Development Towards a Model-dependent Measurement of the ν_μ Charged Current Single π^+ Pion Momentum Spectrum

Kevin Thomas Connolly

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Richard J. Wilkes, Chair

Thompson Burnett

Henry Lubatti

Program Authorized to Offer Degree:
University of Washington Department of Physics

University of Washington

Abstract

Development Towards a Model-dependent Measurement of the ν_μ Charged Current Single π^+ Pion Momentum Spectrum

Kevin Thomas Connolly

Chair of the Supervisory Committee:
Professor Richard J. Wilkes
Department of Physics

In this thesis the development and study of a method for a model-dependent measurement of the muon neutrino charged current (CC) single pion interaction's positive pion momentum spectrum is proposed. The measurement is part of the Tokai-to-Kamioka (T2K) experiment and was performed in Tokai at the near detector hall, using an event selection from a fine-grained detector (FGD) on a carbon target. The T2K neutrino beam has a peak energy of about $\sim$ 0.6 GeV, and the near detector is located 280 m downstream from the proton beam target. Data from 5.9×10^{20} protons on target was collected and used in a CC-inclusive and CC-exclusive event selection. The inclusive event selection contains $\sim 25,000$ CC muon neutrino events, and was subdivided into a single positively charged pion production channel (CC-1-pion) and a CC multi-pion interaction channel (CC-other). The official event selection, applied to both data and Monte Carlo (MC), is presented. Amongst the systematic errors involved in this event selection, the development of a novel detector systematic, created by the author of this thesis and dubbed the "hybrid analysis" is presented in detail. The analysis of published CC muon interaction data from deuterium bubble chamber experiments is presented in detail and used to tune a model of charged pion multiplicities. Utilizing this model, a systematic error on the number of charged pions produced in CC multi-pion events is characterized for the first time and implemented. The official T2K MC corresponding to the presented event selection is used extensively in the

development, study and validation of a fitting method proposed to extract the a model-dependent measurement of the CC-1-pion differential cross section with respect to true positive pion momentum. The studies to date suggest tuning of input cross section model parameters may need to be re-evaluated to ensure representative validation tests, and that the fit method may need to include a regularization method.

TABLE OF CONTENTS

Page

List of Figures iii

Chapter 1: Introduction 1

Chapter 2: A Brief Introduction to Neutrino Physics 3
2.1 The Elements of the Standard Model 3
2.2 Emergence of the Neutrino 5
2.3 Neutrino Oscillations 5
2.4 Neutrino Oscillation Theory 6
2.5 Neutrino Interactions 8

Chapter 3: The T2K Experiment 16
3.1 J-PARC Accelerator 17
3.2 The T2K Neutrino Beamline 19
3.3 Muon Monitor and Emulsion Tracker 21
3.4 Beam Operation Summary 21
3.5 Near Detectors: ND280 22
3.6 Far Detector: SK 37
3.7 Time Synchronization 37

Chapter 4: Software 39
4.1 Flux Simulation 39
4.2 Neutrino Interaction Simulation 40
4.3 Detector Simulation 42
4.4 Electronics Simulation 42
4.5 Calibration 42
4.6 Real Data 44
4.7 Reconstruction 44

Chapter 5: The Charged Current Multiple Pion Selection 45
5.1 Data Summary 45
5.2 CC Inclusive Selection 46
5.3 Selection Summaries 53
5.4 Systematics 56
5.5 Summary 70

Chapter 6: Pion Multiplicity Uncertainty 71
6.1 Motivation 72
6.2 The Fermilab Data 76
6.3 Interlude: How Many Charged Hadrons? 76
6.4 Fermilab Data: Analysis 79
6.5 Fit Results 81
6.6 The Hadron Multiplicity Model 82
6.7 Reweighting the NEUT Nominal 84
6.8 Error Estimation 85
6.9 Summary 87

Chapter 7: The Fit 90
7.1 Selection 90
7.2 Likelihood Definition 91
7.3 Analysis Method 103
7.4 Extract CC-1-pion Normalization 109
7.5 Application to NEUT Nominal 110
7.6 Validation Studies 112
7.7 Bad Fit Examples 119

Chapter 8: Summary and Conclusions 123

Appendix A: Toy Fitting 125
A.1 The Likelihood Function 125
A.2 Toy Model 129

LIST OF FIGURES

Figure Number Page

2.1 The Elements of the Standard Model[1] 4

2.2 Global fit results for oscillation parameters, taken from [17] Table 13.7. The mass values correspond to ($m_1 < m_2 < m_3$), and parenthetical values assume ($m_3 < m_1 < m_2$). The mass splitting convention used is $\Delta m^2 = m_3^2 - (m_2^2 + m_1^2)/2$. 8

2.3 The total CC ν_μ neutrino interaction cross section per nucleon, taken from [18]. 10

2.4 Comparisons between the NOMAD and MiniBooNE results for the measured parameter M_A^{QE} [21]. 11

2.5 MiniBooNE differential cross section and model predictions, from [27]. In this case a RFG nucleus model was used and a clear discrepancy between the simulation and data was found. 14

2.6 Contemporary and traditional model comparisons against the MiniBooNE differential cross section measurement. Figure taken from and additional details given in [28]. The T2K experiment uses the NEUT model, though this figure demonstrates that there is active development in modeling neutrino interactions. 15

3.1 A sketch up of the earth's cross section showing the neutrino production point at J-PARC, the near detector location, and the far detector Super-Kamiokande [30]. 16

3.2 On the left the impact of neutrino energy spectrum as a function of off-axis angle is shown. Note that the on-axis flux is a broader, higher energy beam, whereas the off-axis beam is narrowly peaked at the expected oscillation maximum [31]. On the right is the off-axis neutrino flux broken down by neutrino type with systematic errors [32]. The beam is mostly pure ν_μ 17

3.3 On the left is an overview of the primary and secondary T2K neutrino beamlines. On the right is a schematic showing the location and type of the various beamline monitors for intensity, profile, position and beam loss [30]. 20

3.4 On the left is the schematic of the secondary beamline and its primary components. The baffle is a graphite collimator in place to protect the horns. On the right is a cross sectional view of the three horns, with the target placed inside the first horn[31]. 21

3.5 The T2K beamline history of delivered protons on target (POT) and protons per pulse [31]. 22
3.6 A 3-D model of the pit. The open magnet and basket regions are shown on the first floor, B1 at 24 m deep. The horizontal modules of INGRID can be seen on the service floor, and the vertical modules of INGRID extend down to B2 at 37 m deep [30]. 23
3.7 Exploded view of the ND280 off-axis detector suite. The surrounding magnet, when closed, entirely surrounds and magnetizes the "basket" which houses the PØD, TPCs, FGDs, and ECals [30]. 24
3.8 At the left is an image of an MPPC device, showing the pixel face which is 1.3×1.3 mm^2 in a 26×26 array of 667 [30]. On the right is a charge amplitude spectrum measured with an LED source, demonstrating the single photon resolution [33]. 25
3.9 Magnetic field map (in Gauss) at $x = 0$ taken prior to ND280 detector installation with the magnet operating at 0.07 T [30]. The neutrino beam goes from negative to positive z in this coordinate system. 26
3.10 Cross sectional view of one magnet yoke (left), demonstrating the air gaps between the iron layers in which the SMRD scintillator modules (right) are instrumented [35]. 27
3.11 The PØD, showing the mounting of the electronics on the detector and the detailed elements on the right. Note that in the figure on the right the beam direction is from left to right, with the ND280 tracker downstream of the PØD[36]. 28
3.12 Simplified diagram of the primary TPC features. The cathode is in the middle of each TPC and drifted electrons are read out at the walls at $\pm x$, where z is along the beam direction and y is upward [37]. 30
3.13 On the left is a comparison of spatial resolution, and on the right a comparison of perpendicular momentum resolution. Both are show data and simulated results [37]. 31
3.14 Energy deposition as a function of momentum for negatively charged tracks (left) and positively charged tracks (right) [37]. Expectations for simulated particle type are overlaid. 32
3.15 An example event display of a neutrino interaction in FGD1, with subsequent particles entering the tracker [39]. 33
3.16 A CCD image of a typical scintillator bar. The hole center and TiO_2 thickness is determined by automated image analysis [39]. 34
3.17 The INGRID modular array is shown at left, and an individual module is shown at right. In the right-hand figure, the blue layers represent tracking segments, and the black planes are veto layers [30]. 36

3.18 Schematic of Super-Kamiokande, the far detector in the T2K experiment [30]. 38

4.1 Flowchart of the various stages and their dependencies within the ND280 software suite that is responsible for MC generation and data analysis [30]. . 40

4.2 An example demonstrating the effect of MIP calibration of scintillator bars in the PØD. Using cosmic ray events groups of scintillator bars can be calibrated to ensure a uniform MIP response across the detector. The before image (left) shows a varied response across the detector, whereas the post-calibration response (right) is mostly uniform. 43

5.1 The target and fiducial volume used in FGD1 used in this event selection [32]. The asymmetry in the y coordinate is due to the detector coordinates being centered relative to the magnet, which the FGD is 55 mm above. The beam comes from the left in this figure, with TPC1 being at the left as a veto and TPC2 being flush at the right and used for additional tracking. 48

5.2 An cartoon illustration of a broken track event [58]. An background event enters FGD1, and "breaks" in reconstruction. The second half of the track is taken as a CC muon candidate starting in the fiducial volume of FGD1, leaving behind the first segment as a track entirely contained in FGD1. . . . 49

5.3 CC exclusive channel topologies broken down by selection criterion [58]. . . . 51

5.4 The same CC inclusive selection shown in two different ways: true event topology on the left, and true interaction mode on the right. Due to final state interactions and detector efficiencies and resolution, different combinations of topologies and modes will enter observed sub-selections [58]. 53

5.5 A pictographic summary of the systematic studies associated with this selection [58]. 61

5.6 The cartoon sketch up on the left shows an inserted set of hits forming a short "stub," which have some opening angle θ with respect to the original event's CC μ^- candidate (highest momentum negative track, HMNT). On the right, the parameter α is the fraction of inserted hits that were absorbed into the muon track's hits for failed reconstructions. For these events you can see a trend for low-energy proton tracks failing, in addition to an increasing "theft" of hits as the angle between the tracks decreases. 63

5.7 Proton-Hybridized Efficiency Overlay and Ratio: On the left the efficiency verse momentum for the proton-hybridized sample is shown for data and MC. On the right is the data/MC ratio. 64

5.8 Pion-Hybridized Efficiency Overlay and Ratio: On the left the efficiency verse momentum for the pion-hybridized sample is shown for data and MC. On the right is the data/MC ratio. 64

6.1 Plots created from published data on various charged hadron multiplicity measurements on different targets. The data points presented here come from the CHORUS experiment, which also published the table of aggregated data upon which these plots are based [64]. Neutrino interactions are on the left, and anti-neutrino interactions on the right, with data fit to the empirical relation $\langle n_{ch} \rangle = a + b \log W^2$. 71

6.2 Data and theory comparison of MiniBooNE results[67]. 73

6.3 Different modes of CC 1-pion event identification. 74

6.4 CC-1-Pion, FGD1-TPC1 matched sample, charged pion candidate momentum categorized by true particle type. 74

6.5 CC-1-Pion, FGD1-TPC1 matched sample, charged pion candidate momentum categorized by true event topology. 75

6.6 CC-1-Pion, FGD1-TPC1 matched sample, charged pion candidate momentum categorized by true NEUT interaction mode. 76

6.7 Examining just the CC-other *background* from Figure 6.5, it can be seen that the DIS interaction mode dominates. 77

6.8 MINOS [68] experiment's CC inclusive cross section results, from which the error on the CC-other interactions is derived. The error around 4 GeV is 10%. 78

6.9 Raw data from Fermilab [65], ν-neutron interactions, $\nu + n \rightarrow \mu^- + n + X^+$, with absolute scale on the left and column-normalized on the right. 78

6.10 Raw data from Fermilab [65], ν-proton interactions, $\nu + p \rightarrow \mu^- + n + X^{++}$, with absolute scale on the left and column-normalized on the right. 79

6.11 Analysis of the Fermilab data for ν-proton interactions, producing mean charged hadron multiplicity as a function of $\log W^2$. The red dashed lines are the published fit values, and the black overlay is the re-analysis. It was discovered that the published fit values for the ν-p interactions was incorrect. The correct fit values are shown in the plot at the right. 80

6.12 Analysis of the Fermilab data for ν-neutron interactions, producing mean charged hadron multiplicity as a function of $\log W^2$. The red dashed lines are the published fit values, and the black overlay is the re-analysis. 81

6.13 KNO analysis of the ν-proton and ν-proton interactions in regular and log scales. 82

6.14 Analysis of the dispersion of the mean charged multiplicity from the Fermilab data, with a linear fit. The x-intercept value is taken as the parameter α used in the adjusted KNO analysis. 83

6.15 An adjusted KNO analysis of the ν-proton and ν-proton interactions in regular and log scales, where the possible effects of the target not being an ideal nucleon are accounted for in the parameter α. 84

6.16 Result of the toy charged hadron multiplicity model as compared to the Fermilab data. The model produces charged pion mean multiplicities, as well as total charged hadron multiplicities. 85

6.17 Example generation of the weight histograms for each nucleon target. The top row is the NEUT nominal output, the second row is the toy model's tuned output, and the third row is the TOY/NEUT ratio. It should be noted this is only an example set of weights from an earlier stage of development. 86

6.18 Result of the application of the charged hadron multiplicity tuning to the NEUT sample. The NEUT nominal distribution is shown against the post-tuning NEUT histogram, overlaid with the Fermilab bubble chamber data and linear fits. 87

6.19 The nominal tracker CC-1-pion (top row) and CC-other (bottom row) samples, based on NEUT MC, before and after tuning. The nominal distributions on the left demonstrate the harsh transitions between NEUT's hadron multiplicity model and the out-sourced Pythia/JETSET-based model. The reweighted samples alleve this transition somewhat. 88

6.20 Projection of statistical throws of the 4-D parameter space for the hadron multiplicity model into the 2-D parameter space of the linear parameters A and B. The black bands are due to the Fermilab statistical error, and the gray bands are due to an additional systematic derived from differences with the BEBC experimental results. 89

6.21 Error band due to the charged pion multiplicity uncertainty. The statistical errors are from the Fermilab published data, and the systematic error is attributed to the difference between the BEBC and Fermilab results. 89

7.1 Detector fractional error covariance, bins 1-9 correspond to pion momentum bins in the CC-1-pion sample, bins 10-18 correspond to pion momentum bins in the CC-other sample. At the right is the fractional error in each sample. . 94

7.2 Fractional error covariance matrix (left) for flux parameters. The binning is in parent neutrino energy and flavor, as described in the text and delineated by the black bars. On the right is the flux error on the primary neutrino component, ν_μ. 96

7.3 The pion multiplicity parameters input covariance matrix. The parameters correspond to true pion momentum normalization in the true CC-1-pion and CC-other topologies. These fractional errors may seem large, but the sub-samples are quite small. For example, within the CC-1-pion selection, the true topology of CC-1-pion events with are due to multi-pion or DIS interactions is very small. The effect of these parameters on the selected *samples* can be seen to be quite small, as shown in Figure 7.4. 101

7.4 Fractional error per parameter category. Note the y-scale for the cross section parameters is larger, and that the cross section parameters are the dominant source of prior error. 102

7.5 The nominal distribution for the CC-1-pion and CC-other samples. 104

7.6 An example throw. In this case, all parameters were thrown. The original nominal distribution is shown, and the variation away from is also shown, along with the signal in this throw. 107

7.7 An example fit result to a toy data set. The toy data set is shown as the "throw," and the stacked histogram is the best fit result. 107

7.8 Post-fit results on the left, with the nominal CC-1-pion component, the throw, the post-fit result, and result of fitting a gaussian to 500 cross section throws (all parameters) in each bin. The result of one nuisance throw and one cross section throw on the right. The interpretation of the results in the right-hand plot are as follows: Examining the first bin, we see the nominal signal (black-dashed) is decreased by this throw (green). The effect of all the post-fit parameters that are not cross-section related are shown in red, these are the "nuisance" throws and in this case, in the first bin, the nuisance parameters pull the nominal down to just below the thrown signal. However, when the cross section parameters are included the prediction in the first bin goes down significantly (blue). The change in the prediction from the nuisance throws to the cross section throws represents the effect of *only* the cross section parameters in a given bin. After performing 500 throws similar to this example, the effect of the cross section parameters on each bin is extracted by taking the ratio of the cross section to the nuisance throw bin contents. 109

7.9 Distributions of the CC-1-pion normalization parameters X_j which are extracted from forming the ratios of the nuisance and cross section throws.. . . 110

7.10 NEUT interaction generator's nominal output for the observed CC-1-pion cross section, as a function of the pion momentum and parent neutrino energy. This was generated from a sample of 500,000 ν_μ interactions using the T2K flux. The binning on the left is the neutrino binning used in the flux parameters, and the pion momentum binning used in the pion parameters. The output on the right is in a fine binning for an alternative view of the phase space. 112

7.11 The nominal NEUT output for ν_μ CC-1-pion interactions on a carbon target, in the pion momentum binning used in this analysis (overflow is included in the last bin) and in a fine binning (20 MeV) that is drawn smoothly. 113

7.12 The NEUT nominal output overlaid with the final cross section measurement. In this example a fake data set was used. 114

7.13 Pion parameter pull distributions for toy experiments with all fit parameters thrown. Each parameter corresponds to one bin in true pion momentum space. The pull definition is given in Equation 7.12. For a good fit the distributions should be centered at zero with a gaussian width of 1.0. 115

7.14 Pion parameter pull distributions for toy experiments *without* the cross section fit parameters thrown. Each parameter corresponds to one bin in true pion momentum space. The pull definition is given in Equation 7.12. For a good fit the distributions should be centered at zero with a gaussian width of 1.0. 116

7.15 Fitted $\Delta\chi^2$ distribution for 100 toy experiments with all fit parameters included. 117

7.16 Fitted $\Delta\chi^2$ distribution for 100 toy experiments *without* the cross section fit parameters included. 118

7.17 Post-fit marginalization pulls for each bin, calculated by Equation 7.15, for 100 toy experiments using all fit parameters. 119

7.18 Post-fit marginalization pulls for each bin, calculated by Equation 7.15, for 100 toy experiments generated without cross section parameters thrown and fit with the cross section parameters fixed. 120

7.19 An example of a throw that breaks the fitter's free pion parameter. In this case, the pion parameter corresponding to the last true pion momentum bin was fit to zero. 121

7.20 An example of a throw that breaks the fitter's free pion parameter by inducing extreme anti-correlations. This is most likely due to CC-other sample (bottom row) having pairs of bins that sequentially have large deviations from the nominal followed by nearly matching the nominal. 122

A.1 On the left the the measured toy data and simulated data, both in the reconstructed variable space (after detector smearing). On the right is the distribution which nature follows, a Poisson function with mean $\mu = 4.0$, which for simplicity is also taken to be the predicted model, in the true variable space. 131

A.2 The transfer matrix for a toy model with reconstructed space variable $y = x + \delta$, where x is the true variable and δ is a gaussian smearing. A projection along the vertical axis gives the true distribution, whereas a projection along the horizontal axis gives the observed, "reconstruction" distribution. 132

A.3 Fit results for toy with model and nature equal to a Poisson function with mean $\mu = 4.0$. Fit result parameters $\mathbf{C} = \{1.03, 0.97, 0.96, 1.05, 0.95, 1.02, 0.92, 1.07, 1.62, 0.43\}$. The fit covariance shows no abnormal parameter correlations. 133

A.4 The final fit result shown in dark teal data points, with each data point representing the model prediction scaled by the fit parameter result for the associated bin. 133

A.5 On the left is a comparison of thrown data, a nominal model prediction, and fit results for gaussian smearing equal to **0.2 times the bin width**. In the middle is the resultant transfer matrix, demonstrating minimal smearing. On the right is the fit parameter covariance matrix. For minimal smearing the parameters are well-behaved. 134

A.6 On the left is a comparison of thrown data, a nominal model prediction, and fit results for gaussian smearing equal to **0.4 times the bin width**. In the middle is the resultant transfer matrix, demonstrating minimal smearing. On the right is the fit parameter covariance matrix. For mid-scale smearing the fit parameters demonstrate anti-correlation pairs. 135

A.7 Example post-fit covariances for a toy fit where the smearing was made very large, purposefully inducing over-fitting. The standard fit's post-fit covariance is at the left, showing wild oscillatory correlations, and the Tikhonov regularized covariance is shown at right, where the oscillations have been quelled. 136

ACKNOWLEDGMENTS

If I think about why I went into physics to begin with, it was, to a certain degree, a selfish decision, a shameless pursuit of my own curiosity. However, as I came to learn, in collaborative science the pursuit of knowledge is a social process. Through my own pursuit I made close friendships and shared a lot of wonderful experiences. I wish I could express all of my thanks, but the reality is that I can't possibly capture the entirety of my appreciation for the many individuals along the way, so this process of expressing gratitude is, in someway, a simultaneous apology for its insufficient nature. With that said, there are a few people I would like to mention and I trust that those not mentioned here know they are in my thoughts.

I am, naturally, indebted to my family: my mother, father, brother, grandparents. You let me pursue this project tirelessly and never expressed anything but support and encouragement.

The Holt Family: Brian, Bre, Harper and Julia, you formed my "second family" in Seattle, were always welcoming, always supportive, always caring and always there to make sure that I was doing well. Graduate school was a long process and brought with it the vicissitudes of life – some of those times were very hard, but I always knew that I was welcome in your home. Thank you for everything over the years.

Damon Hanson: What good fortune of being housed together at random! By the time we had talked over mutual interests in the arts and sciences, it was obvious we'd be good friends. Thanks for all the great times at work in Tokai and abroad traveling Japan.

Richard Calland: There's nothing quite like living in a derelict building with nobody else around except one other person to really build the bonds of a friendship.

It was great working with you over the years, it was great cycling through the infinity of small roads between the rice fields and forests of Tokai, and it was great drinking coffee at midnight.

Roger Wendell: It's hard to express my gratitude for all the support and encouragement you've given over the years – you've been a great friend and mentor, always there to listen and provide sound advice when needed. We never had the chance to share our research endeavors directly, so I know all of your support came out of your own time, which makes it all the more appreciated.

Mike Dziomba: We started out in physics together, but by far the most enjoyable part of our friendship was our shared introduction to Japan and the language. Taking up the study of Japanese was a life-changing endeavor and I'm glad I got to go through that with you.

Dave Forbush: Who knows how many days we spent in the pit working on the commissioning and maintenance of the PØD water system? I don't know what I would have resorted to if you weren't there – your endless humor and wit made those days not only enjoyable but memorable.

William Terrano: When I mentioned I was writing acknowledgements you threw both of your hands up and yelled "me me me! I want in!" So, here you are. You've been a great friend over the years and are, in many ways, are a person beyond description. We've had a lot of interesting times together and you've enriched my life. I wish I had had more time in Seattle, but hopefully we'll cross paths in the near future.

Mark Hartz: You were my closest collaborator over the last year of my work efforts, and listened to and answered question after question. I could not have come this far without your support and I am grateful for having had the opportunity to work in your presence and glean what I could of your vast experience and knowledge. More so than anything, your patience in dealing with me is much appreciated.

Finally, I would like to thank my advisor, Jeff Wilkes, for supporting my being in

Japan over the years. You brought me into your research group and let me find my own way without any pressure. This allowed me to explore and work with a variety of groups on a variety of projects and gain an immeasurable amount of knowledge on not just the topics of research but the process, the dynamics, the *entirety* of research. Thank you.

DEDICATION

To everyone that put the hours in to make this possible, be it family, friends or collaborators.

Chapter 1

INTRODUCTION

This thesis will endeavor to present the the context and methods related to the measurement of the muon neutrino charged current single pion production (ν_μ CC-1-pion) by the Tokai-to-Kamioka (T2K) experiment.

Chapter 2 will briefly introduce the topic of particle physics and focus on the discovery and properties of neutrinos. The property known as neutrino oscillations will be presented and an introduction to neutrino interactions will be given. Specifically, the importance of cross section measurements within the realm of oscillation experiments will be presented in this chapter. The most pertinent contemporary experimental results for CC-1-pion cross section measurements, which provide substantial motivation for the work on the development of this measurement within T2K, will be presented in Section 2.5.1.

The summary of the T2K experiment is broken into two chapters. Chapter 3 will introduce the facilities, hardware and design of the T2K experiment. The beamline and far detector site, Super-Kamiokande, will be covered, but particular attention will be given to the near detector site as it is the primary location of the data selection and analysis within this thesis. Chapter 4 will summarize the software side of the experiment, again with particular attention given to the operations and management of data at the near detector site.

Chapter 5 will introduce T2K's near detector data selection, starting with the ν_μ CC inclusive event selection. The CC inclusive sample will be broken down into two sub samples used in the Cc-1-pion cross section measurement. A summary of the systematics associated with the event selection will also be also presented.

Chapter 6 will cover the analysis of external data sets relevant to characterizing the uncertainty on the number of charged pions produced in neutrino interactions. This chapter will begin with motivation for the development of this uncertainty. A summary of data from

earlier experiments will be presented, and the use of these prior results in the formulation and application of a pion multiplicity model will also be given.

Chapter 7 will give a core description on the elements involved in the development of the CC-1-pion measurement. A description of the fit method and implementation will be given in this chapter. The fit parameters used, their sources and their interpretation will be discussed and on-going validation methods will be summarized. A discussion of the current state and future prospects will also be given. Reference to the fitting studies and methods presented in Appendix A, which may have utility in future studies, will be given in this chapter.

Chapter 8 will provide a brief summary of the contents of this thesis.

Chapter 2

A BRIEF INTRODUCTION TO NEUTRINO PHYSICS

Many of the great discoveries throughout the history of science can be told through the process of encountering the unexpected. In many cases these historic discoveries were made by mistake. Particle physics is no exception, and is a field of study which is perpetually up against the frontiers of knowledge, trying to comprehend what elements of our understanding are accurate and what elements are lacking. The research leading up to the discovery of neutrinos, from their origins and through our contemporary understanding of neutrino properties, has been full of the unanticipated. In addition to the rich history of neutrino physics, the contemporary experiments and discoveries have also proven exceptional. Neutrinos are now established to have properties beyond description of the Standard Model of particle physics, with a substantial international research interest pushing to extend our understanding of their properties remaining to be understood.

In the following sections I will give a brief introduction to elementary particle physics, with the goal of providing some historical context, focusing on the efforts that have lead to the contemporary understanding of neutrino physics.

2.1 The Elements of the Standard Model

The reductionist approach to understanding nature that embodies the field of particle physics concerns itself with finding a minimal set of knowledge from which the observable properties of nature can be derived. A mathematical model, in general, has constants, parameters, and the formulae that dictate their relationships. The efforts of modern physics to date have provided what is referred to as the Standard Model (SM), which is a description of the forces that govern particle interactions. At present the standard model incorporates three forces of nature: The electrodynamic, the weak, and the strong force. The most familiar force, namely the gravitational force, is not treated in the standard model, and the

electrodynamic and the weak forces were unified in the latter half of the 20th century.

The elementary particles of the Standard Model are depicted in Figure 2.1. In this figure, the quarks and leptons are organized in "generations," going from left to right with the trend being an increase in mass. In the case of the quarks each generation contains a pair of charged particles, and in the case of leptons each generation has one charged particle and an associated neutral particle. The neutral leptons are called neutrinos. The gauge bosons are the so-called "force carriers" of the various fundamental forces, which are responsible for mediating interactions between particles. The photon is responsible for mediating the electromagnetic force, the W and Z boson mediate the weak force, and the gluon mediates the strong force.

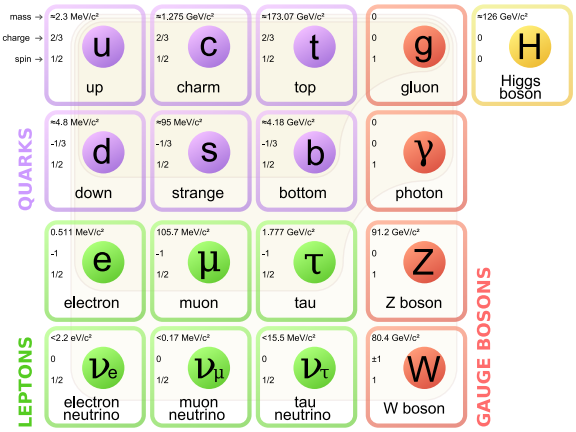

Figure 2.1: The Elements of the Standard Model[1]

The most recent addition to this table was made in 2012, with the discovery of the Higgs boson by both the ATLAS and CMS experiments [2, 3], thus completing the set of hypothesized particles that make up the basic building blocks of nature proposed by the Standard Model.

As the focus of this thesis will be with regard to neutrinos and their properties, it would be apt to take a step back and summarize the historical context which necessitated the existence of such particles.

2.2 Emergence of the Neutrino

The 19th century came to a close with J. J. Thompson discovering the electron in 1897. Not long after that the nucleus was discovered in Rutherford's famous gold-foil experiment. Bohr's model of the nucleus as a positively charged proton paired to a negatively charged electron was proposed in 1914, but had difficulty explaining the mass of heavier atoms. In 1932 Chadwick discovered the neutron, providing a solution to the mass discrepancy. As for addressing the question of what all matter was composed of, one author put it as [4]: "In 1932 it was just protons, neutrons and electrons."

However, there were some earlier indications that something was amiss. It was found that in measurements of nuclear beta decay, the emitted electron was observed to have a spectrum of energies. The decay process, then viewed as a two-body problem, required the electron be emitted with a kinematically predetermined and fixed energy. In 1930, to remedy the discrepancy between observation and expectation Pauli suggested the possible existence of an additional neutral particle taking place in the interaction.

At the time of his proposal, Pauli had apologized for the introduction of a neutral particle, imagining that it would be impossible to detect. His expectations of the difficulty in detecting such a particle were not wrong, and it was not until 1956, almost 30 years later, that Reines and Cowan observed the reactor neutrino interactions $\bar{\nu}_e + p \rightarrow e^+ + n$ [5]. The measurement by Reines and Cowan opened the door to neutrino physics measurements, but it was not long thereafter that neutrinos were found to be not quite as expected.

2.3 Neutrino Oscillations

In 1968 Ray Davis made a measurement of solar ν_e neutrinos at the Homestake Solar Neutrino Experiment [6]. Davis's measurement inferred a rate of solar neutrino interactions that, when compared to the Standard Solar Model, came out significantly below expectations. This deficit was referred to as the Solar Neutrino Problem. The solar neutrino deficit was observed by the SAGE [7], GALEX [8] and Kamiokande-II [9] experiments.

Theory for neutrino oscillations had been proposed by Pontecorvo [10], which explained the transition of one neutrino type into another and thus explained the "missing" flux

observed in experiments. A combination of efforts from neutrino theorists, namely the extension of Pontecorvo's theory by Maki, Nakagawa, and Sakata [11], combined with the theory of matter effects on neutrinos proposed by Mikheyev, Smirnov and Wolfenstein (the so-called MSW effect), could account for the observed discrepancies.

In 1998 the Super-Kamiokande experiment [12] observed a deficit of muon neutrinos expected to originate from cosmic ray interactions in the atmosphere. As these neutrinos were expected to originate from pion decays in the atmosphere, a path length derived from the reconstructed direction of the neutrino could be attributed to each interaction and the observed rate of muon neutrinos could be measured as a function of distance traveled (via the zenith angle). The 1998 measurements firmly establishing the existence of neutrino oscillations and paved the way for contemporary neutrino oscillation experiments.

2.4 Neutrino Oscillation Theory

The derivations presented in this section are based on the work of Giunti and Kim [13]. Neutrino oscillation theory can be derived from a single foundational concept: neutrinos can be described as generations of particles that exist in two distinct bases, the flavor basis or the mass basis, with the following transformation between the eigenstates:

$$|\nu_\alpha\rangle = \sum_i U^*_{\alpha k} \, |\nu_i\rangle \, , \tag{2.1}$$

where the greek index is over the flavor basis states, $\alpha = \{e, \mu, \tau\}$, and the roman index is over the mass basis states, $k = \{1, 2, 3\}$, and the matrix U is the PMNS mixing matrix [11].

The time evolution of a neutrino takes place in the mass eigenstate, where the momentum of the particle is well-defined with energy eigenvalues

$$E_k = \sqrt{\mathbf{p}^2 + m_k^2}. \tag{2.2}$$

For a neutrino created in a flavor state α at some time $t = 0$, the time evolution of the state is given by,

$$|\nu_\alpha(t)\rangle = \sum_k U^*_{\alpha k} e^{-iE_k t} \, |\nu_k\rangle \, . \tag{2.3}$$

The likelihood of finding a neutrino that started in a flavor state ν_α in a different flavor state ν_β at some time t later is obtained from the amplitude,

$$A_{\alpha\to\beta}(t) = \langle\nu_\beta|\nu_\alpha(t)\rangle = \sum_k U^*_{\alpha k} U_{\beta k} e^{-iE_k t}. \tag{2.4}$$

The square of the amplitude gives the transition probability:

$$P_{\alpha\to\beta}(t) = |A_{\alpha\to\beta}|^2 = \sum_{k,j} U^*_{\alpha k} U_{\beta k} U_{\alpha j} U^*_{\beta j} e^{-i(E_k - E_j)t}. \tag{2.5}$$

In the case of ultra-relativistic neutrinos the energy eigenvalue in Equation 2.2 can be approximated by

$$E_k \approx E + \frac{m_k^2}{2E}, \tag{2.6}$$

where the difference $(E_k - E_j)$ in the exponential of the transition probability then becomes

$$E_k - E_j \approx \frac{\Delta m^2_{kj}}{2E}. \tag{2.7}$$

The squared mass difference is defined as

$$\Delta m^2_{kj} = m_k^2 - m_j^2, \tag{2.8}$$

and the energy $E = |\mathbf{p}|$ is the energy without any contribution from the neutrino mass. With the final substitution that the time of travel t can be approximated by the distance traversed, L, the transition probability can be written as

$$P_{\alpha\to\beta}(L,E) = |A_{\alpha\to\beta}|^2 = \sum_{k,j} U^*_{\alpha k} U_{\beta k} U_{\alpha j} U^*_{\beta j} e^{-i\frac{\Delta m^2_{kj} L}{2E}}. \tag{2.9}$$

The values of L and E are typically determined by experimental design, and the values of U and the mass splittings are the constants of nature which are to be measured. The PMNS matrix can be parameterized in the following form,

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{-i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^{-i\alpha_1} & 0 & 0 \\ 0 & e^{-i\alpha_2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{2.10}$$

where $c_{ij} \equiv \cos\theta_{ij}$, $s_{ij} \equiv \sin\theta_{ij}$, with θ_{ij} known as the "mixing angles," δ the CP-violating phase, and the α_i the Majorana phases. It has been shown that the Majorana phases have no effect on neutrino oscillations [14]. The full derivations are beyond the scope of this thesis, but the cited survival [15] and $\nu_\mu \to \nu_e$ transition probabilities [16] are

$$P(\nu_\mu \to \nu_\mu) \approx 1 - 4\cos^2(\theta_{13})\sin^2(\theta_{23}) \times [1 - \cos^2(\theta_{13}) \times \sin^2(\theta_{23})]\sin^2\left(1.267\frac{\Delta m^2_{32}L}{E_\nu}\right) \quad (2.11)$$

$$P(\nu_\mu \to \nu_e) \approx \sin^2(\theta_{23})\sin^2(2\theta_{13})\sin^2\left(\frac{\Delta m^2_{32}L}{4E_\nu}\right). \quad (2.12)$$

Efforts from a variety of experiments that are on-going and in design phases provide footholds on measuring the different neutrino physics parameters. The various oscillation experiments and their primary sensitivities are divided into four general categories: Solar neutrino experiments (θ_{12} and Δm^2_{21}), atmospheric neutrino (θ_{23} and Δm^2_{31}), reactor (θ_{12} and θ_{13}) and long baseline accelerator experiments (θ_{13}, θ_{23} and Δm^2_{32}). A summary of the global fit results for currently measured oscillation parameters is shown in Figure 2.2.

Parameter	best-fit ($\pm 1\sigma$)	3σ
Δm^2_{21} [10^{-5} eV ^{2}]	$7.54^{+0.26}_{-0.22}$	$6.99 - 8.18$
$\lvert\Delta m^2\rvert$ [10^{-3} eV ^{2}]	$2.43^{+0.06}_{-0.10}$ ($2.42^{+0.07}_{-0.11}$)	$2.19(2.17) - 2.62(2.61)$
$\sin^2\theta_{12}$	$0.307^{+0.018}_{-0.016}$	$0.259 - 0.359$
$\sin^2\theta_{23}$	$0.386^{+0.024}_{-0.021}$ ($0.392^{+0.039}_{-0.022}$)	$0.331(0.335) - 0.637(0.663)$
$\sin^2\theta_{13}$ [173]	0.0241 ± 0.0025 ($0.0244^{+0.0023}_{-0.0025}$)	$0.0169(0.0171) - 0.0313(0.0315)$

Figure 2.2: Global fit results for oscillation parameters, taken from [17] Table 13.7. The mass values correspond to ($m_1 < m_2 < m_3$), and parenthetical values assume ($m_3 < m_1 < m_2$). The mass splitting convention used is $\Delta m^2 = m^2_3 - (m^2_2 + m^2_1)/2$.

2.5 Neutrino Interactions

With the rapid progress of contemporary, international neutrino physics programs the world of neutrino physics has now entered a precision era. The mixing angles have been measured,

and although there still remain other significant questions to address, such as determination of the mass hierarchy and measurement of the CP-violating phase δ, a significant effort must be put forward to increase the precision of the measured quantities.

Oscillation experiments are typically designed in a way such that knowledge of the neutrino source, and hence flux, is constrained and well-known, and the observed measurements are the product of the flux times the neutrino interaction cross section. Primary sources of uncertainty in results are thus attributable to flux uncertainty and interaction cross section uncertainty. With the precision era ahead of neutrino physics, increasing interest in refining the existing cross section models has become a new priority.

Neutrino interactions are governed by the weak force and occur with an exchange of either a W boson or a Z boson. In the case of exchanging a W, the interaction is referred to as a Charged Current (CC) interaction, and in the case of exchanging a Z boson the name Neutral Current (NC) is used. As neutrinos themselves are neutral particles, their existence and properties can only be inferred by the observed byproducts of their interactions. The CC interactions produce a charged lepton associated with the flavor of the interacting neutrino, which makes determination of the neutrino flavor possible. In contrast, the outgoing state of the NC interaction is a neutrino thus determination of the interacting neutrino type is not possible.

The NC and CC interactions can be classified into general categories based on the interaction features. At neutrino energies below $\sim$ 1 GeV, the interactions are usually referred to as quasi-elastic (QE). In CC-QE interactions the neutrino scatters off a target nucleon and produces an outgoing lepton with an ejected nucleon. As the parent neutrino energy increases, the target nucleon can be excited into a baryon resonance which most often decays into a final state with a nucleon and pion. These interactions are typically labeled CC-RES. Finally, in the energy range around and above 5 GeV, deep inelastic scattering (DIS) processes are possible. In DIS interactions the neutrino interacts with a quark in the target nucleon producing a hadronic shower. Figure 2.3 shows a the CC cross section as a function of neutrino energy, highlighting the QE, RES and DIS regions.

