) mediates the electromagnetic interaction, the gluon (g) mediates the strong interaction, and the W^\pm and Z^0 bosons mediate the weak interaction. The Higgs boson is responsible for imbuing particles with mass. Image reproduced courtesy of [4].

held together by the strong force, although they also interact through the weak and electromagnetic force. Antiparticle compliments of quarks are known as antiquarks [3].

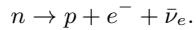
Leptons, on the other hand, do not interact through the strong force. The most familiar example of a lepton is the electron, which make up the exterior region of atoms in matter. The electron (e) has two heavier counterparts, the muon (μ) and tau (τ). These three particles, along with their antiparticles, represent the charged leptons. Each of the charged leptons can couple to a corresponding neutrino in weak interactions; thus the neutrinos are named: electron neutrino (ν_e), muon neutrino (ν_μ), and tau neutrino (ν_τ).

1.2 History of Neutrino Physics

1.2.1 The Neutrino Hypothesis

Mankind's first glimpse at the neutrino came in the early part of the 20th century along with studies of nuclear beta decay. Discovered in 1896 by Henri Bequerel, beta decay was characterized by the mysterious "beta rays" that were emitted by various types of radioactive material [7]. By 1900, however, Becquerel had categorized beta rays as electrons, suggesting that beta decay involved a heavy atom emitting a relatively light, but fast-moving electron² [11]. For a two-body decay, energy conservation can predict the energy of each particle. In the case of beta decay, emission of a relatively lightweight electron from a heavy atom should leave the atom roughly stationary, thus approximately determining the energy of the electron. James Chadwick measured the energy spectrum of beta decay electrons in 1914 and found that the spectrum was continuous rather than discrete and single-valued [12]. In other words, the measured range of electron energies was much larger than the uncertainty in atomic recoil energy. This result was clearly inconsistent with the two-particle process that had previously been hypothesized.

² We now know beta decay to involve a neutron within a nucleus decaying into a proton, electron and electron antineutrino, as follows:



Even though electrons were discovered in 1897 by J.J. Thompson in his examination of cathode rays [8], atomic structure was still shrouded in uncertainty. Geiger, Mardsen and Rutherford's gold foil experiment found evidence of the nucleus 1909, but neutrons were not hypothesized until 1920 by Rutherford [9] and remained undiscovered until work done in 1932 by James Chadwick. [10]

As a solution to this problem, in 1930 Wolfgang Pauli proposed that this decay could in fact be a three-body decay involving a third particle, which he then called a neutron, electrically neutral and also considerably lighter than a nucleus [13]. Two years later the actual neutron we refer to today was discovered by Chadwick [10] and in 1934 Enrico Fermi proposed that Pauli's particle be called a “neutrino.” [14]

1.2.2 Neutrino Discovery – Methods and Observations

Cowan and Reines made the first observation of neutrinos in 1956 [15]. The neutrinos they observed were produced through beta decays in a nuclear reactor. Since the field of particle physics had advanced considerably by this point, Cowan and Reines had enough information to expect the signal:

$$\bar{\nu}_e + p \rightarrow n + e^+. \quad (1.1)$$

Detectors used in their experiment employed scintillation counters to register energy deposition from candidate events. Neutrino events were selected by coincidence between two signals: a prompt electron-positron annihilation signal and a larger delayed signal produced when the neutron is captured by an atomic nucleus, as shown in Figure 1.2. The delay of the second signal spans the time for the fast-moving neutrons to thermalize – that is, to lose energy through nuclear scattering until reaching the energy scale required for capture.

Cowan and Reines measured a neutrino rate of 2.88 ± 0.22 counts/hour, representing a signal rate 3 times larger than the measured background [16]. The neutrinos observed in this experiment were later classified as electron “flavor” neutrinos.

Neutrino physics grew in scope with the discovery of a second class of neutrinos in 1962 [17]. Lee and Yang made a prediction that there existed neutrinos associated with muons in addition to those associated with electrons [18]. They also made detailed calculations predicting cross sections for interaction (1.1) as well as:

$$\nu_e + n^0 \rightarrow p^+ + e^- \quad (1.2)$$

$$\nu_\mu + n^0 \rightarrow p^+ + \mu^- \quad (1.3)$$

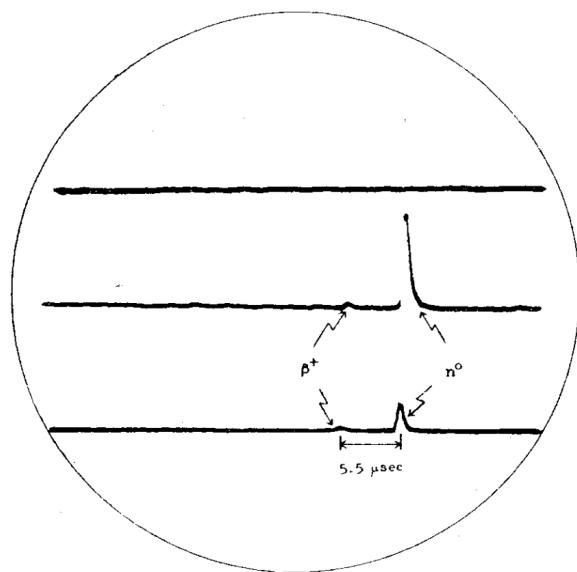


Figure 1.2: Oscilloscope trace from Cowan and Reines' experiment
The small bump is the electron/positron annihilation signal, the large spike is the neutron capture signal. Figure reproduced from [15].

$$\bar{\nu}_\mu + p^+ \rightarrow n^0 + \mu^+. \quad (1.4)$$

An experiment to observe the muon neutrino was devised based on the Alternating Gradient Synchrotron (AGS) at Brookhaven National Lab. AGS accelerated protons for collision upon a fixed target, resulting in production of charged pions. The pions would then decay to yield muons and muon neutrinos as follows:

$$\begin{aligned} \pi^+ &\rightarrow \mu^+ + \nu_\mu \\ \pi^- &\rightarrow \mu^- + \bar{\nu}_\mu. \end{aligned} \quad (1.5)$$

An arrangement of lead and steel was placed in the beamline to absorb the muons. Beyond the absorbers, spark chambers were deployed that would create sparks along the path traversed by charged particles. Neutrinos, unaffected by the muon absorbers, would very occasionally interact within the spark chambers according to (1.3) and (1.4) to produce muons. Activity in the spark chamber triggered a camera to produce photographs for analysis. To eliminate cosmic ray background, events analyzed were required to pass a preselection by satisfying two criteria—that the muon track must start and stop well within the spark chamber, and that their angle relative to the beam direction be less than 60° —leaving behind 113 events over a background of 1800. After surviving preselection, the remaining events were subdivided into four groups. 49 events were labeled as “short tracks,” and eliminated as likely being caused by neutron background. Another group of 22 were designated “vertex events” because they contained more than one track; these events were excluded since they were not representative of interactions (1.3) and (1.4). Yet another 9 events were classified as “showers” due to irregular track activity uncharacteristic of muons. The final group 34 events consisted of was characterized by a single clean muon track; these formed the first observed sample of muon neutrino interactions. In 1988, Leon Lederman, Melvin Schwartz, and Jack Steinberger were awarded the Nobel Prize in Physics for discovery of the muon neutrino.

The force carriers for the weak force, the W^+ and Z^0 bosons, were discovered at the Super Proton Synchrotron (SPS) at CERN, thus opening up a new realm of inquiry for particle physics [19, 20]. Along with the observation of these bosons came a measurement of their cross sections and decay rates. Z^0 bosons are expected to decay to neutrinos and contribute to its decay rate, but these decays are typically unobservable in collider

detectors; this deficit is typically called the invisible rate. By measuring the total Z^0 decay rate as well as the rate observed through visible processes, the difference can be taken to obtain the invisible rate. The measurement – made using the ALEPH detector on the Large Electron Positron collider (LEP) at CERN – determined of the total number of low-mass neutrino types that can interact via the weak force. ALEPH determined that number to be consistent with three light, active neutrino flavors, ruling out the possibility of four flavors at the 98% confidence level [21]. Measurement of invisible Z^0 width, however, only dictates the number of neutrinos that interact weakly. Neutrinos that do not interact weakly are known as sterile neutrinos.

The ALEPH measurement, on top of the previous associations of neutrinos with electrons and muons, suggested that the third neutrino type should be associated with taus. An experiment called Direct Observation of Nu Tau (DONUT) was constructed at Fermilab to detect ν_τ through interactions analogous with the electrons and muons:

$$\nu_\tau + n^0 \rightarrow p^+ + \tau^- \quad (1.6)$$

$$\bar{\nu}_\tau + p^+ \rightarrow n^0 + \tau^+. \quad (1.7)$$

Energetic protons taken from the Tevatron accelerator were directed into a tungsten beam dump to yield a shower of particles including heavy mesons. The primary source of ν_τ in the experiment was the D_S meson through its decay:

$$D_S \rightarrow \tau + \nu_\tau \quad (1.8)$$

Similar to the Brookhaven muon neutrino experiment, DONUT used an array of steel absorbers beyond the beam dump, but additionally implemented a set of deflecting magnets to steer muons and other light particles out of the beamline. Downstream of the absorbers was a detector comprised of alternating layers of two types of tracker modules: emulsion cloud chambers (ECC) and scintillating fiber trackers (SFT). The ECC were assembled with sheets of photographic emulsion that would become exposed along the track of a fast-moving charged particles. This detection scheme provided precise spatial resolution for tracks required for identification of taus by a characteristic kink in their track 2 mm from the ν_τ interaction point. Interleaved between the ECC,

the SFT was comprised of scintillating fibers that would produce light in the presence of charged particle radiation and transmit that light to image intensifiers for digitization. The SFT was used to identify regions containing activity within the ECC for analysis. In 2001 the DONUT collaboration published the results of this analysis; their paper reported the observation of four interactions compatible with (1.6) and (1.7) above an expected background of 0.34 events [22].

1.2.3 The Solar Neutrino Problem

In the 1960's Ray Davis and others set out to determine the flux of neutrinos incident upon the earth as a result of nuclear fusion within the sun [23]. Davis and his collaborators designed and built a detector based on a beta process that converts chlorine atoms to argon atoms, namely:



The cross section for this interaction was calculated in 1964 by John Bahcall [24]. Davis' detector was a large tank (placed underground at the Homestake mine to reduce cosmic ray background) filled with an organic compound called perchloroethylene³ (C_4Cl_8) [23]. Over time, the solar neutrino flux converted a measurable fraction of the chlorine atoms to argon. Every few weeks, Davis would bubble helium through the tanks to collect the argon atoms and measure the amount collected. Upon analysis, it was found that Davis' measured fluxes were a factor of three smaller than those predicted from solar models. Many doubted Davis' measurement because it relied on radiochemical processes rather than event-by-event analysis; however, errors could not be found in the experiment. Flux measurements were also made by the Kamiokande-II experiment and the Sudbury Neutrino Observatory [25, 26]. All results were found to be consistent with Davis' measurements but inconsistent with the prediction. The deficit in flux measured in these experiments became known as the "solar neutrino problem."

The solution to the solar neutrino problem came in the form of neutrino oscillation, that is, the concept that neutrinos can change flavor as they propagate. If neutrinos can change flavor, then some of the electron neutrinos produced in the sun would transform to muon and tau neutrinos before reaching earth. As a result, the aforementioned

³Perchloroethylene is commonly used as a dry cleaning agent

experiments would indeed measure a diminished neutrino flux relative to the prediction.

Bruno Pontecorvo first described neutrino oscillation in 1957 as a process that mixed neutrinos and antineutrinos [27]. These ideas were later expanded by Maki, Nakagawa and Sakata in 1962 to explain the small leptonic decay rate of hyperons⁴ through the mixing of electron and muon neutrinos [28]. The theory of neutrino oscillation was later expanded to include tau neutrinos. A detailed description of neutrino oscillation will come in the following section.

⁴Hyperon is jargon for a particle comprised of one or more strange quarks, but no charm quarks or bottom quarks.

Chapter 2

Neutrino Physics

2.1 Neutrino Interactions

Neutrinos are neutral¹ to both the electromagnetic and strong forces. Gravitational interactions are far too weak to observe. Thus, the only neutrino interactions observed in practice are mediated by the weak force [3].

Weak interactions are mediated by either the W^+ or Z^0 boson. Diagrams depicting neutrino interactions can be seen in Figure 2.1. Since the Z^0 is neutral, the interactions it mediates are called neutral current (NC) interactions, while those mediated by W^+ are called charged current (CC) interactions. The distinction between NC and CC is an important one due to charge conservation. A particle interacting through a neutral boson is left with its charge unchanged, while an interaction involving a charged boson must raise or lower the charge of the interacting particle. Thus, in an NC interaction, an interacting neutrino remains a neutrino. CC interactions, on the other hand, will have an interacting neutrino converted to its lepton partner. To the observer, the outgoing lepton in a CC interaction serves as a signature of the flavor of the interacting neutrino: for instance, the presence of a muon implies that a muon neutrino interacted. NC interactions, on the other hand, carry no flavor signature of the interacting neutrino [29].

The vast majority of neutrino interactions observed in matter involve scattering

¹In the Standard Model, neutrality with regard to a force implies that the particle does not couple to the force carriers of that force.

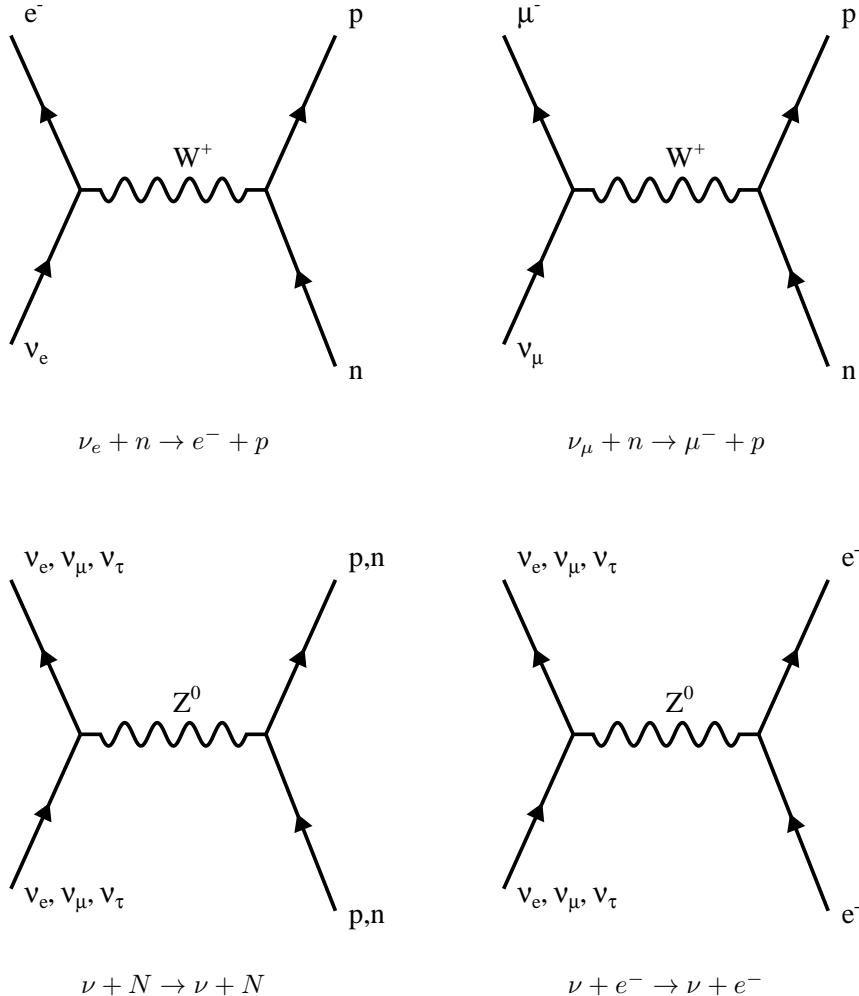


Figure 2.1: Neutrino interaction diagrams

CC and NC interactions are mediated by W^+ or Z^0 bosons, respectively. All figures above are meant to be interpreted with the time axis running from bottom to top, indicating a neutrino and some target exists in the initial state. The top panes show CC interactions, while the bottom panes show NC interactions. The top left (right) pane shows an electron (a muon) neutrino interacting with a neutron through a W^+ to produce an electron (a muon) and a proton in the final state. The bottom left pane shows an NC interacting with nucleon, while the bottom right pane shows an NC interaction with an electron. Note that CC interactions with leptons, such as an electron, are permitted, although they are far rarer. An example of such an interaction can be seen in Figure 2.2c.

against a nuclear target. Scattering with atomic electrons is permitted, although it happens at a far lower rate due to the relatively small mass of the electron. On the leptonic side of the scattering process, the incident neutrino is either converted to a charged lepton in the CC case, or remains a neutrino in the NC case. The nuclear side of the interaction, however, can be far messier due to the structure of the nucleus. The simplest case is called when the neutrino interacts with a single nucleon without significant coupling to the nucleus. For the NC case, this is called an elastic interaction [30], while the CC version is called a quasi-elastic (QE) interaction since the nucleon changes form (e.g. neutron to proton) [31]. A neutrino interaction can also excite a resonance within the nucleus to produce a hadron, for instance, a pion; these scattering events are referred to as resonant production [32]. More complicated scattering events which cause significant disturbance of the nuclear system (and typically complete breakup) are called deep-inelastic scattering (DIS) events [18, 33, 34].

2.2 Neutrino Oscillation

2.2.1 A Straightforward Interpretation

When a neutrino interacts, it does so in one of its three weak flavor states: ν_e , ν_μ or ν_τ . But in contrast to our intuition, the flavor states do not have definite mass; each is a mixture of three components with distinct masses: ν_1 , ν_2 and ν_3 . A neutrino produced in a particular flavor state—for instance, an electron neutrino produced in the sun—is composed of some fraction of ν_1 , a different fraction of ν_2 and yet another fraction of ν_3 . The mass components propagate as three distinct waves, summing together to make a total wave that is the neutrino. As these waves travel through space, they move in and out of phase with each other to cause interference; as a result, the composition fractions of the neutrino change. For every composition of ν_1 , ν_2 and ν_3 , there are distinct probabilities of interacting as ν_e , ν_μ or ν_τ . Thus, when a neutrino reaches its interaction point—say, the solar neutrino arriving at Homestake mine—its makeup will have changed to something that may be more likely to interact as a different flavor. The theory that describes this phenomenon is neutrino oscillation.

2.2.2 Oscillation in Vacuum

The weak flavor states are the eigenstates of the Hamiltonian of weak interactions. A neutrino produced as flavor $l = e, \mu, \tau$ can be written as a superposition of mass states $\alpha = 1, 2, 3$ in terms of the elements of a unitary 3×3 matrix $U_{l\alpha}$ as follows:

$$|\nu_l\rangle = \sum_{\alpha=1}^3 U_{l\alpha} |\nu_\alpha\rangle \quad (2.1)$$

This superposition describes a change of basis and can be likened to rotations in a three dimensional space. $U_{l\alpha}$ can be written in terms of three Euler angles – θ_{12} , θ_{23} , and θ_{13} – as well as a complex phase, δ , which allows for the possibility of CP violation². The matrix $U_{l\alpha}$ (written here with c_{ij} and s_{ij} as shorthand for $\cos \theta_{ij}$ and $\sin \theta_{ij}$, respectively) is then the product of the three rotations, commonly referred to as the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [29]:

$$U_{l\alpha} = \begin{pmatrix} c_{13}c_{12} & c_{13}s_{12} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & c_{13}s_{23} \\ s_{23}s_{12} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{13}c_{23} \end{pmatrix}. \quad (2.2)$$

This matrix, along with the superposition in (2.1) gives the fractions of ν_1 , ν_2 and ν_3 that make up ν_e , ν_μ or ν_τ . To see how neutrinos oscillate we must consider their time evolution:

$$|\nu_l(t)\rangle = \sum_{\alpha=1}^3 U_{l\alpha} e^{-iE_\alpha t} |\nu_\alpha\rangle, \quad (2.3)$$

where E_α represents the energy³ of the eigenstate, which depends on its mass and momentum:

$$E_\alpha = \sqrt{p_\alpha^2 + m_\alpha^2} \quad (2.4)$$

²CP (charge-parity) symmetry is a property which dictates a physical process is invariant under inversion of charge and spatial orientation (parity). A non-zero value of δ would indicate violation of CP symmetry[29].

³Physical quantities are displayed in natural units, $\hbar = c = 1$

For highly relativistic neutrinos, the momentum is much larger than the mass, so we have $|p_\alpha| \simeq |p| \simeq E$ allowing us to write (2.4) as:

$$E_\alpha \simeq p(1 + \frac{m_\alpha^2}{2p^2}) \simeq (E + \frac{m_\alpha^2}{2E}). \quad (2.5)$$

We can then substitute this into (2.3):

$$|\nu_l(t)\rangle = e^{-iEt} \sum_{\alpha=1}^3 U_{l\alpha} e^{-i\frac{m_\alpha^2}{2E}t} |\nu_\alpha\rangle. \quad (2.6)$$

To find the probability of observing the transition between flavor states $\nu_l \rightarrow \nu_m$ (also valid for $l = m$, as in the case of $\nu_\mu \rightarrow \nu_\mu$) we square the inner product between the two states:

$$P_{l \rightarrow m}(t) = |\langle \nu_m | \nu_l(t) \rangle|^2. \quad (2.7)$$

More explicitly, we have

$$P_{l \rightarrow m}(t) = \sum_{\alpha=1}^3 \sum_{\beta=1}^3 U_{l\alpha}^* U_{m\alpha} U_{l\beta}^* U_{m\beta} e^{i\frac{m_\alpha^2}{2E}t} e^{-i\frac{m_\beta^2}{2E}t}, \quad (2.8)$$

which, through the unitarity of U and use of some trigonometric identities, can be written as:

$$\begin{aligned} P_{l \rightarrow m}(t) &= \delta_{lm} - 4 \sum_{\alpha>\beta} \Re(U_{l\alpha}^* U_{m\alpha} U_{l\beta}^* U_{m\beta}) \sin^2 \left(\frac{\Delta m_{\alpha\beta}^2}{4E} t \right) \\ &\quad + 2 \sum_{\alpha>\beta} \Im(U_{l\alpha}^* U_{m\alpha} U_{l\beta}^* U_{m\beta}) \sin \left(\frac{\Delta m_{\alpha\beta}^2}{2E} t \right). \end{aligned} \quad (2.9)$$

Here, δ_{lm} is the Kronecker delta and $\Delta m_{\alpha\beta}^2 = m_\alpha^2 - m_\beta^2$. In equation (2.9), the prefactors on the \sin and \sin^2 terms are real numbers that depend on the Euler angles θ_{12} , θ_{23} and θ_{13} as well as the CP violating phase δ . For relativistic neutrinos, we can take $t \simeq L$, the distance between the source and the interaction point; substituting this in (2.9) yields

a form relevant for oscillation experiments:

$$\begin{aligned} P_{l \rightarrow m}(L) = & \delta_{lm} - 4 \sum_{\alpha > \beta} \Re(U_{l\alpha}^* U_{m\alpha} U_{l\beta}^* U_{m\beta}) \sin^2 \left(\frac{\Delta m_{\alpha\beta}^2}{4E} L \right) \\ & + 2 \sum_{\alpha > \beta} \Im(U_{l\alpha}^* U_{m\alpha} U_{l\beta}^* U_{m\beta}) \sin \left(\frac{\Delta m_{\alpha\beta}^2}{2E} L \right). \end{aligned} \quad (2.10)$$

It is interesting to note the dependence on the mass squared differences, $\Delta m_{\alpha\beta}^2$ rather than the masses themselves. As a result, oscillation experiments are not sensitive to absolute neutrino masses. Measurement of these parameters will be discussed in detail in later sections.

2.2.3 Oscillation in Matter

Neutrinos propagating in matter can scatter upon components of atoms through the exchange of W^\pm and Z^0 bosons. In analogy to light propagating in matter, neutrinos can undergo coherent scatterings with the surrounding particles. Coherent scatterings are those which do not change the state of the system, preserving the momentum and energy of the particles involved.

A quantitative treatment of this phenomenon imparts particles with an effective mass that modifies their physical mass. In the case of light, the effective mass manifests itself in the familiar index of refraction that diminishes the speed of light in matter. For neutrinos, coherent scattering interactions are mediated through the exchange of W^\pm and Z^0 bosons; the Feynman diagrams for these interactions can be seen in Figure 2.2. The neutral current interactions (those moderated by Z^0 exchange) are the same for all three neutrino flavors and have no effect on oscillation probabilities. On the other hand, since terrestrial matter contains electrons but not muons and taus, the charged-current interaction with electrons only occurs for ν_e , giving rise to an effective potential [29]:

$$V_{CC} = \sqrt{2}G_F n_e, \quad (2.11)$$

where n_e is the number density of electrons in the medium and G_F is Fermi's constant. Since this effective potential only exists for electrons, it modifies the oscillation

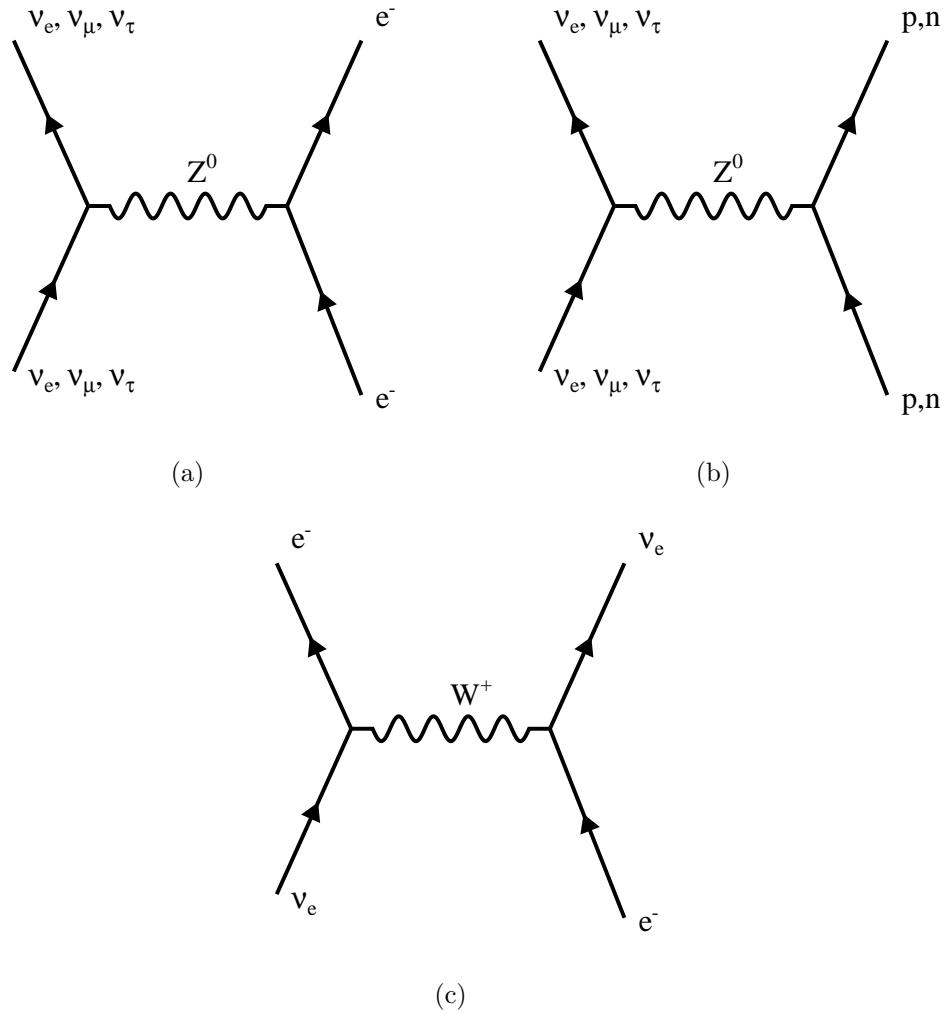


Figure 2.2: Interactions which permit coherent scattering

Coherent interactions are those which leave the momentum and flavor of the scattering products unchanged. In all three diagrams, the time axis runs from bottom to top. The top left pane shows NC neutrino scattering electron scattering against an atomic electron, whereas the top right shows scattering with an atomic nucleon. The bottom shows a CC scattering with an atomic electron; while the electron and ν_e have switched positions, the initial and final state are identical.

probabilities given in equation (2.10), specifically changing the sine term [29]:

$$\sin\left(\frac{\Delta m_{\alpha\beta}^2}{2E}L\right) \rightarrow \frac{\Delta m_{\alpha\beta}^2}{\Delta m_{\alpha\beta}^2 \pm EV_{CC}/2} \sin\left(\left(\frac{\Delta m_{\alpha\beta}^2}{2E} \pm \frac{EV_{CC}}{2}\right)L\right). \quad (2.12)$$

with the \pm corresponding to a difference in sign for neutrinos and antineutrinos, respectively. This effect is known as the Mikheyev-Smirnov-Wolfenstein (MSW) effect [35].

2.2.4 Recent Experiments and Results

In the last two decades experimental neutrino physics has turned the corner from an era of discovery to an era of precision. The field has made this transition through the use of plentiful sources of neutrinos as well as the development of kiloton scale detectors. Interestingly enough, despite the growth in scale of neutrino experiments, the techniques have remained largely the same as those of the discovery era.

Reactor neutrino experiments [36, 37, 38] have become sensitive enough to measure the survival probability for electron antineutrinos, $P(\bar{\nu}_e \rightarrow \bar{\nu}_e)$. Accelerator neutrino experiments [39, 40, 41] can now measure similar probabilities for muon neutrinos, namely, $P(\nu_\mu \rightarrow \nu_\mu)$ and $P(\nu_\mu \rightarrow \nu_e)$. In addition to these common sources, a third plentiful source comes in the form of atmospheric neutrinos. Cosmic-ray protons interacting in the upper atmosphere produce showers of predominantly pions and kaons that subsequently decay into neutrinos. If a detector can reconstruct the zenith angle of an atmospheric neutrino, it can determine the distance between its creation and interaction points. Therefore, a detector that can sample all zenith angles in turn samples a wide range of neutrino baselines ranging from the height of the atmosphere up to the diameter of the earth.

The mixing angles have now all been measured to be nonzero by measuring either survival or appearance probabilities. Precise measurements of θ_{12} , the “solar” mixing angle, have been performed by solar neutrino experiments such as SNO and reactor experiments such as KamLAND [26, 36, 42]. These measurements yield $\theta_{12} \approx 34^\circ$ with about one 1° of uncertainty. The “atmospheric” mixing angle, θ_{23} has been measured as $\sin^2(\theta_{23}) = 0.514^{+0.055}_{-0.056}$ by neutrino observatories such as SNO and Super-Kamiokande (Super-K) as well as long baseline experiments such as MINOS [26, 43, 44, 42]. Most

recently, Daya Bay, RENO, MINOS, T2K, and NOvA have made measurements of θ_{13} that are not consistent with zero [37, 38, 39, 40]. The PDG best fit gives $\sin^2(2\theta_{13}) \approx 0.085 \pm 0.005$

Measurements have also been made of the squared-mass differences. It is known that Δm_{21}^2 is $(7.53 \pm 0.18) \times 10^{-5}$ eV² and that $|\Delta m_{32}^2|$ is much larger at $(2.42 \pm 0.06) \times 10^{-3}$ eV². This means that ν_1 and ν_2 could be similar in mass and that ν_3 is relatively disparate in mass. A remaining question is the sign of Δm_{32}^2 , commonly called the ordering of the mass hierarchy. In other words, we do not know whether ν_3 is heavier (referred to as a normal hierarchy) or considerably lighter (inverted hierarchy) than its counterparts.

The work presented here will make measurements of $\sin^2(\theta_{23})$ and Δm_{32}^2 .

Chapter 3

NOvA

Neutrinos at the Main Injector (NuMI) [45]. is a high-intensity neutrino source located at Fermilab. The NuMI Off-axis ν_e Appearance (NOvA) experiment employs this source as one of the latest efforts in the class of long baseline neutrino experiments. Long baseline neutrino experiments aim to observe neutrino oscillation in a detector placed at some long distance – known as the baseline – from a neutrino source. NOvA places two detectors for observation of neutrinos, one near the source and another 810 km away in Ash River, Minnesota. The Near Detector (ND) is used to observe the initial NuMI flux at a distance where oscillation are negligible; the Far Detector (FD) observes the NuMI flux subject to neutrino oscillation. NuMI will be discussed in section 3.1, and the NOvA detectors in 3.2. The detectors are comprised of long cells which are arranged horizontally and vertically to give orthogonal views of particle interactions.

NOvA aims to measure and θ_{23} , δ , and the neutrino mass hierarchy though sensitivity to the probabilities $P(\nu_\mu \rightarrow \nu_\mu)$ and $P(\nu_\mu \rightarrow \nu_e)$. Preliminary measurements of nonzero measurement of θ_{13} are consistent with measurements from Daya Bay, RENO and T2K [41]. It is also hoped that NOvA will hone the measurement of θ_{23} to determine whether it is greater or less than 45° . This work aims to further techniques which will aid in both of these efforts, although we focus on the application to muon neutrino identification for measurement of θ_{23} and $|\Delta m^2_{32}|$.

The NuMI beam can run in modes which predominantly produce either neutrinos or antineutrinos. As described in Section 2.2.3, matter effects lead to differences in the oscillation probabilities observed for neutrinos and antineutrinos. By comparing the

respective appearance probabilities, it could be possible to determine whether the mass hierarchy is normal or inverted. The difference between neutrinos and antineutrinos could also be used to measure the CP violating phase, δ .

Oscillation analysis requires determination of the flavor state of an interacting neutrino, in particular ν_e which have appeared or ν_μ which have survived. Flavor state can be ascertained in charged-current (CC) interactions which leave a charged lepton in the final state; an electron in the case of ν_e or muon in the case of ν_μ . Neutral-current (NC) interactions, on the other hand, bear no flavor signature and serve as a background in ν_e appearance or ν_μ disappearance analyses, although they can serve as a signal in more exotic searches. Such identification is at the core of the NOvA detector design.

3.1 NuMI and Off-axis Alignment

In a process similar to the Brookhaven experiment that discovered the muon neutrino, the NuMI beam is generated by protons accelerated to 120 GeV by the Main Injector at Fermilab. A schematic of the NuMI configuration can be seen in Figure 3.1.

Protons from the Main Injector are directed into a graphite target producing a shower of hadrons including pions and kaons. The hadron showers are focused into a beam by the magnetic field produced in “focusing horns,” which are tuned to select charged pions by steering them forward while steering other particles out of the beam. The magnetic field in the horns reaches a peak strength of 30 kG resulting from a horn current of 200 kA. Downstream of the focusing horns, particles are free to decay in flight inside a 675 m decay pipe. Muon neutrinos are then produced when charged pions decay as shown in equation (1.5). Alternatively, anti-pions can be focused by switching the direction of the electric current in the focusing horns, thus NuMI can serve as either a neutrino or antineutrino beam. In addition pions, a small component of kaons and other heavier mesons pass through the focusing horns. Heavy mesons can decay to electrons and ν_e at higher rates, which introduces electron neutrino background in the predominantly ν_μ beam. Electron neutrinos are also introduced into the beam through the decay of muons in the decay pipe. Charged particles such as muons and electrons produced in the decay pipe are absorbed in rock upstream of the NOvA Near Detector, eliminating them as background. The NuMI facility can achieve an intensity of 700 kW

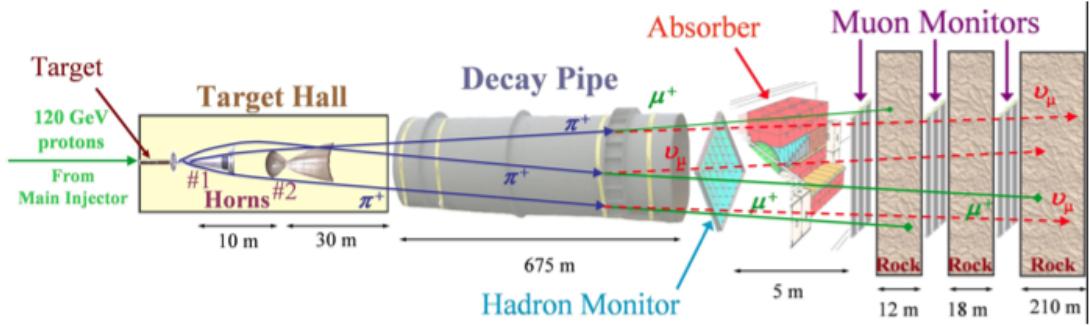
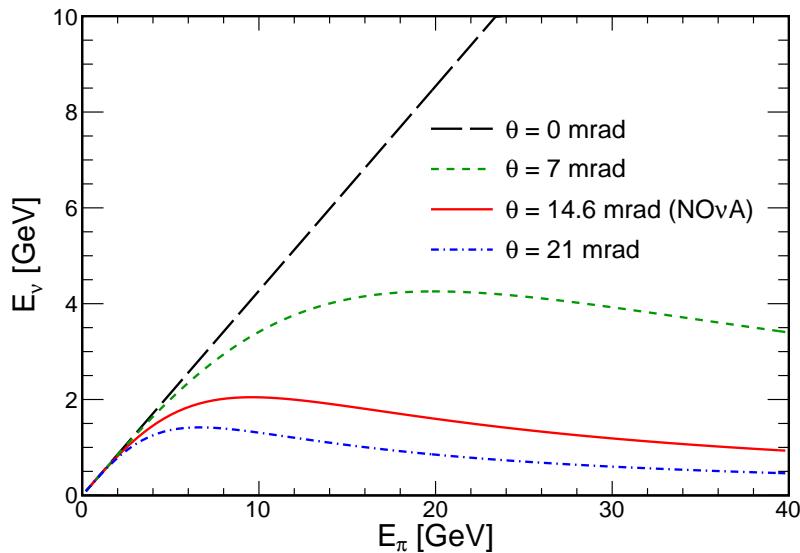


Figure 3.1: A schematic of the NuMI neutrino source

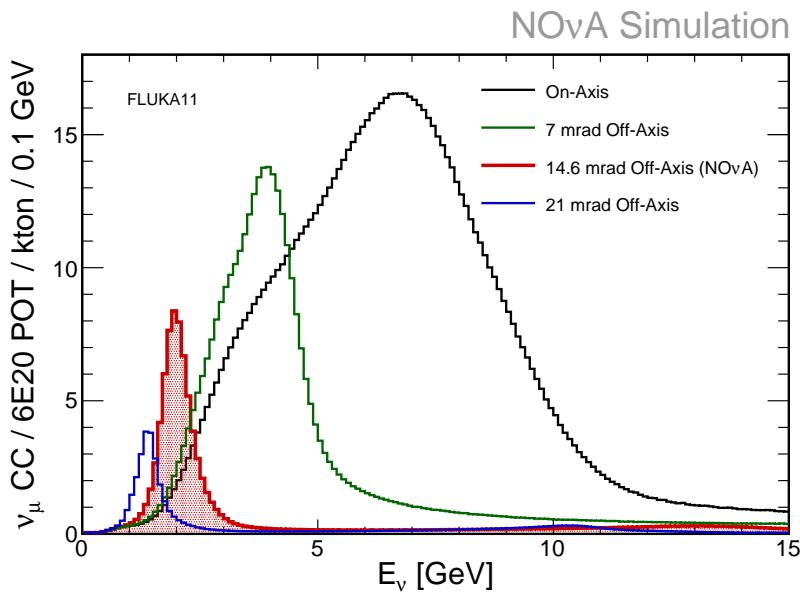
Protons extracted from the Main Injector are directed into a graphite target. Magnetic focusing horns are used to focus charged hadrons, primarily pions. Pions decay in the decay pipe to produce muons and muon neutrinos. The muons are absorbed in rock downstream, but the neutrinos are not. Image reproduced courtesy of [47].

by delivering spills of 4.9×10^{13} protons to the target every 1.3 seconds [46]. Each beam spill spans a duration of $10 \mu\text{s}$.

Rather than being placed directly in the NuMI beam line, the NOvA detectors are placed 14.6 milliradians off of the beam axis. As displayed in Figure 3.2a, the neutrinos produced on-axis demonstrate a strong dependence on the parent pion energy, as opposed to the weak energy dependence of those produced slightly off-axis. Since the pions produced at the target have a spread of energies, this weak dependence helps to narrow the neutrino energy spectrum, as seen in Figure 3.2b. The neutrino energy spectrum at 14.6 milliradians is characterized by a sharp peak near 2 GeV. With a baseline of 810 km, the NOvA FD observes a ν_μ disappearance minimum between 1-2 GeV, which makes that region the most sensitive to oscillation. The narrowness of the peak is desirable in order to reduce the component of the spectrum above 2 GeV which can lead to excess NC background contamination. NC interactions produce only a small amount of visible energy since a large fraction is carried away by the outgoing neutrino. As such, high energy NC events can be misidentified as CC events, so the narrow energy peak helps to reduce this background.



(a) Neutrino energy vs. pion energy at various angles relative to the beam axis



(b) Expected ν_μ event counts vs. neutrino energy at various angles relative to the beam axis

Figure 3.2: Effects of off-axis design

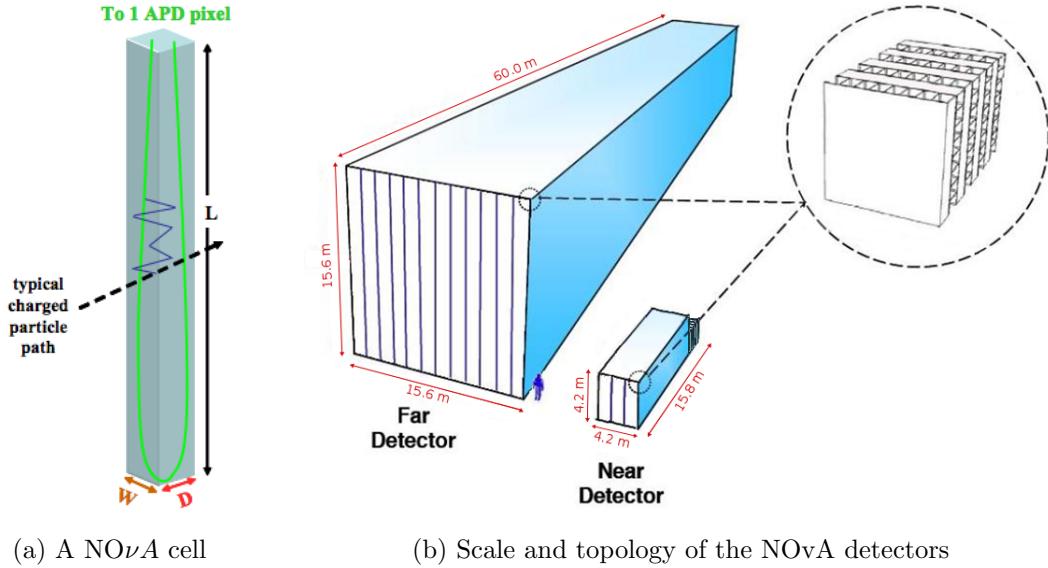


Figure 3.3: Drawings of the NOvA detectors

3.2 NOvA Detectors

The NOvA detectors are designed to allow fine-grained reconstruction of electromagnetic and hadronic showers. Such a detailed view can allow identification of ν_μ and ν_e CC interactions, and help to tell them apart from NC interactions. To achieve this goal, the detector topology is a repeated structure of unit cells which are much smaller than the electromagnetic and strong interaction lengths of the composite materials. Charged particles produce light as they traverse the cells, which enables detection. Cells are long, but arranged in alternating horizontal and vertical planes to provide measurements of Cartesian coordinates. The detector structure can be visualized in Figure 3.3b. The detectors are designed to be functionally identical to each other; they are built with the same materials and geometry to constrain systematic uncertainties due to variation in detector efficiency.

The NOvA FD is in a detector hall on the surface of the earth in northern Minnesota. The ND, on the other hand, is roughly 300 m underground at Fermilab in Illinois.

3.2.1 Cell Structure

The rigid structure of the cells is comprised of tubular *cells* extruded from poly-vinyl chloride (PVC) resin. Each cell is approximately 4 cm wide and 6 cm deep; FD cells are nearly 15 m long, while ND cells are about 4 m long. Cells are arrayed into planes, with adjacent planes alternating between horizontal and vertical alignment. Configured in this way, cells allow Cartesian coordinates to be inferred; vertical cells provide a measurement of the *X*-coordinate, horizontal cells the *Y*-coordinate, and planes the *Z*-coordinate. A visualization of the detector topology can be seen in Figure 3.3b. Planes are arranged transverse to the beam direction, which implies the *Z*-axis is directed roughly along the beam axis. More precisely, the *Z*-axis is in the plane of the horizon, and the beam is directed about 3° above the horizon. As such, the *Y*-axis is aligned with the zenith and *X*-axis chosen to complete a right handed coordinate system [46].

Cells are aligned side-by-side (along the 6 cm side) into units of 32 cells referred to as a module. Planes are thus a collection of adjacent modules which are glued face-to-face to produce detector *blocks*. In the FD, planes are 12 modules wide, blocks contain 32 planes, and the complete detector is 28 blocks. The ND has three blocks of 24 planes, each of which is 3 modules wide. Additionally, the ND is augmented with an extra downstream block known as the *muon catcher*. As its name suggests, this block is designed to *range-out* (or stop) muons, which can traverse a range of several meters from the ν_μ CC interactions seen in NOvA. Interleaved between each pair of horizontal and vertical modules in the muon catcher is a 10 cm thick steel plane. Steel absorbs extra energy from the muons so that they will range-out within the detector, thus allowing their total energy to be estimated based on their range. The muon catcher is comprised of 22 active planes interleaved with 10 steel planes. Muon catcher planes are three modules wide for vertical planes, but only two modules tall for horizontal planes.

3.2.2 Light Collection

Each cell is filled with scintillator-doped mineral oil to produce light in the presence of swiftly moving charged particles. The mixture is roughly 95% mineral oil by mass, blended with a primary scintillant which produces UV light and two wavelength shifting components [48]. A strand of wavelength-shifting fiber runs up and down the length of

each cell with both ends of fiber emerging at one end of the cell, visualized in figure 3.3a. The fiber serves the purpose of capturing and transmitting the scintillation light in two steps; first by lengthening its wavelength through molecular absorption and reemission, and then trapping it within the fiber by total internal reflection.

At the end of the cells, photons trapped in the WLS fiber are converted to photo-electrons on an avalanche photodiode (APD) to produce an electrical signal. One APD board contains 32 pixels to read out each cell of module separately. APD signals are read out by four electronic components built into a front end board (FEB) designed specifically for NOvA. An application-specific integrated circuit (ASIC) shapes the pulses. The ASIC is CR-RC circuit which amplifies and widens the signals in time to better distinguish physics signals from noise. Shaped pulses are passed to an analog-to-digital converter (ADC) which samples at 64 MHz time-to-digital (TDC) clock-ticks. The sampling of the 32 pixels is multiplexed, eight-fold in the FD and two-fold in the ND. Zero suppression of the ADC samples is performed by a field programmable gate array (FPGA) which employs a dual correlated sampling (DCS) technique to trigger a hit to be passed to the data acquisition (DAQ) system. Under the DCS scheme, each ADC sample is compared to the third latest sample. If the difference between those two samples is greater than a fixed threshold, a hit is passed along to be recorded by the DAQ.

3.2.3 Data Acquisition

The NOvA data concentrator module (DCM) is responsible for collecting data from up to 64 APD/FEB pairs. A DCM is a device mounted on the detector with an embedded CPU running Linux and custom NOvA DAQ software. Input and output is transmitted using a standard CAT5e link network link. Output from a DCM is transmitted to a buffer farm over 200 computing nodes. These nodes are arranged to store data in a circular ring buffer, or in other words, when new data arrives, the oldest data is overwritten. The buffer farm is capable of storing FD data for upwards of several minutes, during which time any subset of the buffered data can be written to disk and stored permanently. Trigger decisions are issued through the custom NOvA global trigger software, which can receive external triggers as well as data-driven triggers based on analysis of data stored in the buffer farm. This work primarily relies on external

triggers. Neutrino data is recorded in the ND and FD by the so-called NuMI trigger, which records $550\ \mu\text{s}$ of data centered around the $10\ \mu\text{s}$ NuMI beam spill. FD cosmic-ray data – used for calibration and background estimation – is recorded using a 10 Hz minimum bias trigger, commonly referred to as the cosmic trigger. Cosmic events in the ND are a commodity since the detector is underground; a minimum bias trigger at any sampling rate simply does not record enough activity. In place of such a trigger, a simple activity-based data-driven trigger is used to collect cosmic-ray data in the ND. This trigger reads out data whenever a certain number of hits are recorded above a certain threshold.

Data is written to computer hard disks at the respective ND and FD sites. Each site is capable of storing roughly one month worth of data. Ultimately, the data is transferred to Fermilab for permanent storage, where it is archived on tape and made available for offline reconstruction and physics analysis.

3.2.4 Event Topologies in NOvA

The finely segmented structure of NOvA detectors provides a relatively detailed view of the GeV scale neutrino interactions from the NuMI beam. Muons, electrons, and hadrons all leave distinct signatures in NOvA, as demonstrated in Figure 3.4. All charged particles passing quickly through matter lose energy through ionization¹ [42]. Since they are heavy and don't interact strongly with atomic nuclei, muons leave long tracks in the detector. Electrons, on the other hand, are prone to forming electromagnetic showers, which tend to diffuse energy over a shorter distance and with a wider transverse spread. Hadrons are prone to interactions with atomic nuclei, leading to hard scatters along their trajectory in addition to the standard energy loss from ionization. Exploiting these characteristic differences is essential to measuring oscillation parameters with NOvA since the ν flavor signature is indicated by the presence and type of charged lepton in the final state of an interaction.

Cosmic rays are a prevalent background in the NOvA FD. The vast majority of cosmic-ray activity is the result of muons which traverse the entire detector, although a small fraction of them will range-out within the detector. Cosmic-ray activity can also

¹ Put more simply, charged particles disturb atomic electrons in matter as they pass through, causing energy to be exchanged.

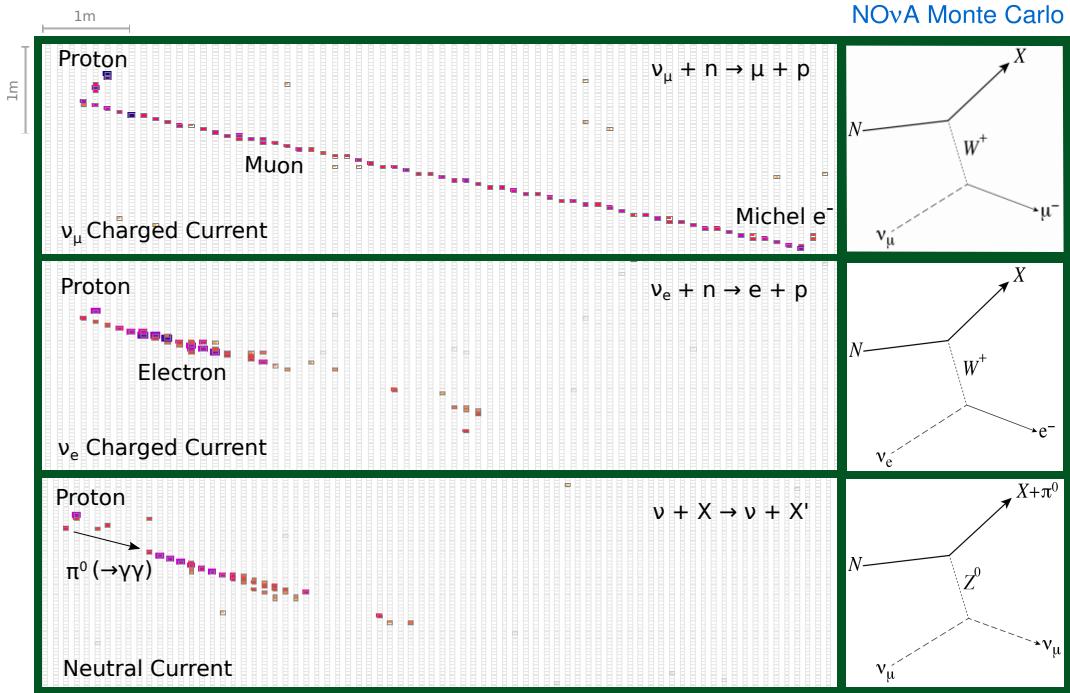


Figure 3.4: Event topologies in NOvA

Shown are three example event topologies in NOvA. The top pane shows a ν_μ CC interaction with a long muon track and visible activity from a proton. An electron shower from a ν_e CC interaction is visible in the middle pane. The third pane shows a NC interaction with a neutral pion decaying to two photons, which forms the dominant background in NOvA's ν_e appearance search. The electromagnetic radiation length in NOvA is relatively long compared to the size of a cell, which aids in distinguishing such events.

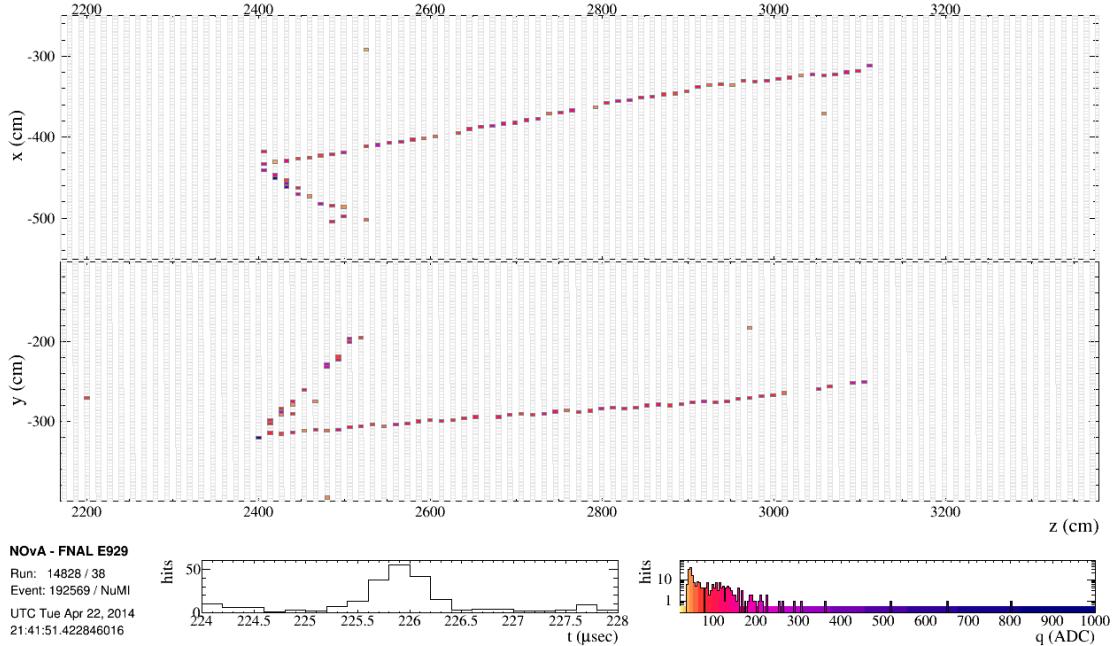


Figure 3.5: An example NOvA event display

The top projection is an X vs. Z view, the bottom is Y vs. Z . The bottom left portion of the display This is a candidate ν_μ charged-current event from the Far Detector.

consist of hadrons or electromagnetic showers, although it is mostly concentrated at the top of the detector, as will be discussed in Chapter 8.

3.2.5 The NOvA Event Display

The NOvA event display provides a detailed representation of NOvA detector readout. An example event display can be seen in Figure 3.5. The largest component in the event display is the cell-hit readout itself, shown separately through the X - Z (vertical module) and Y - Z (horizontal module) views. Additionally, the event display shows information about the data collection; for instance, a timestamp and trigger information. The event display also shows histograms depicting the time and charge distributions for the hits displayed. Hits are typically colored by their time relative to the start of the readout window or the deposited photoelectron charge in ADC units. In either case, the corresponding histogram is colored to display the color scale.

3.2.6 Exposure

In order to scale distributions of selected events, careful tracking of exposure is necessary. The important measures of exposure for NOvA are livetime, protons on-target (POT), and detector mass. Livetime is used for scaling constant processes like cosmic rays and electronic noise. Exposure to neutrinos, however, is proportional to the number of decayed pions in the NuMI beam. The number of pions depends on the number of protons delivered to the target, which is tracked spill-by-spill and recorded in a database. NOvA detectors can also be operated in configurations which leave blocks out of the readout, so careful tracking of active detector mass is also necessary for accurate normalization.

Chapter 4

Monte Carlo Simulation

The NOvA Monte Carlo (MC) simulation chain involves many components to deliver an accurate representation; starting with protons on the NuMI target and running all the way through APD readout in the NOvA detectors. FLUKA and FLUGG are used to simulate the flux in the NuMI beamline [49]. Neutrino interactions within the detectors are simulated using GENIE [50]. Geant4 is used to propagate products of the neutrino interactions through a detailed model of the NOvA detectors [51] Custom NOvA simulation software converts energy depositions into electronic signals [52]. While this chapter aims to describe the methods involved in simulating detector effects, it also provides a detailed description of the effects themselves.

4.1 Beam Simulation

The NuMI beam simulation is used to predict the neutrino flux in both the ND and FD. Simulated protons from the Main Injector are directed upon a model of the target and its surroundings; the resulting shower of hadrons is tracked in detail in order to predict neutrino production.

While the absolute flux is predicted in both detectors, it is really the ratio of the predicted fluxes which enters into an oscillation analysis. ND data are extrapolated to the FD to mitigate systematic errors the absolute flux estimate, so the beam simulation mainly serves to transfer the observed ND spectrum into an FD prediction. The extrapolation procedure will be described in detail in Section 9.1, although it should be noted

here that the primary difference in the fluxes seen between the two detectors arise from geometric effects. The ND is close to the NuMI source; despite its relatively small size, it subtends a much larger solid angle, and thus a spectrum of incident neutrino angle and energy. Also note that the detectors themselves do not observe the neutrino flux directly, but rather its convolution with cross section, which is discussed in the following section.

A detailed representation of the NuMI target facility forms the basis of beam simulation. The geometry of the target hall is encoded in the Geant4 geometry format to capture the positions and compositions of the target, collimator, focusing horns, and decay pipe. The FLUKA [49] particle simulation package is used to track the ancestry of hadrons through the target and their decay downstream. An interface between FLUKA and the Geant4 geometry is provided by FLUGG [53]. In this framework, neutrinos produced in the beam can be traced back through the hadronization chain to the primary proton which interacted in the target. Such precision allows detailed studies to estimate uncertainties based on particle production and interaction models. The FLUKA/FLUGG simulation has been cross-checked by an independent Geant4 simulation (based on the same geometry) and shown to produce comparable results. Comparisons have also been made to data both from the NA49 [54] and MIPP [55] experiments in order to estimate systematic errors.

The output of the beam simulation is a representation of the NuMI flux in terms of flavor and momentum which can be passed to downstream simulation components. A depiction of the observed event rate in both detectors can be seen in Figure 4.1.

4.2 GENIE– Neutrino Interaction Simulation

GENIE is a comprehensive neutrino MC generator for the experimental physics community. According to the GENIE collaboration mission statement, the generator aims to simulate processes for all neutrino species and all nuclear targets, from the MeV to PeV scale. The input to GENIE is the simulated detector flux from FLUKA/FLUGG; the output is a collection of neutrino interactions described as GENIE event records. GENIE event records catalog the flavor and kinematics of the incoming neutrino, struck nucleus, and daughter particles which leave the nucleus.

