A Study of Neutrino Interactions in Argon Gas

Pip Asher Hamilton
Imperial College London

Submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy in Physics

Abstract

The T2K (Tokai to Kamioka) experiment uses an intense off-axis muon neutrino beam to study neutrino oscillations through muon neutrino disappearance and electron neutrino appearance. As T2K and other neutrino oscillation experiments move beyond the regime in which they are statistically limited, uncertainties on neutrino-nucleus cross-sections have become increasingly significant as a limiting factor on the precision of these experiments, and on the next generation of detectors. Reducing these uncertainties requires a better understanding of neutrino-nucleus interactions through empirical study.

This thesis describes the first ever measurement of neutrino-nucleus interactions on a gaseous target, using the intense T2K beam and the three gaseous argon time-projection chambers in the T2K near detector, ND280. I identify 63 neutrino interaction candidates, and make a comparison of the charged-current ν_μ differential cross-section with respect to proton multiplicity between data and the simulation packages NEUT and GENIE.

This thesis also describes the methods developed to select the gas interactions data sample, which by virtue of the detailed reconstruction available in the ND280 TPCs offers further opportunities to test the predictions of nuclear models than those that are covered by this thesis. Future generations of this analysis are expected to expand both the size of the sample and the range of variables tested.

Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

This thesis represents my own work over four years working on the T2K experiment. As with all work on experiments of this size, I could not have completed the analysis without the help and hard work of many other collaborators. In particular, the development and implementation of the TREx software described in Chapter 4 was primarily the work of Dr. Anthony Hillairet and Edward Larkin, with assistance from Dr. Tom Feusels and Dr. Stefania Bordoni; my contribution to this software was primarily testing during development, both to debug new features and to assess how well the software was meeting the goals of the analysis.

The development and testing of the preselection and final selection, the evaluation of the systematic uncertainties associated with our measurement, and the carrying out of the Bayesian unfolding procedure were my own work, subject to review and suggestions from the previously mentioned collaborators, along with other members of the collaboration.

Pip Hamilton

Acknowledgements

First of all, I would like to thank my supervisor Dr Morgan Wascko, both for offering me this analysis and for his continued advice and encouragement through my time at Imperial.

I would like to specially thank Dr Anthony Hillairet, Eddy Larkin, Dr Tom Feusels and Dr Stefania Bordoni for their tireless work in developing, testing and delivering the TREx software, without which we could never have begun to select neutrino interactions in gas.

I would like to thank Dr Per Jonsson, Dr Ben Smith, Dr Peter Sinclair, Dominic Brailsford, Luke Pickering, Wing Ma, and particularly Dr Asher Kaboth for their interest and advice while I was working in the Blackett Laboratory, as well as for creating such a friendly atmosphere in the office. I consider myself extremely lucky to have had the opportunity to work with such talented and agreeable people, and I'm sure owl be seeing you all again.

I would like to thank Ayako-san and the staff at the Masago dormitory for all their kind help while I was in Japan, from helping me obtain a visa to taking me to the doctor when I was ill. Similarly, thank you to Dr Phill Litchfield, Dr Matt Lawe, Linda Cremonesi, Andy Furmanski, Steve Dennis, and all the other members of the gaijin bicycle gang for their company, which made Tokai a much easier place to live.

I would like to thank Dr Alfons Weber for introducing me to the world of neutrino oscillations and pointing me towards the T2K experiment.

I would like to thank Sam Cunliffe, Philip Senior and David Hughes for their patience in putting up with me while I was living with them at various points during my PhD.

I would like to thank Ryan North for allowing us to use the likeness of his webcomic character T-Rex in the logo for our reconstruction.

Finally, thank you Mum, Dad, Robbie and Isobel, for your support throughout my PhD.

Contents

List of figures **xiii**

List of tables **xxiii**

1 Neutrino Oscillation Theory **3**
1.1 A Brief History of Neutrino Physics . . . 3
1.2 The PMNS Mixing Matrix . . . 4
1.3 The State of the Field . . . 7

2 Nuclear Effects in Neutrino Experiments **11**
2.1 GeV-scale Neutrino Interactions . . . 12
2.2 Modelling the Nucleus . . . 15
2.2.1 Elements of Nuclear Models . . . 16
2.2.2 Neutrino Generators . . . 17
2.3 Model Testing Through Observations of Neutrino Interactions in Gas . . 18
2.3.1 Proton Studies . . . 18