A feature of Figure 2.3 worth noting is the significant increase in error bars for decreasing neutrino energy. The cross section at high energy is well-measured, but at neutrino energies

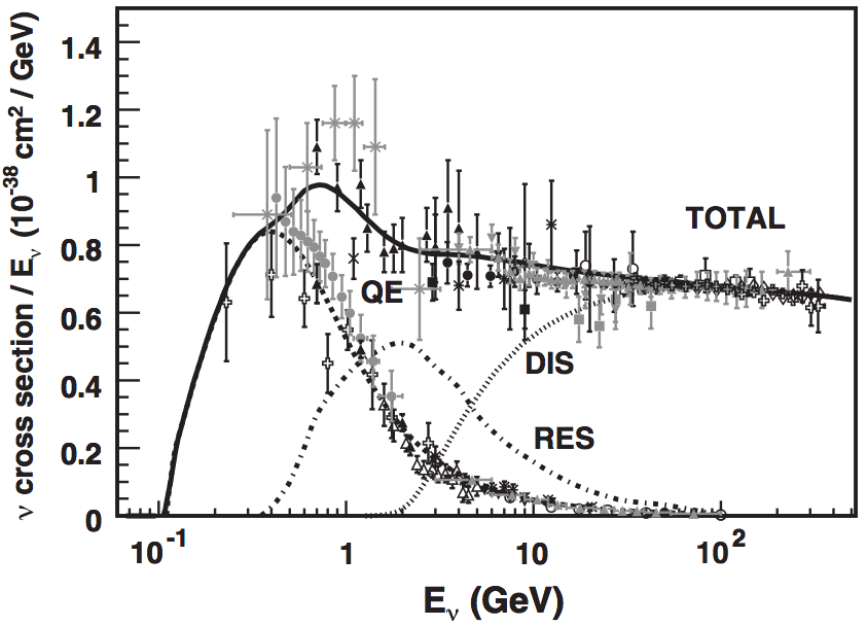

Figure 2.3: The total CC ν_μ neutrino interaction cross section per nucleon, taken from [18].

associated with most oscillation experiment's mean neutrino energy (T2K, for example, is around 0.7 GeV) the uncertainty becomes significant. In the case of T2K, the near-detector measurements of the neutrino interactions are used to cross-check the flux prior to oscillation and to constrain the predicted flux at the far detector, so a precise cross section model is of great importance.

Additional difficulties in cross section measurements are introduced by the fact that many detector targets are not free nucleons. In some cases liquid hydrogen or deuterium targets were used, but more common amongst modern experiments are scintillator targets, primarily composed of carbon, and oxygen targets from water Cherenkov detectors, with additional materials such as lead or iron typically being part of detector designs. This necessarily introduces complicated nuclear re-interactions, known as Final State Interactions (FSI).

Even amongst these difficulties, the CC-QE interaction ($\nu_\mu + n \rightarrow \mu^- + p$) is of particular utility given that the parent neutrino energy can be reconstructed from the outgoing lepton

kinematics and the incident neutrino direction. Additionally, in the lower energy ranges the CC interactions are dominated by this channel. The CC-QE interaction has been studied since the 1970s, with original interest in testing the vector-axial vector nature of the weak interactions and later interest being in measuring the axial vector form factor of the nucleon. The details of the models and development are reviewed in [18], but the pertinent conclusion is that leading model parameterizes the axial vector form factor as

$$F_A(Q^2) = \frac{F_A(0)}{\left(1 + Q^2/\left(M_A^{QE}\right)^2\right)^2}. \tag{2.13}$$

In this equation the parameter M_A^{QE} is the "axial mass," and Q is the momentum transfer from the neutrino to the target nucleon. The value for $Q = 0$, $F_A(0)$, is constrained from nuclear beta decay measurements. This left neutrino scattering experiments in a position to measure M_A^{QE}. The earlier experiments had contributed to a world average of $M_A^{QE} = 1.026 \pm 0.021$ GeV [19], yet in more recent results from NOMAD [20] and MiniBooNE [21] had noteworthy discrepancies. The NOMAD experiment sampled neutrinos in the range of $3 - 100$ GeV and measured $M_A^{QE} = 1.05 \pm 0.02(\text{stat}) \pm 0.06(\text{syst})$ GeV, whereas the MiniBooNE experiment had a mean neutrino energy of 800 MeV and found $M_A^{QE} = 1.35 \pm 0.17$ GeV. The results of these measurements are shown in Figure 2.4.

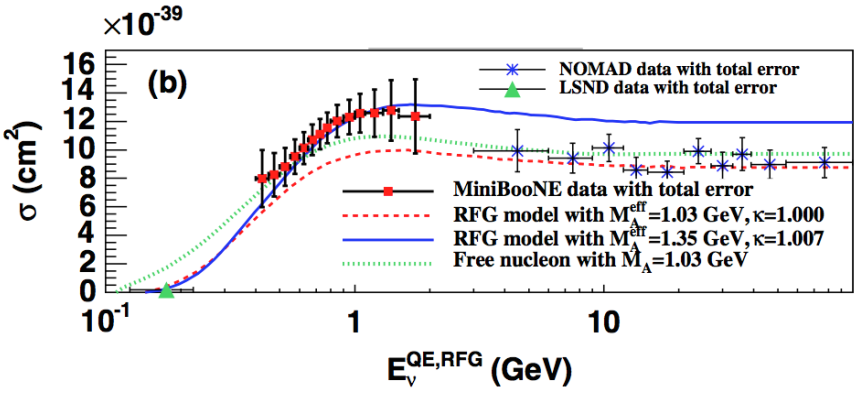

Figure 2.4: Comparisons between the NOMAD and MiniBooNE results for the measured parameter M_A^{QE} [21].

More recently, the MINERνA experiment has produced significant results analyzing both CC-QE interactions on a CH target [22] and cross section ratios on different nuclear targets [23]. Both results have deep implications. The CC-QE paper focuses on the fact that for contemporary experiments using nuclear targets, a model of the nucleus is necessary in addition to a model of the nucleon. To date, the typical simulation models the nucleus as a relativistic Fermi gas (RFG) [24], in which the nucleons are quasi-free and bound in a uniform potential. This MINERνA result found disagreement between the CC-QE differential cross section and the RFG model, but also observed a distribution of vertex activity (energy deposited by ejected nucleons around the interaction vertex). The vertex activity measurement is significant in that it agrees with expectations for an RFG model augmented by accounting for correlations between the nucleons within the target nucleus.

The MINERνA results for cross section ratio measurements are also significant in that they are the first instance of high-statistics neutrino interactions on different nuclear targets in the same neutrino beam, demonstrating clearly the effect of nucleus. As the authors of the publication suggest, the results of this measurement are profound in their implication: our modeling of the nucleus is incomplete, yet for many oscillation experiments there is often dependence in the form of extrapolating near detector data on one target material (e.g. Cabon) to a far detector prediction on a different target (e.g. Oxygen). The MINERνA results thus necessitate a new class of systematic uncertainty to be accounted for.

The results of contemporary neutrino cross section measurements seem to be at odds with the earlier experiments, and in particular the measurement of M_A^{QE} is of special importance as this parameter alone will directly impact the model predictions for interaction rates. A higher value of M_A^{QE} leads to a larger cross section, thus increased event rate predictions. A need for more data is necessary not only to disentangle the value of M_A^{QE} from the effects of the different nuclear targets from the neutrino interaction cross sections, but also to reach the cross section precision necessary for next generation oscillation experiments.

2.5.1 The CC-1-pion Cross Section

The neutrino-induced charged current single pion production (CC-1-pion[1]) interaction has always been recognized as an important cross section to constrain as these interactions form a primary background in neutrino oscillation measurements. This background comes from the fact that CC-1-pion events in which the pion is not reconstructed appear to be CC-QE-like events ($\nu_\mu + n \rightarrow \mu^- + p$). As CC-QE events are used as the signal interaction in oscillation analyses, mis-reconstructing CC-1-pion events into this category contributes a significant background. However, the CC-1-pion interaction has been the subject of concern since some of the earliest measurements first made on free nucleons. The ANL [25] and BNL [26] experiments both made bubble chamber measurements of charged pion production on deuterium for muon neutrinos of energy about 1 GeV, but found significant discrepancy in their results. For experiments like T2K with a peak neutrino flux around 0.6 GeV, developing our understanding of pion production at neutrino energies less than 1 GeV is a priority.

Amongst contemporary experiments the MiniBooNE collaboration has provided a high-statistics sample of CC-1-pion measurements [27] in mineral oil (CH_2). The MiniBooNE charged positive pion differential cross section was found to disagree with simulations – these simulations also used an RFG model of the target nucleus with no further augmentations. The discrepancy between the MiniBooNE differential cross section, shown in Figure 2.5, and the well known discrepancies amongst the older experiments, has stimulated model development to incorporate the cross section, nucleus, and final state interactions. A summary of the various contemporary model development and comparisons to the MiniBooNE data sets is given by [28], from which an example demonstrating the different predictions for the differential CC-1-pion cross section with respect to the pion kinetic energy is given in Figure 2.6.

This thesis will present an event selection of CC-1-pion events and background samples to be fit, and report on the progress towards developing a measurement of the CC-1-pion

[1]Though single pion production of neutral and negatively charged pions is also possible, in the context of this these, "CC-1-pion" will always refer to CC single positively charged pion production, CC-1π^+.

differential cross section with respect to pion momentum. It is expected that this measurement will be of significant interest to the neutrino physics community and is highly anticipated.

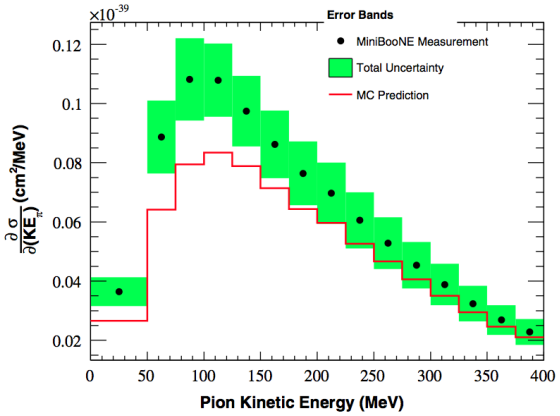

Figure 2.5: MiniBooNE differential cross section and model predictions, from [27]. In this case a RFG nucleus model was used and a clear discrepancy between the simulation and data was found.

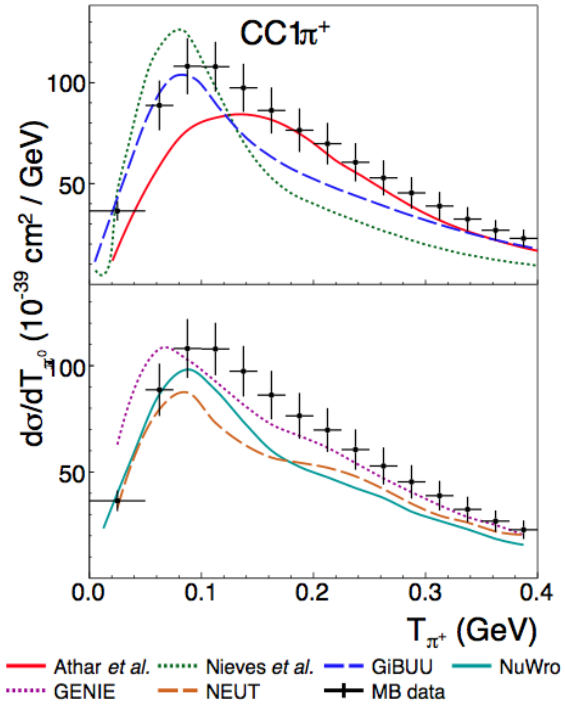

Figure 2.6: Contemporary and traditional model comparisons against the MiniBooNE differential cross section measurement. Figure taken from and additional details given in [28]. The T2K experiment uses the NEUT model, though this figure demonstrates that there is active development in modeling neutrino interactions.

Chapter 3

THE T2K EXPERIMENT

The T2K experiment is a second-generation[1] long-baseline neutrino oscillation experiment. According to the experiment proposal [29] the primary goal of the T2K experiment was to measure the last remaining oscillation parameter θ_{13} through the observation of a ν_e appearance signal in from a mostly pure ν_μ beam. Additional goals of the experiment were to make precision measurements of the known oscillation parameters Δm^2_{23} and $\sin^2(2\theta_{23})$ through ν_μ disappearance studies. The study of neutrino cross section measurements using the suite of near detectors was also proposed.

The T2K experiment uses a proton beam generated at the Japan Proton Accelerator Research Center (J-PARC) located on the east coast of Japan in the village Tokai in Ibaraki prefecture. The beam properties and neutrino flux is sampled near the production point and directed at the far detector, Super-Kamiokande, 295 km away on the west coast of Japan near the village Higashi-Mozumi in Gifu prefecture. A schematic of the baseline is shown in Figure 3.1. The experiment began taking data in January of 2010, with the near detectors running stably by Feburary 2010.

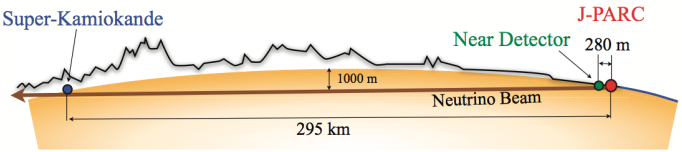

Figure 3.1: A sketch up of the earth's cross section showing the neutrino production point at J-PARC, the near detector location, and the far detector Super-Kamiokande [30].

One of the distinguishing features of the T2K neutrino beam is the use of an off-axis

[1] In Japan the first-generation long baseline experiment was the KEK-to-Kamioka (K2K) experiment.

beam. This choice was motivated by the fact that the neutrino energy spectrum becomes more narrowly peaked as one moves off the primary axis. The exchange is a loss of flux, by the gain is a a neutrino beam narrowly peaked at the expected first oscillation maximum with a reduced background from high-energy neutrino interactions. The T2K beam is fixed at 2.5° off-axis. The spectrum and beam components can be seen in Figure 3.2.

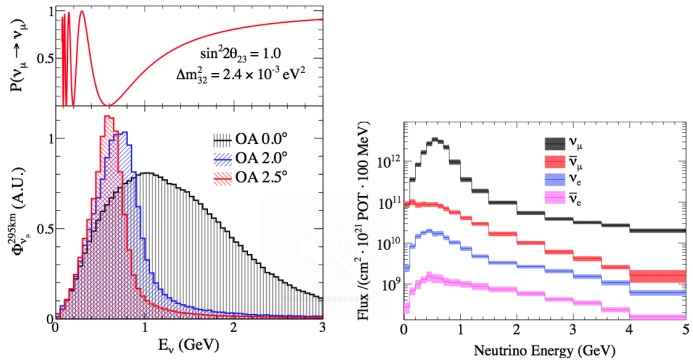

Figure 3.2: On the left the impact of neutrino energy spectrum as a function of off-axis angle is shown. Note that the on-axis flux is a broader, higher energy beam, whereas the off-axis beam is narrowly peaked at the expected oscillation maximum [31]. On the right is the off-axis neutrino flux broken down by neutrino type with systematic errors [32]. The beam is mostly pure ν_μ .

At the time of publication of [30], the T2K collaboration consisted of over 500 physicists and technical staff members from 59 institutions. Twelve countries were in participation: Canada, France, Germany, Italy, Japan, Poland, Russia, South Korea, Spain, Switzerland, the United Kingdom and the United States.

A summary of the accelerator, neutrino beam production, near detectors and far detector will be presented in the following sections. The presentation here is based on the information within [30], and in sub-sections specific references with additional content are given.

3.1 J-PARC Accelerator

Construction the J-PARC accelerator facility began in April 2004 and was finished and commissioned by 2009. The accelerator is comprised of three distinct stages: the linear

accelerator (LINAC) is first, followed by the rapid-cycling synchrotron (RCS) which feeds into the final main ring (MR) synchrotron. The LINAC is responsible for starting with a H^- beam which is accelerated up to 181 MeV (400 MeV in the final design after upgrades). An top-down schematic of the three main accelerator sections is shown in Figure 3.1.

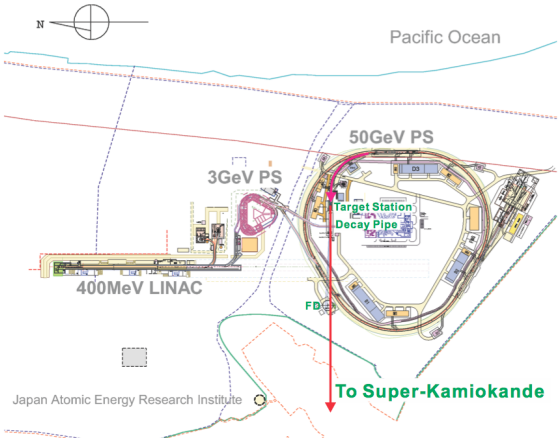

Table 3.1: Three main components J-PARC's accelerator: linear accelerator, rapid-cycling synchrotron, and main ring [29]. The original LOI called for a 50 GeV MR.

The LINAC beam is converted by charge-stripping foils to a H^+ beam as it enters the RCS wherein the beam is accelerated to 3 GeV. The RCS has a harmonic number of two and two bunches per cycle. Only about 5% of the of the RCS bunches continue to the MR (the majority are used for the Material and Life Science Facility at J-PARC) where they are accelerated to 30 GeV. The MR has a harmonic number of 9 and circulates protons in bunches of 8 (prior to 2010 the MR operated with bunches of 6). The main ring has two extraction points: A slow extraction point used for the hadron beamline facilities, and a fast extraction point used for the neutrino beamline. The T2K beam is taken from the fast extraction approximately once every 2 seconds, with all 8 bunches in a single beam "spill" of $\sim 5\mu s$. The MR design parameters are shown in Table 3.2.

Circumference	1567 m
Beam power	~750 kW
Beam kinetic energy	30 GeV
Beam intensity	$\sim 3 \times 10^{14}$ p/spill
Spill cycle	~0.5 Hz
Number of bunches	8/spill
RF frequency	1.67 – 1.72 MHz
Spill width	~5 μsec

Table 3.2: The J-PARC accelerator main ring synchrotron nominal design parameters [30]. The maximum power reached up through data taking at the end of 2012 was about 200 kW.

3.2 The T2K Neutrino Beamline

The neutrino beamline can be considered as two sequential sections: first the primary beamline, followed by the secondary beamline. The primary beamline is responsible for the extraction of protons from the MR and their redirection towards the far detector, Super-Kamiokande. The secondary beamline is made from the secondary particles (mostly pions) which are generated from the proton-target collisions. The neutrino beam is born out of the secondary beam through the secondary particle decays.

3.2.1 Primary Beamline

The primary beamline is made of three segments, shown in Figure 3.3: The preparation section (54 m), the arc section (147 m) and the focusing section (37 m). The preparation section uses normal conducting magnets to guide the beam towards the arc section. In the arc section super-conducting magnets are used to bend the proton beam 80.7° in the direction of Super-Kamiokande. In the final focusing section normal conducting magnets are used to focus the beam onto the target and steer the beam 3.637° into the ground (with respect to the horizontal).

To maintain the tuning and stability of beam operation a variety of monitors are installed in the primary beamline. A summary of the location of the beam intensity (5 current trans-

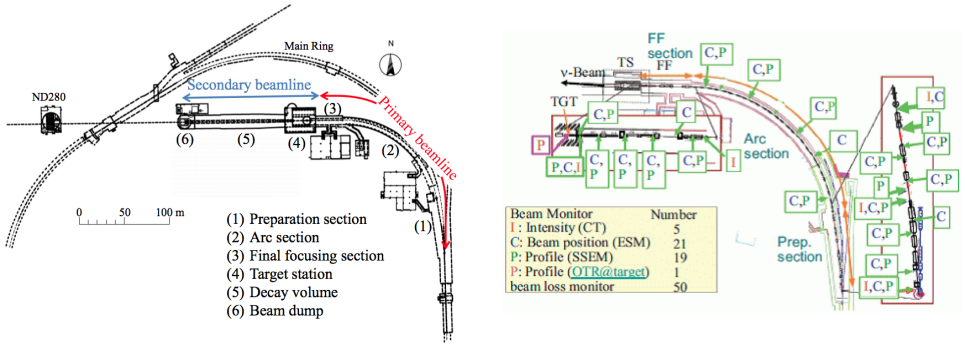

Figure 3.3: On the left is an overview of the primary and secondary T2K neutrino beamlines. On the right is a schematic showing the location and type of the various beamline monitors for intensity, profile, position and beam loss [30].

formers), beam position (21 electrostatic monitors), beam profile (19 segmented secondary emission monitors) and beam loss monitors (50) is shown in Figure 3.3.

3.2.2 Secondary Beamline

The secondary beamline is composed of three section: The target station, the decay volume and the beam dump, as shown in Figure 3.4. The target station is in a helium vessel and houses the proton target, the magnetic horns and an optical transition radiation monitor (OTR). The OTR monitors the beam profile just upstream of the target. The target, a 91.4 cm long (1.9 interaction lengths) graphite core surrounded by a 2 mm thick graphite tube, is sealed in a 0.3 mm thick titanium case and located inside the first horn.

The three horns are pulsed at 250 kA (320 kA design max) and focus the secondary pions with a 1.7 T toroidal magnetic field. The flux at the peak of the neutrino energy ($\sim$ 600 MeV) is increased by a factor of about 17. A cross section of the horns is shown in Figure 3.4.

The decay volume for the focused pions is a 96 m long steel tunnel filled with helium. For additional shielding 6 m of reinforced concrete surrounds the decay volume. At the end of the decay volume is the beam dump, which is 75 tons of graphite about 3 meters in length along the beam direction. An additional 2.4 m of iron plates are used downstream

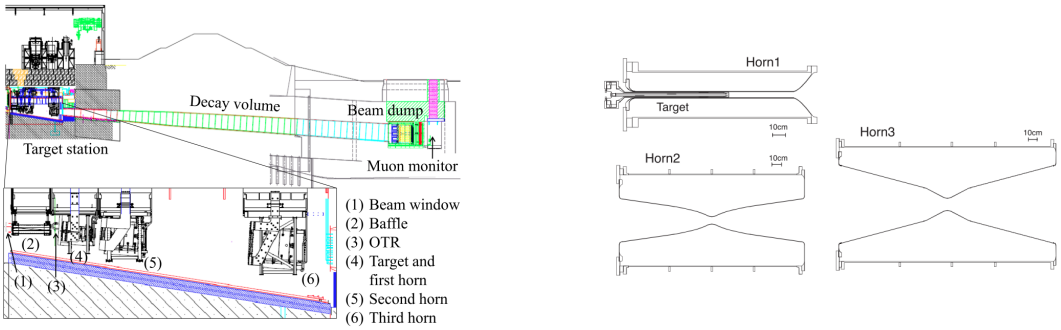

Figure 3.4: On the left is the schematic of the secondary beamline and its primary components. The baffle is a graphite collimator in place to protect the horns. On the right is a cross sectional view of the three horns, with the target placed inside the first horn[31].

of the beam dump, allowing only muons above ~ 5 GeV/c to pass.

3.3 Muon Monitor and Emulsion Tracker

To measure the beam intensity and direction on a bunch-by-bunch basis the high-energy muons which punch through the beam dump are measured by the Muon Monitor (MUMON). These muons originate with the neutrinos during charged pion two-body decay. The beam direction is taken to be the direction from the proton target to the center of the muon profile. The target precision is 0.25 mrad, or approximately 3 cm on the muon profile center. The expected flux for nominal beam operation (3.3 $\times 10^{14}$ protons/spill at 320 kA horn current) is 1×10^7 charged particles/cm^2/bunch.

The MUMON consists of two detector arrays: ionization chambers (117.5 m from the target) and silicon PIN photodiodes (118.7 m from the target). Downstream of the MUMON is an emulsion-based tracker which measures flux and momentum distribution of muons. Additional details and references on the design of each of these detectors can be found in [30].

3.4 Beam Operation Summary

A history of the neutrino beamline delivery is given in Figure 3.5. The cumulative protons on target (POT) is the standard measure of data taking in neutrino experiments, and is

shown by the monotonic increasing blue line, using the scale on the left-hand side of the graph. The first flat period in 2010 shows the typical summertime shut down. The second, year-long flat period is due to the 2011 March 11^{th} earthquake which caused serious damage to all of north east Japan, including the J-PARC facilities. The number of protons per spill as measured at the current transformer closest to the target (CT5) is shown by the red scatter plot with the scale on the right-hand side. Data is deemed good if all equipment is working properly, the horn currents are within 5 kA of the mean and if MUMON reports good alignment and beam intensity within 5% of the mean. A "good spill" list of data suitable for analysis is maintained and distributed by the beam group.

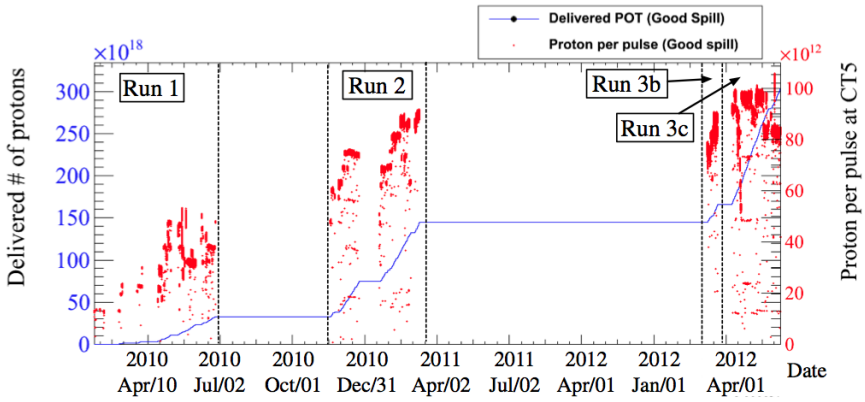

Figure 3.5: The T2K beamline history of delivered protons on target (POT) and protons per pulse [31].

3.5 Near Detectors: ND280

For any oscillation experiment it is important to measure the neutrino flux, energy spectrum and interaction rates prior to oscillation. This is achieved by detectors that are placed close to the neutrino production point. In the case of T2K, the off-axis nature of the neutrino beam requires the additional efforts of stringent alignment measurement and monitoring. To achieve these means the T2K experiment has a suite of near detectors located 280 m downstream from the beam target referred to as ND280. The ND280 detectors are housed

in a pit of diameter 17.5 m and a total depth of 37 m. The pit itself has three different floors for separate usages. A 3-D model of the pit is presented in Figure 3.6.

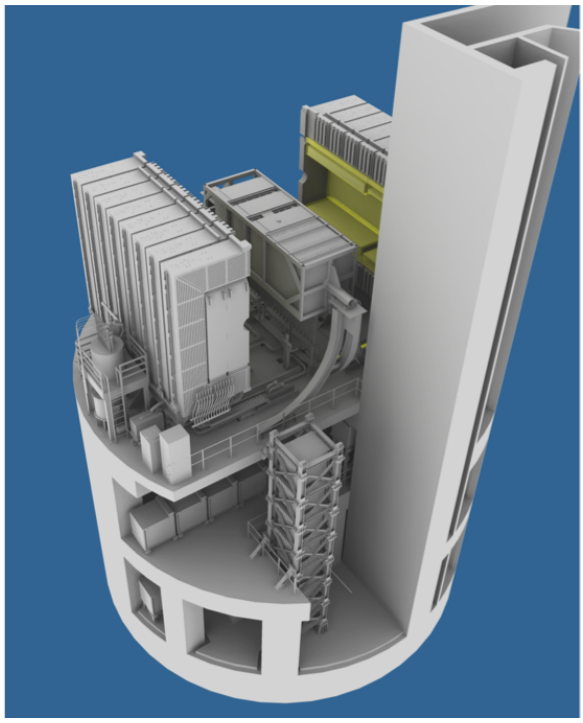

Figure 3.6: A 3-D model of the pit. The open magnet and basket regions are shown on the first floor, B1 at 24 m deep. The horizontal modules of INGRID can be seen on the service floor, and the vertical modules of INGRID extend down to B2 at 37 m deep [30].

The primary ND280 detectors, the off-axis pi-zero detector and tracking detectors sample the neutrino flux and are responsible for measuring neutrino interaction rates and cross sections. The tracker consists of three time projection chambers and two fine-grained detectors and together with the pi-zero detector reside inside a magnetized space surrounded by a recycled magnet from the UA1 experiment. The magnetized region is referred to as "the basket." A muon range detector envelopes the basket to capture any exiting muons. A grid of modular detectors, INGRID, are arranged and centered on-axis to continuously monitor the beam alignment and and intensity.

The off-axis detectors are staged on the B1 floor, which is approximately 24 m below

the surface and on the line of sight from the proton target to the far detector center, 2.5° off-axis. The bottom modules of ingrid are staged on the B2 floor, approximately 37 m deep. The third "floor"[2] is between B1 and B2 and houses the horizontal modules of INGRID, readout electronics and additional services.

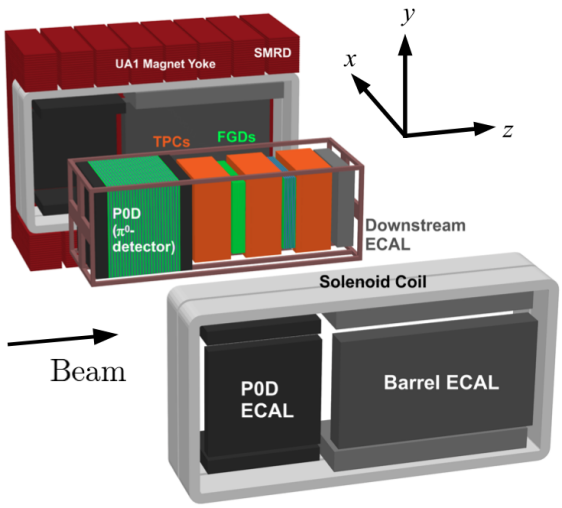

Figure 3.7: Exploded view of the ND280 off-axis detector suite. The surrounding magnet, when closed, entirely surrounds and magnetizes the "basket" which houses the PØD, TPCs, FGDs, and ECals [30].

3.5.1 Multi-Pixel Photon Counter

All of the scintillation-based detectors in the ND280 complex use the Hamamatsu[3] Multi-Pixel Photo Counter (MPPC). Given the limited space availability in the ND280 pit and the magnetization of the basket detectors, a device which was both compact and insensitive to magnetic fields was necessary. The MPPC device, pixel detection plane are shown in

[2] As a piece of T2K history, the three floors of the ND280 pit used to be arranged in the more logical B1, B2, and B3. Due to mysterious regulations, only two "basement" floors were allowed, to which the easiest solution was to *rename* the original second floor a *service* floor. The elevator button was changed from B2 to S, and B3 became B2 and any discussion amongst the old guard and the new researchers became very complicated.

[3] http://www.hamamatsu.com

Figure 3.8, and design parameters are given in Table 3.3. In total 64,000 devices were manufactured for T2Ks use.

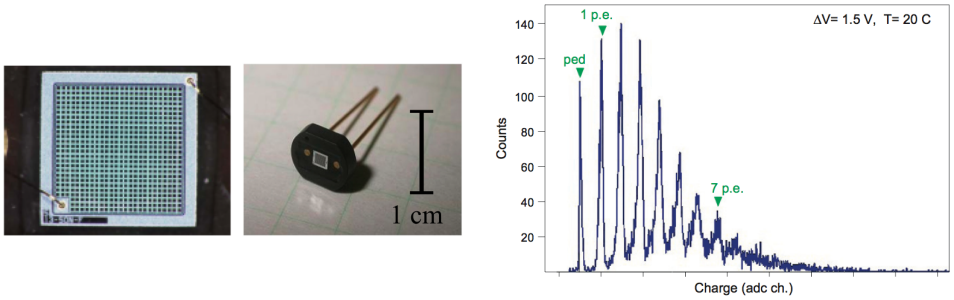

Figure 3.8: At the left is an image of an MPPC device, showing the pixel face which is 1.3 $\times$ 1.3 mm^2 in a 26 $\times$ 26 array of 667 [30]. On the right is a charge amplitude spectrum measured with an LED source, demonstrating the single photon resolution [33].

The detectors which use MPPCs use identical electronics to read them out, based on the Fermilab Trip-T chips [34] mounted on Trip-T Front end Boards (TFBs) [34].

Number of pixels	667
Active area	1.3×1.3 mm^2
Pixel size	50×50 μm^2
Operational voltage	68 – 71 V
Gain	$\sim 10^6$
Photon detection efficiency at 525 nm	26 – 30%
Dark rate, threshold = 0.5 p.e., T = 25 °C	$\leq$ 1.35 MHz

Table 3.3: Primary parameters of the MPPC design [30].

The MPPC operates by treating each pixel in its detection surface as a binary (i.e. *hit* or *not-hit*) element that is insensitive to the incident amplitude in that particular pixel. If a pixel is hit a Geiger avalanche is created and the signal is counted, thus the MPPC signal is a linear sum of the number of fired pixels and its dynamic range is limited by the total number of pixels on the device itself. An example signal demonstrating the single photon resolution of a standard MPPC is shown in Figure 3.8. Additional details of the MPPC device can be found in [33].

3.5.2 The Magnet and the Side Muon Range Detector

When taking neutrino beam data the ND280 tracker and PØD are magnetized, which allows for particle momentum and charge sign measurement. The field is provided by the recycled CERN UA1/NOMAND experiments magnet. The 850 ton magnet is made up of 16 c-shaped iron yokes and an aluminum coil, and is divided into two mirror-symmetric halves which can open and close as necessary to access the inner detectors. When operating, the magnet provides a 0.2 T dipole magnetic field in the horizontal direction along the $-x$-axis in detector coordinates and requires 2900 A with a voltage drop of 155 V. The outer dimensions are 7 m $\times$ 5.6 m $\times$ 6.1 m. After installation the magnetic field was mapped before the installation of the remaining ND280 detectors, as can be seen in Figure 3.9. Additional details on the operation of the magnet can be found in [30].

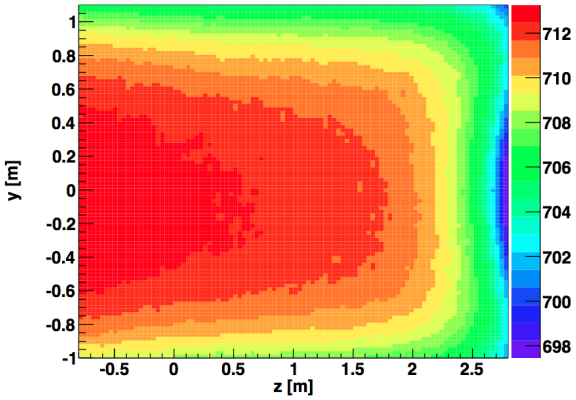

Figure 3.9: Magnetic field map (in Gauss) at $x = 0$ taken prior to ND280 detector installation with the magnet operating at 0.07 T [30]. The neutrino beam goes from negative to positive z in this coordinate system.

The UA1 magnet also houses the side range muon detector (SMRD). As can be seen in Figure 3.10, the yokes of the magnet have air gaps between the iron layers. Each yoke consists of 16 iron played and 15 air gaps, 48 mm thick and 17 mm thick, respectively. The SMRD consists of scintillator panels placed in the horizontal and vertical air gapes of the magnet yokes. The primary goal of the SMRD is to measure the muon momenta of high-angle muons exiting the tracker. The SMRD naturally serves as an external veto for

neutrino interactions in the surrounding materials, and as a cosmic ray trigger [35].

The SMRD utilizes polystyrene scintillator slabs doped with 1.5% PTP and 0.01% POPOP, and is the only detector to use a distinct S-shaped threading of wave length shifting fibers. The standard scintillator module can be seen in Figure 3.10. The modules are read out at both ends by MPPCs and in total comprise 4016 channels.

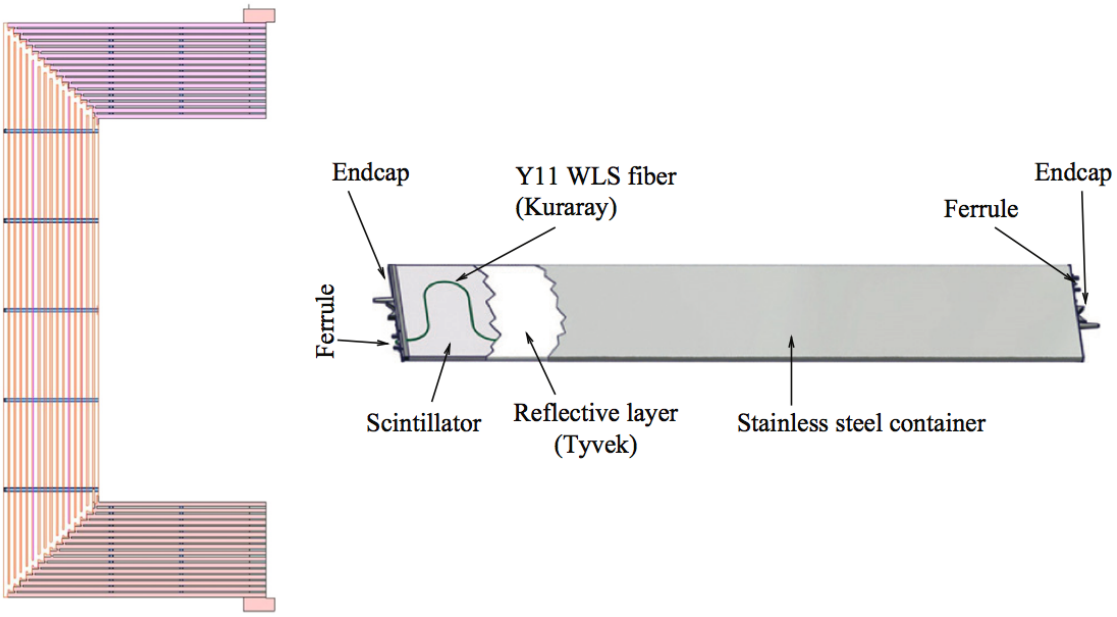

Figure 3.10: Cross sectional view of one magnet yoke (left), demonstrating the air gaps between the iron layers in which the SMRD scintillator modules (right) are instrumented [35].

3.5.3 The Pi-zero Detector

The pi-zero detector (PØD) is the most up-stream detector in the ND280 basket with the primary goal of measuring the neutrino interaction cross sections that generate π^0's. Design was focused on the measurement of the neutral current process $\nu_\mu + N \rightarrow \nu_\mu + N + \pi^0 + X$ which is the primary background in the $\nu_\mu \rightarrow \nu_e$ appearance signal. Salient features of the PØD design and operation are discussed here, but additional technical details can be found in the detector paper [36].

The PØD is shown in figure 3.11, where both an external view showing the mounting of the electronics is given, and a cross-sectional view showing the detailed breakdown of the individual detector regions is also shown. There are four main regions of the PØD, the

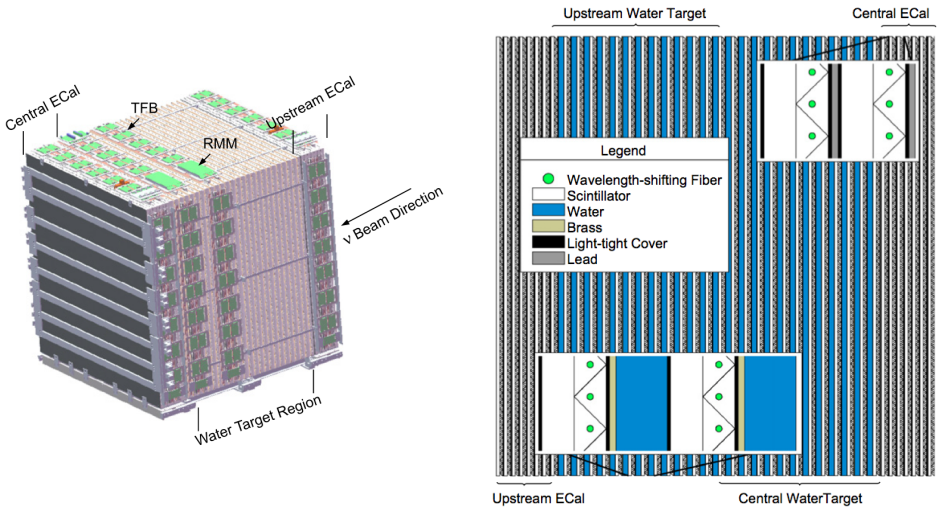

Figure 3.11: The PØD, showing the mounting of the electronics on the detector and the detailed elements on the right. Note that in the figure on the right the beam direction is from left to right, with the ND280 tracker downstream of the PØD[36].