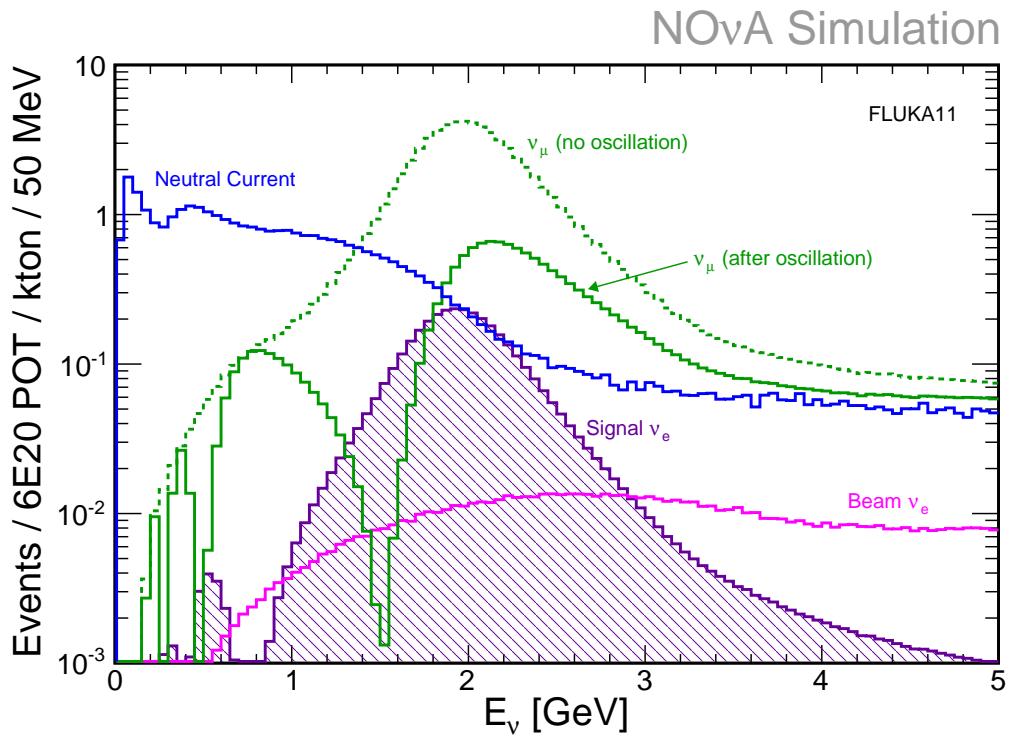


Figure 4.1: Simulated spectrum of observed FD events

The observed neutrino event spectrum in NOvA is the convolution of NuMI beam flux and neutrino interaction cross section. The spectra here show the predicted rate of neutrino interactions in NOvA as a function of interaction energy. Depicted in green is the ν_μ CC interaction spectrum; the dashed line is the prediction in the absence of neutrino oscillation, while the solid line includes the effect. The purple distribution shows the appeared ν_e CC spectrum, while the magenta distribution shows the intrinsic ν_e component of the beam. NC events, shown in blue, are binned in terms of their visible energy (excluding the energy of the outgoing neutrino) in order to represent the energy which would be deposited in the detector.

Nuclear structure is at the core of neutrino interaction simulation, for which GENIE uses the Relativistic Fermi Gas model of Bodek and Ritchie [56]. At low energies, this gives a reasonable approximation of nuclear binding energies and struck nucleon momentum. At higher energies, it provides a model for nuclear shadowing and other effects.

A principal component in neutrino event generation is interaction cross sections. In order to generate events across such a wide range of energies, GENIE must stitch together a variety of cross section models. GENIE calculates a total cross section by integrating over the differential cross sections from the various models. The total cross section is combined with the flux (from FLUKA and FLUGG simulations in this case) and detector geometry to produce the overall energy spectrum of interacting neutrinos. Since this step is relatively computationally expensive, GENIE computes the energy spectra as a separate process and stores the interpolated result based upon cubic splines [57]. For each generated event, the energy splines are sampled to select the interaction energy, but differential cross sections are used to sample the remaining kinematics.

While GENIE employs several cross section models, only a few are important for this work. The selected ν_μ CC events in NOvA are dominated by just three interaction modes: Quasi-Elastic scattering (QE), baryon resonance production (RES) and deep-inelastic scattering (DIS). QE interactions can be interpreted as a neutrino scattering off a single nucleon, RES interactions involve an excitation of the nucleus by the neutrino, and DIS represented as an interaction with the entire nucleus which can induce complete nuclear disintegration. QE events dominate the low energy tail of the energy spectrum, while DIS events account for most of the high energy tail.

The QE model in GENIE is an implementation of the Llewellyn-Smith model [31] with a dipole form factor. In that implementation, the axial vector mass is the sole free parameter. RES events are generated using from the Rein-Segal model [32], which employs the Feynman-Kislinger-Ravndal (FKR) model of baryon resonances [58]. FKR parameterizes resonance wave functions as excited states of a 3-quark system in a relativistic harmonic oscillator potential. The GENIE implementation of Rein-Segal neglects interference between resonances as well as outgoing lepton masses. Similar to the QE model, this RES implementation leaves the axial vector mass. For DIS interactions, GENIE uses a leading order model with corrections from Bodek and Yang [59] to better

describe the low Q^2 region. This crossover energy region where DIS interactions begin to dominate is of principal importance to NOvA since it accounts for most of the events on the high side of the energy peak. GENIE calculates DIS cross sections at the parton level, or in other words, the cross section involves all sea and valence quarks. This implementation of DIS involves many free parameters, although most of them have small effects on the ν_μ CC event selection in NOvA.

Free parameters in GENIE will be discussed further in chapter 10. For now, it is important to note that GENIE includes a reweighting feature for all free parameters in the models it includes. For any given event, a weight can be determined by calculating the cross section with a modification of one of the free parameters. The weights are calculated simply by taking the ratio of this modified cross section to the original cross section.

4.3 Geant4– Interaction Product Propagation

Geant4 [51] is a popular software suite for simulating the passage of particles through matter. In the NOvA simulation chain, GENIE produces daughter particles leaving the nucleus, then Geant4 simulates the passage of those particles through matter. Geant4 is thus responsible for the energy deposition, interaction and decay of those particles within the detector and surrounding materials.

This propagation relies on a detailed representation of the detector geometry which describes the position of assorted materials. The locations of the blocks have been taken from the as-built locations measured by a laser scan system, although it does not include the precisely measured tilt and stager on a plane-by-plane basis. Certain features are impossible or difficult to measure and incorporate into the geometry, for instance the location of the fiber within the cells and glue voids between modules. The geometry also includes materials surrounding the detector. These materials include the steel support structure around the detectors, the rock and concrete comprising the detector halls, as well as the overburden above the FD.

The models which govern interaction, energy deposition and decay of particles in GEANT is highly configurable through so-called *physics lists*. NOvA simulation uses the QGSP_BERT_HP physics list by default; although other physics lists have been

used to study systematic effects, namely FTF_BIC, QGSP_BIC_HP and QGSC_BERT. QGS refers to the quark-gluon string model for hadronic interactions above 20 GeV [60]. QGSC uses the Chiral Invariant Phase Space (CHIPS) model for nuclear de-excitation [61]. QGSP, on the other hand, uses the G4Precompound model for nuclear de-excitation. BERT implies that the Bertini cascade model is used for hadronic interactions below 10 GeV [62, 63]. The low energy parametrized (LEP) model is used for intermediate energies in all cases. BIC refers to the Binary cascade model, which can be used for interactions below 10 GeV in place of BERT [64]. FTF is uses the FRITOF description of string excitation and fragmentation to model interactions in the high energy regime instead of the quark-gluon string model [65]. HP represents the Geant4 high precision neutron simulation.

The Geant4 simulation process propagates particles by stepping them through their trajectories. For each particle at each step, the software selects the next action from a variety of possibilities, e.g. decay, interact with various components of the medium, or simply move forward and deposit some energy. Geant4 employs an adaptive stepping procedure which takes larger steps when appropriate in order to improve computing performance. The states of the particles at each of these steps forms the output of Geant4 which is utilized within the NOvA simulation chain. For each particle, a list is recorded which stores the position and energy deposition for each of the steps.

From the Geant4 particle trajectories, the energy depositions are converted into light output in each cell. The energy depositions of each particle in each cell are summed to produce a fiber and liquid scintillator (FLS) hit. The units of these FLS hits are arbitrary, but they do take into account an important scintillator effect known as Birks suppression [66]. This effect is must be handled before the energy depositions are summed across steps in order to correctly handle rapid changes in particle energy deposition, particularly at the end of their trajectories.

Birks suppression is caused by a local saturation of scintillation centers in the neighborhood of a swift charged particle. In other words, the light output per unit of deposited energy is diminished as the deposited energy increased since there are a finite number of scintillation centers in the vicinity of the particle. Without Birks suppression, one might expect the light output per unit length, $\frac{dL}{dX}$ of the scintillator to be directly

proportional to the energy deposition per unit length, $\frac{dE}{dX}$:

$$\frac{dL}{dX} = L_0 \frac{dE}{dX}. \quad (4.1)$$

Here, L_0 is a constant which depends on the scintillator. Birks law modifies this expression to include the suppression of light output as the energy deposition increases:

$$\frac{dL}{dX} = L_0 \frac{\frac{dE}{dX}}{1 + k_B \frac{dE}{dX}}. \quad (4.2)$$

Birks constant, k_B , is the coefficient which governs the suppression and is also dependent on the type of scintillator. The form used by NOvA includes a correction introduced by Chou [67] which adds a third parameter, k_C , and a quadratic term:

$$\frac{dL}{dX} = L_0 \frac{\frac{dE}{dX}}{1 + k_B \frac{dE}{dX} + k_C \frac{(dE)^2}{dX}}. \quad (4.3)$$

This leaves three parameters to be tuned in the simulation, L_0 , k_B , and k_C . L_0 can simply be absorbed into the photon collection step, to be discussed in section 4.4. The Birks and Chou parameters, k_B and k_C , have been tuned using muon and protons from selected ν_μ CC QE interactions in the ND. Here, the data and MC distributions of energy deposited per plane were used to produce a χ^2 surface in the space of k_B and k_C . The optimal values were chosen as the minimum point in that surface.

4.4 Photon Collection and Attenuation

The next step in the simulation chain is to convert light depositions into discrete photons which arrive at the APD. Since NOvA cells are closed, it is impossible to make direct measurements of the light collection and transport efficiency *in situ*. To overcome this lack of information, a toy ray-tracing model has been used to produce light collection templates. Some basic assumptions about the scintillator, fiber and cells are used in this model. The characteristic scintillator emission time is assumed to be 9 ns. Bench measurements specify a reflectivity of 87.7% for the cell walls [52] and an index of refraction of 1.46 for the mineral oil [48]. Photons are captured in fiber with a

characteristic length of 30.66 cm, which is tuned based on studies of cosmic ray muon attenuation curves to provide the best data/MC agreement. The ray tracing model produces a two dimensional collection rate distribution as a function of energy deposition and position along the fiber relative to where the energy was deposited. From this template distribution, the number of photons collected is modeled for each particle energy deposition produced by Geant4. Since the ends of the cells are not reflective, the template distributions are truncated at the point where the cell ends. An overall scale factor for the number of photons produced per unit of deposited energy is a free parameter in the simulation.

Once the number of collected photons is determined, the next step is propagate each of those photons through the fiber. The collected photons are divided in two to be sent in either direction along the fiber. An attenuation curve is used to determine the mean number of photons which would arrive as a function of the position of capture. Measurements of the attenuation length have been taken from the quality control data generated at the time of module construction. Finally, a Poisson smearing and 85% quantum efficiency is applied to produce the final number of photons which arrive at the APD.

4.5 Electronic Readout

Ultimately, simulation of cell hits recorded by the detector is based on a digitization of the photo-electron (PE) response of the APD. This digitization is broken down into many steps. First, analog traces are created for each cell which had photons arriving at the APD. The dual correlated sampling (DCS) algorithm is then applied to the traces in order to trigger hits in the same way that they would be in real data.

The analog traces for each APD pixel are represented as a series of analog-to-digital converter (ADC) samples spaced at 62.5 ns. Each sample is determined by scaling the number of PE relative to the PE value which is known to saturate the ADC scale from bench measurements. That ADC value is then shaped corresponding to the rise time, R , and fall time, F , of a CR-RC amplifier:

$$f(t) = \frac{F}{FR} \left(e^{-(t-t_0)/F} - e^{-(t-t_0)/R} \right), \quad (4.4)$$

where $t - t_0$ corresponds to the time between the PE response and the ADC sample in question.

Another effect, affectionately dubbed “APD sag”, is handled in the step where ADC traces are generated. Bench measurements have shown that there is a negative crosstalk between pixels on a given APD. In other words, charge deposited in one pixel is known to subtract a small amount from the charge recorded in all of other pixels in the same APD. The scale of the negative crosstalk is 1.86% of the total charge deposited across all pixels on the same APD. This effect is simulated by producing trace which sums over all pixels on each APD, scaling it down to 1.86% of the total, then subtracting that shape from each pixel-by-pixel ADC trace. Prior to the subtraction, each ADC trace is scaled up by 1.86% to remove “self-sag”, or the sag which is produced by that particular pixel. Since the effect accumulates over pixels, the effect on calorimetric energy can be larger than 1.86%, but is sensitive to the topological distribution of particle or shower in question. Empirically, the effect on calorimetric energy has been observed at the scale 4-8% depending on the event topologies.

Noise is added to the ADC traces by a Gaussian-Markov model representing the current current and voltage noise in the amplifier circuit. The final trace is then sampled by the DCS algorithm, which triggers a hit to be recorded if the the difference a given ADC sample differs by more than a fixed threshold from value recorded three samples earlier. For real data, the thresholds are determined for each channel based on the RMS of its ADC distribution. For simulation, however, the thresholds are drawn randomly for each channel from a distribution based on data.

It is interesting to note that this method does not provide a definite link between simulated hits and simulated particles. Energy depositions of particles produce photons which are propagated to APD to produce a trace. That trace, however, can also include contributions from other particles, hence the broken link. Hits thus can only be approximately connected to particular energy depositions by considering the time of the energy deposition relative to the hit in question.

Parameter	Value
$\Delta m_{32}^2 [10^{-3} \text{ eV}^2]$	2.37
$\Delta m_{21}^2 [10^{-5} \text{ eV}^2]$	7.53
$\sin^2(2\theta_{12})$	0.846
$\sin^2(2\theta_{13})$	0.086
$\sin^2(2\theta_{23})$	0.5
δ_{CP}/π	0
L [km]	810
$\rho [\text{g/cm}^3]$	2.84
Hierarchy Assumption	Normal

Table 4.1: Oscillation parameters used to weight MC prediction

The values of Δm_{21}^2 and $\sin^2(2\theta_{12})$ match the values from the Particle Data Group. The value of $\sin^2(2\theta_{13})$ was chosen to match [41]. For simplicity, δ_{CP} was set to 0, although it has little impact on ν_μ CC disappearance. L is the distance the neutrinos travel between the ND and FD. The density of the earth, ρ , is necessary for the matter effects discussed in Section 2.2.3. The value for ρ was taken from [69]. The hierarchy assumption used makes no difference to ν_μ CC disappearance; for simplicity, the normal assumption was used.

4.6 Oscillated Spectra

Several parts of this work will rely on presenting an *oscillated spectrum* using the FD MC prediction. In all of those cases, ν events are weighted by an oscillation probability determined based on the interaction energy recorded in MC truth ¹. Note that the parameters Δm_{32}^2 and $\sin^2(2\theta_{23})$ are included in the oscillation fit for this analysis, where they are free to vary rather than fixed.

The oscillation parameters used in this weighting are taken from [42, 41, 68, 69]. A complete listing of the parameters used can be seen in Table 4.1.

¹In MC simulation, it is possible to record detailed information regarding physics activity and detector response. This collection of recorded information is colloquially called *MC truth*. In NOvA simulation, stored information includes, but is not limited to, interaction kinematics from GENIE, step-by-step particle trajectory and energy depositions from Geant4, and photoelectron signals from the detector simulation.

Chapter 5

Event Reconstruction

In order to extract useful physics information from the NOvA detector, the raw data (cell hits) must be processed into higher level forms. This processing step is known as reconstruction. For this analysis, the primary goals of reconstruction are to identify ν_μ charged-current (CC) interactions and estimate the energy of the neutrinos involved in the interactions. Identification of ν_μ CC interactions is really a task of rejecting backgrounds induced by both cosmic rays and the NuMI beam. The cosmic ray background dominantly composed of down-going muons, although other particles can also be present. Backgrounds in the NuMI beam include both neutral current (NC) interactions and CC interactions from ν_e or ν_τ .

Traditional reconstruction efforts involve detection of lines and other shapes in raw data. Classification is then a process of extracting manually engineered features from those events which discriminate between signal and background. This analysis uses machine learned features in a convolutional neural network to replace manual feature engineering. The approach presented here differs in that the hits themselves are presented to the neural network as two separate images, corresponding to x (vertical) and y (horizontal) views. The goal with this approach is to sidestep pathological failures incurred in each reconstruction and feature engineering step, thus allowing more of those pathological events to be more correctly classified.

The NOvA *event display* provides a visualization of the raw detector readout which serves as the input to reconstruction. An example event display is shown in Figure 5.1. The data corresponds to 550 μs of Far Detector (FD) readout, with hits colored

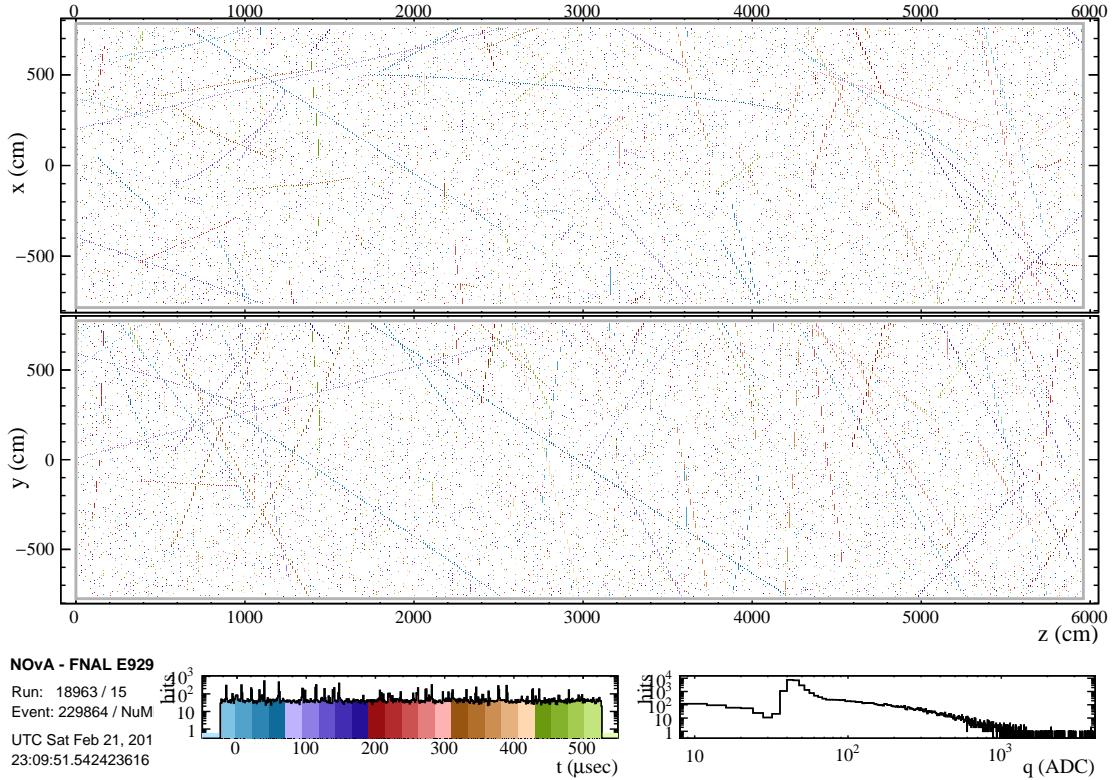


Figure 5.1: An example NOvA event display

The top projection is an x vs. z view, the bottom is y vs. z . Hits are colored by the time at which they were recorded relative to the start of the readout window; the color scale is visible in the bottom center pane.

according to their time recorded relative to the start of the readout window. Visible activity displayed falls into two primary groups: randomly distributed electronic noise and correlated activity from cosmic rays. Since cosmic rays cross the detector quickly, hits along each track are displayed as a uniform color.

5.1 Slicing

Reconstructing events in NOvA first requires resolving activity within the continuous readout windows. In other words, the continuous readout of noise and physics activity

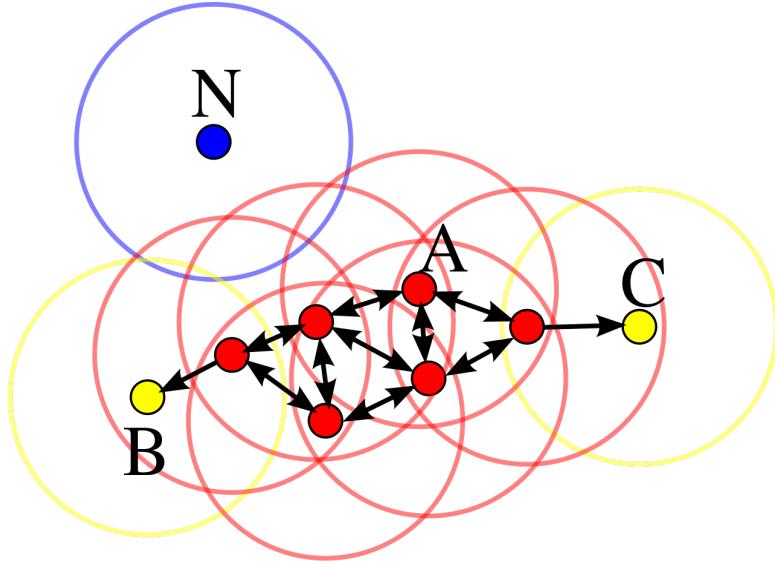


Figure 5.2: Depiction of clustering with the DBSCAN algorithm

The open circles around each point are drawn with a radius to represent the neighbor distance threshold. The red points, for instance point A, are considered to be *core* points since they have as many or more neighbors than the minimum required to form a cluster. The yellow points, including B and C, are called *edge* points since they have fewer neighbors than the minimum required to make a cluster, but included in the cluster with the red points since they are below the neighbor distance threshold. The blue point is not part of the cluster since its distance from all other points exceeds the neighbor threshold. Image reproduced from [70].

must be separated into groups of hits which correspond to individual particle interactions. This task is referred to as *slicing*.

Slicing in NOvA is achieved through an implementation of the DBSCAN algorithm [71, 72]. The algorithm evaluates the distance between every pair of points based on a distance metric which suits the problem at hand. If the distance between a pair of points is below a threshold, ϵ , the points are added to the same cluster. Clusters can be disregarded if they do not include a minimum number of hits, which can be optimized based on the amount of noise in the data. Boundaries between clusters are effectively formed in low density regions. A depiction of how points are clustered can be seen in Figure 5.2.

The distance metric used for slicing in NOvA is based on the time, t , and spatial

position, \vec{r} , of hits

$$L = \left(\frac{\Delta t - \Delta \vec{r}/c}{T} \right)^2 + \left(\frac{\Delta \vec{r}}{D} \right)^2, \quad (5.1)$$

where c is the speed of light. The parameters T and D represent length scales which dictate the relative weight of separation in time and space. T is taken to be the timing resolution of the two hits in the pair summed in quadrature. Timing resolution for hits is estimated based on the pulse height. In the FD, the timing resolution is roughly 10 ns, while the ND timing resolution is roughly 5 ns [73]. D is configurable and set to 100 cm for the FD and 75 cm for the ND [72].

The slicing algorithm for NOvA has been tuned such that it provides good separation between cosmic rays while leaving entire neutrino interactions intact. Certain delayed activity, such as Michel electrons (from muon decay) and neutron activity is prone to being excluded from a slice. Generally, however, a slice in NOvA data is treated as a potential cosmic ray or entire neutrino interaction with all daughter particles.

5.2 Tracking

Once interactions have been resolved, individual particle trajectories can be isolated. Cosmic ray interactions are typically characterized by a single particle trajectory, but neutrino interactions can involve several particles with distinct trajectories. Isolating and fitting those trajectories is commonly referred to as *tracking*.

This analysis applies the results of two different tracking algorithms. The first is a simple, least-squares based approach called *CosmicTrack*; the second, *KalmanTrack*, is based on Kalman filtering [74] and optimized to identify muons in ν_μ CC interactions [75].

5.2.1 CosmicTrack – Least-squares Regression

CosmicTrack was designed to quickly reconstruct the long, mostly vertical tracks which are produced by the vast majority of cosmic rays which are visible in NOvA. The approach is to simply fit lines to the activity each view (x and y) of the detector. Fitting is restricted to hits within a particular slice, and precisely one track is reconstructed for each slice. Hits which are far from the track in each view are iteratively removed

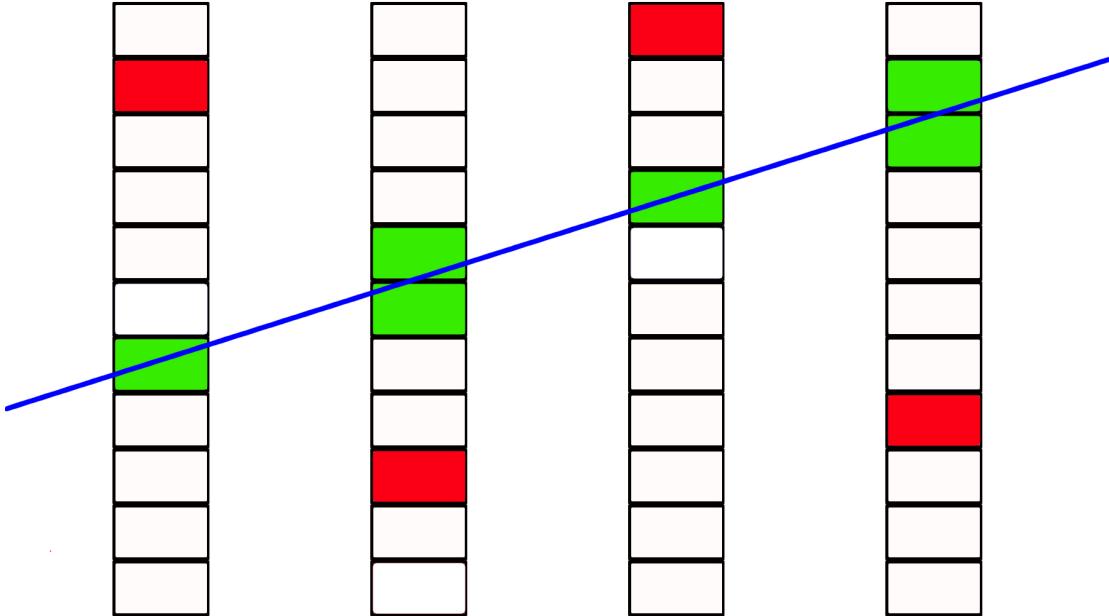


Figure 5.3: Illustration of the tracking concept in NOvA
The *CosmicTrack* algorithm forms tracks based on a linear regression approach. The *KalmanTrack* algorithm, inspired by the Kalman filtering technique, expands tracks based on the χ^2 of a piecewise linear fit.

to improve robustness against noise hits. The tracks found in either view are then interleaved in a plane-by-plane fashion in order to form 3D tracks from the separate 2D tracks. CosmicTrack assumes particles to be downgoing, so the start point of the track is always deemed to be the highest position of the track within the detector.

5.2.2 KalmanTrack – Kalman Filtering

KalmanTrack [75] is an algorithm inspired by the Kalman filtering technique [74]. In a similar fashion to CosmicTrack, tracks are first found separately for vertical (X) and horizontal (Y) views, then merged. Track seeds are formed for pairs of hits which fall within a three plane window in the given view. Starting with the most downstream (in the plane dimension, running along the beam direction) seed, the seeds are expanded

upstream As the track is expanded, the hit positions are fit with piecewise linear segments to allow deviations resulting from multiple scattering. A hit is added to the track if it does not change the χ^2 of the track fit by more than eight units. Tracks are terminated when they cross more than three planes without picking up any additional hits. If the track traverses a cell with a dead readout channel, that plane is not included in the missed plane total. The procedure is then repeated in the other direction in order to pick up any hits which may have been missed. Since tracks emanating from a neutrino interaction vertex are generally directed downstream; thus working from the downstream side allows the longest track to be found first. In a vast majority of cases, the muon track in a ν_μ CC interaction is the longest track. As such, the KalmanTrack procedure is well suited for ν_μ CC interactions.

Once tracks have been found in either view, they must be stitched together to form 3D tracks. Tracks are paired based on a score which describes how they coincide in the Z -dimension, i.e. plane extent. For a each track in the X -view with start z_{x1} and end z_{x2} , a score is calculated for each overlapping track in the Y -view with start z_{y1} and end z_{y2} :

$$\text{score} = \frac{|z_{x1} - z_{y1}| + |z_{x2} - z_{y2}|}{\min(z_{x2}, z_{y2}) - \max(z_{x1}, z_{y1})}. \quad (5.2)$$

The quantity in the numerator describes the proximity of the start and end points, while the denominator represents the extent of the track overlap. The pair of tracks with the lowest score is matched and removed from the pool of 2D tracks, then the pair with the next lowest score is matched, and so on. Unmatched 2D tracks are also recorded and labeled as such for downstream reconstruction and analysis.

5.3 Calibration

An important input to many reconstruction efforts is an accurate estimate of the energy deposited in a cell. Calibration procedures aim to provide an the estimation of energy deposition based on the recorded ADC value and position of a particular cell hit. The NOvA calibration accounts for three primary effects: attenuation of light in the wavelength shifting fiber, cell-to-cell variations, and conversion conversion from ADC scale to physical energy units.

5.3.1 Attenuation and Cell-to-cell Correction Procedure

The attenuation correction is based upon cosmic ray muons since they are abundant and their energy deposition properties are well understood. A reference sample of cell hits is selected from muon tracks which satisfy a few simple characteristics. First, tracks must touch two detector faces, indicating that they traversed the detector rather than stopping inside. The tracks are required to cross at least ten planes to ensure that path length (distance traversed by the muon within each cell) can accurately be estimated. Hits on selected tracks are only used for calibration if they satisfy the so-called tri-cell criterion; that is, that there are also hits present in adjacent planes. The tri-cell criterion is schematically depicted in Figure 5.4. This requirement helps ensure that a reasonable path length estimate can be obtained. Certain cells fail to produce enough hits to allow the subsequent attenuation fits to be obtained, for instance cells on the edges of the detector and those adjacent to dead cells. Such cells are calibrated with by enriching the sample with hits which have neighbors in adjacent planes.

For each reference hit, the ADC value is converted to PE based on a scale factor. The path length and position along the cell (i.e. distance from the APD readout) are both estimated from the muon track trajectory. A 2D histogram of the ratio of PE to path length vs. position along cell is constructed for each cell. A fit is to the mean PE per path length response function of position is used to characterize each cell. The functional form for the middle part of the cells is a sum of two exponentials, one corresponding the short trip to the APD and the other for the long way to the far end of the cell and back around the fiber loop.

$$y(W) = C + \left(\exp\left(\frac{W}{X}\right) + \exp\left(-\frac{L+W}{X}\right) \right) \quad (5.3)$$

Above, W is the position along the cell, X is the fiber attenuation length, and L is the length of the cell. The attenuation length as well as constants C and A are free parameters in the fit. At the end of cells, a diminished response known as *roll-off* has been observed, presumably due to light being absorbed at the ends rather than in the

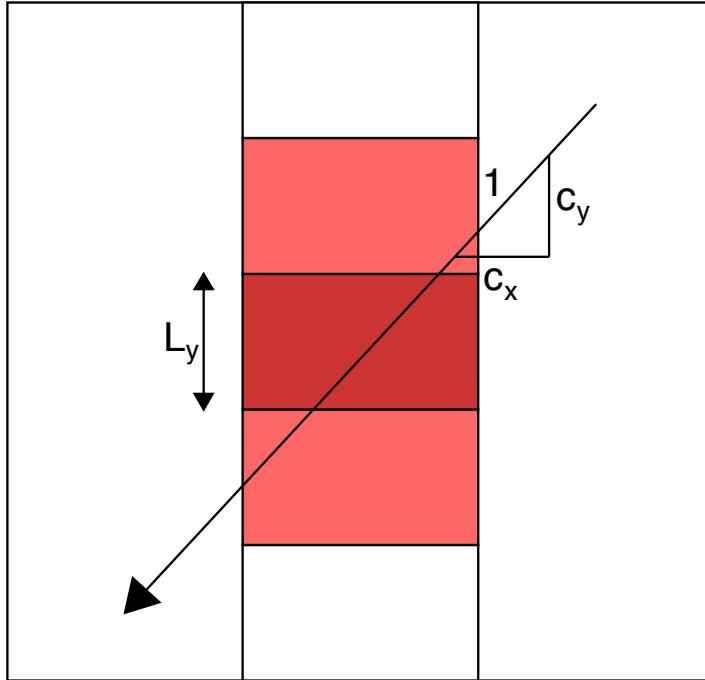


Figure 5.4: Depiction of a tri-cell hit used for calibration
The presence of activity in the two adjacent cells helps ensure an accurate estimate of path length traversed by a muon through a cell.

fiber. The deficit of light is empirically well described by an x^4 form

$$y = \begin{cases} 1 - \alpha_R(x - x_R)^4 & x > +x_R \\ 1 & \text{otherwise} \\ 1 - \alpha_L(x - x_L)^4 & x < -x_L \end{cases} \quad (5.4)$$

where free parameters x_R and x_L indicate the start of the roll-off region on either side and α_R, α_L determine the scale. These fits are often not robust, however, so they are not used. Alternatively, an interpolation method called LOWESS [76] is used to correct the residuals from the exponential form. In addition to picking up the roll-off behavior, the

LOWESS fits also serve to correct other significant deviations observed in the data far from the ends of the cells. These deviations are believed to be caused by any peculiar positioning of the fiber within a cell which could affect light collection. The curve produced by LOWESS at any point is the weighted mean of the deviations,

$$w_i = \left(1 - \left| \frac{x - x_i}{\sigma} \right|^3 \right)^3 \quad (5.5)$$

where the smoothing length scale σ is 30cm. Examples of the final fits with LOWESS corrections can be seen in figure 5.5.

The attenuation fits determine the mean response to cosmic ray muons is can be determined for a hit at any position in the detector. Thus, the calibrated response for any hit is the ratio of its PE to result of the fit at that position. At this point, the energy in absolute units is still unknown, but the effects of attenuation and cell-by-cell variations have been mitigated.

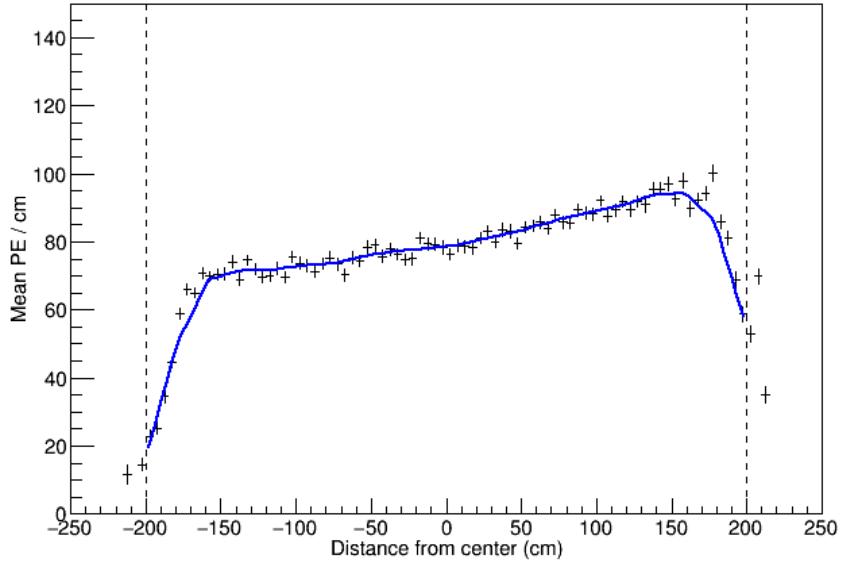
5.3.2 Absolute Energy Calibration Procedure

The absolute energy calibration is responsible for converting the corrected PE value described in the previous section into a value which corresponds to meaningful physical units. Muon tracks which stop inside the detector form a reference sample with well understood energy deposition characteristics. Near the end of particle tracks, the energy deposition per unit length (dE/dX) is well described by the Bethe-Bloch formula [42]. The formula predicts a stable minimum in dE/dX near the track end, which for muons in organic materials covers a wide range of several hundred cm. This stable region is demonstrated in Figure 5.6. For NOvA, the region between 100 and 200 cm from the track endpoint is used to determine the absolute energy scale.

5.3.3 Calibrated Energy Depositions for Hits

The calibration procedure provides an estimation of the actual deposited energy as a function of recorded ADC and the location of the hit within the detector. Accurately locating a hit in the X -view or Y -view, however, requires an estimation of the unmeasured coordinate from the other view. For a hit in the X -view, the unknown Y position (and

ND cosmic data - plane 151 (vertical), cell 51



FD cosmic data - plane 419 (vertical), cell 219

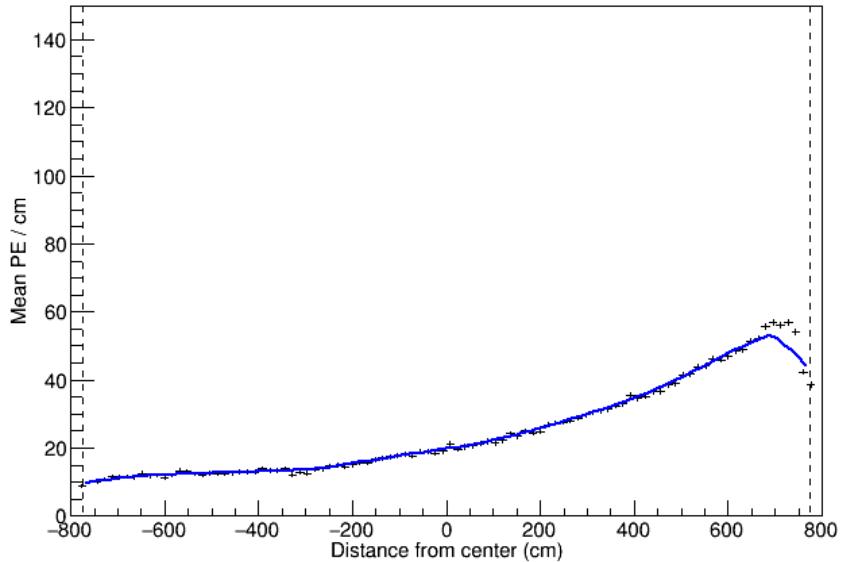


Figure 5.5: Example attenuation fits

Above are good Near (left) and Far (right) detector fits, showing attenuation fits with LOWESS corrections. The fits account for attenuation in the WLS fiber and effectively remove cell-to-cell variations. The LOWESS [76] smoothing procedure is used to correct for residuals in the fit.

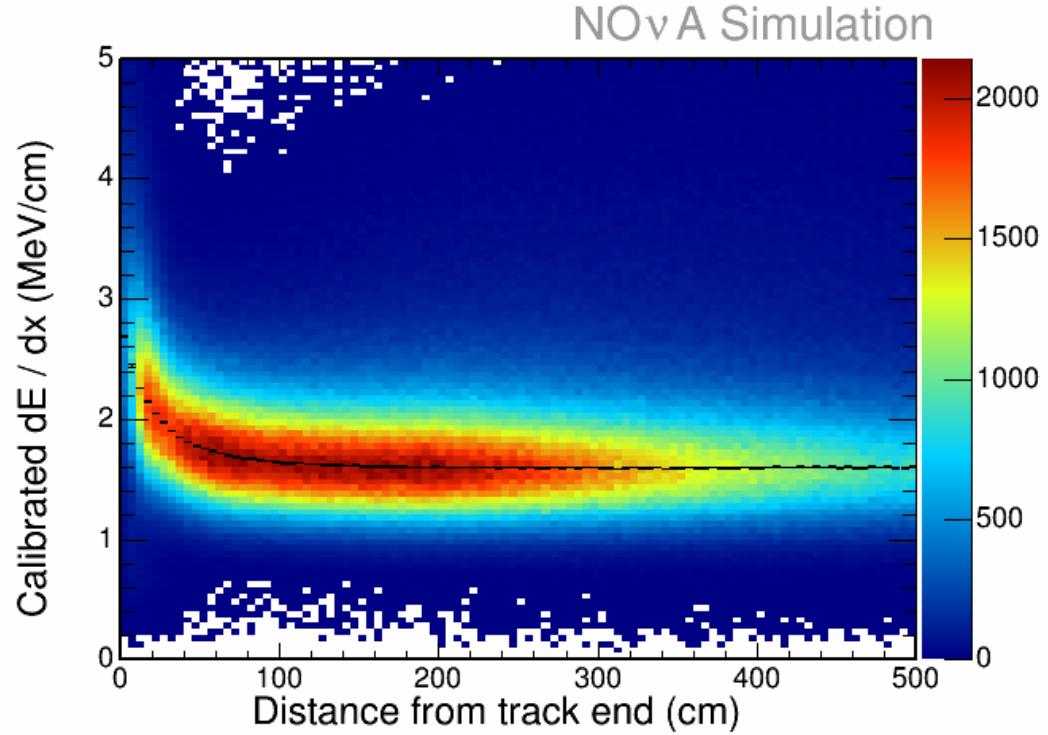


Figure 5.6: 2D histogram depicting deposited energy per path length (dE/dX). Each entry in the histogram above corresponds to the dE/dX of a cell hit after attenuation and cell-to-cell correction. The color scale indicates the number of measurements in each bin. The rise in energy at the end of muon track (left) is well described by the Bethe-Bloch formula [42].

vice versa) of the hit is essential in applying the attenuation correction. Estimation of the unmeasured coordinate must rely on reconstruction, namely slices and tracks. For slices, the unmeasured coordinate is determined to be the mean position of all hits in the slice from the other view. For tracks, the unmeasured coordinate can be inferred by interpolating the track trajectory through the plane in question. Tracks, as a result, allow for much more precisely calibrated energy deposition.

5.4 Muon Identification

To estimate the energy of a ν_μ CC interaction, it is beneficial to identify the track produced by the outgoing muon [77]. Muon identification is applied to the collection of tracks which are output from the *KalmanTrack* algorithm. Muon tracks are identified using a k-Nearest Neighbor (kNN) [78] classification algorithm with four features extracted from tracks as input [75]. The kNN algorithm populates an n dimensional hyperspace with labeled examples (tracks in this case, labeled as muon or not-muon) where n corresponds to the number of input features. For each test example to be evaluated, some number, k , of nearest neighbors in the hyperspace are located. Proximity is ascertained using the L_2 norm in \mathbb{R}_n . The output of the algorithm is the fraction of the k neighbors which belong to the desired class, in this case tracks which were truly produced by muons. For muon identification in NOvA, $k = 80$. The library of examples which populates the hyperspace is taken from NOvA simulation.

The inputs which form the kNN hyperspace are track length, scattering log-likelihood, dE/dX (energy deposition per unit path length) log-likelihood, and non-hadronic plane fraction [75]. The track length is path length traversed by the track taken from summing between interpolated trajectory points. Log-likelihoods for both scattering and dE/dX are constructed using probability distributions taken from simulation for both muons and charged pions. The non-hadronic plane fraction is the fraction of track which remains after an algorithm has identified and removed a region of the track which is contaminated from overlap with the hadronic shower [79].

5.5 Energy Estimation

Estimation of the energy of selected ν_μ CC interactions is crucial to the measurement of oscillation parameters θ_{23} and Δm_{32}^2 ; the distribution of estimated energy is fit in order to produce likelihood contours. The fitting process will be described in detail in the 9. For this analysis, the neutrino energy is estimated by splitting apart activity from the identified track and hadronic shower, estimating the energy of those components, then summing the two components [77]. The energy estimation procedure is tuned on events with muon identification kNN output greater than 0.75.

The muon track energy is obtained from a fit to the 2D distribution of muon track

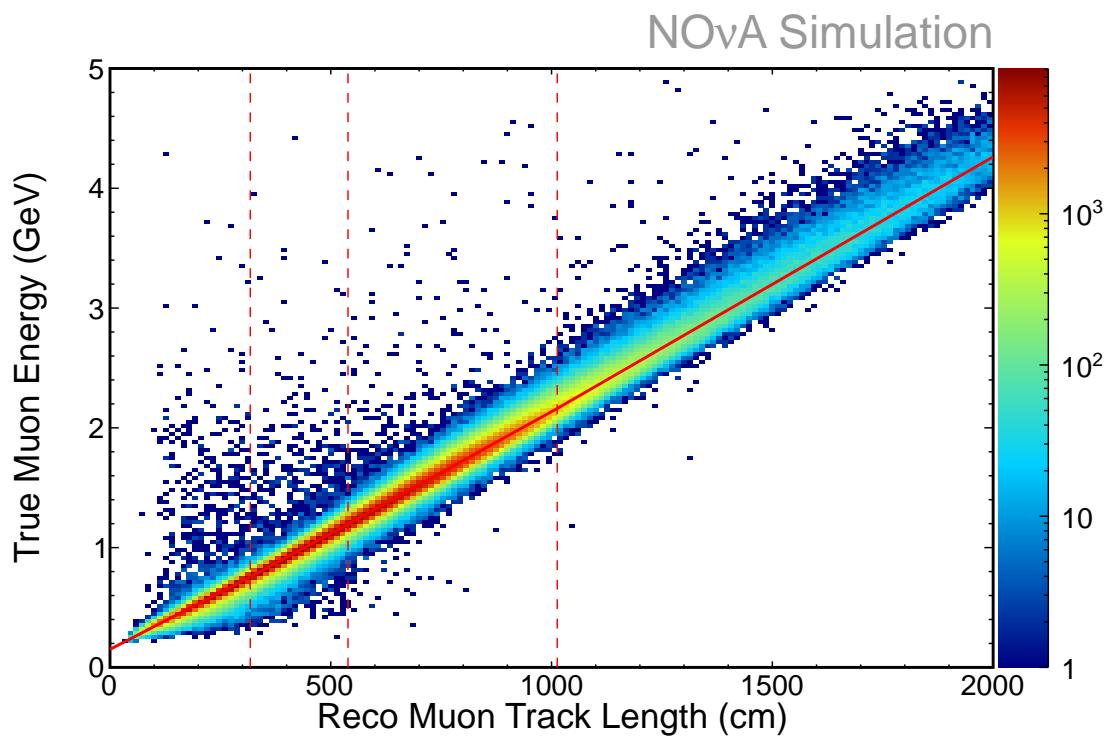


Figure 5.7: Muon energy fit

For muon energy, the 2D distribution of true muon energy and reconstructed track length is fit. The fit function is a piecewise linear spline with four segments. The fit is drawn as a solid red line, and the spline join points illustrated by the vertical dashed red line.

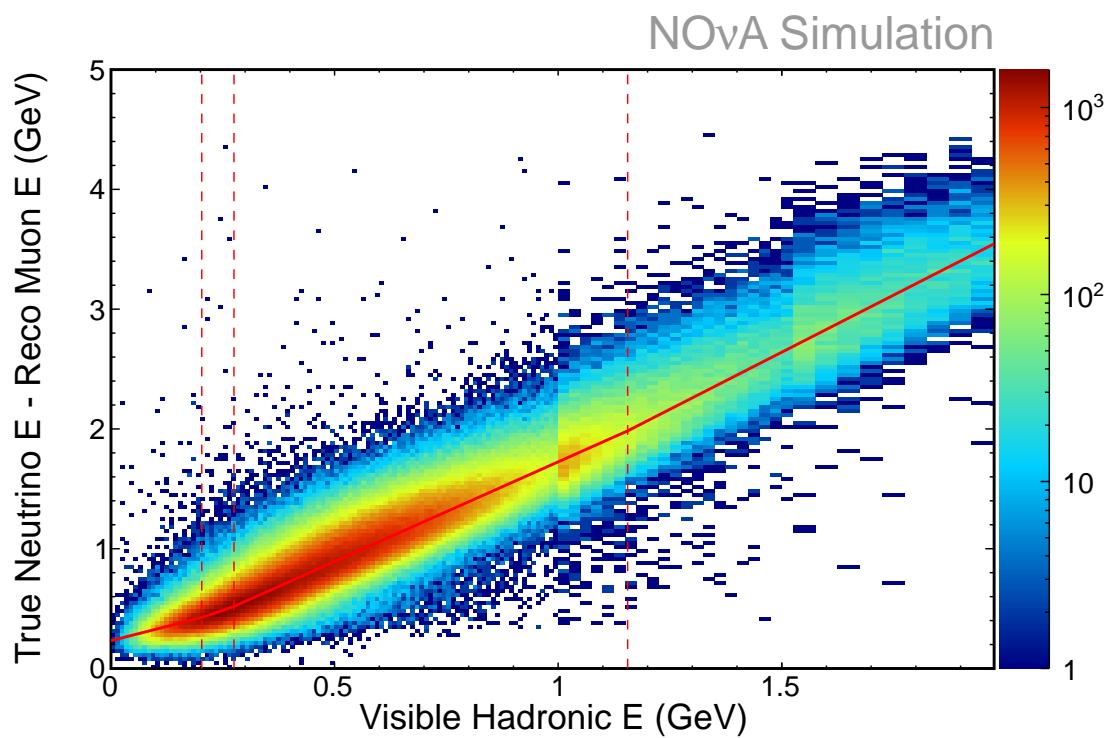


Figure 5.8: Hadronic energy fit

For hadronic energy, the 2D distribution of calorimetric energy and true hadronic energy is fit. In both cases, the fit function is a piecewise linear spline with four segments. The fit is drawn as a solid red line, and the spline join points illustrated by the vertical dashed red line. The vertical discontinuities around 1 GeV and 1.5 GeV are the result of a change in the bin horizontal bin width at those points.

length and true muon energy from simulation. For muons in NOvA, the track length is more tightly correlated with true muon energy than than the sum of calibrated energy depositions. The functional form for the fit is piecewise linear spline with four segments, as seen in Figure 5.7. The offset, four slopes and three join points are all free parameters in the fit.

The energy of the hadronic shower is estimated from any hits which are not on the muon track. The calorimetric energy (sum of hit energy depositions) is obtained using the slice method for calibration, namely using average position of all hits from the other view. Hadronic energy which overlaps with the muon track [79] is also added to the calorimetric energy sum. A fit to 2D distribution of calorimetric energy and true hadronic energy from simulation is used to determine the hadronic energy contribution. The hadronic energy is taken as the true neutrino energy minus the muon energy, effectively absorbing any contribution from the Fermi momentum of the nucleus involved in the neutrino interaction. Fitting in this manner also accounts for the non-active material (PVC and glue) in the detector. Just as in the muon track case, the functional form for the fit is piecewise linear spline with four segments, as seen in Figure 5.8. The offset, four slope values and three join points are all free parameters in the fit.

Chapter 6

Neural Networks

The field of machine learning is concerned with algorithms which can learn to make predictions from data. Typically this involves some form of multivariate function approximation [80]. In other words, given some set of input variables, predict some set of output variables. If the goal is to predict one or more outputs of a continuous function, the task is referred to as regression. The task of separating examples into groups is called classification.

It is also common to separate machine learning approaches into two categories: supervised and unsupervised [81]. The supervised approach involves training an algorithm using a collection of examples for which the function output is known. In unsupervised learning, algorithms can work to extract meaningful output from data where the function output is not known.

The approach presented in this dissertation applies a supervised learning strategy for classification of signal events.

6.1 Artificial Neural Networks

An artificial neural network is a common tool used in classification and regression. This class of algorithms draws inspiration from biological neurons. Nervous systems of organisms are built up from a repeated structure of cells, called *neurons*, which are connected to each other in order to transmit signals. An example of a neuron can be seen in Figure 6.1. The base of a neuron is formed by a branching tree of *dendrites*

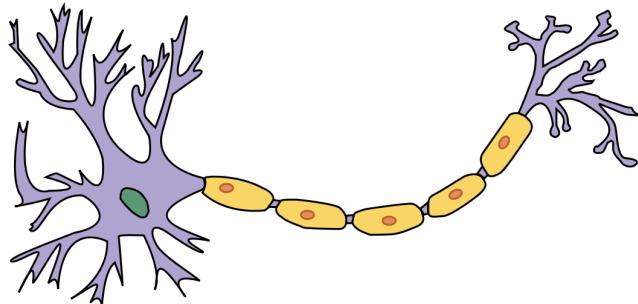


Figure 6.1: Basic rendering of a biological neuron

Signals are received through the branching dendrites to the left of the image. The cell body integrates and retransmits signals to other cells through the axon, which runs horizontally from left to right in the figure. Biological nervous systems are based upon a repeated structure of neurons to carry signals over long distances. Image reproduced from [82].

which receive signals from other neurons. Signals are integrated and amplified in the cell body, then retransmitted through a long shaft, or *axon*, to other neurons. Intelligence is achieved by connecting many neurons which produce a wide variety of responses based on given stimuli.

Artificial neural networks are designed to crudely approximate the structure of a biological nervous system [80]. The basic unit of a neural network is called a *node*. Each is a function which computes an output value from a set of inputs. Nodes are arranged into *layers*, as seen in Figure 6.2. In graph theory terms, a neural network is a *feedforward graph*, implying that information is passed forward to subsequent layers, but not backward. Input variables to the classification or regression problem are passed through *input layer*. Any number of downstream *hidden layers* can be trained to perform the classification or regression task. The predicted output is then extracted from the final *output layer*. For multivariate regression, the output is a vector of decimal values. For multi-class classification, the output is usually a vector of as many values as there are classes, each ranging between zero and one. Each value can be interpreted as how well the input represents that particular class.

Each node in a layer can receive input from all nodes in the previous layer, although more sparsely connected networks have been investigated [80]. The response to each input is characterized by a vector of weights, $\mathbf{w}^{i,j}$, for node i in layer j . Each layer produces a vector of outputs, \mathbf{a}^j , so the response in for the i th node in layer j is the inner product between the weight vector of the node and the outputs of the previous layer in addition to a bias term, $b^{i,j}$:

$$z^{i,j} = \mathbf{w}^{i,j} \circ \mathbf{a}^{j-1} + b^{i,j} \quad (6.1)$$

The node response is commonly passed through some nonlinearity, $f(x)$ to produce the output to be passed to the next layer, thus the element in \mathbf{a}^j for the i th node is:

$$a^{i,j} = f(z^{i,j}) \quad (6.2)$$

Common choices for the nonlinearities are the sigmoid function, hyperbolic tangent (tanh), rectified linear unit (ReLU) or softmax [83, 81, 80, 84].

$$\begin{aligned} f(x) &= \frac{1}{1 + e^x} && \text{(Sigmoid)} \\ f(x) &= \frac{e^x - e^{-x}}{e^x + e^{-x}} && \text{(tanh)} \\ f(x) &= \max(0, x) && \text{(ReLU)} \\ f(x) &= \frac{e^{x_i}}{\sum_k e^{x_k}} && \text{(Softmax)} \end{aligned} \quad (6.3)$$

The indicies i and k in softmax enumerate the outputs of the other nodes in the same layer. As a result, the softmax is thus normalized such that the sum of outputs from that layer is equal to one. This feature makes the softmax a popular choice for output nodes in classification tasks since it provides a convenient pseudo-probabilistic interpretation of the output values.

Sigmoid and tanh are attractive because they are bounded and differentiable¹. Rectified linear units are more analogous to the thresholded linear response of a biological

¹The importance of differentiability arises in the context of backpropagation, which is discussed in the following section

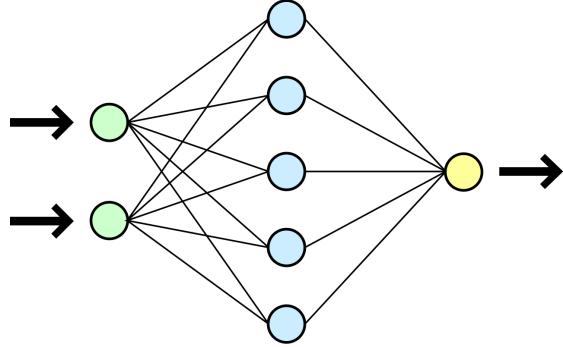


Figure 6.2: Visualization of a simple neural network

This network has an input layer (green) with two nodes, a hidden layer (blue) with five nodes, and an output layer (yellow) with one node. In a neural network, nodes in hidden and output layers receive input from the upstream layer, starting with the input layer. Image reproduced from [85].

neuron. The negative component of range of the tanh is thus poorly motivated based on biological neurons, which do not produce a negative response. Even still, tanh non-linearities have been observed to produce good results. ReLU is additionally attractive since it does not involve the exponential function, which is expensive to evaluate from a computational standpoint. Meanwhile, the unbounded derivative of ReLU can lead to problems during gradient-based training, discussed in section 6.2. This trouble can be averted by imposing an artificial bound on the derivative of the ReLU. It should also be noted that while the ReLU is not differentiable at zero, it is differentiable arbitrarily close to zero, which is sufficient for training.

The input layer to the neural network depends on the problem at hand. Basic applications of neural networks involve some number of input variables ranging from a few to a few hundred. In some cases, the inputs are a set of engineered features extracted from some raw data, while other cases input raw data directly into a network. The capacity for a network to approximate functions has been proven to increase with the number of hidden layers [80, 81]. For practical reasons which will be discussed in section 6.3, diminishing returns are typically seen with more than a few hidden layers unless special techniques are applied [32, 81].

6.2 Supervised Learning – Backpropagation

While neural networks were first proposed in the 1950s, practical applications were out of reach until a supervised learning strategy called backpropagation was introduced by Werbos in 1974 [86]. Backpropagation is a supervised learning algorithm; the network is presented with a set of training examples with a known output as an iterative process to learn optimal weight parameters. The major step forward in the backpropagation is a clever application of the chain rule to differentiable nonlinearities, thus allowing steepest descent stochastic gradient descent (SGD) to be used for training.

SGD is a common optimization technique. If there is some objective function, E , to be optimized, SGD seeks optimal weights using the gradient that function. The form of the objective function depends on the problem at hand, but it is generally constructed to quantify the error resulting in the network result relative to the desired output. Since it depends on the network output, the $E(\mathbf{w})$ is a function of the weights in the network. Over many iterations, a function can be approximately minimized by adjusting the weights opposite the gradient of the objective function at a given iteration step. This gives the update rule for the weight vector \mathbf{w}

$$\mathbf{w} := \mathbf{w} - \eta \nabla E(\mathbf{w}). \quad (6.4)$$

The parameter η is referred to as the *learning rate*, which governs the step size. Generally, η is set to some value between 0 and 1, although there is no straightforward way to determine a value which will give the quickest or optimal convergence. Large values for the learning rate may speed up convergence, but might also cause the algorithm to repeatedly step over the minimum location in the surface given by the objective function.

There are many variants of SGD which are popular. Backpropagation itself does not specify which variant of SGD should be applied, it merely provides a mechanism to calculate the partial derivative with respect to each weight parameter. At the core of backpropagation is a simple, but clever, application of the chain rule to find the derivative at an arbitrary node upstream of the output in a feedforward network.

First, the predicted output is calculated for a set of training examples are calculated in a feedforward pass. An objective function is chosen to quantify the error between the

target and predicted outputs. Minimizing the objective function through an iterative process thus trains the network to predict the target outputs. In regression problems, the objective function is commonly some mean-square error on the Euclidean distance between the target outputs and predicted outputs. A simple implementation, for a network with m outputs and p training examples is

$$E = \frac{1}{2} \sum_p \sum_m (t_{mp} - y_{mp})^2 \quad (6.5)$$

where t_{mp} and y_{mp} respectively correspond to the target and predicted output of the network. The factor of $\frac{1}{2}$ is simply present for cancellation after taking the derivative. Classification problems commonly use another form of the objective function called the cross-entropy. For output classes m , and again training examples p , target t and prediction y , the cross entropy is [80]:

$$E = \sum_p \sum_m t_{mp} \ln(y_{mp}) + (1 - t_{mp}) \ln(1 - y_{mp}) \quad (6.6)$$

Backpropagation can be applied to any feedforward network of nodes with differentiable nonlinearities. A layered structure is not required. Therefore, the process can be described by indexing w_{ij} as the weight from node j to node i . A feedforward network implies $j > i$ is not allowed. The forward pass thus defines the inner product for each node, a_i , as a function of the weights and upstream nodes

$$a_i = \sum_{j < i} w_{ij} y_j. \quad (6.7)$$

Above, y_i is the result of the nonlinearity for each node

$$y_i = f(a_i). \quad (6.8)$$

Backpropagation finds a partial derivative for each weight by summing over a batch of training examples. The overall partial derivatives for the weight to node i from weight

j is the sum of the gradients from training examples p :

$$\frac{\partial E}{\partial w_{ij}} = \sum_p \frac{\partial E_p}{\partial w_{ij}} \quad (6.9)$$

The partial derivative for a particular training example can be written in terms of the upstream node outputs a_k

$$\frac{\partial E_p}{\partial w_{ij}} = \sum_k \frac{\partial E_p}{\partial a_k} \frac{\partial a_k}{\partial w_{ij}} = \sum_k \delta_{kp} \frac{\partial a_k}{\partial w_{ij}}, \quad (6.10)$$

with δ_{kp} defined for convenience:

$$\delta_{kp} = \frac{\partial E_p}{\partial a_{kp}} = \frac{\partial E_p}{\partial y_{kp}} \frac{\partial y_{kp}}{\partial a_{kp}}. \quad (6.11)$$

The second term is just the derivative of the nonlinearity, f

$$\frac{\partial y_{kp}}{\partial a_{kp}} = f'(a_{kp}). \quad (6.12)$$

The output nodes have no dependence on upstream nodes. For the mean-square error objective function, for instance, we have:

$$\delta_{kp} = -(t_{pk} - y_{pk})f'(a_{kp}). \quad (6.13)$$

The partial derivative of the cross-entropy objective, on the other hand, is:

$$\delta_{kp} = \left(\frac{t_{pk}}{y_{pk}} - \frac{1 - t_{pk}}{1 - y_{pk}} \right) f'(a_{kp}). \quad (6.14)$$

The derivative of the objective function with respect to the weights in the hidden layers depends on downstream layers, so the chain rule must be applied. The derivative for the k th hidden node relative to the p th training example is

$$\delta_{kp} = \frac{\partial E_p}{\partial a_{kp}} = \sum_l \frac{\partial E_p}{\partial a_{lp}} \frac{\partial y_{lp}}{\partial a_{kp}}. \quad (6.15)$$

The first factor in the sum is simply δ_{lp} . The second factor is zero if node k does not

connect to node l , otherwise it simply the product of the node connection weight and the slope of the nonlinearity for that node, evaluated for that training sample:

$$\frac{\partial a_{lp}}{a_{kp}} = f'_{lp} w_{lk} \quad (6.16)$$

Thus, for hidden nodes, we have:

$$\delta_{kp} = f'_{kp} \sum_l w_{lk} \delta_{lp} \quad (6.17)$$

Backpropagation drove a major shift in the field of neural networks [80]. Prior to the innovation of backpropagation, researchers focused their time on the theoretical capacity of networks of binary hidden units. Armed with backpropagation, research could be devoted to applications, but now using differentiable hidden units trained with SGD [80].