3 T2K and ND280 **21**
3.1 An Introduction to T2K . . . 21
3.2 Creating the Neutrino Beam . . . 22
3.2.1 The Off-Axis Method . . . 22
3.2.2 Proton Accelerator . . . 23
3.2.3 Neutrino Beamline . . . 24
3.3 Near Detectors . . . 26
3.3.1 The ND280 Complex . . . 26
3.3.2 INGRID . . . 26
3.3.3 ND280 . . . 27
3.4 Far Detector: Super-K . . . 34

4 ND280 Software and Simulation **37**

4.1 Overview . . . 37
4.2 Monte Carlo simulation . . . 38
4.2.1 Beam simulation . . . 38
4.2.2 Neutrino interaction simulation: NEUT and GENIE . . . 40
4.2.3 Detector simulation . . . 41
4.3 TPC calibration . . . 41
4.4 TPC reconstruction . . . 42
4.5 TREx . . . 43
4.5.1 Pathfinding . . . 45
4.5.2 Output Structure . . . 47
4.5.3 δ-rays . . . 49

5 Selection **51**
5.1 Preselection . . . 52
5.1.1 Preselection Background Topologies . . . 52
5.1.2 Preselection Cuts . . . 53
5.1.3 Preselection Performance . . . 55
5.2 Final Selection . . . 56
5.2.1 Cuts . . . 57
5.2.2 Performance . . . 65

6 Systematic Uncertainties **73**
6.1 Flux systematics . . . 73
6.2 Cross-section systematics . . . 74
6.3 TPC Detector systematics . . . 76
6.3.1 Track-Finding Efficiency . . . 77
6.3.2 Muon Identification Efficiency . . . 78
6.3.3 Vertex Connection Efficiency . . . 79
6.3.4 Proton Identification Efficiency . . . 80
6.4 Out of fiducial volume systematics . . . 81
6.5 Propagation of Systematic Uncertainties . . . 82

7 Argon Cross-Section **87**
7.1 Bayesian Unfolding . . . 88
7.1.1 Unfolding Method . . . 88
7.1.2 Fake Data Tests . . . 90
7.2 Uncertainties . . . 94

7.3 Results 97

8 Investigation of Results and Future Improvements 105

8.1 Investigation of Data-Monte Carlo Discrepancies 105

8.2 Sample Selected Data Events 106

8.2.1 Noisy Tracks 110

8.2.2 Incorrect Vertex Finding 113

8.2.3 Differing Angular Distributions 113

8.3 Future Improvements 114

8.3.1 Improved Preselection 114

8.3.2 Improved Reconstruction 115

8.3.3 Improved Selection 116

8.3.4 Improved Monte Carlo Modelling 117

9 Conclusions 119

A Copyright Waiver 121

Bibliography 125

List of figures

2.1 Feynman diagrams for charged-current quasi-elastic scattering ("CCQE", left) and neutral-current quasi-elastic scattering ("NCQE", right). 12

2.2 Feynman diagrams for charged-current resonant pion production ("CCRES"). 13

2.3 Feynman diagram for deep inelastic scattering ("DIS"). 14

2.4 Plots showing the total neutrino (left) and antineutrino (right) per nucleon cross-section divided by neutrino energy, using data from multiple experiments (figure taken from [1]; see reference for details). Predictions for the different interaction modes are drawn from the NUANCE simulation package. 14

2.5 Data from MiniBooNE and NOMAD compared to model predictions for the values of M_A^{QE} that best fit each dataset. Figure taken from [2] . . . 15

2.6 Plots the predicted range of protons in the drift gas of the ND280 TPCs as a function of kinetic energy (left) and momentum (right), from particle gun simulation. Limits are drawn on to show the anticipated minimum track length for successful reconstruction, showing that in this medium there is the potential to reconstruct protons with as little kinetic energy as 0.4 MeV. 19

2.7 Plots showing the predictions from NEUT and GENIE for the momentum (left) and multiplicity (right) of protons leaving the target nucleus for charged-current ν_μ interactions on Ar in the T2K beam. Of particular interest are the model disagreements at high multiplicity and low momentum. The momentum distributions are annotated to show how the extra sensitivity of a gas detector covers the region of maximum tension between the two generators. 19

3.1 A diagram illustrating the layout of the T2K experiment [3]. Modified for clarity by T. Ishida. 22

3.2 A plot showing the neutrino beam spectra for various off-axis (OA) angles [4]. The ν_μ survival probability is drawn along the top, showing that the peak of the 2.5° OA spectrum coincides with the first oscillation maximum. 22