"upstream" and "central" ECals are composed of alternating scintillator planes and lead sheets. These two regions provide a veto against external particles such as sand muons and cosmics entering the central region of the detector. The central region houses the water targets, and is broken into two sections ("upstream" and "downstream" water targets). The water target region is composed of alternating scintillator planes, water bags, and brass sheets. The details of these material arrangements can be seen in Figure 3.11.

In total there are 40 scintillator modules (PØDules). Each PØDule is composed of 134 horizontal bars (2133 mm long) and 126 vertical bars (2272 mm long). The individual bars are threaded with wavelength-shifting fiber, with bars mirror-coated at one end and read out by Hamamatsu multi-pixel photon counters (MPPCs [reference to MPPC paper]) at the other end. In total there are 10,400 channels in the PØD.

In the water target region there are 25 water target layers, with each layer being com-

posed of two individual water bladders. Each bladder can be filled and drained independently, feeding from or into an external 3000 liter water tank. The water bladder depth and stability is monitored continuously. The author of this thesis contributed to the installation, monitoring, and maintenance of the water system, as well as in the analysis for the calculation of the fiducial volume target mass.

The active region of the PØD has dimensions 2103 mm × 2239 mm × 2400 mm (width × height × length) and has a mass of 15,800 kg when filled with water, and 12,900 kg when dry. The specifications of each detector region are give in Table 3.4

Super-PØDule	Mass (kg)	Dimensions (mm × mm × mm)	Depthin R.L.
Upstream ECal	2900	2298 × 2468 × 305	4.946
Upstream water Target:		2298 × 2468 × 888	
Empty	3600		1.370
Filled	5100		2.379
Central water Target:		2298 × 2468 × 854	
Empty	3500		1.356
Filled	4900		2.287
Central ECal	2900	2298 × 2468 × 304	4.946

Table 3.4: PØD mass, dimensions and radiation lengths for each detector region [36].

3.5.4 Time Projection Chambers

The ND280 basket contains three TPCs, with TPC1 immediately downstream of the PØDand with the two fine-grained detectors located between TPC2 and TPC3. The TPCs are filled with an ionizing gas mixture which emanates electrons when charged particles pass through the active region. The electrons are then drifted to the read out planes by an electric field. The basket region is magnetized and the electric field runs parallel to the magnetic field. From the ionization the curvature of the particles can be measured to allow determination of the momentum component perpendicular to the magnetic field, and from the charge deposition as a function of energy a particle ID can discriminate electrons, muons and protons. Additional details can be found in [37].

The TPC structure uses a double box design, as shown in Figure 3.12. The inner box

is coated with copper strips and forms the field cage. The outer box acts as the ground potential. The maximum drift distance is 897 mm (center cathode out along x), with a drift field of approximately 275 V/cm. Along the direction of the neutrino beam (z), the active region of the TPCs are 720 mm in length after excluding regions near the walls where the field is not uniform.

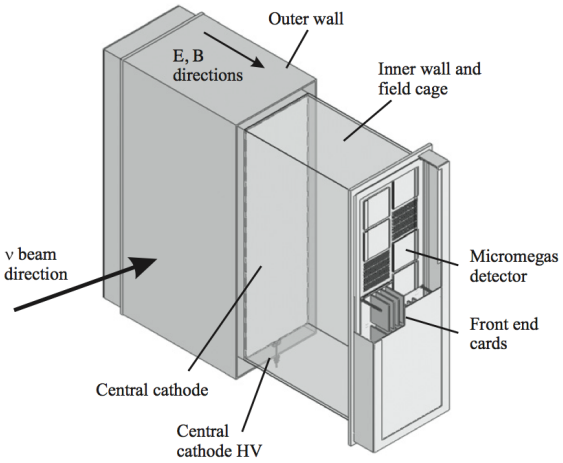

Figure 3.12: Simplified diagram of the primary TPC features. The cathode is in the middle of each TPC and drifted electrons are read out at the walls at $\pm x$, where z is along the beam direction and y is upward [37].

Between the inner and outer box is a CO_2 gas, and within the inner box is the "TPC gas, which is nominally 95% Argon, 3% CF_4, and 2% iC_4H_{10}. A sophisticated gas system is used to maintain temperatures and pressures, as well as minimize contamination from other gases (such as O_2, which absorbs electrons) which impact drift velocity and gas gain. The TPCs are equipped with a laser-based photoelectron calibration system which operates by shining a 266 nm light on an aluminum target. With this method electrons can be emitted in a controlled pattern to monitor drift velocity and distortions in the electron drift.

Unlike the rest of the scintillator-based detectors in ND280, the TPCs use micromegas [38] detectors to read out the clusters of drifted electrons. A single module contains a grid of 1,728 channels over an area of 36 cm $\times$ 34 cm. Each TPC has 12 modules on each x-wall (the electron drift is from the *central* cathode out along x in both directions), so the three

TPCs together make up 124,416 channels.

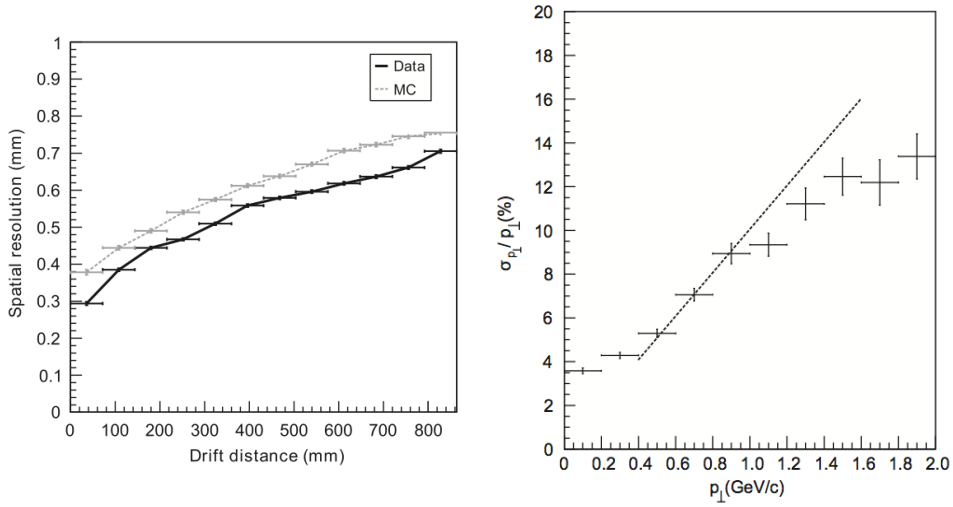

Figure 3.13: On the left is a comparison of spatial resolution, and on the right a comparison of perpendicular momentum resolution. Both are show data and simulated results [37].

The spatial resolution depends on the drift distance for a particular track, but the average for clusters that hit multiple pads in a micromegas is well below 1 mm. The design goal of a perpendicular momentum resolution of $\delta(p_\perp)/p_\perp < 0.1p_\perp$ (GeV/c) was also met. The resolution of the momentum measurement decreases with higher energy tracks as their curvature across the active region decreases. Data and monte carlo comparisons for spatial and momentum resolution are shown in Figure 3.13. The energy loss as a function of momentum for negatively and positively charged particles, along with the monte carlo expectations for particle types is shown in Figure 3.14. The particle ID algorithm compares the energy loss per unit length to expected values for a given particle type hypothesis. An example event display demonstrating the tracking of the TPCs can be seen in Figure 3.15.

3.5.5 Fine-Graind Detectors

Located between the TPCs are two fine-grained detectors (FGDs). The first FGD, FGD1, is a fully-active, scintillator-based detector and is located between TPC1 and TPC2. The

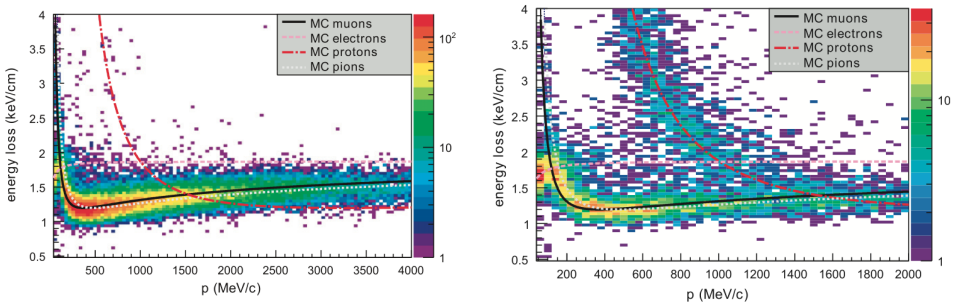

Figure 3.14: Energy deposition as a function of momentum for negatively charged tracks (left) and positively charged tracks (right) [37]. Expectations for simulated particle type are overlaid.

second FGD, FGD2, is composed of alternating scintillator and water layers and is located between TPC2 and TPC3. The primary features and design goals discussed here are taken from [39], wherein more details can be found.

The design of the FGDs were driven by a few different goals. Firstly, FGD1's primary purpose is to provide an interaction target mass for the tracker subsystem of ND280. The first TPC acts as a veto, followed by the target FGD1, and then the subsequent TPCs and FGD2 form the core of the tracker. To utilize the features of the TPCs it was necessary to ensure that the FGDs themselves were not too thick along the axis of the neutrino beam (z) such that charged particles would not leave the detector volume. An example event display is shown in Figure 3.15, showing the FGDs and TPCs working as a tracking system.

Neutrino oscillation analyses rely on the CCQE interaction ($\nu_l + n \rightarrow l^- + p$) to obtain the parent neutrino energy, which is calculable from the outgoing lepton four momentum. A primary background to this interaction is the CC-1pi interaction ($\nu_l + N \rightarrow l^- + N' + \pi$). To better understand this background, FGD1 was designed to be capable of detecting all charged particles exiting an interaction vertex with good reconstruction and particle ID efficiencies.

To bypass the difficulties of calculating the variation in nuclear re-interaction effects (e.g. pion absorption, production and charge exchange) across different nuclear targets, the FGD2 was designed to have interleaved water target layers. By comparing the interaction

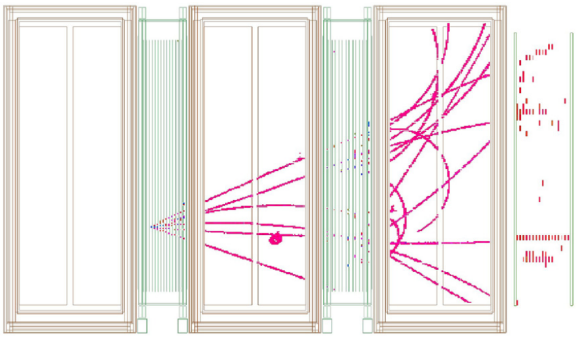

Figure 3.15: An example event display of a neutrino interaction in FGD1, with subsequent particles entering the tracker [39].

rates between FGD1 and FGD2, the neutrino interaction rates on a water target can be measured and used to better our understanding of results at the far detector, Super-K, which uses a water target.

The scintillator bars used in the FGDs are made of polystyrene doped with PPO (1%) and POPOP (0.03%), and were co-extruded with a TiO_2-doped reflective coating. Unlike the PØD, the FGD design used square bars with an edge length of 9.6 mm. Each bar is threaded lengthwise with a wavelength shifting fiber. A cross sectional picture of a bar is shown in Figure 3.16.

A plank of 192 bars in the horizontal and 192 bars in the vertical directions form an "XY module, with dimensions 186.2 cm $\times$ 186.2 cm $\times$ 2.02 cm. Within an individual x-layer or an individual y-layer, alternating bars are read out by MPPCs at alternating ends, thus the FGDs' readout electronics are instrumented on all four sides of the detector. FGD1 is composed of 15 XY modules, providing a target mass of 1.1 tons. FGD2 is made of 7 XY modules and 6 water layers. The total number of channels is 8448, with FGD1 having 5760 and FGD2 having 2688. Each FGD has outer dimensions of 230 cm (width) $\times$ 240 cm (height) $\times$ 36.5 cm (depth).

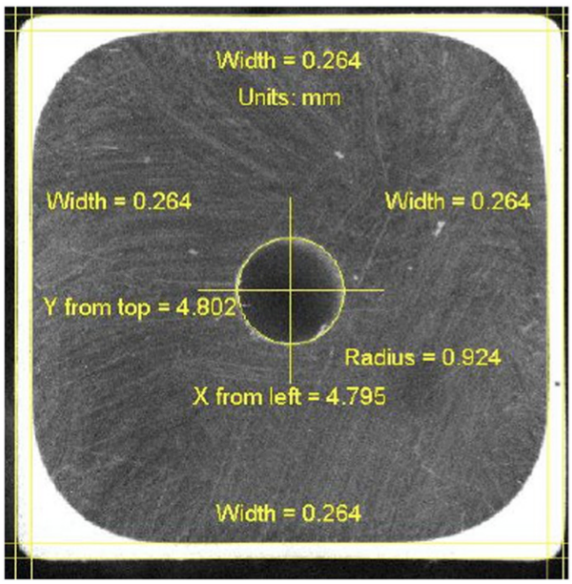

Figure 3.16: A CCD image of a typical scintillator bar. The hole center and TiO_2 thickness is determined by automated image analysis [39].

3.5.6 Electromagnetic Calorimeters

The entirety of the basket is enveloped by electromagnetic calorimeters (excluding the PØDs upstream face, which utilizes the PØDs upstream section which is also an electromagnet calorimeter). There are 13 individual modules total, which can be grouped into three primary categories: The PØD ECal, the Barrel ECal, and the Downstream ECal (DS ECal). The orientation and arrangement of these modules is most easily conveyed through the exploded view of ND280 in Figure 3.7. The primary goal of the ECal modules is to capture particles exiting the PØD and tracker detectors, assisting in particle identification and tracking, as well as acting as an external veto.

Each ECal module is a composed of alternating layers of polystyrene scintillator co-extruded with a reflective TiO_2 coating, and sheets of lead, which acts as the target and radiator. The ECal scintillator is doped with 1% PPO and 0.03% POPOP fluors. The cross section of each scintillator bar is identical in all modules and rectangular in cross section, measuring 40 mm $\times$ 10 mm. Each scintillator bar has a hole running lengthwise along the

bar in which a wavelength shifting fiber is located and read out by MPPC. Depending on the module the fibers are read out from either both ends or a single end. The length of each bar depends on both the orientation and the module. The design details of each module are summarized in Table 3.5.

The differences in each category of ECals is slightly different. In the case of the PØD, the goal is to capture exiting particles and distinguish between muons and photons, therefore a design using thicker lead sheets was used for the PØD ECal. In contrast, the Barrel ECal had the goal of providing a finer resolution and tracking ability to enhance the tracker analysis, so a thinner choice of lead sheets and additional scintillator layers was used.

All ECal modules are equipped with a light injection calibration system. This system uses a puled LED to illuminate the wavelength shifting fibers and enables routine integrity and calibration tests.

	DS-Ecal	Barrel ECal	P0D ECal
Length (mm) Width (mm) Depth (mm) Weight (kg)	2300 2300 500 6500	4140 1676 top/bottom 2500 side 462 8000 top/bottom 10000 side	2454 1584 top/bottom 2898 side 155 1500 top/bottom 3000 side
Num. of layers	34	31	6
Bar orientation	x/y	Longitudinal and Perpendicular	Longitudinal
Num. of bars	1700	2280 Longitudinal top/bottom 1710 Longitudinal sides 6144 Perp top/bottom 3072 Perp sides	912 Longitudinal top/bottom 828 Longitudinal sides
Bars per layer	50	38 Longitudinal top/bottom 57 Longitudinal side 96 Perp top/bottom/sides	38 Longitudinal top/bottom 69 Longitudinal sides
Bar length (mm)	2000	3840 Longitudinal 1520 Perp top/bottom 2280 Perp sides	2340 Longitudinal
Pb thickness (mm)	1.75	1.75	4.0

Table 3.5: Summary of the various ECal modules and their design properties [40].

3.5.7 Interactive Neutrino Grid

The Interactive Neutrino Grid (INGRID), Figure 3.17, is a modular detector array centered on the beam axis. A defining feature of the T2K beam and experimental setup is the use of the off-axis beam and INGRIDs primary role is to monitor what is defined as the beam center. The summary of the basic features of INGRID presented here are based on [30, 41].

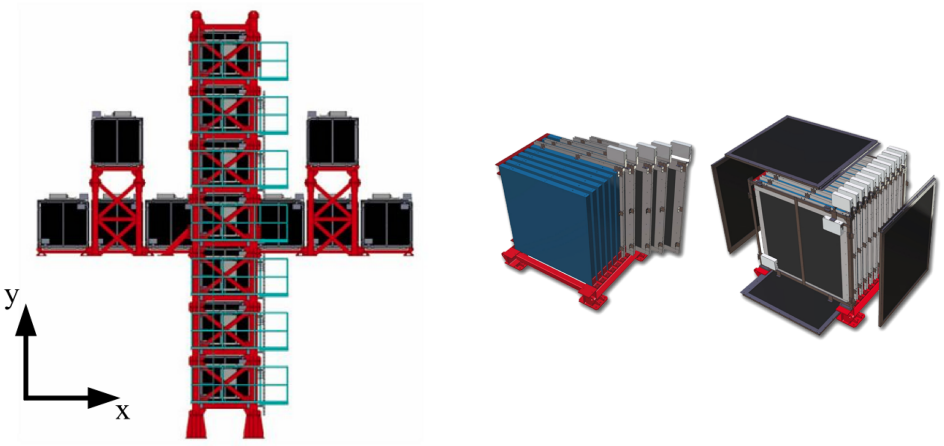

Figure 3.17: The INGRID modular array is shown at left, and an individual module is shown at right. In the right-hand figure, the blue layers represent tracking segments, and the black planes are veto layers [30].

The INGRID detector is arranged in a crossed grid, consisted of 7 horizontal modules and 7 vertical modules, with two off-axis modules included to measure axial symmetry. The overlap of the two central modules is aligned with the direction of the proton beam and defined to be the on-axis center of the neutrino beam. The INGRID detector covers a 10 m $\times$ 10 m transverse area and collects daily statistics on the number of neutrino interactions in each module to monitor both the beam intensity stability and the beam alignment within 1 mrad (approximately 10 cm at INGRID's location).

The INGRID modules are made of alternating scintillator layers and iron plates. The scintillator is polystyrene doped with 1% PPO and 0.03% POPOP, co-extruded with a TiO_2 coating and has a wavelength shifting fiber threaded through the center. Each scintillator layer is made of two planks of 24 scintillator bars (one oriented in the x direction, one in

the y). The iron plates are 124 cm $\times$ 124 cm and 6.5 cm in the axis along the neutrino beam. Each module consists of 11 scintillator layers and and 9 iron layers (there is no iron between the last two scintillator layers). The iron mass per module is 7.1 tons. Additionally, each INGRID module is also surrounded by a veto plane made of 22 scintillator bars. The INGRID layout and individual module design can be seen in Figure 3.17.

The total number of channels is 9,592 (1,144 for the veto planes), with each channel being read out by an MPPC at one end. Additional characteristics of the MPPCs characterized for INGRID can be found in [42]. INGRID was calibrated with cosmic rays on the surface prior to installation in the pit, and during data taking after installation.

3.6 Far Detector: SK

Super-Kamiokande (SK) [43], the 50 kton water Cherenkov detector instrumented with approximately 13,000 photomultiplier tubes (PMTs), is used at the far detector for the T2K experiment. SK is located 295 km away from the neutrino production point in Gifu prefecture and is built 1 km deep inside Mt. Ikenoyama. The detector is cylindrical in shape and separated into an inner detector (ID) and an outer detector (OD). The OD acts as a veto and provides additional information to categorize events that originate and remain within the ID. The OD is instrumented with 1,885 outward-facing 20 cm diameter PMTs. The ID is instrumented with 11,129 50 cm diameter PMTs and has a diameter of 33.8 m and a height of 36.2 m. A schematic of the detector is shown in Figure 3.18. The SK detector has been operating 1996 and is a stable, well-understood detector. Additional details of the calibration and detector simulation can be found in [43].

3.7 Time Synchronization

The time synchronization system provides $\sim 50ns$ synchronization between the beam spill at J-PARC and event times at Super-Kamiokande. At both ends of the experiment a pair of commercial GPS receivers (primary and backup) is used to provide GPS-based 1 pulse per second signals. These 1 pps signals are referenced by local rubidium clocks which provide timing to the electronics which are integrated with the data acquisition systems. With this system, in the case of brief GPS signal loss accurate timing can be maintained. The systems

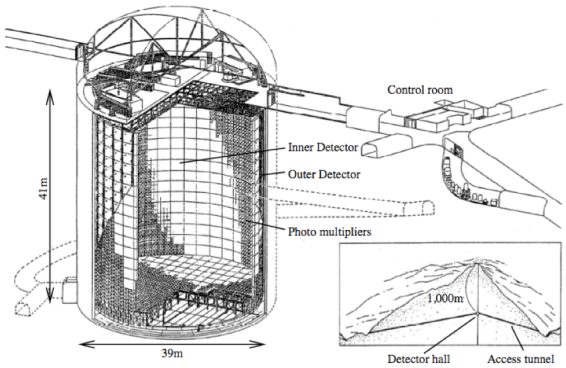

Figure 3.18: Schematic of Super-Kamiokande, the far detector in the T2K experiment [30].

at each site run in parallel continuously to minimize downtime. Additional details on the commercial units used can be found in [30].

Chapter 4

SOFTWARE

The analyses performed at the near detector sites depend on a range of software for various tasks of the data collection, simulation and analysis. These responsibilities are divided into different working groups. Most pertinent to this thesis are the various stages necessary for data analysis of ND280 results. A summary of the near detector data collection process will be given, and an outline of the steps performed in the data simulation which produces T2K's official fake data sets will be presented.

Naturally, one goal is to create a monte carlo (MC) data set as accurate and representative as possible of the real data. This priority necessitates that the storage, treatment, and analysis of the real and MC data sets is identical. The simulation steps are, for better or worse, outlined in one monumental flow chart shown in Figure 4.1, from which the salient points will be presented. The names of most software packages have the prefix "oa," stemming from "off-axis," and will be parenthetically referenced in the descriptions below.

4.1 Flux Simulation

The neutrino MC generation begins with the primary proton beam. A package called FLUKA[44] is used to simulate the primary proton beam interactions with the target, the surrounding horn material, the upstream baffle (a collimator). FLUKA2008 is used, as it has good agreement with external data. The secondary particle kinematic information is passed to a package called JNUBEAM, which is based on Geant3 [45] and GCALOR. JNUBEAM is responsible for propagating the particles through the target station materials, the decay volume and onto the beam dump. The decays of the propagated particles are then calculated using the published branching fractions and decay rates from the PDG [17]. The kinematic information of the entire neutrino production chain, starting from the initial proton, is saved to allow reweighting of different physics.

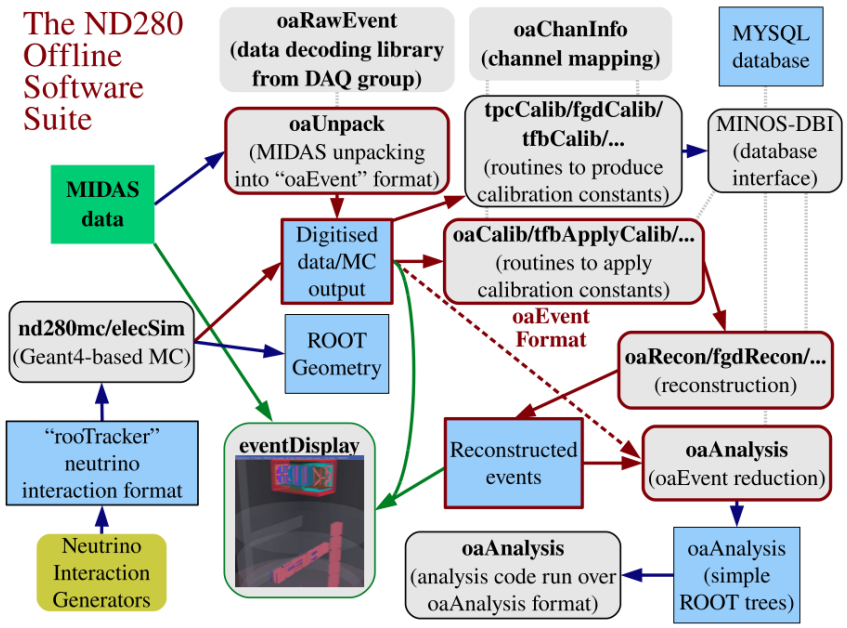

Figure 4.1: Flowchart of the various stages and their dependencies within the ND280 software suite that is responsible for MC generation and data analysis [30].

A significant amount of effort has been put forth by the beam group in collaboration with the NA61/SHINE experiment. The NA61/SHINE experiment took place at CERN and made measurements of both charged pion and charged kaon production [46, 47] using a thin target (2 cm) and a replica of the T2K target with a 31 GeV/c proton beam. The results from the NA61 experiment were used to tune the beam simulation.

In addition to the external data, the beam group also uses the many measurements from the beamline monitors to provide MC tuning parameters that correspond to each data period. Additional details on the flux prediction can be found in [31].

4.2 Neutrino Interaction Simulation

Neutrino interaction experiments use a special class of physics software often referred to as "neutrino generators," or within context simply "generators." Generators are responsible for taking the cumulative knowledge of all participating physics interactions and producing outgoing particle "vectors." The typical input to a generator is a target atom (e.g. carbon) or

free nucleon and a hypothesized neutrino flux. The output put is a collection of interactions, and for each interaction all pertinent information is stored. Information starting with the incident neutrino all the way through the to the outgoing final state particles is stored, including the internal re-interactions of particles as they traverse the target nucleus before exiting.

The de facto generator for the T2K experiment is NEUT [48]. However, as a variety of experiments publish results based on the GENIE [49] generator, this package is also used. All T2K official MC productions are made with both NEUT and GENIE. For the simulations and analyses presented in this thesis, NEUT v5.4.1.2 was used.

To apply a generator to T2K, the flux provided by the beam group is used in combination with a geometry description of the ND280 detectors. The geometry content is based in the ROOT [50] framework. With the proper input neutrino flux and the distribution of materials in all of the detectors, neutrino interactions at specific coordinates on specific target elements can be simulated.

The NEUT generator uses the Llewellyn Smith model [51] to simulate CCQE interactions. To model the momentum distribution of nucleons within target atoms the relativistic fermi gas (RFG) model of Smith and Moniz is implemented [24]. Pion production is implemented with the Rein and Sehgal model [52], with the additional implementation that 20% of delta resonances decay without producing a pion (pionless delta decay).

The multi-pion and DIS interactions, which are a topic explored in detail in Chapter 6, are managed by a few different methods. For multi-pion events with a hadronic energy below a certain threshold NEUT uses an internal routine to calculate hadron multiplicities. For DIS and multi-pion production above an energy threshold, the interactions are handled by the PYTHIA/JETSET [53] package. The studies of this thesis revealed that the transition between the internal NEUT methods and the PYTHIA/JETSET package is less than ideal.

The GENIE generator is an open-source project with an excellent user's manual and documentation which details the implementation of the various models referenced here [49]. Noteworthy is the implementation of the transition between their internal multi-pion model [54] and the PYTHIA/JETSET package: a linearly weighted transition between the two models is used, as compared to NEUT's hard threshold at 2 GeV.

The neutrino vectors produced at the generator level are passed forward to the detector-specific simulation stages.

4.3 Detector Simulation

The ND280 detector simulation (nd280mc) is based on a geometry in the ROOT [50] framework. Particle trajectories through the detector are simulated with the Geant4 [45] package. The detector simulation keeps track of the deposition of energy throughout the detector components, re-interactions such as scattering, and secondary particle decays within the detector.

Neither the neutrino generator stage nor the detector simulation itself is dependent on the beam timing that is actually used within the experiment. To account for this, the number of neutrino interactions per beam spill is assigned based on the beam intensity, and they are distributed in a timing that simulates the individual bunches within the spill.

The energy deposition information is passed next to the electronics simulation.

4.4 Electronics Simulation

The electronics simulation (elecSim) is responsible for taking the simulated depositions of energy and producing an expected response of the electronics readout. A majority of the ND280 detectors are scintillator based with wavelength shifting fibers (which are included in the geometry description). The properties of these elements, such as the propagation and attenuation of light along the fibers, and the electronics response is simulated. Electronic effects and the MPPC response such, as noise and saturation are involved at this stage. The TPC's ionization drift and readout electronics are also simulated.

The output of the electronics simulation is saved in the same format as the real data.

4.5 Calibration

Entire calibration suites (oaCalib) are designed for each sub-detector and managed by individual groups. However, conceptually there are similar tasks involved. As a majority of the detectors are scintillator based and use similar electronics. Characterizing the response of the electronics during dark periods (i.e. between events when no energy deposition is

found), which are referred to as "pedestal" triggers, and during cosmic triggers is carried out between all beam spills over all data taking periods.

The pedestal data allows calibration and characterization of dark noise. The cosmic triggers, which can be analyzed to select high-energy through-going cosmic rays can be used to calibrate the scintillator response to minimum ionizing particles (MIP). The cosmic trigger data turns out to be large enough that high-statistics can be aggregated for small regions of most detectors, allowing bar-to-bar MIP calibrations over time.

The author of this thesis was responsible for realizing that the PØD calibration scheme, which calibrated only ~ 40 regions, could be vastly improved. A new system was developed and a fine-grained calibration of the detectors MIP response in 640 distinct regions was established. An example of MIP calibration in the PØD is shown in Figure 4.2.

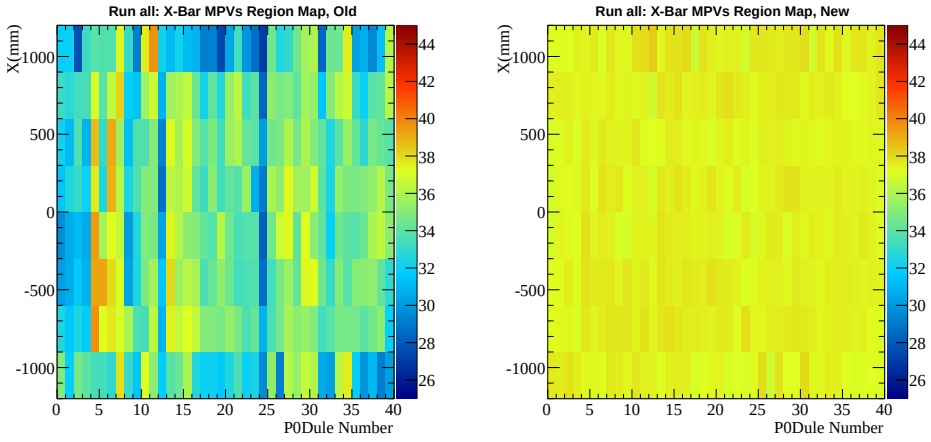

Figure 4.2: An example demonstrating the effect of MIP calibration of scintillator bars in the PØD. Using cosmic ray events groups of scintillator bars can be calibrated to ensure a uniform MIP response across the detector. The before image (left) shows a varied response across the detector, whereas the post-calibration response (right) is mostly uniform.

The calibration constants for all sub detectors across all run periods is stored in a MySQL database (along with the slow-control data such as ND280 pit temperatures, humidities, power supply voltages and so on).

4.6 Real Data

The real beam data is saved in the MIDAS [55] format and processed by an specific package which "unpacks" the data into the save software event class (oaEvent) used by the MC, which is based upon the ROOT framework.

After calibrations are applied the data and the MC are treated by the reconstruction packages.

4.7 Reconstruction

The data and MC are treated equally by the reconstruction algorithms. The reconstruction processes can be broken down into "local" and "global" reconstruction. The local processes are the reconstruction routines unique to each sub-detector, e.g. PØD Recon, FGD Recon, and so on. Each sub-detector is responsible for performing a reconstruction routine on the local data, and the results of those algorithms are forwarded to the Global Reconstruction package (oaRecon). The global reconstruction attempts to "stitch" together the local objects into global tracks that traverse multiple detectors and produce events that are common to the entirety of ND280. The global reconstruction process is based on a package called RECPACK [56], and the reconstruction packages of each sub detector are detailed in [57]. As the TPCs provide excellent momentum and position resolution they are, when possible, taken as the seed tracks in global reconstruction, which are then matched to FGD tracks.

Chapter 5

THE CHARGED CURRENT MULTIPLE PION SELECTION

The analysis presented in this thesis is based on a selection of data from the ND280 tracker, using primarily the FGDs and TPCs presented in Section 3.5. This selection is called the Charged Current Multiple Pion (CC Multi-Pi) ν_μ event selection, and uses the ND280 data collected up through May of 2013.

As the primary goals of T2K are focused on oscillation analyses, the original intent of this analysis was to provide a near detector-based constraint on the far detector flux and cross section parameters.

This selection is based on CC interactions that occur in the fiducial volume of the first FGD and produce tracks that enter the downstream tracker (TPCs and the second FGD). An inclusive CC selection is made first by selecting events with a negative muon that is matched between the FGD and the TPC. This sample is then categorized into three samples:

- **CC-0-Pion/CCQE Sample** Contains no pions
- **CC-1-Pion Sample** Contains one and only one positive pion
- **CC-Other** Contains a single negative or neutral pion, or multiple pions

This chapter will focus on the inclusive CC event selection, and will examine the CC-1-Pion and CC-Other samples. The systematic studies will be summarized. The information presented here is based entirely on the T2K internal technical note [58].

5.1 Data Summary

The data used in this analysis is from T2K data runs 1 through 4, which correspond to the start of data taking in 2010 up until the end of data in May 2013. The metric for data in

a neutrino experiment is protons on target (POT). Of the POT delivered a subset which pass all data quality flags is selected. For the beam the data quality conditions are with regard to beam alignment and intensity and horn operations. For the ND280 complex a series of checks for each sub-detector is made (e.g. event rates, energy depositions, vertex distributions, dark noise rates, timing checks, voltages, temperatures, and so forth). During nominal operations each event is tagged with a good data quality flag. The POT delivered per run period is shown in Table 5.1.

To do comparisons between data and Monte Carlo (MC) a sample of MC specific to the conditions of each run period was generated. About an order of magnitude more MC than data was produced, corresponding to 9.545×10^{21} POT.

An additional source of events found in ND280 is due to neutrinos interacting in the environment surrounding ND280. The ND280 pit itself has concrete walls, but as J-PARC is near the Pacific ocean it is surrounded by a sand environment. Therefore these beam-induced events from the environment are referred to as "sand muons," and produce mostly forward-going muons parallel to the neutrino beam. A dedicated sample of sand muon simulation was also made for each run period and included. The summary of data, beam MC and sand muon MC is given in Table 5.1.

5.2 CC Inclusive Selection

CC neutrino interaction will produce a charged lepton associated with the flavor of the neutrino. For ν_μ interactions, the CC interaction should produce a μ^-. Therefore the CC ν_μ selection looks for events that produce a μ^- that can be associated with a starting vertex in the fiducial volume of the first FGD. This is called an "inclusive" CC selection, as it does not discriminate between the specific sub-samples such as CC-1-Pion versus CC-0-Pion. The inclusive selection will first be presented and then broken down into exclusive channels.

The target detector, FGD1, and its fiducial volume are shown in Figure 5.1. The most upstream XY layer of of scintillator bars is excluded, and on the x and y faces, five scintillator bars at each end are excluded.

The CC inclusive selection cuts are:

Period	Data POT	MC POT	Sand POT
Run 1 - Water In	1.66891×10^{19}	9.985×10^{20}	9.49×10^{20}
Run 2 - Water In	4.2858×10^{19}	1.1305×10^{21}	9.49×10^{20}
Run 2 - Water Out	3.55196×10^{19}	1.0025×10^{21}	9.49×10^{20}
Run 3b - Water Out	2.14598×10^{19}	3.0×10^{20}	9.49×10^{20}
Run 3c - Water Out	1.34821×10^{20}	1.5×10^{21}	9.49×10^{20}
Run 4 - Water In	1.62488×10^{20}	3.0×10^{21}	9.49×10^{20}
Run 4 - Water Out	1.76246×10^{20}	1.6135×10^{21}	9.49×10^{20}
Total	5.90082×10^{20}	9.545×10^{21}	6.643×10^{21}

Table 5.1: Number of POT for each data set after data quality checks, the beam MC and sand muon MC. About an order of magnitude more MC than data is created. The "water in" and "water out" names refer to the PØD status, which has water targets which can be filled and drained.

1. **Data quality flag:** Check the global ND280 and beamline quality to ensure the event comes from a good spill during nominal detector operation.

2. **Event Bunching:** Each beam spill is composed of 8 bunches. Events are binned into their associated bunches to avoid "pile up." For example, if two clean CC events occur in different bunches, without this step the spill would see a single event producing two muons from two vertices, and, most likely, discard it. Event pile up will be a larger issue for nominal beam power, but at 200 kW about 1 event per spill is seen on average.

3. **TPC Candidate:** Require at least one reconstructed TPC track to be present. At this time there is no FGD1-only CC analysis, so without a muon candidate entering the TPC the event cannot be considered.

4. **FGD-TPC Candidate & TPC Quality:** Extending the previous cut, every event has to match at least one TPC muon candidate to an associated FGD1 track which has a vertex within the FGD1 fiducial volume. Additional comments on the FGD1

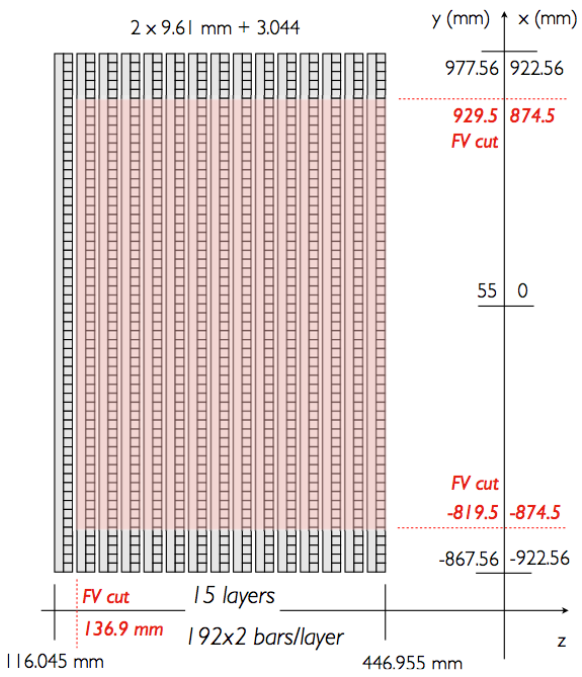

Figure 5.1: The target and fiducial volume used in FGD1 used in this event selection [32]. The asymmetry in the y coordinate is due to the detector coordinates being centered relative to the magnet, which the FGD is 55 mm above. The beam comes from the left in this figure, with TPC1 being at the left as a veto and TPC2 being flush at the right and used for additional tracking.

fiducial volume will be given. An additional quality cut on the TPC track is made here, rejecting very short tracks which have poor reconstruction.

5. **Upstream Background:** Events with particles entering FGD1 from upstream (typically sand muons, PØD or magnet events) are removed. This is achieved by vetoing on TPC1 activity. As particles can also enter the FGD from the side (skipping TPC1), the highest momentum track that is not the candidate muon for the CC event must have a starting position compatible with the muon candidate starting position.

6. **Broken Tracks:** Events with particles that enter FGD1 from the upstream region of the tracker can result in reconstructed tracks being "broken," i.e. reconstructed as two

separate tracks. Events with a muon candidate start position near the downstream end of the FGD, and a second track entirely contained in the FGD but with a start position outside the fiducial volume, are rejected. A mock-up of a broken track event is shown in Figure 5.2.

7. **Muon PID:** The μ^- candidate track is passed to the TPC particle identification (PID). The PID is based on likelihoods of dE/dx for given particle hypotheses.