6.3 Deep Learning

From a theoretical perspective [80], deep neural networks (those with many layers) offer increased capacity for problem solving by forming higher order nonlinear functions. There are technical difficulties involved in the training of deeper networks, however. The increase in complexity for larger networks can lead to serious constraints surrounding the requisite computing power. Further, backpropagated derivatives can become small in upstream layers, causing very little learning to occur in those layers [80, 84]. For decades, these constraints required applications to be developed around a small set of features extracted from raw data rather than the raw data itself. Engineering of those features is commonly a meticulous process involving considerable domain knowledge. Recent innovations in network architecture and training strategy have led to major advances in performance, particularly in the field of computer vision and recognition [81]. Applications of neural networks now commonly involve deep architectures with many inputs. In the case of image processing problems, it is not uncommon to use upwards of several thousand raw pixels to a network with ten or more hidden layers [81].

Convolutional neural networks have been highly successful in the field of computer vision for classification and many other tasks [87, 83, 88]. The technique is inspired by studies of the visual cortex of animals [81], in which groupings of neurons scan over small sub-regions of the visual field. Independent treatment of sub-regions can allow the network as a whole to detect learned features in a position-independent fashion. In other words, classification of an image would be possible even if the subject appears in a different location within the frame.

This section is not a complete survey of techniques used in deep learning, which are many, but merely a description of the techniques used in this work.

6.3.1 Convolution Layers

Convolution layers [81] produce many alternative representations of the input image, each serving to extract some learned feature. This is achieved using a discrete convolution operation. The convolution of discrete sequences f and g is defined as

$$(f * g)_i = \sum_{j=-\infty}^{\infty} f_j g_{i-j}. \quad (6.18)$$

For convolution layers, the sequence g is taken to be the image and f is called a filter. The filters in a convolutional architecture are restricted to have non-zero values over a finite range corresponding to an $n \times n$ pixel sub-region of the image. The square sub-region of the image is mapped into a one dimensional vector in order to support the convolution. The filter weights, f_j , are learned through backpropagation to learn features within the image. The output of a convolution layer has each pixel replaced with the output of f convolved over the surrounding $n \times n$ sub-region. A convolution layer can have many such filters, leading to multiple representations which are sensitive to different features. A visualization of such an architecture can be seen in Figure 6.3.

6.3.2 Pooling

With a large number of input pixels and many convolutional filters, the number of weights which must be learned can grow dramatically. Pooling [87] is a powerful technique for down-sampling the number of parameters and thus reduce the complexity of

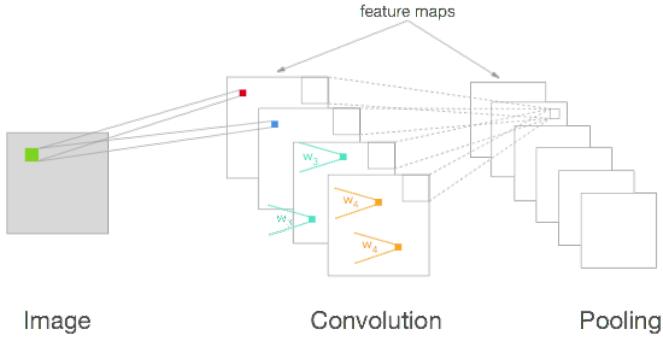


Figure 6.3: Conceptual diagram of a convolutional neural network

The architecture above includes a convolutional layer and a pooling layer. Here, four separate convolutional filters are trained and pooled into a single feature map. The structure can be repeated for several downstream feature maps. Image reproduced from [89].

the network. In so-called *max pooling*, down-sampling is achieved by taking the maximum value of each filter output in $n \times m$ sub-regions of the image. An alternative to max pooling is *average pooling*, where the average value is used instead of the maximum. Pooling regions can be non-overlapping [87], or be allowed to overlap [84, 83]. Pooling has also been used to combine the output of semantically similar² feature maps by taking the maximum (or average) output over a group of feature maps[81]. Semantic similarity is imposed during training through the bottleneck of information flowing backwards through the pooling layers. Layers which pool over feature maps rather than pixel regions are referred to as *maxout* layers.

6.3.3 Network-in-Network and GoLeNet

The Network-in-Network (NIN) [90] approach has been used to augment the learning capacity of convolution layers while also reducing dimensionality. Rather than using maxout feature pooling to reduce dimensionality, the NIN approach recombines features using a weighted sum of filter outputs with a ReLU nonlinearity to produce a new set of features. While NIN layers could be used with more output filters than input

² Semantic similarity is the notion that representations which are grouped together illustrate similar features [81, 90].

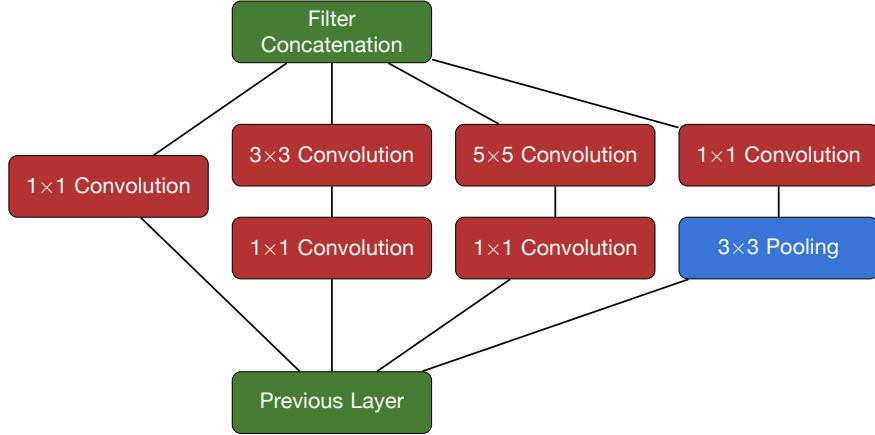


Figure 6.4: Inception module architecture

The inception module 1×1 , 3×3 , and 5×5 filters trained side-by-side [84]. Prior to each convolution operation, 1×1 convolutions with NIN filter recombination to reduce dimensionality.

filters, they are more commonly used with fewer. In that case, the NIN layers can be seen as synthesizing output across filters to produce a higher level, but more compact representation of the input filters.

The NIN approach has been used heavily in a leading network architecture known as GooLeNet [84]. The GooLeNet architecture is built up from a repeated structure dubbed the *inception module*, as seen in Figure 6.4. In the inception module, all 1×1 convolution layers are used for NIN recombination to reduce dimensionality. Prior to more expensive 3×3 and 5×5 convolutions, the input filters are reduced down to 96 and 16 filters, respectively. The 3×3 and 5×5 respectively include 128 and 32 filters, which are concatenated with the output of 64 1×1 filters and another 32 1×1 filters which follow 3×3 max pooling. The full GooLeNet, seen in Figure 6.5 includes nine repetitions of the inception module.

Another important feature of GooLeNet is the use of multiple softmax outputs to better train early layers [84]. GooLeNet includes two extra outputs at upstream points in the network to propagate information backwards during training. In the final

classification result, these outputs are ignored, but during training they provide a less attenuated version of the error. At any point in the network, the gradient used for training is the average of the gradient from each of the downstream softmax outputs.

6.3.4 Regularization

Regularization refers to the prevention of overfitting during network training, that is, ensuring that the trained model can reasonably generalize beyond the training sample. A common regularization technique is to include penalty terms in the objective function which proportional to the square or absolute value of the weights, thus constraining them them to remain small during training. Keeping the weights small prevents the model from responding too strongly to any one input feature.

Recently, a new regularization technique called *dropout* [91] has gained in popularity. Dropout sets a subset of the weights in a layer to zero for a given training iteration. On each iteration, each weight is set to zero with probability r , so a new different sample of weights is excluded at each iteration. The inner product of the weights with the input is scaled up by a factor of $1/r$ to roughly maintain the overall scale of the value passed through the nonlinearity. Dropout with $r \approx 0.5$ proves to be a very robust regularizer and has quickly become the most popular technique used in the field [81].

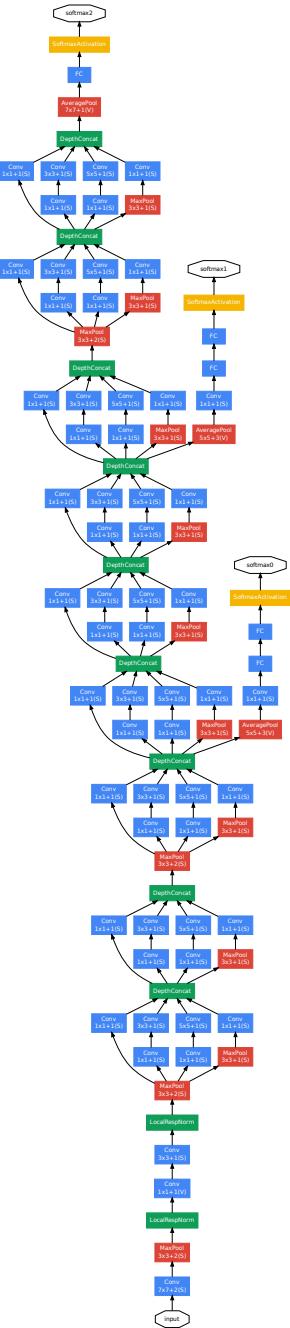


Figure 6.5: Complete GoLeNet architecture

The complete GooLeNet architecture contains 9 inception modules and upstream softmax outputs. Image reproduced from [84] with courtesy of the authors.

Chapter 7

Convolutional Neural Network Event Classifier

At the core of any particle physics analysis is the selection of signal events, that is, the events which represent the physical process under study. Traditionally, features are extracted from event reconstruction and passed to a machine learning algorithm [81], e.g. k-nearest neighbor [78], neural network [80], or decision tree [92]. While this approach is generally successful, those classifiers can often be fooled by reconstruction failures [81]. Even the most robust reconstruction failures will fall victim to more pathological event topologies. Some examples include overlapping particle activity which cannot be resolved and particles which don't travel far enough to make a recognizable track.

An approach which relies on less reconstruction is thus desirable [93]. In the case of NOvA, this means building a classifier which receives the raw detector output as input, so as to cut out the reconstruction as a middleman.

Since NOvA detector output is essentially a pair of images with discrete pixels, it pays to draw inspiration from the computer vision community. The recent advances in image classification [83, 81, 84] discussed in Chapter 6 lend themselves well to the task at hand. The implementation discussed in this chapter involves a convolutional neural network trained on events, treated as images, from NOvA simulation and data. In this approach, the feature extraction units which lead to classification are trained as convolutional filters in a deep architecture.

7.1 NOvA Events as Images

Naively, completely removing reconstruction from the classification pipeline would mean taking the entire detector readout for a given time window as input. The X and Y view could each be treated as an image and passed to a convolutional neural network. The neural network, with the positional independence afforded by convolution, could arguably learn to find neutrino events regardless of where they lie within the detector. Sadly, this naive approach falls down for a few reasons. First, while the ND with roughly 20,000 pixels would produce images of a manageable scale, the FD with nearly 350,000 pixels is far to large. Scaling the images down to an acceptable scale would eliminate considerable detail. Additionally, calibration of hit energy depositions is not straightforward when looking at the entire detector. As discussed in section 5.3.3, calibrated energy deposition from a hit requires an estimate of distance along the fiber path between the hit location and readout electronics. This calibration is important in order to enhance the uniformity of the event images, leaving behind the effects of attenuation and cell-to-cell variations would add unnecessary diversity to the training sample.

Rather than using the readout of the entire detector, it is sensible to use the reconstructed slices discussed in section 5.1. The performance of DBSCAN in isolating cosmic rays and neutrino interactions is robust and efficient enough [72] that relying on this simple reconstruction step will does significantly impact classification results. In the case of slices, the distance from the readout electronics is determined by the mean position of all activity in the opposite view.

Each slice is transformed into an image by locating a region of interest to capture activity. The images are 80 cells wide and 100 cells deep in either view. The upstream side of the image is the first plane with a hit in the slice. In either view, the center of the image is defined by the median cell position among all hits within the 100 plane window spanned by the image. The size and placement of this window ensures that the majority of beam events are well contained, including ν_μ CC interactions which are characteristically extensive as a result of long muon tracks. Example images can be seen in Figures 7.1, 7.2, and 7.3.

Pixel intensity in the image is proportional to the calibrated energy deposition in

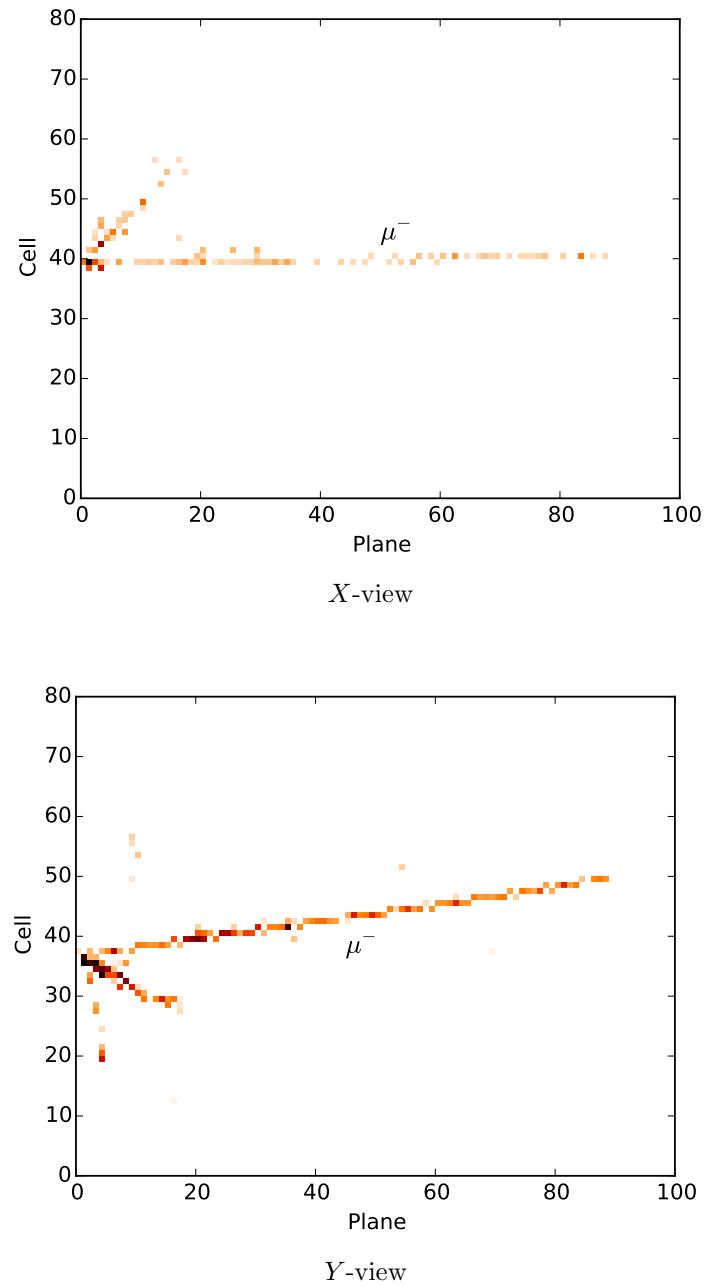


Figure 7.1: Image formed from a ν_μ CC interaction

Images such as this one are passed as input to the convolutional neural network. The top and bottom panes show the *X*-view and *Y*-view, respectively. The region of interest is determined by the furthest plane upstream with activity and the median cell position of all activity in the 100 plane window spanned by the image. The pixel intensity is proportional to the calibrated energy deposition of each cell hit. This particular image consists of a long muon track and a hadronic shower.

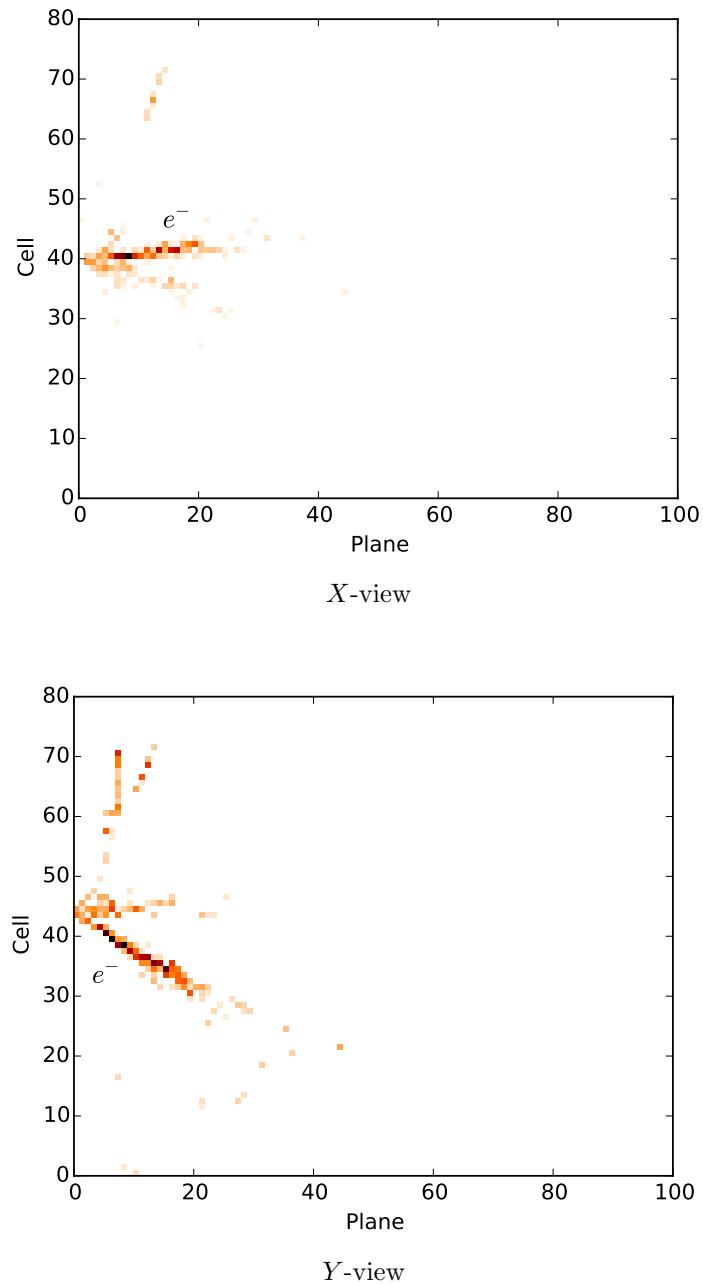


Figure 7.2: Image formed from a ν_e CC interaction

Images such as this one are passed as input to the convolutional neural network. The top and bottom panes show the X-view and Y-view, respectively. The region of interest is determined by the furthest plane upstream with activity and the median cell position of all activity in the 100 plane window spanned by the image. The pixel intensity is proportional to the calibrated energy deposition of each cell hit. This particular image consists of a long muon track and a hadronic shower.

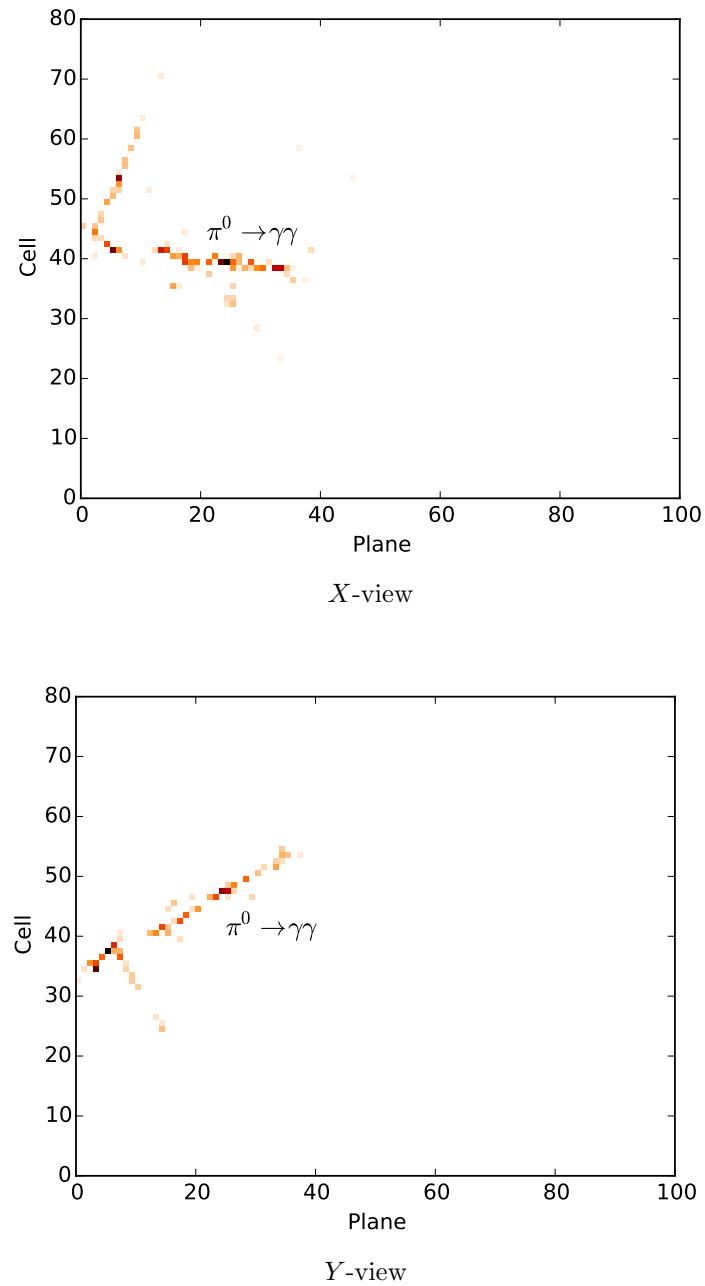


Figure 7.3: Image formed from an NC interaction

Images such as this one are passed as input to the convolutional neural network. The top and bottom panes show the *X*-view and *Y*-view, respectively. The region of interest is determined by the furthest plane upstream with activity and the median cell position of all activity in the 100 plane window spanned by the image. The pixel intensity is proportional to the calibrated energy deposition of each cell hit. This particular image consists of a hadronic shower including a neutral pion which decays to two photons.

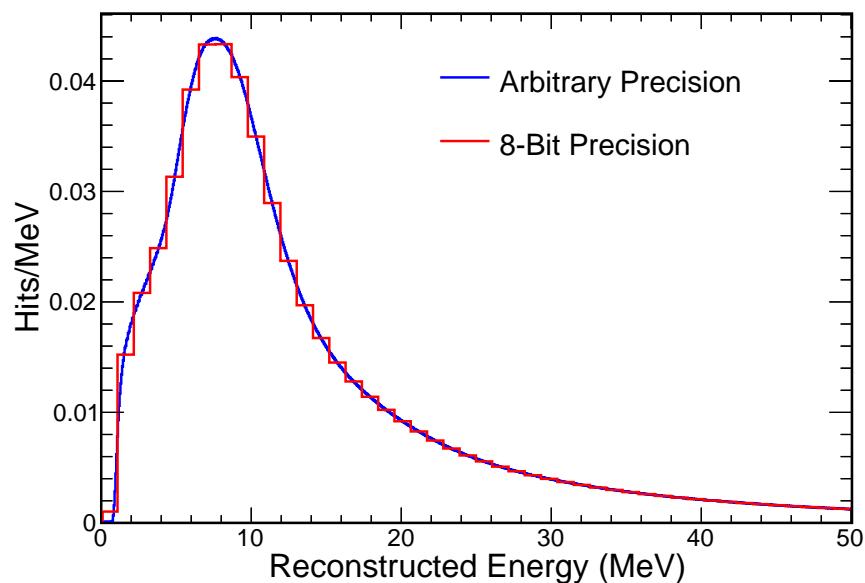


Figure 7.4: Comparison of continuous hit intensity scale and discretized scale
Data volume is reduced by converting the continuous intensity scale of cell hit energy depositions into a scale with 256 discrete values. The discrete scale can be represented with eight bits, offering a significant savings over a floating point representation.

each cell. Images are thus interpretable as grayscale with the shade of each pixel corresponding to the amount of energy recorded in that cell. In order to optimize data storage and transfer in the training stage, the pixel intensities are converted to a scale of 256 discrete values from the double precision floating point representation calibration result. This conversion offers a factor of eight savings (eight bit versus 64 bit) in savings over a floating point representation without significantly compromising the representational capacity. The importance of this savings is especially visible in reading data from disk into memory during training. A comparison original continuous scale and discrete scale can be seen in figure 7.4.

7.2 Architecture

The network architecture used in for this analysis is primarily based on the GooLeNet architecture [84] introduced in section 6.3.3. Initial experiments were performed with a full implementation of GooLeNet for each view, with only the final fully-connected layers merged. It was observed during training, however, that the classification of the second intermediate loss unit showed equivalent classification accuracy to the final loss unit. The architecture was then pruned and restructured in order to reduce the total number of operations, and thus the amount of computation required for training and evaluation. No degradation was observed in the classification results with the pruned architecture. A depiction of the architecture can be seen in Figure 7.5.

Input from the X -view and Y -view is handled separately in two branches of the network which are only merged in the final few layers. Early in the network, images are subjected to 7×7 convolution and 3×3 pooling, both with a stride of two. Using stride two in these early layers scales the images down by a factor of four. Local response normalization (LRN) [83] is used to smooth the filtered images. Two more convolution operations, 1×1 and 3×3 are used, then another pass of smoothing with LRN, and 3×3 pooling with stride two. The 1×1 convolution serves as dimensionality reduction in the number of filters. A pair of inception modules is then used, then another 3×3 pooling layer with stride two and another inception module. The filter outputs for each view are then concatenated and passed through another inception module. A final pass of average pooling is employed before a softmax layer, which serves as the output of the

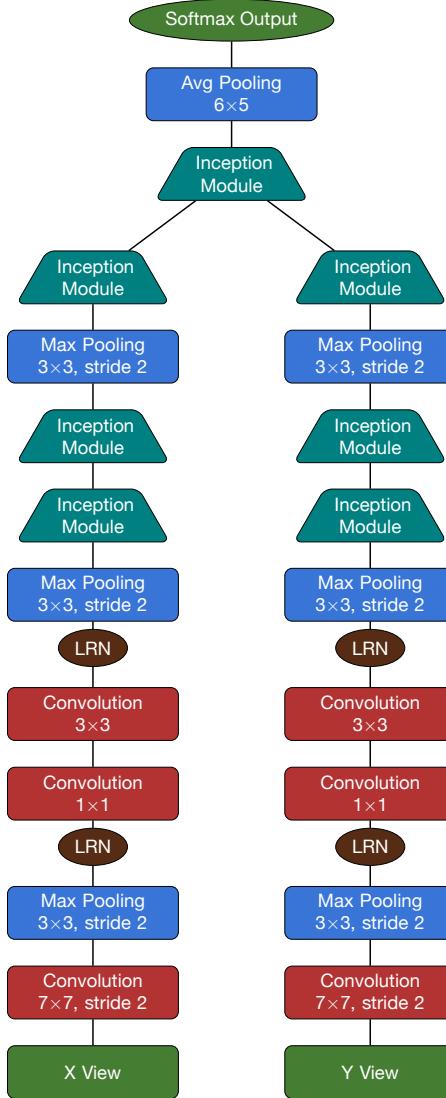


Figure 7.5: Diagram of the convolutional neural network architecture used for event classification

Starting with the input at the bottom, the network has separate branches for the X -view and Y -view. Each branch undergoes successive convolution, pooling, and local response normalization (LRN). Inception modules are used in downstream layers. The two views are merged and passed through a final inception module, and pooled. The output of the network comes from softmax units.

network.

In the interest of time, the architecture presented here has only been coarsely optimized, leaving plenty of room for further study. It seems quite likely that a significantly more compact network could perform equally as well. Classification results could also perhaps be improved through fine tuning of the architecture. Eliminating the aggressive downsampling with stride two in the early layers could significantly increase the level of detail which is available in the inception modules. Merging views earlier in the architecture could provide higher levels of abstraction and greatly reduce the number of operations required. The effect of LRN should be studied by removing those layers and comparing results.

7.3 Regularization

A few regularization techniques were applied in order to enhance the ability of the network to generalize beyond the training sample. To keep the network weights small, weight decay term was included in the loss function equal to 2×10^{-4} times the squared *L₂* norm of the weight vector. The softmax output layer was also subjected to dropout [91], described in section 6.3.4, with a probability of 0.4.

Perhaps the most robust defense against over training is more training data. In the absence of infinitely large and variegated training samples, however, we can augment our sample by deliberately randomly changing the input image. Two techniques were employed in order to add variation to the dataset. First, Gaussian noise was applied to all pixel intensities with a standard deviation of 1%. Since no Monte Carlo simulation is perfect, adding noise has the added benefit of training the network to rely less heavily on the simulated intensity in each pixel. Second, events were randomly selected to be reflected in the cell dimension, which is roughly transverse to the beam direction. Symmetry about in the cell dimension is not perfect; the beam axis is directed 3° above detector horizon and attenuation in the optical fiber causes thresholding to become more significant for hits further from the readout electronics. These effects are small, however, and their presence in fact aids in enhancing variation of the training sample and deweighting precise details of the Monte Carlo simulation.

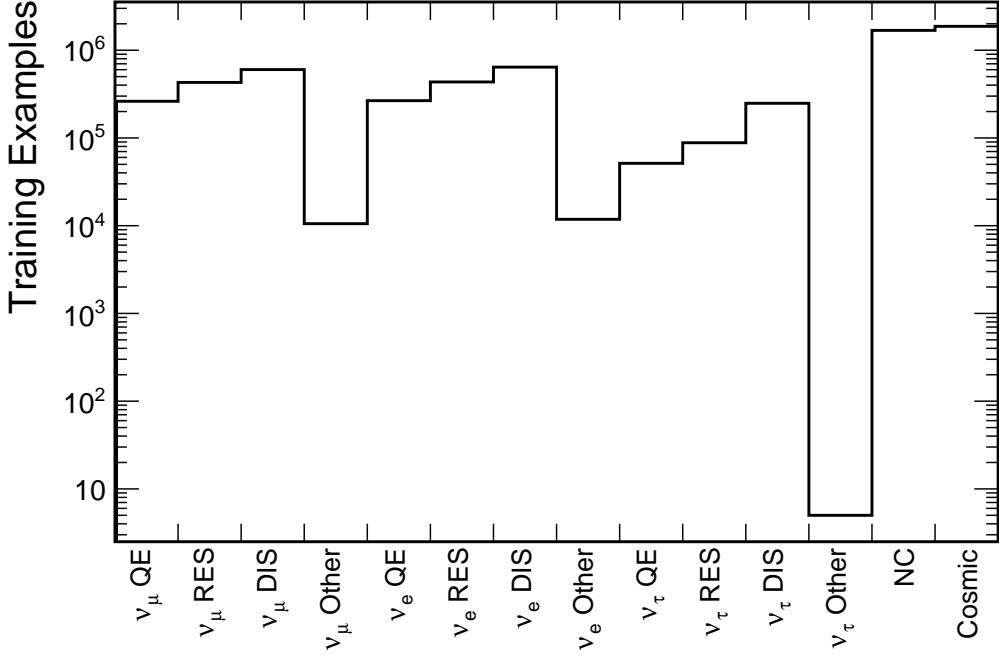


Figure 7.6: Histogram depicting class distribution for training dataset

Training events are taken from NOvA MC. Events are simulated with the NuMI energy spectrum and flavor distribution, ignoring oscillation. Samples of ν_e and ν_τ events are enhanced by respectively interchanging the ν_e and ν_τ components with the ν_μ components. Cosmic rays were only included in the sample during the fine-tuning phase of the training.

7.4 Training

Training was performed using the Caffe framework [94] using NVIDIA Tesla K20 and K40 graphics processing units (GPU). Weight updates were performed using the basic stochastic gradient descent (SGD) [80] procedure described in Chapter 6.

The training sample was taken from NOvA Monte Carlo simulation. Events are simulated with an energy, interaction, flux distribution corresponding to the NuMI beam spectrum convoluted with GENIE cross section model. Enhanced samples of ν_e and ν_τ events are generated by respectively interchanging the ν_e and ν_τ of the NuMI beam flux with the vastly dominant ν_μ component. Flavor interchange is based on the

neutrino flux upstream of GENIE cross section convolution to allow proper interaction weighting. More details on the simulated neutrino spectrum can be found in section 4.2. Cosmic ray data were added to the training sample for fine-tune training phase. The events were required to pass a preselection based on track containment and angle of the reconstructed track from the CosmicTrack algorithm described in section 5.2.1. The cosmic ray preselection will be described in section 8.1.

Events were split into classes based on the interacting ν flavor and GENIE interaction mode. NC interactions are grouped into a single class regardless of ν flavor. For CC interactions, a class is formed for each flavor (ν_e , ν_μ , ν_τ) and GENIE interaction mode combination. The QE, RES, and DIS modes each form their a distinct class, with all other interactions placed in an “other” category. Separating events by interaction mode is motivated as an attempt to give the network a hint at the underlying physics to aid in training. Another class is reserved for cosmic rays. In total, the training sample included 4.7 million neutrino events from NOvA simulation and 1.8 million events from cosmic ray data. The class distribution can be seen in table 7.1 and visualized as a histogram in table 7.6.

Due to GPU memory constraints, a mini-batch training scheme was employed. In mini-batch training, gradients used in weight updates are averaged over small subsamples of the training dataset rather than the entire sample. After calculating the gradient for a batch, it is flushed from memory and the next batch is loaded. Batches are reused after each full pass over the training dataset. A batch size of 32 was used to train the network. Each pass over a mini-batch is referred to as an *iteration*; a full pass over the dataset is called an *epoch*.

The initial learning rate for SGD was set to 0.01. During the training, the learning rate was decreased by a factor of 0.96 according to a prescribed schedule. The learning rate was decreased after the first 100,000 iterations, again after the next two sets of 50,000 iterations, then again after four sets of 25,000 iterations, and subsequently after each 10,000 iterations.

Training was interrupted after 640,000 iterations after an error was discovered in the architecture. Rather than having the intended 15 nodes in the softmax output layer, the architecture had accidentally been configured with 1,000 nodes in that layer. The architecture was modified to have the correct number of nodes, and training was

Class Label		Number of Training Examples
ν_μ	QE	261,823
	RES	429,470
	DIS	601,497
	Other	10,546
ν_e	QE	266,111
	RES	434,720
	DIS	641,813
	Other	11,829
ν_τ	QE	51,372
	RES	88,185
	DIS	248,408
	Other	5
NC		1,683,050
Cosmic Ray		1,872,258

Table 7.1: Class distribution for training dataset

Training events are taken from NOvA MC. Events are simulated with the NuMI energy spectrum and flavor distribution, ignoring oscillation. Samples of ν_e and ν_τ events are enhanced by respectively interchanging the ν_e and ν_τ components with the ν_μ components. Cosmic rays were only included in the sample during the fine-tuning phase of the training.

resumed by reloading trained weights from all layers but the output. This method for resolving the error was justified by an approach commonly used for image classification competitions known as *fine-tuning* [83]. Since learning occurs much more quickly in downstream layers relative to upstream filters, downstream layers can be quickly fine-tuned based on filters which have been trained in advance. Upon resuming training, the learning rate was set to 0.002, roughly equal to the learning rate reached after 640,000 iterations. The decrease schedule, however, was started from the beginning rather than reverting to the once every 10,000 iteration schedule that would have been employed beyond iteration 640,000.

After 970,000 total iterations, the cosmic ray data were added to the training sample. With this enhanced training sample, the network was trained for another 990,000 iterations. A separate version of the network was trained for 990,000 iterations using cosmic ray data from the beginning, but was outperformed by the other network in terms of overall accuracy. In particular, the network originally trained on a beam-only training set allowed for more pure ν_e CC selection.

7.5 Training Results

The trained network serves as a good classifier for ν oscillation analysis involving both ν_μ disappearance and ν_e appearance. Classifier outputs ν_e , ν_μ , and ν_τ are formed by summing across the softmax outputs for QE, RES, DIS, and Other subclasses. The performance of this classifier has been demonstrated to outperform all other classification approaches developed for the NOvA oscillation analyses. The ν_μ and ν_e classifier outputs can be seen in figures 7.7 and 7.8 respectively. The application to ν_μ disappearance will be discussed in detail throughout the rest of this dissertation.

7.6 Filter Visualization

Since convolution and pooling layers preserve the spatial orientation of information within an image, visualizing filters and filter output for particular images can provide insight with regard to the network operation. As observed in many convolutional neural network examples, early layers in the network respond to low-level characteristics in an

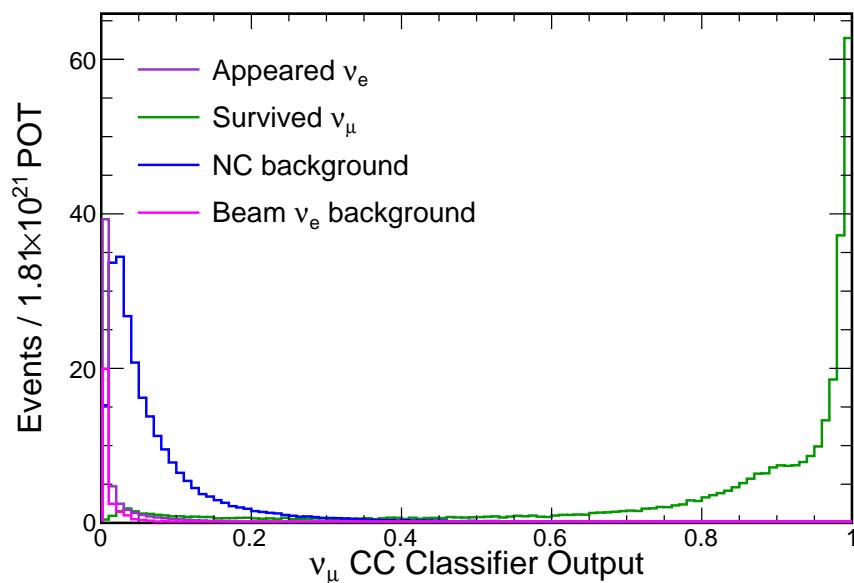


Figure 7.7: Histogram of the ν_e classifier output

The survived ν_μ signal can be seen as a strong peak at the right. Both ν_e , and NC interactions all enter as a background in ν_μ disappearance analysis. Clear separation in the classifier can be seen between the signal and backgrounds.

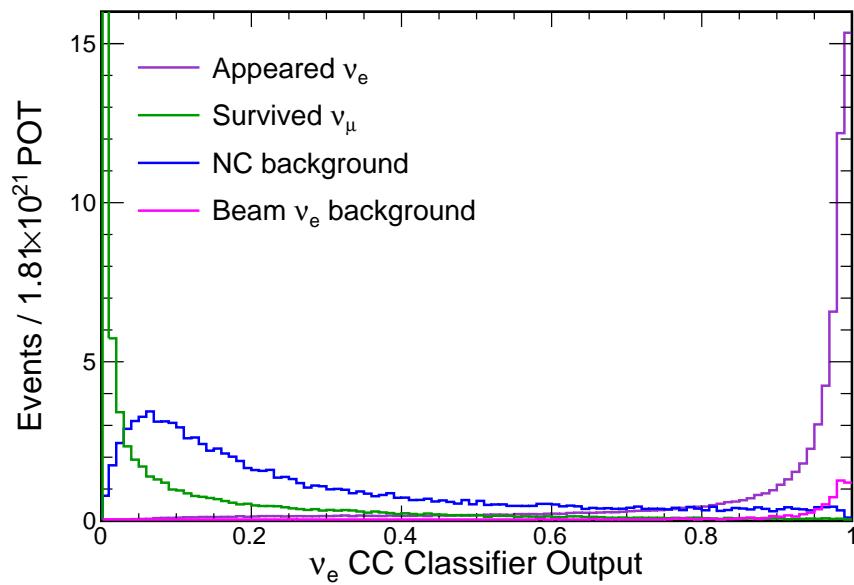


Figure 7.8: Histogram of the ν_μ classifier output

The appeared ν_e signal can be seen as a strong peak at the right. Survived ν_μ , beam ν_e , and NC interactions all enter as a background in ν_e appearance analysis. Clear separation in the classifier can be seen between the signal and backgrounds.

image [81]. The 7×7 convolutional filters trained in the first layer can be seen in Figure 7.9. The output of those filters can be seen in Figures 7.10, 7.11, and 7.12. Downstream layers in the network start to pick up higher level concepts, like the presence of track-like or shower-like activity. Output from the first inception module in the Y -view branch can be seen in Figure 7.13.

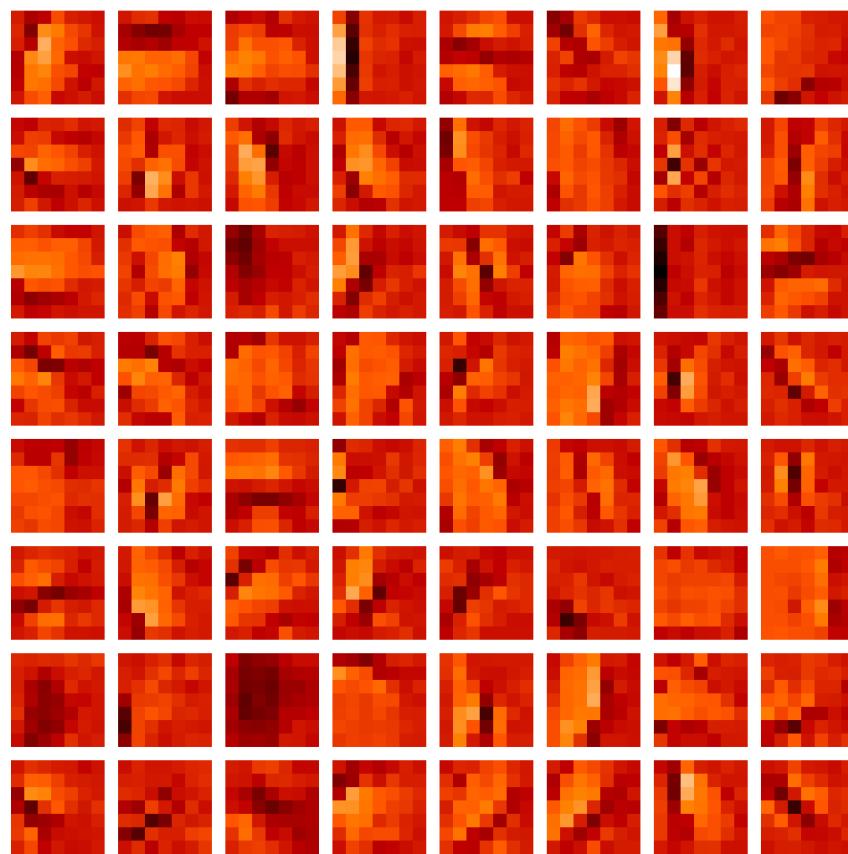


Figure 7.9: 7×7 Convolutional Filters

The first layer of the network used for classification performs 64 different 7×7 convolution operations for both the X -view and Y -view. Each entry in the grid above shows one of those trained filters from the first layer of the Y -view branch of the network.

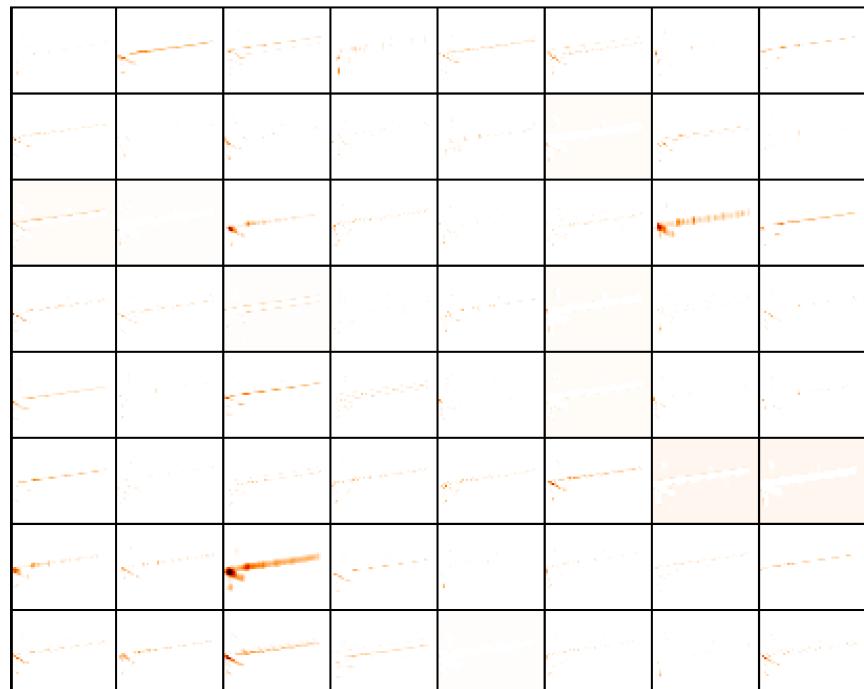


Figure 7.10: 7×7 convolutional filter outputs for a ν_μ CC event

Convolutional filters produce many alternative representations of an image. Each entry in the grid shows the output in the corresponding location of the grid in figure 7.9. The input image for these filtered representations was the Y -view image of the ν_μ CC event from Figure 7.1

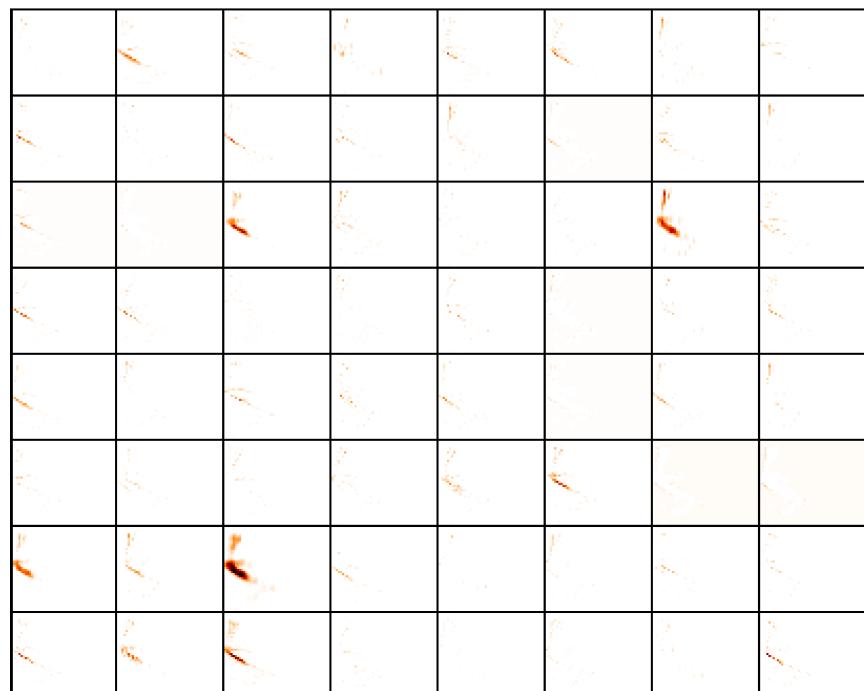


Figure 7.11: 7×7 convolutional filter outputs for a ν_e CC event

Convolutional filters produce many alternative representations of an image. Each entry in the grid shows the the output in the corresponding location of the grid in figure 7.9. The input image for these filtered representations was the Y -view image of the ν_e CC event from Figure 7.2

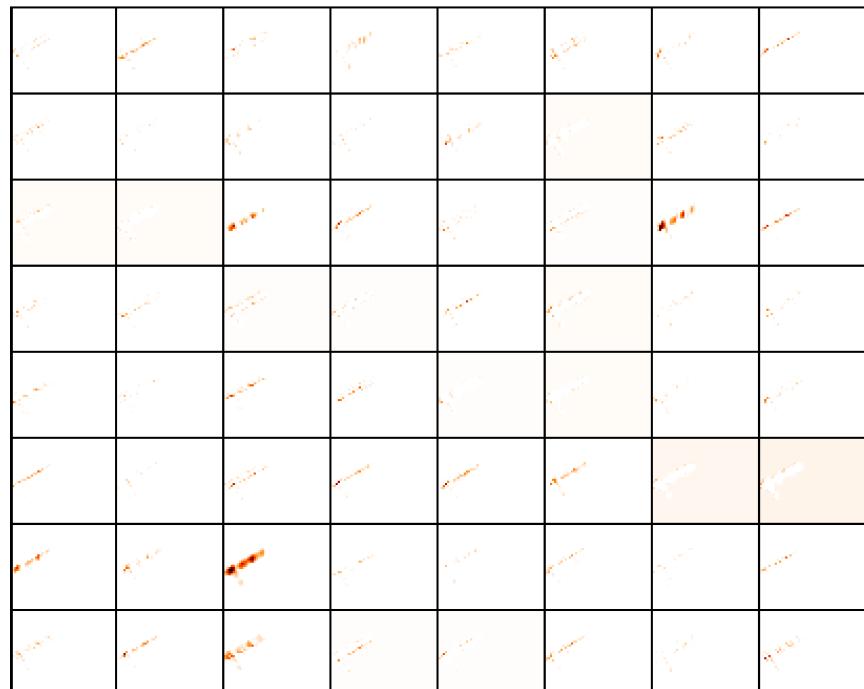
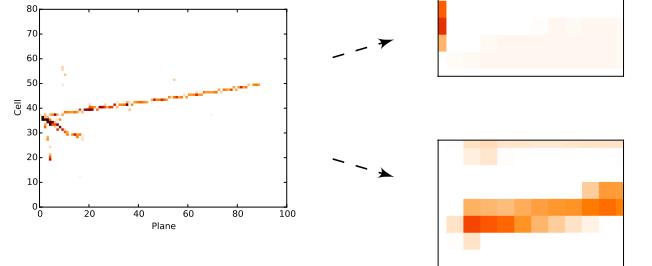
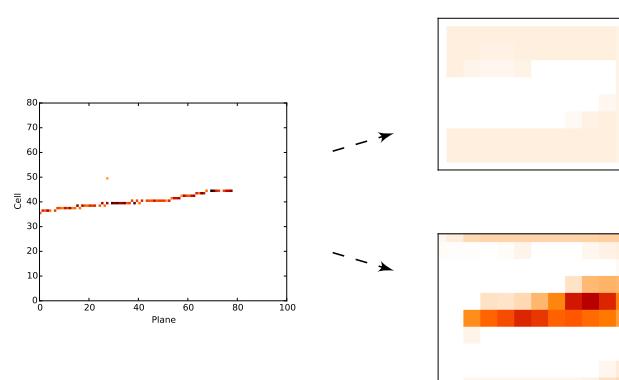


Figure 7.12: 7×7 convolutional filter outputs for an NC event

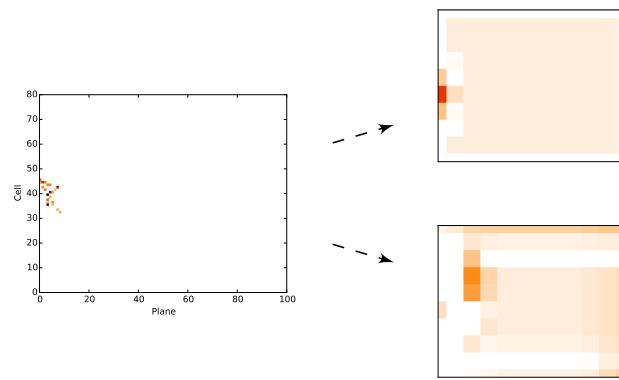
Convolutional filters produce many alternative representations of an image. Each entry in the grid shows the the output in the corresponding location of the grid in figure 7.9. The input image for these filtered representations was the Y -view image of the NC event from Figure 7.3



ν_μ CC DIS event with significant hadronic activity



ν_μ CC QE event with little hadronic activity



NC event with solely hadronic activity

Figure 7.13: Inception module filter output

Input images and corresponding filter output from two particular filters in the first inception module in the Y-view branch of the network. In each pane, the input image at left and filter outputs are at right. The top filter in each pane appears to respond to hadronic activity, while the bottom filter appears to correspond to more track-like activity.

Chapter 8

Event Selection

Measuring the parameters which govern a physical process requires a clean sample of events which exhibit that process. The definition of clean, of course varies wildly depending on the experiment and parameters in question. For this analysis, the goal is to accurately determine the reconstructed energy spectrum for ν_μ CC events. Thus, the goal of event selection is twofold: the selected sample should be relatively background free and consist of events for which the energy of the interacting neutrino can be determined.

Estimating the energy of events requires that they be fully contained within the detector. If all particles borne in the interaction deposit their energy within the detector, then an estimate of the neutrino energy can be obtained using the method described in section 5.5. If one or more of the particles escape the detector, it can be difficult to estimate how much energy is missing. Events with poor energy resolution can provide some sensitivity to oscillation parameters, in particular θ_{23} , which governs the amplitude of oscillation. This analysis, however, makes no attempt at including escaping events.

Two primary backgrounds must be suppressed for this analysis. The first is the vastly dominant cosmic-ray background; the FD is on the surface of the earth rather than underground and is thus exposed to cosmic rays at a rate in excess of 100 kHz [46]. The second source of background is events from the NuMI beam which are not ν_μ CC interactions. NC interactions are the trickiest beam background to reject due to their abundance, especially those with a charged pion in the final state which can produce a relatively long track. Both appeared and survived ν_e in the beam are relatively rare

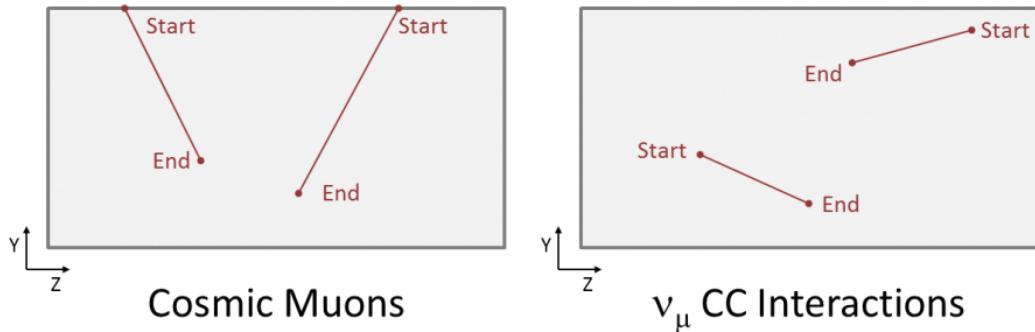


Figure 8.1: Comparison of track topologies for cosmic-ray and neutrino events
 Cosmic-ray muon tracks (left) will tend to touch at least one detector face. Neutrino interactions (right) which occur within the detector will tend to be well contained. Cosmic-ray tracks which touch two detector faces can be rejected without appreciable signal loss.

and tend not to fake ν_μ CC signal.

The event classifier described in Chapter 7 serves as to resolve both cosmic-ray and beam-induced backgrounds. Additional selection cuts described in this chapter serve to ensure reconstruction quality and containment. A handful of extra cuts are required to further reject cosmic background.

8.1 Cosmic-Ray Preselection

The vast abundance of cosmic-ray activity in the FD necessitates the use of an initial preselection upstream of heavier reconstruction algorithms. Running the entire reconstruction chain over all cosmic-ray activity would simply be too computationally expensive. Instead, an initial preselection has been applied based on the output of the slicing and CosmicTrack algorithms described in chapter 5. The preselection rejects roughly 90% of cosmic-ray activity while preserving nearly all contained ν_μ CC signal.

A large majority of cosmic-ray activity consists of muons which traverse the detector, as depicted in Figure 8.1. If the muon tracks are well reconstructed, this activity can easily be rejected by identifying tracks which touch two faces of the detector. Proximity to the detector faces is determined by measuring the distance between the track start and stop points as projected along the track direction. The track projections method

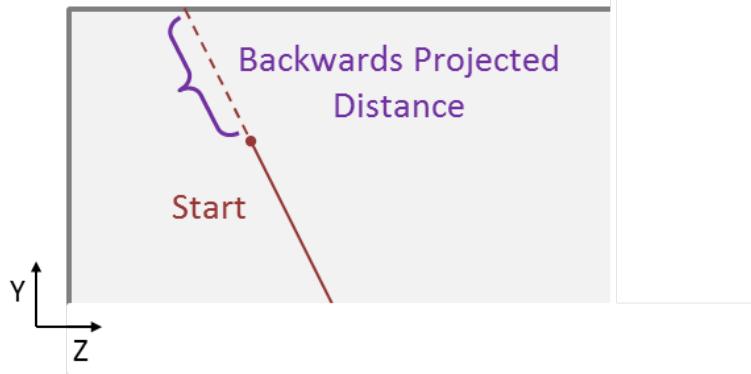


Figure 8.2: Illustration of track distance projection

The proximity of a track endpoint to the edge of the detector is determined by measuring the distance projected along the track direction.

is depicted in Figure 8.2. This measure is only reliable when the simple CosmicTrack fit has succeeded, however. The fit result is checked by assessing the fraction of hits in the slice which were included in the track fit, with high values indicating a successful fit. As such, the event is rejected if both projected distances are less than 35 cm and the hit fraction is greater than 0.8.

The cosmic-ray preselection also requires events to have at least 20 hits and fewer than 400. Events outside of that range are typically of an interaction energy which carries little oscillation sensitivity. An additional selection variable is constructed from the track direction cosines relative to the beam, c_{beam} , and vertical, c_y :

$$a = |c_{beam}| * (c_y + 1) \quad (8.1)$$

Tracks from the CosmicTrack algorithm are constructed as downgoing, so c_y is always negative. Thus, the variable a is large, the track is generally directed along the beam axis and not highly vertical. Events with a greater than 0.2 will pass the preselection.

8.2 Reconstruction Quality

Events of which have not been well reconstructed can cause problems in energy estimation. In particular, identification of the muon track within a ν_μ CC interaction requires the KalmanTrack algorithm to have reconstructed a 3D track. Events with very little activity are also not especially intelligible. As such, the reconstruction quality selection requires a reconstructed track from the CosmicTrack algorithm as well as a 3D track from the KalmanTrack algorithm. The quality selection also requires events to have at least 20 hits spanning at least 4 contiguous planes.

8.3 Muon Track Designation

As described in Section 5.5, reconstructed neutrino energy depends on the identification of a muon track. This distinction is made by the Muon ID algorithm described in Section 5.4. Tracks reconstructed by the KalmanTrack algorithm (Section 5.2.2) are each assigned a Muon ID score. The track with the highest Muon ID score is dubbed the muon track and treated with particular selection requirements. All remaining activity in the slice is grouped into a hadronic cluster, which also receives particular selection requirements.

8.4 Containment

8.4.1 Far Detector Containment

Containment in the FD is based on the distance of the slice and track endpoints from the edges of the detector. Slices are required to be at within the detector with a buffer of two planes upstream and downstream as well as two cells in both views. Track containment is based on projected distances in a similar fashion to the cosmic-ray preselection, however the distances are measured in the number of cells spanned by the projection rather than the distance itself. The muon track is required to have a projected span of at least 11 cells in the forward and backward direction, and at least one cell for the track from CosmicTrack.