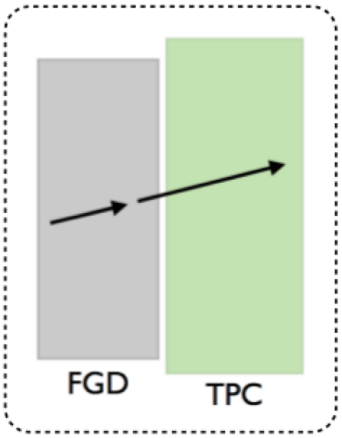

Figure 5.2: An cartoon illustration of a broken track event [58]. An background event enters FGD1, and "breaks" in reconstruction. The second half of the track is taken as a CC muon candidate starting in the fiducial volume of FGD1, leaving behind the first segment as a track entirely contained in FGD1.

5.2.1 Exclusive Channels

With the CC inclusive selection complete, the full selection is carried out by counting the number of observed pions. This is achieved through three primary methods:

- **TPC PID**: The TPC can be used to identify charged pions, muons and electrons and positrons by looking comparing dE/dx measurements. The presence of electrons and positrons are used to infer the existence of neutral pions. For a valid PID to be made,

the secondary tracks must be in the same timing bunch as the candidate CC muon, start within the FGD fiducial volume, and pass an overall TPC track quality cut (i.e. not too short such that reconstruction is unreliable).

- **FGD-Only Tracks:** Tracks that are entirely contained within the FGD fiducial volume are called FGD-only tracks. In this selection an event that has no michel electron a search for FGD-only tracks is performed. Due to reconstruction issues with broken tracks, one and only one pion candidate is counted (in the case of allowing multiple tracks, an out-of-fiducial background could enter, break, and result in at least one good candidate track).

- **FGD Michel Electrons:** Michel Electrons (ME), electrons which come from muon decay, can be used to tag low energy pions via the two-step decay: $\pi^+ \rightarrow \mu^+ + \nu_\mu$, $\mu^+ \rightarrow e^+ + \nu_\mu + \nu_e$ (similarly for π^-. To tag a ME, a cluster of hits not in the same timing bunch of the muon candidate and of minimum threshold energy is searched for.

With these pion counting capabilities the CC inclusive sample is broken down by the following sub-selections:

- **CC-0-pi:** A negative muon candidate with zero charged or neutral pions (inferred by TPC electrons/positrons) observed. Events with FGD-only tracks or michel electrons are rejected.

- **CC-1-pi:** A negative muon candidate with only one positively charged pion. Events with neutral or negative pions reconstructed in the TPC are rejected. FGD michel electrons and FGD-only tracks can contribute to this category, under the condition that the sum of the number of FGD michel electrons, FGD-only pion tracks, and TPC positive pions must be equal to one.

- **CC-Other:** Events that enter this category can have one or more positrons or electrons in the TPC (i.e. a neutral pion), one or more negative pions in the TPC, or

more than one positive pion (again, using the sum of the three positive pion sources: FGD-only tracks, FGD michel electrons, and TPC tracks).

These sub-selections are summarized in Figure 5.3.

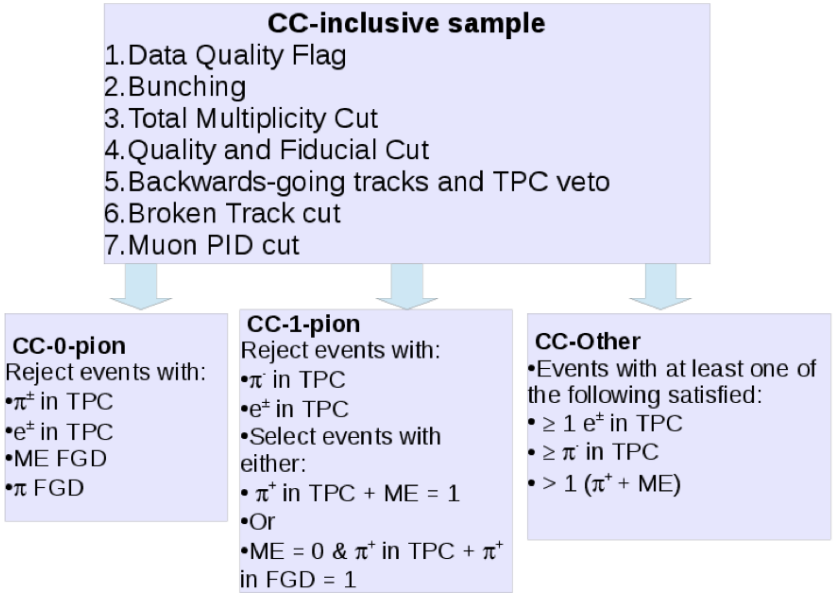

Figure 5.3: CC exclusive channel topologies broken down by selection criterion [58].

It is important to note that these categories are *observed* topologies. The neutrino interaction, as we understand it today, occurs on an individual nucleon within a nucleus (except for coherent interactions which interact with the entire nucleus). This point of interaction is usually referred to as the interaction "vertex." However, due to final state interactions, the primary particles produced in an interaction might differ from the final state topology, i.e. the distribution and type of particles that exit the nucleus. A simple example of this is true CC-multi-pion interaction in which one pion is re-absorbed in the nucleus. The final state topology in this case is that of CC-1-pion. The neutrino interaction generator, and our neutrino interaction theories, discuss true interaction types at the vertex. Our experiments measure the observed topology. We have models and attempt to understand

the so-called Final State Interactions (FSI) which transform the interaction topology at the interaction vertex to an observed topology which may differ.

When working with simulated data we can access the truth information of an event and determine the interaction type at the vertex, the initial state particles, and the post-FSI final state topology. This categorization of event types is cumbersome, so I will re-iterate their definitions.

- **Event Reaction/Mode:** Sometimes referred to as the reaction, or the "mode" of an event is the type of neutrino interaction as given by the simulation model. These modes are CC-Resonant, CC Deep Inelastic Scattering (DIS), neutral current (NC) and so on. At the vertex an outgoing set of particles is produced.

- **True Topology:** After the particles produced at the interaction vertex are propagated through the nuclear medium of the target nucleus and have undergone final state interactions, the modified, nucleus-exiting topology of particles defines the true final state topology.

- **Observed Topology:** The observed topology, when dealing with simulated data is defined as the result of detector simulation and analysis cuts. In the case of real data, it is simply our observed result after analysis cuts.

Due to detector inefficiencies and resolution, it is possible that *true* topologies of one sort end up in event selection categories of a different *observed* topology. Again, as a simple example, a CC-1-pion event where the pion reconstruction fails will be categorized as a CC-0-pion event. The movement of events of a given true topology between different observed topologies is referred to as *event migration*. The primary types of event migration can be summarized as follows:

- **CC-0-Pion Mis-classification:** The largest contributor to this category is CC-1-pion events where the charged pion is missed. The pion can be missed due to secondary interactions in the FGD making the reconstruction of the track difficult. Other possibilities are the michel electron being missed by reconstruction, or the TPC

PID failing to identify the charged pion. Pions that are sideways-going and exit the FGD into the surrounding ECal can also be missed.

- **CC-1-Pion Mis-classification:** As in the CC-0-pion mis-classification, events from the CC-other category where pions are missed end up in the CC-1-pion category. In addition, the CC-1-pion selection incurs a background from CC-other events where an electron or positron from a π^0 is missed in TPC.

  CC-0-pion events that have spurious reconstruction of short proton or muon tracks, or have michel electron clusters from near-by ν-magnet interactions, also end up migrating into the CC-1-Pion category.

An example demonstrating how an MC selection can be represented by true event topology or true neutrino interaction mode is given in Figure 5.4, which shows the CC inclusive selection in two different formats.

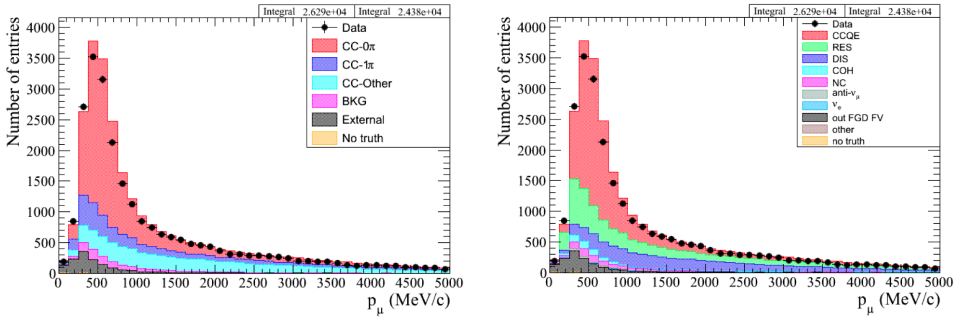

Figure 5.4: The same CC inclusive selection shown in two different ways: true event topology on the left, and true interaction mode on the right. Due to final state interactions and detector efficiencies and resolution, different combinations of topologies and modes will enter observed sub-selections [58].

5.3 Selection Summaries

A detailed summary of the selection efficiencies and purities will be given here. First a summary of the CC-inclusive selection will be given. The analysis presented in this thesis will

focus on the CC-1-pion and the CC-other samples, so additional details on the composition of those samples will be presented. The MC event numbers presented are after normalization to the data POT.

5.3.1 CC-Inclusive Selection Summary

The CC-inclusive selection, which is based primarily on identifying the highest momentum negative muon track, has a muon purity of about 90%. The purity is defined as the number of selected events in an observed topology that are of that true topology. The dominant background is due to π^- contamination. The muon candidate shown by true particle type is shown in Table 5.2. The total number of events (data and MC) at each cut level, and the sample reduction is shown in Table 5.3. The fractional contribution of each *true* event topology is shown in Table 5.4.

True particle	Fraction (%)
μ^-	89.8
μ^+	0.36
e^-	0.44
e^+	0.08
π^-	7.5
π^+	1.3
p	0.44
other	0.1

Table 5.2: True particle type of the CC-inclusive muon candidate, which is about 90% muons with a primary background doing from π^-[58].

The CC-sample can be broken down into sub-selections, which we will discuss in the next sections. The relative fraction of the total CC selection given by sample, purity and efficiency of those samples is shown in Table 5.5. The efficiency is defined as the number of true selected events for a given topology, divided by the number of true events of that

Cut Level	Composition (%)	Sample Reduction (%)	Data Events	MC Events
TPC Candidate	2.27	100	1996217	108972.33
FGD-TPC Match & Quality	55.73	43.8	48731	47751.9
Upstream Background	68.7	34.7	35942	37786.1
Broken Track	70.2	33.8	34804	36833.2
Muon PID	90.72	24.8	25917	27082.1

Table 5.3: Composition of CC-inclusive selection, relative to true CC events, with sample size and data events per cut level.

topology in the FGD1 fiducial volume.

5.3.2 CC-1-Pion Selection Summary

The CC-1-pion selection is difficult to isolate mostly due to CC-other contamination. Any detector efficiency impacting the identification of pions will contribute to the likelihood that a multi-pion event which belongs in CC-other may end up in the CC-1-pion selection. The composition of the CC-1-pion sample by neutrino interaction type is given in Table 5.6.

As described in Section 5.2.1, the pion in the CC-1-pion event can be identified by the presence of a FGD michel electron, an FGD-only pion track, or an FGD-TPC matched track. The different methods of pion identification have different purities. The purity of the CC-1-pion selection, and the different selection methods is given in Table 5.7.

The number of data and MC events per pion identification method is given in Table 5.8.

5.3.3 CC-Other Selection Summary

The CC-Other selection is a mostly pure sample. The sample breakdown relative to true event topology is shown in Table 5.9. The composition of the CC-other sample by interaction mode is given in Table 5.10. The number of data and MC events is given in Table 5.11.

CC-Inclusive Sample	Fractional Composition (%)
True CC-0-Pion	51.48
True CC-1-Pion	15.01
True CC-Other	24.24
Background	4.03
Out of FGD 1 FV	5.23

Table 5.4: CC-inclusive selection broken down by *true* topology and backgrounds. The background is defined as anti-neutrino and neutral current events. There is about a 5% out-of-fiducial background [58].

Sample	Efficiency (%)	Purity (%)	Fraction of CC Selection(%)
Observed CC-0-Pion	47.81	72.43	68.4
Observed CC-1-Pion	28.37	49.24	16
Observed CC-Other	29.71	73.6	15.5

Table 5.5: Sub-sample selections with their efficiencies and purities, and what fraction of the total CC-inclusive selection they comprise.

5.4 Systematics

The systematics associated with the CC-multipion selection are pictorially summarized in Figure 5.5. A brief description of each systematic will be given, with particular focus on the so-called "hybrid" analysis, which was developed by the author of this thesis. The evaluation of the systematics relies on "control samples," namely real data samples with very well known properties. Examples are cosmic rays, which have a dedicated trigger which reliably selects MIP-like particles. Through-going "sand muons" are also used, or events that begin in the PØD and continue on into the tracker.

CC-1-Pion Sample	Composition (%)
CCQE	5.3
Resonant	39.4
DIS	31.3
Coherent	10.6
Neutral Current	4.7
$\overline{\nu}_\mu$	1.7
ν_e	0.4
External	6.6
other	0.04

Table 5.6: The CC-1-pion sample broken down by neutrino interaction mode and background events. Note the two dominant contributions are resonant pion production (signal), and then Deep Inelastic Scattering, which may contribute to both the signal via nuclear re-interactions that produce a true CC-1-pion topology, or to background due to detector effects [58].

5.4.1 TPC Calibration

The official data processing and calibration phase adjusts for TPC gas densities and temperature changes. However, some data/MC differences still remain. A calibration factor based on the dE/dx measurement of muons as a function of time over data taking periods is applied. The control sample used is a sample of muons between 350 and 550 MeV/c that start in the PØD and cross all three TPCs.

5.4.2 TPC PID Systematics

A TPC PID systematic for muons is computed from the difference between data and MC. The control sample used is a selection of sand muons crossing the TPCs. The difference in the mean muon PID likelihood as a function of momentum is compared between data and MC. Similarly, the width of the distribution of PID likelihood as a function of momentum is also computed.

CC-1-Pion Sample	Total (%)	FGD-TPC Track (%)	FGD ME (%)	FGD-Only Track(%)
True CC-0-Pion	6.4	4.1	4.8	26.3
True CC-1-Pion (Purity)	49.2	50.3	51.4	34.4
True CC-Other	31.0	34.1	26.6	27.6
Background	6.8	8.8	4.0	4.7
External	6.6	2.7	13.1	7.0

Table 5.7: Composition of the CC-1-pion broken down by true topology type and for each method of identifying the pion. The row corresponding to the true CC-1-pion topology gives the purity. Note the differences in samples: for example, the michel electron (ME) method is subject to a large external contaminations, the FGD-only sample has a large CC-0-pion contamination (i.e. CCQE protons mis-ID'ed as pions).

Pion ID Method	MC Events	Data Events
FGD-TPC Matched Track	2532.1	2316
FGD Michel Electron	1491.6	1420
FGD-Only Track	399.1	349
Total	4422.8	4085

Table 5.8: Selection of CC-1-pion events based on different pion identification methods.

A high-purity proton sample is used to perform an identical analysis for the TPCs' proton PID. The proton control sample is made by selecting the highest momentum positive tracks starting from a vertex in the FG1 fiducial volume that have momentum between 300 MeV/c and 1.1 GeV/c.

Differences in the PID pull central values and widths, as a function of momentum, are corrected for bias and scale by examining the data-MC differences.

CC-Other Sample	Composition (%)
True CC-0-Pion	5.8
True CC-1-Pion	7.8
True CC-Other (Purity)	73.6
Background	8.7
External	4.1

Table 5.9: Composition of the CC-other sample as given by true interaction topology.

5.4.3 FGD-only Track Efficiency

This systematic was designed to assess the efficiency of the FGD-only reconstruction. This is a reconstruction process that examines only FGD hits and tries to reconstruct a FGD-contained track. The control sample used was a selection of stopping protons that pass through TPC1 and into FGD1. The TPC1 information is used to establish a known direction and momentum prior to entering FGD1.

5.4.4 FGD-only Track PID

The FGD PID functions comparing the measured energy deposition along a track to the expected energy deposition for a given particle hypothesis. This systematic examined the FGD PID by selecting a control sample of stopping muons and protons that crossed TPC1 and entered FGD1. The muon sample is used to model the pion systematic.

5.4.5 FGD-only Track Hybrid Efficiency

Hybrid, in the context of analyses, is the nomenclature used to describe an event or collection of events that result from a process of combining physics information from multiple, independent sources. In this case, the two sources are real data events and MC events. The hybrid analysis in the FGD that is described here involves the combination of energy deposition information (from here on referred to as a "hit") from individual, single-particle

CC-Other Sample	Composition (%)
CCQE	3.9
Resonant	14.2
DIS	67.7
Coherent	1.4
Neutral Current	6.8
$\overline{\nu}_\mu$	0.9
ν_e	0.9
External	4.1
other	0.2

Table 5.10: Composition for the CC-Other sample, according to the generator interaction types.

Sample	MC Events	Data Events
CC-Other	4332.36	4297

Table 5.11: The number of MC (scaled to data) and data events for the CC-Other selection.

event simulations, into beam CC-0-pion vertices selected from official data and Monte Carlo productions.

This hybrid analysis was designed to examine the FGD-only track reconstruction efficiency in the presence of at least one long muon track and an additional, optional long proton track. FGD reconstruction efficiencies for isolated tracks have been performed in the past, but no study that incorporates the impact on FGD-only Track efficiency due to the presence of other tracks has been performed until now.

The outline of the hybridization sequence is as follows:

1) Vertex Skimming Select from official data and MC productions a set of CC-0-pion vertices.

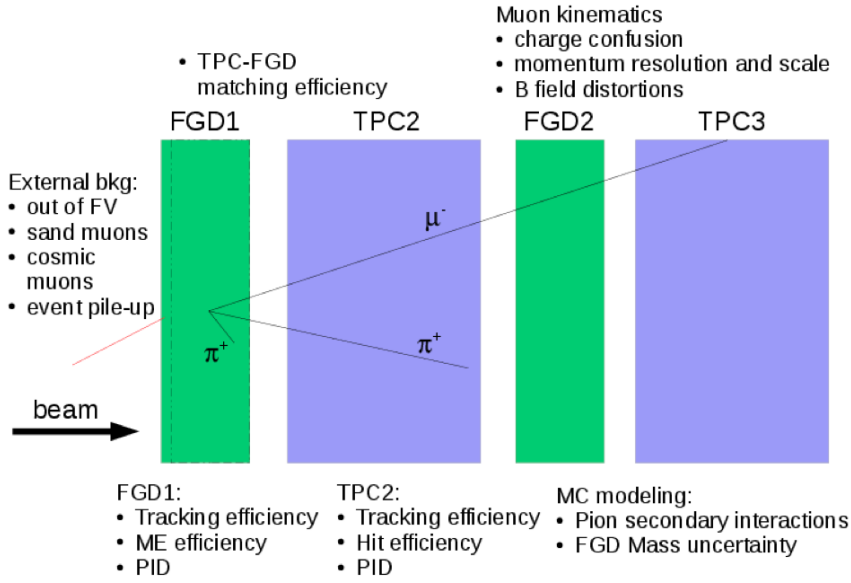

Figure 5.5: A pictographic summary of the systematic studies associated with this selection [58].

2) Candidate Vertex Selection From the event list a selection based on the reconstruction-level files was performed. The candidates vertices are processed to reject events that have FGD-only results. This step produced a set of candidate CC-0-pion vertices with only a long muon, or a long muon and a long proton track but ensured that no event had any FGD isolated track results.

3) Particle Creation At each of the data and MC vertices selected in the previous step, single-particle events were created. For each vertex 100 proton and100 positive pion events, uniform in kinetic energy and isotropic in angle were created. For the proton sample, kinetic energies that correspond to a momenta range of 400-1000 MeV/c were used, with the lower bound being selected by the FGD-only efficiency having a sharp activation at about 400 MeV/c for protons. For the pion sample, kinetic energies that correspond to a momenta range of 100-300 MeV/c were used.

4) Single-Particle Event Selection For each particle event generated a selection program is run to ensure that each individual event does not result in a FGD-TPC matched

track. The goal of this step is to ensure that a sample of single-particle events with hits that remain entirely contained in the FGD is selected. The check rejects events that have any FGD-TPC match results, but does not require an FGD-only result. In other words a short track that deposits hits in the FGD but does not successfully reconstruct is accepted.

5) Hybridization For each set of 100 particle events generated at a given candidate beam event vertex, 100 hybrid events are created (assuming all 100 events pass the non-matching requirements of step 4). The hybridization process is carried out by extracting the hit information from the single-particle event and adding it to the associated beam event.

The above steps generate four hybrid samples, namely CCQE data and MC events each of which are separately proton-hybridized and pion-hybridized. After the hybrid samples were created they were re-processed with the FGD reconstruction software and the efficiencies examined.

The results of the efficiency study for the proton- and pion-hybridized samples are shown in Figures 5.7 and 5.8, respectively. The straight line fit in the data/MC ratio is used to evaluate the size of the systematic for each sample.

The primary reason for the decrease in overall efficiency is suspected to be the "absorption" of inserted hits into the pre-existing track(s) during FGD-TPC matching and track finding, i.e. short tracks that are aligned with the muon track will have a majority of their energy deposition attributed to the muon track and the remaining hit information is difficult to reconstruct into a short isolated track. This is demonstrated in Figure 5.6, which demonstrates the fraction of short track hits "stolen" by the muon track for failed reconstruction events as a function of proton momentum and angle between the two tracks.

5.4.6 FGD Mass Uncertainty

An analysis of the density of the elemental components of the XY modules that make up the scintillator target of the FGDs was performed [59]. The conclusion was that the FGD module mass is known to 0.67%.

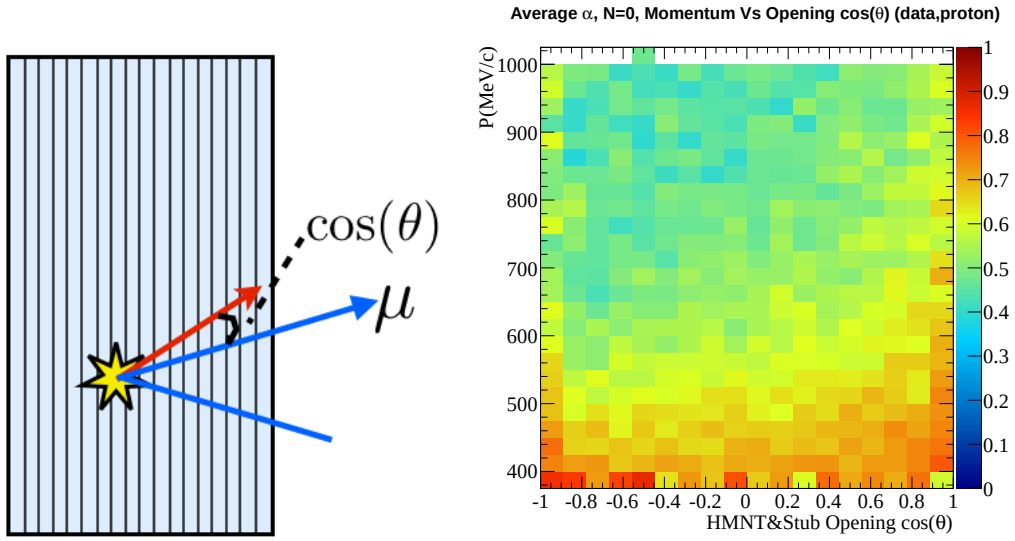

Figure 5.6: The cartoon sketch up on the left shows an inserted set of hits forming a short "stub," which have some opening angle θ with respect to the original event's CC μ^- candidate (highest momentum negative track, HMNT). On the right, the parameter α is the fraction of inserted hits that were absorbed into the muon track's hits for failed reconstructions. For these events you can see a trend for low-energy proton tracks failing, in addition to an increasing "theft" of hits as the angle between the tracks decreases.

5.4.7 Michel Electron Efficiency

This analysis used FGD-triggered cosmic events to examine the Michel electron (ME) detection efficiency differences between data and MC. The FGD cosmic trigger functions by looking for correlated hits between FGD1 and FGD2, and therefore selects cosmics that are roughly aligned with the tracker z-axis. This analysis exploits this feature of the trigger and requires a cosmic trigger to have a muon in TPC2 (between the two FGDs), stopping in FGD1 (with the extra veto on activity in TPC1). The ME detection efficiency ratio between data is thus assessed as a function of data taking time. The MC ME efficiency is $61.9 \pm 1.10\%$, and a typical data ME efficiency is $58.1 \pm 0.8\%$. The systematic assigned to this efficiency is $\sim 4\%$.

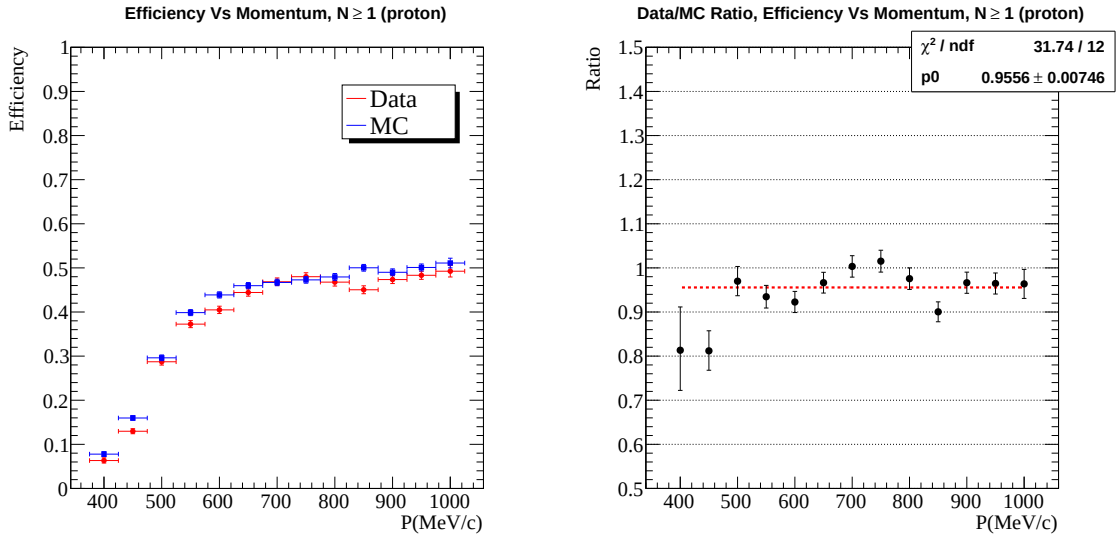

Figure 5.7: Proton-Hybridized Efficiency Overlay and Ratio: On the left the efficiency verse momentum for the proton-hybridized sample is shown for data and MC. On the right is the data/MC ratio.

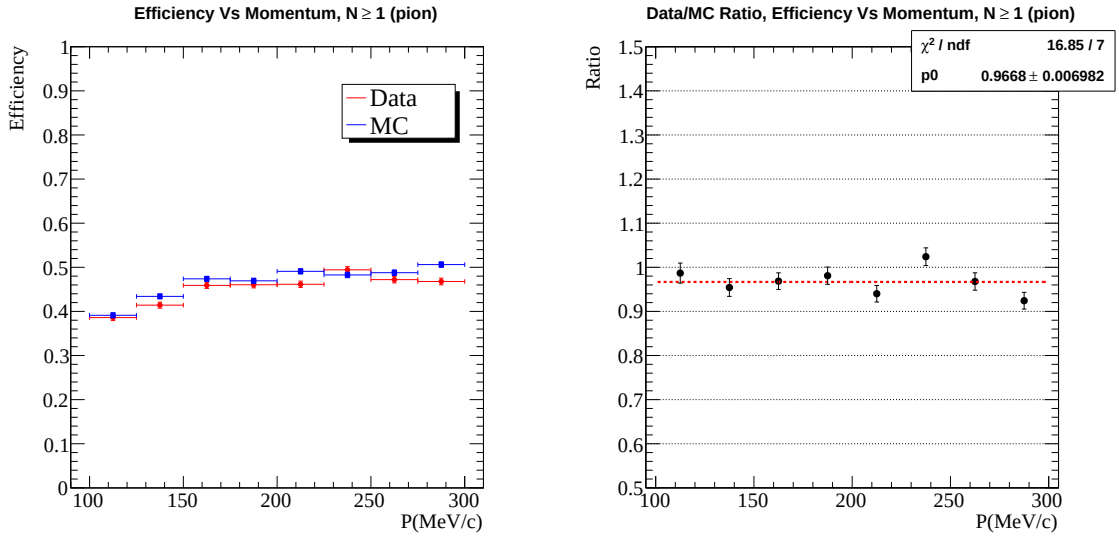

Figure 5.8: Pion-Hybridized Efficiency Overlay and Ratio: On the left the efficiency verse momentum for the pion-hybridized sample is shown for data and MC. On the right is the data/MC ratio.

5.4.8 Out-of-Fiducial-Volume Events

This systematic aims to address the problem of neutrino interactions inside the ND280 detector suite but outside the FGD1 fiducial volume. The core concept of this analysis is that OOFV events are reconstruction failures, so the approach taken is to study the most common failure modes of reconstruction and assign uncertainties to the rates. Examples of some categories are: Events attributed to neutral particles entering the FGD and creating a π^- mistaken for a muon candidate, or backwards-going π^+ from the barrel e-cal reconstructed as forward-going μ^-. In total there is about a 5% OOFV background.

5.4.9 TPC Cluster Efficiency

The TPC functions by drifting ionized electrons onto a readout plane that is segmented into "pads," and a collection of pads that have energy deposition above a threshold are considered a "cluster." A cluster efficiency difference between data and MC will cause a systematic shift in the CC-multipion selection as the muon candidate TPC track has to pass a TPC quality cut, which is defined as having a minimum number of clusters. The MC was found to have a $1 \sim 3\%$ larger efficiency depending on the readout pad location.

5.4.10 TPC Single and Double Track-finding Efficiency

A separate analysis to study the TPC track-finding efficiency was carried out, specifically for the case of single-track events and double-track events. The TPCs, for single tracks, were found to have reconstruction efficiency statistically consistent with 100% for tracks with 16 hit pads or more, and dropped only to 97% for tracks of 10 hit pads.

A similar result was found for the double-track systematic studies, namely that if two overlapping tracks had more than 18 hit pads (the basic TPC quality cut), the reconstruction efficiency for both MC and data were found to be 100%.

5.4.11 TPC Momentum Resolution

The TPC momentum resolution is examined with a control sample of through-going muons from beam spill events with a high-quality, negative track starting within ND280 and muon-

like. By examining the momentum measurement of the same track at different TPCs (while accounting for the energy loss in the FGDs), it was found that the MC events had smaller momentum resolution than data events. The MC resolution is smeared as a function of the transverse momentum measured by the TPC, P_t, and is on the order of 10% to 30% depending on the momentum range.

5.4.12 TPC Charge Confusion

Charge confusion is the mis-identification of a track charge in the TPC, which determines charge based on curvature and direction (direction meaning either "forward-going" or "backward-going"). A study examining the data and MC events with tracks crossing in all TPCs examined, checking the consistency of the charge identification across the three detectors. It was found that below 5 GeV/c, charge confusion difference between data and MC is less than 1%, but for higher momentum tracks the difference is 3.10%±3.1%. Very high energy tracks are hard to charge ID because of their minimal curvature, from which the charge is inferred.

5.4.13 TPC Momentum Scale

The TPC momentum scale is set by the B-field measurement mapping uncertainties, carried out in 2009 [60]. The results of that mapping gives a momentum scale uncertainty of 0.5%. To check this one study done evaluating the momentum-by-range measurement in the FGD1 was done using stopping FGD-cosmic rays. The TPC2 momentum and direction measure was used to predict the range of the particle entering FGD1, and this range was compared with hypothesized variations in momentum on the order of ±5%. The variations were found to strongly disagree with the data and any momentum bias was found to be disfavored by this study.

5.4.14 Event Pile Up

Event "pile up" is the case in which tracks from separate events end up in the same event record and cause a mis-reconstruction. Usually this is caused by some sort of upstream

interaction that produces tracks entering the downstream tracker. The only significant contributor to pile up is due to sand muons. Typically these sorts of events are rejected based on TPC1 activity. However, as sand muons are not included in the official beam MC productions, the pile up effect is not accounted for.

A separate sand muon MC was produced and the rate of TPC1 vetoes was estimated and then applied as a weight to the official MC. The official beam MC is thus reweighted to be slightly *decreased* as we are accounting for an effect that would rejects if it had been accounted for.

The pileup correction is directly related to sand muons, and the sand muon rate is a function of the beam intensity. As the beam intensity varied over different runs, the pileup correction is unique to each run period. Larger beam intensities will have a larger correction as more sand muons are produced. Furthermore, uncertainty in the surrounding materials and sand muon simulation contribute to a systematic uncertainty on this correction. Run 1, the lowest intensity run, is corrected by $C = 0.51 \pm 0.09\%$, and Run 4, the highest intensity run, is corrected by $C = 1.3 \pm 0.13\%$. The MC weight per run is applied as $(1 - C)$.

5.4.15 Cosmic Ray Background

A dedicated sample of cosmic ray MC is generated for every official beam MC period. The cosmic MC is used to perform studies of event rates of cosmics that pass the CC inclusive cuts. The frequency of MC events passing the CC cuts is very small, and due to the 120 ns timing window per bunch, the integrated time of good spill data is only 6.9 s. These studies found that after data/MC corrections, the total number of expected cosmic events passing these cuts was less than 1 event for the entire data taking period, and thus this systematic is ignored.

A false TPC1 study was also performed using "empty spill" data. Empty spill data corresponds to times where full beam data taking was underway but there was no beam operation, so full timing and conditions are identical. The TPC1 cosmic rate with reconstructed TPC1 components was found to be 458 ± 28 Hz, but given the bunch timing window the probability of such events during beam is less than 0.01%.

5.4.16 Sand Muon Background

The same sand muon sample discussed in the event pile up study is used to evaluate a background by examining events that pass the CC-inclusive and pion multiplicity cuts for each CC sub-sample. The number of sand muon events relative to the number selected in the CC, CC-0-pion, and CC-1-pion samples was found to be about 0.4% for each sub-sample. The CC-other sample had a contamination of about 0.1%.

The systematic uncertainty on this correction stems from the fact that in the data, neutrino interactions occur in the PØD casing, magnet and surrounding structure of ND280. These events are selected in magnet interaction MC (which accounts for the interactions in the magnet and the structure around ND280, but does not extend to the pit walls and surrounding sand), and added to the sand muon rate to estimate the total sand muon-like contributions that should be comparable to data. A 10% discrepancy is found and included as a systematic uncertainty.

5.4.17 Pion Secondary Interactions

Pion secondary interactions are re-interactions of pions outside of the nucleus in which they were produced. This study accounted for differences between the Geant4 models and differences in external data. The secondary interactions covered were pion absorption, charge exchange, and quasi-elastic scattering. Absorption results in a pion being completely absorbed by a nucleus. Charge exchange occurs when a charged pion interacts with a nucleus and produces a π^0. Quasi-elastic scattering occurs when the pion interacts inelastically with the nucleus, producing a same-charge pion and nuclear remnants.

Any effects that modify the pion multiplicity will impact the CC-multipion event categorization. Pion absorption and charge exchange will modify the final state topology, and scattering can cause the pion to change direction and momentum in ways that make reconstruction less reliable. To account for these effects, two sets of event weights are generated in this study, one to adjust the MC to match the external data, and one to allow for variations based on uncertainty in the external data. Additional details of this study are given in [61].

5.4.18 TPC-FGD Matching

To evaluate the TPC-FGD matching efficiency a set of through-going muons was used, with the requirement that long tracks be found in TPC1 and TPC2, implying that a track passed entirely through FGD1 (which is between TPC1 and TPC2).

The matching efficiency of a TPC track to some part of the FGD hits is 99.9% for particles above 200 MeV/c, 98.9% down to 100 MeV/c, and parameterized as $(30\% + 68.8(p/100))$ for momentum $p \leq 100$ MeV/c. The systematic uncertainty on these three regimes is taken as 0.2%, 0.4%, and 15%, respectively.

A second study which evaluates the matching efficiency of the TPC track to the *entire* set of expected FGD hits was also performed. This efficiency is important for reweighting the OOFV events (only $\sim 5\%$ of CC inclusive), which enter the FGD but due to reconstruction failures a few hits are missed and it appears that the track starts *within* the FGD FV. This study evaluates the same through-going muons to examine the FGD matching is, and also evaluates cosmic events. More details can be found in [62].

5.4.19 TPC Field Distortions

TPC field distortions can cause variations in the reconstructed momentum, which manifests itself in the migration of events from one momentum bin to another. A treatment of the magnetic field distortions in the TPC region was performed [63], which had two main components: a field correction, and a distortion correction.

The field correction comes from analysis of the magnetic field map which was performed prior to installation of the ND280 detectors [60]. The distortion correction is based on measurements of the TPC laser system which flashes aluminum targets on the central cathode. The illuminated targets produce ionization patterns from known configurations and allows correction for distortions in the magnetic and electric fields. The field correction alone is used in the reconstruction of the data, and the size of the distortion correction defines the systematic error.

5.5 Summary

This chapter has summarized the selection of a CC inclusive sample using the FGD as a target volume and the downstream tracker detectors. This sample was broken down into sub-categories based on pion multiplicities, giving exclusive selections of the CC-0-pion, CC-1-pion and CC-other samples.

A comprehensive list of detector and selection systematics was summarized. The implementation of these systematics will be discussed in more detail in the fit description.

Chapter 6

PION MULTIPLICITY UNCERTAINTY

This chapter will describe the first attempts at characterizing a new systematic uncertainty that up until this point has not been considered: The uncertainty on the number of charged pions produced in a neutrino interaction. The concept behind this uncertainty is neatly indicated by the results in Figure 6.1. The y-axis in each plot is the mean number of charged particles, and data presented is from different experiments on different targets. The details of these plots will be discussed in this chapter, but the qualitative behavior is that as you increase the invariant mass of the interaction, W, you have more charged hadrons. That's not exactly surprising, but with regard to error on the number of produced particles the point is this: there is significant spread in the measured mean charged particle production. This is given by the spread in y-values for all the different measurements.

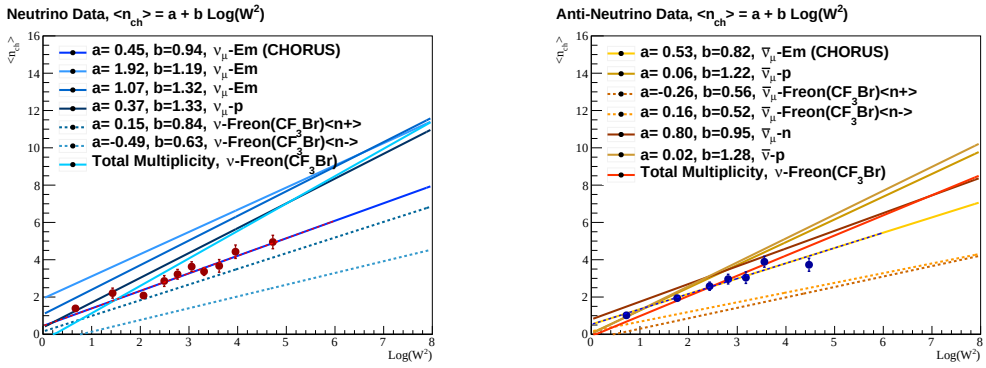

Figure 6.1: Plots created from published data on various charged hadron multiplicity measurements on different targets. The data points presented here come from the CHORUS experiment, which also published the table of aggregated data upon which these plots are based [64]. Neutrino interactions are on the left, and anti-neutrino interactions on the right, with data fit to the empirical relation $\langle n_{ch} \rangle = a + b \log W^2$.

The approach taken to assess this uncertainty was to examine bubble chamber data

from the Fermilab bubble chamber experiment [65] on a deuterium target. From this data a charged hadron model was created and tuned to the data. Once the model was created the variation of the effect of the variation of the model parameters was possible and the impact on the ND280 detector was assessed.

The rest of this chapter will involve some comments on motivation for this study and then present the analyzed external data. A description of the charged hadron model developed, and the tuning of this model will also be presented.

6.1 Motivation

The necessity of assessing all known systematic errors is taken as a principle, validating any efforts to carry out the necessary work. Therefore, the fact that we have a natural process, charged hadron production, with an inherent uncertainty on the number of charged hadrons produced is, alone, a valid argument to necessitate the characterization of this uncertainty. However, there are a few additional noteworthy points that support the efforts presented in this chapter.

6.1.1 MiniBooNE Results

The results from MiniBooNE have caused great interest in the measurement of the CC-1-pion cross section. Due to the noteworthy discrepancy between data and theory the community is in need of new results of this measurement. Shown in Figure 6.2 is a study demonstrating the MiniBooNE results compared with a neutrino interaction generator called GiBUU [66]. The study demonstrates that although there is some agreement between the model and the data, this is only the case when final state interactions are not enabled. Turning on final state interactions results in a significant discrepancy.

The CC-1-pion interaction is under close scrutiny, therefore fully addressing the systematic uncertainty on this process is important.

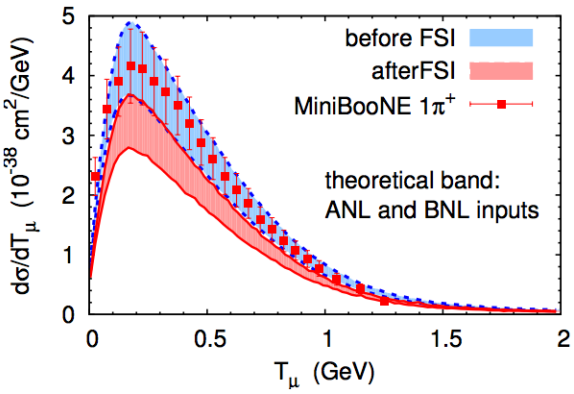

Figure 6.2: Data and theory comparison of MiniBooNE results[67].

6.1.2 ND280's CC Tracker Sample

The measurement of interest is the CC single-pion interaction's pion momentum spectrum. At T2K's near detector complex a CC multipion sample was measured in the tracker, with the scintillator of FGD1 as the target. This event selection produces three main categories of events: CC 0-pion (also known as "CCQE-like"), CC 1-pion, and CC multi-pion (also referred to as "CC-Other"). The details of this event selection were discussed in Chapter 5.

As the name suggests, the CC 1-Pion sample has a single positively charged pion, and it is the momentum of this pion that is of particular interest. However, the method of identifying a single-pion event is threefold: The pion can be inferred by the existence of a Michel Electron in FGD1, the pion can be measured as a short but fully-contained track within the scintillator volume of FGD1, or, finally the pion can be detected as an FGD1-TPC1 matched track with a momentum measurement being made in TPC1. These modes of identification are depicted in figure 6.3.

With the interest of having a well-measured momentum, an analysis that uses only events from the FGD1-TPC1 matched event categories is being developed, and thus an understanding of the dominant backgrounds in this subsample is necessary. The primary background turns out to be contamination from the miscategorization of CC-other events. For the FGD1-TPC1 matched events, the composition of event topologies is shown in Table

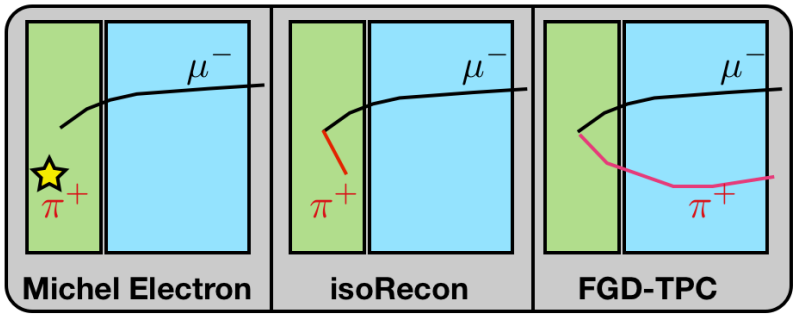

Figure 6.3: Different modes of CC 1-pion event identification.

5.7.

In figures 6.4, 6.5, and 6.6, the breakdown of the CC 1-Pion pion momentum is shown with respect to true particle type of the positively charged pion candidate, true topology, and NEUT interaction mode of the event, respectively.

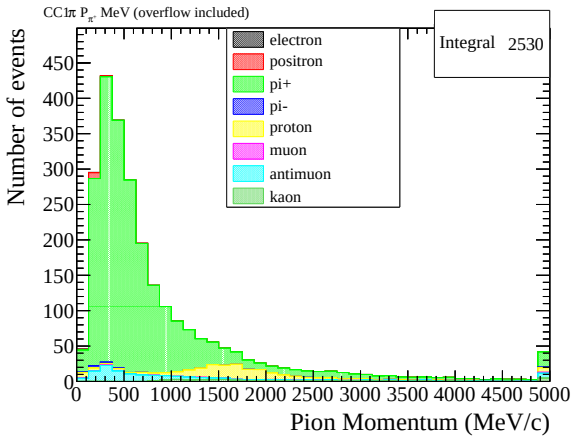

Figure 6.4: CC-1-Pion, FGD1-TPC1 matched sample, charged pion candidate momentum categorized by true particle type.

Qualitatively, the most noteworthy aspects of these three figures is that the the signal CC-1-pion selection is good at selecting true charged pions as the candidate, but that there is still a large background. The background comes from the CC-other category, which is

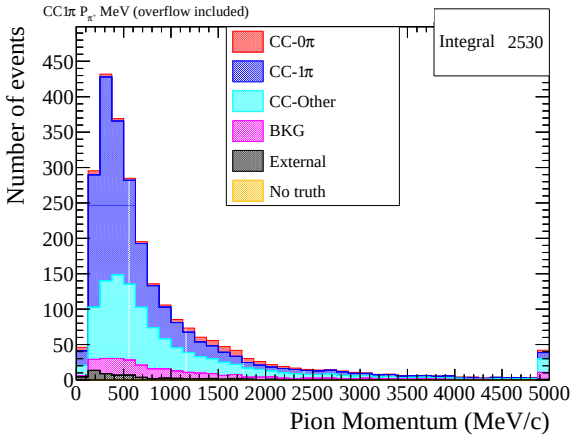

Figure 6.5: CC-1-Pion, FGD1-TPC1 matched sample, charged pion candidate momentum categorized by true event topology.

almost entirely DIS and multi-pion interaction modes.

We expect that any uncertainty that can be attached to the DIS and multi-pion modes, e.g. an uncertainty on the charged pion multiplicity, to contribute to the uncertainty in the CC-1-pion selection.

6.1.3 Motivation: Current Uncertainties

Currently there is only one uncertainty assigned to the CC-other interactions. The error is derived from the total CC inclusive cross section reported by MINOS [68], shown in Figure 6.8, with additional details given in [69]. This uncertainty is implemented as an energy-dependent function that applies to NEUT's CC multi-pion, CC DIS, and CC resonant $K/\eta/\gamma$ production according to the function

$$\sigma_{CC-Other} = 4.0\text{GeV}/E_{\nu}. \tag{6.1}$$

The dependence of this error on the inverse of the neutrino energy may cause alarm, but as the CC-other interactions have a threshold of around 0.6 GeV there is no concern over this error becoming ill-behaved at low energies.

With only a single error on the CC-other sample it was decided to try and obtain

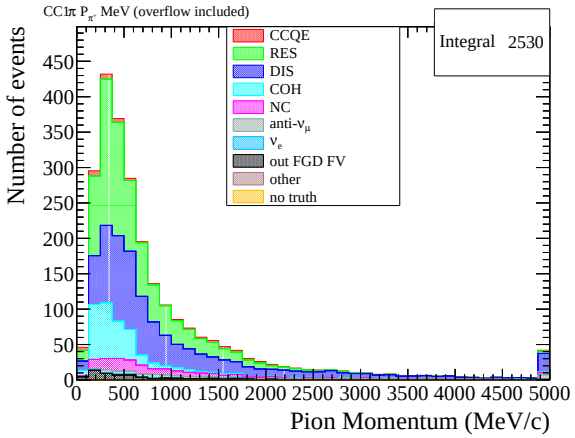

Figure 6.6: CC-1-Pion, FGD1-TPC1 matched sample, charged pion candidate momentum categorized by true NEUT interaction mode.

additional error on the DIS and multi-pion interactions.

6.2 The Fermilab Data

A deuterium-target experiment on a 15-foot bubble chamber was carried out at Fermilab in 1980s [65]. A 350 GeV/c proton beam produced a wide-band neutrino beam with average neutrino energy $\langle E_\nu \rangle = 50$ GeV. Events in the data sample were separated into $\nu - p$ and $\nu - n$ interactions. The data was separated into different target nucleons by consideration that neutron targets produce an odd number of prongs ($\nu + n \rightarrow \mu^- + n + X^+$), and proton targets produce an event number of prongs ($\nu + p \rightarrow \mu^- + n + X^{++}$).

The raw data for charge multiplicities a function of $\langle W^2 \rangle$ and target, taken from Table 1 of [65] was digitized[1] and re-analyzed. The data is shown in Figures 6.9 and 6.10.

6.3 Interlude: How Many Charged Hadrons?

Before the analysis of the Fermilab data will be presented it would be apt to describe the motivation behind the particular analyses carried out. The question at hand is the development of a model that answers the following question: For a neutrino-nucleus interaction

[1]Manual data entry by hand, lots of fun.

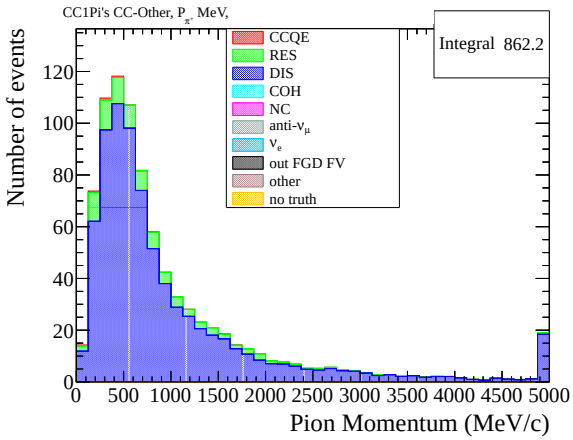

Figure 6.7: Examining just the CC-other *background* from Figure 6.5, it can be seen that the DIS interaction mode dominates.

at some invariant mass W, what is the number of hadrons produced? In particular, what is the number of charged hadrons produced?

The most straight-forward approach involves three primary steps:

1. **Calculate $\langle n_{ch} \rangle$:** As alluded to in Figure 6.1, an empirical relation between the mean number of charged hadrons produced and the invariant mass of the interaction was discovered. The relation is linear in the log of the invariant mass squared:

$$\langle n_{ch} \rangle = A + B \log W^2. \tag{6.2}$$

2. **Calculate $\langle n \rangle$:** From the mean number of charged hadrons, the mean number of all hadrons produced is calculated. The published methods employed in GENIE's AGKY model for hadron multiplicities [54], and the internal methods of NEUT both employ the following argument: All pions are created with equal probability, so the total multiplicity is simply $1.5 \times \langle n_{ch} \rangle$.

3. **Calculate Probability of n Hadrons:** To go from the *mean* number of hadrons to the probability of an interaction which produces n hadrons, the formalism of Koba,

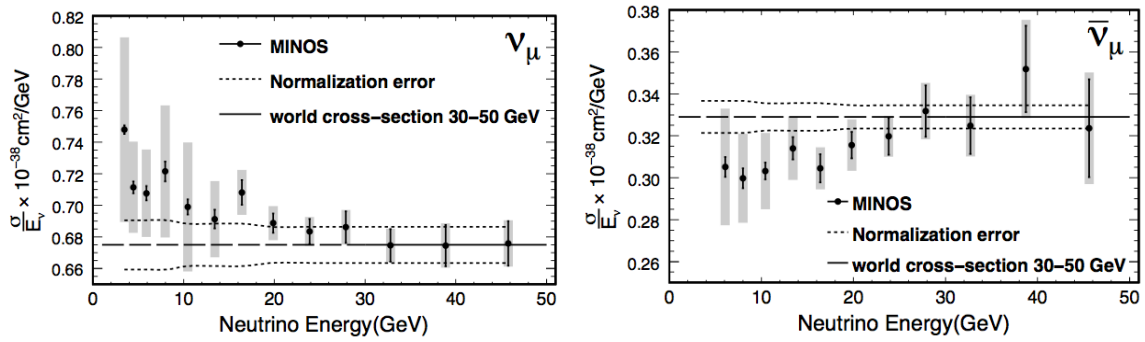

Figure 6.8: MINOS [68] experiment's CC inclusive cross section results, from which the error on the CC-other interactions is derived. The error around 4 GeV is 10%.

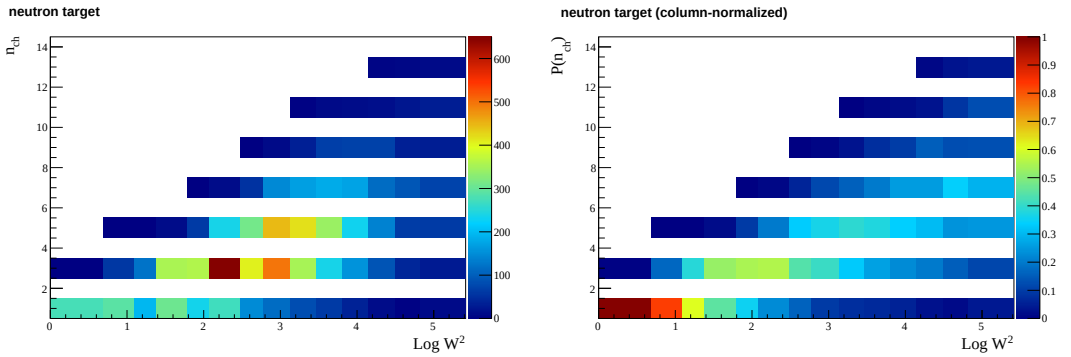

Figure 6.9: Raw data from Fermilab [65], ν-neutron interactions, $\nu + n \rightarrow \mu^- + n + X^+$, with absolute scale on the left and column-normalized on the right.

Nielsen and Oleson, known as KNO scaling [70], is employed. This formalism says that there is a universal function f of the form

$$f(z = n/\langle n \rangle) = \langle n \rangle \times P(n). \tag{6.3}$$

A common parameterization for the KNO scaling function f is the Levy function [54]

$$Levy(z; c) = \frac{2e^{-c}c^{cz+1}}{\Gamma(cz+1)}, \tag{6.4}$$

Where Γ is the Gamma function and the variable $z = n/\langle n_{ch} \rangle$ and there is a single free parameter c.

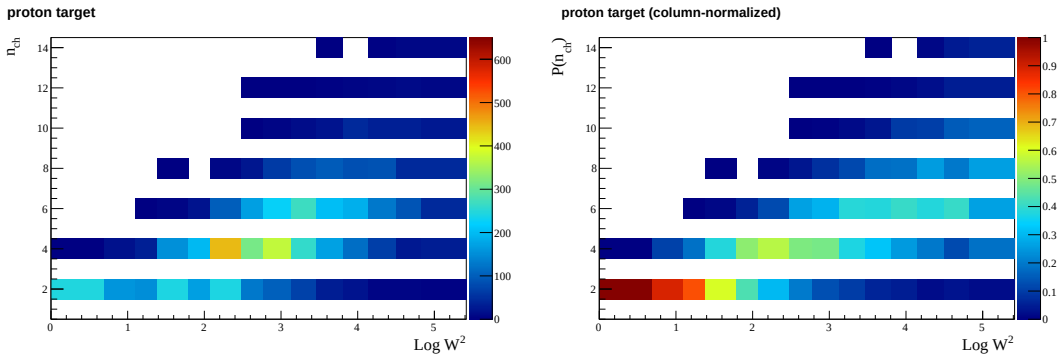

Figure 6.10: Raw data from Fermilab [65], ν-proton interactions, $\nu + p \to \mu^- + n + X^{++}$, with absolute scale on the left and column-normalized on the right.

With this formalism, one can begin with the invariant mass of the interaction, W, calculate the mean charged hadron multiplicity which can then be scaled to the total mean hadron multiplicity, from which the probability distribution for the interaction producing exactly n hadrons can be formulated. We can now return to the discussion of the Fermilab data.

6.4 Fermilab Data: Analysis

The re-analysis of the Fermilab data to produce the charged hadron multiplicity as a function of $\log W^2$ is shown in Figures 6.11 and 6.12.

A standard KNO-based analysis of the data is shown in Figure 6.13.

6.4.1 Comment on KNO Scaling

The KNO scaling relation is, in effect, a measure on the dispersion of multiplicities. We produce a distribution which is peaked at $z \sim 1$ where $z = n/\langle n_{ch} \rangle$, but it has a characteristic *width*. The dispersion of the multiplicity,

$$D_{ch} = \sqrt{\langle n_{ch}^2 \rangle - \langle n_{ch} \rangle^2} \tag{6.5}$$

is also found to follow a linear relation,

$$D_{ch} = A_d + B_d \langle n_{ch} \rangle . \tag{6.6}$$

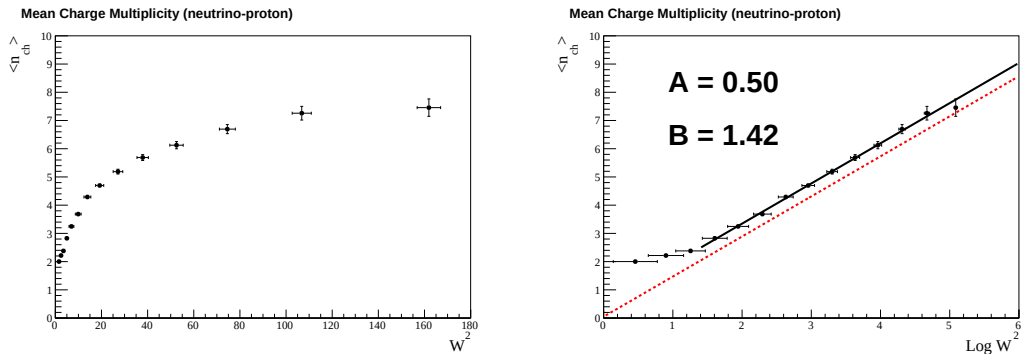

Figure 6.11: Analysis of the Fermilab data for ν-proton interactions, producing mean charged hadron multiplicity as a function of $\log W^2$. The red dashed lines are the published fit values, and the black overlay is the re-analysis. It was discovered that the published fit values for the ν-p interactions was incorrect. The correct fit values are shown in the plot at the right.

For a free nucleon target, it is expected from KNO scaling that the dispersion should go to zero as the mean charged multiplicity goes to zero. The CHORUS experiment found that this was not the case in their data, but their target was heavy emulsion and they attributed the difference to nuclear effects impacting the dispersion [64].

The non-zero dispersion can be addressed by a change of variables

$$z = \frac{n}{\langle n \rangle} \rightarrow z = \frac{n - \alpha}{\langle n \rangle - \alpha}, \tag{6.7}$$

where the parameter α is the x-intercept (not y!) of the dispersion fit. This amounts to taking a linear fit and shifting it to the right until the x-intercept is at $x = 0$.

In the case of the Fermilab data it was not expected that this effect would be very large. The advantage of deuterium is that it is *almost* a free-nucleon target, meaning the effect of FSI is minimal (or at least hoped to be!). Nevertheless, the dispersion of the data was calculated and the fit results are shown in Figure 6.14, and using the derived parameter α from these fits an adjusted KNO-based analysis was performed, show in Figure 6.15.

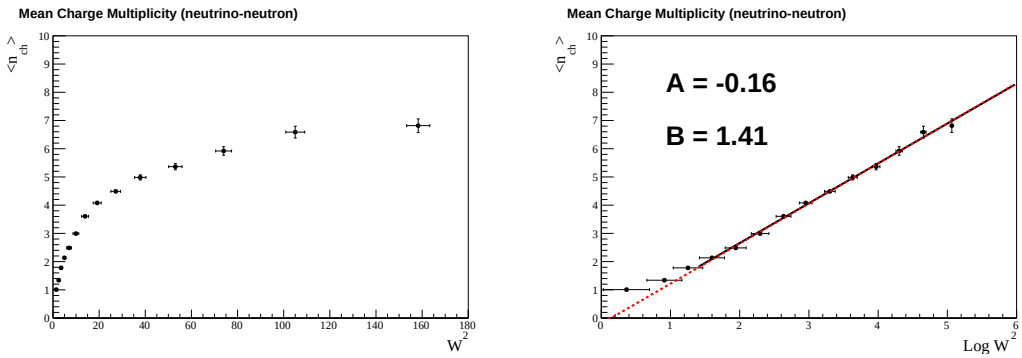

Figure 6.12: Analysis of the Fermilab data for ν-neutron interactions, producing mean charged hadron multiplicity as a function of $\log W^2$. The red dashed lines are the published fit values, and the black overlay is the re-analysis.

6.5 Fit Results

The final approach taken with the Fermilab data in Figures 6.9 and 6.10 was to do a single 2-D fit for the neutron target and proton target data individually. The full fit combines the linear relation in Equation 6.2 (parameters A and B), and the KNO parameterization via the Levy function in Equation 3 (parameter c). Starting with a re-arrangement of the KNO scaling law in Equation 6.3:

$$P(z;A,B,\alpha,c) = \frac{1}{\langle n \rangle} \times f(z) \tag{6.8}$$

$$= \frac{1}{1.5 \times (A + B\,\langle n_{ch} \rangle)} \frac{2e^{-c}c^{cz+1}}{\Gamma(cz+1)}, \tag{6.9}$$

where z, with the change of variables in Equation 6.7, is also a function of A and B and α, $z(A,B,\alpha)$.

The fit was performed under two conditions: A standard KNO fit with α fixed at zero, and a modified fit with α free. The 2-D fit results for the Levy parameter c are shown in Table 6.1, and the linear parameters A and B along with the FSI-like parameter α are shown in Table 6.2.

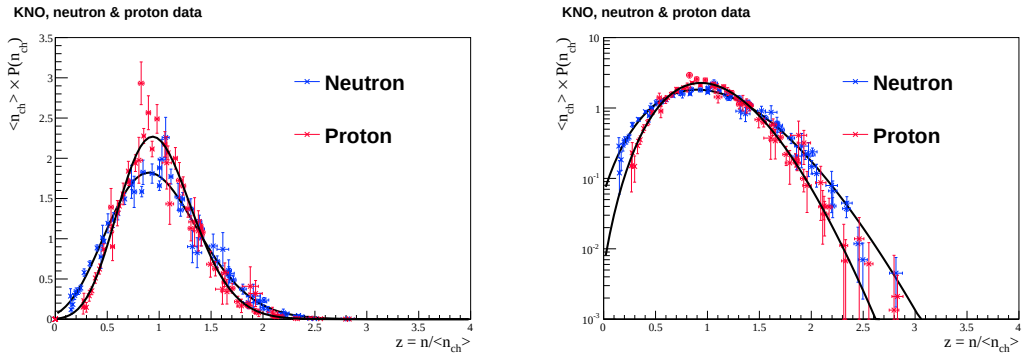

Figure 6.13: KNO analysis of the ν-proton and ν-proton interactions in regular and log scales.

Table 6.1: Results of KNO fits with Levy function parameterization on each set of data from Fermilab. A fit with the parameter α fixed at zero and a fit with α free were performed. The Levy function parameter c is shown in comparison with the published AGKY model's values [54] used by GENIE.

	c	χ^2	c_α	χ^2_α	c_{GENIE}
νn	5.21 ± 0.08	114.6/57	6.22 ± 0.43	106.5/56	5.22 ± 0.15
νp	7.98 ± 0.17	57.5/60	10.2 ± 1.29	53.3/59	7.93 ± 0.34

6.6 The Hadron Multiplicity Model

A simple hadron multiplicity model was developed to utilize the fit results derived from external data sets. The algorithmic outline of the program is as follows:

1. Select a target nucleon, either proton or neutron, with equal probability.

2. Select a random invariant mass W, enforce $\log W^2 > 1$.

3. Use the fit A and B parameters and thrown W to calculate $\langle n_{ch} \rangle$. For proton targets, the value 0.66 is subtracted from the mean to account for proton contributions, and for neutron targets the value 0.50 is subtracted from the mean to account for proton targets. These numbers are based on the claim that for proton targets, the outgoing

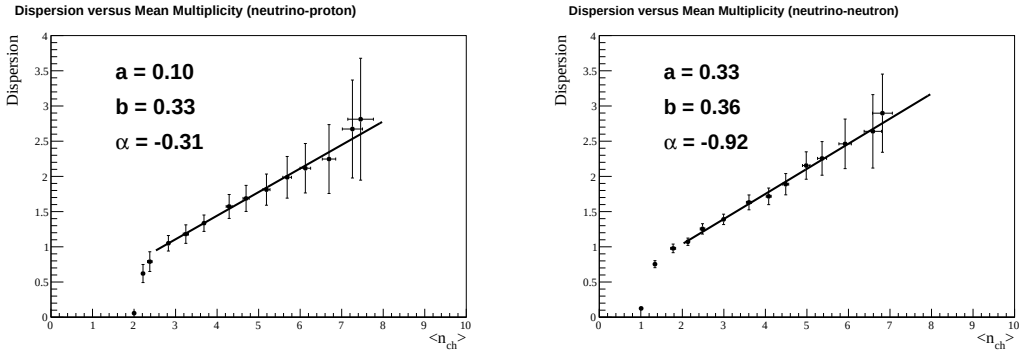

Figure 6.14: Analysis of the dispersion of the mean charged multiplicity from the Fermilab data, with a linear fit. The x-intercept value is taken as the parameter α used in the adjusted KNO analysis.

Table 6.2: Linear parameters and FSI parameter α results of KNO fits with Levy function parameterization on each set of data from Fermilab.

	A	B	α
νn (α fixed)	-0.05 ± 0.05	1.38 ± 0.02	0.0
νn	-0.08 ± 0.05	1.39 ± 0.02	0.35 ± 0.15
νp (α fixed)	0.68 ± 0.06	1.36 ± 0.03	0
νp	0.61 ± 0.07	1.38 ± 0.03	0.62 ± 0.34

proton:neutron ratio is 2:1, whereas for neutron targets the ratio is 1:1 [54]. This now produces the mean number of *charged pions*, as opposed to the charged *hadron* mean.

4. Calculate the total pion mean multiplicity as $1.5 \times \langle n_{ch} \rangle$.

5. Use the KNO distribution with the fit parameters A, B, α and $\langle n \rangle$ to throw a random number of pions produced in an event, n_{pions}.

6. Generate pairs of neutral pions with probability 1/3 and $\{\pi^+, \pi^-\}$ pairs with probability 2/3 until n_{pions} have been generated. If n_{pions} is an odd number, generate the last charged pion randomly with equal probability.

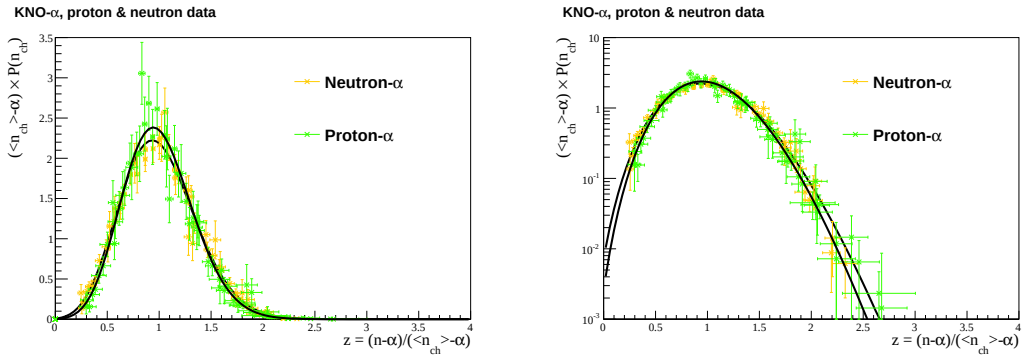

Figure 6.15: An adjusted KNO analysis of the ν-proton and ν-proton interactions in regular and log scales, where the possible effects of the target not being an ideal nucleon are accounted for in the parameter α.

7. Select one out-going nucleon with equal probability.

8. Enforce conservation of charge given the target nucleon, ejected nucleon, and generated pions.

Random throws of the last three steps in this model are repeated if the charge conservation criterion fails. The results of this model with nominal tuned parameters are shown in Figure 6.16. As the model keeps track of the total charged hadrons produced (ejected nucleon and pions), both the mean charged hadron multiplicity and the charged pion multiplicity can be calculated. The model is compared to the bubble chamber data for ν-neutron and ν-proton interactions, where the total charged hadron multiplicity agrees well. It is then assumed that the underlying charged pion multiplicity is a good approximation.

6.7 Reweighting the NEUT Nominal

The final goal of this endeavor will be to produce a 2-D histogram that represents the number of charged pion versus $\log W^2$ for any given assignment of the parameters estimated in the previous section. If this is done, then the parameters can be varied to assess the change in this distribution of charged pion multiplicity. Furthermore, the distribution can be taken as a ratio against NEUT's nominal distribution and a weighting scheme can be generated.

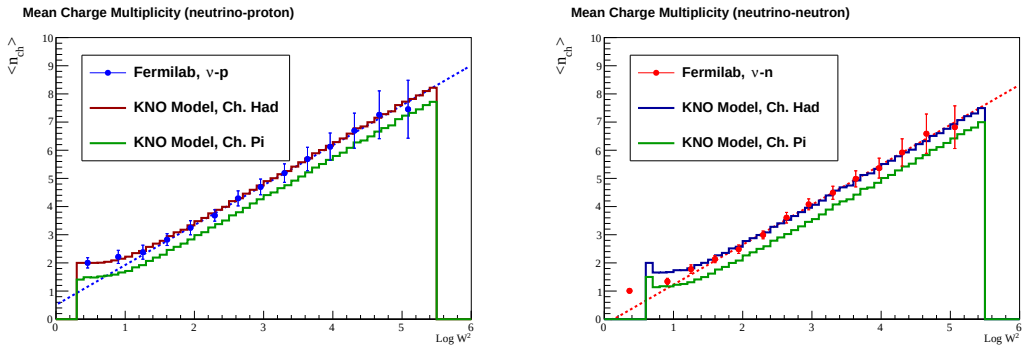

Figure 6.16: Result of the toy charged hadron multiplicity model as compared to the Fermilab data. The model produces charged pion mean multiplicities, as well as total charged hadron multiplicities.

With these weights established, the change in multiplicity-related parameters can be applied to the NEUT model or to the ND280 tracker selections.

An example of the weight generation histograms is shown in Figure 6.17, where the toy values are generated, the NEUT nominal output is generated, and then the ratio of the histograms is taken to form a final weighting histogram. This is done for protons and neutrons individually.

The application of these weights to the NEUT sample are shown in Figure 6.18. A majority of the tuning takes place in the higher region of $\log W^2$, and this turns out to be the region where NEUT switches from its internal hadron multiplicity model to the Pythia/JETSET package. However, the T2K beam produces neutrino interactions that are no greater than $\log W^2 \sim 3$ GeV$^2/c^4$. The nominal and reweighted tracker samples are shown in Figures 6.19. Note especially in the un-tuned tracker sample the harsh and completely non-physical transition between NEUT's hadron multiplicity model and the Pythia/JETSET methods.

6.8 Error Estimation

The final endeavor with the hadron multiplicity model is to estimate the impact on the tracker samples. This is where the full 4-parameter fit to the 2-D Fermilab data is utilized.

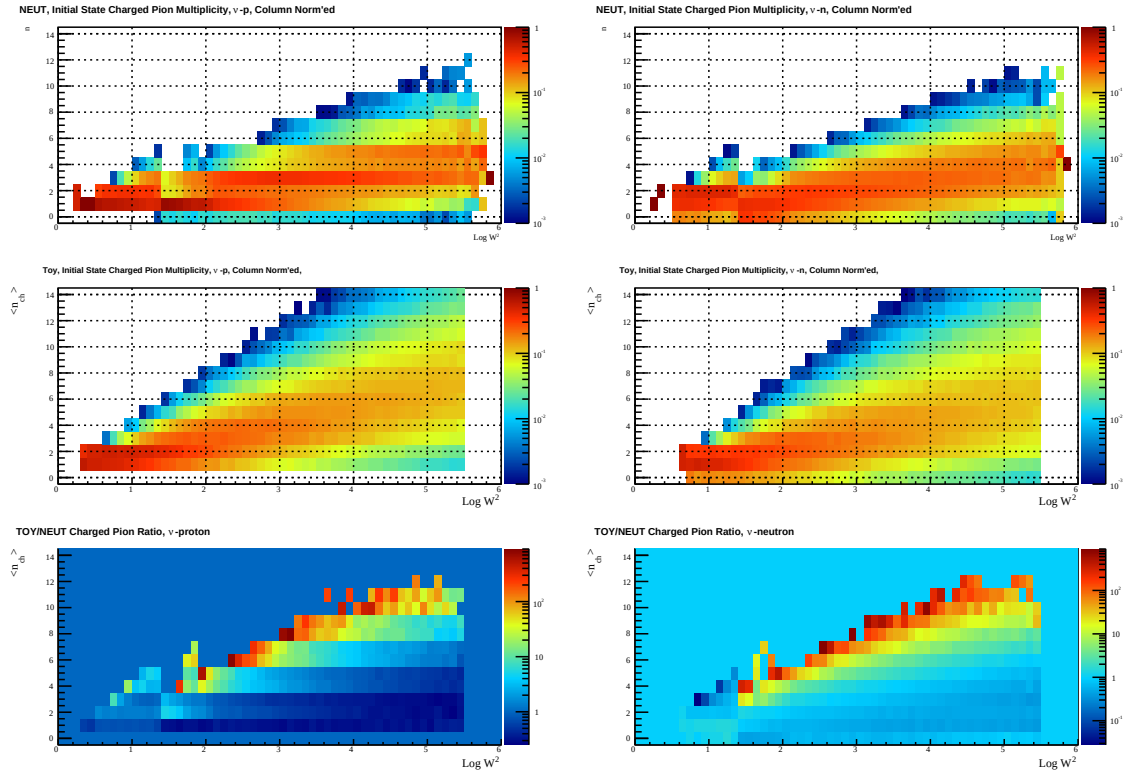

Figure 6.17: Example generation of the weight histograms for each nucleon target. The top row is the NEUT nominal output, the second row is the toy model's tuned output, and the third row is the TOY/NEUT ratio. It should be noted this is only an example set of weights from an earlier stage of development.

The fit produces a covariance matrix for the parameters, so correlated throws can be made in accordance with parameter errors for A, B, c and α. However, it was found that as the data was published with only statistical errors that the variations of the parameters were minimal.

To address this concern, one more set of published data from the Big European Bubble Chamber (BEBC) experiment's [71] liquid hydrogen data was referenced. A measurement of the linear relation $\langle n_{ch} \rangle$ was made, and the difference between the BEBC results and the Fermilab results was taken as an additional systematic error and added to the uncertainty on the fit parameters A and B. The BEBC result for ν-proton interactions is

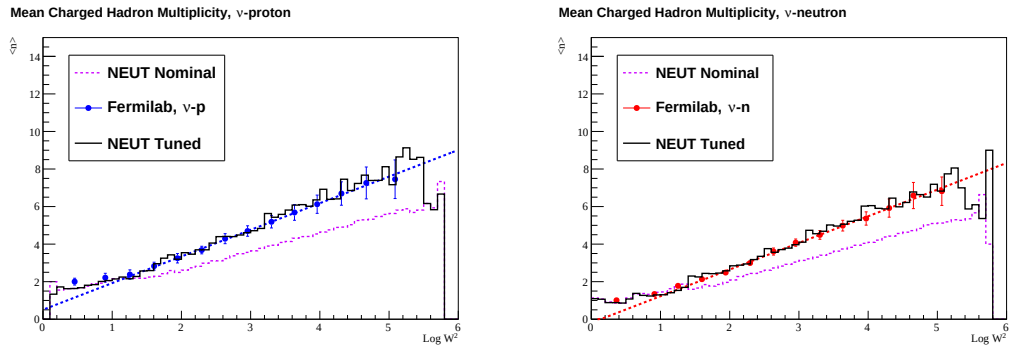

Figure 6.18: Result of the application of the charged hadron multiplicity tuning to the NEUT sample. The NEUT nominal distribution is shown against the post-tuning NEUT histogram, overlaid with the Fermilab bubble chamber data and linear fits.

$\langle n_{ch} \rangle = (0.37 \pm 0.02) + (1.33 \pm 0.02) \log W^2$. A projection of the 4-D parameter space into the two dimensions of the linear parameters A and B is shown in Figure 6.20, where the black band shows throws from just the Fermilab statistical error and the gray band shows throws when the additional systematic error from the difference with BEBC is included.

For every throw of the 4-D parameter space a new set of weights is generated, as in Figure 6.17, and then the entire tracker sample is re-processed. For each event the true invariant mass is calculated and the charged pion multiplicity is counted, and then a weight from the appropriate bin is applied. This reweighting process is only applied to NEUT modes that correspond to DIS and Multi-pion interaction modes in the initial state, i.e. before FSI. From 1,000 throws of parameters the re-weighted samples are generated, and for each bin a gaussian fit to the distribution of throws is used to obtain the 1-σ error. An example of the effect of this error on the tracker CC-1-pion and CC-other samples is shown in Figure 6.21.

6.9 Summary

A study of external data was completed, involving the complete re-analysis of published results. A hadron multiplicity model was developed, based on the internal code of the NEUT

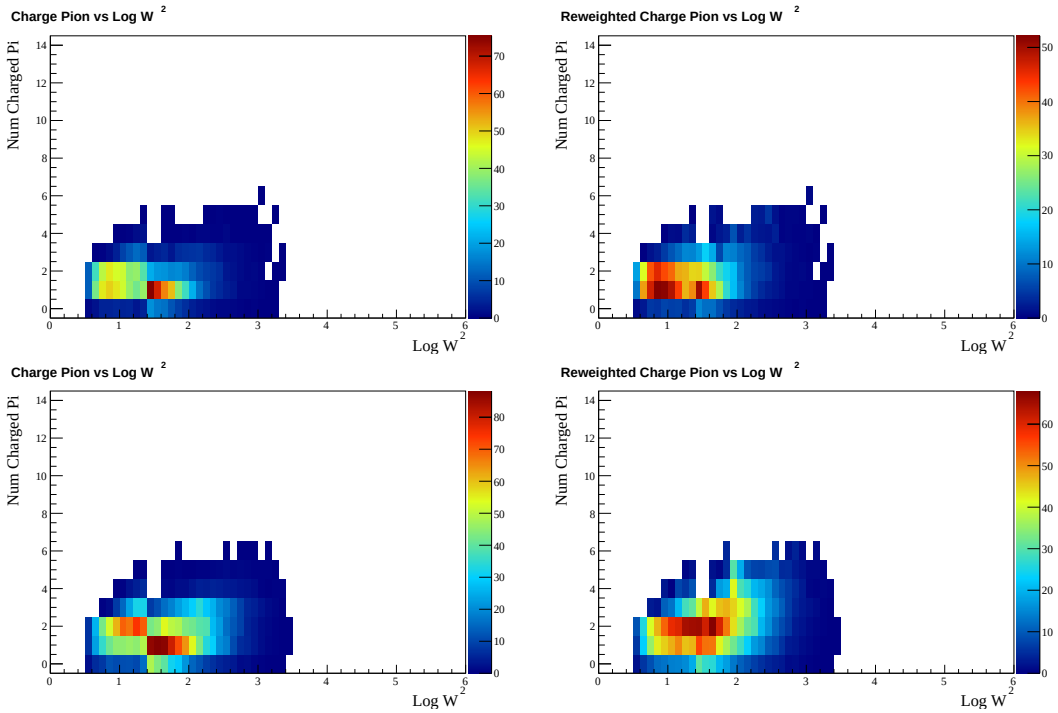

Figure 6.19: The nominal tracker CC-1-pion (top row) and CC-other (bottom row) samples, based on NEUT MC, before and after tuning. The nominal distributions on the left demonstrate the harsh transitions between NEUT's hadron multiplicity model and the out-sourced Pythia/JETSET-based model. The reweighted samples alleve this transition somewhat.

model and the AGKY [54] model. A parameterization of charge multiplicity distributions was fit to the Fermilab data to obtain tuned model parameters, and this tuning was applied to NEUT via a weighting scheme that was also developed. The published errors on various model parameters were used to estimate the systematic error associated with charged hadron multiplicities, and this error was propagated to the ND280 tracker samples.