8.4.2 Near Detector Containment

Containment in the ND is more complicated than the FD due to the consideration of the Muon Catcher and its truncated height. Slices are rejected if they start in the first two planes of the detector or end within the last two planes. The muon track is required to start in active region by requiring that the start Z position is within 1175 cm from the front of the detector. If a track enters the muon catcher, the height of the track at the first muon catcher plane is required to be less than the height of the muon catcher ($Z = 55$ cm) to ensure that it did not leave and re-enter the detector. The track projection is also required to span at least five cells in the forward direction nine cells in the backward direction. Finally, the hadronic cluster (all hits in the slice not on the muon track) is required to have calorimetric energy less than 30 MeV.

8.5 Background Rejection

The selections above ensure events are of sufficient quality for analysis and reduce level of cosmic-ray activity to an acceptable level. The primary rejection of NC and cosmic-ray events is handled by the convolutional neural network classifier and some additional requirements to handle cosmic-ray events which would remain in the sample otherwise.

The convolutional neural network discriminant is formed by the sum of the four softmax outputs corresponding to ν_μ CC (QE, RES, DIS, other). Events are rejected if the softmax output is less than 0.5. Figure 8.3 shows the distribution of the discriminant for signal and background. A fair amount of cosmic-ray background slips past the softmax requirement. The vast majority of these events are near the top of the detector and can be rejected by requiring the CosmicTrack start Y position to be less than 600 cm, as seen in Figure 8.4. Additional cosmic background can be rejected using the transverse momentum of the event. Two measures of transverse momentum are used, one which is based on the reconstructed slices and another based on the reconstructed prongs described in [73]. In the slice method, a unit vector is formed in the direction running from the start of the muon track to the energy centroid of the slice; the transverse momentum fraction reconstructed as the sine of the angle between that vector and the beam axis. The prong method is similar, except that the the momentum vector is formed as the energy weighted sum of all reconstructed prongs for that slice. Both of

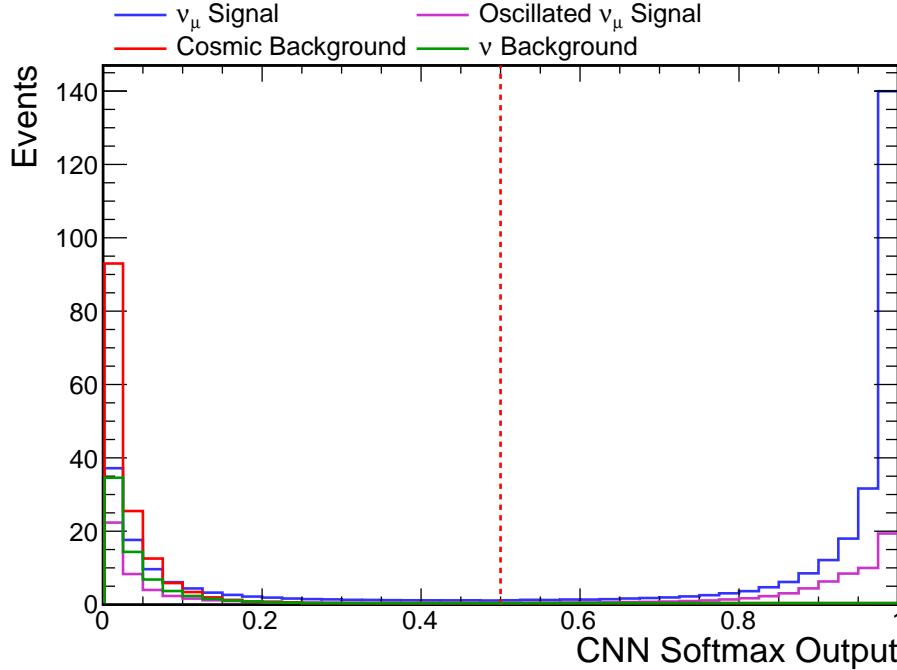


Figure 8.3: Distribution of softmax output from convolutional neural network for signal and backgrounds

The convolutional neural network output serves as the primary discriminant for rejecting background. The blue distribution shows ν_μ CC signal from the unoscillated beam spectrum, magenta shows the oscillated spectrum (weighted according to Section 4.6). The red and green distributions show the cosmic-ray and beam backgrounds, respectively. Selection applied to these distributions include all FD analysis cuts with the exception of the cut on the variable shown here. Events are rejected if they fall to the left of the vertical dashed red line.

these transverse momentum measures, seen in Figures 8.6 and 8.5, are required to be less than 0.8.

8.6 FD Selected Sample

The figures at the end of this chapter give many representations on the FD selected sample. Figure 8.7 shows FD MC prediction for the reconstructed energy spectrum for the final selected energy spectrum. Beyond that, two perspectives can be used to interpret the selection: the selected spectrum itself and the efficiency of the selection.

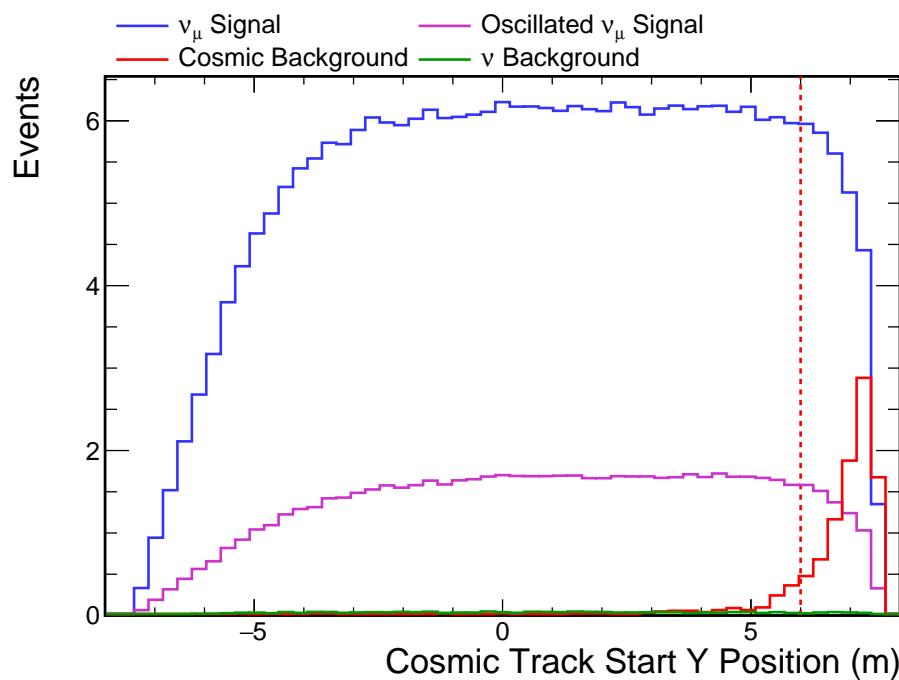


Figure 8.4: Distribution of CosmicTrack start Y position for signal and backgrounds. The blue distribution shows ν_μ CC signal from the unoscillated beam spectrum, magenta shows the oscillated spectrum (weighted according to Section 4.6). The red and green distributions show the cosmic-ray and beam backgrounds, respectively. Selection applied to these distributions include all FD analysis cuts with the exception of the cut on the variable shown here. Events are rejected if they fall to the right of the vertical dashed red line.

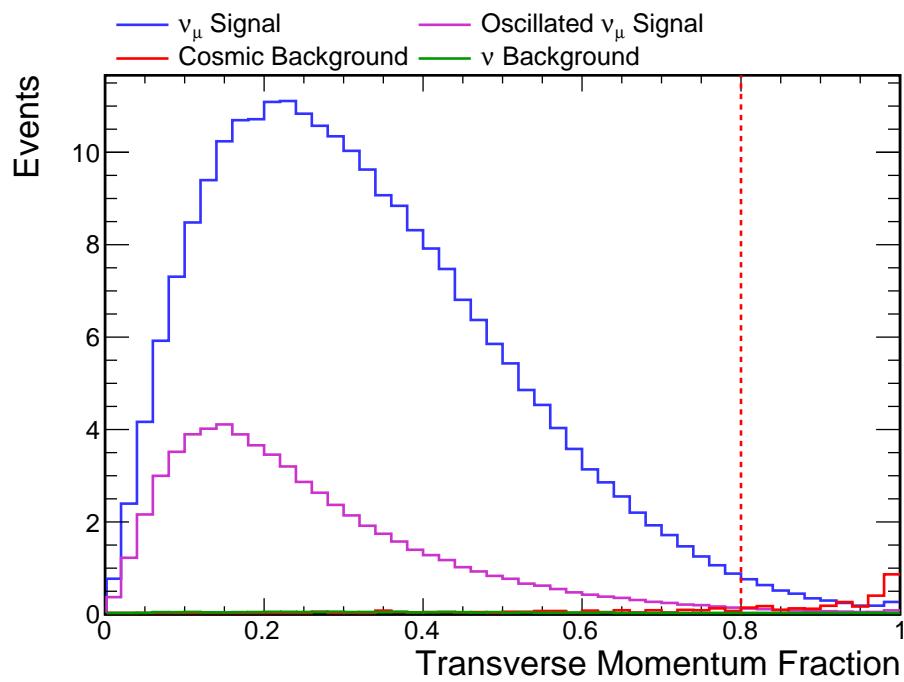


Figure 8.5: Distribution of transverse momentum estimated from slice mean position for signal and backgrounds

The blue distribution shows ν_μ CC signal from the unoscillated beam spectrum, magenta shows the oscillated spectrum (weighted according to Section 4.6). The red and green distributions show the cosmic-ray and beam backgrounds, respectively. Selection applied to these distributions include all FD analysis cuts with the exception of the cut on the variable shown here. Events are rejected if they fall to the right of the vertical dashed red line.

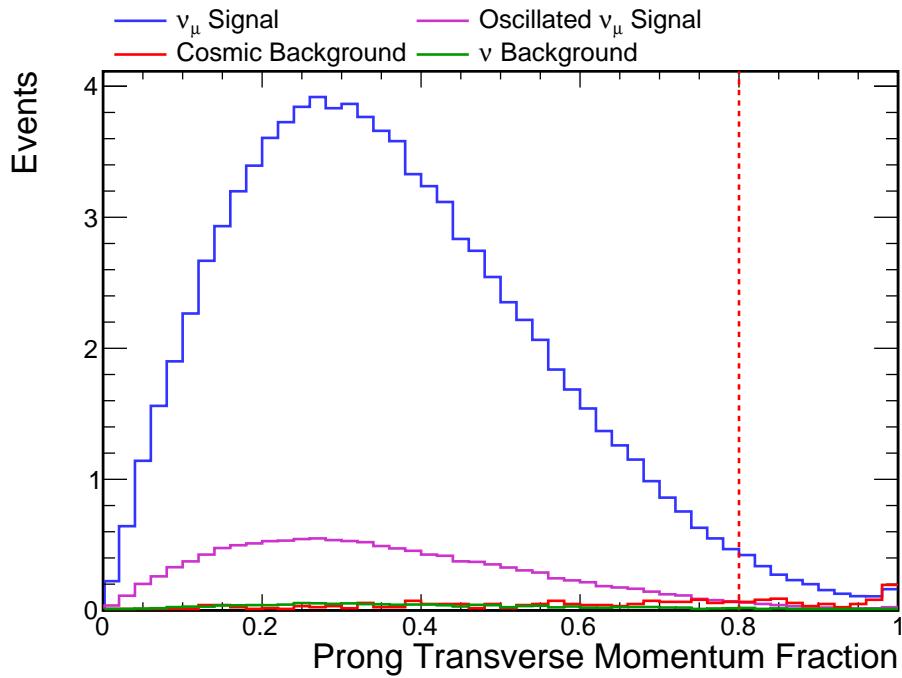


Figure 8.6: Distribution of transverse momentum estimated from reconstructed prongs for signal and backgrounds

The blue distribution shows ν_μ CC signal from the unoscillated beam spectrum, magenta shows the oscillated spectrum (weighted according to Section 4.6). The red and green distributions show the cosmic-ray and beam backgrounds, respectively. Selection applied to these distributions include all FD analysis cuts with the exception of the cut on the variable shown here. Events are rejected if they fall to the right of the vertical dashed red line.

In each bin, efficiency is the ratio of the selected events to the total events which could have been selected. The distribution denominator in this ratio – events which could have been selected – is taken to be the sample of events with are contained in the detector based on MC truth. Simulated events are deemed to be contained if all true simulated energy depositions are contained within the detector volume. Figures 8.8, 8.9, 8.10, and 8.11 all include both MC prediction for the selected FD spectrum and the efficiency ratio. In the spectrum, the true contained events is shown as a black distribution, which serves as the denominator for the efficiency distribution.

Figure 8.8 shows distributions for true signal events for the samples remaining after each cut (quality, containment, CNN classifier, CosmicTrack start Y , and the two transverse momentum cuts) are successively applied. Figure 8.9 shows the same plots for the true background events. For the CNN classifier, the signal efficiency is generally higher (70%) for events with higher energy. The CosmicTrack Start Y cut leads to a flat reduction in signal efficiency as a function of energy, which is to be expected since it simply reduces the effective detector volume. Transverse momentum, on the other hand, impacts low-energy events more significantly since small reconstruction is more difficult for events with less activity. Cutting on the CNN classifier output reduces the background rate significantly, as seen in figure 8.9.

Figures 8.10 and 8.11 show schematically similar plots, but the final two samples are the result of different selections rather than applied separately. The *CNN + Cuts* sample uses the selection described here, while *Muon ID + BDT* uses the selection described in [68]. That alternative selection uses the presence of an identified muon to select ν_μ CC events, as well as a boosted decision tree (BDT) [92] to reject cosmic-ray background. The CNN approach provides improved efficiency in the low energy bins which are especially sensitive to neutrino oscillation. Losses in the high energy region are driven mainly by the CosmicTrack start Y cut. Senitivity enhancement achieved by the CNN approach is quantified in Figure 8.12.

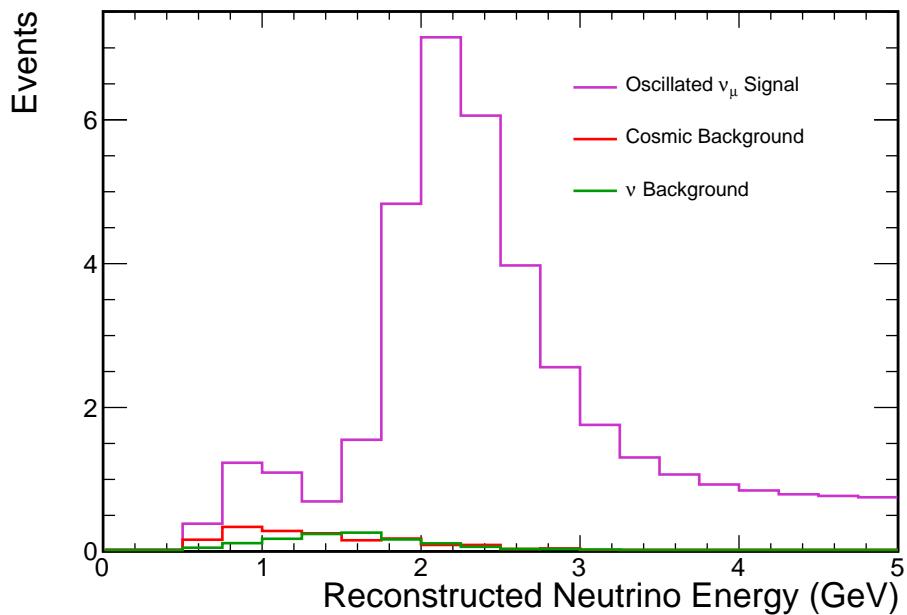


Figure 8.7: Final selected spectrum for ν_μ CC signal and backgrounds
The FD MC prediction of the final selected spectrum resulting from all cuts described in this chapter. The ν_μ CC signal distribution has had oscillation probabilities applied based on the true energy of selected events and oscillation parameters from described in 4.6.

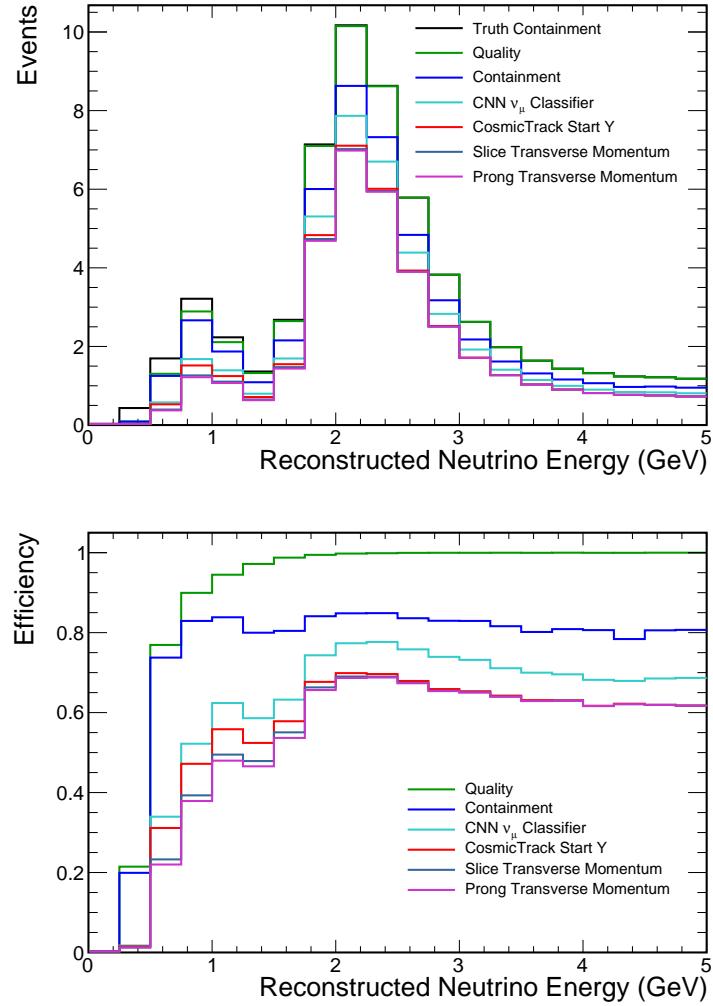


Figure 8.8: Selected ν_μ CC signal event spectrum and efficiency

The top pane shows the reconstructed neutrino energy spectrum for selected ν_μ CC signal events from FD MC simulation. Truth containment is the sample of events which deposit no energy outside of the detector, this serves as the denominator for the efficiency ratio shown in the bottom pane. Each subsequent spectrum adds an additional selection cut to the sample. The ν_μ CC signal distribution has had oscillation probabilities applied based on the true energy of selected events and oscillation parameters from described by 4.6.

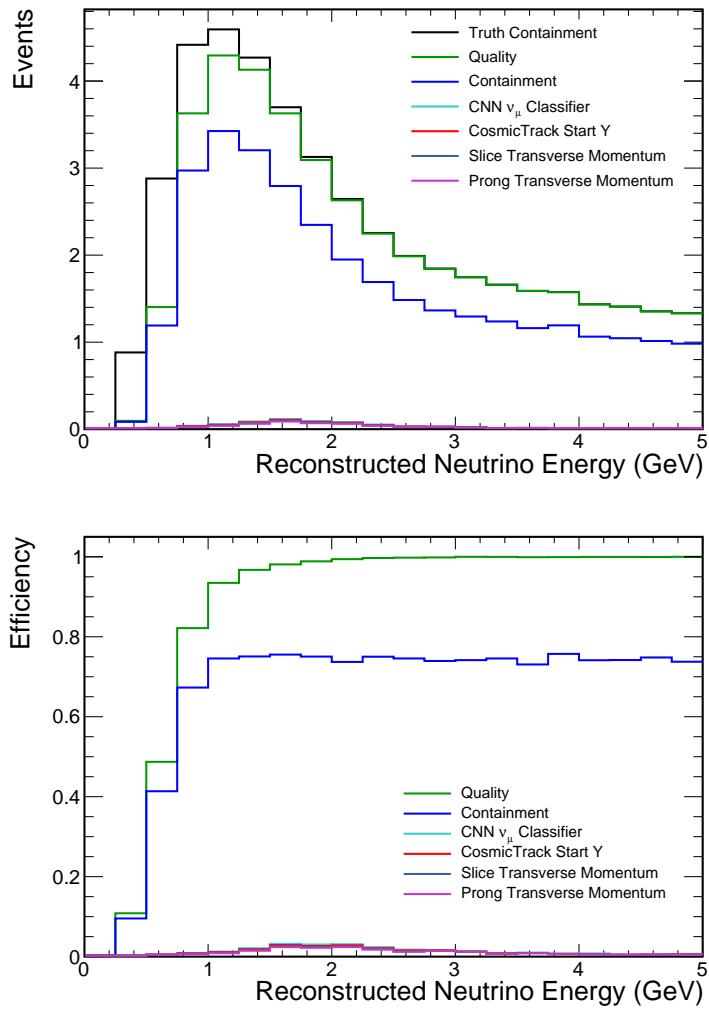


Figure 8.9: Selected ν background event spectrum and efficiency

The top pane shows the reconstructed neutrino energy spectrum for selected background events from FD MC simulation. Truth containment is the sample of events which deposit no energy outside of the detector, this serves as the denominator for the efficiency ratio shown in the bottom pane. Each subsequent spectrum adds an additional selection cut to the sample.

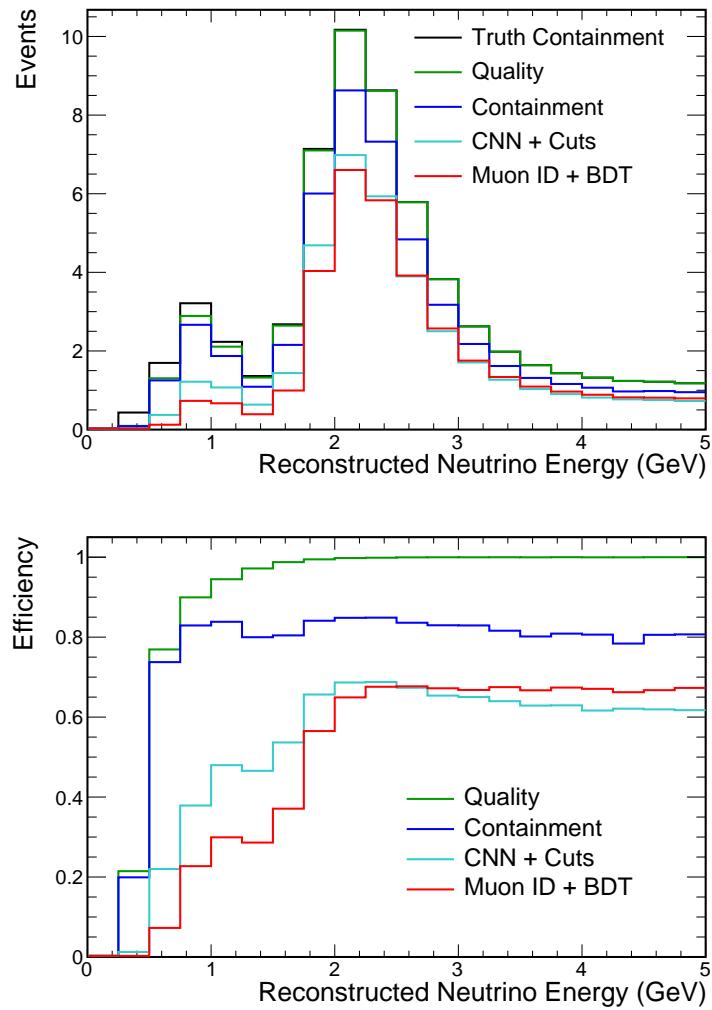


Figure 8.10: Comparison of ν_μ CC signal event spectrum and efficiency to existing approach

The top pane shows the reconstructed neutrino energy spectrum for selected ν_μ CC signal events from FD MC simulation. Truth containment is the sample of events which deposit no energy outside of the detector, this serves as the denominator for the efficiency ratio shown in the bottom pane. Each subsequent spectrum adds an additional selection cut to the sample. The ν_μ CC signal distribution has had oscillation probabilities applied based on the true energy of selected events and oscillation parameters from described in 4.6.

The two most selective spectra, *CNN + Cuts* and *Muon ID + BDT*, compare the CNN approach to the existing approach used in [68]. The existing approach relied on the Muon ID described in Section 5.4, along with a boosted decision tree [92] to reject cosmic background. The CNN selection presented here improves the selection efficiency compared to the existing approach in the low energy bins which are most sensitive to neutrino oscillation.

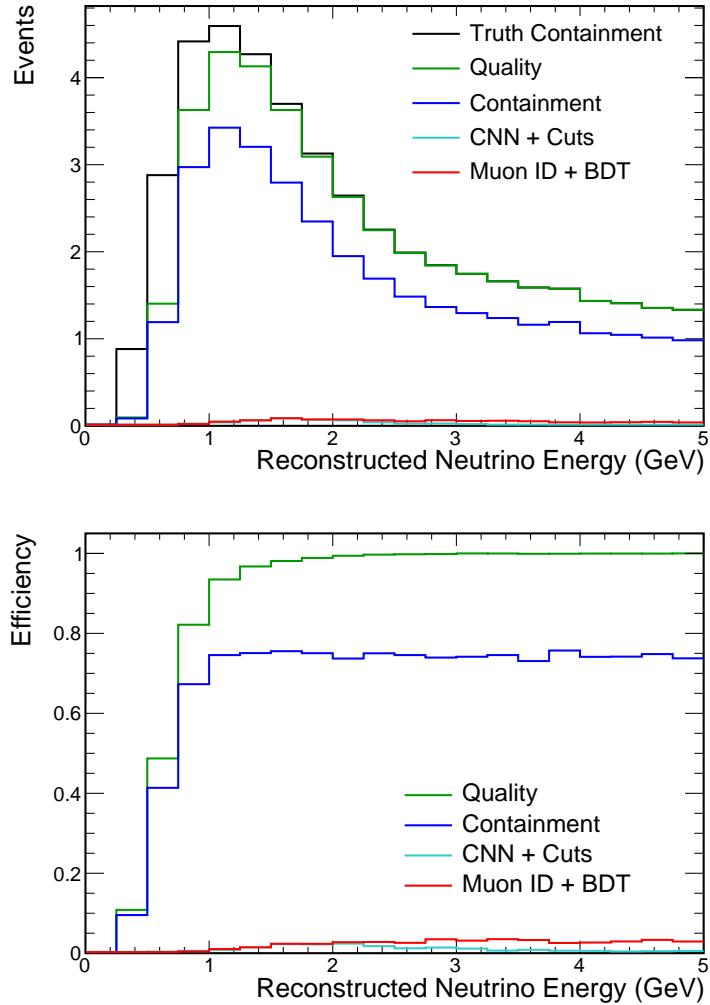


Figure 8.11: Comparison of ν background event spectrum and efficiency to existing approach

The top pane shows the reconstructed neutrino energy spectrum for selected background events from FD MC simulation. Truth containment is the sample of events which deposit no energy outside of the detector, this serves as the denominator for the efficiency ratio shown in the bottom pane. Each subsequent spectrum adds an additional selection cut to the sample.

The two most selective spectra, *CNN + Cuts* and *Muon ID + BDT*, compare the CNN approach to the existing approach used in [68]. The existing approach relied on the Muon ID described in Section 5.4, along with a boosted decision tree [92] to reject cosmic background. The CNN selection presented here improves the background rejection efficiency compared to the existing approach.

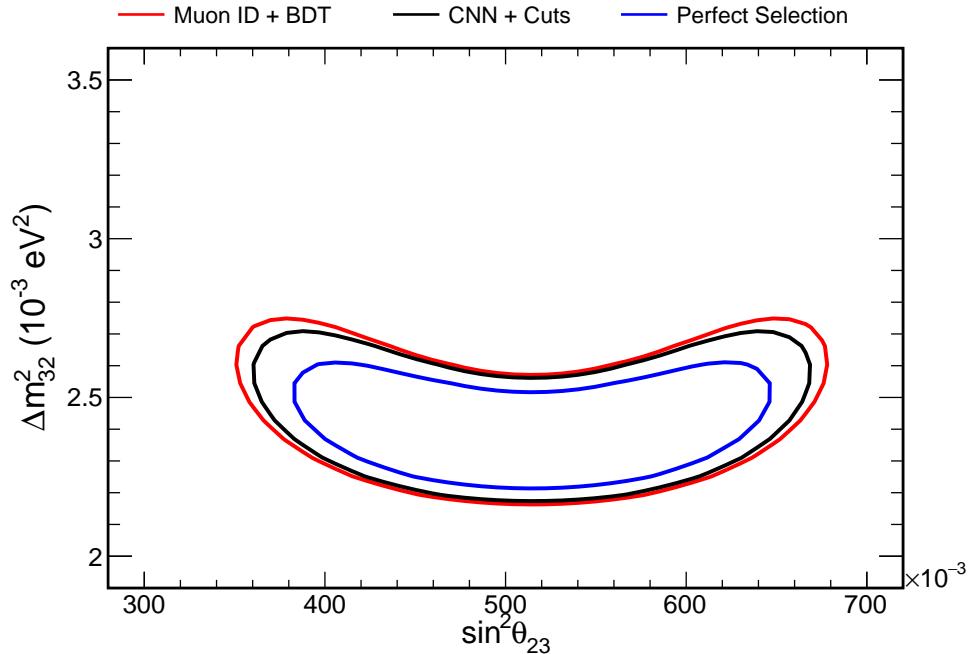


Figure 8.12: Sensitivity comparison between existing analysis approach and CNN technique

The contours shown above are expected 90% confidence limits for Δm_{32}^2 and θ_{23} using NOvA MC simulation. Confidence intervals are formed using the procedure described in Section 9.2, although it should be noted that no extrapolation is used here; A “fake data” scaled oscillated from the FD MC prediction is fit to the FD MC prediction. No systematic errors are included. The red contour uses the selection described in [68], the black is the CNN approach, and the blue is the result of a perfect selection where ν_μ CC events are selected by MC truth. Contours were generated with oscillation probabilities applied based on the true energy of selected events and oscillation parameters from described in 4.6.

8.7 ND Selected Sample

The ND selection gives an opportunity to compare the MC prediction to data. This gives us our first glimpse at detector data induced by real neutrino interactions.

The agreement between the MC prediction and the data is generally good. Reconstructed slices (described in section 5.1) show a slight bias in the data toward smaller clusters, as seen in Figure 8.13. The Muon ID data, shown in Figure 8.14, display a slight excess in the most signal-like bins relative to the MC prediction. That excess in signal translates to a normalization enhancement in the muon track length (Section 5.2.2) distribution in Figure 8.15. The singal excess is also visible in the distribution of CNN softmax output seen in Figure 8.16. The shift in Slice N_{Hit} is driven primarily by a discrepancy in the number of hits in the hadronic cluster (as described in Section 5.5), which can be seen in Figure 8.17. Calorimetric hadronic energy picks up that same shift toward lower energy relative to the MC prediction. Since the hadronic energy enters directly into reconstructed neutrino energy, seen in Figure 8.19, a shift is seen there as well.

These discrepancies, especially in the energy spectra, can be troubling from the perspective of oscillation analysis. The extrapolation procedure described in Section 9.1 will pick up the normalization and shape discrepancies and propagate them to the FD prediction, but that will be incorrect if the discrepancies are not correlated to the FD observation. Thus, we must include systematic errors which account for the discrepancies between the ND MC prediction and ND data. The discussion of systematic errors will come in Chapter 10.

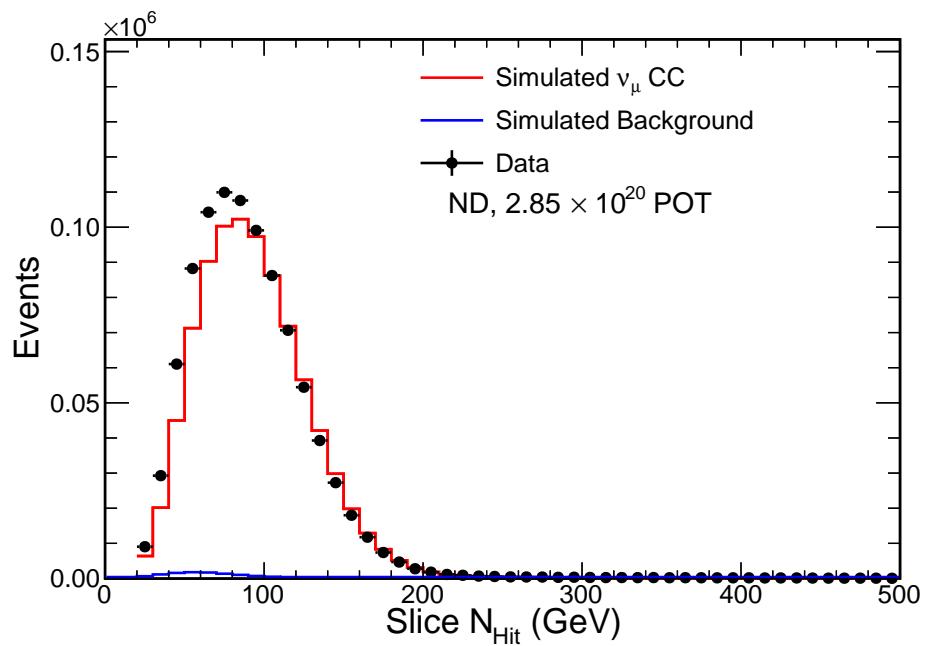


Figure 8.13: Slice N_{Hit} distribution for selected ND sample

Comparing ND data to the ND MC prediction provides a window on simulation performance. The MC prediction is shown in red, with the background component shown in blue. The selected ND data are shown as black points. Slice N_{Hit} is the number of hits in the reconstructed slice, as described in Section 5.1. The data show a slight shift toward smaller clusters.

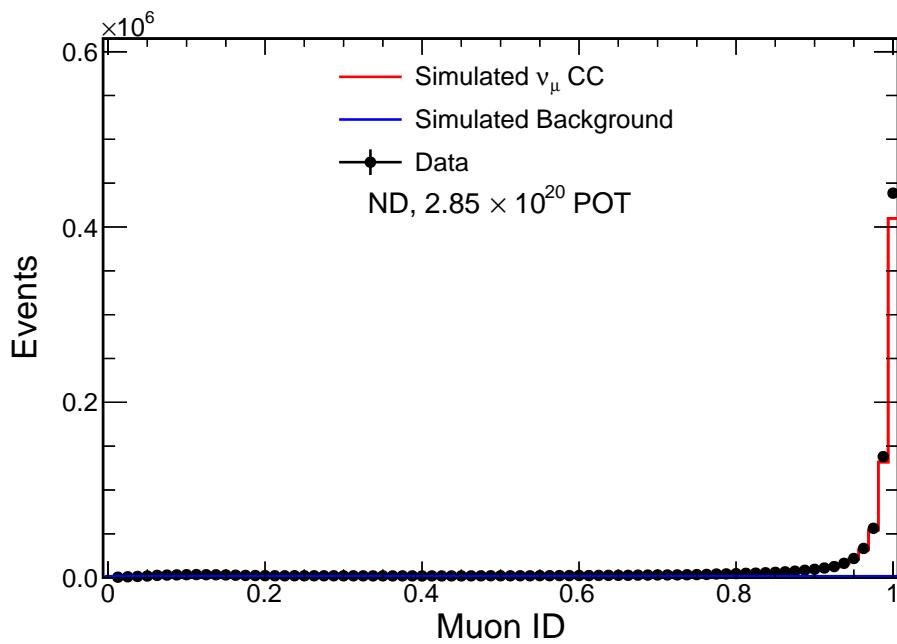


Figure 8.14: Muon ID distribution for selected ND sample

Comparing ND data to the ND MC prediction provides a window on simulation performance. The MC prediction is shown in red, with the background component shown in blue. The selected ND data are shown as black points. The Muon ID variable is the output of a kNN based on track length and likelihood estimates for observed energy deposition and multiple scattering, as described in Section 5.4.

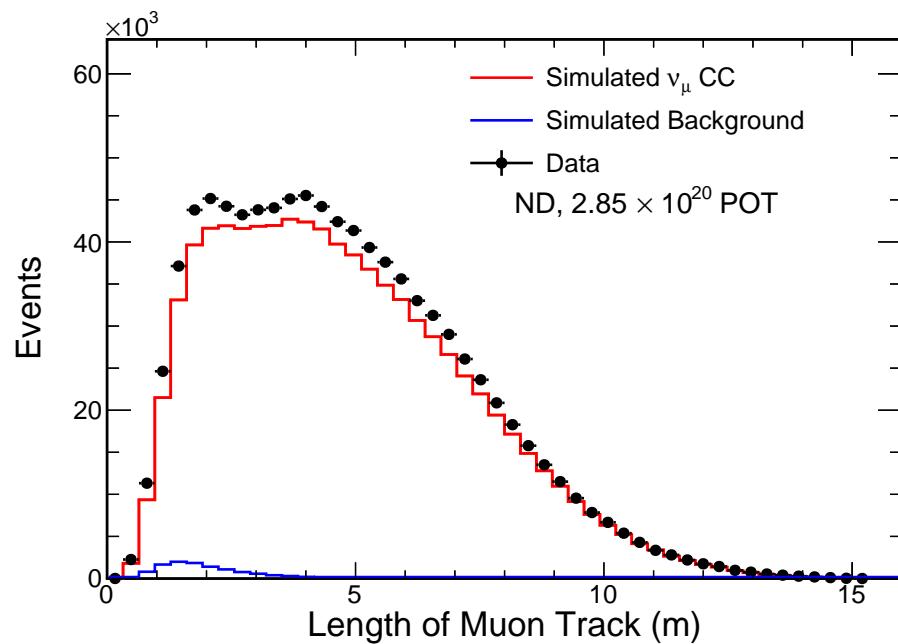


Figure 8.15: Muon track length distribution for selected ND sample
Comparing ND data to the ND MC prediction provides a window on simulation performance. The MC prediction is shown in red, with the background component shown in blue. The selected ND data are shown as black points. The muon track length is the length of the muon track (Section 5.2.2) with the highest muon ID score (Section 5.4).

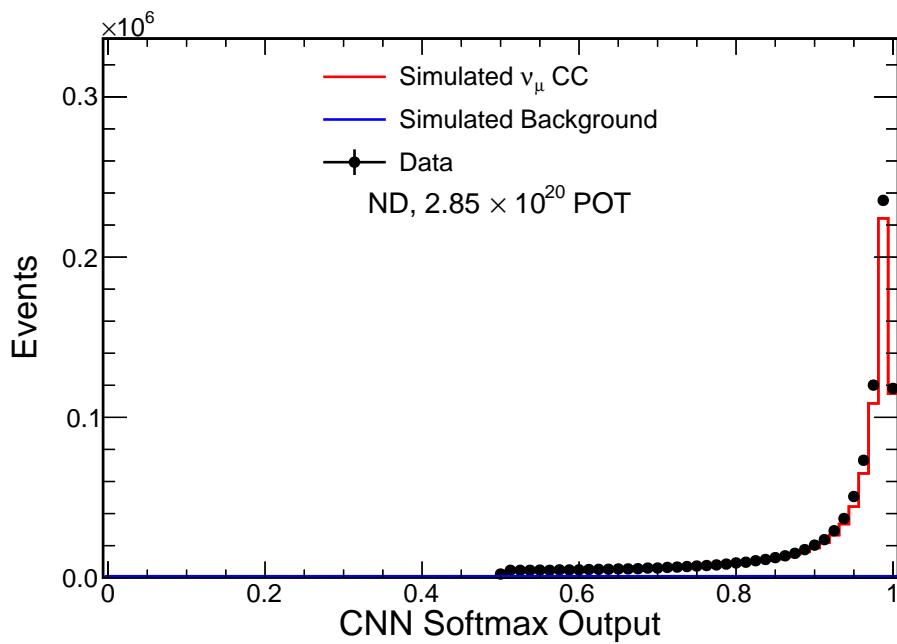


Figure 8.16: CNN classifier output distribution for selected ND sample
Comparing ND data to the ND MC prediction provides a window on simulation performance. The MC prediction is shown in red, with the background component shown in blue. The selected ND data are shown as black points. The CNN classifier shows a slight enhancement in the signal region, similar to the Muon ID variable shown in Figure 8.14.

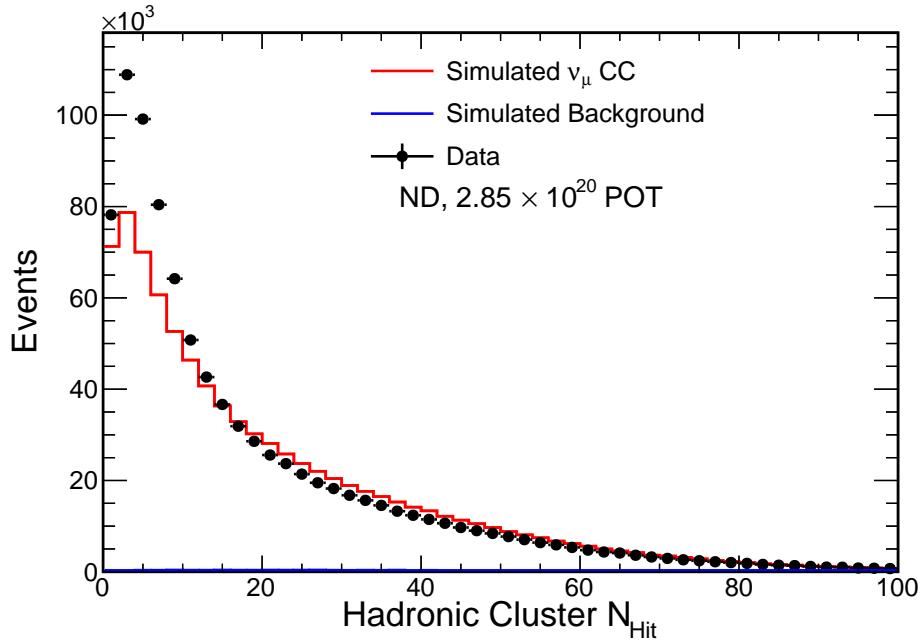


Figure 8.17: Hadronic cluster N_{Hit} distribution for selected ND sample
Comparing ND data to the ND MC prediction provides a window on simulation performance. The MC prediction is shown in red, with the background component shown in blue. The selected ND data are shown as black points. This distribution shows the number of hits in the hadronic cluster, that is, all hits in the slice which are not on the selected muon track, as described in Section 5.5. The data show a trend toward fewer hits in the hadronic cluster, which is consistent with [68].

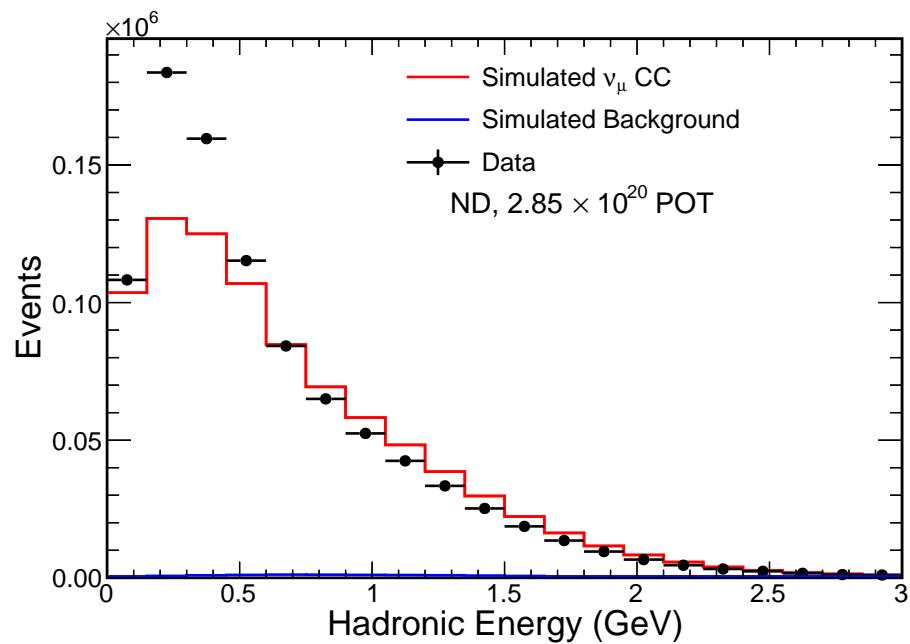


Figure 8.18: Hadronic energy distribution for selected ND sample
Comparing ND data to the ND MC prediction provides a window on simulation performance. The MC prediction is shown in red, with the background component shown in blue. The selected ND data are shown as black points. Hadronic energy is the result of the spline fit to the calorimetric energy of the hadronic cluster as described in Section 5.5. The data show a shift toward lower hadronic energy, similar to that observed in from [68].

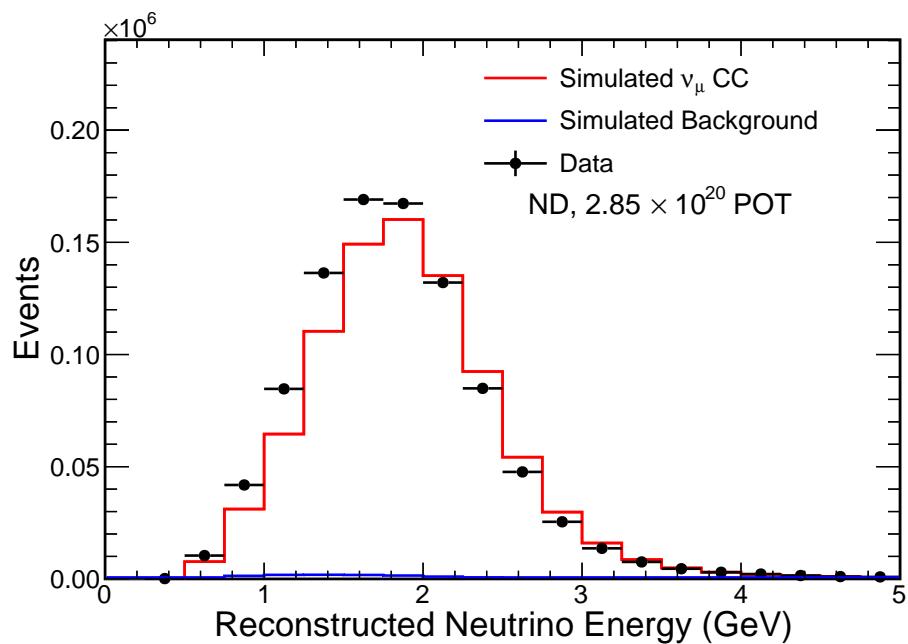


Figure 8.19: Reconstructed energy distribution for selected ND sample

Comparing ND data to the ND MC prediction provides a window on simulation performance. The MC prediction is shown in red, with the background component shown in blue. The selected ND data are shown as black points. Reconstructed energy is the sum of muon track energy and hadronic energy, as described in section 5.5. The data show a shift toward lower energy, consistent with the observation in [68].

Chapter 9

Analysis

A measurement of parameters Δm_{32}^2 and θ_{23} is obtained from a likelihood fit to the FD data, namely the reconstructed energy spectrum. Performing such a fit requires a prediction of the FD spectrum which can be varied in terms of oscillation parameters. A naive approach would be to generate a FD prediction straight from FD Monte Carlo (MC) simulation. That approach, however, would be quite sensitive to systematic uncertainties, in particular those in NuMI beam flux and neutrino cross sections. NOvA’s two-detector design is aimed precisely at mitigating the effects of such uncertainties; the ND can be used to determine the expected spectrum at the FD in a way which mostly cancels the large uncertainties the MC prediction. The procedure of generating the FD prediction in terms of the observed ND spectrum, referred to as *extrapolation*, is described in Section 9.1. The fitting procedure is described in Section 9.2.

9.1 Prediction and Extrapolation

The FD prediction is built from signal and background components. In total, the prediction is built up from ν_μ CC signal, beam-induced background components, and cosmic-ray background. As the dominant component, the ν_μ CC signal is extrapolated from ND data to take the observed spectrum into account. Beam backgrounds, on the other hand, are taken straight from MC simulation. The cosmic-ray background component is taken from data recorded outside of the time window of NuMI spills, referred to as the *out-of-time sideband*.

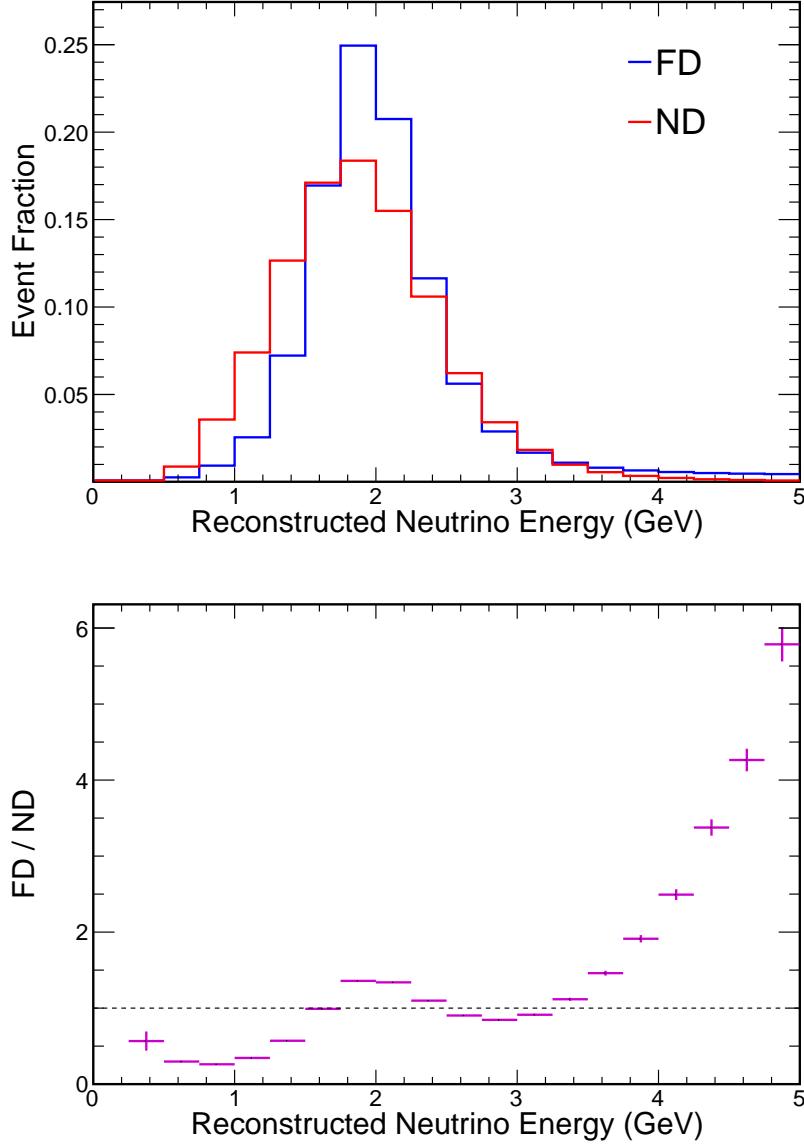


Figure 9.1: MC Prediction for ND and FD

The observed ND and FD event energy spectra display different characteristics primarily due to geometric effects. The small size of the ND gives a bias towards containment of lower energy events. Also, proximity to the NuMI source causes the ND to observe a wider range of solid angle, effectively broadening the energy peak. The top pane shows the MC prediction for both ND and FD overlaid, with each distribution scaled to unit area to highlight shape differences. The bottom pane shows the ratio of the two spectra. Note that the ratio is not used in the extrapolation procedure; instead, the ratio of the true spectra is used to transfer from ND to FD.

9.1.1 ν_μ CC Signal Extrapolation

One might expect that the ratio of the FD spectrum to the ND spectrum could be directly interpreted as an oscillation probability. This naive approach is complicated by two factors. First, the ND and FD spectra are independently shaped by various geometric and other systematic effects. Second, the imperfect resolution in energy estimation causes distortion of the measured spectra which is inconsistent across detectors. The extrapolation procedure aims to mitigate both of these factors.

Geometric effects form the majority of the discrepancy between the ND and FD spectra. Foremost, the ND is a small detector. The energy spectrum observed by ND is thus shaped by the topologies of events which can be contained within the detector. As a result, the ND spectrum is biased towards events with smaller hadronic showers and shorter muon tracks, although the steel planes of the muon catcher helps to mitigate the latter effect. Additionally, even though the ND is considerably smaller than the FD, it subtends a much larger solid angle relative to the NuMI source. As a result, it sees pions decaying over a wider range of off-axis angle and thus a broader energy spectrum than the FD.

The geometric effects which shape the ND and FD spectra is well modeled by the MC simulation. Selection effects induced by the detector size are handled by simulating with an accurately sized detector. The beam acceptance issue is covered well by the beam simulation described in Section 4.1; uncertainties in the beam situation are treated as a systematic uncertainty. The discrepancy between the FD and ND spectra is shown in Figure 9.1.

The extrapolation procedure also aims to capture the smearing effect of imperfect energy resolution independently in each detector. This is accomplished by forming a 2D histogram of reconstructed energy vs. true energy in each detector, referred to as the *reco-true matrix*. These histograms can be used to reweight between the reconstructed and true spectra in either detector [95].

Ultimately, extrapolation from ND to FD is performed in bins of true energy [95]. First, the ND data spectrum is reweighted (bin-by-bin) using the ND reco-true matrix to reproduce the observed ND spectrum in bins of true energy. Each bin in that spectrum is then reweighted in terms of FD/ND true energy ratios to produce an FD prediction in bins of true energy. Oscillation probabilities are applied to the FD true bins. The

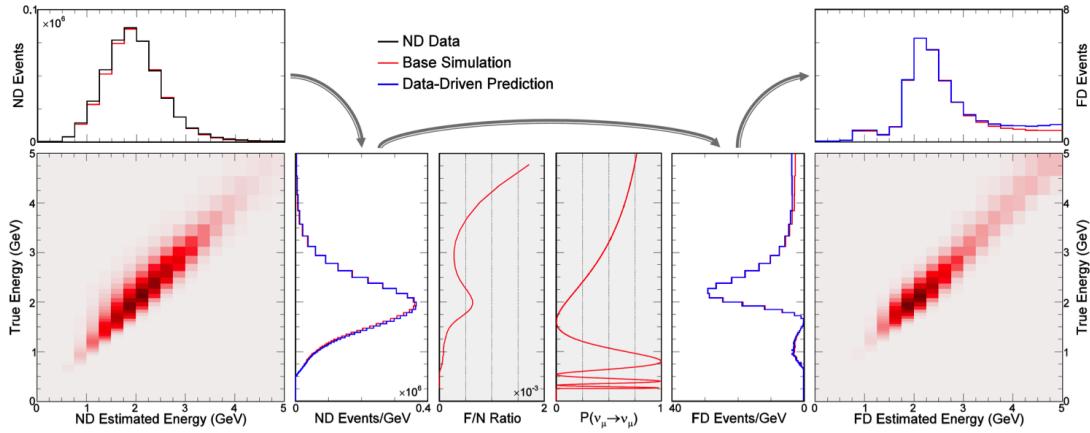


Figure 9.2: Schematic view of the extrapolation procedure

Extrapolation of ν_μ CC signal is performed in bin-by-bin. The reconstructed ND energy spectrum is first reweighted to true energy using the ND reco-true matrix. Reweighted ND true energy is then transferred to the FD spectrum using the F/N ratio. Oscillation probabilities are applied to true energy bins, then a prediction in reconstructed energy is produced using the FD true-reco matrix. Selected event spectra shown here do not correspond to those involved in this analysis, graphic only serves as a schematic overview.

bins in the oscillated true prediction are then reweighted in using the FD reco-true matrix to produce the final FD prediction in bins of true energy. A schematic view of the extrapolation procedure can be seen in Figure 9.2.

9.1.2 Background Prediction

Only the ν_μ and $\bar{\nu}_\mu$ components are given the full extrapolation treatment. The beam-induced background components – NC, ν_e , $\bar{\nu}_e$, ν_τ and $\bar{\nu}_\tau$ – are all taken directly from the MC prediction [95]. Since these components are small, taking the MC prediction does not incur a significant effect. In practice, this means that the ND MC prediction for each background component is subtracted from the ND data prior to extrapolation. Once the extrapolated FD prediction has been generated, each predicted MC background component is added.

Cosmic-ray background is estimated using selected events from the out-of-time sideband in FD data. While each NuMI beam spill spans only 10 μ s, the functional readout

window spans $450\ \mu\text{s}$. The excess data which surrounds the beam spill forms the out-of-time sideband. Sideband data is convenient for this purpose because it trivially matches the exposure of the in-time data scaling by a simple factor f :

$$f = \frac{10\ \mu\text{s}}{450\ \mu\text{s} - 10\ \mu\text{s}} = \frac{1}{44} \quad (9.1)$$

Similar to the FD beam-induced background prediction, the cosmic-ray background prediction is added directly to the extrapolated FD signal prediction [96].

9.1.3 Detector Configurations

The NOvA FD installation procedure and Data Acquisition (DAQ) system were designed with an ability to provide readout of a partial detector. It is not uncommon for the detector to operate in a configuration where one or more *blocks* (as described in section 3.2.1), of the detector are left out of the readout. Much of the data used in this analysis was recorded while the NOvA FD was still under construction, meaning that absent blocks are common in the data.

An accurate FD prediction thus relies on an accurate representation of the detector configurations in the MC simulation. This is achieved enumerating the detector configurations used in the data and aggregating the exposure (POT) recorded in each configuration. The FD MC sample is then constructed by proportionally allocating the appropriate exposure to each detector configuration.

9.2 Fitting

9.2.1 Maximum-likelihood Fitting

A binned maximum-likelihood procedure is used to measure oscillation parameters Δm_{32}^2 and θ_{23} by comparing the extrapolated FD prediction to the observed FD data [96]. The prediction is cast as a function of the oscillation parameters by applying oscillation probabilities to the FD true energy spectrum as described in Section 9.1.1. Oscillation probabilities are calculated using the full three-flavor formulation. With the exception of Δm_{32}^2 and θ_{23} , which are used in the fit, the values for all other oscillation parameters are identical to those described in section 4.6. The result of the fit is a two

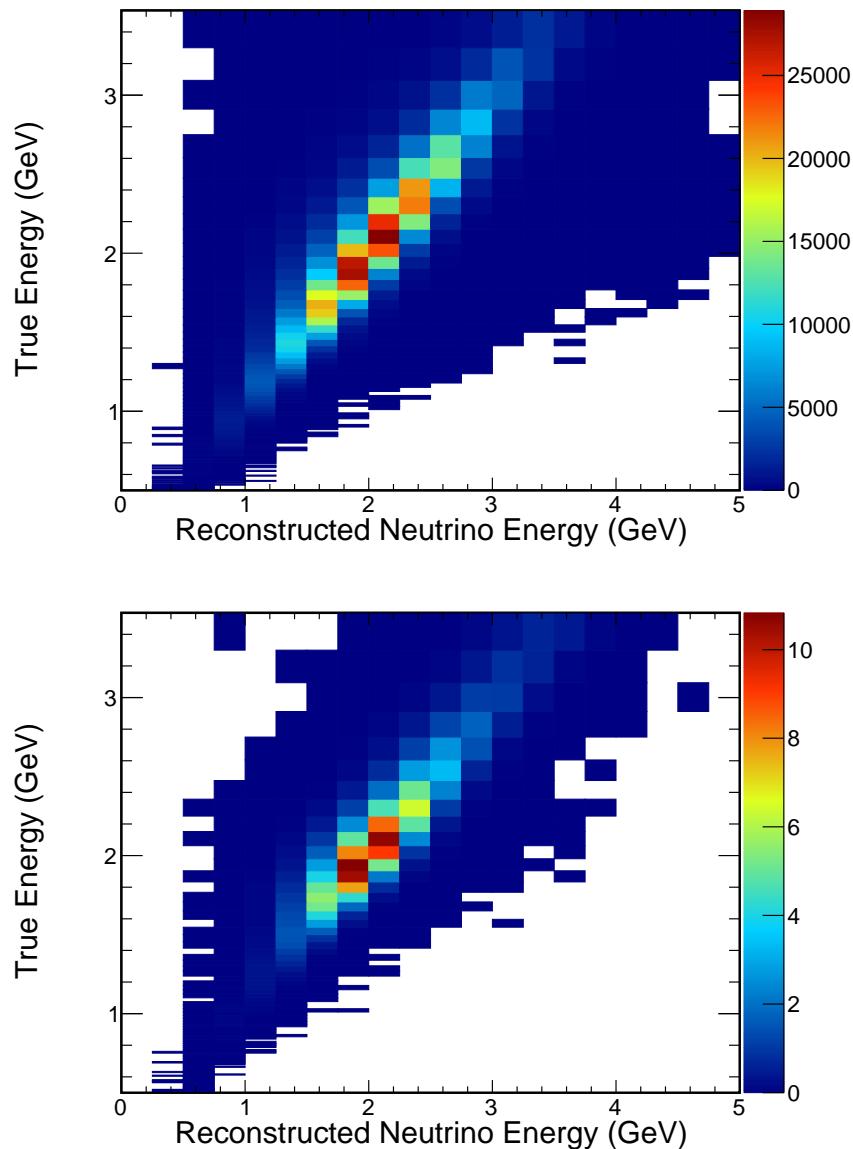


Figure 9.3: ND and FD reco-true matrices

The ND and FD reco-true matrices are used to spectra between reconstructed and true energy. The ND reco-true matrix, top, is used to convert the ND data reconstructed energy spectrum to a reweighted true energy representation. The FD reco-true spectrum is used to convert a predicted FD extrapolated true energy spectrum to a reconstructed energy spectrum.

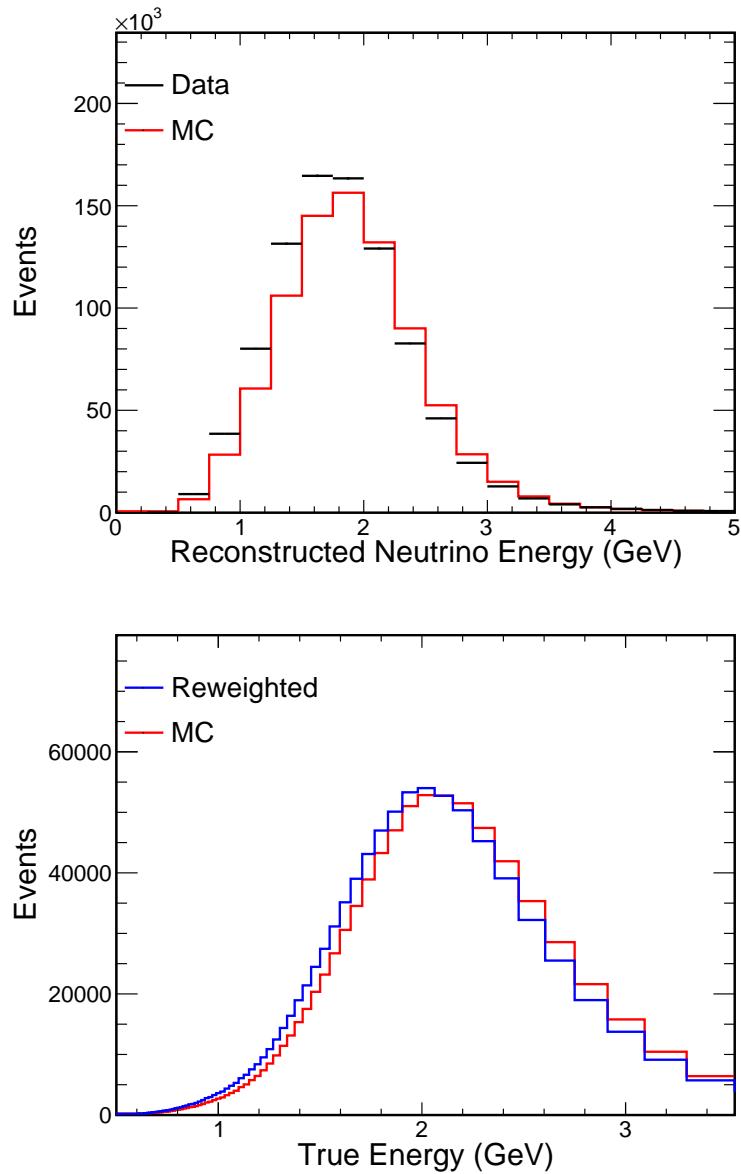


Figure 9.4: ND spectra: reconstructed, data, true and reweighted

The reconstructed energy spectrum in ND data (top, black) is used to adjust the MC prediction (bottom and top, red). The MC prediction is reweighted bin-by-bin based on the reco-true matrix (Figure 9.3) to produce the reweighted prediction, blue.

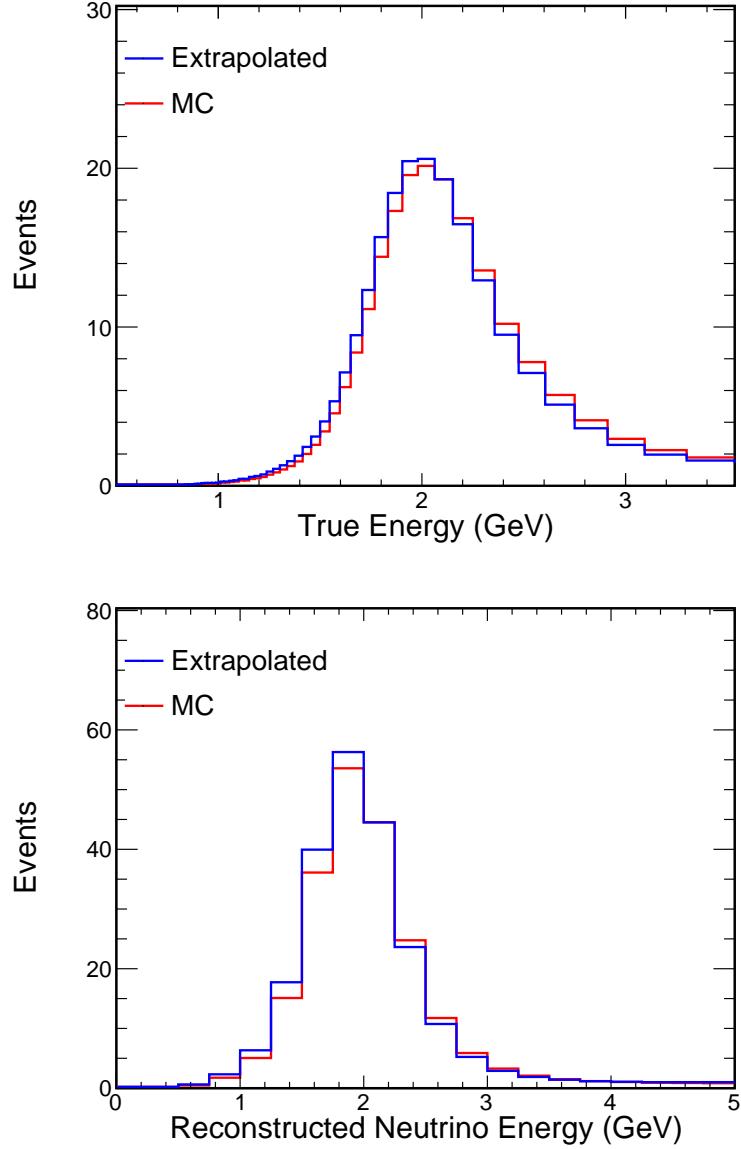


Figure 9.5: FD prediction: reconstructed, true and reweighted

The red spectra in both the top and bottom pane show the straight FD MC prediction in bins of true and reconstructed energy, respectively. In the top pane, the blue spectrum is constructed by applying the FD/ND ratio to the ND true energy spectrum shown in Figure 9.4. The corresponding blue spectrum in the bottom pane has been reweighted based on the reco-true matrix in Figure 9.3 to produce the final FD prediction. In this case, no oscillation probabilities have been applied to the true bins, thus showing the predicted spectrum for the case that neutrino oscillation did not occur.

dimensional confidence interval for the parameters $|\Delta m_{32}^2|$ and $\sin^2(\theta_{23})$.

The likelihood function used is the log-likelihood for Poisson distributed data [42]:

$$\chi^2 = 2 \sum_i e_i - o_i + o_i \ln \left(\frac{o_i}{e_i} \right), \quad (9.2)$$

where the sum runs over bins i , e_i is the predicted number of events in that bin, and o_i is the observed number of events in FD data. Note that maximizing the likelihood actually corresponds to minimizing χ^2 . Confidence levels can be drawn in the two dimensional dimensional space based on $\Delta\chi^2$ relative to the minimum value. The critical $\Delta\chi^2$ values for 68% and 90% confidence intervals in the case of both 1D and 2D contours can be found in [42].

Systematic uncertainties are included in the fit by adding a term to the likelihood function as follows [42]:

$$\chi^2 = 2 \sum_i e_i - o_i + o_i \ln \left(\frac{o_i}{e_i} \right) + \sum_j \frac{s_j^2}{\sigma_j^2}. \quad (9.3)$$

The sum over j is over systematic contributions. Each systematic uncertainty is allowed to shift each bin, with σ_j reflecting our prior knowledge of the allowed freedom in that systematic shift. The value s_j is determined by a process known as *marginalization*. At each grid point for Δm_{32}^2 and θ_{23} , the fitter is allowed to scan over values of the systematic shift; s_j is taken the value of the shift which minimizes χ^2 . The sum over j provides a penalty for large shifts, which disfavors values which deviate strongly from our prior knowledge of the systematic parameter. The result of the marginalization process is to generally produce a better fit at each grid point, effectively reducing the depth of the χ^2 surface. A shallower surface results in larger confidence intervals, which fits with the intuition that systematic uncertainties reduce the confidence in a measurement.

9.2.2 The Feldman-Cousins Procedure

An alternative fitting strategy is employed based on [97]. This approach uses a Frequentist approach which is concerned with proper *coverage*; that is, the notion that a 90% confidence interval should include the true values in 90% of trials if the experiment were

to be repeated many times. At each grid point, a large number of pseudo-experiments are generated. Each pseudo-experiment is based on the oscillated MC prediction, but with Poisson distributed fluctuations in each bin. The χ^2 is determined for each pseudo-experiment and 90th percentile forms the critical value at each grid point. When fitting the final data, the χ^2 is compared to the grid points. Points in the data grid where the χ^2 exceeds the critical value at that point grid are excluded from the contour, while those which fall below are included. In each pseudo-experiment, values of the systematic shift are chosen from their true distribution, rather than assuming a Gaussian shape. This method ensures proper coverage by construction, even in the presence of large deviations from Gaussian behavior.

Chapter 10

Systematic Uncertainties

The extrapolated prediction depends on our knowledge of particle propagation and detector behavior, mainly through the MC simulation. Such dependencies introduce systematic uncertainties when that knowledge is incomplete. The design of NOvA and its reconstruction chain is attempts to mitigate systematic uncertainties. For instance, a two detector design helps cancel flux and cross section uncertainties; estimating muon energy from range helps eliminate some dependence on calibration accuracy. Despite these efforts, the degree to which systematic effects are mitigated must be tested and residual behavior must be incorporated into the analysis framework.

10.1 Treatment of Systematic Uncertainties

Systematic uncertainties are incorporated by allowing event records to be altered in by an arbitrary procedure before they are handled by the selection and analysis framework. This process allows the analysis framework to produce selected spectra as if each event were different in some systematic fashion. Each event can also be assigned an alternate weight based on entries in the event record, such as MC truth information or the output of the reconstruction algorithms. For each systematic effect, alternate versions of the FD prediction are generated corresponding to $\pm 1\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$ shifts [96].

The event records which are allowed to be shifted are populated by the output of the reconstruction algorithms, not the raw hit-by-hit detector readout itself. Unfortunately, many of the systematic effects must be assessed by altering the hit-by-hit detector

readout. In these cases, alternative MC samples are generated which correspond to some systematic change in the behavior of the simulation and/or reconstruction. The effect of that systematic change is then and parametrized in terms of some entry in the event record which captures the behavior [96].

For each systematic effect, the shape of the $\pm 1\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$ shifts is interpolated using a cubic spline [57]. The interpolation gives the fitter a smooth profile for marginalization. In the Feldman-Cousins procedure [97] (described in Section 9.2.2) that shape is used for sampling the uncertainties in the pseudo-experiments [96].

10.1.1 Relative vs. Absolute Uncertainties

In parametrizing systematic uncertainties between two detectors, there is a distinction between uncertainties which are correlated between detectors and those which have freedom to move independently. Neutrino cross section uncertainties, for instance, should be correlated between the ND and FD; the physics which governs neutrino interactions is not expected to change between Fermilab and northern Minnesota. Uncertainty in the calorimetric energy scale, however, might not be correlated between the two detectors. There are systematic effects in the MC simulation and calibration procedure which we have no evidence to believe behave similarly across detectors. In this analysis, systematic uncertainties which are correlated between are referred to as *absolute* uncertainties as opposed to those which could be uncorrelated, which are called *relative* uncertainties. Parameterized absolute uncertainties adjust the ND and FD prediction simultaneously by the same amount, Relative uncertainties, on the other hand, adjust the prediction by the same magnitude, but in opposite directions.

10.2 Flux Uncertainty

The NuMI beam simulation is used to predict the ν flux in both the ND and FD, as described in Section 4.1. Ultimately, these simulations enter in the analysis through the extrapolation of the ND spectrum to produce the FD prediction. It is thus necessary to estimate the systematic uncertainties in the flux prediction.

Flux uncertainties are broken down into two categories; hadron-production uncertainties and beam-transport uncertainties [98]. The hadron production uncertainty

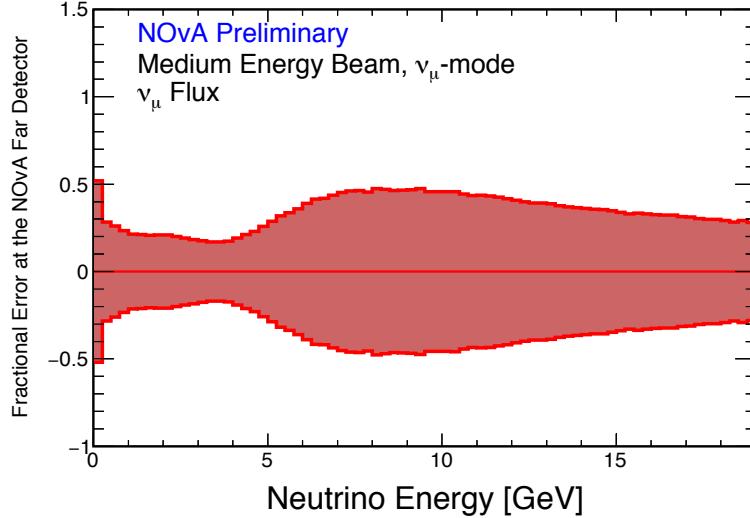


Figure 10.1: FD hadron production uncertainty band

The systematic uncertainty in the flux prediction has been estimated using an alternative configuration of the FLUKA generator which has been tuned to minimize the discrepancy between NA49 [54] data and MC simulation. This alternative tune leads to a relatively large uncertainty band in the flux prediction, shown above.

represents incompleteness in the modeling of hadrons produced in the NuMI target. The beam transport uncertainties, on the other hand, relate to mis-modeling of target facility components (horns, decay pipe, etc.) which could affect the flux prediction.

Hadron-production uncertainty has been estimated by comparing the standard FLUKA [49] cross-section model configuration to an alternative tune which resolves discrepancies between NA49 [54] data and MC simulation [98]. The shift between the output of the two configurations was used to produce band of systematic uncertainty as a function of true neutrino energy in both detectors. The FD uncertainty band can be seen in Figure 10.1, ND in Figure 10.2, and the ratio of the two in Figure 10.3. While the FD and ND bands are relatively large, the systematic uncertainty cancels significantly in the ratio.

The beam-transport uncertainties have been estimated in a similar fashion. Elements of the detector hall have been alternatively configured in order to estimate the resulting flux uncertainty [98]. The following alternative configurations were tested:

- Horn current shifted by ± 1 kA (0.5%)

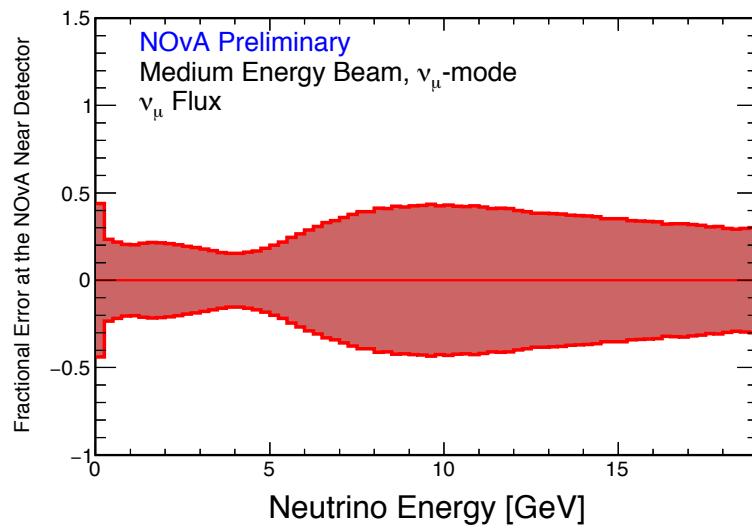


Figure 10.2: ND hadron production uncertainty band

The systematic uncertainty in the flux prediction has been estimated using an alternative configuration of the FLUKA generator which has been tuned to minimize the discrepancy between NA49 [54] data and MC simulation. This alternative tune leads to a relatively large uncertainty band in the ND flux prediction, shown above.

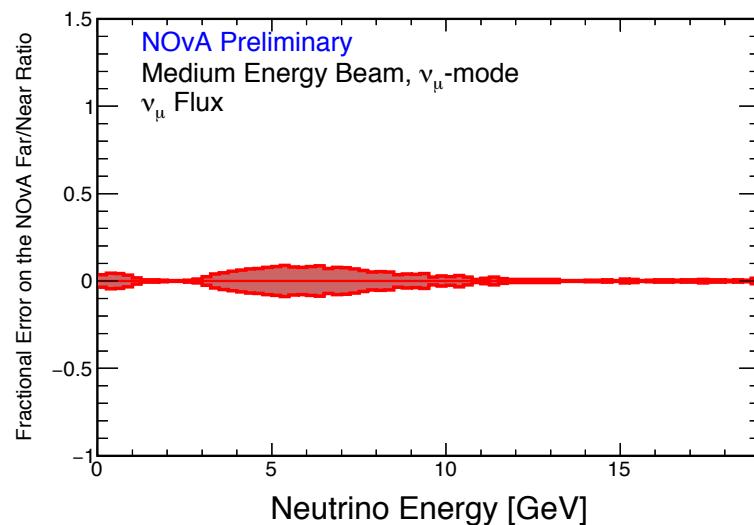


Figure 10.3: Hadron production uncertainty in FD/ND ratio

The systematic uncertainty in the flux prediction has been estimated using an alternative configuration of the FLUKA generator which has been tuned to minimize the discrepancy between NA49 [54] data and MC simulation. While the uncertainty in the FD and ND are respectively large, the correlation between the two leads to significant cancellation.

- Horn 1 position shifted by ± 2 mm in both transverse dimensions
- Horn 1 position shifted by ± 2 mm in both transverse dimensions
- Current distribution in horn switched from linear to exponential with skin depth 0.77 cm
- Beam position on target shifted by ± 0.5 mm in both transverse dimensions
- Beam spot size on target shifted by ± 0.2 mm in both transverse dimensions
- Target position shifted by ± 2 mm in longitudinal dimension

These alternative configurations have been used to produce systematic uncertainty bands analogous to the hadron production bands shown in Figures 10.1, 10.2, and 10.3.

For both the ND and FD, the systematic uncertainty bands have been combined to produce one total flux uncertainty band [98]. The band is used to reweight the true energy spectrum in forming the FD prediction. The effect of the flux uncertainty reweighting can be seen in Figure 10.4.

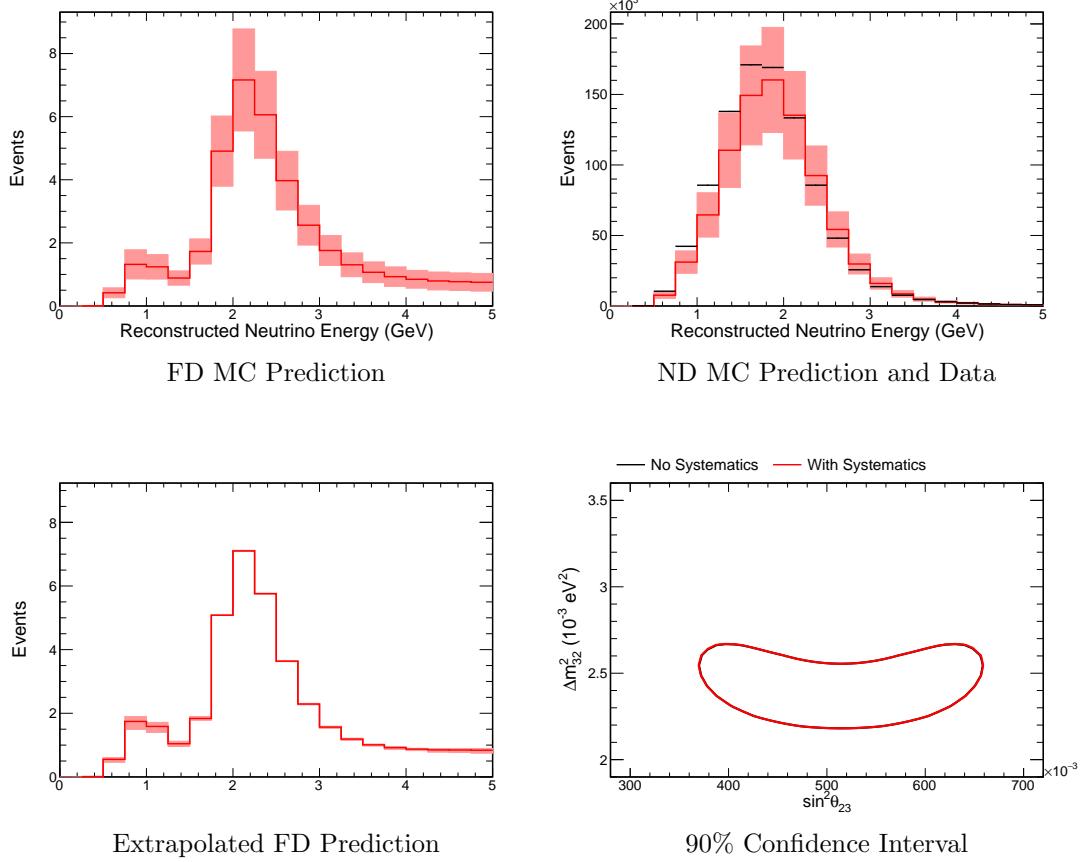


Figure 10.4: Systematic effect of combined flux uncertainty
 Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

10.3 Cross Section Uncertainty

GENIE includes a built in utility for reweighting events based on uncertainties in the free parameters of its cross section models [50]. For each free parameter, event weights are determined by recalculating the cross section with that parameter altered; the weight is taken to be the ratio of the altered cross section to the altered one. At analysis time, selected events in the ND and FD MC predictions can be reweighted on an event-by-event basis. There are 67 parameters which can be adjusted independently which are listed in Table 10.2.

Parameter	Description
AGKY_pT1pi	AGKY transverse momentum in single pion states
AGKY_xF1pi	AGKY Feynman x for single pion states
AhtBY	Higher-twist parameter(A_{HT}) in Bodek-Yang model
AhtBYshape	Higher-twist parameter(A_{HT}) in Bodek-Yang model, shape only
BR1eta	Branching ratio for single- η resonance decays
BR1gamma	Branching ratio for radiative resonance decays
BhtBY	Higher-twist parameter (B_{HT}) in Bodek-Yang model
BhtBYshape	Highter-twist parameter (B_{HT}) in Bodek-Yang model, shape only
CCQE Mom Distro FG to SF	CCQE Nucleon Momentum Distribution
CCQE Pauli Sup Via KF	CCQE Pauli suppression (changes Fermi level k_F)
CV1uBY	C_{V1u} valence GRV98 PDF correction in Bodek-Yang Model
CV1uBYshape	C_{V1u} valence GRV98 PDF correction in Bodek-Yang Model, shape only
CV2uBY	C_{V2u} valence GRV98 PDF correction in Bodek-Yang Model
CV2uBYshape	C_{V2u} valence GRV98 PDF correction in Bodek-Yang Model, shape only

Continued on next page

Table 10.1 – continued from previous page

Parameter	Description
DISNuclMod	DIS nuclear modification (shadowing, anti-shadowing, EMC)
EtaNCEL	Strange axial form factor for NC elastic
FormZone	Formation zone
FrAbs_N	Nucleon absorption probability
FrAbs_pi	Pion absorption probability
FrCEEx_N	Nucleon charge exchange probability
FrCEEx_pi	Pion charge exchange probability
FrElas_N	Nucleon elastic reaction probability
FrElas_pi	Pion elastic reaction probability
FrInel_N	Nucleon inelastic reaction probability
FrInel_pi	Pion inelastic reaction probability
FrPiProd_N	Nucleon-pion production probability
FrPiProd_pi	Pion-pion production probability
MFP_N	Nucleon mean free path
MFP_pi	Pion mean free path
MaCCQE	Axial mass for CC quasi-elastic
MaCCQEshape	Axial mass for CC quasi-elastic, shape only
MaCCRES	Axial mass for CC resonance production
MaCCRESShape	Axial mass for CC resonance production, shape only
MaCOHpi	Axial mass for CC coherent pion production
MaNCEL	Axial mass for NC elastic
MaNCRES	Axial mass for NC resonance production
MaNCRESShape	Axial mass for NC resonance production, shape only
MvCCRES	Vector mass for CC quasi-elastic
MvCCRESShape	Vector mass for CC quasi-elastic, shape only
MvNCRES	Vector mass for CC resonance production

Continued on next page

Table 10.1 – continued from previous page

Parameter	Description
MvNCRESshape	Vector mass for CC resonance production, shape only
NormCCQE	CCQE normalization
NormCCQEenu	CCQE normalization
NormCCRES	CC resonance production, normalization
NormDISCC	CC DIS production, normalization
NormNCRES	NC resonance production, normalization
R0COHpi	Nuclear size parameter controlling pion absorption in Rein-Sehgal model
RnubarunuCC	$\nu\bar{\nu}$ CC ratio
RvbarnCC1pi	Non resonance background in $\bar{\nu}nCC1\pi$ interactions
RvbarnCC2pi	Non resonance background in $\bar{\nu}nCC2\pi$ interactions
RvbarnNC1pi	Non resonance background in $\bar{\nu}nNC1\pi$ interactions
RvbarnNC2pi	Non resonance background in $\bar{\nu}nNC2\pi$ interactions
RvbarpCC1pi	Non resonance background in $\bar{\nu}pCC1\pi$ interactions
RvbarpCC2pi	Non resonance background in $\bar{\nu}pCC2\pi$ interactions
RvbarpNC1pi	Non resonance background in $\bar{\nu}pNC1\pi$ interactions
RvbarpNC2pi	Non resonance background in $\bar{\nu}pNC2\pi$ interactions
RvnCC1pi	Non resonance background in $\nu nCC1\pi$ interactions
RvnCC2pi	Non resonance background in $\nu nCC2\pi$ interactions
RvnNC1pi	Non resonance background in $\nu nNC1\pi$ interactions
RvnNC2pi	Non resonance background in $\nu nNC2\pi$ interactions
RvpCC1pi	Non resonance background in $\nu pCC1\pi$ interactions
RvpCC2pi	Non resonance background in $\nu pCC2\pi$ interactions
RvpNC1pi	Non resonance background in $\nu pNC1\pi$ interactions
RvpNC2pi	Non resonance background in $\nu pNC2\pi$ interactions
Theta_Delta2Npi	Pion angular distribution in Rein-Sehgal

Continued on next page

Table 10.1 – continued from previous page

Parameter	Description
VecCCQEshape	Choice of CCQE vector form factors (BBA05/Dipole), shape only

Table 10.2: All reweightable free parameters in GENIE

GENIE includes a built-in utility for reweighting events to account for cross section uncertainties [50]. Free parameters are adjusted in order to calculate an alternative cross section; weights are formed by the ratio of the alternative cross section to the original cross section. This table shows the complete list of parameters which can be varied in order to produce event weights.

Redundant Parameter	Counterpart
MaCCQEshape	MACCQE
MaCCRESshape	MaCCRES
MvCCRESshape	MvCCRES
MaNCRESshape	MaNCRes
MvNCRESshape	MvNCRes
AhtBYshape	AhtBY
BhtBYshape	BhtBY
CV1uBYshape	Cv1uBY
CV2uBYshape	CV2uBY
NormCCQE	MaCCQE
NormCCQEenu	MaCCQE
NormCCRES	MaCCRES
NormNCRES	MaNCRes

Table 10.3: Redundant knobs in GENIE Reweight

Certain parameters in GENIE are paired with shape-only or normalization-only counterparts. These counterparts are excluded in this analysis since they are redundant. This table shows the list of parameters which are excluded and the parameter with which they are redundant.

Certain parameters are matched with shape-only and/or normalization-only counterparts to separately illustrate those effects. Since these effects would be redundant, these parameters are excluded from the reweighting process. All redundant parameters are listed in Table 10.3.

Marginalizing over the large number of free parameters in GENIE would be an expensive operation. Rather than including all parameters, only the ones with the most significant effect were given an independent treatment. These parameters are MaCCQE, MaCCRES, MvCCRES, MaNCRES, and MaNCEL. The effect of MaCCQE on the prediction and confidence intervals can be seen in Figure 10.5, MaCCRES in Figure 10.6, MvCCRES in Figure 10.7, MaNCRES in Figure 10.8, and MaNCEL in Figure 10.10.

The remaining parameters were grouped together by summing the $\pm 1\sigma$ bin shifts in quadrature. Between \pm shifts, the larger of the two was included in the sum. Sums were determined separately for signal and background spectra and for each detector. These sums can be visualized in Figure 10.11 for the FD and Figure 10.12 for the ND. The

effect of the combined systematic uncertainty on the prediction and confidence intervals can be seen in Figure 10.13.

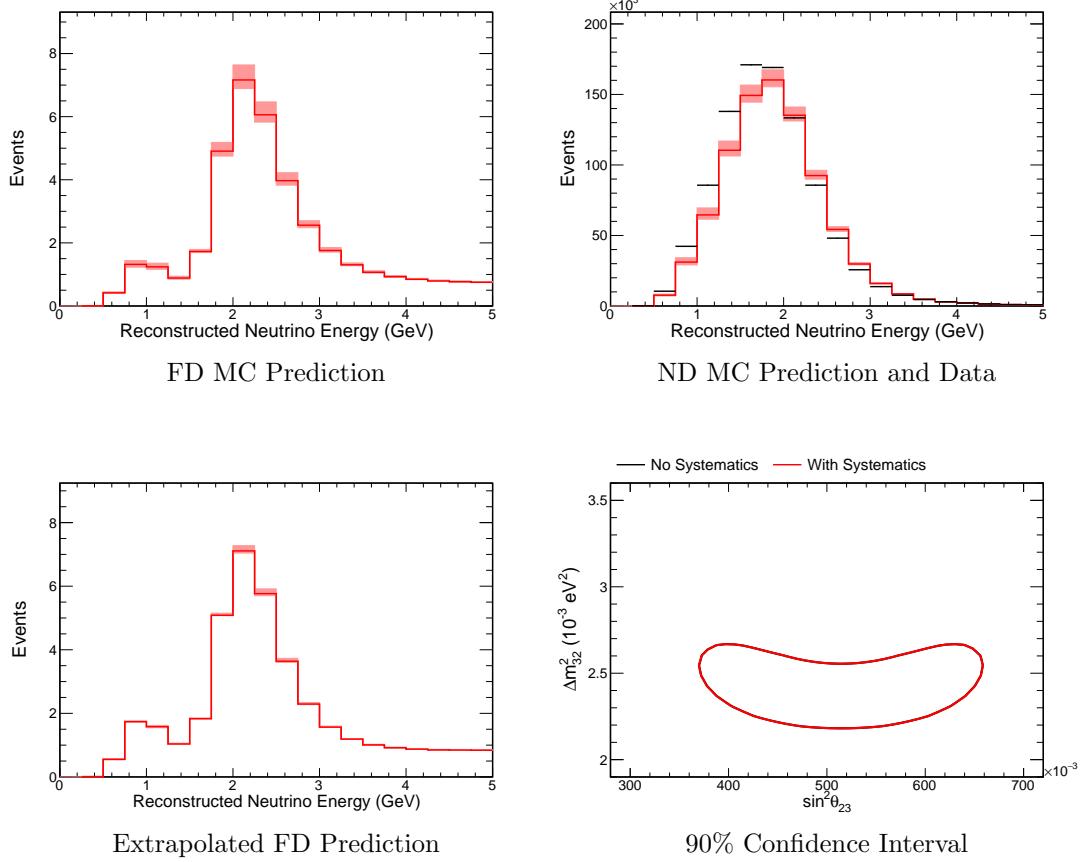


Figure 10.5: Systematic effect of MaCCQE uncertainty

Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

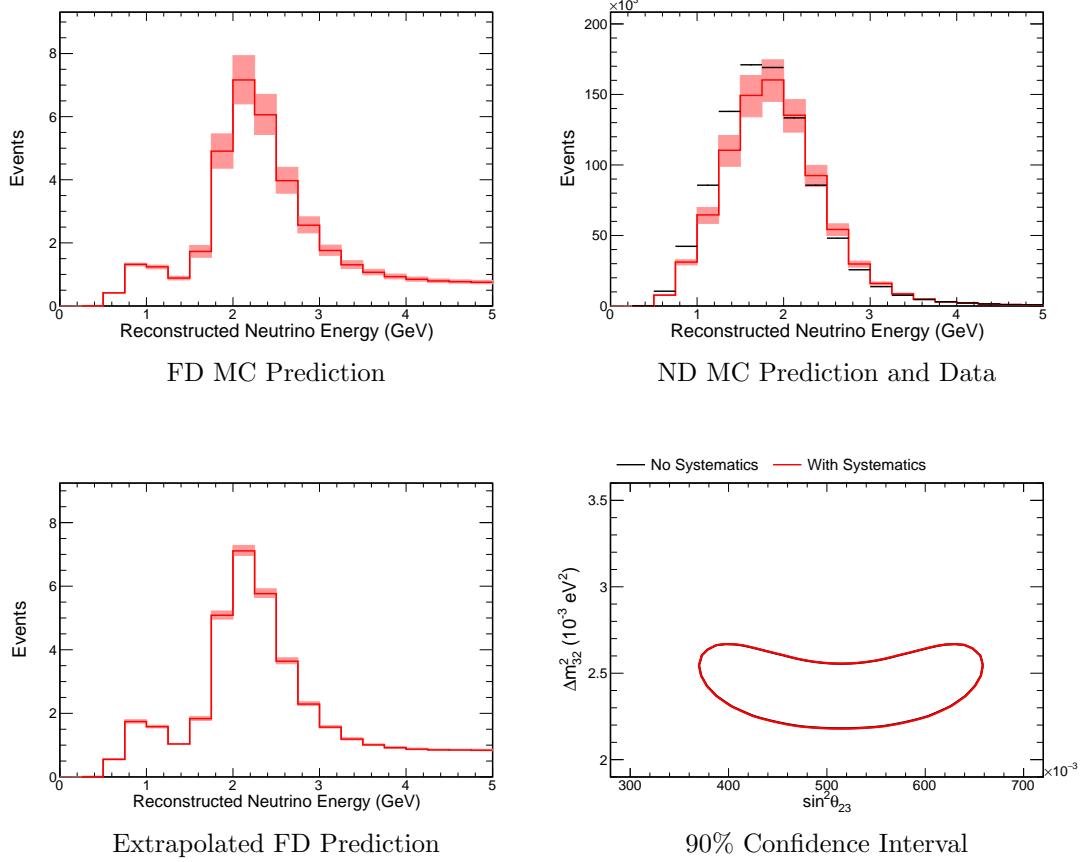


Figure 10.6: Systematic effect of MaCCRES uncertainty

Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

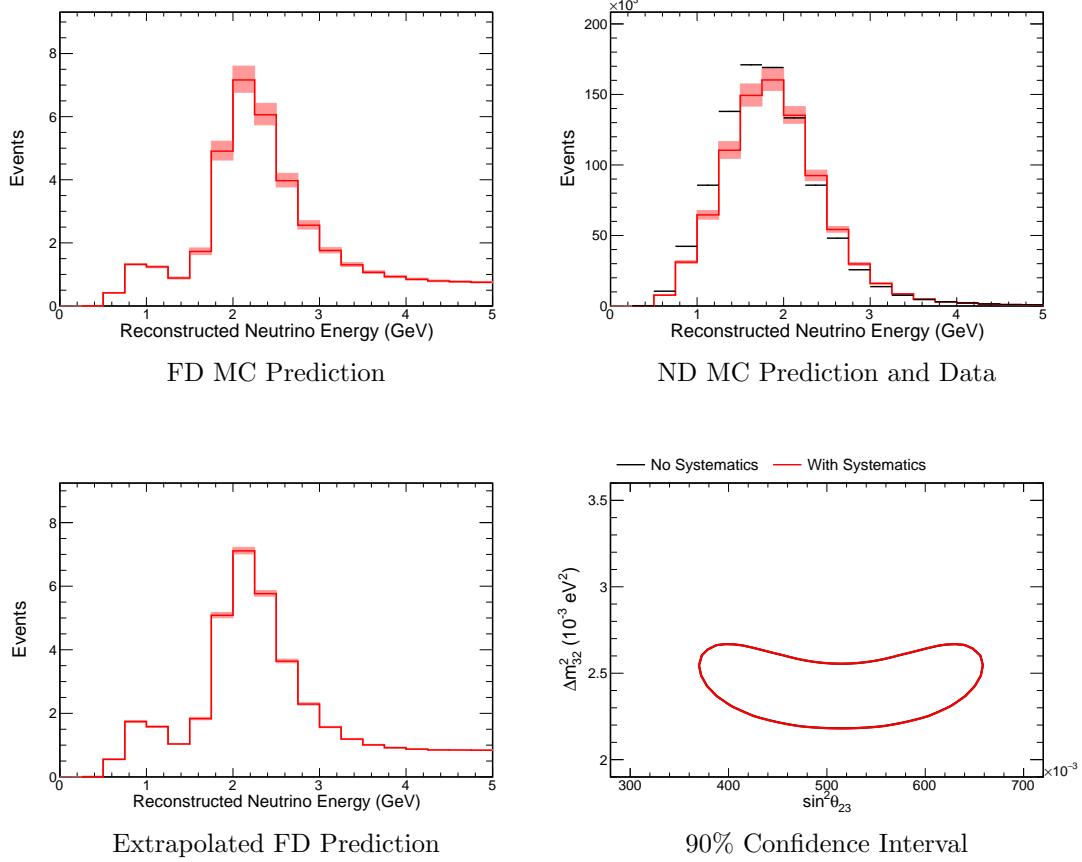


Figure 10.7: Systematic effect of MvCCRES uncertainty

Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

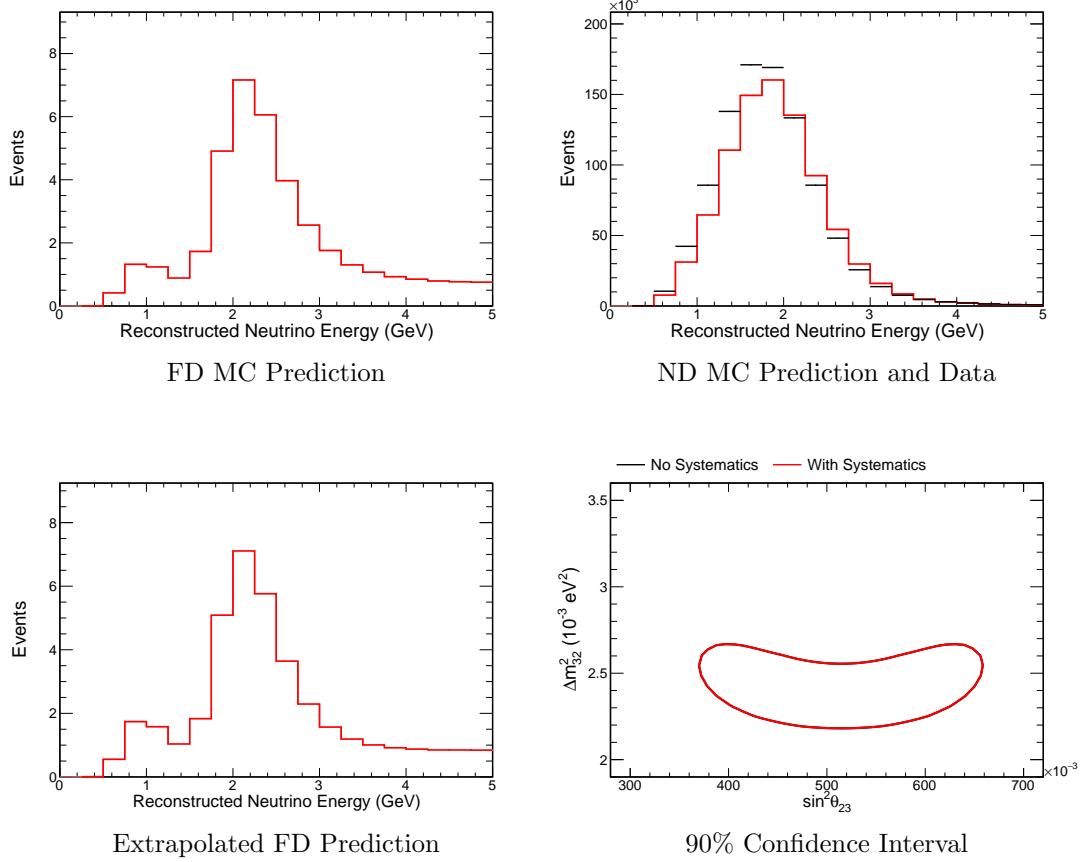


Figure 10.8: Systematic effect of MaNCEL uncertainty

Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

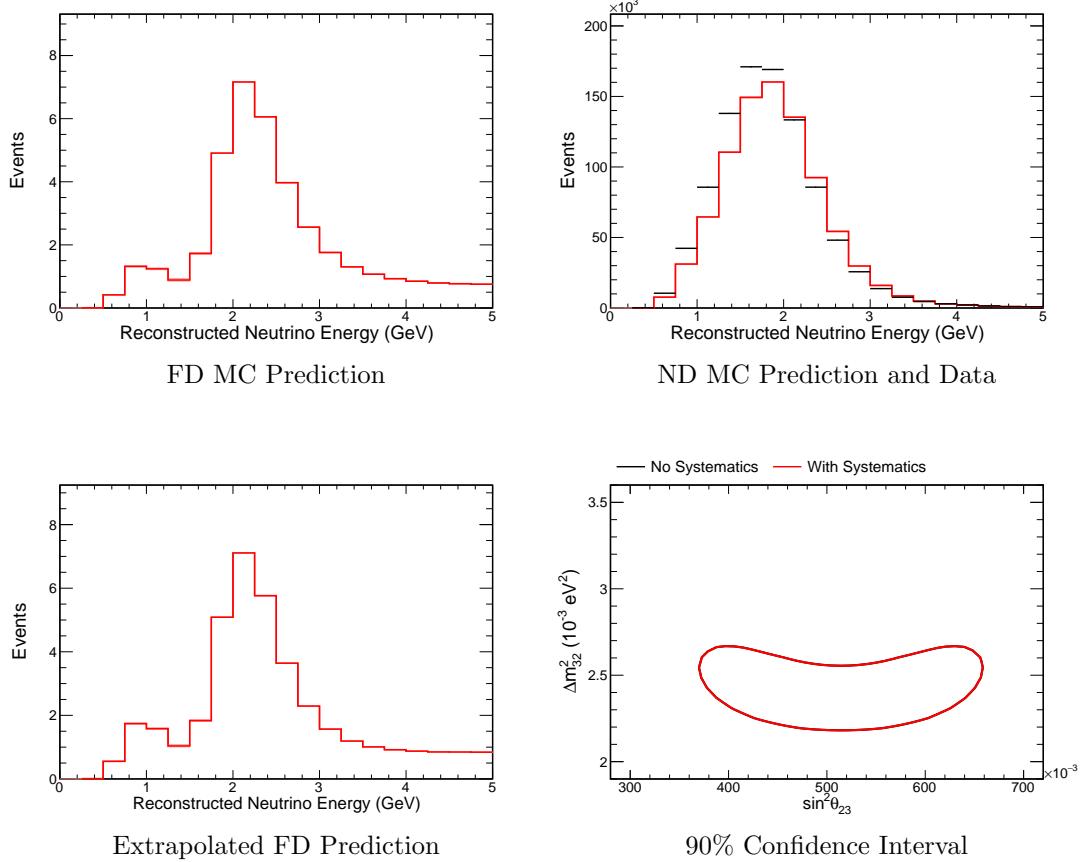


Figure 10.9: Systematic effect of MaNCRES uncertainty

Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

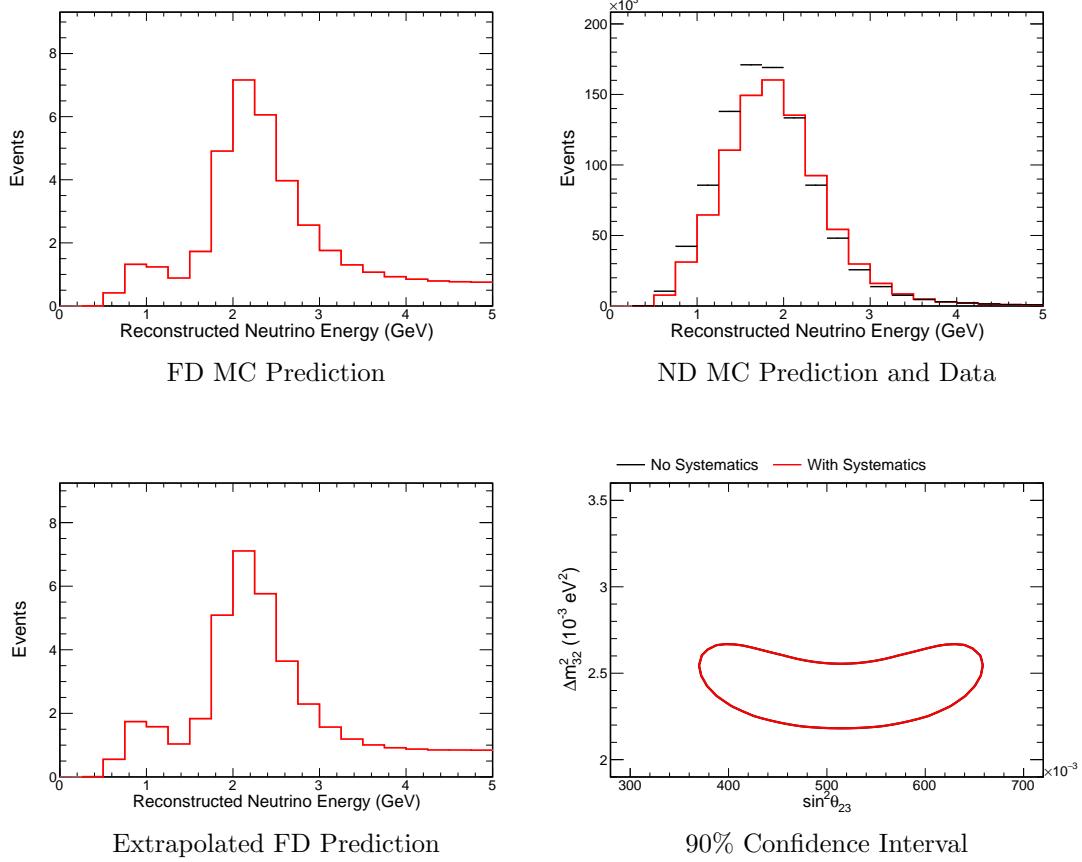


Figure 10.10: Systematic effect of MvNCRES uncertainty

Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

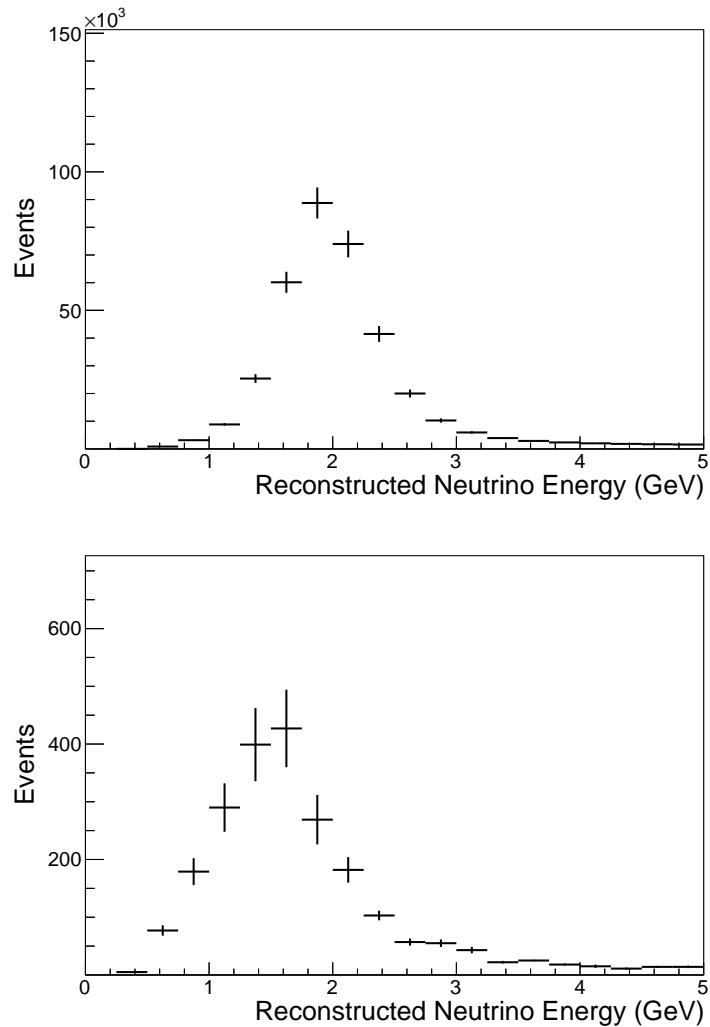


Figure 10.11: Summed GENIE uncertainty band for FD
The effects of free parameters with small systematic shifts were combined by summing the shifts in quadrature. The top pane shows the magnitude of the systematic uncertainty for the FD signal spectrum, while the bottom pane shows the uncertainty on the background spectrum.

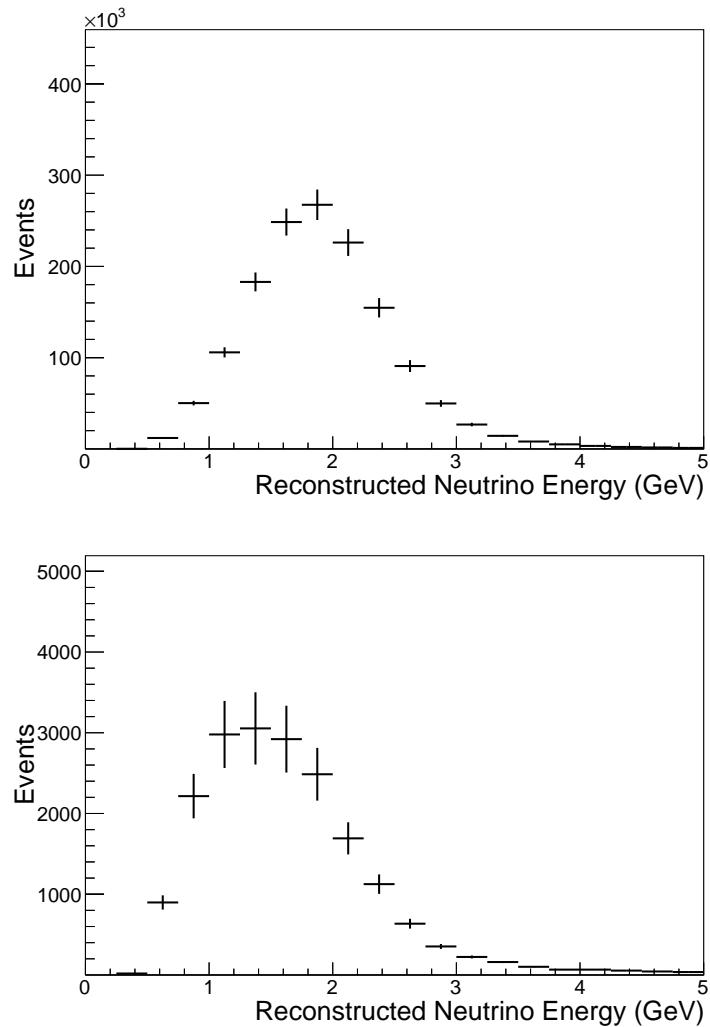


Figure 10.12: Summed GENIE uncertainty band for ND
 The effects of free parameters with small systematic shifts were combined by summing the shifts in quadrature. The top pane shows the magnitude of the systematic uncertainty for the ND signal spectrum, while the bottom pane shows the uncertainty on the background spectrum.

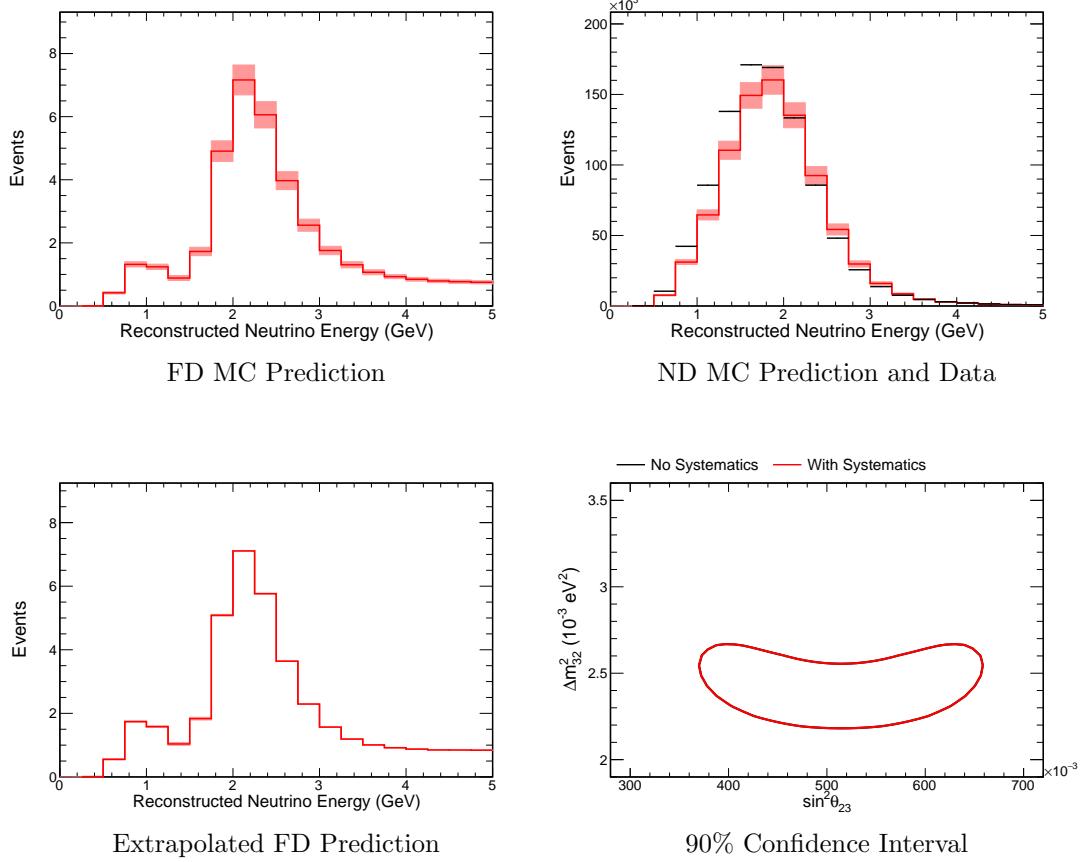


Figure 10.13: Systematic effect of combined GENIE uncertainty
 Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

Physics List	Mean Ratio	Normalization Ratio
FTFP_BERT	0.999	1.001
FTF_BIC	0.994	1.008
QGSC_BERT	0.998	0.999
QGSP_BIC_HP	0.991	0.993

Table 10.4: Shifts induced by alternative Geant4 physics lists

The ND MC prediction was formed using alternative MC samples with altered Geant4 physics lists. The spectra were fit with a truncated normal distribution in order to parametrize the systematic uncertainty as a shift in mean and normalization. The second column shows ratio of the shifted mean to the nominal for each alternative sample; the third column shows the ratio of the normalization between the shifted and nominal.

10.4 Particle Propagation Uncertainty

Simulating neutrino interactions in NOvA includes propagating secondary particles¹ through the detectors. This process is subject to uncertainties in the models involved, which mainly those involving interactions with atomic nuclei [64]. The Geant4 package can be configured to operate with different *physics lists* which each represent a different set of interaction models. In order to estimate the systematic uncertainty, the NOvA MC was configured to run with four different alternative physics lists: QGSP_BIC_HP, FTF_BIC, FTFP_BERT, and QGSC_BERT. Each of these physics lists is described in Section 4.3.

Using the sample generated with each physics list, an alternative ND MC prediction was generated. The beam peak was fit with a truncated normal distribution in the range of 1.25 GeV to 2.75 GeV to extract the mean and normalization. The parameters extracted were compared to the nominal ND MC prediction and used to estimate the effect of the systematic uncertainty. See Figure 10.17 for the fit and prediction for QGSC_BIC_HP, Figure 10.14 for FTF_BIC, Figure 10.15 for FTFP_BERT, and Figure 10.16 for QGSC_BERT. The largest shifts came from QGSC_BIC_HP, which were just shy of 1%. Thus, the systematic was included in the analysis framework as a 1% shift in normalization weight and a 1% energy scale factor. The effect of the energy scale on the prediction is displayed in Figure 10.18 and the normalization shift in Figure 10.19.

¹ Particles produced in the neutrino interaction

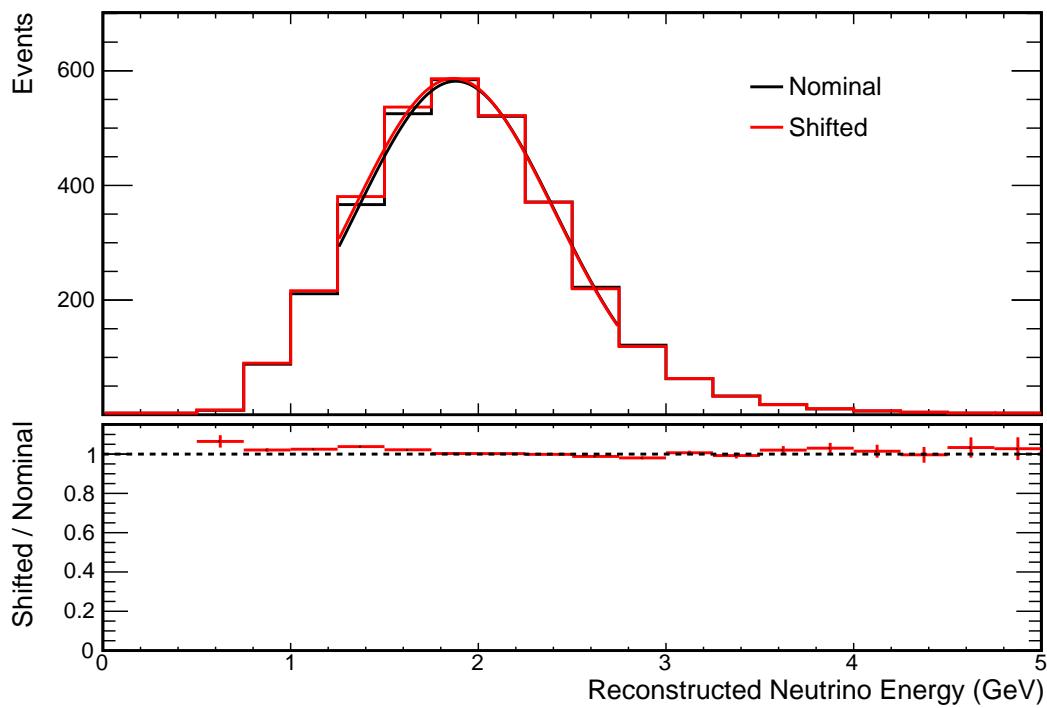


Figure 10.14: Systematically shifted spectrum alternative Geant4 physics list: FTF_BIC. The systematic uncertainty for particle propagation modeling was estimated by configuring Geant4 with alternative physics lists; in this case FTF_BIC was used. The ND prediction for each sample was compared to the nominal prediction by fitting both with a truncated normal distribution. The fits are overlaid as smooth curves on top of the spectra for both the nominal and shifted prediction.