The work presented in this chapter was presented to the authors of NEUT and GENIE, and developed in conjunction with their feedback [72]. The work presented herein is on going and is under review for inclusion in the Super-Kamiokande atmospheric oscillation studies. The atmospheric neutrino sample has a significantly different distribution in W, namely it is peaked at a higher value, so it is anticipated that the tuning to the NEUT model may be more significant therein.

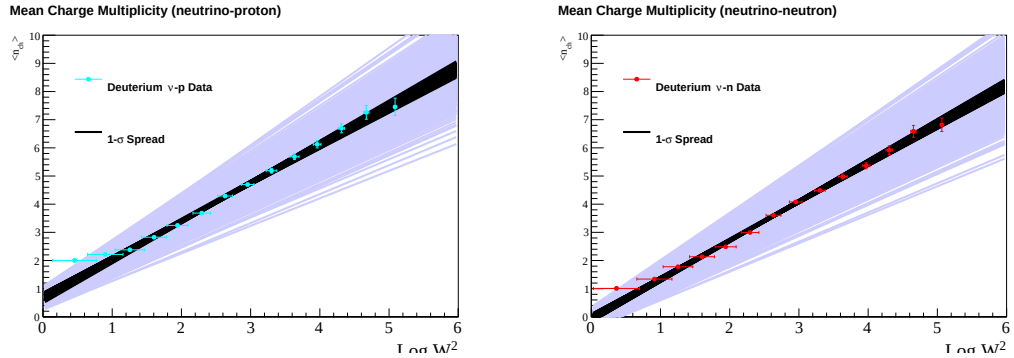

Figure 6.20: Projection of statistical throws of the 4-D parameter space for the hadron multiplicity model into the 2-D parameter space of the linear parameters A and B. The black bands are due to the Fermilab statistical error, and the gray bands are due to an additional systematic derived from differences with the BEBC experimental results.

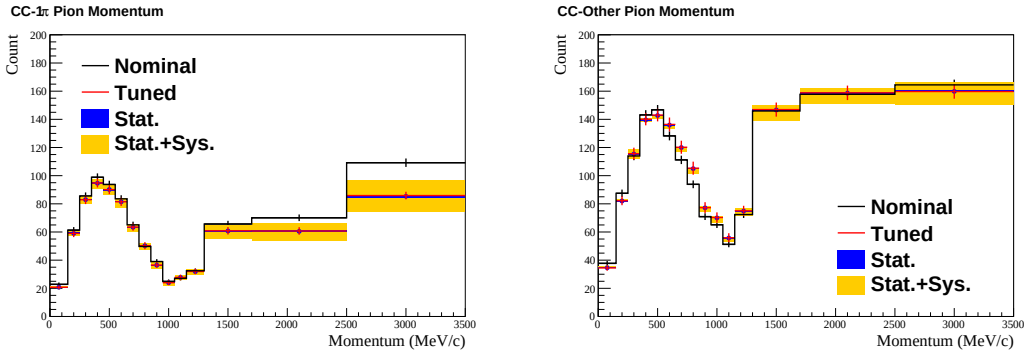

Figure 6.21: Error band due to the charged pion multiplicity uncertainty. The statistical errors are from the Fermilab published data, and the systematic error is attributed to the difference between the BEBC and Fermilab results.

Chapter 7

THE FIT

In this chapter an analysis of the data presented in Chapter 5 will be discussed. A model-dependent fit of the CC-1-pion and CC-other samples will be described. The goal of this measurement will be to extract normalization parameters that represent the variation of the NEUT model's observable CC-1-pion cross section on carbon. NEUT is the official T2K model for neutrino interactions and is well-established in the larger neutrino physics community. The CC-1-pion normalization parameters will give measure of the constraint the ND280 data has on any adjustments to the CC-1-pion cross section necessary. A more detailed description of the fit method and fit parameters will be given.

7.1 Selection

As presented in Chapter 5, a thorough CC inclusive selection was broken down into exclusive channels. In particular, the CC-1-pion sample was broken into three sub-categories defined by the method of pion detection. For this measurement the FGD-TPC matched pion sub-sample will be used for the CC-1-pion selection. This is the sub-selection in which the positive pion candidate in the CC-1-pion selection is a matched track from FGD1 to TPC2. Events where the pion is detected by Michel Electron or isolated FGD-only tracks are not included here. The FGD-TPC matched sample was selected due to the good momentum resolution of the TPC, and, similarly, the ME and FGD-only samples were ignored due to the in-development status of their momentum reconstruction routines.

The CC-other selection used here also requires there to be at least one positive FGD-TPC matched positive pion track. Events such as multi-pion interactions where only negative pions make it into the TPC, which are part of the general CC-other category, are not included.

All events in the two selected samples are binned in positive pion momentum with the

following binning (9 bins total, per sample):

- Pion Momentum (MeV/c): $[0-250, 250-350, 350-450, 450-550, 550-650, 650-850, 850-1150, 1150-1700, 1700-30000]$

No angle binning of the pion is used in this analysis. For display purposes, in all plots of variables using the above binning (e.g. Figure 7.1) have the final plot draw with the last bin as $1,700-2,500$ MeV/c, but **overflow out to 30 GeV/c is included**.

7.2 Likelihood Definition

In the process of studying the full fitting method and testing some of the proposed ideas that are now applied in this analysis, a series of toy studies were performed. The toy studies use simple distributions to illustrate the likelihood fitting methods used in the full fit. For readers unfamiliar with the process, Appendix A provides a more verbose and conceptual discussion on the fit methods applied here.

The pion momentum data for each sample is binned in the corresponding sample's histogram. Taking both samples together, let the bin index be i. Then, for each bin we have the predicted number of events in bin i, N_i^p, and similarly the measured data in in bin i, N_i^d. The predicted number of events is the culmination of our model's prediction and it incorporates parameters for neutrino flux ($\mathbf{b}$), cross section models ($\mathbf{x}$), detector systematic uncertainties ($\mathbf{d}$) and pion multiplicity uncertainties ($\mathbf{m}$). The CC-1-pion normalization parameters ($\mathbf{c}$) are free parameters. The bold face indicates that each category of model parameter is a vector of parameters, and the predicted number of events can be written explicitly as a function of these parameters: $N_i^p(\mathbf{b}, \mathbf{x}, \mathbf{d}, \mathbf{m}, \mathbf{c})$. Let the collected set of parameters be take as the shorthand $\mathbf{s} = \{\mathbf{b}, \mathbf{x}, \mathbf{d}, \mathbf{m}, \mathbf{c}\}$.

We expect the number of events in each bin to be Poisson distributed, giving a joint probability density function (PDF):

$$L_{bins}(\mathbf{s}) = \prod_{i}^{bins} \frac{[N_i^p(\mathbf{s})]^{N_i^d} \, e^{-N_i^p(\mathbf{s})}}{N_i^d!} \tag{7.1}$$

The uncertainty and model parameters are constrained and the fit parameters are taken to have multivariate normal distributions. The flux parameters prior PDF, for example,

takes of the form

$$\pi(\mathbf{b}) = \frac{1}{(2\pi)^k |V_b|^{1/2}} e^{-\frac{1}{2}\Delta\mathbf{b}(V_b^{-1})\Delta\mathbf{b}^T}, \tag{7.2}$$

where $\Delta\mathbf{b} = \mathbf{b} - \mathbf{b_{nom}}$ is the deviation of the parameter vector of dimension k from the nominal parameter values, and V_b is the covariance matrix. Identical distributions are taken for the cross section, detector and pion multiplicity parameters and represent the parameters that are varied in the fit.

In total we have the joint likelihood function

$$\begin{aligned} L(\mathbf{s}) &= \pi(\mathbf{b}) \times \pi(\mathbf{x}) \times \pi(\mathbf{d}) \times \pi(\mathbf{m}) \times L_{bins}(\mathbf{s}) \\ &= \pi(\mathbf{b}) \times \pi(\mathbf{x}) \times \pi(\mathbf{d}) \times \pi(\mathbf{m}) \times \prod_i^{bins} \frac{[N_i^p(\mathbf{s})]^{N_i^d} e^{-N_i^p(\mathbf{s})}}{N_i^d!}. \end{aligned} \tag{7.3}$$

In the fit algorithm, the quantity that is maximized is the likelihood ratio. The likelihood ratio is the likelihood function relative to the likelihood evaluated at the nominal parameter values, $\mathbf{s}_{nom}$ and with $N_i^p = N_i^d$,

$$L^r = \frac{L(\mathbf{s})}{L(\mathbf{s}_{nom})}. \tag{7.4}$$

Maximizing the likelihood ratio is equivalent to minimizing the negative log-likelihood ratio [17]:

$$\begin{aligned} -2\log L^r(\mathbf{s}) &= 2\sum_i^{bins}\left[N_i^p(\mathbf{s}) - N_i^d + N_i^d \log\frac{N_i^d}{N_i^p(\mathbf{s})}\right] && \text{(Bin)} \\ &+ \sum_i^{E_\nu\ bins}\sum_j^{E_\nu\ bins} \Delta b_i (V_b^- 1)\Delta b_j && \text{(Flux)} \\ &+ \sum_i^{xsec\ pars}\sum_j^{xsec\ pars} \Delta x_i (V_x^- 1)\Delta x_j && \text{(Cross Section)} \\ &+ \sum_i^{det\ pars}\sum_j^{det\ pars} \Delta d_i (V_d^- 1)\Delta d_j && \text{(Detector)} \\ &+ \sum_i^{mult\ pars}\sum_j^{mult\ pars} \Delta m_i (V_d^- 1)\Delta m_j && \text{(Pion Multiplicity)} \end{aligned} \tag{7.5}$$

Where the first term is due to the binned likelihood and the subsequent terms are due to each prior parameter type. In this form it is easier to see how our prior knowledge is incorporated: any deviation from our best estimate of the parameter values will generate a

non-zero Δ, thus increase the negative log-likelihood. The covariance matrices which encode parameter correlations and our error estimates will determine the size of the penalty added for a given deviation of the parameter from the nominal parameter values.

The predicted number of events per bin is obtained through an event-by-event weighted sum over all MC events. The fit parameters are varied and then the distribution of events is built via

$$N_i^p = \sum_j^{MCEvents} b_j \times x_j^{norm} \times w_j^x(\mathbf{x}) \times m_j \times c_j \times w_j^d(\mathbf{d}) \times \delta_i \left(p_j(\mathbf{d}), s_j(\mathbf{d})\right). \tag{7.6}$$

Each term in the product represents an event-specific weight associated with tunable parameters representing different physics components of our overall model. Variations in the expected flux produce a flux-related weight to be applied, via the weight b_j. The cross section-related parameters x^{norm} are fractional weights that represent changes in cross section normalization for different interaction types, and the cross section weights $w^x(\mathbf{x})$ are functions of final state interaction (FSI) parameters (e.g. weights that reflect the effect of changing pion absorption, charge exchange, etc). The free CC-1-pion normalization parameters c_j reweight only true CC-1-pion events and are functions of true pion momentum. The detector systematic parameters can modify the observed pion momentum p, meaning an event can migrate between samples s or bins. The effect of varying the detector parameters is propagated by the weights $w^d(\mathbf{d})$. The function δ_i examines the observables as a function of detector parameter variation and matches the event to the observable bin N_i^p, to which it is added with the appropriate net weight. Each of these parameters will be discussed in more detail, but it is through this event-by-event reweighting that the fit routine builds predictions based on parameter variations.

A brief description of each parameter category will now be given.

7.2.1 Detector Error Parameters

A covariance matrix approach is used to model the detector systematic errors. The results of the various detector systematic uncertainty studies provide tunings and bounds for all of the detector errors. With these inputs, random but correlated throws of the detector error

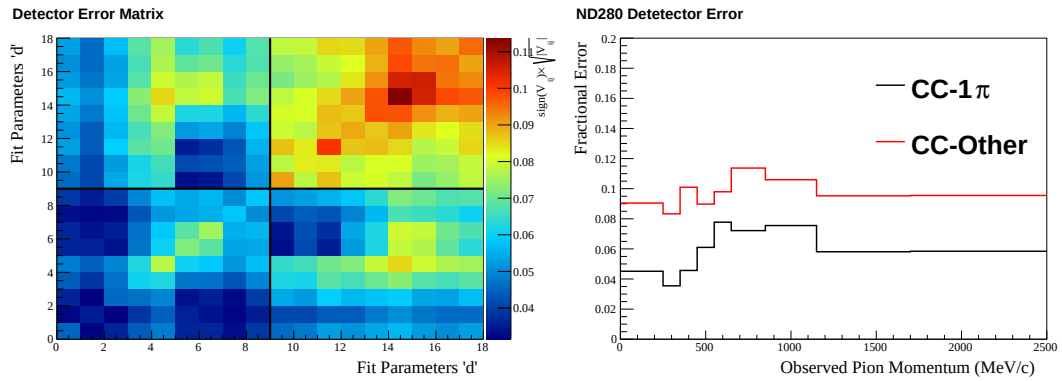

Figure 7.1: Detector fractional error covariance, bins 1-9 correspond to pion momentum bins in the CC-1-pion sample, bins 10-18 correspond to pion momentum bins in the CC-other sample. At the right is the fractional error in each sample.

parameters can be made and the result on the observable distribution can be assessed in the reweighted distribution of events, via Equation 7.6.

To construct the covariance matrix due to the variation of detector errors, 2,000 throws of detector parameters were made. In addition, throws of the MC statistics were incorporated at this stage to account for the MC statistical error. Then, from these throws a mean value per bin was calculated,

$$N_i^{mean} = \frac{1}{2000}\sum_{k=1}^{2000} N_i^k, \tag{7.7}$$

and the covariance was calculated as:

$$(V_d)_{ij} = \frac{1}{1999}\sum_{k=1}^{2000} \frac{(N_i^k - N_i^{mean})(N_j^k - N_j^{mean})}{N_i^{mean} N_j^{mean}}. \tag{7.8}$$

The error on bin i is thus given as $\sqrt{V_{i,i}}$ and the correlation in error between bins due to all of the detector errors is given by the off-diagonal elements. Furthermore, the detector parameters $\mathbf{d}$ are given a new nominal value based on the mean calculated here, $d_i^{nom} = N_i^{mean}/N_i^{nom}$, as the covariance is calculated from the means of each bin. The detector systematic covariance is shown in Figure 7.1.

7.2.2 Flux Parameters

The flux parameters **b** are normalization parameters in bins of parent neutrino energy. These parameters represent the change in event rates due to the uncertainties in the flux. The flux reweighting and tuning group of the T2K collaboration is responsible for flux simulation and modeling, as well as delivering a flux covariance matrices for each detector site. The flux tuning accounts for uncertainties in off-axis angle, total number of POT delivered, kaon production, and hadron multiplicity. Flux tuning is one of the largest and most important endeavors of the T2K experiment, the details of which are beyond the scope of this thesis, but can be found in [73].

The flux covariance matrix has the following binning, in GeV:

- ν_μ (11 bins): 0.0, 0.4, 0.5, 0.6, 0.7, 1.0, 1.5, 2.5, 3.5, 5.0, 7.0, 30.0
- $\bar{\nu}_\mu$ (5 bins): 0.0, 0.7, 1.0, 1.5, 2.5, 30.0
- ν_e (7 bins): 0.0, 0.5, 0.7, 0.8, 1.5, 2.5, 4.0, 30.0
- $\bar{\nu}_e$ (2 bins): 0.0, 2.5, 30.0

On an event-by-event basis the parent neutrino type and energy is evaluated and the associated flux parameter b for that bin is applied. The flux fractional error covariance matrix is shown in Figure 7.2.

7.2.3 Final State Interaction Parameters

A total of six final state interaction (FSI) systematic parameters are accounted for in the fit. They are:

- Pion Production
- Pion Absorption
- Low-energy Charge Exchange

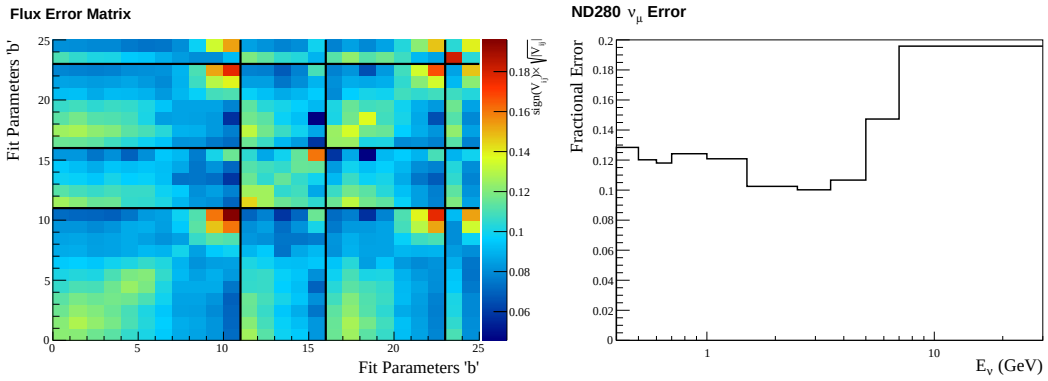

Figure 7.2: Fractional error covariance matrix (left) for flux parameters. The binning is in parent neutrino energy and flavor, as described in the text and delineated by the black bars. On the right is the flux error on the primary neutrino component, ν_μ.

- High-energy Charge Exchange
- Low-energy Inelastic Interactions
- High-energy Inelastic Interactions

These interactions represent the interactions produced pions can undergo as they traverse the nucleus from the interaction vertex. They are "low-level" parameters in the sense the weights associated with these parameter variations correspond to modifications of elementary processes within the NEUT's re-interaction models.

The response to FSI adjustments requires complex calculations of interaction probabilities at incremental steps as pions traverse the nucleus. To bypass this computational expenditure, event weights due to the effect of changing the individual FSI parameters from -3σ to $+3\sigma$ in steps of 1σ are calculated, and then these results are interpolated with a spline[1]. One spline for each FSI parameter is generated per event, and as the fit varies the FSI parameters the effect of these effects can be calculated by evaluating the spline at the associated parameter value to obtain the weight for the event. The FSI reweighting

[1]A spline is an smooth, differentiable function conceptually similar to a polynomial fit to of order n to n data points, but with additional constraints on matching higher order derivatives such that oscillations due to overfitting are suppressed

	Inel. Low	Inel. High	π Prod.	π Abs.	CEX Low	CEX High
CEX High	0.000	-0.304	0.373	0.000	0.000	0.278
CEX Low	0.071	-0.046	0.000	-0.048	0.567	0.000
π Abs.	0.151	-0.036	0.000	0.412	-0.048	0.000
π Prod.	0.000	-0.408	0.500	0.000	0.000	0.373
Inel. High	-0.053	0.338	-0.408	-0.036	-0.046	-0.304
Inel. Low	0.412	-0.053	0.000	0.151	0.071	0.00

Table 7.1: The prior error matrix ($sign(V_{ij}) * \sqrt{|V_{ij}|}$ where V_{ij} is the covariance matrix) on the FSI systematic parmeters, from [74].

mechanism is setup in such a way that the weight due to each individual interaction can be factored, such that the total FSI weight is due the product of each individual interaction reweighting. Additionally, the splines are designed in a way such that they preserve the overall cross section and total number of events.

The T2K Neutrino Interactions Working Group (NIWG) is responsible for creating and delivering the parameter covariance matrix that constrains these parameters. The full details of FSI parameter development can be found in [75]. The error matrix is shown in Table 7.1, and the overall effect of these parameters on the selection samples is shown in Table 7.2.

7.2.4 Cross Section Parameters

Similarly to the FSI parameters, the cross section parameters are defined by the NIWG and complete details can be found in [69, 76]. The original efforts behind the implementation of the FSI and cross section parameters was to gain some variational flexibility in $p - \theta$ phase space used in the full oscillation analyses. As the oscillation analyses rely on the CC-QE interaction as their "golden channel" from which they reconstruct the parent neutrino energy, many parameters were implemented with the CC-QE channel in mind. In the case of CC-1-pion and CC-other interactions, some of these parameters have very little influence as the CC-QE content in these two samples is marginal.

The cross section parameters and their uncertainties are derived by fits to data from the MiniBooNE experiment for the respective interaction [77, 21, 78, 27]. The cross section parameters used in this fit are:

- $\mathbf{M_A^{QE}}$: A dipole approximation of the axial form factor of the nucleus, described in more detail in Section 2.5.

- **CCQE**: A set of normalization parameters on the CC-QE interactions, applied in parent neutrino energies of different regimes.

- $\mathbf{P_F, E_B}$: A pair of parameters that describe the target nucleus in the CC-QE interaction. The model describes the nucleus as a Relativistic Fermi Gas (RFG) with two parameters: The Fermi momentum of the nucleus (P_F), and the binding energy E_B. For Carbon, $P_F = 217 \pm 30$ MeV/c and $E_B = 25 \pm 9$ MeV/c [24].

- **Spectral Function (SF)**: A parameter that attributes an uncertainty due to the difference between the RFG model of the nucleus, and the Spectral Function, which is a new model of the nuclear potential believed to be more accurate.

- **Resontant Pion Production**: A few parameters pertinent to the resonant pion production processes are used: M_A^{RES}, a dipole approximation of the axial form factor of the nucleons (analogous to M_A^{QE}), two normalization uncertainties on the CC-1-pion rate, binned in parent neutrino energy, and an overall NC-1-π^0 applied over all neutrino energies.

- **CC-Other Shape**: A normalization uncertainty on the CC-other interactions, applied in true parent neutrino energy

- **NC-Other**: A normalization uncertainty on the NC elastic interactions, NC resonant $\gamma/\eta/K$ production, NC DIS and NC Multi-pion production.

- **Pionless Δ-Decay (PDD)**: A parameter that controls the branching fraction of Δ decay in which the pion is absorbed. The default value gives a PDD rate of 20%.

- **Coherent Pion Production:** A normalization of the CC-coherent pion production rate, which due to the large experimental uncertainty is assigned a default error of 100%. At present, it was decided that this parameter would fixed for the analysis presented here as this channel is largely degenerate with the free pion parameters.

The effect of each of the cross section parameters on the fit samples is shown in Table 7.2. These numbers were obtained by generating 1,000 fake data sets where only the listed cross section parameter(s) were varied. The resultant variation in the observed pion momentum bins for the selection samples was fit with a gaussian and the width was taken as the uncertainty.

7.2.5 Pion Multiplicity Parameters

The pion multiplicity parameters, **m**, are responsible for reweighting events that come from the NEUT multi-pion (MP) and deep inelastic scattering (DIS) interaction modes. The origin of these parameters is discussed in detail in Chapter 6. The primary goal of these parameters is to try and incorporate an error in the CC-1-pion sample that is associated with the larger CC-other *background*. This background is mostly MP/DIS interactions, and the CC-other selection is almost entirely MP/DIS.

The pion multiplicity error parameters are applied via a covariance matrix, with one parameter m per true pion momentum bin in each sample, and only with application to true DIS/MP interactions.

A subtlety regarding the implementation of the multiplicity parameters is that *some* CC-1-pion events originate from MP/DIS interactions. This is primarily due to FSI effects where what was originally an interaction producing multiple pions ends up being with a final state topology of only one positively charged pion. If one were to apply the multiplicity weights to the CC-1-pion *selection* then both signal and background events would be re-weighted in a way correlated with the MP/DIS interactions found in the CC-other selection. Isolating the impact on the CC-1-pion cross section is non-trivial in this case.

Instead of applying the multiplicity weights to the selected samples, they were applied in the true topology categories. A covariance matrix was generated in true pion momentum for

Parameter	E_ν Range	Nominal Value	CC1π	CC Other
M_A^{QE}	all	1.21 ± 0.45 GeV/c^2	5.1%	1.5%
CCQE E1	$0 < E_\nu < 1.5$	1.0 ± 0.11	0.4%	<0.1%
CCQE E2	$1.5 < E_\nu < 3.5$	1.0 ± 0.30		
CCQE E3	$E_\nu > 3.5$	1.0 ± 0.30		
p_F ^{12}C	all	217 ± 30 MeV/c	<0.1%	<0.1%
E_B ^{12}C	all	25 ± 9 MeV	<0.1%	<0.1%
SF ^{12}C	all	0 (off) 1 (on)	<0.1%	<0.1%
M_A^{RES}	all	1.41 ± 0.25 GeV/c^2	13.7%	6.1%
CC1π E1	$0 < E_\nu < 2.5$	1.15 ± 0.31		
CC1π E2	$E_\nu > 2.5$	1.0 ± 0.4		
NC1π^0	all	0.96 ± 0.43		
CC Coherent	all	1.0 ± 1.0	12.2 %	6.1%
CC Other shape	all	0.0 ± 0.40	3.4%	6.2%
NC Other	all	1.0 ± 0.30	1.7%	2.8%
Pion-less Δ Decay	all	0.0 ± 0.2	8.1%	2.5%
FSI	all	Table 7.1	3.6%	3.3%
Total			15.7%	10.0%

Table 7.2: The nominal values and fractional errors on the ND280 Tracker CC-1-pion and CC-other predictions from each cross section source of uncertainty. The parameters and their range of applicability in terms of the parent neutrino energy are listed. Parameters without units are normalizations.

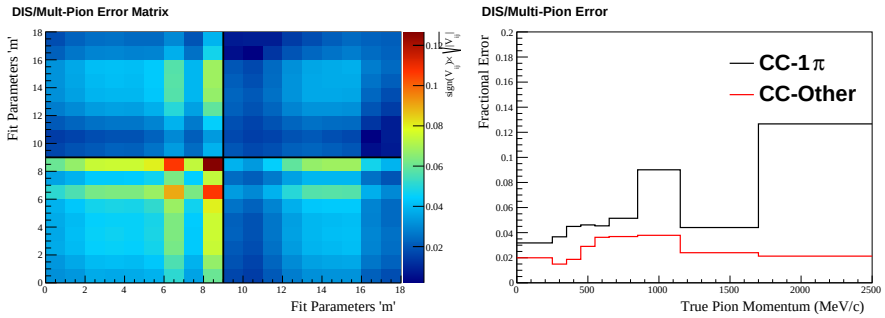

Figure 7.3: The pion multiplicity parameters input covariance matrix. The parameters correspond to true pion momentum normalization in the true CC-1-pion and CC-other topologies. These fractional errors may seem large, but the sub-samples are quite small. For example, within the CC-1-pion selection, the true topology of CC-1-pion events with are due to multi-pion or DIS interactions is very small. The effect of these parameters on the selected *samples* can be seen to be quite small, as shown in Figure 7.4.

true CC-1-pion and CC-other topologies. The fractional error matrix is shown in Figure 7.3.

7.2.6 Free Pion Parameters

The free pion parameters **c** are responsible for scaling the CC-1-pion cross section, and will be referred to as the "pion parameters". There are two distinct features of the pion parameters: they are free (unconstrained), and they are applied in the *true* pion momentum space. There is one pion parameter per true momentum bin, and the true momentum binning is identical to the observed momentum binning.

These two features combine in a subtle way: One might be tempted to say that because you have a number of parameters equal to the number of bins, that you'd get a perfect fit. The subtlety is that these parameters are applied in the true momentum space. As the event-by-event reweighting is applied, any event with a true CC-1-pion topology is reweighted by the parameter c corresponding to the event's true pion momentum. However, as the fitter is minimizing the *observed* pion space, detector effects and resolution which smear events from true space into observable result in the pion parameters for individual true momentum bins being applied to multiple reconstructed bins. This is discussed in detail in the toy

fitting example in Appendix A.

7.2.7 Parameter Error Summary

To assess the total error generated by each category of parameter (flux, detector, cross section, multiplicity) 1,000 toy data sets was generated for each type and the impact on the observed pion momentum bins in each sample was analyzed. For each bin in each sample, and for the total sample size, the distribution of throws was fit with a gaussian and the width of the gaussian reported as an error. The results per bin are shown in Figure 7.4, and the results per sample is shown in Table 7.3.

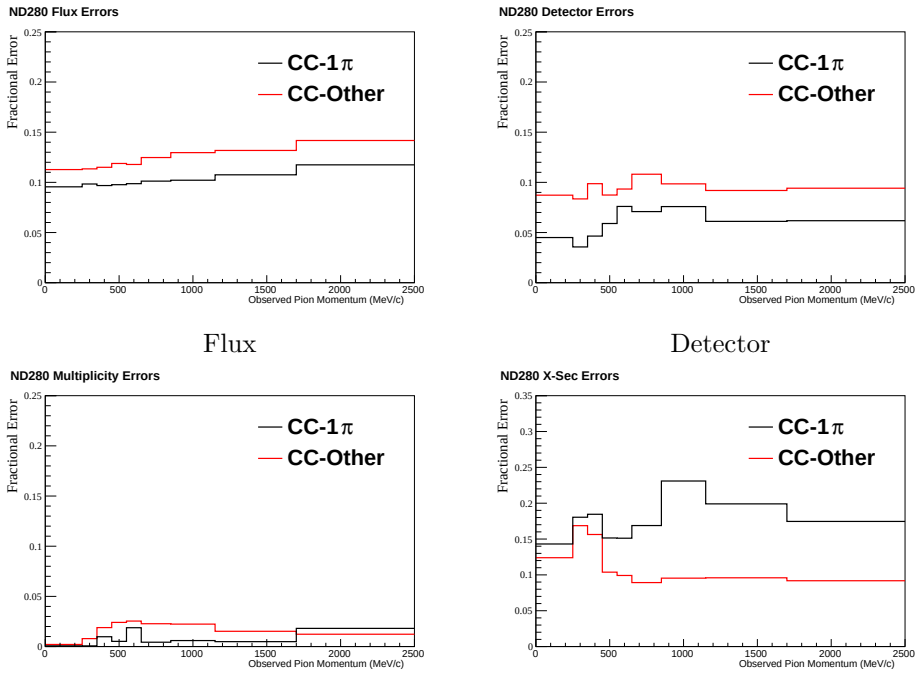

Figure 7.4: Fractional error per parameter category. Note the y-scale for the cross section parameters is larger, and that the cross section parameters are the dominant source of prior error.

Category	CC-1-Pion	CC-Other
Flux	9.8%	12.5%
Detector	4.4%	8.1%
Multiplicity	0.5%	1.4%
Cross Section	15.7%	10.0%
Total	20.6%	17.0%

Table 7.3: Prior error per sample due to each parameter type.

7.2.8 Technical Specifications

This fit was implemented using the ROOT framework's interface to the MINUIT [79] libraries. Specifically, ROOT's Minuit2 libraries in conjunction with the MIGRAD minimization algorithm.

7.3 Analysis Method

Now that the fit method and fit parameters have been introduced, the entirety of the analysis method will now be outlined. Before performing a fit to data the fitter and method is tested on MC. Many fake data sets are generated by making making a random sampling of the fit parameters – all of which are constrained by covariance matrices, so correlations in the sampled fit parameter space are included. One such sampling of the fit parameters is referred to as a "throw," and the MC is reweighted with these parameter values to form the fake data set. The analysis outline here is performed on one such throw. The analysis chain is as follows:

- Perform a throw
- Perform the fit
- Do post-fit throws

- Extract CC-1-pion normalization

- Reweight NEUT's nominal cross section

Each of these steps will now be described.

7.3.1 The Nominal Distribution

For a starting reference point we first examine the nominal distribution. This is the distribution generated when all fit parameters are set to and fixed at their prior values. In Figure 7.5 the breakdown for CC-1-pion (CC-1π) and CC-other (CC-Nπ) samples is shown in two ways: as stacked plots showing the dominant contributing interaction modes, and as overlaid histograms showing the signal, true CC-1-pion interactions on carbon, versus the background (everything else). The reaction codes are given in Table 7.4

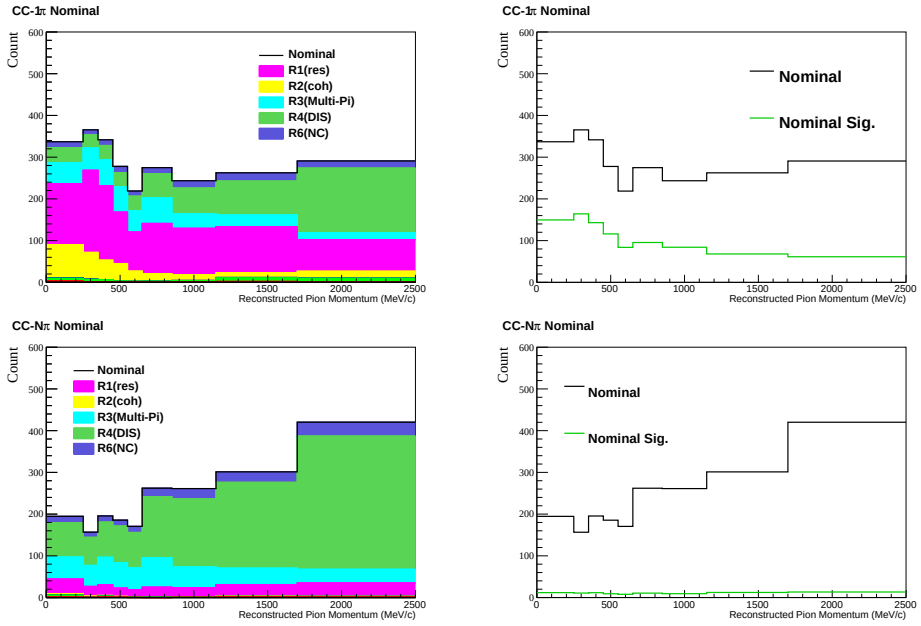

Figure 7.5: The nominal distribution for the CC-1-pion and CC-other samples.

Reaction Code	Category	NEUT modes
R0	CCQE	1
$*$R1	CC1π	11,12,13
$*$R2	CC coherent.	16
$*$R3	CC mult-π	21
R4	CC 1γ	17
	CC η, K, Λ prod.	22,23
	CC DIS	26
R5	NC $1\pi^0$	31,32
$*$R6	NC $1\pi^{\pm}$	33,34
	NC coherent, NC multi-π	36,41,46
R7	other	70
R8	NC other	38,39,42,43,44,45,51,52

Table 7.4: Cross section reaction codes used in fit, categories, and corresponding NEUT modes. Those modes marked with a star are the dominant contributors to the selections in this analysis.

7.3.2 Throw

First perform a throw to generate a fake data set with the following parameter conditions:

- Flux parameters
- Cross section and FSI parameters, **Coherent Interactions fixed at nominal**
- Detector parameters (MC statistical errors incorporated here)
- Pion multiplicity parameters
- Pion parameters **fixed at 1.0**
- Bin statistics thrown

It was decided to fix the CC coherent pion interaction parameter. Coherent interactions, due to their nature, produce pions that undergo no FSI (it is an interaction with the entire nucleus). This makes the behavior of rescaling the coherent interaction rate degenerate with the free pion parameters. This is an approximation, because there are coherent interactions on non-carbon targets that the pion parameters do not reweight.

Throwing "bin statistics" for each fake data set means taking the final histograms and adjusting the bin contents by a random sampling of a poisson distribution with mean equal to the bin content after all reweighting.

An example throw is shown in Figure 7.6. These distributions are given to the fitter as an input toy data set.

7.3.3 Perform the Fit

The fake data set is processed and the full fit is performed. The CC-1-pion and CC-other samples are simultaneously fit for flux, detector, cross section, multiplicity and free pion parameters. This produces a post-fit covariance matrix and best-fit values for each fit parameter. An example fit result is shown in Figure 7.7, where the stacked histogram is the fit result, and the throw is shown for reference. The imperfections in the fit result are easier to see in the CC-other sample.

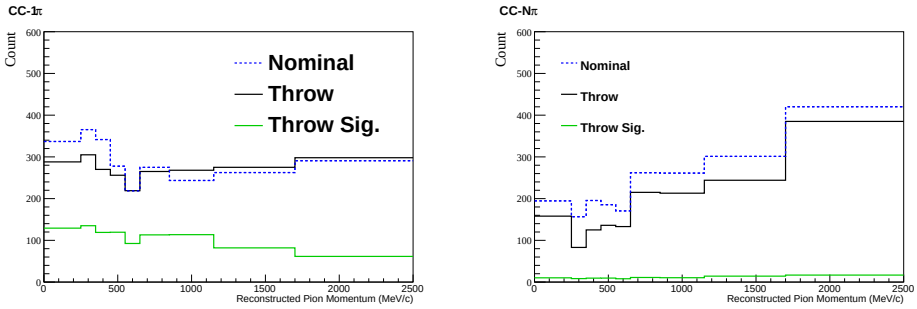

Figure 7.6: An example throw. In this case, all parameters were thrown. The original nominal distribution is shown, and the variation away from is also shown, along with the signal in this throw.

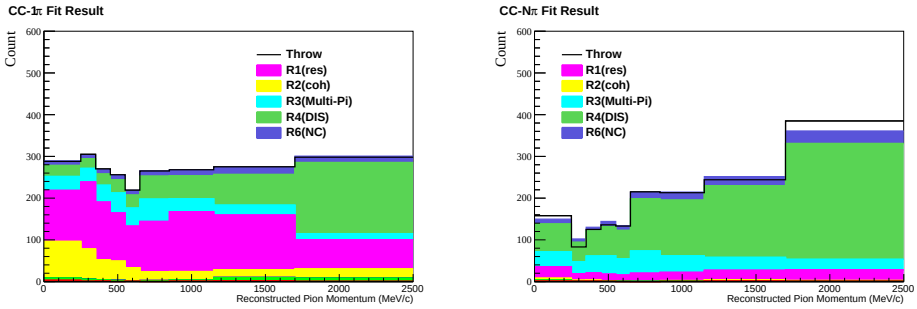

Figure 7.7: An example fit result to a toy data set. The toy data set is shown as the "throw," and the stacked histogram is the best fit result.

7.3.4 Post-fit Throwing

The final goal is to assess the total error on the pion parameters themselves, as these will reweight the nominal cross section. As there is a complex interplay between the pion parameters and the constrained parameters, simply taking the post-fit error on the pion parameters directly is insufficient. The event rate in a given momentum bin can change due to modification of the "nuisance" parameters. For example, if the ν_μ flux parameters were all shifted, the event rate would shift and the pion parameters would adjust to account for *some* of this (in a non-trivial way). Thus a method to disentangle the change in event rate due to the nuisance parameters from the change in event rate due to the cross section and

pion parameters must be applied. This method will now be described.

A set of throws, constrained by the post-fit covariance, with all nuisance parameters set to the post-fit values is made. The conditions for the first set of throws is:

- Throw flux parameters.
- Throw detector parameters.
- Throw multiplicity parameters
- **Fix** the cross section parameters at the **nominal** values
- **Fix** the pion parameters at the **nominal** values

This will produce a set of histograms where the bin content for each throw accommodates the variations induced by the nuisance parameters, i.e. this is a measure of deviation from the nominal that can be attributed to parameters unrelated to the CC-1-pion cross section. A set of 500 such throws is made, call them **Nuisance Throws**, producing a set histograms $\{N\}$.

A second set of throws is made with all of the parameters. This is similar to the first set of throws, except now the cross section and pion parameters are also thrown:

- Throw flux parameters.
- Throw detector parameters.
- Throw multiplicity parameters
- **Throw** the cross section parameters at the post-fit values
- **Throw** the pion parameters at the post-fit values

Call these throws the **Cross Section Throws**, which produce a set of 500 histograms $\{C\}$.

The result of one nuisance throw and one cross section throw is shown in Figure 7.8 (right) (their application is discussed in the next section). In this case the nuisance parameters can be seen to account for only a small portion of the variation in the rate in each bin, with a majority of the change coming from the cross section parameters. Also in Figure 7.8 (left), the thrown rate for this example and the best post-fit estimate are shown. Also shown is the result of fitting a gaussian to the distribution of the 500 cross section throws in each bin, referred to as the post-fit marginalization.