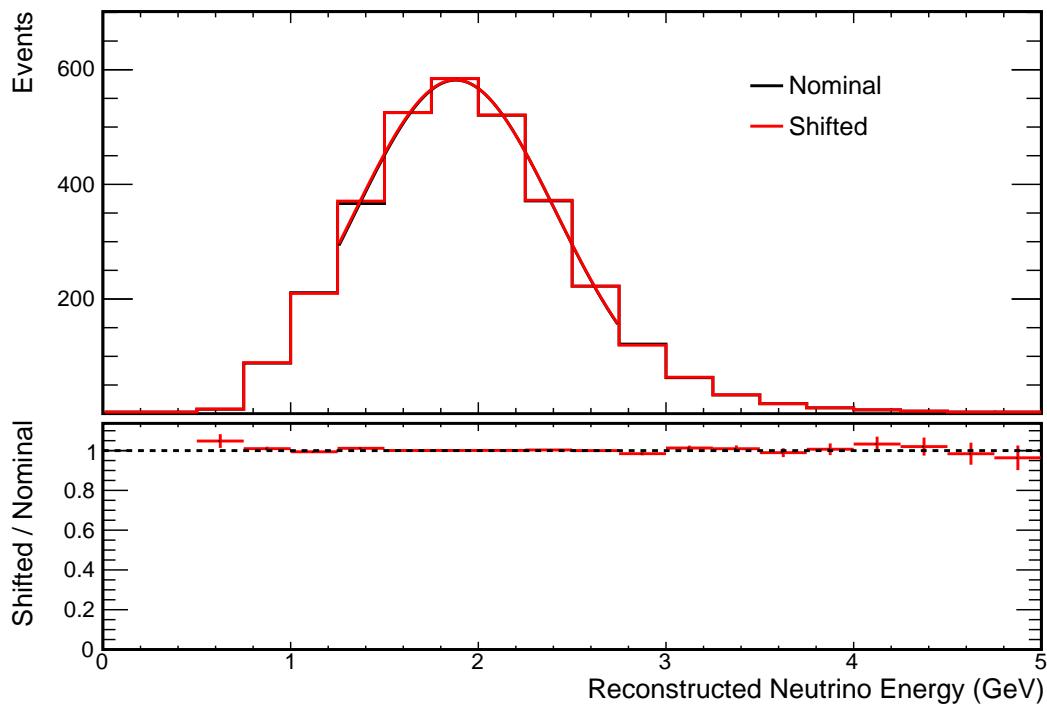


Figure 10.15: Systematically shifted spectrum alternative Geant4 physics list: FTFP_BERT

The systematic uncertainty for particle propagation modeling was estimated by configuring Geant4 with alternative physics lists; in this case FTFP_BERT was used. The ND prediction for each sample was compared to the nominal prediction by fitting both with a truncated normal distribution. The fits are overlaid as smooth curves on top of the spectra for both the nominal and shifted prediction.

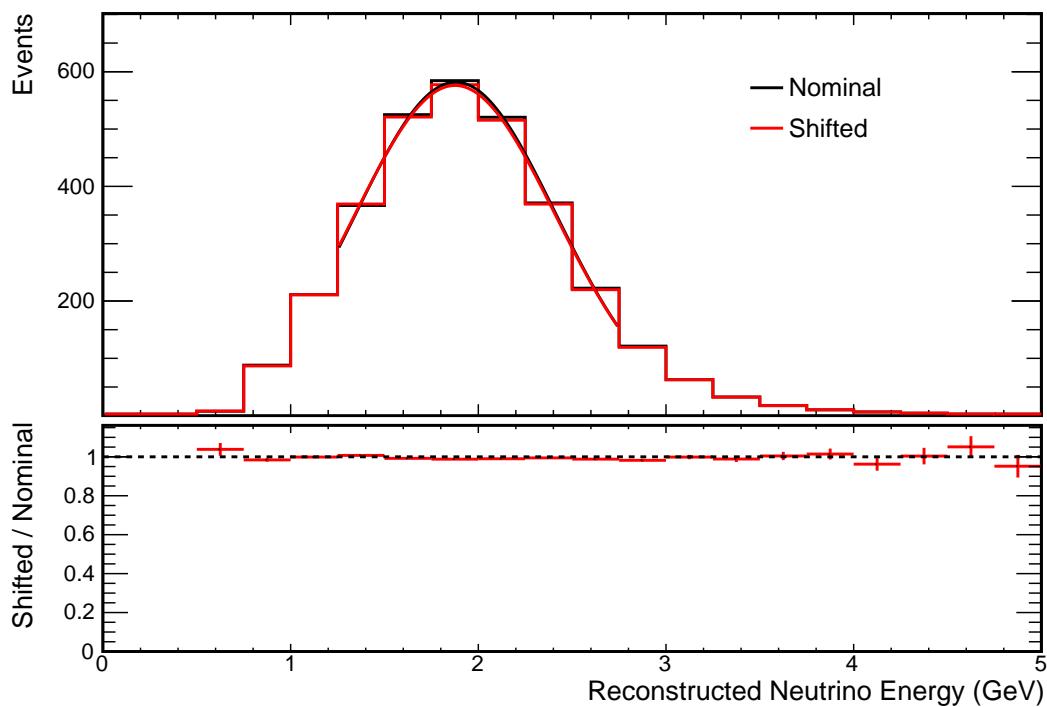


Figure 10.16: Systematically shifted spectrum alternative Geant4 physics list: QGSC_BERT

The systematic uncertainty for particle propagation modeling was estimated by configuring Geant4 with alternative physics lists; in this case QGSC_BERT was used. The ND prediction for each sample was compared to the nominal prediction by fitting both with a truncated normal distribution. The fits are overlaid as smooth curves on top of the spectra for both the nominal and shifted prediction.

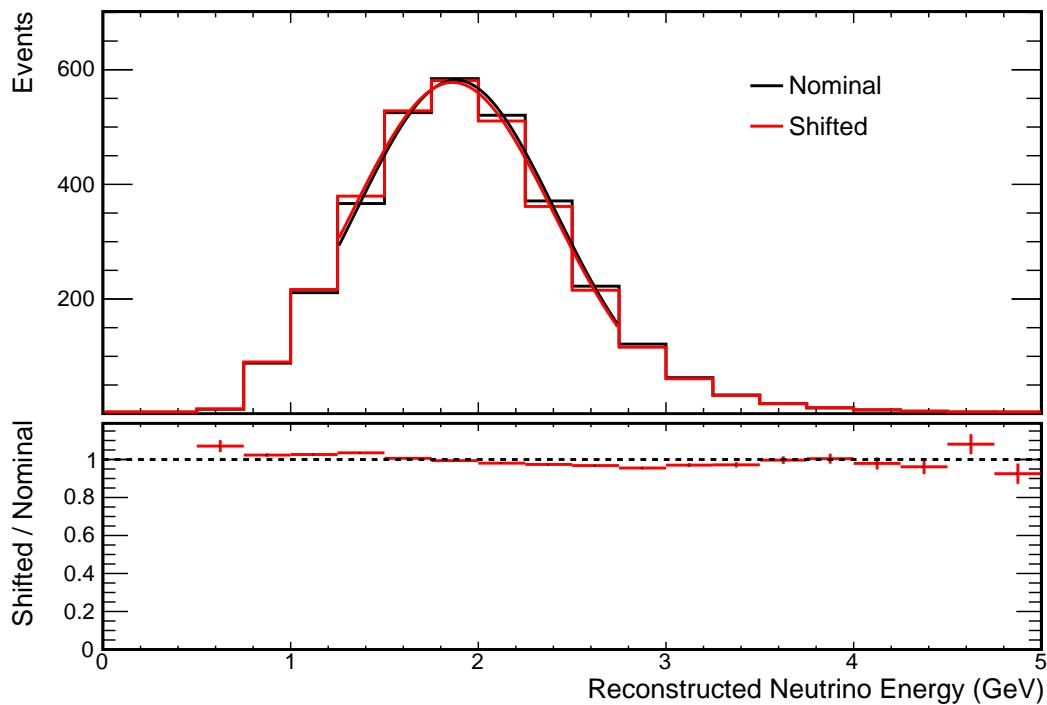


Figure 10.17: Systematically shifted spectrum alternative Geant4 physics list: QGSP_BIC_HP

The systematic uncertainty for particle propagation modeling was estimated by configuring Geant4 with alternative physics lists; in this case QGSP_BIC_HP was used. The ND prediction for each sample was compared to the nominal prediction by fitting both with a truncated normal distribution. The fits are overlaid as smooth curves on top of the spectra for both the nominal and shifted prediction.

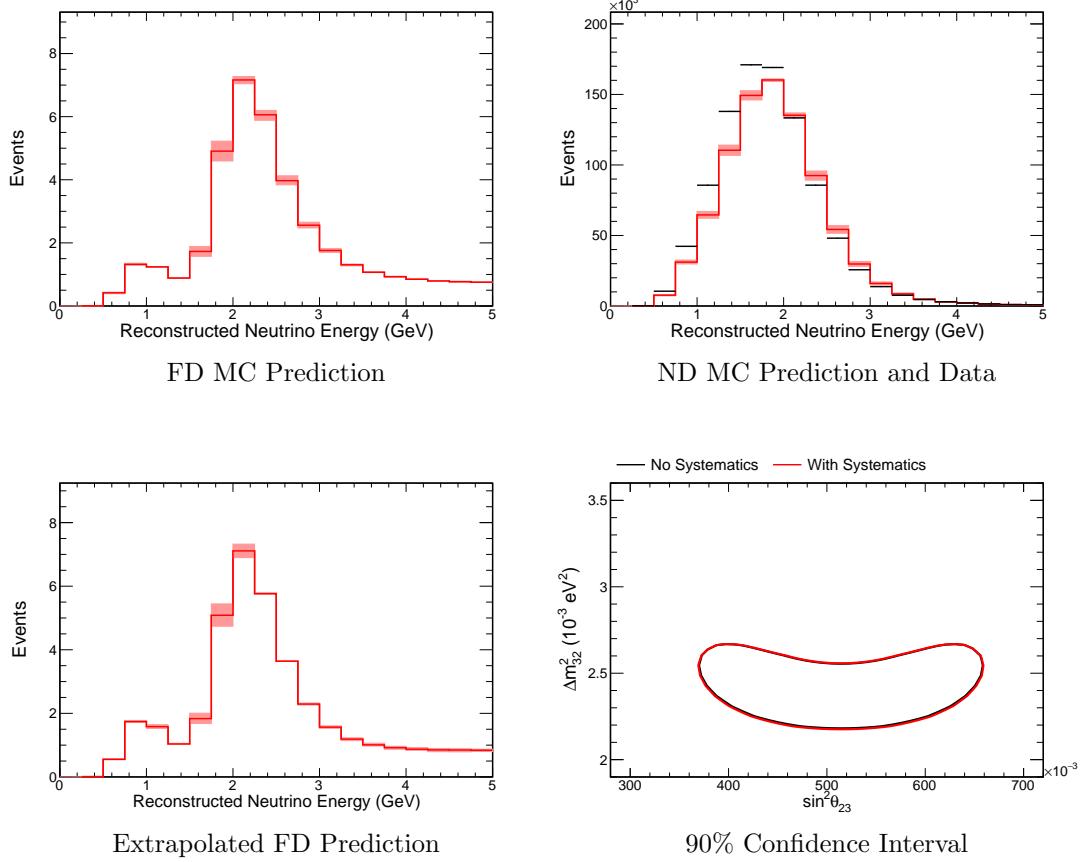


Figure 10.18: Systematic effect of particle propagation energy scale uncertainty. Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

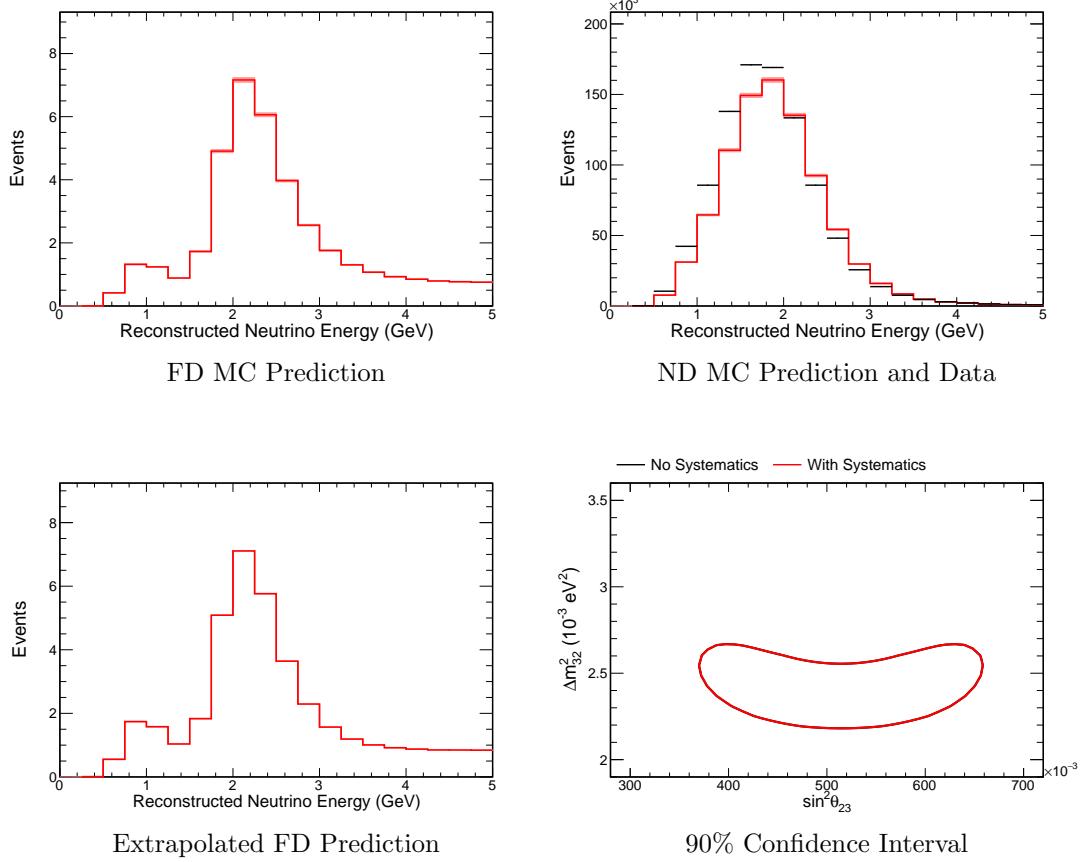


Figure 10.19: Systematic effect of particle propagation energy normalization uncertainty. Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

Alternative Sample	Mean Ratio	Normalization Ratio
$k_B = 0.02 \text{ cm/MeV}$	0.981	0.977
$k_B = 0.01 \text{ cm/MeV}$	0.996	0.983

Table 10.5: Shifts induced by alternative scintillation production properties
The ND MC prediction was formed using MC samples with alternative values for k_B and k_C set to 0. The spectrum for each alternative sample was fit with a truncated normal distribution in order to parametrize the systematic uncertainty as a shift in mean and normalization. The second column shows ratio of the shifted mean to the nominal for each alternative sample; the third column shows the ratio of the normalization between the shifted and nominal.

10.5 Scintillation Production Uncertainty

As discussed in Section 4.3, NOvA simulation models scintillation production using Birks' law [66] with a corrections from [67]:

$$\frac{dL}{dX} = L_0 \frac{\frac{dE}{dX}}{1 + k_B \frac{dE}{dX} + k_C \frac{dE^2}{dX^2}}. \quad (10.1)$$

The free parameter L_0 sets the absolute scale for light production as a function of energy deposition per path length dE/dX , while k_B and k_C are free to capture any nonlinearity. L_0 is effectively absorbed by downstream simulation and covered by the calorimetric energy uncertainty in Section 10.6. Uncertainty in k_B and k_C , however, must be treated independently. The values $k_B = 0.04 \text{ cm/MeV}$ and $k_C = -0.0005 \text{ cm}^2/\text{MeV}^2$ were tuned using comparisons between ND data and MC. Two alternative samples were generated with different samples of k_B and k_C . The first sample has $k_B = 0.02 \text{ cm/MeV}$ and $k_C = 0$; the second has $k_B = 0.01 \text{ cm/MeV}$ and $k_C = 0$. The value of 0.01 is motivated by measurements [99] of k_B in organic scintillators.

In a manner similar to the particle propagation uncertainty, the FD MC prediction for each shifted sample was fit with a truncated normal distribution; the mean and normalization were extracted from the fit and interpreted as an energy scale and normalization shift. A comparison of the fits is shown in Table 10.5. A 2% uncertainty is taken on both the reconstructed energy scale and normalization.

Since the tuning was performed using ND data, we have no knowledge how well

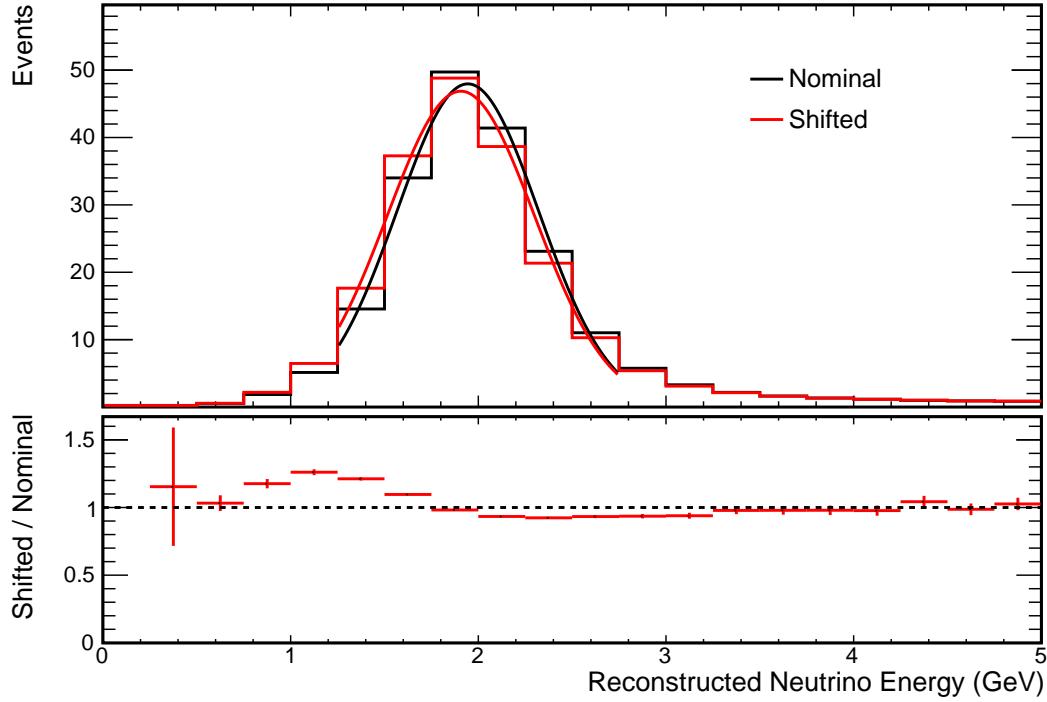


Figure 10.20: Systematically shifted spectrum with $k_B = 0.02 \text{ cm/MeV}$

The systematic uncertainty for scintillation production propagation modeling was estimated by configuring the FD MC with alternative scintillation production properties. In this case $k_B = 0.02 \text{ cm/MeV}$ was used. The FD prediction for each sample was compared to the nominal prediction by fitting both with a truncated normal distribution. The fits are overlaid as smooth curves on top of the spectra for both the nominal and shifted prediction.

these parameters match FD data. As such, the energy scale and normalization factor are each represented by a separate relative and absolute uncertainty. The effect of the absolute normalization uncertainty is shown in Figure 10.23, the relative normalization uncertainty in Figure 10.23, the absolute energy scale uncertainty in Figure 10.24, and the relative energy scale uncertainty Figure 10.25.

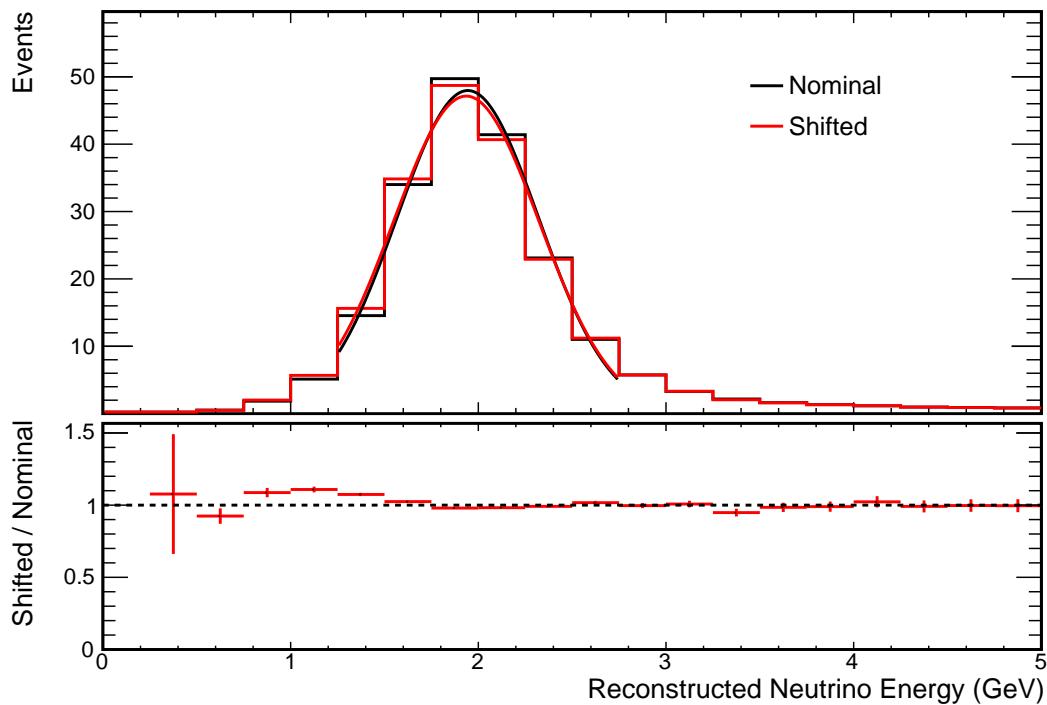


Figure 10.21: Systematically shifted spectrum with $k_B = 0.01 \text{ cm/MeV}$
The systematic uncertainty for scintillation production propagation modeling was estimated by configuring the FD MC with alternative scintillation production properties. In this case $k_B = 0.01 \text{ cm/MeV}$ was used. The FD prediction for each sample was compared to the nominal prediction by fitting both with a truncated normal distribution. The fits are overlaid as smooth curves on top of the spectra for both the nominal and shifted prediction.

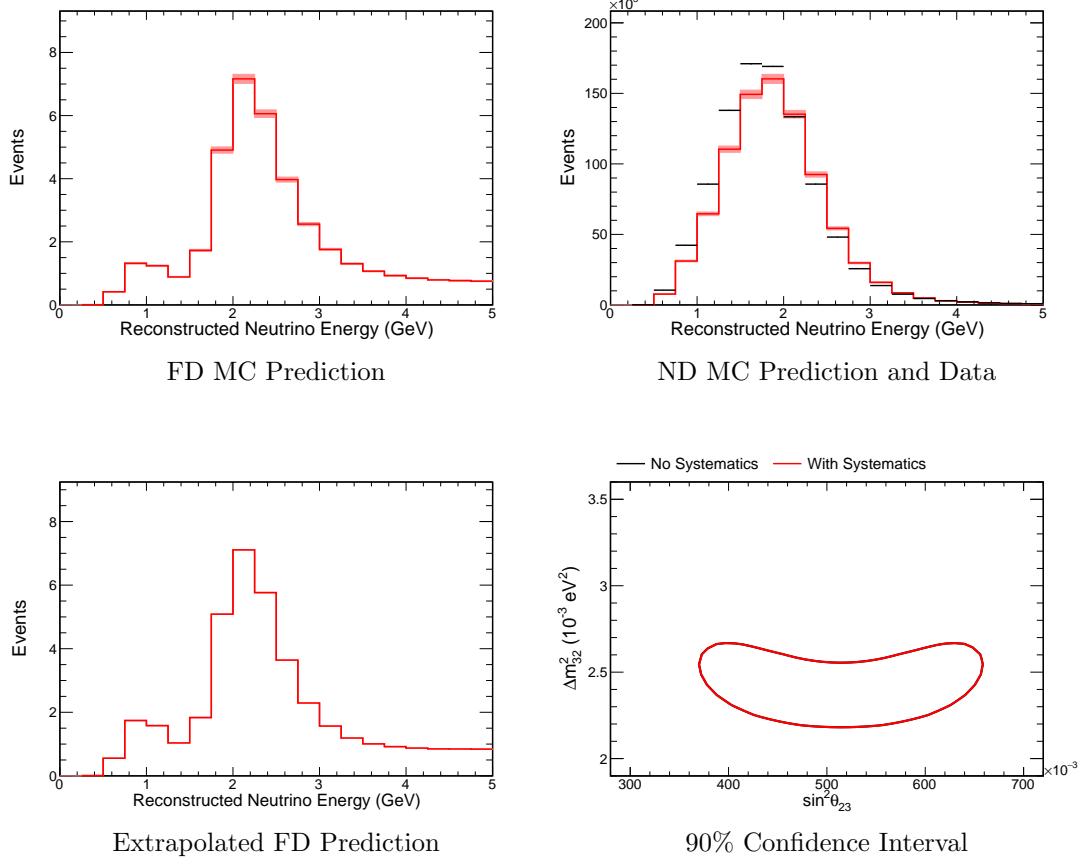


Figure 10.22: Systematic effect of absolute scintillation production normalization uncertainty

Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

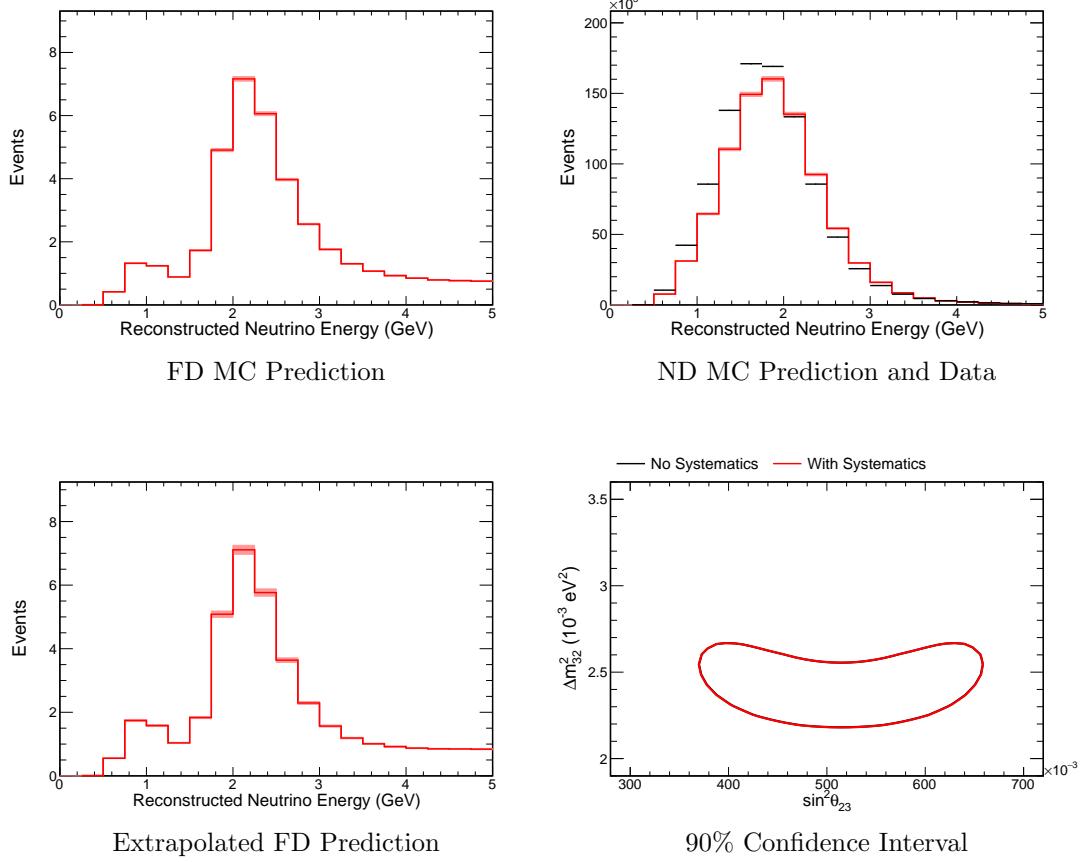


Figure 10.23: Systematic effect of relative scintillation production normalization uncertainty

Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

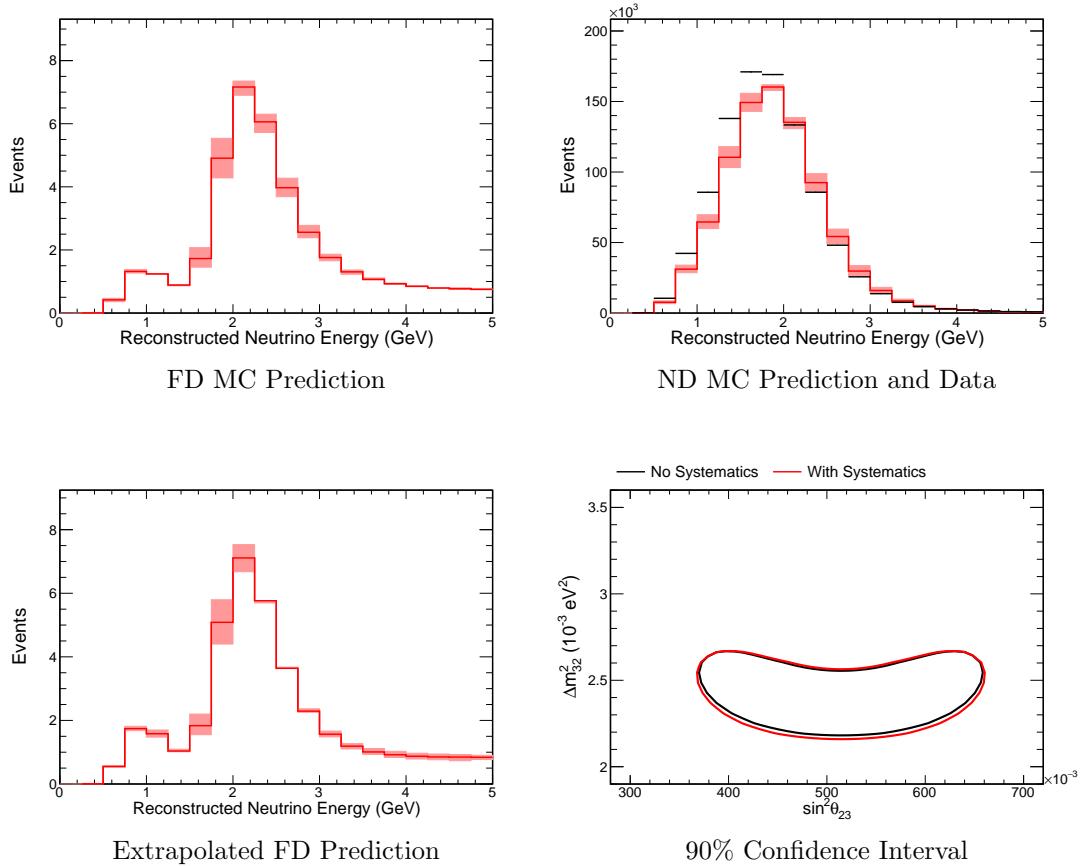


Figure 10.24: Systematic effect of absolute scintillation production energy scale uncertainty

Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty.

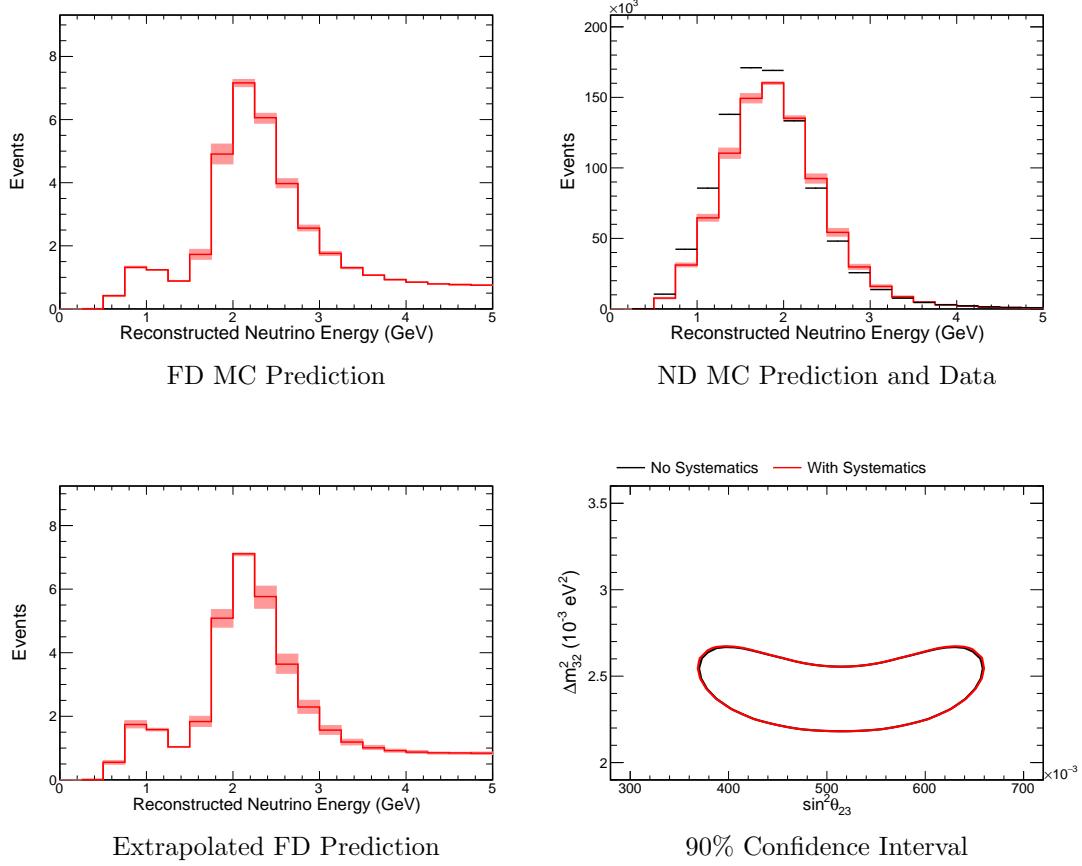


Figure 10.25: Systematic effect of relative scintillation production energy scale uncertainty

Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

Alternative Sample	Mean Ratio	Normalization Ratio
+5% Absolute Energy Scale	1.020	0.981
-5% Absolute Energy Scale	0.980	1.013
Random cell-to-cell variations	1.001	0.996
Attenuation slope adjustment	1.009	0.995

Table 10.6: Shifts induced in alternative calibration samples

The ND MC prediction was formed using MC samples with altered calibration parameters. The spectrum for each alternative sample was fit with a truncated normal distribution in order to parametrize the systematic uncertainty as a shift in mean and normalization. The second column shows ratio of the shifted mean to the nominal for each alternative sample; the third column shows the ratio of the normalization between the shifted and nominal.

10.6 Calorimetric Energy Uncertainty

The calorimetric energy scale is determined independently for both detectors and both for data and MC. In the extrapolation procedure, it is assumed that the calorimetric energy scale between data and MC has been determined accurately. Studies of muon activity in the ND has shown that the calorimetric energy scales can differ by as much as 4%. Accurate calorimetry also relies on the attenuation and cell-to-cell effects to have been properly calibrated out. In order to test the effects of such effects on the reconstruction, a suite of alternative calibration samples have been constructed. Two samples were created by adjusting the absolute energy scale up and down by 5% [77]. Another sample was produced by adjusting the attenuation curve for each cell up and down by a random factor drawn from a normal distribution with a spread of 0.8%. The final sample was produced by modifying the slopes of the attenuation fits amounting to a 20% deviation at the far end of the cell.

Again, the systematic uncertainty has been estimated by observing the effect of these shifts. In each case, a truncated normal distribution was fit to the shifted FD MC prediction. The fit results are compared to a fit to the nominal FD prediction in Table 10.6. The fit for the the $\pm 5\%$ shifts can be seen in Figures 10.26 and 10.27, respectively. Figure 10.28 shows the random calibration shift sample and Figure 10.29 shows the slope shift sample. Both the random and slope shifts show small effects on the mean energy and normalization; this makes sense since the shifts can average

out across the detector. The 5% energy scale shifts, however, lead to a significant systematic effect. Both the normalization and energy scale were seen to deviate from the nominal prediction by $\pm 2\%$. The energy scale shift is smaller than the full 5% since calorimetry only enters in estimation of the hadronic energy; the energy of the muon track is estimated from range and thus insensitive to calorimetric energy variations.

Thus, the calorimetric energy uncertainty is included in the analysis as a 2% normalization shift and 2% energy scale. Both of these are taken as both relative and absolute. The effects of the absolute and relative normalization shift can be seen in Figures 10.30 and 10.31, respectively. Similarly, the absolute and relative calibration scale can be seen respectively in Figures 10.32 and 10.33.

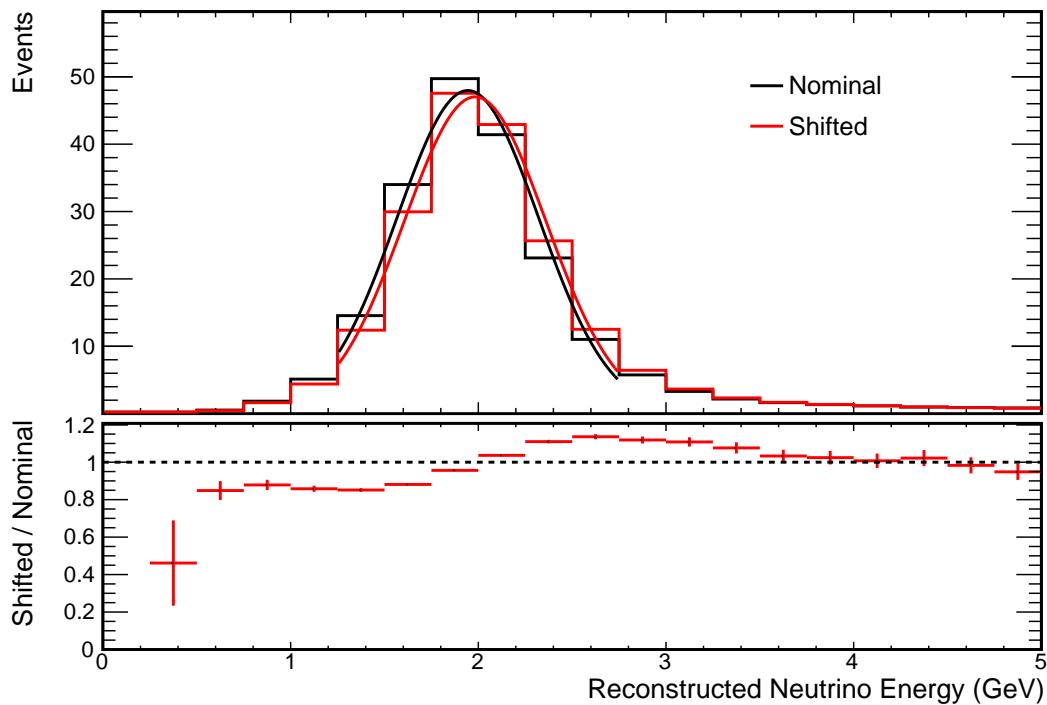


Figure 10.26: Systematically shifted spectrum with -5% absolute energy scale. The systematic uncertainty for scintillation production propagation modeling was estimated by configuring the FD MC with alternative scintillation production properties. In this case the absolute energy scale was increased by 5% . The FD prediction for each sample was compared to the nominal prediction by fitting both with a truncated normal distribution. The fits are overlaid as smooth curves on top of the spectra for both the nominal and shifted prediction.

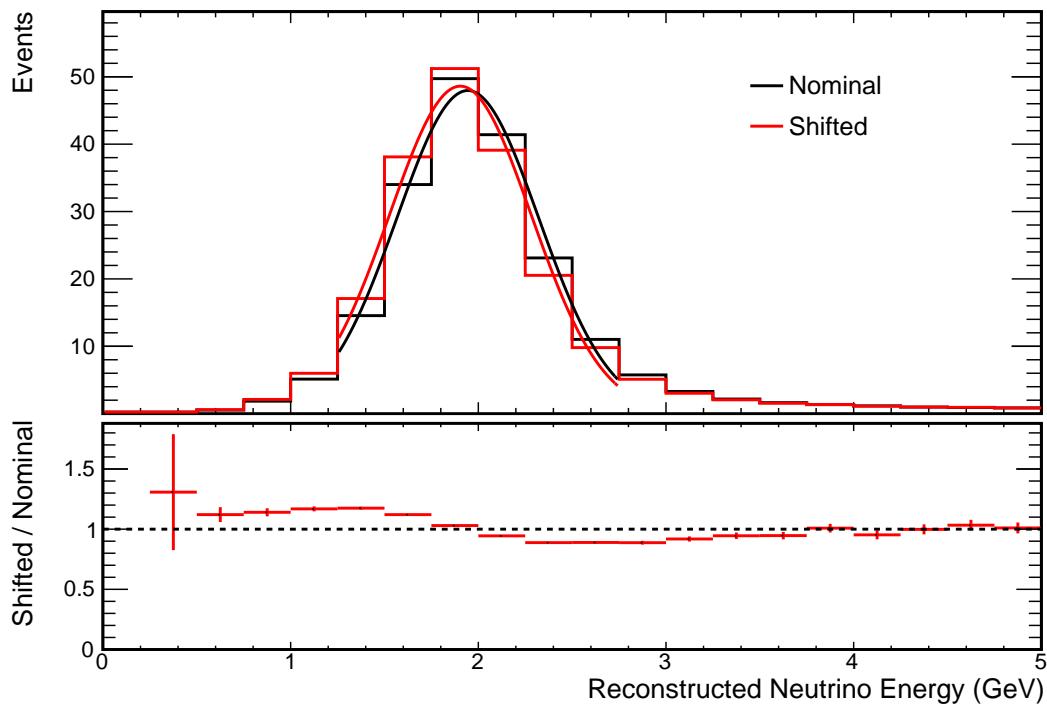


Figure 10.27: Systematically shifted spectrum with -5% absolute energy scale. The systematic uncertainty for scintillation production propagation modeling was estimated by configuring the FD MC with alternative scintillation production properties. In this case the absolute energy scale was decreased by 5%. The FD prediction for each sample was compared to the nominal prediction by fitting both with a truncated normal distribution. The fits are overlaid as smooth curves on top of the spectra for both the nominal and shifted prediction.

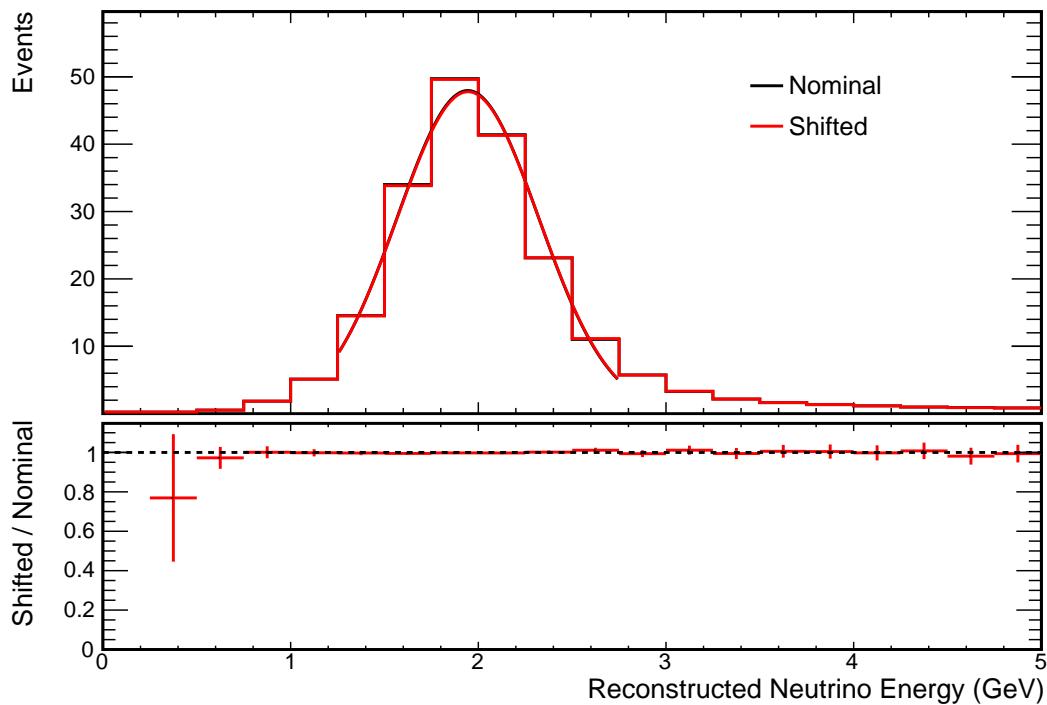


Figure 10.28: Systematically shifted spectrum with random cell-to-cell variations. The systematic uncertainty for scintillation production propagation modeling was estimated by configuring the FD MC with alternative scintillation production properties. In this case the response for each channel was randomly scaled by a normally distributed 8%. The FD prediction for each sample was compared to the nominal prediction by fitting both with a truncated normal distribution. The fits are overlaid as smooth curves on top of the spectra for both the nominal and shifted prediction.

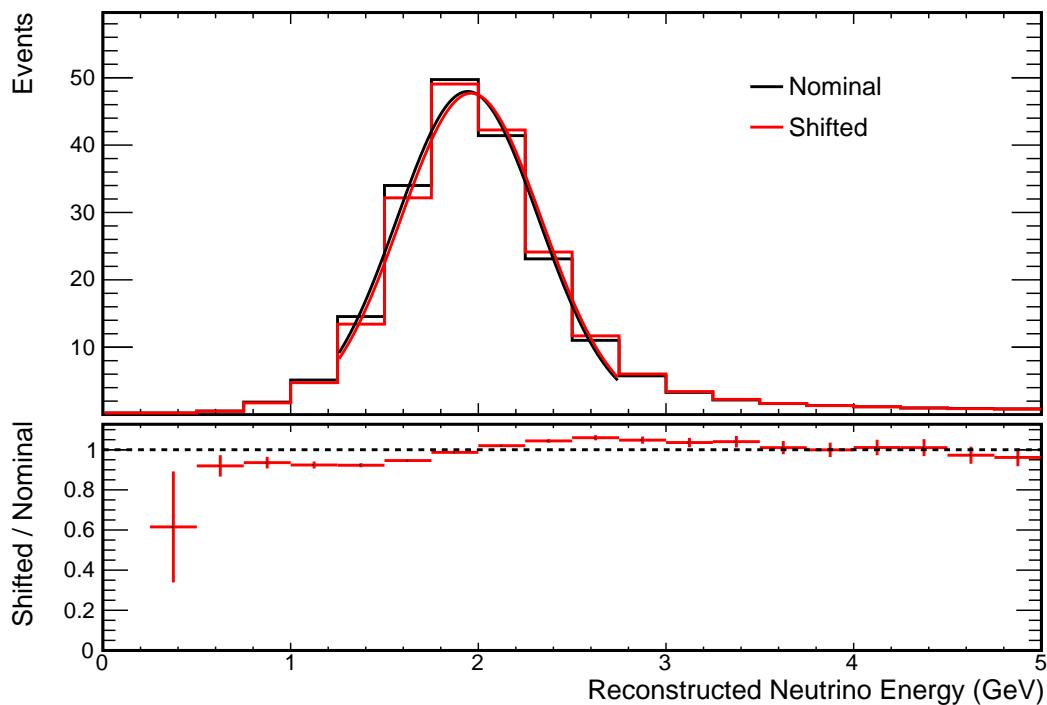


Figure 10.29: Systematically shifted spectrum with artificial attenuation slopes
 The systematic uncertainty for scintillation production propagation modeling was estimated by configuring the FD MC with alternative scintillation production properties. In this case the attenuation slopes were altered with the effect of increasing the response at the far end of the cell by 20%. The FD prediction for each sample was compared to the nominal prediction by fitting both with a truncated normal distribution. The fits are overlaid as smooth curves on top of the spectra for both the nominal and shifted prediction.

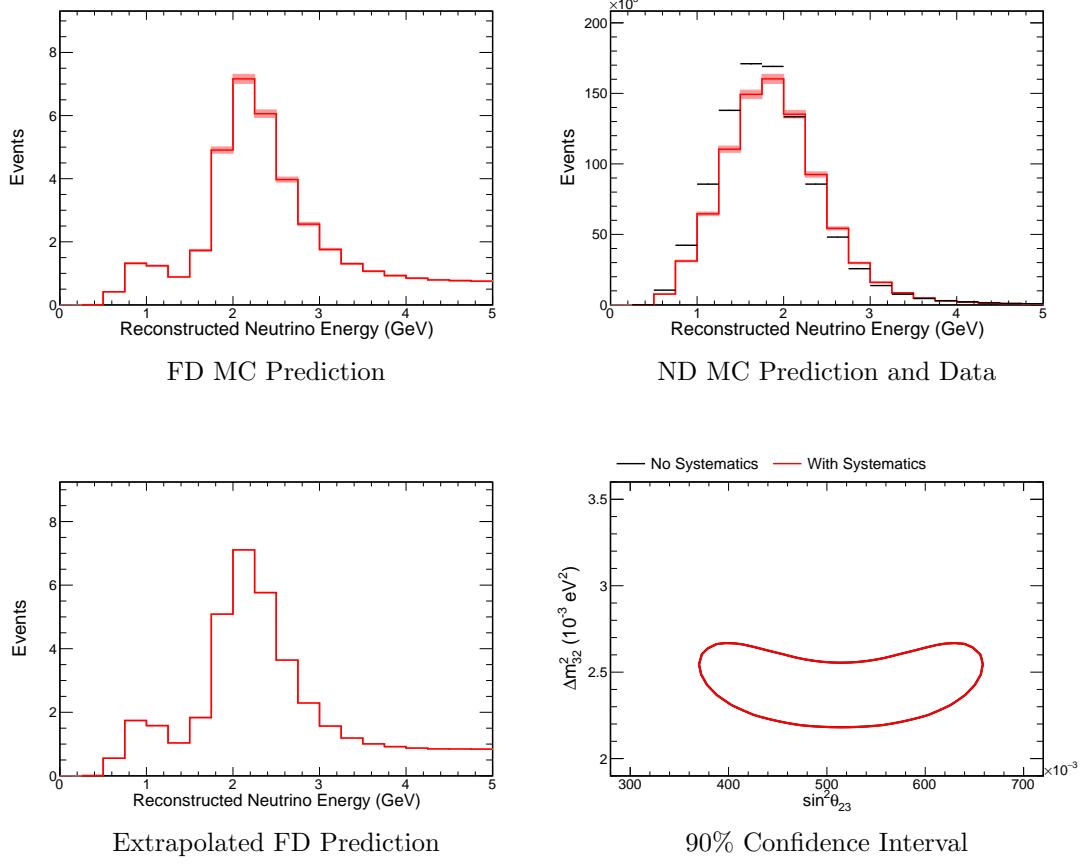


Figure 10.30: Systematic effect of absolute calorimetric energy normalization uncertainty

Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

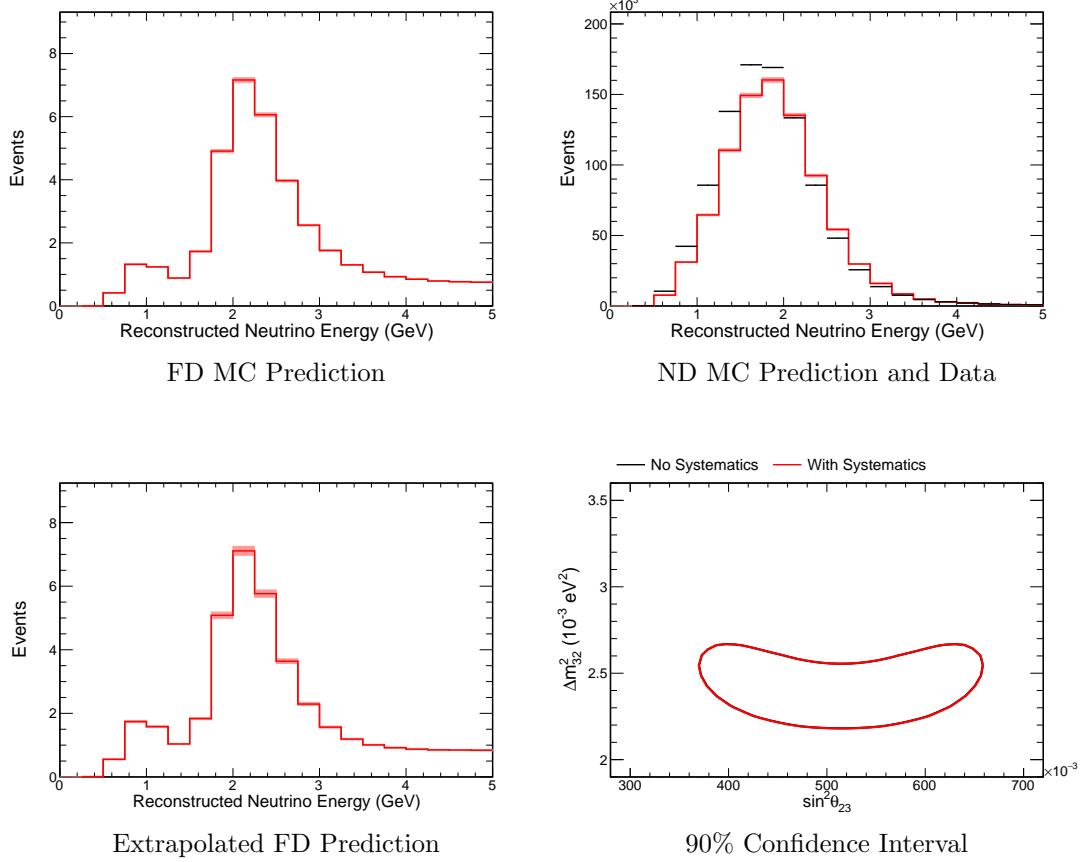


Figure 10.31: Systematic effect of relative calorimetric energy normalization uncertainty. Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

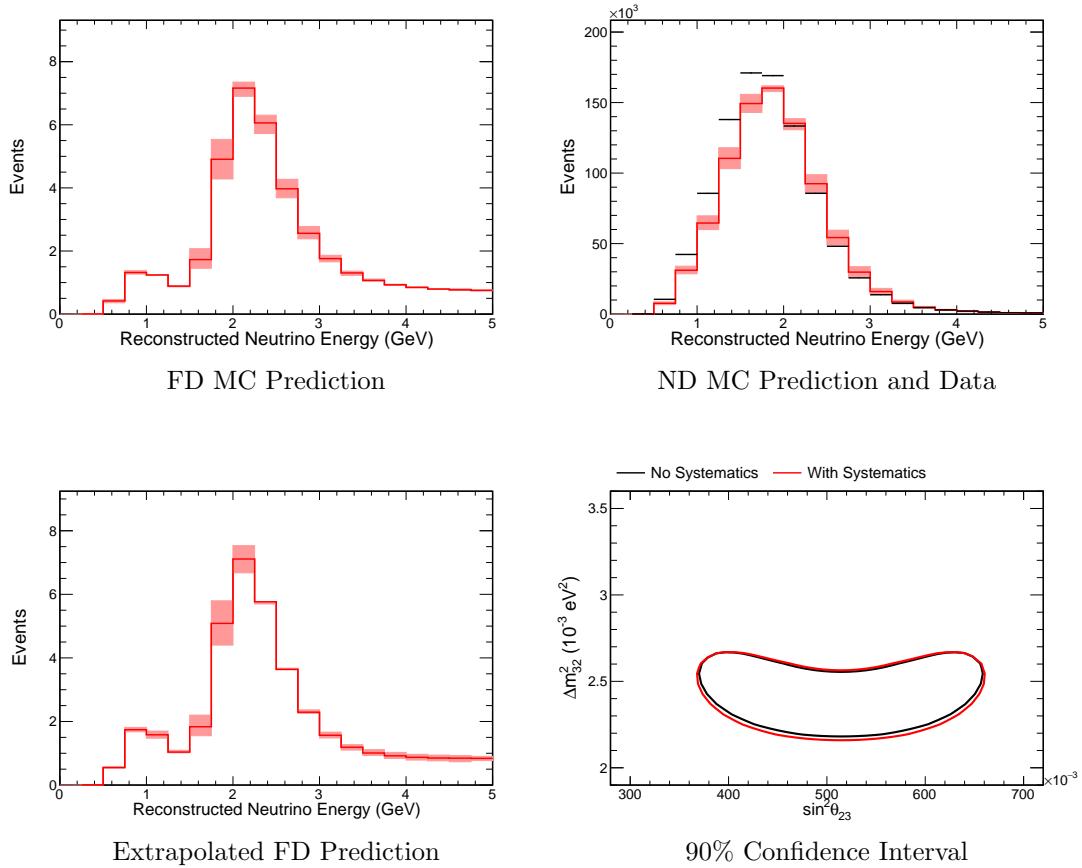


Figure 10.32: Systematic effect of absolute calorimetric energy scale uncertainty. Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty.

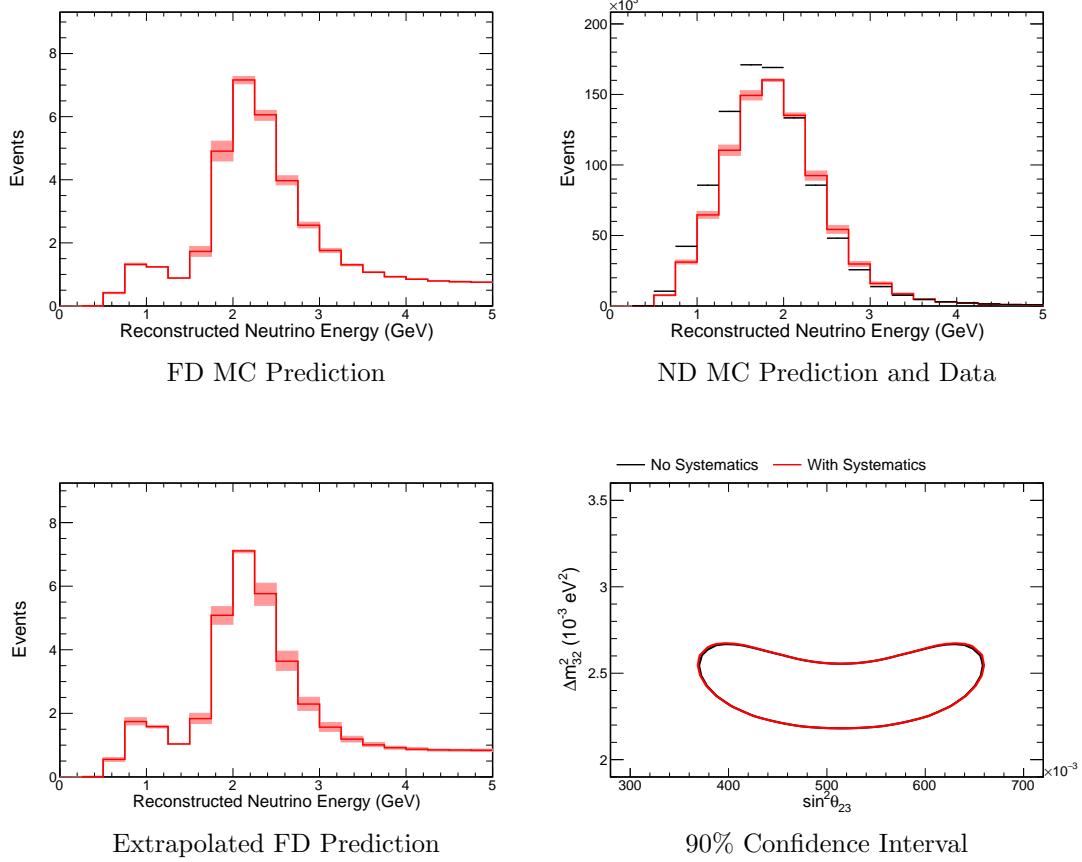


Figure 10.33: Systematic effect of relative calorimetric energy scale uncertainty. Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

Component	FD		ND	
	Mass (kg/m/cell)	Percentage	Mass (kg/m/cell)	Percentage
Extrusions	0.951 ± 0.016	36.4 ± 0.6	0.951 ± 0.016	36.2 ± 0.6
Scintillator	1.641 ± 0.009	62.8 ± 0.3	1.653 ± 0.011	63.0 ± 0.4
Glue	0.019 ± 0.0003	0.07 ± 0.01	0.019 ± 0.0003	0.07 ± 0.01
Fiber	0.001 ± 0.0001	0.04 ± 0.04	0.001 ± 0.0001	0.04 ± 0.04
Total	2.612 ± 0.018	100.0 ± 0.7	2.624 ± 0.019	100.0 ± 0.07

Table 10.7: Uncertainties in density of NOvA cells

The detector mass uncertainty was estimated by aggregating uncertainties density of composite materials. Shown above is the density of each component in both the ND and FD along with the corresponding uncertainty. The values are also displayed in terms of percent composition rather than absolute density.

10.7 Detector Mass Uncertainty

Proper modeling of detector mass is a requirement of the extrapolation procedure; if it were incorrect, exposure of the MC prediction could be inconsistent with data. The uncertainty has been estimated in [75] by taking uncertainties on each detector component in terms of linear density of cells. The result is shown in table 10.7. In each detector, the uncertainty in density is 0.7%.

A separate accounting from [100] compared the total mass of both the simulated and physical FD. That study produced an discrepancy of 0.8%

Those two uncertainties have been combined as a 1% overall normalization uncertainty, both absolute and relative. The effect of the absolute uncertainty is shown in Figure 10.34 and the relative uncertainty in Figure 10.35.

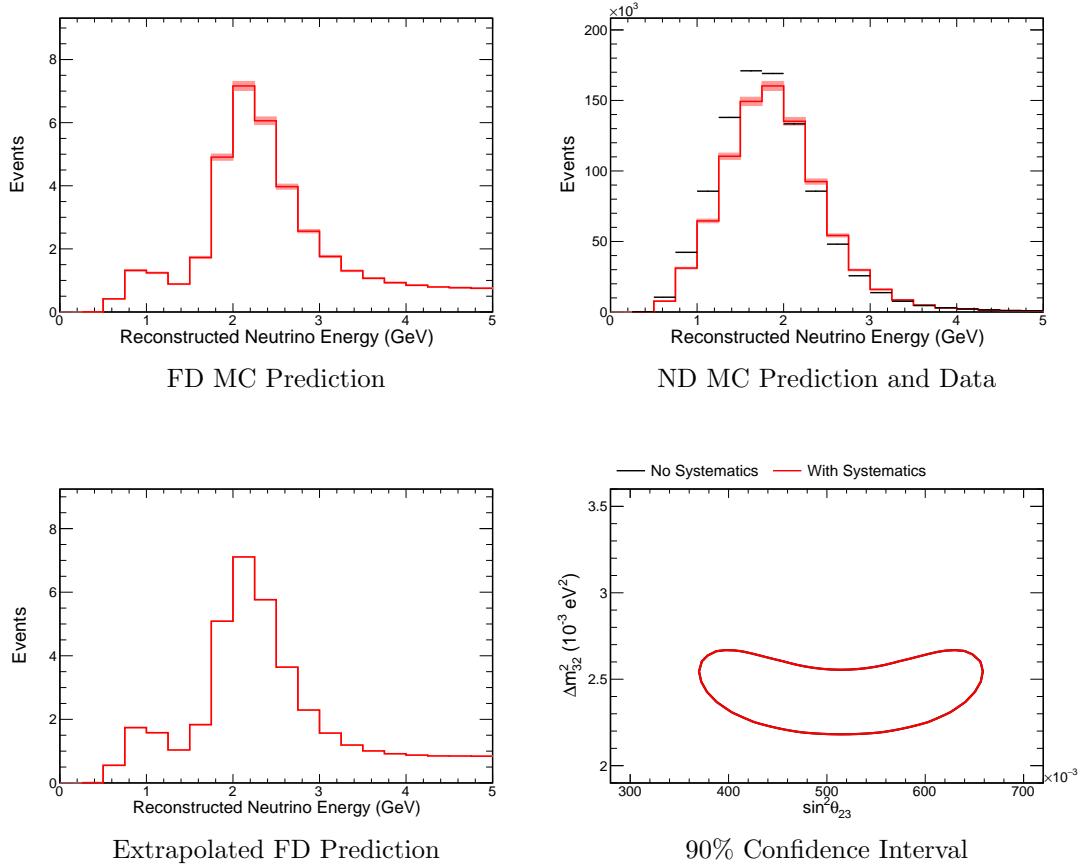


Figure 10.34: Systematic effect of absolute detector mass uncertainty
 Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

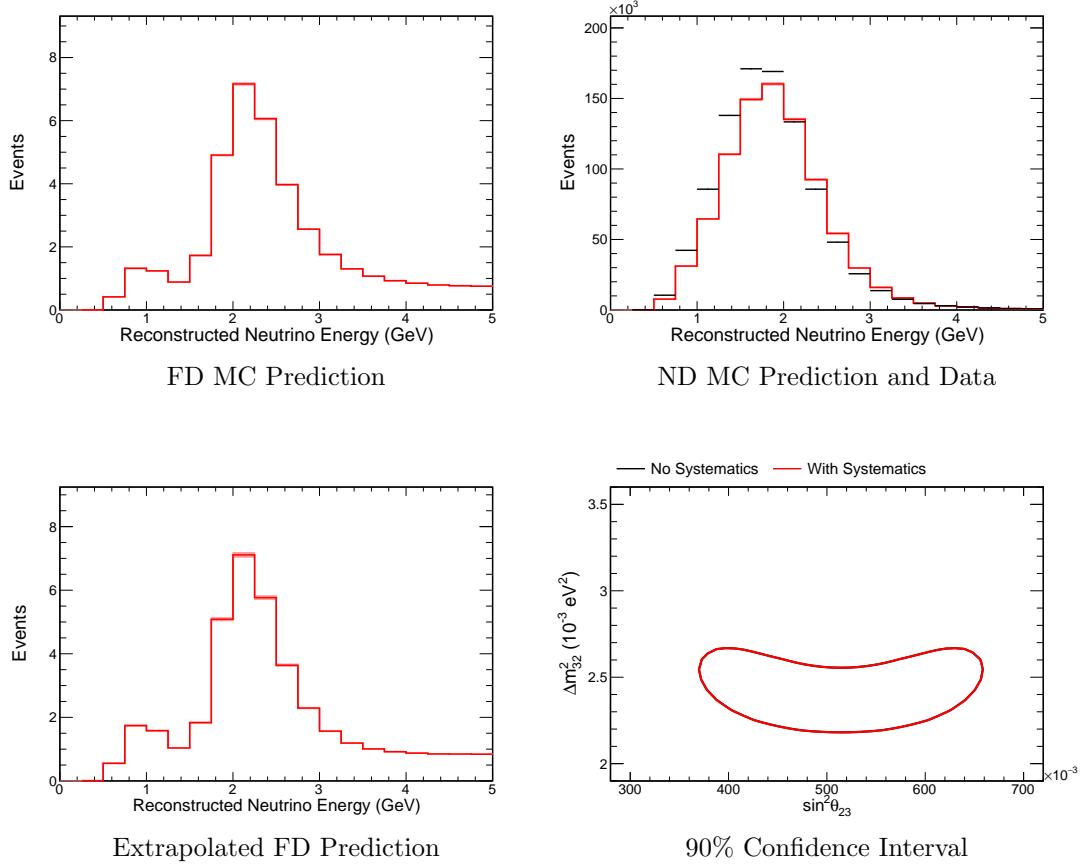


Figure 10.35: Systematic effect of relative detector mass uncertainty
 Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

10.8 Muon Range Uncertainty

The uncertainty in detector composition described in the previous section could induce discrepancy in the observed range traversed by muon tracks. As described in Section 5.5 length of the muon track enters directly into the reconstructed energy estimate. To compensate for this dependence, a systematic uncertainty of 1% of the track energy estimate has been applied. In the ND, the effect is enhanced by a 2% uncertainty on the amount of steel in the muon catcher [75]. As such, the uncertainty in the ND is taken as 2% rather than 1%. The effect of the FD muon range uncertainty is visible in Figure 10.36; the corresponding ND effect is shown in Figure 10.37.

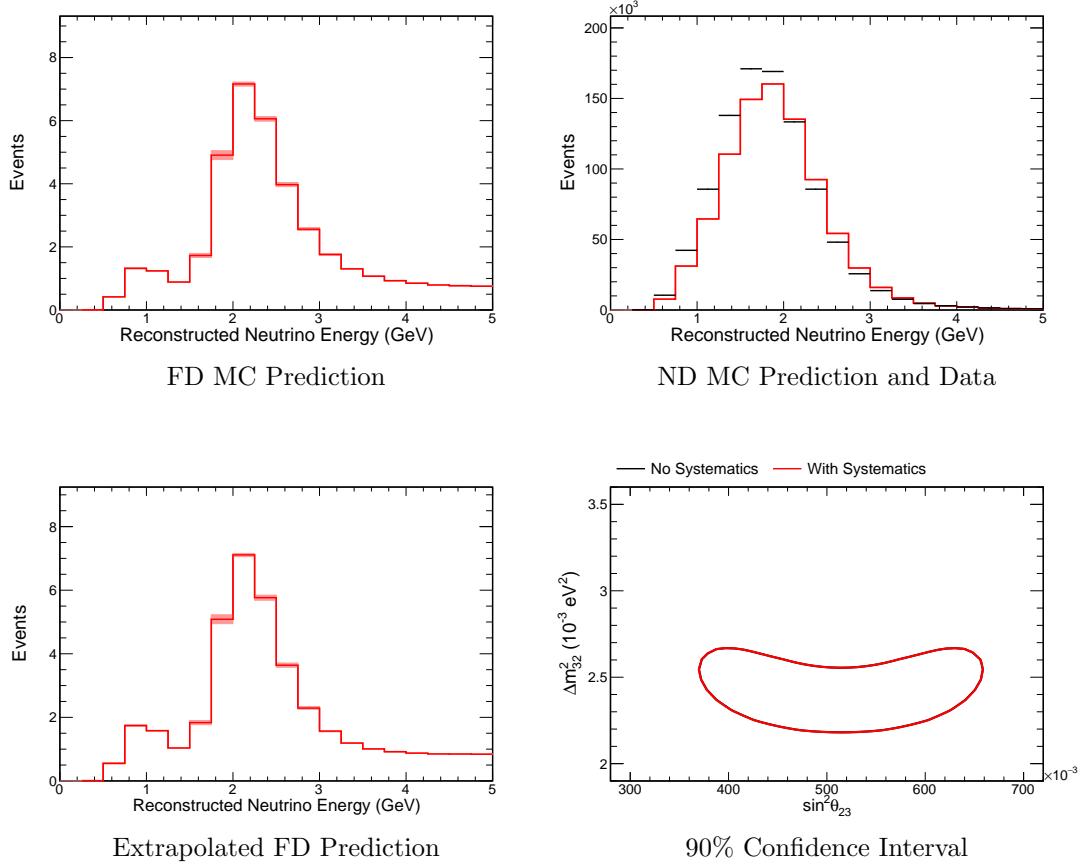


Figure 10.36: Systematic effect of FD muon track range uncertainty. Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

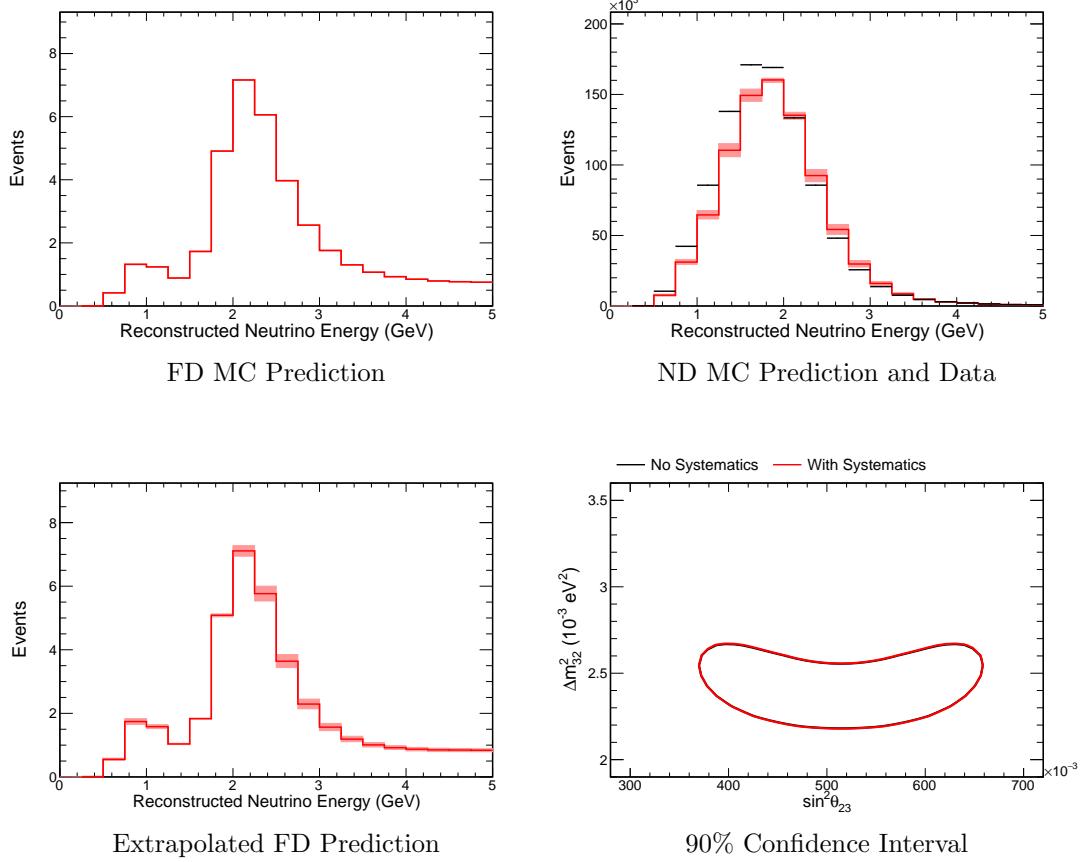


Figure 10.37: Systematic effect of ND muon track range uncertainty
 Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty. The effect is small enough that the contour which incorporates the systematic effect (red) almost completely obscures the version without it.

10.9 Hadronic Energy Uncertainty

As shown in Figure 8.18, the reconstructed hadronic energy distribution agrees poorly between the ND prediction and data. Hadronic energy enters directly in the reconstructed neutrino energy, which is also shown to disagree in Figure 8.19. Depending on which is taken as the denominator, the means of the ND predicted and data hadronic energy spectra differ by 12%. The source of this discrepancy is not understood. It may arise from an incomplete description of the interaction model in GENIE which distorts the observed spectra. Alternatively, the discrepancy may be caused by improper detector modeling. Since the cause is not understood, we take this discrepancy as a systematic uncertainty. The effect is represented as 12% uncertainty on the hadronic component of the reconstructed neutrino energy, both absolute and relative. The effect of the absolute uncertainty is shown in Figure 10.38 and the relative uncertainty in Figure 10.39. The hadronic energy uncertainty is by far the largest uncertainty in this analysis.

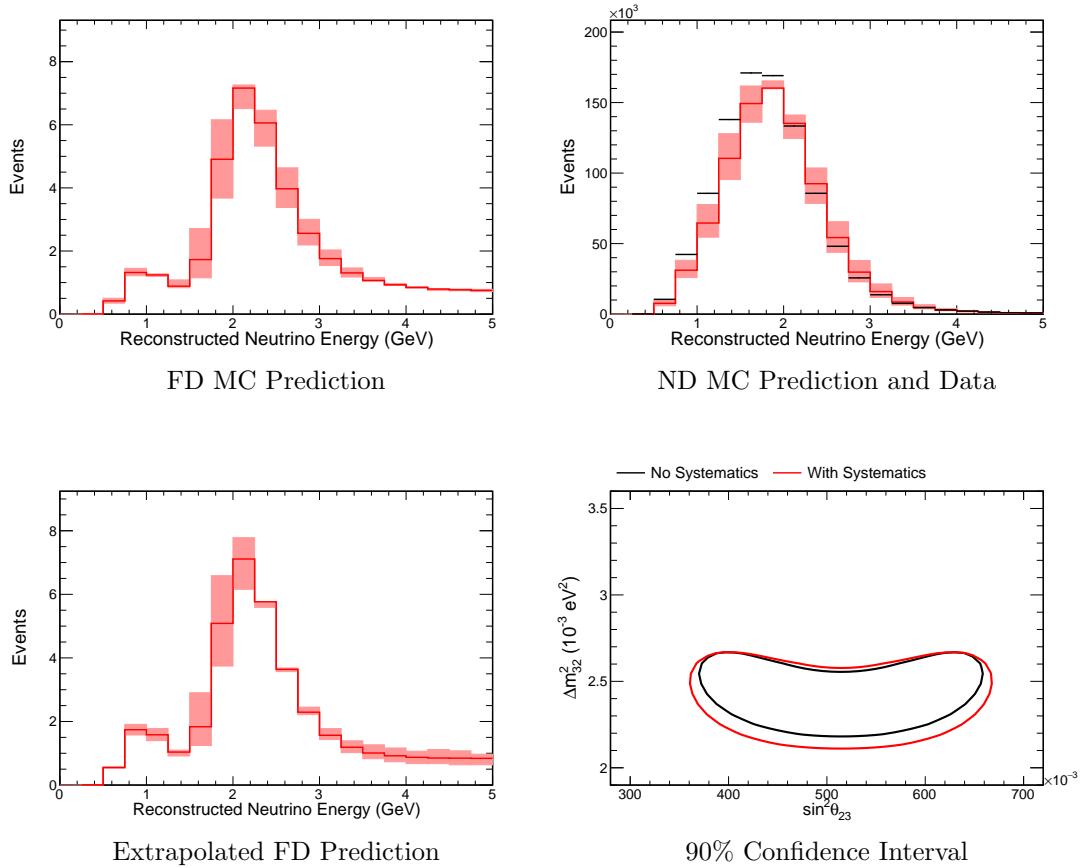


Figure 10.38: Systematic effect of absolute hadronic energy scale uncertainty. Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty.

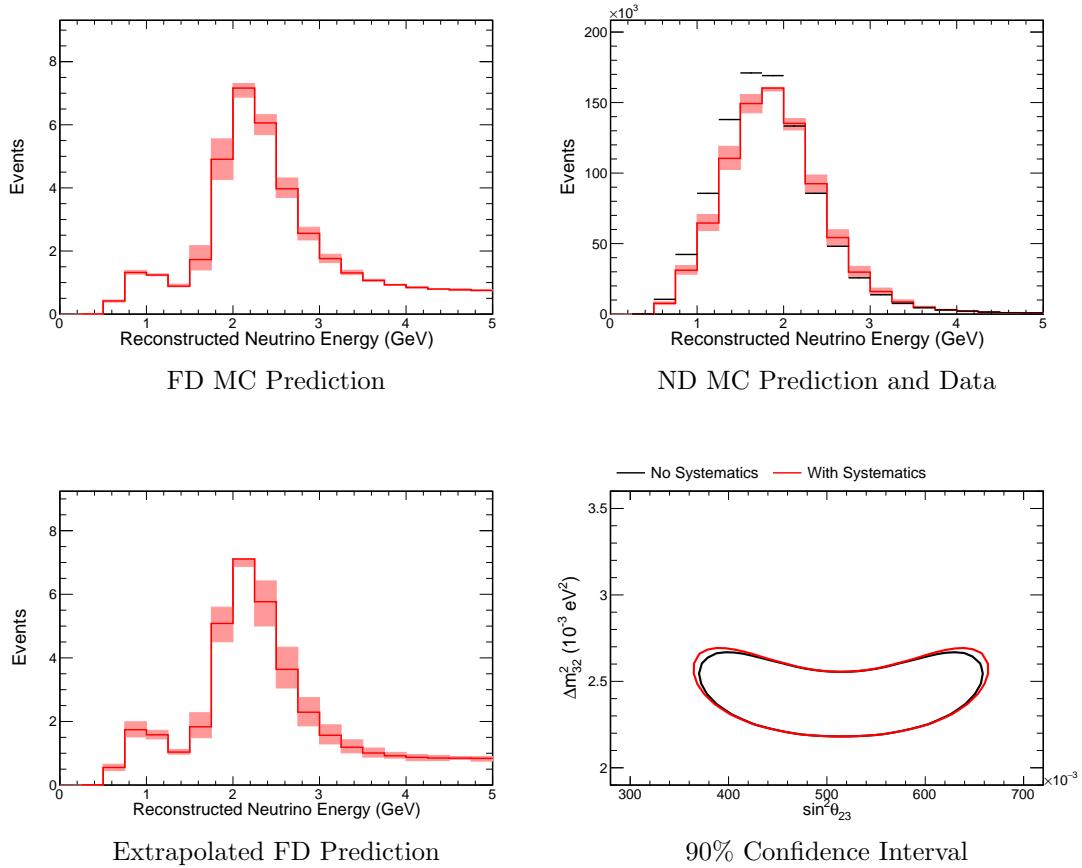


Figure 10.39: Systematic effect of relative hadronic energy scale uncertainty. Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty.

10.10 Negligible Uncertainties

Certain uncertainties have been investigated and deemed to be negligible in this analysis. The uncertainty in protons-on-target (POT) accounting was estimated by comparing the primary NuMI intensity monitor to upstream intensity monitors in the Main Injector and found to be remarkably stable [98]. Studies of the effect of dead channels (APD pixels) involved randomly masking off up to 10% of channels and observing the result in reconstruction [101]. The effect of these masks was found to be negligible.

10.11 Total Systematic Uncertainty

All of the individual systematic uncertainties are combined in the final prediction. The combined effect of all systematic uncertainties can be seen in Figure 10.40.

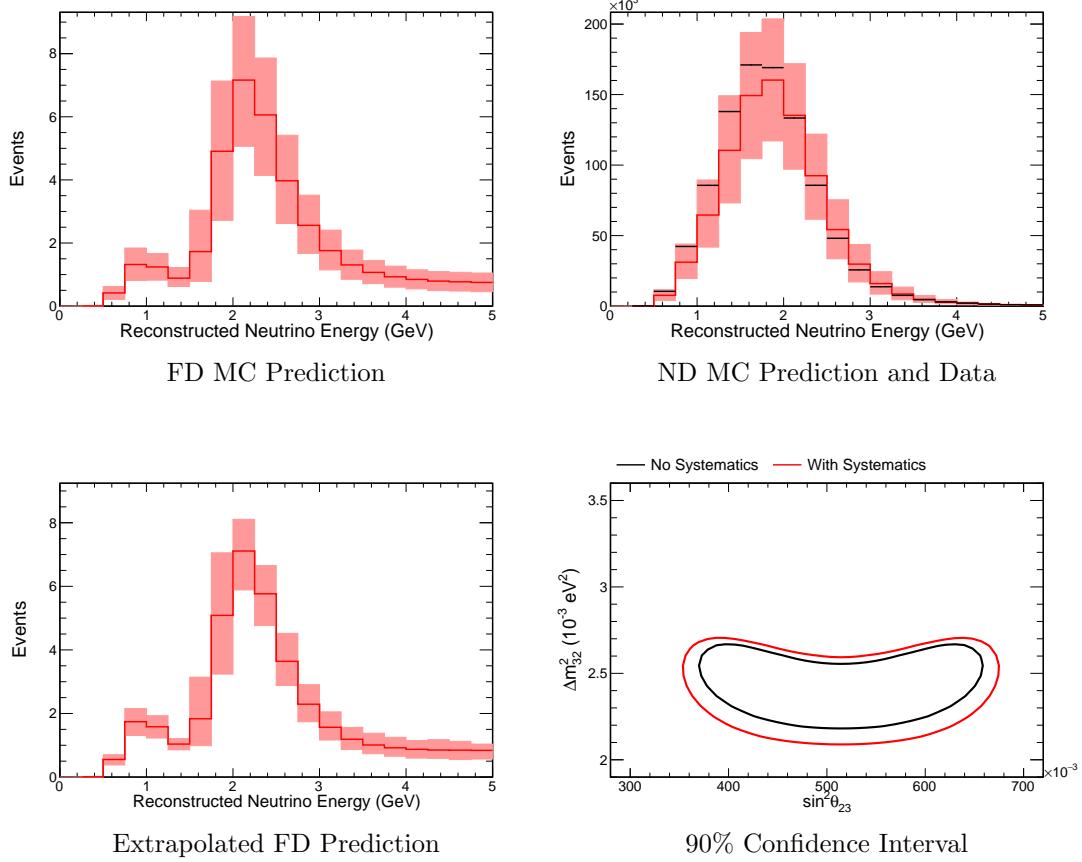


Figure 10.40: Effect of total systematic uncertainty

Systematic effects can be seen in the predictions and confidence intervals which result. The top left pane shows the FD prediction, while the top right shows the ND prediction and ND data overlaid in black. The result of the extrapolation is shown in the bottom left, in which systematic uncertainties can cancel. The bottom right pane shows 90% confidence intervals with and without the effect of the systematic uncertainty.

Chapter 11

Results

Upon opening the signal region for analysis, 39 events were selected. Assuming the default oscillation parameters (shown in Table 4.1) the extrapolated prediction is for 38.1 NuMI beam events to be selected with a cosmic-ray background of 3.4 events. In the absence of neutrino oscillation, 222.0 NuMI beam events would have been predicted. The selected spectrum was fit as described in Section 9.2; the best-fit parameters are $\sin^2(\theta_{23}) = 0.43$ (with a statistically-degenerate compliment near 0.6) and $|\Delta m_{32}^2| = 2.48 \times 10^{-3}$ eV². The fit results will be discussed in more detail in Section 11.3.

11.1 FD Selected Sample

It is important to inspect the selected events to confirm that the selection is consistent with expectation. For a variety of distributions which characterize the selection, we can compare the distribution from selected FD events to the FD MC prediction. The following distributions can all be seen in their respective figures:

- CNN softmax output (Figure 11.2)
- Slice N_{Hit} (Figure 11.3)
- Number of tracks formed by KalmanTrack (Figure 11.4)
- Muon track length (Figure 11.5)
- Muon track start X (Figure 11.6)

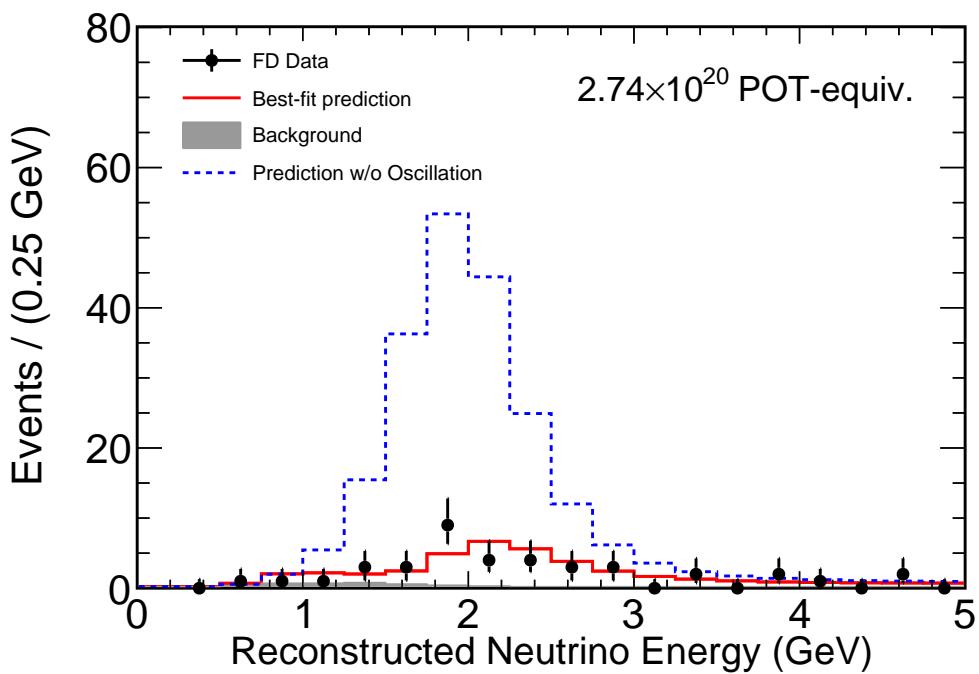


Figure 11.1: FD observed spectrum with prediction in the absence of neutrino oscillation. The distribution of the 39 selected FD events are shown in black. The selected spectrum was fit to obtain the oscillation parameters used for the FD MC prediction shown in red, while the blue dashed line shows the FD MC prediction in the absence neutrino oscillation. The background spectrum is shown in gray.

- Muon track start Y (Figure 11.7)
- Muon track start Z (Figure 11.8)
- Muon track stop X (Figure 11.9)
- Muon track stop Y (Figure 11.10)
- Muon track stop Z (Figure 11.11)
- Muon track Z -direction cosine (Figure 11.12)
- Hadronic cluster N_{Hit} (Figure 11.13)
- Hadronic energy (Figure 11.14)
- Muon ID (Figure 11.15).

None of the distributions show significant excursions relative to the MC prediction. The consistency between FD data and the MC prediction provides a reasonable measure of confidence in the event selection, although it is hard to make a very strong statement with such limited statistics.

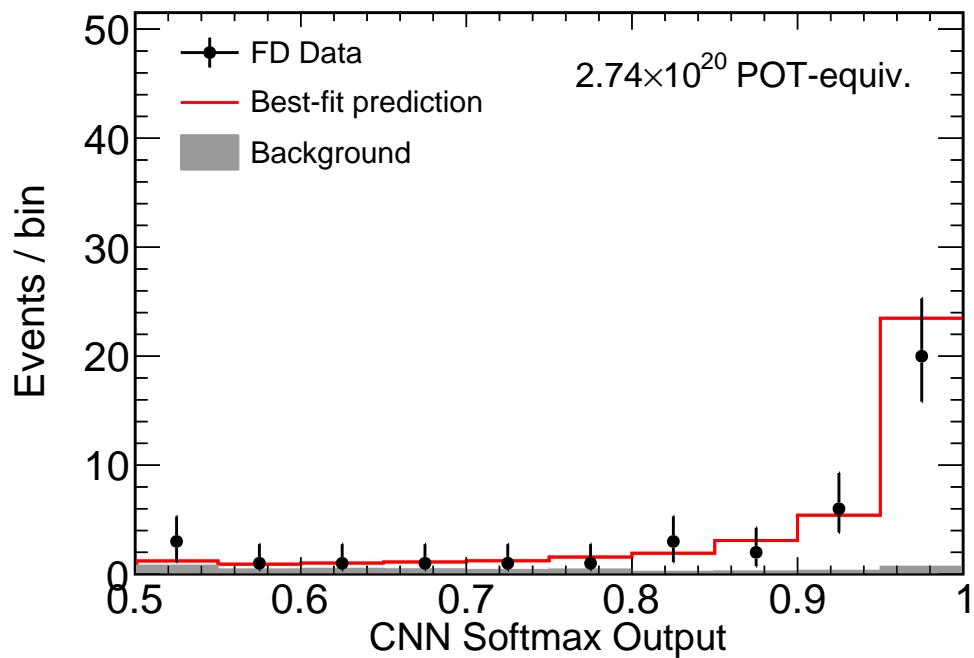


Figure 11.2: CNN softmax output distribution for selected FD events with MC prediction

The distribution of the 39 selected FD events are shown in black. The selected spectrum was fit to obtain the oscillation parameters used for the FD MC prediction shown in red, while the blue dashed line shows the FD MC prediction in the absence neutrino oscillation. The background spectrum is shown in gray.

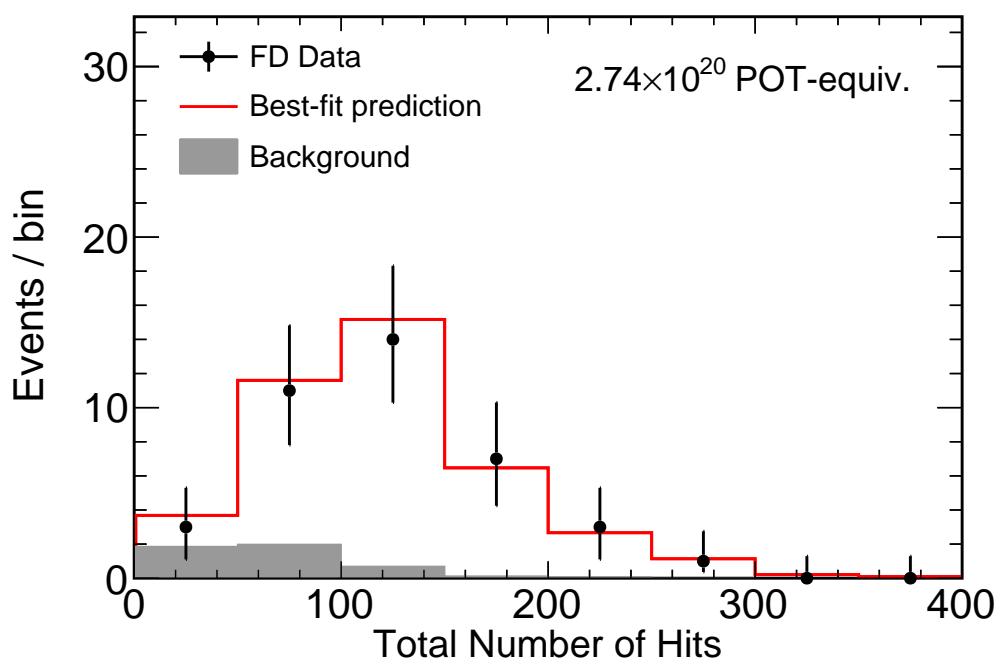


Figure 11.3: Slice N_{Hit} distribution for selected FD events with MC prediction. The distribution of the 39 selected FD events are shown in black. The selected spectrum was fit to obtain the oscillation parameters used for the FD MC prediction shown in red, while the blue dashed line shows the FD MC prediction in the absence neutrino oscillation. The background spectrum is shown in gray.

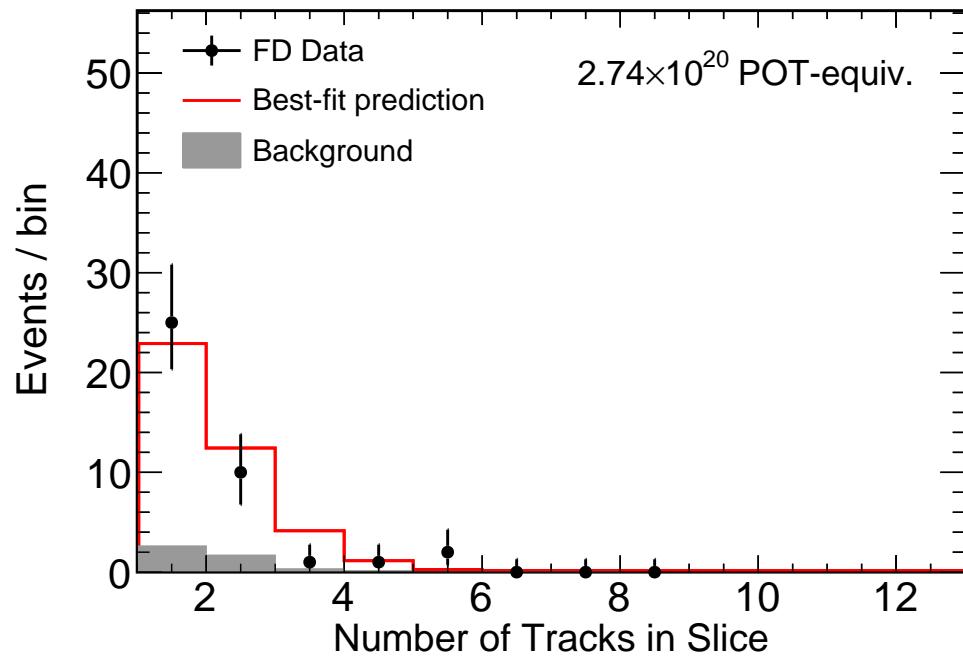


Figure 11.4: Number of tracks formed by KalmanTrack distribution for selected FD events with MC prediction

The distribution of the 39 selected FD events are shown in black. The selected spectrum was fit to obtain the oscillation parameters used for the FD MC prediction shown in red, while the blue dashed line shows the FD MC prediction in the absence neutrino oscillation. The background spectrum is shown in gray.

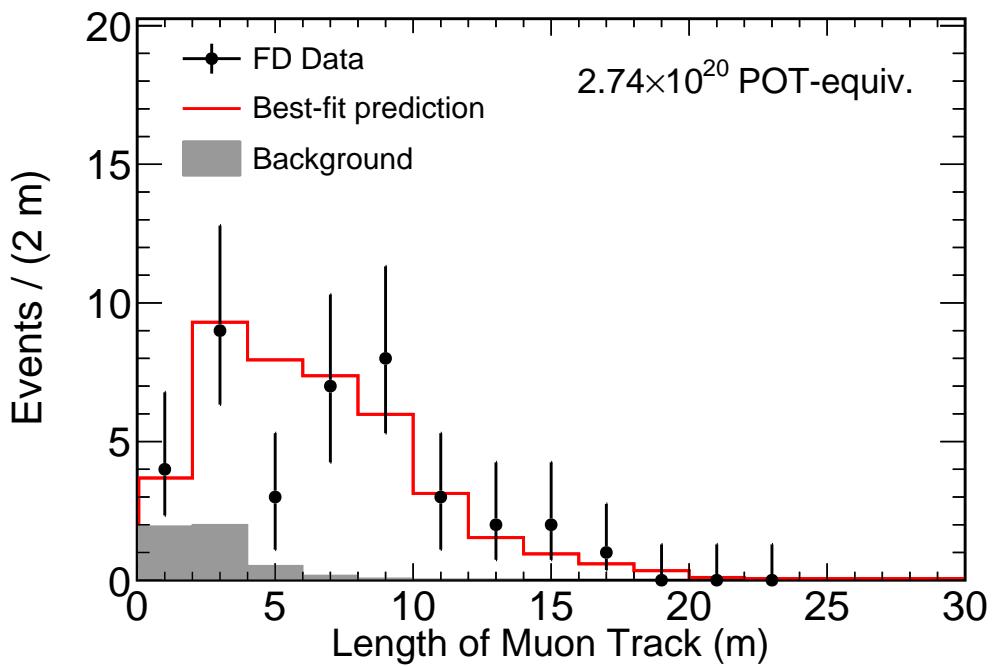


Figure 11.5: Muon track length distribution for selected FD events with MC prediction. The distribution of the 39 selected FD events are shown in black. The selected spectrum was fit to obtain the oscillation parameters used for the FD MC prediction shown in red, while the blue dashed line shows the FD MC prediction in the absence neutrino oscillation. The background spectrum is shown in gray.

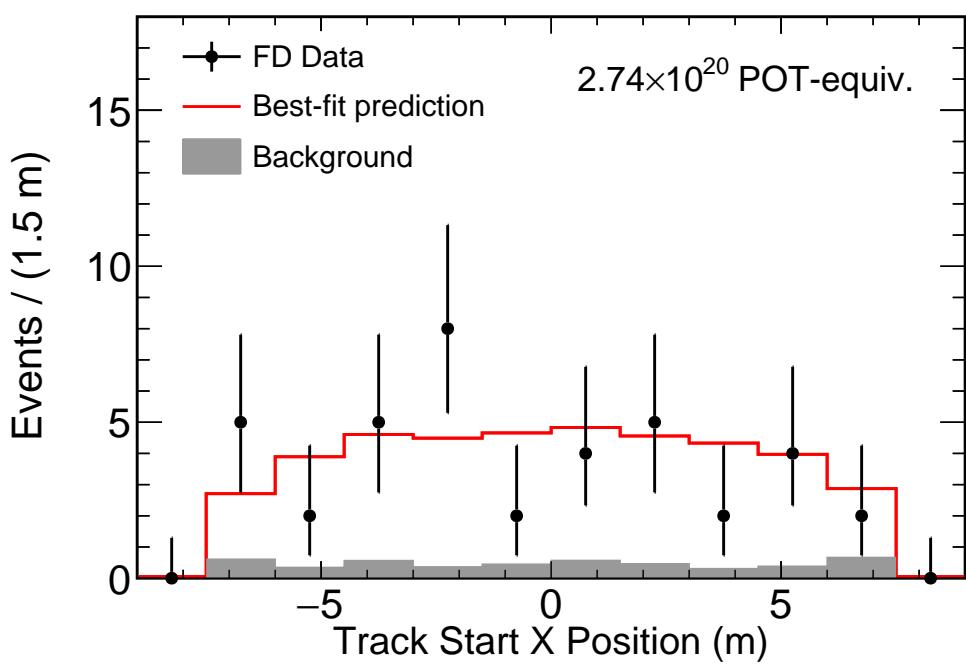


Figure 11.6: Muon track start X distribution for selected FD events with MC prediction. The distribution of the 39 selected FD events are shown in black. The selected spectrum was fit to obtain the oscillation parameters used for the FD MC prediction shown in red, while the blue dashed line shows the FD MC prediction in the absence neutrino oscillation. The background spectrum is shown in gray.

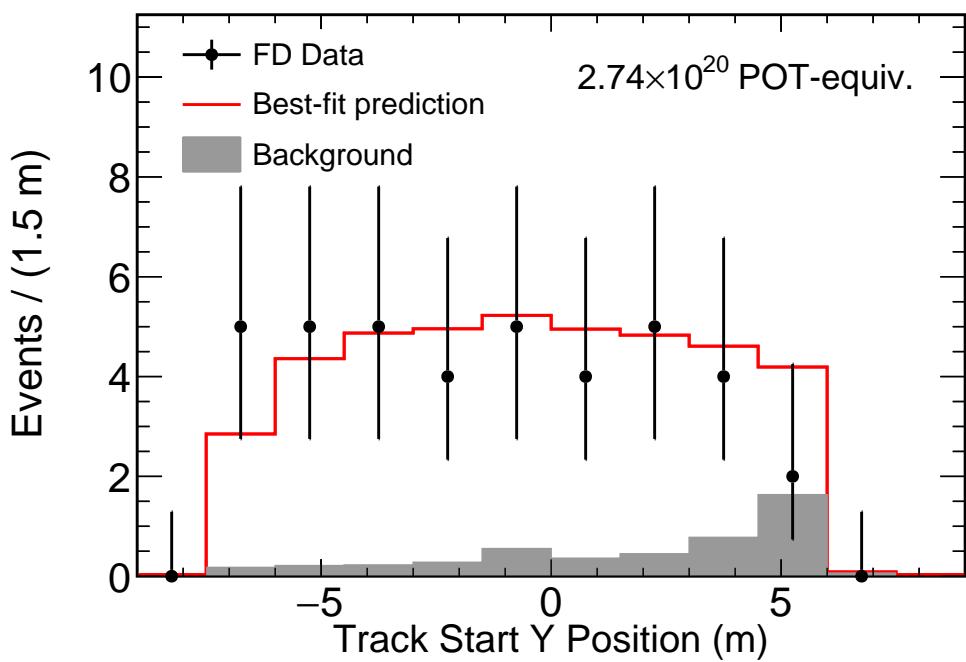


Figure 11.7: Muon track start Y distribution for selected FD events with MC prediction. The distribution of the 39 selected FD events are shown in black. The selected spectrum was fit to obtain the oscillation parameters used for the FD MC prediction shown in red, while the blue dashed line shows the FD MC prediction in the absence neutrino oscillation. The background spectrum is shown in gray.

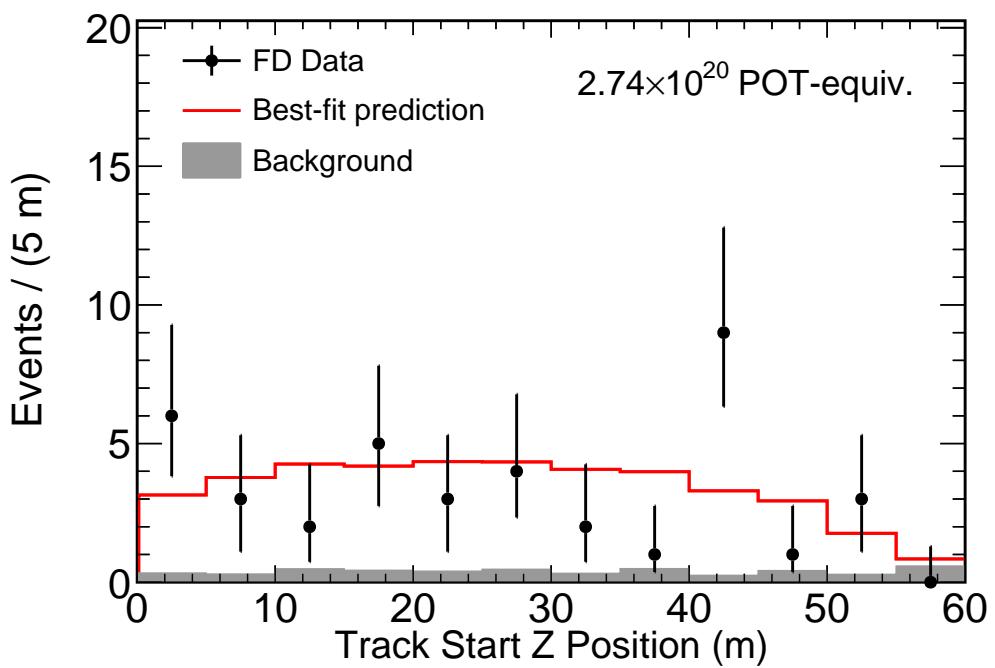


Figure 11.8: Muon track start Z distribution for selected FD events with MC prediction. The distribution of the 39 selected FD events are shown in black. The selected spectrum was fit to obtain the oscillation parameters used for the FD MC prediction shown in red, while the blue dashed line shows the FD MC prediction in the absence neutrino oscillation. The background spectrum is shown in gray.

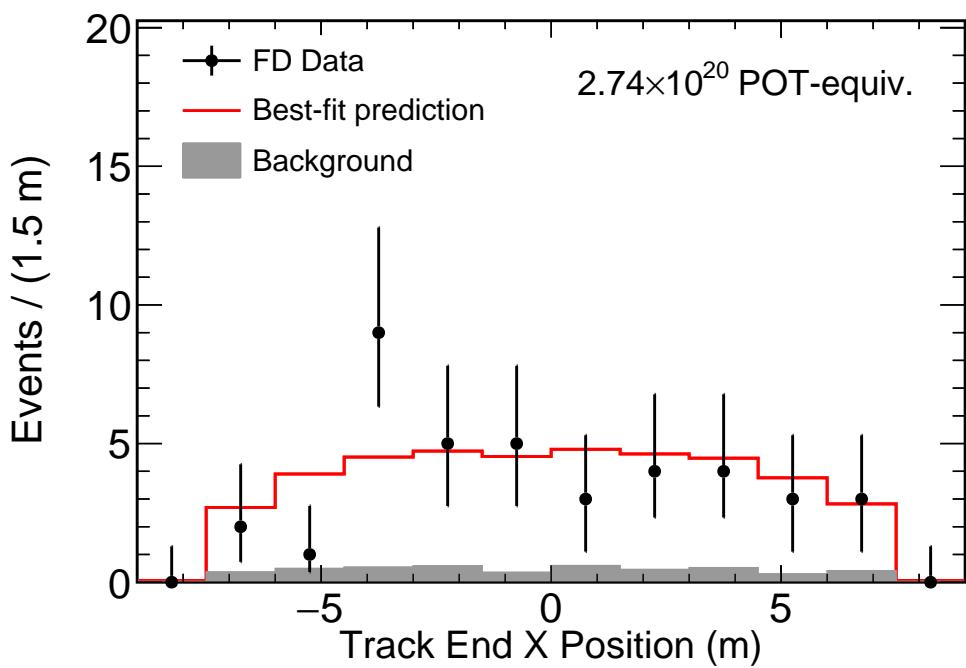


Figure 11.9: Muon track stop X distribution for selected FD events with MC prediction. The distribution of the 39 selected FD events are shown in black. The selected spectrum was fit to obtain the oscillation parameters used for the FD MC prediction shown in red, while the blue dashed line shows the FD MC prediction in the absence neutrino oscillation. The background spectrum is shown in gray.

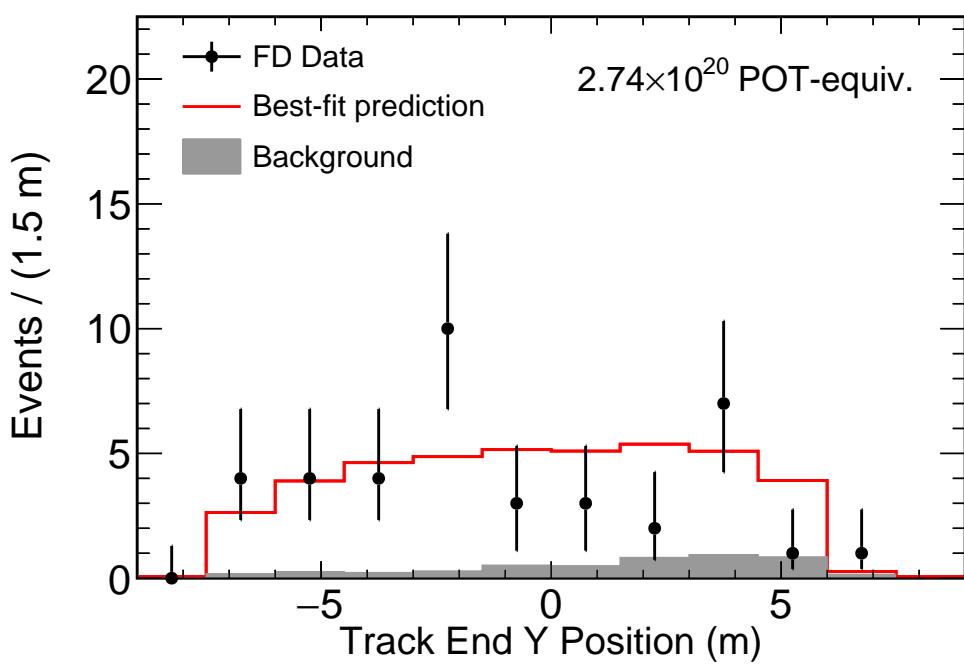


Figure 11.10: Muon track stop Y distribution for selected FD events with MC prediction. The distribution of the 39 selected FD events are shown in black. The selected spectrum was fit to obtain the oscillation parameters used for the FD MC prediction shown in red, while the blue dashed line shows the FD MC prediction in the absence neutrino oscillation. The background spectrum is shown in gray.

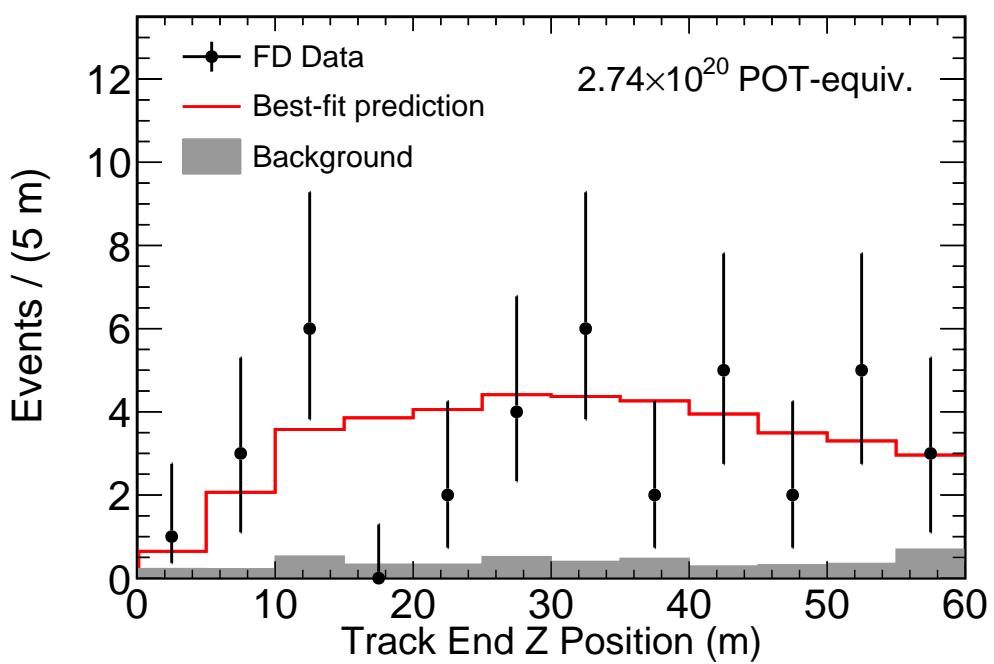


Figure 11.11: Muon track stop Z distribution for selected FD events with MC prediction. The distribution of the 39 selected FD events are shown in black. The selected spectrum was fit to obtain the oscillation parameters used for the FD MC prediction shown in red, while the blue dashed line shows the FD MC prediction in the absence neutrino oscillation. The background spectrum is shown in gray.

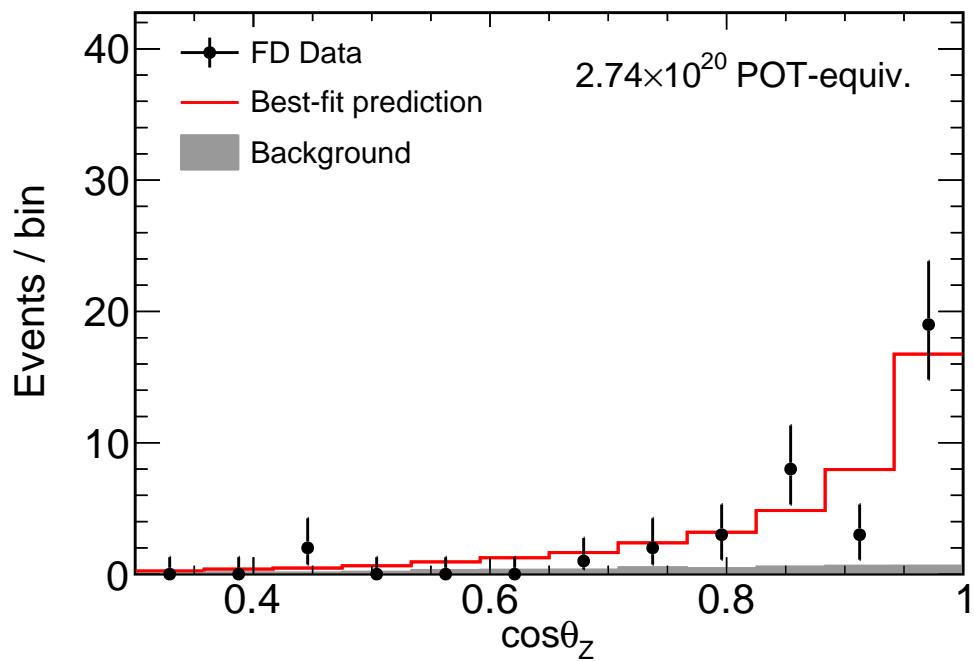


Figure 11.12: Muon track Z -direction cosine distribution for selected FD events with MC prediction

The distribution of the 39 selected FD events are shown in black. The selected spectrum was fit to obtain the oscillation parameters used for the FD MC prediction shown in red, while the blue dashed line shows the FD MC prediction in the absence neutrino oscillation. The background spectrum is shown in gray.

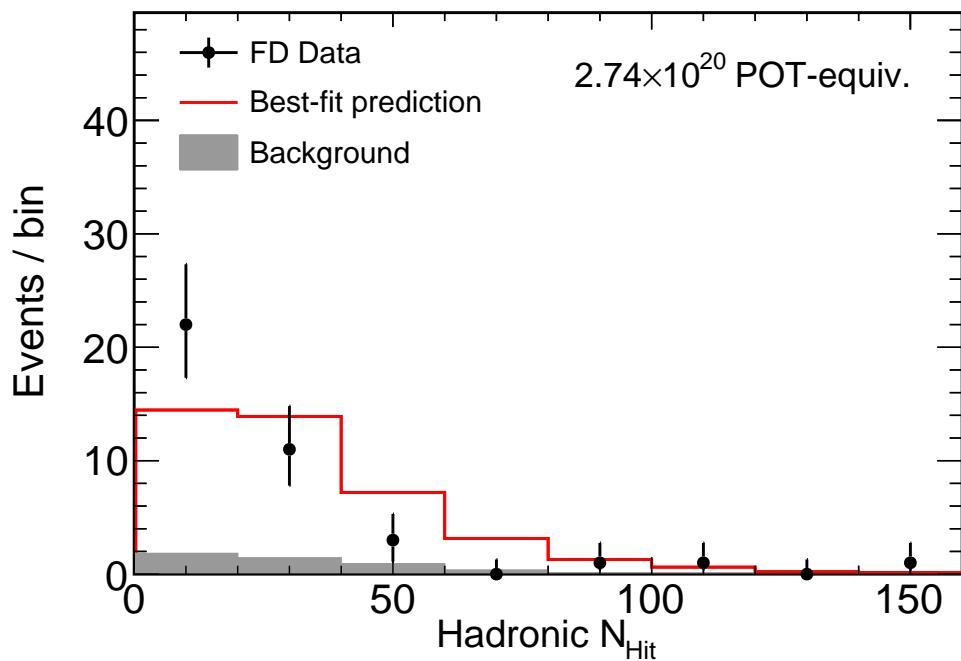


Figure 11.13: Hadronic cluster N_{Hit} distribution for selected FD events with MC prediction

The distribution of the 39 selected FD events are shown in black. The selected spectrum was fit to obtain the oscillation parameters used for the FD MC prediction shown in red, while the blue dashed line shows the FD MC prediction in the absence of neutrino oscillation. The background spectrum is shown in gray.

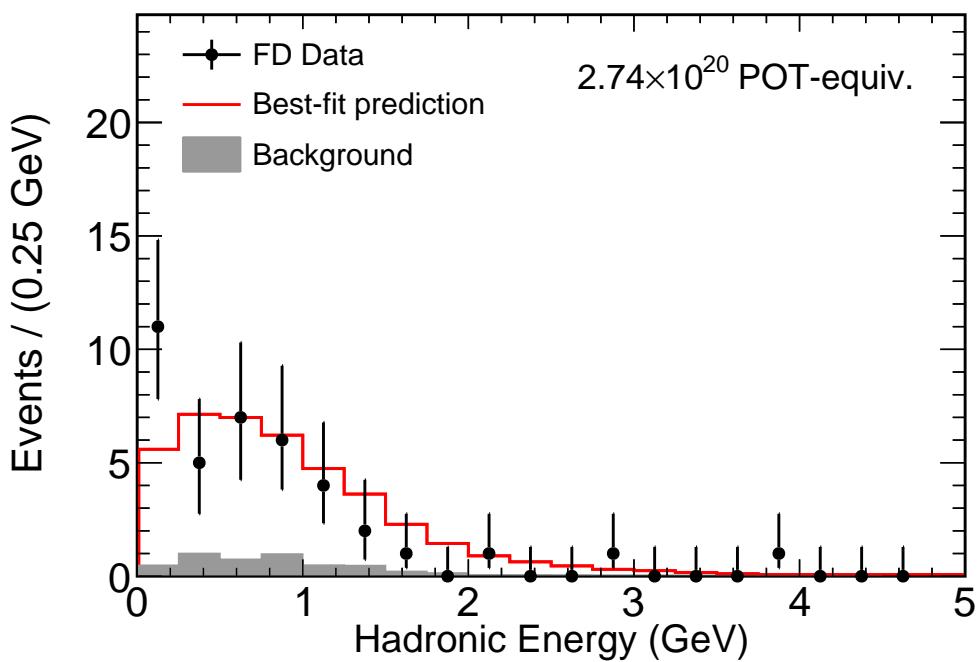


Figure 11.14: Hadronic energy distribution for selected FD events with MC prediction. The distribution of the 39 selected FD events are shown in black. The selected spectrum was fit to obtain the oscillation parameters used for the FD MC prediction shown in red, while the blue dashed line shows the FD MC prediction in the absence of neutrino oscillation. The background spectrum is shown in gray.