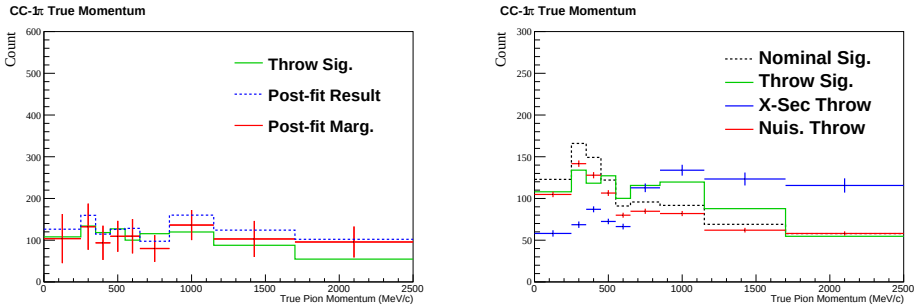

Figure 7.8: Post-fit results on the left, with the nominal CC-1-pion component, the throw, the post-fit result, and result of fitting a gaussian to 500 cross section throws (all parameters) in each bin. The result of one nuisance throw and one cross section throw on the right. The interpretation of the results in the right-hand plot are as follows: Examining the first bin, we see the nominal signal (black-dashed) is decreased by this throw (green). The effect of all the post-fit parameters that are not cross-section related are shown in red, these are the "nuisance" throws and in this case, in the first bin, the nuisance parameters pull the nominal down to just below the thrown signal. However, when the cross section parameters are included the prediction in the first bin goes down significantly (blue). The change in the prediction from the nuisance throws to the cross section throws represents the effect of *only* the cross section parameters in a given bin. After performing 500 throws similar to this example, the effect of the cross section parameters on each bin is extracted by taking the ratio of the cross section to the nuisance throw bin contents.

7.4 Extract CC-1-pion Normalization

At this point two sets of histograms have been made: nuisance throws N_t and cross section throws C_t, where t is the throw index $t = (1, 500)$. An example of one nuisance and one cross section throw is given in Figure 7.8 (right) . The CC-1-pion sample has 9 bins. For

each bin and each throw, the distribution of the ratios of the cross section throws to the nuisance throws is formed: $X_j = \{(C_{t,j}/N_{t,j})\}$, where j is the momentum bin index. The central value and width of the distribution X_j for each bin will indicate the scale and error on the CC-1-pion cross section adjustment to be applied to the NEUT nominal cross section. For the example throw presented here the distributions for each X_j are shown in Figure 7.9. The difference between taking the post-fit CC-1-pion normalization parameters and the extraction method presented here is shown in Table 7.5.

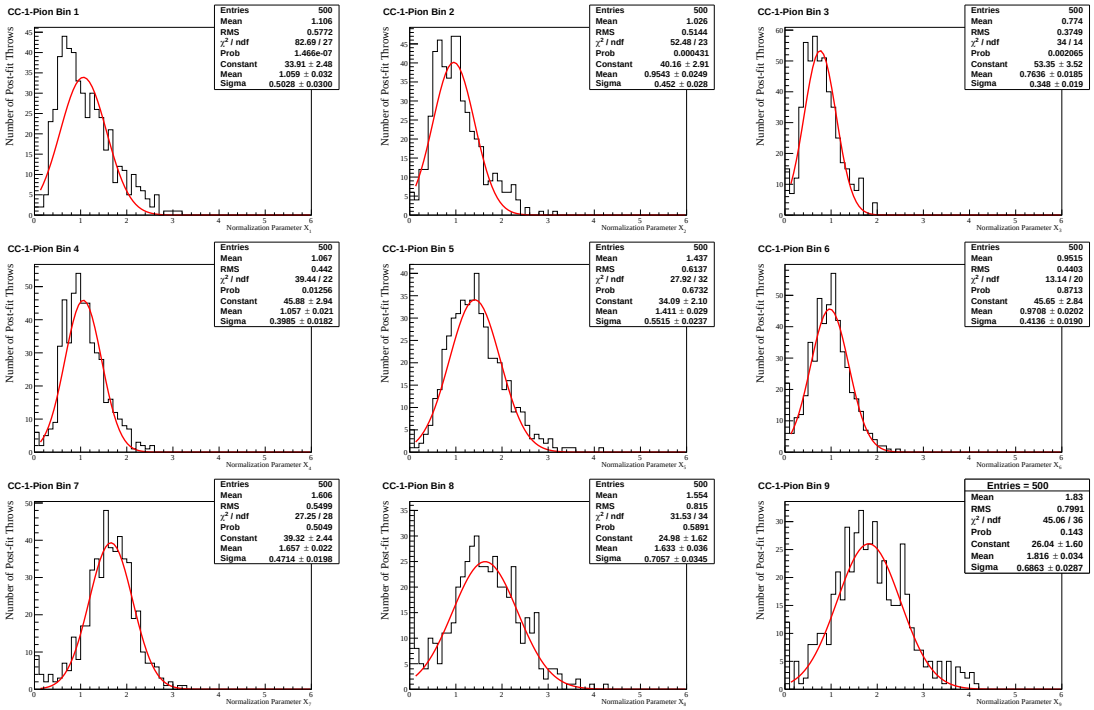

Figure 7.9: Distributions of the CC-1-pion normalization parameters X_j which are extracted from forming the ratios of the nuisance and cross section throws..

7.5 Application to NEUT Nominal

Next, a ν_μ interaction sample on a carbon is generated in NEUT using the official T2K flux. The output neutrino interactions are binned in two dimensions: in parent neutrino energy E_ν versus pion momentum. The binning in parent neutrino energy will allow reweighting

Bin	Fit CC-1π Norm. Parameter	Post-fit Extracted Norm. Parameter
Par 1	1.23 ± 0.40	1.06 ± 0.50
Par 2	1.31 ± 0.52	0.95 ± 0.45
Par 3	1.11 ± 0.56	0.76 ± 0.35
Par 4	1.25 ± 0.50	1.06 ± 0.40
Par 5	1.67 ± 0.65	1.41 ± 0.55
Par 6	1.20 ± 0.67	0.97 ± 0.41
Par 7	2.09 ± 1.00	1.66 ± 0.47
Par 8	2.22 ± 1.45	1.63 ± 0.71
Par 9	2.03 ± 0.95	1.82 ± 0.69

Table 7.5: An example comparison between the fitter's free CC-1-pion parameters and the CC-1-pion normalizations extracted by post-fit processing. There extracted normalization parameters exhibit possibilities of oscillation features which may require regularization methods, but more notable is the reduction in error.

of the flux bins with the post-fit flux parameters. The selection criteria for the observable CC-1-pion topology definition is:

- one and only one μ^-
- one and only one π^+
- zero π^0, zero π^-
- any number of neutrons and protons and other particles

The output of NEUT is in Figure 7.10. This sample was generated with $N = 500,000$ events and the T2K flux. The post-fit flux parameters can be applied to each NEUT bin to adjust the expected number of events in the pion momentum projection. For the nominal 2-D NEUT distribution, H^{NEUT}, the flux-reweighted projection into momentum bins B_j is

$$B_j = \sum_{i}^{FluxBins} b_i \times H_{j,i}^{NEUT}, \tag{7.9}$$

where the sum is over neutrino flux bins with flux normalization parameters b_i. The nominal result (all b parameters equal to 1.0) is shown in Figure 7.11 (left). After this, the CC-1-pion normalization parameters X_j (Section 7.4) can be applied to their respective bins to obtain the reweighted differential cross section,

$$\begin{aligned} \frac{d\sigma_j}{dp_\pi} &= X_j \times B_j \\ &= X_j \times \frac{d\sigma_j^{nom}}{dp_\pi} \end{aligned} \tag{7.10}$$

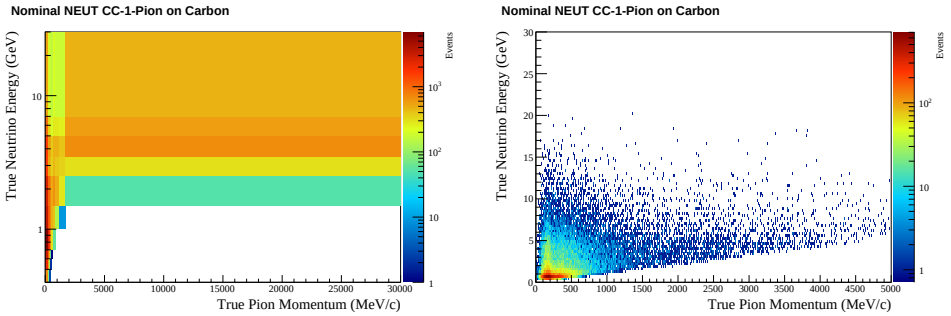

Figure 7.10: NEUT interaction generator's nominal output for the observed CC-1-pion cross section, as a function of the pion momentum and parent neutrino energy. This was generated from a sample of 500,000 ν_μ interactions using the T2K flux. The binning on the left is the neutrino binning used in the flux parameters, and the pion momentum binning used in the pion parameters. The output on the right is in a fine binning for an alternative view of the phase space.

Taking the extracted normalization parameters for this fake data set (Table 7.5) and applying them in conjunction with the post-fit flux parameters as described in Equations 7.9 and 7.10, the cross section measurement can be produced. The final result is shown in Figure 7.12.

7.6 Validation Studies

The typical method for validating this type of fitting procedure to examine what are called "pull distributions." This is done by examining the result of many fits to fake data sets called "toy experiments."

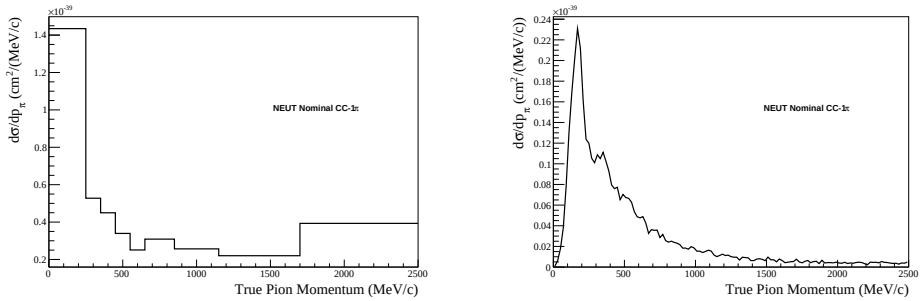

Figure 7.11: The nominal NEUT output for ν_μ CC-1-pion interactions on a carbon target, in the pion momentum binning used in this analysis (overflow is included in the last bin) and in a fine binning (20 MeV) that is drawn smoothly.

A pull distribution for each parameter is made by examining the difference of the post-fit parameter value from the true parameter value. For a given fit parameter τ, the prior constraint gives a nominal value, τ_{nom} and is constrained with a prior error σ_{prior}. The post-fit value of the parameter is denoted τ_{fit}, with post-fit error on the parameter σ_{fit}. One can then form the pull, g:

$$g = \frac{\tau_{fit} - \tau_{nom}}{\sqrt{\sigma^2_{prior} - \sigma^2_{fit}}}. \tag{7.11}$$

The denominator is the error on the numerator taking into account the correlation between errors in the fit result and constraint [80]. Similarly, for an unconstrained (i.e. free) parameter, the pull is defined as:

$$g = \frac{\tau_{fit} - \tau_{nom}}{\sigma_{fit}}. \tag{7.12}$$

In the case of the pion parameters, for example, the nominal values are all equal to 1.0, as we assume that our model (NEUT) has a valid CC-1-pion cross section that does not need adjustment.

After performing many parameter throws and examining the pull distribution of these "toy fits," one can construct the distribution of pulls g for each parameter. For an unbiased fit the pull distribution should be a gaussian centered at 0 with a width of 1.

As discussed in Section 7.3.4, the free pion parameters incorporate some of the "movement" of the constrained parameters. Thus the pull distributions for the constrained pa-

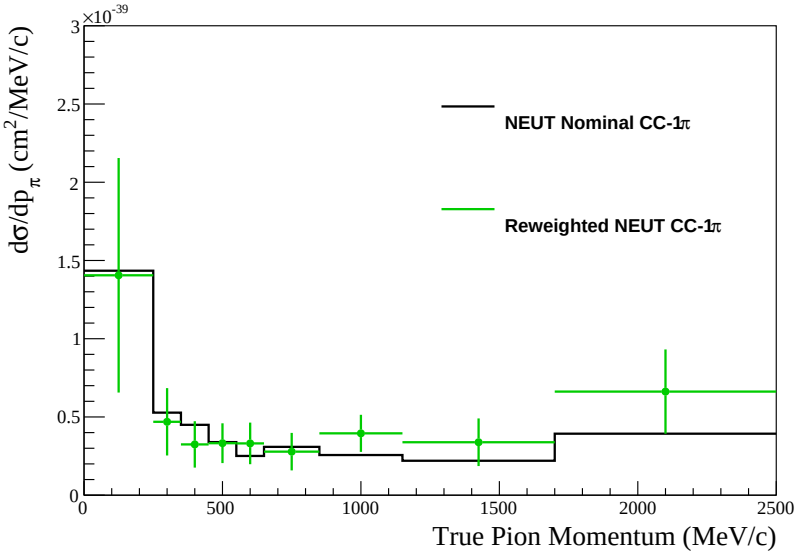

Figure 7.12: The NEUT nominal output overlaid with the final cross section measurement. In this example a fake data set was used.

rameters is expected to be very narrow. The pull distributions of the free pion parameters are examined, as well as the pull distribution of the post-fit marginalization.

7.6.1 Pion Parameter Pulls From Fits With All Parameters

A set of 100 toy experiments was generated with all fit parameters included and analyzed. The pull distributions for the free pion parameter corresponding to each bin in true pion momentum are shown in Figure 7.13. The fitted $\Delta\chi^2$ is shown in Figure 7.15.

For the most part these pull distributions are not bad. Some parameters exhibit bias (not centered at zero), and their widths are a little bit too broad. However, it was found that for 24 of these 100 toy experiments at least one pion parameter was ill-behaved, fixing itself at the lower bound of zero. The exact cause that induces the bad fitting behavior is still under study. Some additional comments and some examples of broken fits are presented in Section 7.7.

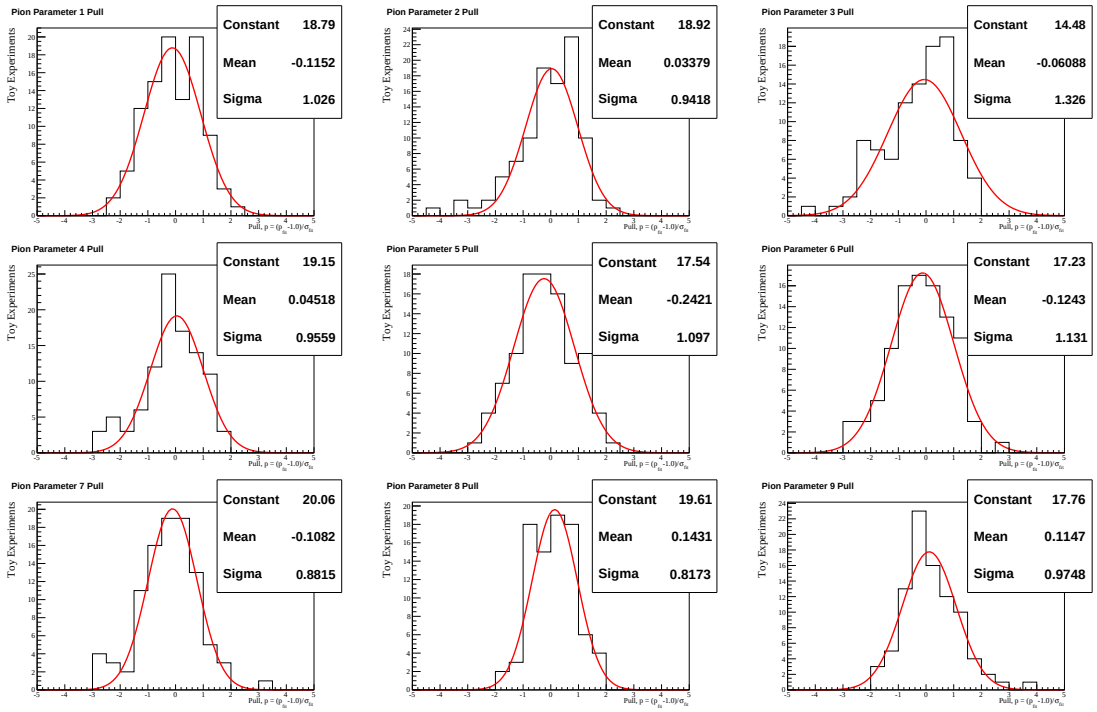

Figure 7.13: Pion parameter pull distributions for toy experiments with all fit parameters thrown. Each parameter corresponds to one bin in true pion momentum space. The pull definition is given in Equation 7.12. For a good fit the distributions should be centered at zero with a gaussian width of 1.0.

7.6.2 Pion Parameter Pulls From Fits With Fixed Cross Section Parameters

As it was known that the cross section parameters can introduce large variations and possible degeneracies with the pion parameters, another set of 100 toy experiments was processed with throws of all but the cross section-related parameters. In this case, 98 of the 100 toy experiments finished without any poor behavior in the fit parameters. The pull distributions for the free pion parameter corresponding to each bin in true pion momentum are shown in Figure 7.14. The fitted $\Delta\chi^2$ is shown in Figure 7.16.

7.6.3 Comment on the Fitted $\Delta\chi^2$

The fit performed here is between two samples, the CC-1-pion sample and the CC-other sample. Each sample has 9 bins in pion momentum, for a total of 18 bins being fit. However,

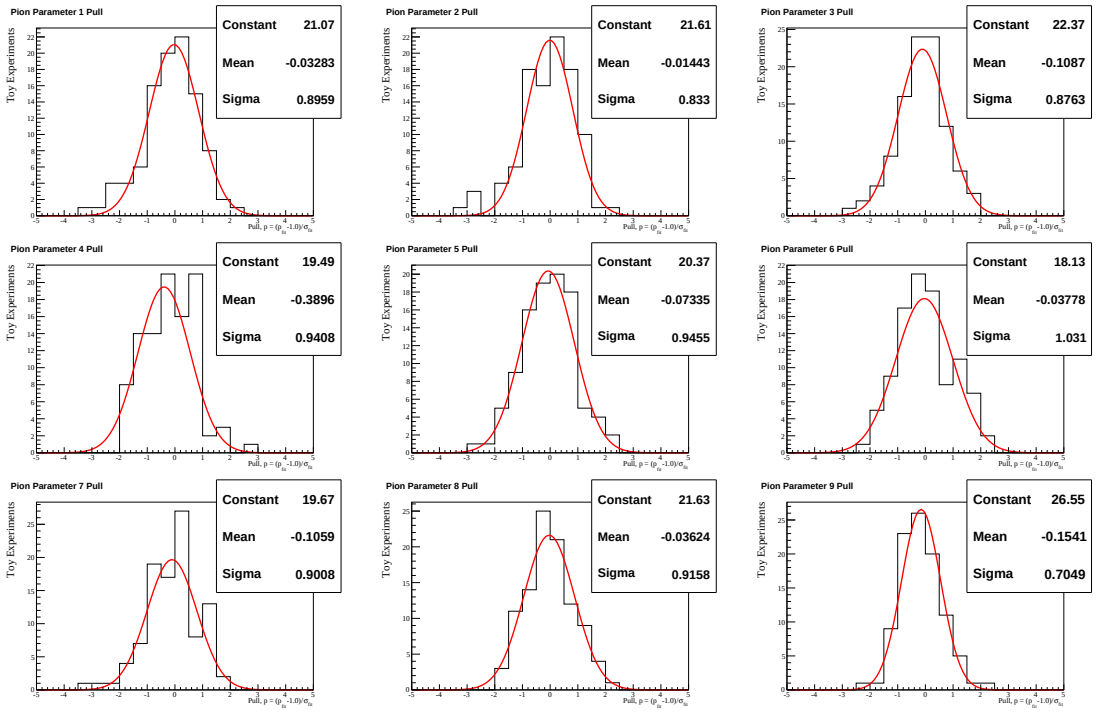

Figure 7.14: Pion parameter pull distributions for toy experiments *without* the cross section fit parameters thrown. Each parameter corresponds to one bin in true pion momentum space. The pull definition is given in Equation 7.12. For a good fit the distributions should be centered at zero with a gaussian width of 1.0.

the fit description gives a total of 91 parameters: 25 flux parameters, 18 detector parameters, 18 multiplicity parameters, 21 cross section parameters and 9 free pion parameters.

All but the free pion parameters come with prior constraints, which is encoded in the associated covariance matrices used to constrain the parameters. As the constraints are derived from additional external data (e.g. various control samples used in generating the detector systematic parameters), they come with additional degrees of freedom. Thus for each constrained fit parameter (i.e. flux, detector, multiplicity and cross section parameters), a degree of freedom is removed by the parameter but then a degree of freedom is

added by the constraining data. Our expected $\Delta\chi^2$ is thus

$$\begin{aligned} \Delta\chi^2_{expected} &= N_{bins} - N_{parameters} + N_{constraints} \\ &= 18 - 91 + 82 = 9 \end{aligned} \tag{7.13}$$

$$\tag{7.14}$$

The fitted $\Delta\chi^2$ are shown in Figures 7.15 and 7.16, for 100 toy experiments with all fit parameters and 100 toy experiments with all but the cross section parameters, respectively. Both are fit with a $\Delta\chi^2 \sim 8$, which is close to expectations.

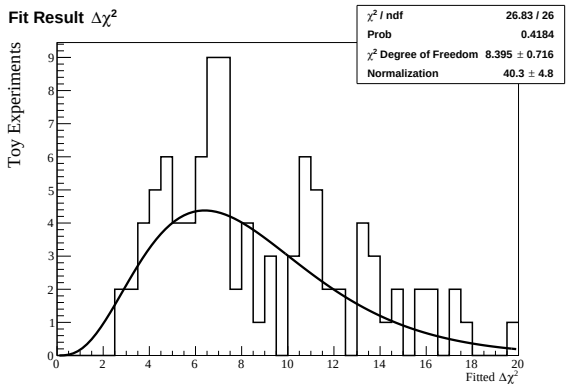

Figure 7.15: Fitted $\Delta\chi^2$ distribution for 100 toy experiments with all fit parameters included.

7.6.4 Fit Pulls With All Parameters

The post-fit marginalization process uses the post-fit covariance matrix to do correlated throws of the fit parameters. This way an ensemble of fake data sets can be generated from which the distribution mean and width of the signal in each bin can be measured. The width given by this process should represent the total error due to all parameters and their correlations. An example of this is shown in Figure 7.8.

For a given throw, denote the true signal in the i^{th} bin by S_i, and the post-fit marginalization generated from 500 post-fit throws as having fit value f_i and width σ_i. Then the for

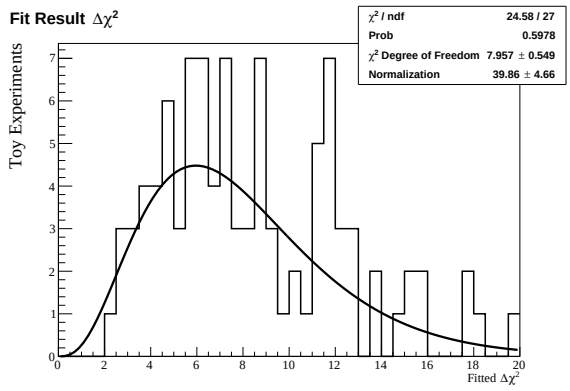

Figure 7.16: Fitted $\Delta\chi^2$ distribution for 100 toy experiments *without* the cross section fit parameters included.

each bin can be calculated by

$$p_i = \frac{f_i - S_i}{\sigma_i}. \tag{7.15}$$

This is the fit accuracy measured in deviations of the fit parameter error. Such a pull can be formed for each bin for each toy experiment. The distribution of these pulls for 100 toy experiments with all fit parameters used is shown in Figure 7.17. The pulls for almost every bin exhibit a slightly negative bias, with slightly narrow widths.

7.6.5 Fit Pulls Without Cross Section Parameters

As was the case in investigating the source of the ill behavior of the free pion parameters, the first candidate was to remove the cross section parameters, which are known to known to introduce large variations in the fake data set generation. In the case of the fit pulls the negative bias and moderately narrow pull width seems to persist, as shown in Figure 7.18. The exact cause of the negative bias and narrow width is under investigation[2].

[2]The persistence of the negative bias between fits with and without cross section parameters makes it tempting to speculate that this is more likely the result of a bug more so than a feature of the fit method.

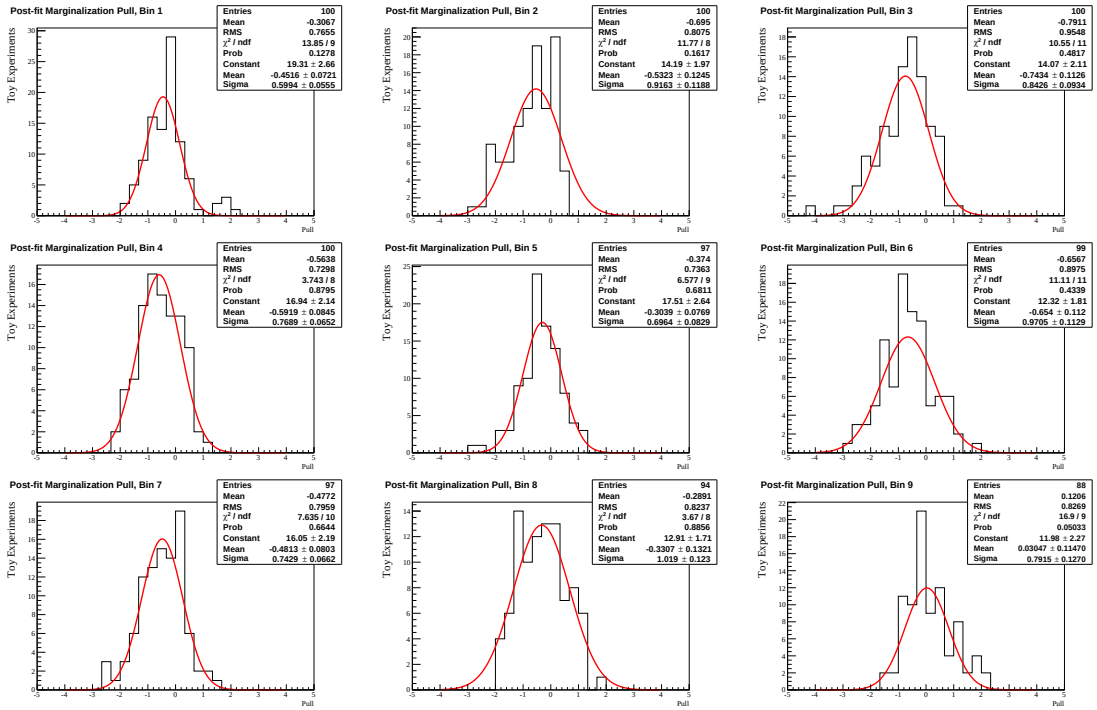

Figure 7.17: Post-fit marginalization pulls for each bin, calculated by Equation 7.15, for 100 toy experiments using all fit parameters.

7.7 Bad Fit Examples

There are two distinct modes of fit failure. They are both defined by the behavior of the free pion parameters which are ultimately used to make the cross section measurement. In one mode, one or multiple pion parameters are forced to a value of zero, the lower bound. This is equivalent to the fitter setting the signal in that bin in true pion momentum space to zero. An example of such a throw is given in Figure 7.19.

This example is representative in that most fits that break in this way end up with the pion parameter for the final bin being fit to zero. The final bin has the poorest signal to background ratio and extends from $1,700 - 30,000$ MeV. It is arguable that as the CC-1-pion cross section becomes exceedingly small above $1,500$ MeV (Figure 7.11) that this parameter should either be fixed or re-binned.

The second mode of failure is when the pion parameters are well-defined in the sense

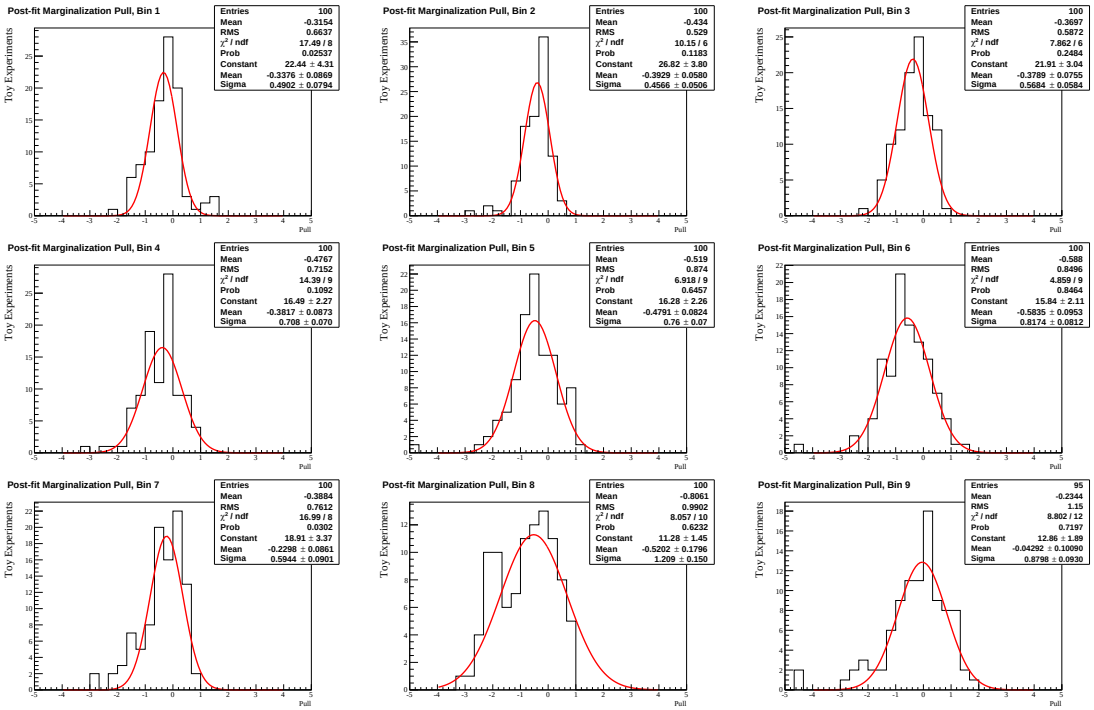

Figure 7.18: Post-fit marginalization pulls for each bin, calculated by Equation 7.15, for 100 toy experiments generated without cross section parameters thrown and fit with the cross section parameters fixed.

that they are not pushed against the parameter lower bound, but the collective behavior is suspect. More precisely, oscillatory behavior which is indicative of over-fitting is seen. In this situation one parameter can be fit to a very high value, while the next is fit to a very low value, and the anti-correlated pair from true phase space form a valid prediction in the smeared reconstructed space. This type of problem is discussed in detail in Appendix A. Over-fitting of this type has a known solution, namely regularization. An example fit that induces oscillatory pion parameters is shown in Figure 7.20

Large sets of fake data sets have been generated and fit. The exact origins of the ill-behaved fits are still under investigation, but there are a few likely candidates. The currently implemented cross section parameters were originally designed for use in oscillation analyses and thus have conservative error estimates. One repercussion of this is that the fake data sets generated with throws of these parameters sometimes end up with radical variations in the

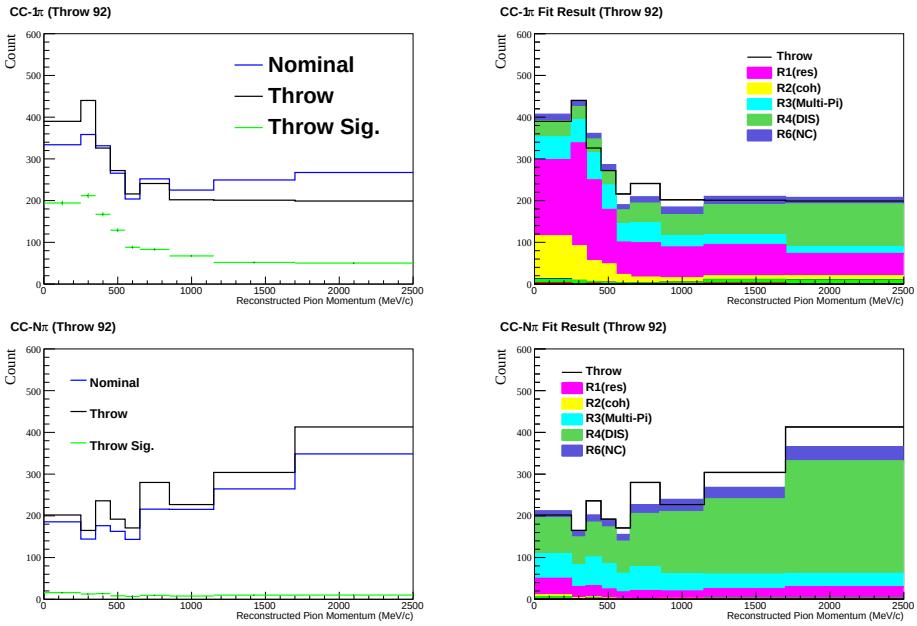

Figure 7.19: An example of a throw that breaks the fitter's free pion parameter. In this case, the pion parameter corresponding to the last true pion momentum bin was fit to zero.

nominal (e.g. overall normalizations changing by more than 100%). For a model-dependent measurement, validating the analysis methods against distributions that are wildly different than the nominal is questionable – it's certainly ideal that the fit method has no known failure modes, but if simply *looking* at the model and being able to tell it is patently wrong, then the fit results in this case hold little value to begin with.

An additional source of potential issues is that the cross section parameters are almost all entirely independent. It is observed that a majority of broken fits result from data sets where multiple parameters are thrown to extreme values in either an anti-correlated or correlated fashion, again contributing to radical departures from the nominal. It is currently under consideration to limit the cross section parameter ranges in the generation of the fake data sets to assess the fit behavior under more representative conditions.

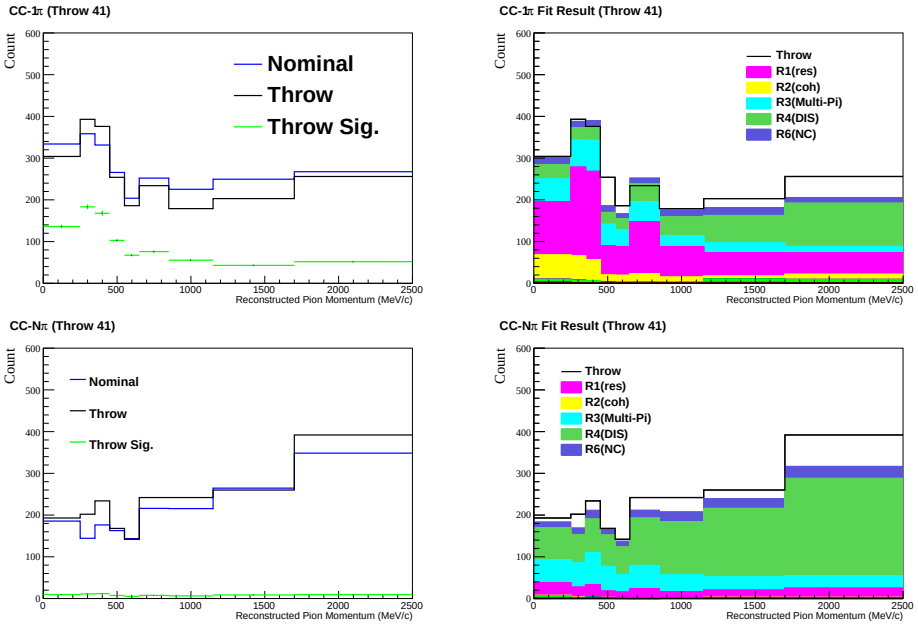

Figure 7.20: An example of a throw that breaks the fitter's free pion parameter by inducing extreme anti-correlations. This is most likely due to CC-other sample (bottom row) having pairs of bins that sequentially have large deviations from the nominal followed by nearly matching the nominal.

Chapter 8

SUMMARY AND CONCLUSIONS

This thesis has focused on the experimental aspects of contemporary neutrino cross section measurements. A description of the T2K experiment was given with particular attention given to the near detector complex (ND280) in Tokai where the produced neutrino beam is sampled prior to oscillation.

A ν_μ CC-inclusive event selection method based in ND280's tracker, which is a combination of scintillator-based fine-grained detectors and time projection chambers was presented. This event selection was categorized into single-pion (CC-1-pion) and multi-pion (CC-other) production channels, forming the two primary data samples to be utilized in the cross section measurement method under development for this thesis. The development of unique software tools, analysis and contribution of a novel systematic uncertainty which assessed the efficiencies of reconstructing short tracks within multi-track event topologies was contributed to this event selection procedure.

Due to the significant CC-other background in the CC-1-pion sample a significant effort was put into developing a CC-other related systematic error. This systematic was applied to the uncertainty on the multiplicity of charged pions, which until this point has not been characterized for any analysis in T2K. The development of this work necessitated the digitization and analyses of historic bubble chamber data, the creation of a pion multiplicity model and implementation of a method for reweighting the standard Monte Carlo.

The above-mentioned endeavors were brought together in a simultaneous fit of the CC-1-pion and CC-other data samples. This fit was designed to extract normalizations of the CC-1-pion cross section in true pion momentum phase space relative to the official T2K model (NEUT) predictions. This fit expands upon an implementation of the event selection and systematic reweighting software interface to develop a ν_μ CC-1-pion cross section measurement method. The components of the fit software and analysis methods are now

implemented and undergoing validation studies prior to performing a fit to the data.

The validation studies to date have suggested that additional studies on the prior errors attributed to some fit parameters, namely the cross section parameters, is necessary. These parameters adopted prior values from the T2K collaboration's working groups that focused on oscillation analyses, wherein conservative (i.e. very large) errors were assigned. Also under consideration is the application of regularization methods in the fit. Regularization was studied and implemented within toy fitting routines as a proof of principle. A regularization implementation was included in the official fit software, but is not currently utilized. If the on-going validation studies indicate regularization may be necessary, additional computational time for validation of the regularized fits will be necessary.

To make a measurement many stages and many diverse contributions are necessary. From the hardware side one begins with the detectors. This involves design, construction, commissioning and operations management. On the software side there is the on-going data acquisitions, calibrations and quality assessments. There is also the development of standardized tools and, ultimately, analysis methods. During my time on the T2K experiment, I have had the opportunity to partake in almost every one of these stages. Those components most pertinent to the development of the measurement itself, namely the data selection, systematic studies, fit methods and validations have been presented in detail within this thesis.

Appendix A

TOY FITTING

This chapter will endeavor to discuss the process of comparing our measurement model with the observed data. In colloquial terms this is referred to as *fitting*. The general goal of an experiment is to measure some distribution $f(x)$ of an observable x (such as pion momentum) and to then make inferences from and comparisons to the observed distribution. The observed distribution is usually binned into a histogram and differs from the true distribution due to statistical and systematic errors. If there is a theoretical prediction for the true distribution which depends on some parameters a, $f_{th}(x|a)$, the best values of the parameters a can be extracted by the fitting process [81].

The implementation of the fit process is rather complex, so first a toy model with all of the elements of the full data fit will be presented in stages. The goal will be to introduce the likelihood function used and focus on the concept of inferring true parameter and kinematic values from observed measurements. Regarding this latter point, the discussion will encounter the problem of over-fitting, which induces unphysical behavior in fit results, and how the process of regularization is applied to mitigate such effects.

A.1 The Likelihood Function

A.1.1 The Principle of Maximum Likelihood

The following section is based on the presentation of the method of maximum likelihood by Cowan [82].

In the most general formation, we can speak of measuring a random variable x which we suppose has values distributed according to a fixed probability. The function that describes that the probability of x having a specific value is referred to as the *probability density function*, or p.d.f., and can depend on one or more parameters [1]. The p.d.f. for a given

[1] If the p.d.f. has no parameters then it is a constant function, with an equal probability for all observa-

variable x which depends on a set of m parameters $\Theta = \{\theta_1, ..., \theta_n\}$ can be written as $f(x|\Theta)$.

With an assumed p.d.f., the probability that of an observed measurement of the variable x, call it x_1, has a value between the range $[x_1, x_1 + dx_1]$ is $f(x_1|\Theta)dx_1$. We can imagine that after an experiment is performed we have a set of n fixed measurements $\mathbf{x} = \{x_1, ..., x_n\}$, to each of which a probability can be assigned, and when taken together an estimate of the total probability for an observed set of data can be calculated as

$$\text{Joint Probability}(\mathbf{x}|\Theta) = \prod_{i=1}^{n} f(x_i|\Theta)dx_i. \tag{A.1}$$

In the application of the p.d.f. the values for the parameters Θ are also assigned. If the p.d.f. and the parameters Θ are correct then we expect the probability for an individual measurement to be high, whereas if either the p.d.f. or the parameter values assigned are incorrect then the probability will be small. As the intervals dx_i do not depend on the parameters Θ, we can define the *likelihood function* L, which shares these features, as

$$L(\Theta) = \prod_{i=1}^{n} f(x_i|\Theta). \tag{A.2}$$

The likelihood function can be viewed as a function of the parameters Θ only as the data from the experiment is fixed. The goal is to estimate the parameters of the p.d.f. under the assumption that the true values give the greatest agreement with the distribution of the observed data and thus the largest probability.

The likelihood function in equation A.2 can be made more amenable to calculation if the logarithm is taken, which converts the product into a sum,

$$L(\Theta) = \sum_{i=1}^{n} \log f(x_i|\Theta). \tag{A.3}$$

The logarithm is a monotonic increasing function, so any maximum value of $L(\Theta)$ will be shared by $\log L(\Theta)$. Additionally, any constant terms can be dropped as they will not influence the likelihood as the parameters are adjusted.

The quantities described here are probabilities, thus they will take on a value greater than or equal to zero and less than or equal to one, meaning the logarithm of the likelihood

---

tions.

will be in the range $(-\infty, 0]$. The principle of maximum likelihood, in what is a bit of a verbal jumble, prescribes that we maximize the likelihood function $L(\Theta)$, or *minimize* the negative log-likelihood function $-\log L(\Theta)$. It is the job of the curious to go about actually *finding* the extremum, which is a process that will be discussed in later sections[reference to actual section].

A.1.2 The Likelihood Function

In the case of a neutrino cross section measurements it is often the case that the analyzer deals with binned data. The experiment collects data on events that pass certain selection cuts and the observables of interest are collected in bins of width limited by detector resolution or statistics. Additional consideration for a minimum desired event rate that may require a width larger than the detector resolution.

The function of an individual bin for an observable is to report the number of events that occur in a given range over a given amount of time. As these are natural events that we assume occur at a constant rate under fixed conditions we expect that an individual bin is governed by Poisson statistics. The Poisson distribution,

$$P(r|\mu) = e^{-\mu}\frac{\mu^r}{r!} \tag{A.4}$$

gives the probability to observe a number of events r in a given time interval. The Poisson distribution has the property that the single parameter μ represents both the mean and the variance of the distribution.

In designing a likelihood function for an observable measurement that is binned with the expectation of individual bin contents governed by Poisson statistics, the single parameter μ represents our best estimate (produced by a model) for the number of events we expect to find in that bin and can be denoted b, for $b$in. Similarly, the number of data events measured in a given bin can be denoted d. One can thus define the likelihood for a single bin as the Poisson distribution in equation A.4 with the substitution $\mu \to b$ and $r \to d$. Assuming we have a model-based prediction for the expected number of events in each bin, the total, or "joint" likelihood for a set of data $\mathbf{d} = \{d_1, ..., d_n\}$ and predicted bin contents

$\mathbf{b} = \{b_1, ..., b_i\}$ can be formed, in accordance with equation A.2:

$$L(\mathbf{b}) = \prod_{i=1}^{n} e^{-b_i} \frac{b_i^{d_i}}{d_i!}. \tag{A.5}$$

Working with the negative log-likelihood function and applying Stirling's approximation[2]

$$\begin{aligned} -\log L(\mathbf{b}) &= -\sum_{i=1}^{n} \log\left(e^{-b_i} \frac{b_i^{d_i}}{d_i!}\right) \\ &= \sum_{i=1}^{n} b_i - d_i \log b_i + \log d_i! \\ &\approx \sum_{i=1}^{n} b_i - d_i \log b_i + (d_i \log d_i - d_i) \\ &= \sum_{i=1}^{n} (b_i - d_i) + d_i \log\left(\frac{d_i}{b_i}\right). \end{aligned} \tag{A.6}$$

Equation A.6 is the the simplest form of a negative log likelihood function for a histogrammed set of data $\mathbf{d}$ where the *parameters* are the predicted bin contents $\mathbf{b}$. Typically when one thinks of a model parameter or parameters one might think of an ejection angle θ or an incident energy E rather than the number of events in a bin. This point deserves a comment.

The calculation of the number of events expected to be found in a histogram is very complex. The entirety of all the processes, starting from proton beam interactions which create the neutrino flux, to neutrino interactions on nucleon targets through secondary particle interactions and detector response simulations are simulated and processed in a way identical to processing real data events. Each stage of simulation involves many models and many parameters that might be conceived of in the more traditional sense. The culmination of simulation and the processing of fake data in the same manner which real data is processed is an "observed" histogram with bin contents that represent our best estimate of *all* models involved in the simulation. We can adjust the more salient model parameters in the simulation process and re-estimate the expected outcome, and the effect of any parameter modification will manifest itself in the adjustment of the predicted bin contents

[2]Alternatively, the logarithm of the factorial can be dealt with by maximizing the likelihood in the form of a ratio relative to the nominal prediction $\hat{\mathbf{b}}$, $L(\mathbf{b})/L(\hat{\mathbf{b}})$.

at the end of the data analysis process. In this way some simulation parameters can be varied and their net effect measured, and as the maximum likelihood search proceeds the best parameter estimates can be found. The likelihood function serves as a mechanism to guide the assessment of our prediction versus the observed data. For a sense of assurance one can see that if the prediction **b** is equal to the data **d** then equation A.6 will be at the minimum value zero.

It should be noted that the likelihood function as shown in equation A.2 can have additional terms impacting each bin. Currently each bin has a likelihood derived from a single p.d.f., namely the Poisson function for that bin, $P(d_i|b_i)$. A joint p.d.f. for each individual can be formed by taking the product of the Poisson function and some other p.d.f. that may depend on additional parameters, $f(d_i|\Theta)$. If this is the case and we construct the negative log-likelihood we see find that the joint p.d.f. results in adding terms:

$$\begin{aligned} -\log L(\mathbf{b}) &= -\sum_{i=1}^{n} \log \left(P(d_i|b_i) f(d_i|\Theta) \right) \\ &= -\sum_{i=1}^{n} \log P(d_i|b_i) + \log f(d_i|\Theta) \\ &= -\sum_{i=1}^{n} \text{Poisson Term} + \text{Additional Terms.} \end{aligned} \tag{A.7}$$

Regularization will be introduced as an additional term, and in the real data fit our prior experimental knowledge will be inserted to the likelihood and appear as additional terms, too.

A.2 Toy Model

As discussed in the introduction to this chapter, it is often the case of high-energy physics experiments that there is a theoretical prediction $f_{th}(x|a)$ from which it is desirable to extract the parameter values a by comparison of the model to the measurement of the true distribution $f(x)$. Due to the realistic limitations of detectors, direct and perfect measurements of a variable x is often not possible. In fact, limited resolution and systematic effects can be conceived of in the following way: the target variable x is not measured, but instead a variable y and its associated distribution $g(y)$ is measured. The process of

measurement is a transformation, taking x to y:

$$g(y) = \int_a^b T(y,x) f(x) dx. \tag{A.8}$$

In this equation the true distribution $f(x)$ is defined over the integration range and the function $T(y,x)$ is called a kernel in the theory of integral equations [81]. The kernel describes the entirety of the detector response which transforms the variables x to y in the measurement process. In the case of binned data this equation takes a matrix form and the kernel is often referred to as the transfer matrix.

The situation we now want to consider is that of the case when the theoretical model proposed comes with substantial confidence. The model makes predictions in the true space of the variable of interest x, and the experiment makes measurements in the transformed space y. Ultimately it is desirable to make statements about the true distribution. Parameters $\mathbf{C}$ that scale the model's prediction in the true space are applied, and through detector simulation the adjusted true variables x are then transformed by the operation $T(x,y)$ and the resultant distribution $g(y)$ is compared to the observed data. If the scaling parameters $\mathbf{C}$ remain close to the value of 1.0, then we conclude that the model is accurate. On the other hand, if the scaling parameters have some significant divergence from 1.0 for specific bins or all, then a negative statement about the quality of the model can be made. This process will now be demonstrated with a minimal toy example.

Suppose an experiment is performed measuring a variable x governed by the true distribution $f(x)$, with $0 \leq x \leq 10$, resulting in a set of measurements $\mathbf{x}$ that are histogrammed into bins $\mathbf{b} = \{b_1, ..., b_{10}\}$ of equal bin width, $\Delta B = 1$. For an ideal detector, exact measurement without error would result in an observed histogram exactly equal to $\mathbf{b}$. Let us consider what happens in the case where the detection process introduces "smearing" of the measured value, taking the variable x to some value $y = x + \delta$. In this case δ will be a randomly sampled number from a gaussian distribution centered at zero with a modest width of $\Delta B/5$.

We first suppose that a data set is collected and obeys some model or distribution which we can refer to as "nature." For the sake of exposition, suppose nature, for this variable, is given by a Poisson function with mean $\mu = 4.0$. The process of experimental measurement

transforms nature's distribution via detector effects and statistics and we get a Poisson-like distribution, shown in Figure A.1.

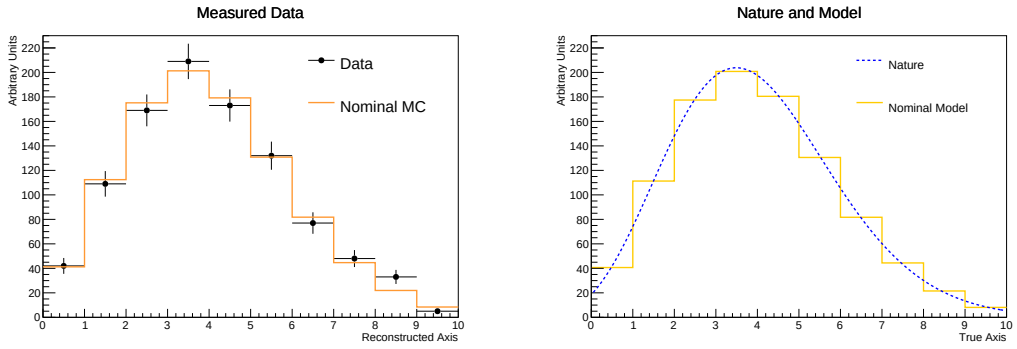

Figure A.1: On the left the the measured toy data and simulated data, both in the reconstructed variable space (after detector smearing). On the right is the distribution which nature follows, a Poisson function with mean $\mu = 4.0$, which for simplicity is also taken to be the predicted model, in the true variable space.

The transfer matrix which shows the transition of the simulated data to the reconstruct space through detector effects is shown in Figure A.2. The transfer matrix is generated from running a large amount of monte carlo (MC) simulation and then normalized to the measured data set. Ideally one produces an order of magnitude or more of MC than data such that the statistical errors are negligible.

The 1-D model predictions shown in Figure A.1 are the projections of the transfer matrix into either the reconstructed or true variable space. The statements of interest will be with regard to the model relative to nature as shown in the true variable space, but the process by which we make those statements involves fitting the nominal model prediction ("Nominal MC") to the observed data.

We introduce one scaling parameter per bin in the true variable space, $\mathbf{C} = \{C_1, ..., C_{10}\}$. The binned results in the true variable space have bin content $\mathbf{b'} = \{b_1 C_1, ..., b_{10} C_{10}\}$. These parameters have no constraints other than the requirement of being positive. At this point it is important to pause and highlight an important subtlety: The fit parameters are applied in the true space, i.e. a single parameter C_i will scale an entire *column* of the transfer matrix shown in Figure A.2. The model prediction which is compared to the data occurs

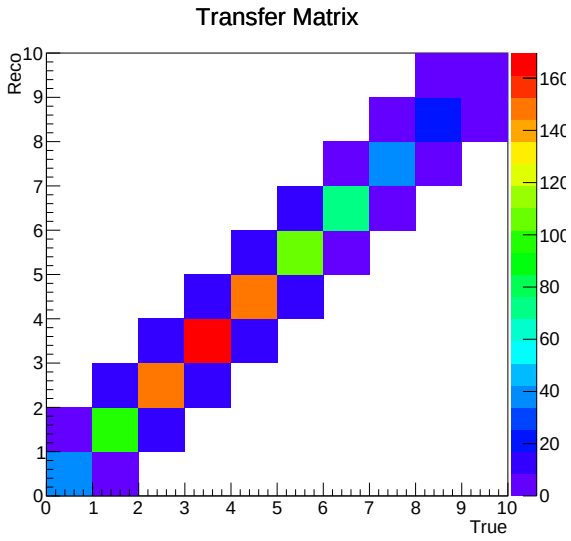

Figure A.2: The transfer matrix for a toy model with reconstructed space variable $y = x+\delta$, where x is the true variable and δ is a gaussian smearing. A projection along the vertical axis gives the true distribution, whereas a projection along the horizontal axis gives the observed, "reconstruction" distribution.

in the *reconstructed* space which is "observed" in an experiment. Therefore the relation between the 10 fit parameters and the 10 data bins is convoluted by the transfer matrix, and the model prediction in an individual observable bin carries dependence of *multiple* fit parameters:

$$b_i^{\text{prediction}} = \sum_{j=1}^{n} T_{i,j} b_j C_j. \tag{A.9}$$

In the case of an ideal detector, the transfer matrix would be exactly diagonal and these fit parameters $\mathbf{C}$ would scale individual bins. In the opposite case, the measurement process smears out events in a true bin uniformly across all measured bins (effectively a counting detector with one bin) and the parameters are degenerate total event rate normalizations. The fit results for the minimal smearing case with both nature and the predicted model being identical is shown in Figure A.3.

The final step in making a statement on the model in the true space is to take the normalization parameters $\mathbf{C}$ and apply them to the model predictions in the true variable

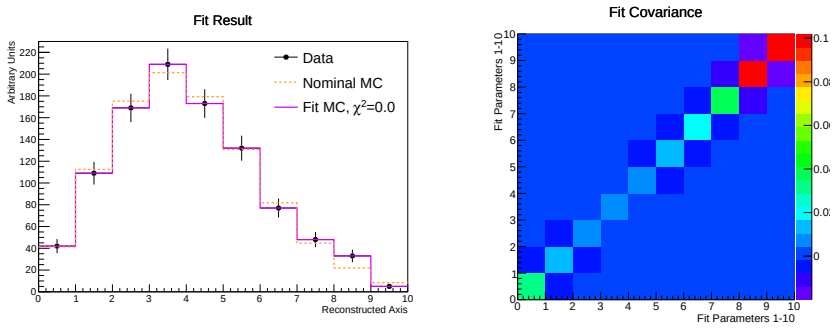

Figure A.3: Fit results for toy with model and nature equal to a Poisson function with mean $\mu = 4.0$. Fit result parameters $\mathbf{C} = \{1.03, 0.97, 0.96, 1.05, 0.95, 1.02, 0.92, 1.07, 1.62, 0.43\}$. The fit covariance shows no abnormal parameter correlations.

space. In this example this is achieved by integrating the model prediction over each bin, and then scaling the result by the associated normalization parameter for that range,

$$M_i = C_i \int_{\Delta B_i} \text{Model}(x) dx. \tag{A.10}$$

Here M_i is the final statement on the model in the i^{th} bin, and C_i are post-fit scaling parameters. The result errors are derived from the post-fit parameter errors. The scaled model result is shown in Figure A.4.

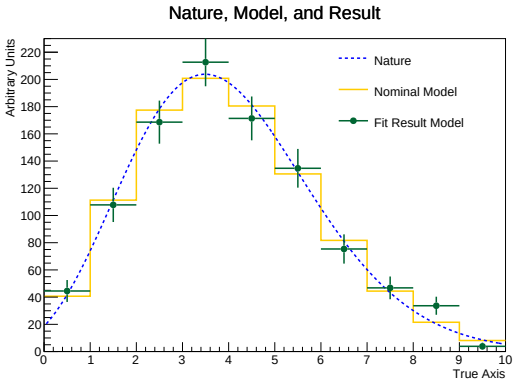

Figure A.4: The final fit result shown in dark teal data points, with each data point representing the model prediction scaled by the fit parameter result for the associated bin.

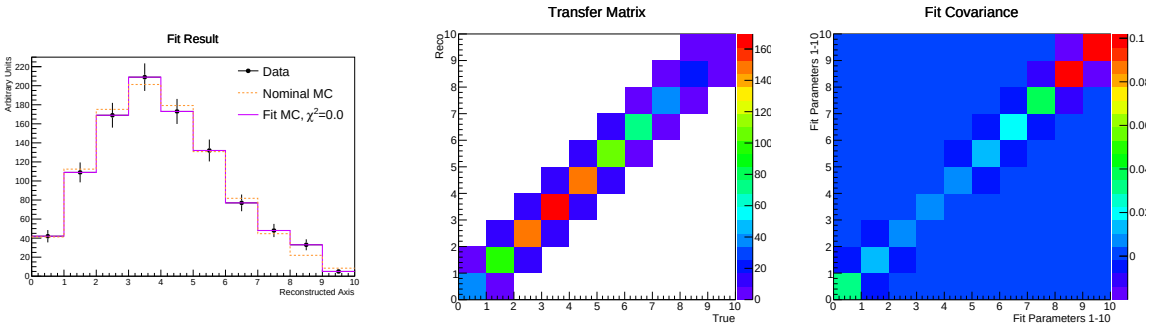

Figure A.5: On the left is a comparison of thrown data, a nominal model prediction, and fit results for gaussian smearing equal to **0.2 times the bin width**. In the middle is the resultant transfer matrix, demonstrating minimal smearing. On the right is the fit parameter covariance matrix. For minimal smearing the parameters are well-behaved.

Proceed with a description of getting to a final model prediction by way of integrating the true distribution with the scaled parameter. End on the note highlighting the reality of unphysical oscillations make the likelihood give a good, or even perfect fit chi2, but that the pathologies manifest themselves in the integrated result. This point necessitates the discussion of regularization.

A.2.1 Regularization

Regularization is the process of smoothing out unphysical oscillations in fit parameters, as was seen in Figure A.6. The general idea is to add another term, R, to the likelihood function that acts as a penalty based on the behavior of the fit parameters. In our examples, the term will be a function of the normalization parameters, $R(\mathbf{C})$.

For example, one can form a vector of the fractional changes in each normalization parameter $\mathbf{C}$, and as the parameters oscillate this norm will become quite large. If such a term is then added to the likelihood with some overall scaling strength α, the oscillations will be penalized.

The regularization strength α will determine the influence of the regularization term. For a very large value of α, the poisson likelihood has very little influence and the regularization dominates the likelihood. For very small values of α the regularization term has a diminished

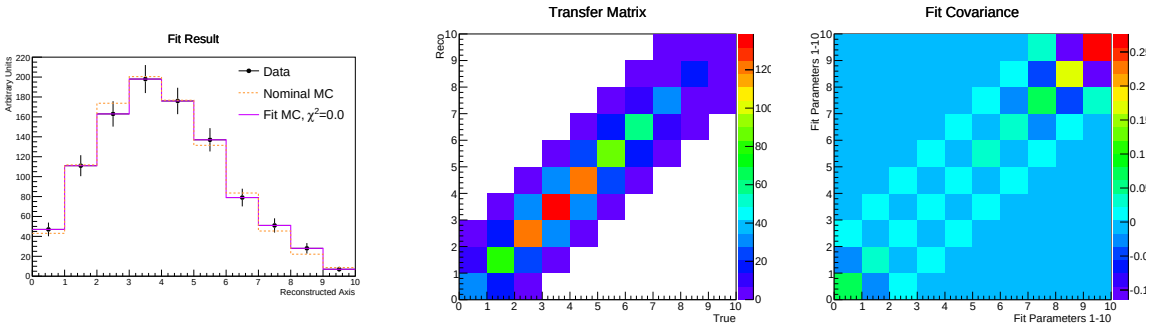

Figure A.6: On the left is a comparison of thrown data, a nominal model prediction, and fit results for gaussian smearing equal to **0.4 times the bin width**. In the middle is the resultant transfer matrix, demonstrating minimal smearing. On the right is the fit parameter covariance matrix. For mid-scale smearing the fit parameters demonstrate anti-correlation pairs.

effect.

Tikhonov [83] regularization is a method of regularization commonly applied. For second order Tikhonov regularization, groups of three adjacent bins are treated and the variation across them is penalized,

$$R(\mathbf{C}) = \sum_{i}^{n-2} (-C_i + 2C_{i+1} - C_{i+2})^2. \tag{A.11}$$

A treatment for various orders and for various bin widths is discussed in both [82, 81]. The post-fit covariance for a toy fit with an extreme smearing applied is shown in Figure A.7, where the regularized covariance has almost no oscillatory correlation in the fit parameters.

A difficulty associated with regularization methods is selecting an appropriate strength α. Various methods are proposed in [82] but were beyond the scope of implementation within these studies.

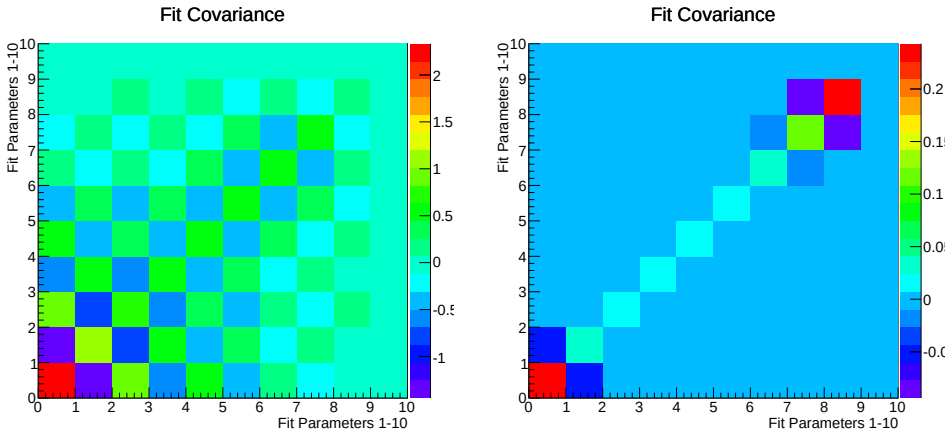

Figure A.7: Example post-fit covariances for a toy fit where the smearing was made very large, purposefully inducing over-fitting. The standard fit's post-fit covariance is at the left, showing wild oscillatory correlations, and the Tikhonov regularized covariance is shown at right, where the oscillations have been quelled.

BIBLIOGRAPHY

[1] Wikimedia Commons. *Standard Model of Elementary Particles*. File: `Standard Model of Elementary Particles.svg`. 2007. URL: `http://en.wikipedia.org/wiki/Standard_Model#mediaviewer/File:Standard_Model_of_Elementary_Particles.svg`.

[2] G. Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1 –29. URL: `http://www.sciencedirect.com/science/article/pii/S037026931200857X`.

[3] S. Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30 –61. URL: `http://www.sciencedirect.com/science/article/pii/S0370269312008581`.

[4] David Griffiths. *Introduction to Elementary Particle Physics*. Second. Wiley, 2008.

[5] C. L. Cowan et al. "Detection of the Free Neutrino: a Confirmation". In: *Science* 124.3212 (1956), pp. 103–104.

[6] Raymond Davis, Don S. Harmer, and Kenneth C. Hoffman. "Search for Neutrinos from the Sun". In: *Phys. Rev. Lett.* 20 (21 1968), pp. 1205–1209. URL: `http://link.aps.org/doi/10.1103/PhysRevLett.20.1205`.

[7] J. N. Abdurashitov et al. "Measurement of the solar neutrino capture rate with gallium metal". In: *Phys. Rev. C* 60 (5 1999), p. 055801. URL: `http://link.aps.org/doi/10.1103/PhysRevC.60.055801`.

[8] P. Anselmann et al. "GALLEX solar neutrino observations: complete results for GALLEX II". In: *Physics Letters B* 357.12 (1995), pp. 237 –247. URL: `http://www.sciencedirect.com/science/article/pii/037026939500897T`.

[9] K. S. Hirata et al. "Results from one thousand days of real-time, directional solar-neutrino data". In: *Phys. Rev. Lett.* 65 (11 1990), pp. 1297–1300. URL: http://link.aps.org/doi/10.1103/PhysRevLett.65.1297.

[10] B. Pontecorvo. "Mesonium and antimesonium". In: *Sov. Phys. JETP* 6 (1957). [Zh. Eksp. Teor. Fiz. 33, 549 (1957)], p. 429.

[11] Z. Maki, M. Nakagawa, and S. Sakata. "Remarks on the unified model of elementary particles". In: *Prog. Theor. Phys.* 28 (1962), p. 870.

[12] Y. Fukuda et al. "Study of the atmospheric neutrino flux in the multi-GeV energy range". In: *Phys. Lett.* B436 (1998), pp. 33–41. eprint: hep-ex/9805006.

[13] C. Giunti and C. Kim. *Fundamentals of Neutrino Physics and Astrophysics*. Oxford University Press, 2007.

[14] Carlo Giunti. "No Effect of Majorana Phases in Neutrino Oscillations". In: *Phys. Lett.* B686 (2010), pp. 41–43. eprint: 1001.0760.

[15] K. Abe et al. "Precise Measurement of the Neutrino Mixing Parameter θ_{23} from Muon Neutrino Disappearance in an Off-axis Beam". In: *Phys.Rev.Lett.* 112 (2014), p. 181801. eprint: 1403.1532.

[16] K. Abe et al. "Evidence of Electron Neutrino Appearance in a Muon Neutrino Beam". In: *Phys.Rev.* D88 (2013), p. 032002. eprint: 1304.0841.

[17] J. Beringer et al. (Particle Data Group). In: *Phys. Rev. D* 86.010001 (2012).

[18] J. A. Formaggio and G. P. Zeller. "From eV to EeV: Neutrino cross sections across energy scales". In: *Rev. Mod. Phys.* 84 (3 2012), pp. 1307–1341. URL: http://link.aps.org/doi/10.1103/RevModPhys.84.1307.

[19] Veronique Bernard, Latifa Elouadrhiri, and Ulf-G Meitner. "Axial structure of the nucleon". In: *Journal of Physics G: Nuclear and Particle Physics* 28.1 (2002), R1. URL: http://stacks.iop.org/0954-3899/28/i=1/a=201.

[20] V. Lyubushkin et al. "A study of quasi-elastic muon neutrino and antineutrino scattering in the NOMAD experiment". In: *The European Physical Journal C* 63.3 (2009), pp. 355–381.

[21] Aguilar-Arevalo et al. "First measurement of the muon neutrino charged current quasielastic double differential cross section". In: *Phys. Rev. D* 81 (9 2010), p. 092005. URL: http://link.aps.org/doi/10.1103/PhysRevD.81.092005.

[22] G. A. Fiorentini et al. "Measurement of Muon Neutrino Quasielastic Scattering on a Hydrocarbon Target at E_ν ~3.5 GeV". In: *Phys. Rev. Lett.* 111 (2 2013), p. 022502. URL: http://link.aps.org/doi/10.1103/PhysRevLett.111.022502.

[23] B. G. Tice et al. "Measurement of Ratios of ν_μ Charged-Current Cross Sections on C, Fe, and Pb to CH at Neutrino Energies $2-20$ GeV". In: *Phys. Rev. Lett.* 112 (23 2014), p. 231801. URL: http://link.aps.org/doi/10.1103/PhysRevLett.112.231801.

[24] E. J. Moniz et al. "Nuclear Fermi Momenta from Quasielastic Electron Scattering". In: *Phys. Rev. Lett.* 26 (8 1971), pp. 445–448. URL: http://link.aps.org/doi/10.1103/PhysRevLett.26.445.

[25] G. M. Radecky et al. "Study of Single Pion Production by Weak Charged Currents in Low-energy Neutrino d Interactions". In: *Phys. Rev. D* 25 (1982), 11611173.

[26] T. Kitagaki et al. "Charged Current Exclusive Pion Production in Neutrino Deuterium Interactions". In: *Phys. Rev. D* 34 (1986), 25542565.

[27] Aguilar-Arevalo et al. "Measurement of neutrino-induced charged-current charged pion production cross sections on mineral oil at $E_\nu \sim 1$ GeV". In: *Phys. Rev. D* 83 (5 2011), p. 052007. URL: http://link.aps.org/doi/10.1103/PhysRevD.83.052007.

[28] Philip Rodrigues. "Comparing pion production models to MiniBooNE data". In: *arXiv:1402.4709 [hep-ex]* (2014).

[29] "Letter of Intent: Neutrino oscillation experiment at JHF". In: (). URL: http://neutrino.kek.jp/jhfnu/loi/loi_JHFcor.pdf.

[30] K. Abe et al. "The T2K experiment". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 659.1 (2011), pp. 106 –135. URL: http://www.sciencedirect.com/science/article/pii/S0168900211011910.

[31] K. Abe et al. "T2K neutrino flux prediction". In: *Phys. Rev. D* 87 (1 2013), p. 012001. URL: http://link.aps.org/doi/10.1103/PhysRevD.87.012001.

[32] K. Abe et al. "Measurement of the inclusive ν_μ charged current cross section on carbon in the near detector of the T2K experiment". In: *Phys. Rev. D* 87 (9 2013), p. 092003. URL: http://link.aps.org/doi/10.1103/PhysRevD.87.092003.

[33] A. Vacheret et al. "Characterization and simulation of the response of Multi-Pixel Photon Counters to low light levels". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 656.1 (2011), pp. 69 –83. URL: http://www.sciencedirect.com/science/article/pii/S0168900211014513.

[34] A. Vacheret et al. "The front end readout system for the T2K-ND280 detectors". In: *Nuclear Science Symposium Conference Record, IEEE* 3 (2007), 19841991.

[35] S. Aoki et al. "The T2K Side Muon Range Detector (SMRD)". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 698.0 (2013), pp. 135 –146. URL: http://www.sciencedirect.com/science/article/pii/S0168900212011242.

[36] S. Assylbekov et al. "The T2K ND280 off-axis pizero detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 686.0 (2012), pp. 48 –63. URL: http://www.sciencedirect.com/science/article/pii/S0168900212005153.

[37] N. Abgrall et al. "Time projection chambers for the {T2K} near detectors". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 637.1 (2011), pp. 25 –46. URL: http://www.sciencedirect.com/science/article/pii/S0168900211003421.

[38] I. Giomataris et al. "Micromegas in a bulk". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 560.2 (2006), pp. 405 –408. URL: http://www.sciencedirect.com/science/article/pii/S0168900205026501.

[39] P.-A. Amaudruz et al. "The T2K fine-grained detectors". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 696.0 (2012), pp. 1 –31. URL: http://www.sciencedirect.com/science/article/pii/S0168900212008789.

[40] D Allan et al. "The electromagnetic calorimeter for the T2K near detector ND280". In: *Journal of Instrumentation* 8.10 (2013), P10019. URL: http://stacks.iop.org/1748-0221/8/i=10/a=P10019.

[41] K. Abe et al. "Measurements of the T2K neutrino beam properties using the INGRID on-axis near detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 694.0 (2012), pp. 211 –223. URL: http://www.sciencedirect.com/science/article/pii/S0168900212002987.

[42] M. Yokoyama et al. "Application of Hamamatsu MPPCs to T2K neutrino detectors". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 610.1 (2009). New Developments In Photodetection {NDIP08} Proceedings of the Fifth International Conference on New Developments in Photodetection, pp. 128 –130. URL: http://www.sciencedirect.com/science/article/pii/S0168900209010444.

[43] S. Fukuda et al. "The Super-Kamiokande detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 501.23 (2003), pp. 418 –462. URL: http://www.sciencedirect.com/science/article/pii/S016890020300425X.

[44] G. Battistoni et al. "The FLUKA code: description and benchmarking". In: *AIP Conf. Proc.* 896.31 (2007).

[45] R. Brun et al. *Report No. CERN-W5013.*

[46] N. Abgrall et al. "Measurements of cross sections and charged pion spectra in proton-carbon interactions at 31 GeV/c". In: *Phys. Rev. C* 84 (3 2011), p. 034604. URL: http://link.aps.org/doi/10.1103/PhysRevC.84.034604.

[47] N. Abgrall et al. "Measurement of production properties of positively charged kaons in proton-carbon interactions at 31 GeV/c". In: *Phys. Rev. C* 85 (3 2012), p. 035210. URL: http://link.aps.org/doi/10.1103/PhysRevC.85.035210.

[48] Y. Hayato. "Neut". In: *Nuclear Physics B - Proceedings Supplements* 112.13 (2002), pp. 171 –176. URL: http://www.sciencedirect.com/science/article/pii/S0920563202017590.

[49] C. Andreopoulos et al. "The GENIE neutrino Monte Carlo generator". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 614.1 (2010), pp. 87 –104. URL: http://www.sciencedirect.com/science/article/pii/S0168900209023043.

[50] Rene Brun and Fons Rademakers. "ROOT - An Object Oriented Data Analysis Framework". In: *Nucl. Inst. & Meth. in Phys. Res. A* 389 (1997), pp. 81–86.

[51] C.H. Llewellyn Smith. "Neutrino reactions at accelerator energies". In: *Physics Reports* 3.5 (1972), pp. 261 –379. URL: http://www.sciencedirect.com/science/article/pii/0370157372900105.

[52] Dieter Rein and Lalit M Sehgal. "Neutrino-excitation of baryon resonances and single pion production". In: *Annals of Physics* 133.1 (1981), pp. 79 –153. URL: http://www.sciencedirect.com/science/article/pii/0003491681902426.

[53] Torbjrn Sjstrand, Stephen Mrenna, and Peter Skands. "PYTHIA 6.4 physics and manual". In: *Journal of High Energy Physics* 2006.05 (2006), p. 026. URL: http://stacks.iop.org/1126-6708/2006/i=05/a=026.

[54] T. Yang et al. "A hadronization model for few-GeV neutrino interactions". In: *The European Physical Journal C* 63.1 (2009), pp. 1–10.

[55] K. Olchanski. S. Ritt P. Amaudruz. *MIDAS (Maximum Integration Data Acquisition System)*. URL: http://midas.psi.ch.

[56] A. Cervera-Villanueva, J.J. Gmez-Cadenas, and J.A. Hernando. "RecPack a reconstruction toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 534.12 (2004).

Proceedings of the {IXth} International Workshop on Advanced Computing and Analysis Techniques in Physics Research, pp. 180 –183. URL: http://www.sciencedirect.com/science/article/pii/S0168900204015220.

[57] A. Hillairet et al. *ND280 Reconstruction*. Tech. rep. 72. T2K Collaboration, 2011.

[58] C. Bojechko, (K. Connolly), et al. *CC-Muliple-Pion ν_μ Event Selections in the ND280 Tracker Using Run 1 - 4 Data*. Tech. rep. 152. T2K Collaboration, 2013.

[59] S. Oser. *Elemental Composition and Masses of FGD XY modules*. Tech. rep. 91. T2K Collaboration, 2010.

[60] E. Frank, A. Marchionni, and M. Messina. *B-field calibration and systematic errors*. Tech. rep. T2K Collaboration, 2009.

[61] Jordan Myslik. *Determination of pion secondary interaction systematics for the ND280 ν_μ analysis*. Tech. rep. 125. T2K Collaboration, 2013.

[62] A. Hillairet et al. *ND280 tracker tracking efficiency*. Tech. rep. 75. T2K Collaboration, 2012.

[63] C. Bojechko et al. *Measurement and Correction of Magnetic Field Distortions in the Time Projection Chambers*. Tech. rep. 61. T2K Collaboration, 2013.

[64] A. Kayis-Topaksu et al. "Charged-particle multiplicities in charged-current neutrino and anti-neutrinonucleus interactions". In: *The European Physical Journal C* 51.4 (2007), pp. 775–785.

[65] D. Zieminska et al. "Charged-particle multiplicity distributions in νn and νp charged-current interactions". In: *Phys. Rev. D* 27.47 (1983).

[66] O. Buss et al. "Transport-theoretical description of nuclear reactions". In: *Physics Reports* 512.12 (2012). Transport-theoretical Description of Nuclear Reactions, pp. 1 –124. URL: http://www.sciencedirect.com/science/article/pii/S0370157311003619.

[67] O. Lalakulich and U. Mosel. "Comparison of GiBUU calculations with MiniBooNE pion production data". In: *arXiv:1304.2409v1 [nucl-th]* (2013). URL: http://arxiv.org/pdf/1304.2409.pdf.

[68] P. Adamson et al. "Neutrino and antineutrino inclusive charged-current cross section measurements with the MINOS near detector". In: *Phys. Rev. D* 81 (7 2010), p. 072002. URL: http://link.aps.org/doi/10.1103/PhysRevD.81.072002.

[69] Patrick de Perio et al. *Cross section parameters for the 2012a oscillation analysis.* Tech. rep. 108. T2K Collaboration, 2013.

[70] Z. Koba, H.B.Nielsen, and P. Olesen. "Scaling of multiplicity distributions in high energy hadron collisions". In: *Nucl. Phys. B* 40.317 (1972).

[71] P. Allen et al. "Multiplicity Distributions in Neutrino-hydrogen interactions". In: *Nucl. Phys. B* 181 (1981), pp. 385–402.

[72] Y. Hayato and C. Andreopoulos. "Private Communication".

[73] Nicolas Abgrall et al. *Flux Prediction and Uncertainties for the 2012a Oscillation Analysis.* Tech. rep. 99. T2K Collaboration, 2013.

[74] Mark Hartz, Asher Kaboth, and Kendall Mahn. *Constraining the Flux and Cross Section Models with Data from the ND280 Detector for the 2013 Oscillation Analysis.* Tech. rep. 166. T2K Collaboration, 2013.

[75] Patrick de Perio. *NEUT Nuclear Effects (FSI).* Tech. rep. 33. T2K Collaboration, 2012.

[76] Patrick de Perio et al. *Implementation of the NIWG Cross Section Parametrization.* Tech. rep. 113. T2K Collaboration, 2013.

[77] Aguilar-Arevalo et al. "Measurement of ν_μ and $\overline{\nu}_\mu$ induced neutral current single π^0 production cross sections on mineral oil at $E_\nu \sim 1$ GeV". In: *Phys. Rev. D* 81 (1 2010), p. 013005. URL: http://link.aps.org/doi/10.1103/PhysRevD.81.013005.

[78] Aguilar-Arevalo et al. "Measurement of ν_μ-induced charged-current neutral pion production cross sections on mineral oil at $E_\nu \in 0.5 - 2.0$ GeV". In: *Phys. Rev. D* 83 (5 2011), p. 052009. URL: http://link.aps.org/doi/10.1103/PhysRevD.83.052009.

[79] F. James. *MINUIT Reference Manual.* Tech. rep. CERN, 2000.

[80] Luc Demortier and Louis Lyons. *Everything you ever wanted to know about pulls*. Tech. rep. 5776. CDF/ANAL/PUBLIC, 2002.

[81] V. Blobel. *Unfolding methods in high energy physics experiments*. CERN 85-09. 1985.

[82] Glen Cowan. *Statistical Data Analysis*. Oxford Science Publications, 1998.

[83] A. N. Tikhonov. "On the solution of improperly posed problems and the method of regularization". In: *Sov. Math.* 5 (1963), p. 1035.