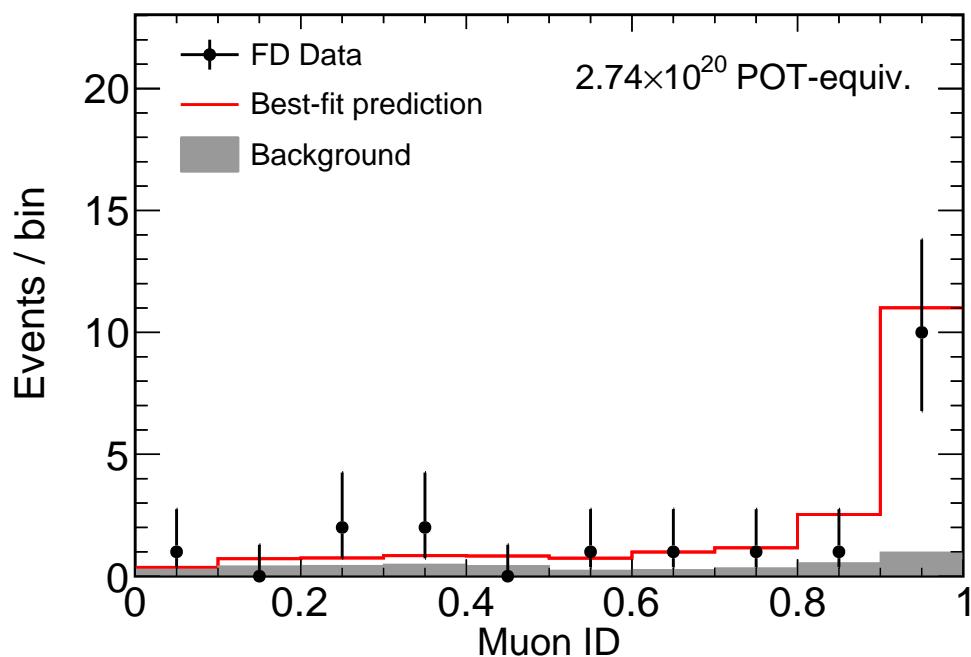


Figure 11.15: Muon ID distribution for selected FD events with MC prediction. The distribution of the 39 selected FD events are shown in black. The selected spectrum was fit to obtain the oscillation parameters used for the FD MC prediction shown in red, while the blue dashed line shows the FD MC prediction in the absence neutrino oscillation. The background spectrum is shown in gray.

11.2 Selected Event Displays

This section shows a gallery of selected FD events. Figures 11.16 and 11.17 are candidate ν_μ CC interactions, Figures 11.18 and 11.19 are candidate ν_μ RES interactions, and Figures 11.20 and 11.21 are candidate ν_μ CC DIS interactions.

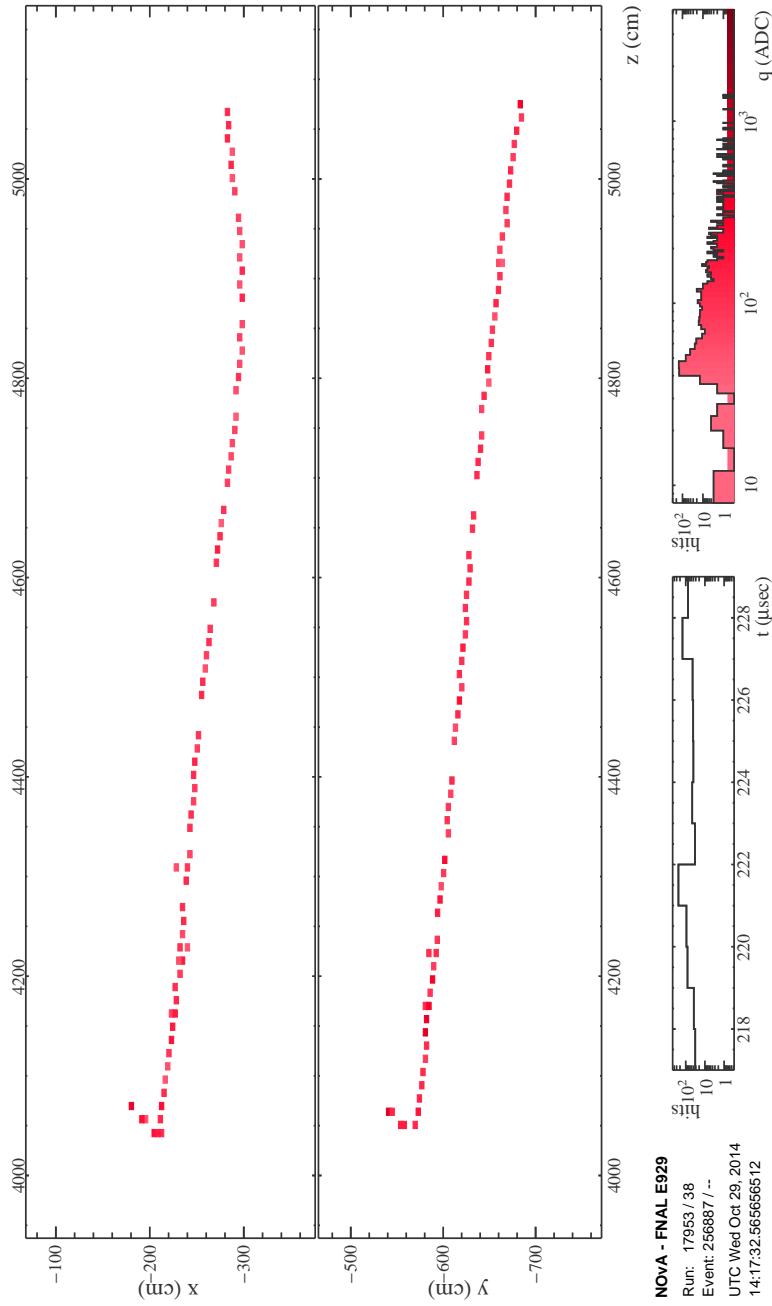
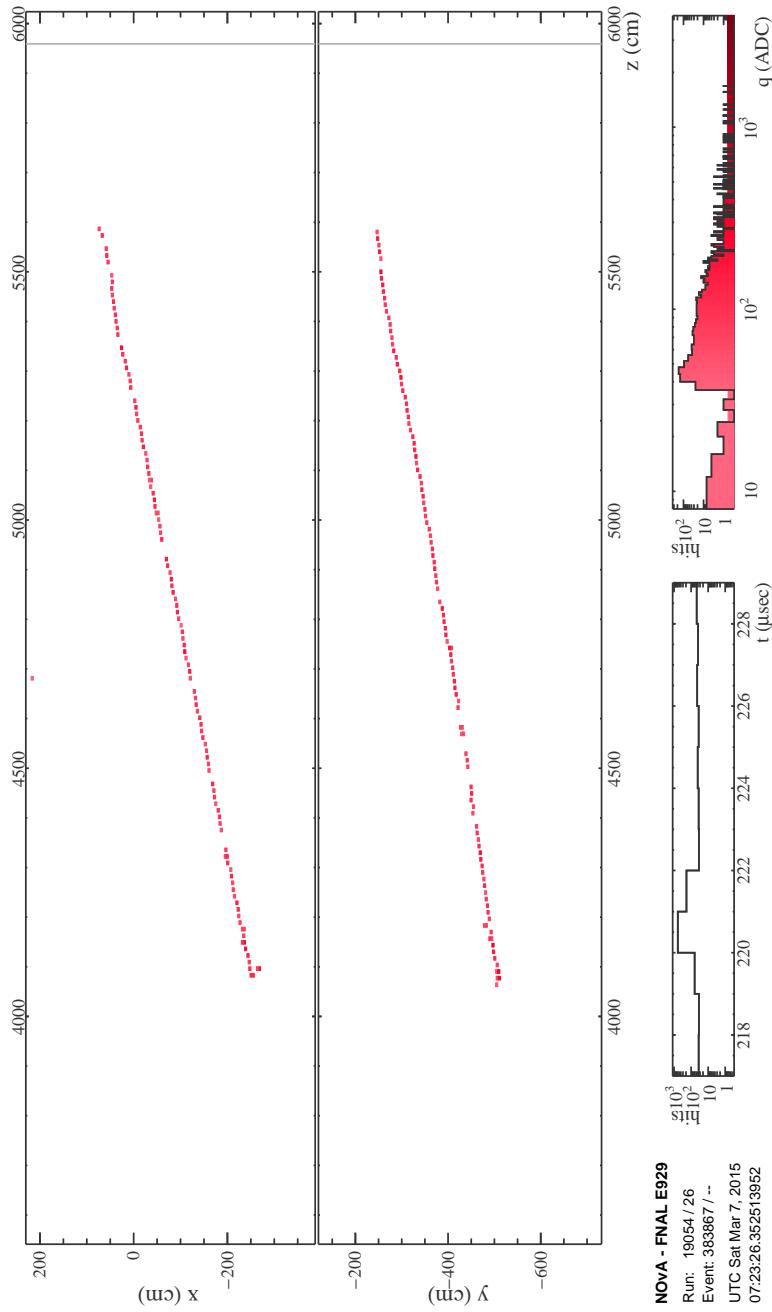


Figure 11.16: Selected FD ν_μ CC QE candidate
(Rotated left) The top projection is an X vs. Z view, the bottom is Y vs. Z . The bottom left portion of the display This is a selected ν_μ charged-current event candidate from the Far Detector. With a long muon track and short secondary track, this event bears the signature of a CC QE interaction.



200

Figure 11.17: Selected FD ν_μ CC QE candidate

(Rotated left) The top projection is an X vs. Z view, the bottom is Y vs. Z . The bottom left portion of the display This is a selected ν_μ charged-current event candidate from the Far Detector. With a long muon track and hardly any hadronic activity, this event bears the signature of a CC QE interaction.

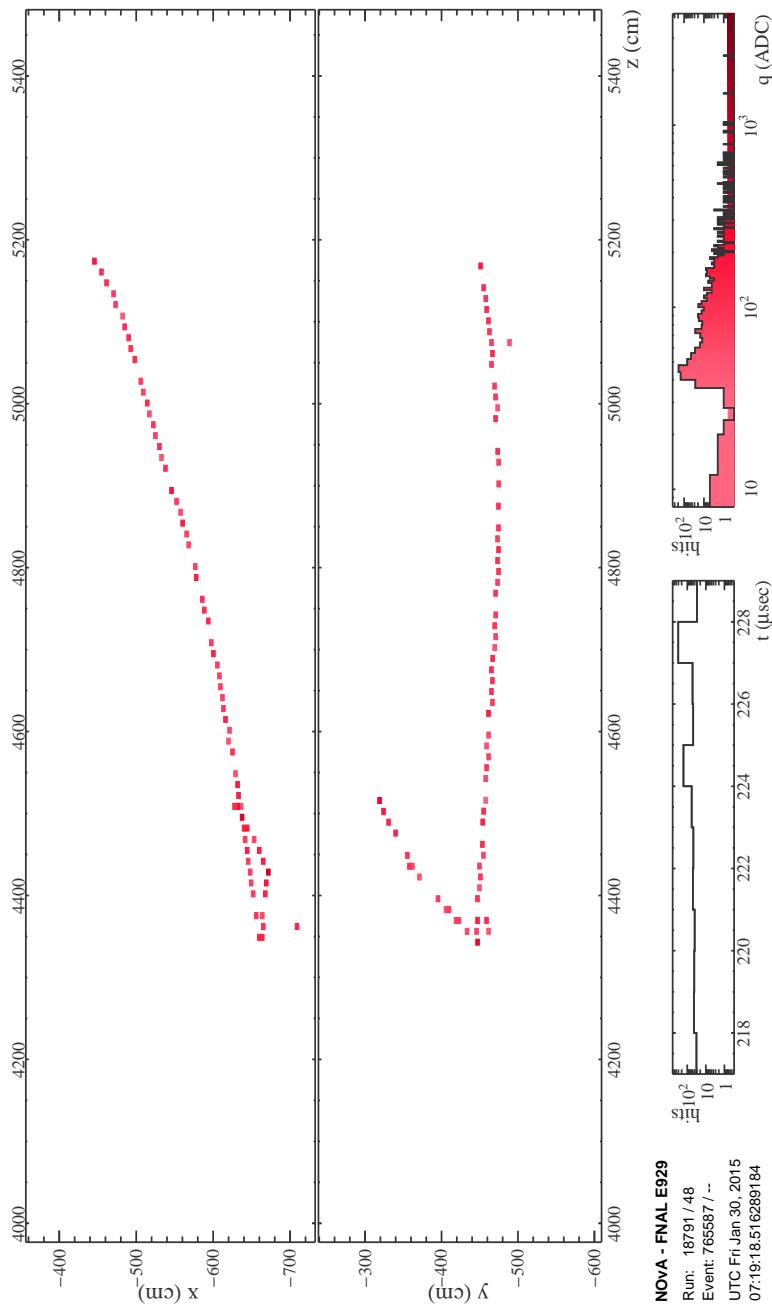


Figure 11.18: Selected FD ν_μ CC QE candidate
(Rotated left) The top projection is an X vs. Z view, the bottom is Y vs. Z . The bottom left portion of the display This is a selected ν_μ charged-current event candidate from the Far Detector. The relatively long secondary track possibly a pion, which suggests this event is a CC RES interaction.

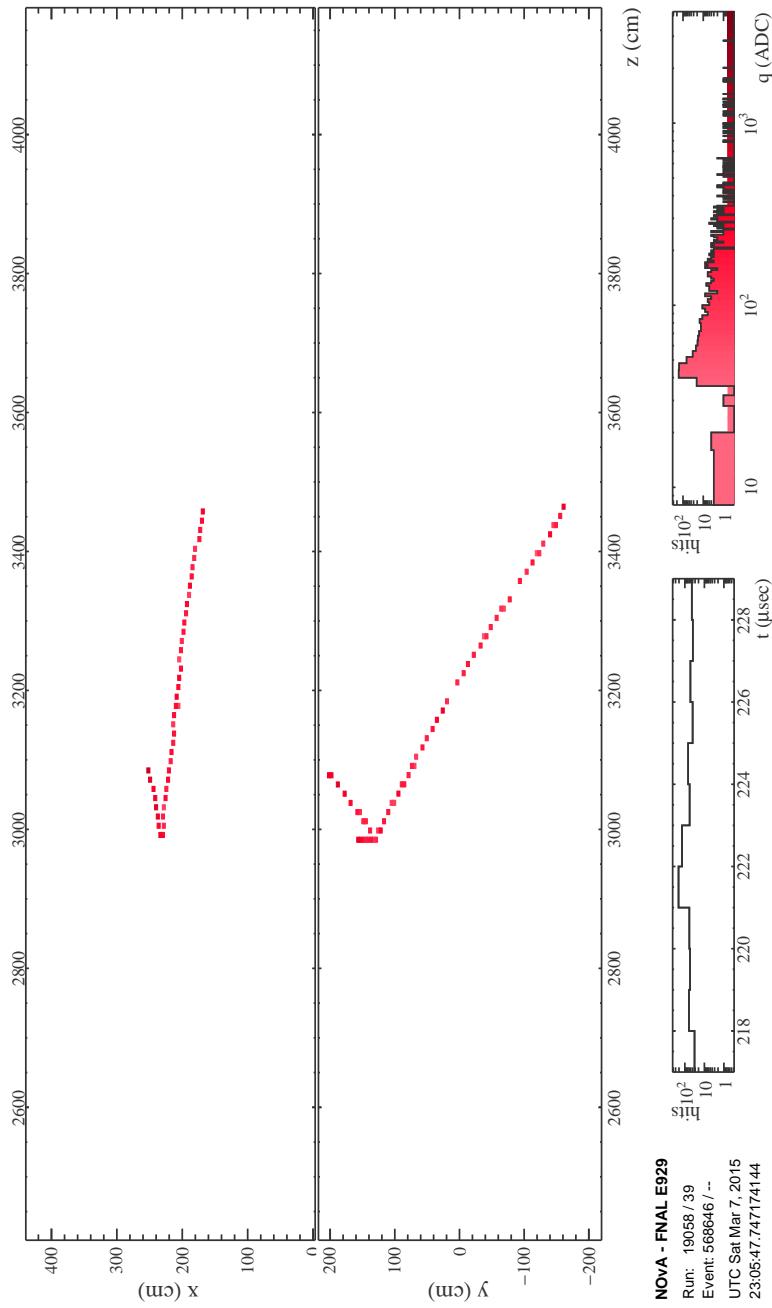


Figure 11.19: Selected FD ν_μ CC QE candidate

(Rotated left) The top projection is an X vs. Z view, the bottom is Y vs. Z . The bottom left portion of the display This is a selected ν_μ charged-current event candidate from the Far Detector. The relatively long secondary track possibly a pion, which suggests this event is a CC RES interaction.

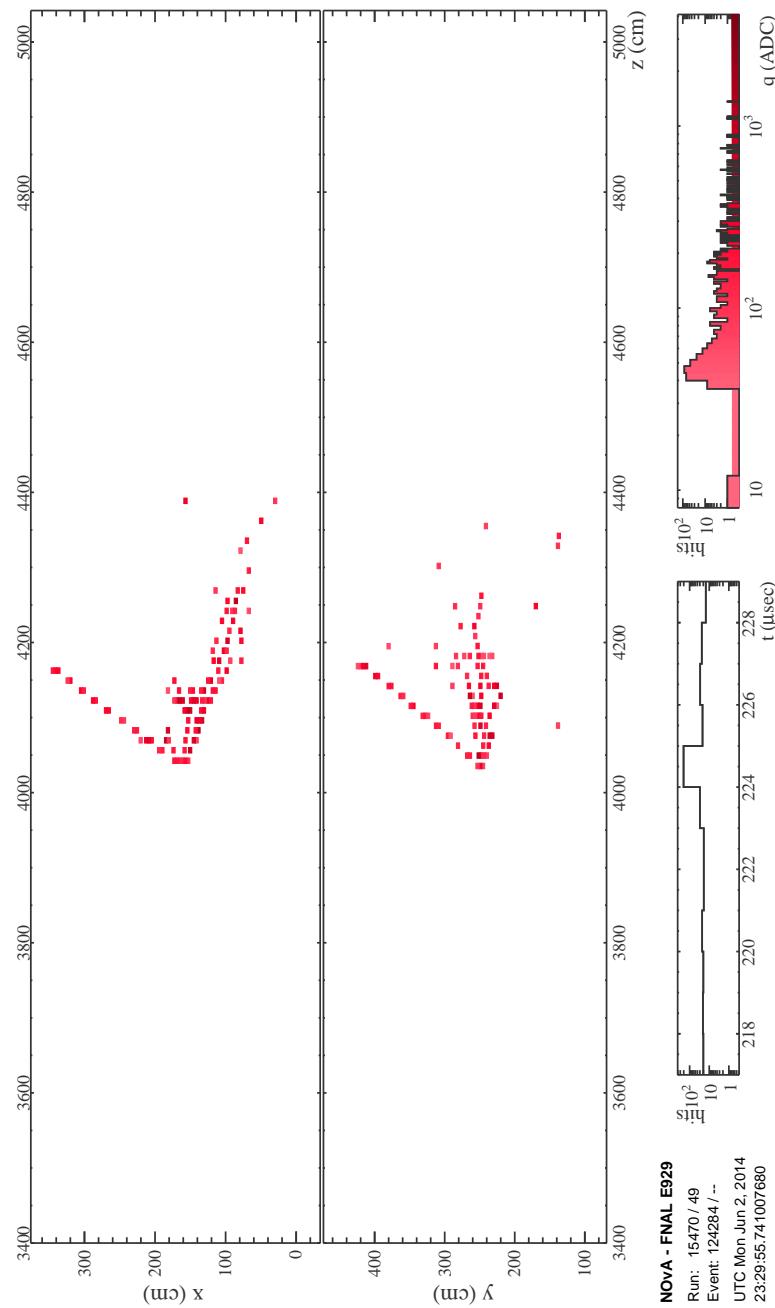


Figure 11.20: Selected FD ν_μ CC QE candidate

(Rotated left) The top projection is an X vs. Z view, the bottom is Y vs. Z . The bottom left portion of the display This is a selected ν_μ charged-current event candidate from the Far Detector. The extensive hadronic shower in this event indicates that it could be a CC DIS interaction.

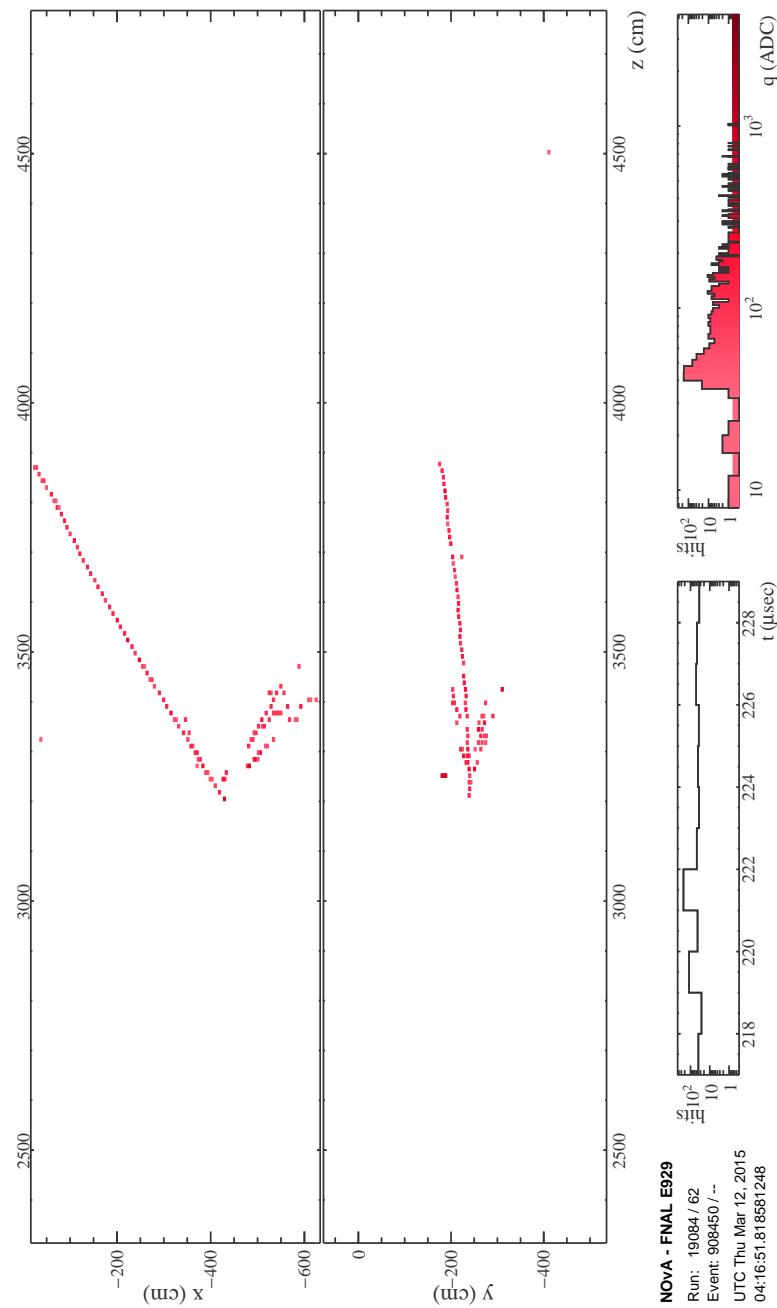


Figure 11.21: Selected FD ν_μ CC QE candidate

(Rotated left) The top projection is an X vs. Z view, the bottom is Y vs. Z . The bottom left portion of the display This is a selected ν_μ charged-current event candidate from the Far Detector. The extensive hadronic shower in this event indicates that it could be a CC DIS interaction.

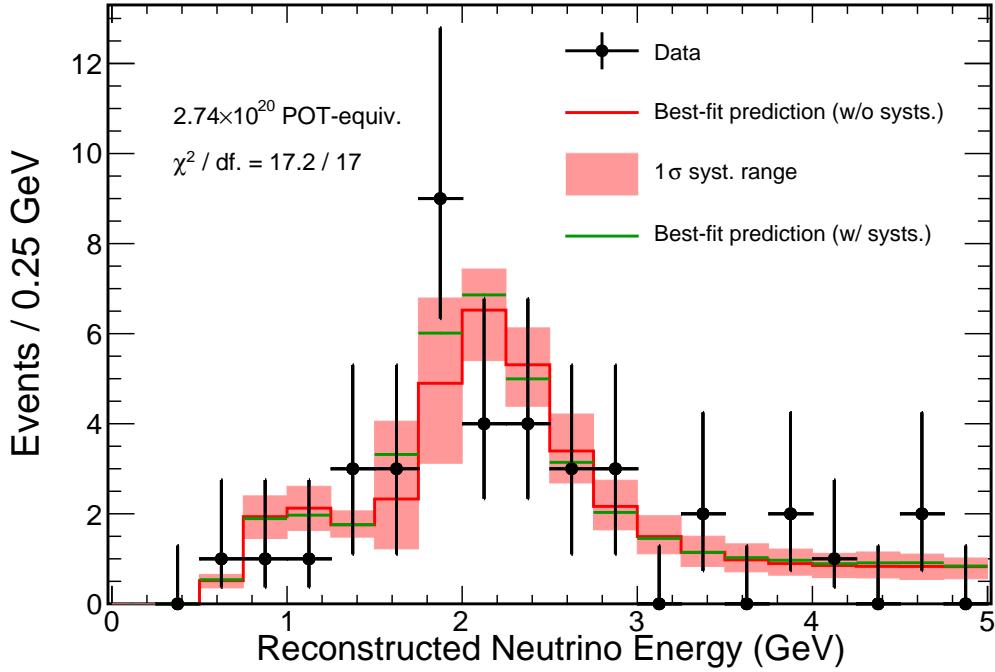


Figure 11.22: FD observed spectrum with best-fit and systematic uncertainties. The distribution of the 39 selected FD events are shown as black points with associated Poisson error bars. The selected spectrum was fit to obtain the oscillation parameters used for the extrapolated prediction shown as a red line, with the 1σ uncertainty band shown as the red shaded region. The green spectrum is the best-fit prediction including the effects of systematic uncertainties.

11.3 Fit Results

The FD data were fit with the extrapolated FD prediction [95] as described in Chapter 9. Systematic uncertainties, listed in Chapter 10, were included in the fit with a penalty term [42, 96]. Figure 11.22 shows the best-fit prediction with and without uncertainties, overlaid with FD data and the predicted 1σ systematic uncertainty band.

Confidence intervals have been extracted from the fit as described in Section 9.2. Figure 11.23 shows the 90% confidence interval with and without the effect of systematic errors. the best-fit parameters are $|\Delta m_{32}^2| = 2.48 \times 10^{-3}$ eV 2 . and $\sin^2(\theta_{23}) = 0.43$. It should be noted that the parameter space is symmetric in the $\sin^2(\theta_{23})$ dimension, which

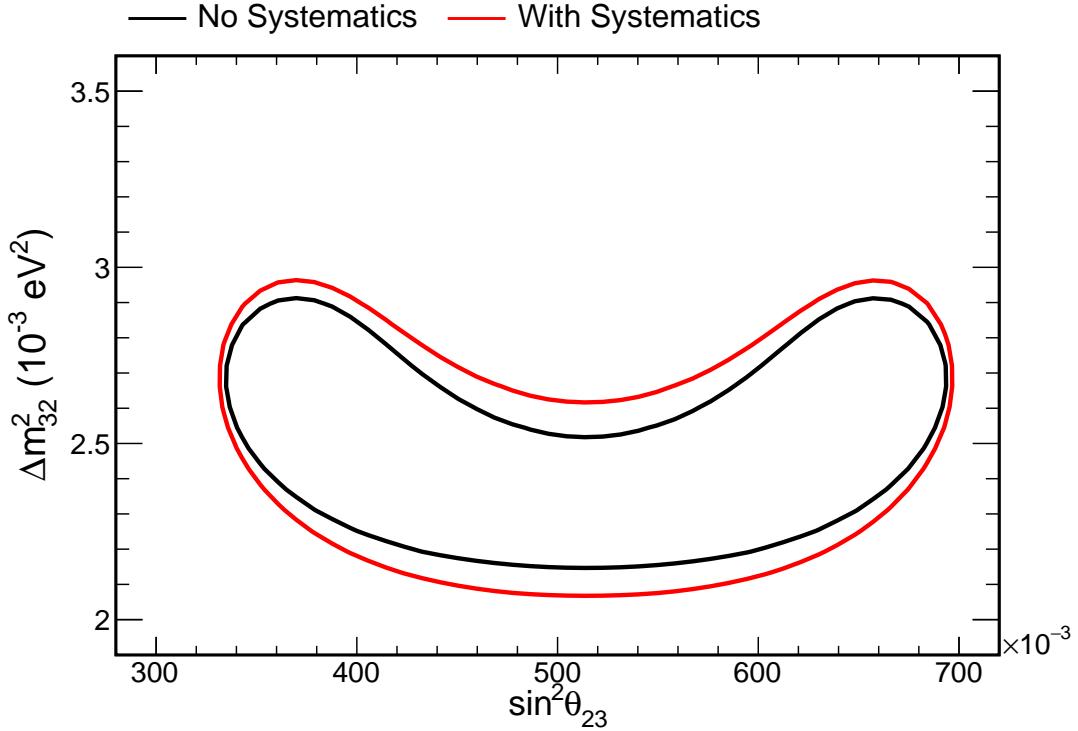


Figure 11.23: Measured 90% confidence interval with and without systematic uncertainties

The confidence interval with the effect of systematic uncertainties is shown in red and without systematic uncertainties in black. The best-fit parameters are $\sin^2(\theta_{23}) = 0.43$ (with a statistically-degenerate compliment near 0.6) and $|\Delta m_{32}^2| = 2.48 \times 10^{-3}$ eV 2 .

means that a statistically equivalent best-fit point is located at roughly $\sin^2(\theta_{23}) = 0.6$

The confidence intervals from this analysis are compared to recent results from T2K [102] and MINOS [103] in Figure 11.24.

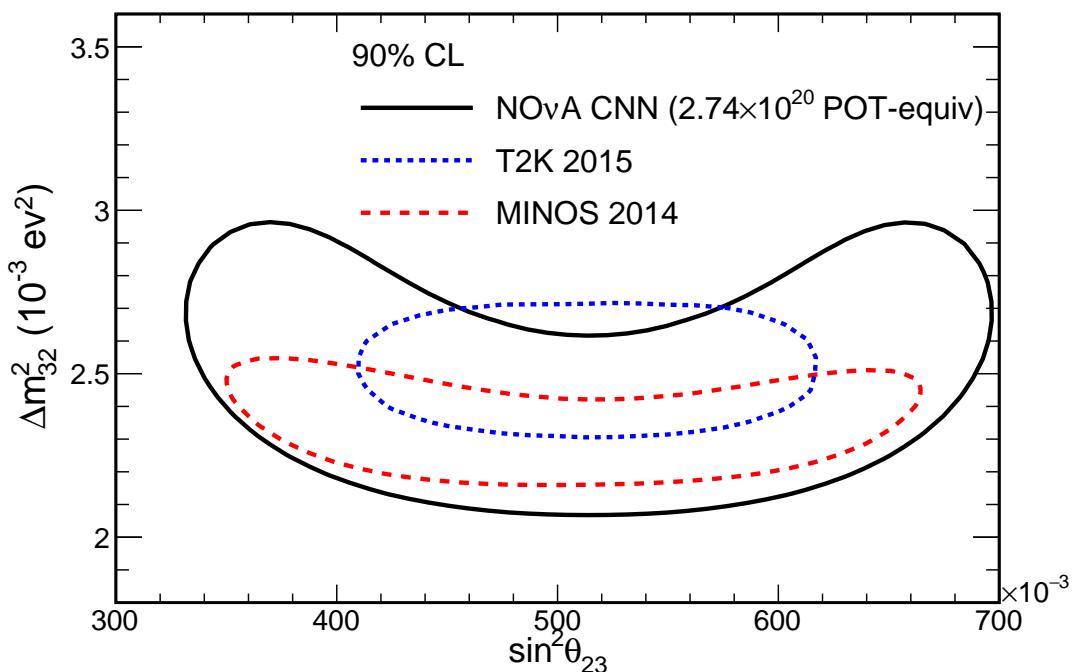


Figure 11.24: Comparison of 90% confidence interval to existing results
The confidence interval from this analysis (including systematic uncertainties) is shown in black. Recent results from T2K [102] and MINOS [103] are overlaid.

11.4 Feldman-Cousins Confidence Interval

An alternative fitting procedure from [97, 96] has been used to produce frequentist confidence intervals with proper 90% coverage. The procedure is described in Section 9.2.2. The results can be seen in Figure 11.25.

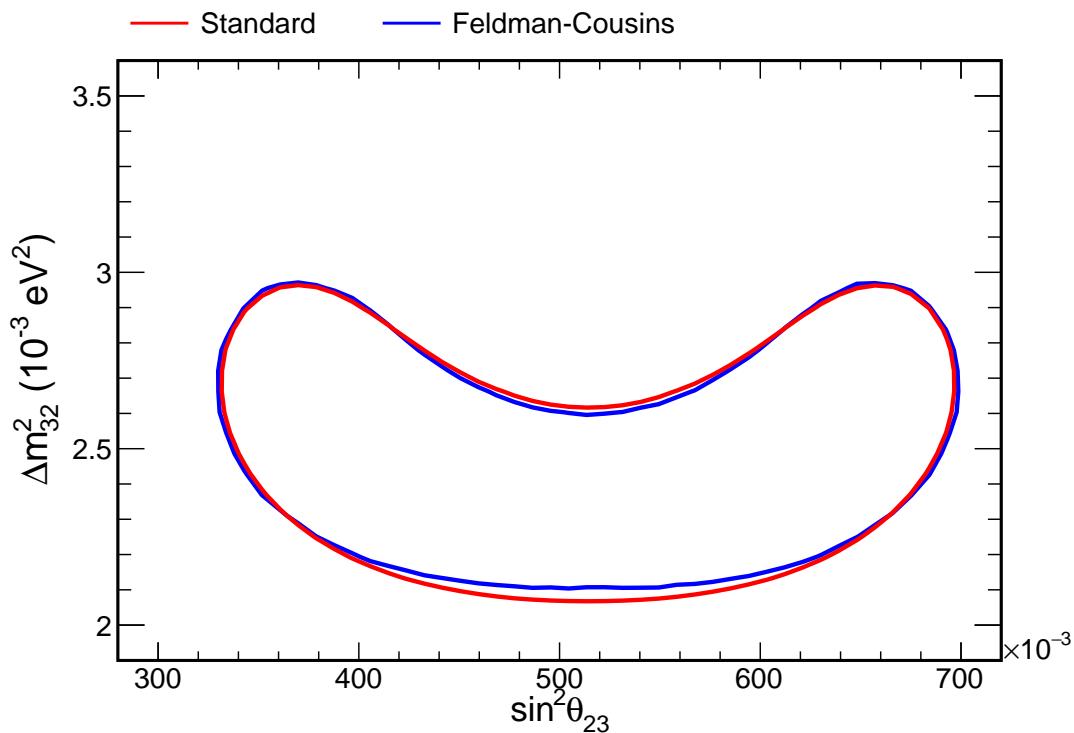


Figure 11.25: Feldman-Cousins 90% confidence interval

The Feldman-Cousins procedure [97] produces confidence intervals with proper coverage by running many pseudo-experiments, each with alternate statistical fluctuations. A surface critical χ^2 values is determined based on the percentile of pseudo-experiments. Points on the likelihood surface are compared to the critical value at that point to determine inclusion in the confidence interval. In this case, the Feldman-Cousins confidence interval is quite similar to the standard Bayesian approach.

Chapter 12

Conclusion

Even though it employs just a fraction the designed NOvA exposure, the neutrino oscillation result presented in Chapter 11 is nonetheless compelling. The fit shows a slight preference for *non-maximal* mixing, that is, $\sin^2(\theta_{23})$ not equal to 0.5. If the best-fit value holds, future NOvA measurements should be able to resolve maximal mixing from non-maximal mixing. Predicted sensitivity contours for the exposure in the NOvA design (six years, 14 kt, 700 kW) [46] are shown in Figure 12.1.

The selection technique presented here demonstrates a novel approach to event classification. While convolutional neural networks have delivered considerable improvements to the field of computer vision and recognition [81], the application to high-energy physics is completely new. As shown in Chapter 8, this technique offers improved physics sensitivity through enhanced signal efficiency and background rejection relative to traditional feature-extraction based approaches.

While the result shown here is compelling, additional effort could provide further improvement. The network architecture and training described in Chapter 7 has only been coarsely optimized. Reconfiguration of the network – for instance, removal of the aggressive pooling in the early layers – could improve the recognition capability of the network. It is also not clear that the learning capacity of the network has been saturated; simply training the network through more iterations could improve its performance. Further, the event selection presented in Chapter 8 has also been only crudely optimized. A grid-search optimization of the selection criteria based on physics sensitivity could surely narrow the confidence intervals. Perhaps better yet, a

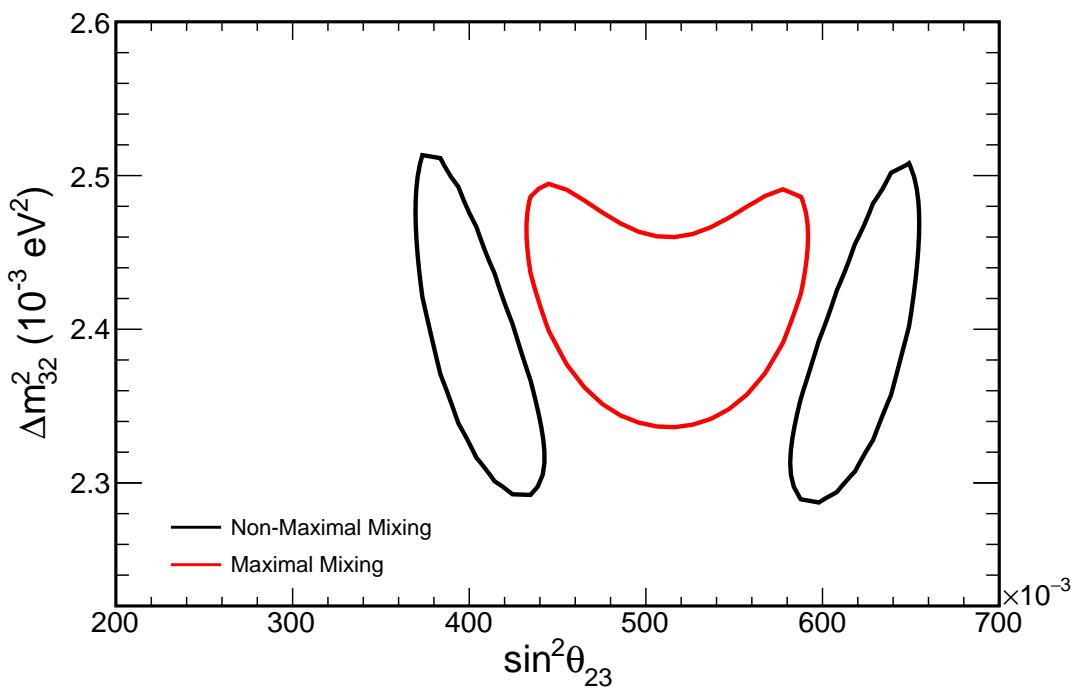


Figure 12.1: Predicted sensitivity of this analysis assuming NOvA design exposure. NOvA has been designed to run for six years with 14 kt of detector mass and 700 kW operation of the NuMI beam. Assuming that exposure, the prediction above shows an ability to resolve maximal mixing ($\sin^2(\theta_{23}) = 0.5$), from a particular non-maximal mixing ($\sin^2(\theta_{23}) = 0.4$) scenario.

multivariate technique based on the background rejection variables shown in Chapter 8.5 could be trained to take advantage of correlations between those variables.

It should also be noted that the existing network can also be used to improve the NOvA ν_e appearance measurement. Studies have shown that this approach also promises improved physics sensitivity for the measurement of δ and potential determination of the neutrino mass hierarchy [104]. It may also be possible to train convolutional neural networks to solve other problems, for instance, identification of reconstructed tracks and prongs. The technique of semantic segmentation [105], where each pixel is given a classification label, could be especially useful for that task. Location of particular features in an event, such as an interaction vertex, could also be achieved with a convolutional neural network. Such an application would be analogous to facial-keypoint detection [81].

The work presented here is clearly just the beginning. NOvA, with its fine-grained structure, provides a very interesting testing ground for many techniques which are successful in the field of computer vision. From a physics perspective, the experiment is also poised to deliver many compelling measurements in the coming years. The result shown here demonstrates a vast potential for exciting developments on the road ahead.

References

- [1] Charles Petit. Heart of the matter: Neutrinos' shifty behavior might help explain why the universe has so much stuff in it. *Science News*, 183(2):18–23, 2013.
- [2] Sylvia Berryman. Ancient atomism. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2008 edition, 2008.
- [3] F. Halzen and A.D. Martin. *Quarks and leptons: an introductory course in modern particle physics*. Wiley, 1984.
- [4] Wikimedia Commons. Standard model of elementary particles, 2006.
- [5] JM Dias, FS Navarra, M Nielsen, and C Zanetti. $z-\{c\}(3900)$ as a tetraquark state: decay width in qcd sum rules. *arXiv preprint arXiv:1311.7591*, 2013.
- [6] J Barth, W Braun, J Ernst, K-H Glander, J Hannappel, N Jöpen, H Kalinowsky, F Klein, E Klempert, R Lawall, et al. Evidence for the positive-strangeness pentaquark θ^+ in photoproduction with the saphir detector at elsa. *Physics Letters B*, 572(3):127–132, 2003.
- [7] H. Becquerel. *Compt. Rend.*, 1896.
- [8] J.J. Thompson. *Philosophical Magazine*, 44, 1897.
- [9] E. Rutherford. Bakerian lecture. nuclear constitution of atoms. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 97(686):pp. 374–400, 1920.
- [10] J. Chadwick. The Existence of a Neutron. *Royal Society of London Proceedings Series A*, 136:692–708, June 1932.

- [11] H. Becquerel. Deviation du rayonnement du radium dans un champ électrique. *Comptes rendus de l'Academie des sciences*, 1900.
- [12] J. Chadwick. *Verh. Dtsch. Phys. Ges*, 1914.
- [13] W. Pauli. *Pauli Letter Collection, CERN*, 1930.
- [14] E. Fermi. Versuch einer Theorie der β -Strahlen. I. *Zeitschrift fur Physik*, 88:161–177, March 1934.
- [15] C. L. Cowan Jr. and F. Reines. The neutrino. *Nature*, 178:pp. 446–449, 1956.
- [16] C. L. Cowan Jr., F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire. Detection of the free neutrino: A confirmation. *Science*, 124(3212):pp. 103–104, 1956.
- [17] G. Danby, J-M. Gaillard, K. Goulian, L. M. Lederman, N. Mistry, M. Schwartz, and J. Steinberger. Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos. *Phys. Rev. Lett.*, 9:36–44, Jul 1962.
- [18] Tsung D Lee and Chen-Ning Yang. Parity nonconservation and a two-component theory of the neutrino. *Physical Review*, 105(5):1671, 1957.
- [19] G. Arnison et. al. (UA1 Collaboration). Experimental observation of isolated large transverse energy electrons with associated missing energy at $s=540$ gev. *Physics Letters B*, 122(1):103 – 116, 1983.
- [20] G. Arnison et. al. (UA1 Collaboration). Experimental observation of lepton pairs of invariant mass around 95 gev/c² at the cern sps collider. *Physics Letters B*, 126(5):398 – 410, 1983.
- [21] D. DeCamp et. al. (ALEPH Collaboration). Determination of the number of light neutrino species. *Physics Letters B*, 231(4):519 – 529, 1989.
- [22] K. Kodama et. al. (DONUT Collaboration). Observation of tau neutrino interactions. *Physics Letters B*, 504(3):218 – 224, 2001.
- [23] R. Davis, D. Harmer, and K. Hoffman. Search for neutrinos from the sun. *Phys. Rev. Lett.*, 20:1205–1209, May 1968.

- [24] John N. Bahcall. Solar neutrino cross sections and nuclear beta decay. *Phys. Rev.*, 135:B137–B146, Jul 1964.
- [25] Y. Suzuki. Kamiokande and super-kamiokande. *Progress in Particle and Nuclear Physics*, 40(0):427 – 441, 1998.
- [26] A. B. McDonald et. al. (SNO Collaboration). First neutrino observations from the sudbury neutrino observatory. *Nuclear Physics B - Proceedings Supplements*, 91(13):21 – 28, 2001.
- [27] B. Pontecorvo. Mesonium and antimesonium. *Zhur. Eksptl'. i Teoret. Fiz.*, 1957.
- [28] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. Remarks on the unified model of elementary particles. *Progress of Theoretical Physics*, 28(5):870–880, 1962.
- [29] Q. Ho-Kim and X.Y. Pham. *Elementary Particles and Their Interactions: Concepts and Phenomena*. Springer Berlin Heidelberg, 2010.
- [30] Stuart Charles Fuess. Neutrino-proton and anti-neutrino-proton elastic scattering. Technical report, Illinois Univ., Urbana, IL (United States), 1981.
- [31] C.H. Llewellyn Smith. Neutrino reactions at accelerator energies. *Physics Reports*, 3(5):261 – 379, 1972.
- [32] Dieter Rein and Lalit M Sehgal. Neutrino-excitation of baryon resonances and single pion production. *Annals of Physics*, 133(1):79–153, 1981.
- [33] LW Whitlow, Stephen Rock, A Bodek, S Dasu, and EM Riordan. A precise extraction of $r = \sigma_0 / \sigma_0 t$ from a global analysis of the slac deep inelastic ep and ed scattering cross sections. *Physics Letters B*, 250(1):193–198, 1990.
- [34] MAG Aivazis, Fredrick I Olness, and Wu-Ki Tung. Leptoproduction of heavy quarks. i. general formalism and kinematics of charged current and neutral current production processes. *Physical Review D*, 50(5):3085, 1994.
- [35] Lincoln Wolfenstein. Neutrino oscillations in matter. *Physical Review D*, 17(9):2369, 1978.

- [36] K. Eguchi et. al. (The KamLAND Collaboration). First results from kamland: Evidence for reactor antineutrino disappearance. *Phys. Rev. Lett.*, 90:021802, Jan 2003.
- [37] F. P. An et. al. (The Daya Bay Collaboration). Observation of electron-antineutrino disappearance at daya bay. *Phys. Rev. Lett.*, 108:171803, Apr 2012.
- [38] J. K. Ahn et. al. (The RENO Collaboration). Observation of reactor electron antineutrinos disappearance in the reno experiment. *Phys. Rev. Lett.*, 108:191802, May 2012.
- [39] P. Adamson et. al. (The MINOS Collaboration). New constraints on muon-neutrino to electron-neutrino transitions in minos. *Phys. Rev. D*, 82:051102, Sep 2010.
- [40] K Abe, J Adam, H Aihara, T Akiri, C Andreopoulos, S Aoki, A Ariga, T Ariga, S Assylbekov, D Autiero, et al. Observation of electron neutrino appearance in a muon neutrino beam. *Physical review letters*, 112(6):061802, 2014.
- [41] P Adamson, C Ader, M Andrews, N Anfimov, I Anghel, K Arms, E Arrieta-Diaz, A Aurisano, D Ayres, C Backhouse, et al. First measurement of electron neutrino appearance in nova. *arXiv preprint arXiv:1601.05022*, 2016.
- [42] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014.
- [43] Y. Ashie et. al. (The Super-Kamiokande Collaboration). Measurement of atmospheric neutrino oscillation parameters by super-kamiokande i. *Phys. Rev. D*, 71:112005, Jun 2005.
- [44] P. Adamson et. al. (The MINOS Collaboration). Observation of muon neutrino disappearance with the minos detectors in the numi neutrino beam. *Phys. Rev. Lett.*, 97:191801, Nov 2006.
- [45] P Adamson, K Anderson, M Andrews, R Andrews, I Anghel, D Augustine, A Aurisano, S Avvakumov, DS Ayres, B Baller, et al. The numi neutrino beam. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 806:279–306, 2016.

- [46] J. Cooper et. al. (The NO ν A Collaboration). No ν a technical design report.
- [47] Robert Miles Zwaska. *Accelerator systems and instrumentation for the NuMI neutrino beam*. PhD thesis, University of Texas, Austin, 2005.
- [48] S Muffson, B Baugh, C Bower, T Coan, J Cooper, L Corwin, J Karty, P Mason, A Pla-Dalmau, and M Proudfoot. Liquid scintillator production for the nova experiment. *arXiv preprint arXiv:1504.04035*, 2015.
- [49] A. Ferrari Fasso‘, J. Ranft, and P.R. Sala. Fluka: present status and future developments. *Proc. IV Int. Conf. on Calorimetry in High Energy Physics*, pages 493–502, Sep 1993.
- [50] C. Andreopoulos et al. The GENIE Neutrino Monte Carlo Generator. *Nucl. Instrum. Meth.*, A614:87–104, 2010, 0905.2517.
- [51] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, F. Behner, L. Bellagamba, J. Boudreau, L. Broglia, A. Brunengo, H. Burkhardt, S. Chauvie, J. Chuma, R. Chytracek, G. Cooperman, G. Cosmo, P. Degtyarenko, A. Dell’Acqua, G. Depaola, D. Dietrich, R. Enami, A. Feliciello, C. Ferguson, H. Fesefeldt, G. Folger, F. Foppiano, A. Forti, S. Garelli, S. Giani, R. Giannitrapani, D. Gibin, J.J. Gmez Cadenas, I. Gonzlez, G. Gracia Abril, G. Greeniaus, W. Greiner, V. Grichine, A. Grossheim, S. Guatelli, P. Gumplinger, R. Hamatsu, K. Hashimoto, H. Hasui, A. Heikkinen, A. Howard, V. Ivanchenko, A. Johnson, F.W. Jones, J. Kallenbach, N. Kanaya, M. Kawabata, Y. Kawabata, M. Kawaguti, S. Kelner, P. Kent, A. Kimura, T. Kodama, R. Kokoulin, M. Kossov, H. Kurashige, E. Lamanna, T. Lampn, V. Lara, V. Lefebure, F. Lei, M. Liendl, W. Lockman, F. Longo, S. Magni, M. Maire, E. Medernach, K. Minamimoto, P. Mora de Freitas, Y. Morita, K. Murakami, M. Nagamatu, R. Nartallo, P. Nieminen, T. Nishimura, K. Ohtsubo, M. Okamura, S. O’Neale, Y. Oohata, K. Paech, J. Perl, A. Pfeiffer, M.G. Pia, F. Ranjard, A. Rybin, S. Sadilov, E. Di Salvo, G. Santin, T. Sasaki, N. Savvas, Y. Sawada, S. Scherer, S. Sei, V. Sirotenko, D. Smith, N. Starkov, H. Stoecker, J. Sulkimo,

- M. Takahata, S. Tanaka, E. Tcherniaev, E. Safai Tehrani, M. Tropeano, P. Truscott, H. Uno, L. Urban, P. Urban, M. Verderi, A. Walkden, W. Wander, H. Weber, J.P. Wellisch, T. Wenaus, D.C. Williams, D. Wright, T. Yamada, H. Yoshida, and D. Zschiesche. Geant4a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003.
- [52] A Aurisano, C Backhouse, R Hatcher, N Mayer, J Musser, R Patterson, R Schroeter, and A Sousa. The nova simulation chain. In *Journal of Physics: Conference Series*, volume 664, page 072002. IOP Publishing, 2015.
- [53] M Campanella, PR Sala, S Vanini, and A Ferrari. First calorimeter simulation with the flugg prototype. Technical report, ATL-COM-SOFT-99-004, 1999.
- [54] Marek Gazdzicki, NA49 Collaboration, et al. Report from na49. *Journal of Physics G: Nuclear and Particle Physics*, 30(8):S701, 2004.
- [55] JM Paley, MD Messier, R Raja, U Akgun, DM Asner, G Aydin, W Baker, PD Barnes Jr, T Bergfeld, L Beverly, et al. Measurement of charged pion production yields off the numi target. *Physical Review D*, 90(3):032001, 2014.
- [56] A. Bodek and J. L. Ritchie. Further studies of fermi-motion effects in lepton scattering from nuclear targets. *Phys. Rev. D*, 24:1400–1402, Sep 1981.
- [57] K.E. Atkinson. *An introduction to numerical analysis*. Wiley, 1978.
- [58] R Ph Feynman, M Kislinger, and F Ravndal. Current matrix elements from a relativistic quark model. *Physical Review D*, 3(11):2706, 1971.
- [59] Arie Bodek and UK Yang. Higher twist, ξ_W scaling, and effective lo pdfs for lepton scattering in the few gev region. *Journal of Physics G: Nuclear and Particle Physics*, 29(8):1899, 2003.
- [60] AB Kaidalov and KA Ter-Martirosyan. Pomeron as quark-gluon strings and multiple hadron production at sps-collider energies. *Physics Letters B*, 117(3):247–251, 1982.

- [61] MV Kossov. Chiral-invariant phase space model. *The European Physical Journal A-Hadrons and Nuclei*, 14(3):265–269, 2002.
- [62] Hugo W Bertini and Miriam P Guthrie. News item results from medium-energy intranuclear-cascade calculation. *Nuclear Physics A*, 169(3):670–672, 1971.
- [63] MP Guthrie, RG Alsmiller, and HW Bertini. Calculation of the capture of negative pions in light elements and comparison with experiments pertaining to cancer radiotherapy. *Nuclear Instruments and Methods*, 66(1):29–36, 1968.
- [64] G Folger, VN Ivanchenko, and JP Wellisch. The binary cascade. *The European Physical Journal A-Hadrons and Nuclei*, 21(3):407–417, 2004.
- [65] Bo Andersson, G Gustafson, and Hong Pi. The fritiof model for very high energy hadronic collisions. *Zeitschrift für Physik C Particles and Fields*, 57(3):485–494, 1993.
- [66] J B Birks. Scintillations from organic crystals: Specific fluorescence and relative response to different radiations. *Proceedings of the Physical Society. Section A*, 64(10):874, 1951.
- [67] CN Chou. The nature of the saturation effect of fluorescent scintillators. *Physical Review*, 87(5):904, 1952.
- [68] P Adamson, C Ader, M Andrews, N Anfimov, I Anghel, K Arms, E Arrieta-Diaz, A Aurisano, D Ayres, C Backhouse, et al. First measurement of muon-neutrino disappearance in nova. *arXiv preprint arXiv:1601.05037*, 2016.
- [69] C Bassin. The current limits of resolution for surface wave tomography in north america. *Eos Trans. AGU*, 2000.
- [70] Wikimedia Commons. Illustration von clusteranalyse mit de:dbSCAN (minpts=3)., 2011.
- [71] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.

- [72] Michael David Baird. *An Analysis of Muon Neutrino Disappearance from the NuMI Beam Using an Optimal Track Fitter*. PhD thesis, Indiana University, 2015.
- [73] Evan Niner. *Observation of electron neutrino appearance in the NuMI beam with the NOvA experiment*. PhD thesis, INDIANA UNIVERSITY, 2015.
- [74] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.
- [75] Nichloas J. Raddatz. *Muon Neutrino Disappearance in NOvA*. PhD thesis, University of Minnesota, 2016.
- [76] William S Cleveland. Lowess: A program for smoothing scatterplots by robust locally weighted regression. *American Statistician*, pages 54–54, 1981.
- [77] Susan M. Lein. *Muon Neutrino Contained Disappearance in NOvA*. PhD thesis, University of Minnesota, 2015.
- [78] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.
- [79] Kanika Sachdev. *Muon Neutrino to Electron Neutrino Oscillation in NOvA*. PhD thesis, University of Minnesota, 2015.
- [80] R.D. Reed and R.J. Marks. *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*. A Bradford book. MIT Press, 1999.
- [81] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [82] Wikimedia Commons. Diagram of neuron with arrows but no labels., 2008.
- [83] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

- [84] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *arXiv preprint arXiv:1409.4842*, 2014.
- [85] Wikimedia Commons. A simplified view of an artifical neural network., 2006.
- [86] P. J. Werbos. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard University, 1974. Department of Applied Mathematics.
- [87] Yann LeCun, Koray Kavukcuoglu, and Clément Farabet. Convolutional networks and applications in vision. In *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, pages 253–256. IEEE, 2010.
- [88] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.
- [89] Wikimedia Commons. A diagram of a convolutional layer and a pooling layer applied to an image., 2014.
- [90] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *CoRR*, abs/1312.4400, 2013.
- [91] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.
- [92] Jerome H Friedman. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4):367–378, 2002.
- [93] Chris Backhouse and RB Patterson. Library event matching event classification algorithm for electron neutrino interactions in the *nova* detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 778:31–39, 2015.

- [94] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.
- [95] J Lozier. ModularExtrap Technical Note, NOvA Collaboration Internal Document 12563-v1, 2014.
- [96] C Backhouse. The CAFAna framework - technote, NOvA Collaboration Internal Document 9222-v4, 2013.
- [97] Gary J Feldman and Robert D Cousins. Unified approach to the classical statistical analysis of small signals. *Physical Review D*, 57(7):3873, 1998.
- [98] H Duyang, R Hatcher, K Maan, S Mishra, A Radovic, and R Schroeter. A Technote Describing the Derivation and Size of NuMI Flux Uncertainties Used in the First NOvA Analyses, NOvA Collaboration Internal Document 13584-v6, 2015.
- [99] VI Tretyak. Semi-empirical calculation of quenching factors for ions in scintillators. *Astroparticle Physics*, 33(1):40–53, 2010.
- [100] J Musser. Detector Mass Modeling Errors and Systematics, NOvA Collaboration Internal Document 13655-v3, 2015.
- [101] J Sepulveda-Quiroz. Bad Channels Systematic Technote, NOvA Collaboration Internal Document 13663-v3, 2015.
- [102] K Abe, J Adam, H Aihara, T Akiri, C Andreopoulos, S Aoki, A Ariga, S Asylbekov, D Autiero, M Barbi, et al. Measurements of neutrino oscillation in appearance and disappearance channels by the t2k experiment with 6.6×10^{20} protons on target. *Physical Review D*, 91(7):072010, 2015.
- [103] P Adamson, I Anghel, A Aurisano, G Barr, M Bishai, A Blake, GJ Bock, D Bogert, SV Cao, CM Castromonte, et al. Combined analysis of $\nu \mu$ disappearance and $\nu \mu \nu e$ appearance in minos using accelerator and atmospheric neutrinos. *Physical review letters*, 112(19):191801, 2014.

- [104] A Radovic, D Rocco, A Aurisano, E Niner, F Psihas, and A Himmel. A CNN Based NuE ID, NOvA Collaboration Internal Document 14437-v3, 2015.
- [105] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

Appendix A

Glossary and Acronyms

Care has been taken in this dissertation to minimize the use of jargon and acronyms, but this cannot always be achieved. This appendix defines jargon terms in a glossary, and contains a table of acronyms and their meaning.

A.1 Glossary

- **Analog to Digital Converter (ADC)** – A device which measures an analog response and produces a discretized value as output.
- **Baryon** – A particle comprised of three valence quarks. Protons and neutrons are examples of baryons.
- **Cell** – The fundamental unit of NOvA detectors. Cells are filled with liquid scintillator and use a wavelength shifting fiber to transmit light to an avalanche photodiode (APD).
- **Cell Hit** – A recorded pulse of light recorded in a NOvA cell.
- **Charge-Parity (CP)** – The combined operation of charge reversal (particles to antiparticles) and spatial inversion (reversal of coordinates).
- **Hadron** – A composite particle made of at least two quarks. *Mesons* such as pions and kaons are composed of one quark and one antiquark. *Baryons* are formed by some combination of three quarks or antiquarks. Recently, more exotic

states with four or five quarks have been observed in medium energy electron accelerators [5, 6].

- **Meson** – A particle comprised of two valence quarks. Pions and kaons are examples of baryons.
- **Monte Carlo (MC) truth** – Detailed information regarding particle interactions and detector response in MC simulation.
- **Muon Catcher** – The block furthest downstream in the NOvA ND which is used to stop (range-out) muons before they exit the detector. Active NOvA planes are interleaved with 10 cm steel planes in order to provide additional stopping power.
- **Quantum Efficiency** – The fraction of photons which produce photo-electrons in a piece of light detection hardware such as a photo-diode or photomultiplier tube.
- **Slice** – A cluster of hits which is produced by the NOvA implementation [72] of the DBSCAN [71] algorithm (Section 5.1).
- **Track** – A reconstructed particle trajectory. Algorithms which locate tracks are described in Sections 5.2.1 and 5.2.2.
- **Range-out** – A particle is said to range-out when it runs out of energy after depositing it all through ionization.

A.2 Acronyms

Table A.1: Acronyms

Acronym	Meaning
ADC	Analog-to-digital converter
APD	Avalanche photodiode
BDT	Boosted decision tree
Continued on next page	

Table A.1 – continued from previous page

Acronym	Meaning
CC	Charged current
CNN	Convolutional neural network
CP	Charge-Parity
DAQ	Data acquisition
DCM	Data concentrator module
DCS	Dual correlated sampling
DIS	Deep-inelastic scattering
FD	Far Detector
FLS	Fiber and liquid scintillator
GPU	Graphical processing unit
kNN	k-Nearest neighbor
LRN	Local response normalization
MC	Monte Carlo
NC	Neutral current
ND	Near Detector
PVC	Polyvinyl chloride
POT	Protons on-target
QE	Quasi-elastic (scattering)
RES	Resonant-elastic scattering
SGD	Stochastic gradient descent
TDC	Time-to-digital converter
WLS	Wavelength shifting