3.3 An aerial photograph of the J-PARC accelerator complex. 23

3.4 Official plot of the POT delivered by the T2K beam up to the present time, in both ν and anti-ν mode. 25

3.5 A schematic of the secondary neutrino beamline. Official T2K figure provided by T. Ishida. 26

3.6 A diagram showing the layout of the INGRID detector as seen from the beam direction [3] (left), a diagram showing the placement of INGRID in the pit (centre), and a photograph of INGRID's vertical modules (right). 27

3.7 A muon track in an INGRID module, seen from both the side and top views [3]. 27

3.8 An exploded cross-section of ND280 [3] (left), and a photograph of ND280 with the magnet closed (right). 28

3.9 A diagram showing the construction of the P0D [3], with the beam entering from the left. The triangular segmentation of the scintillator bars is shown in close-up in the insets. 29

3.10 An MPPC of the type used in the ND280 P0D, FGDs and ECals. Left: magnified view of MPPC face, showing 26×26 pixel array. Right: ceramic package of the same MPPC. 29

3.11 Illustration of the perpendicular layer configuration for the FGDs and ECals.The shaded bar on the left has its y coordinate defined by the bar and its z coordinate defined by its layer; the shaded bar on the right has its x coordinate defined by the bar and the z coordinate defined by its layer. Combining the 2D information from successive alternating layers allows for the reconstruction of a complete 3D track. Figure drawn by Luke Pickering. 30

3.12 Cutaway drawing showing the TPC design [3]. The outer box has dimensions 2.30 × 2.40 × 0.97 m ($x \times y \times z$, where z is the beam axis and y is normal to the surface of the earth); the active gas volume contains 3,000 L of gas mixture, with a further 3,300 L in the gap volume between the field cage and the outer wall. 31

3.13 ND280 MicroMEGAS module in close-up. 33

3.14 A diagram illustrating a neutrino interaction in Super-K from the Super-K website (left), and a photograph of the detector's interior wall (right). . . 34

3.15 Event displays from Super-K, showing the sharp ring produced by a muon (left) and the "fuzzy" ring produced by an electron (right). Event display provided by B. Still. 35

4.1 A diagrammatic overview of the steps used by the ND280 software. The prefix "oa" stands for "off-axis" and is used ubiquitously throughout the ND280 software. 38

4.2 A plot of the nominal ν_μ flux from simulation (top), and a plot of the weights that are applied during the tuning (bottom). Different weights are applied for different neutrinos (i.e. $\overline{\nu}_\mu$, ν_e). 39

4.3 Plots showing the distribution of the energy loss as a function of the momentum for negatively (left) and positively (right) charged particles produced in neutrino interactions, compared to the expected curves from Monte Carlo for muons, electrons, protons and pions [5]. 43

4.4 Two simulated ND280 event displays showing the event topologies looked for in this analysis (left) and for which the ND280 reconstruction software was designed (right). On the left, a neutrino interacts in the TPC drift gas, producing a vertex in the fiducial volume of the TPC. On the right, a neutrino interacts in the FGD, producing a muon track which propagates through the TPC downstream. 44

4.5 A diagram illustrating the A* pathfinding procedure, connecting activated MicroMEGAS pads to navigate between a start point (green) and an end point (red). The small green arrows represent the connection cost ($c_{\text{connection}} = |\vec{r}_{\text{next}}|^4$) at each step; the long red arrow represents the heuristic cost $c_{\text{heuristic}} = |\vec{r}_{\text{end}}|$. Diagram provided by E. Larkin. 46

4.6 A diagram illustrating the procedure for finding junctions, provided by E. Larkin. 47

4.7 A diagram illustrating the difference in hit profile between kink-like vertices and curved tracks, provided by E. Larkin. 47

4.8 A cartoon illustrating the hierarchy of patterns, junctions and paths. In this example there are two patterns in the TPC: pattern 1 consists of three paths connected by a junction, while pattern 2 consists of a single isolated path. 48

4.9 A diagram illustrating the δ-ray tagging procedure in the YZ (left) and XZ (right) views, provided by E. Larkin. Multi-peak hits are shown in deep red, while other hits tagged as belonging to the δ-ray are shaded in pink. 49

5.1 A simulated event with no activity in the TPCs. Hits in the bottom ECal show a contained neutrino interaction, with no interaction products reaching the TPCs. 52

5.2 A simulated event in which a neutrino interaction in FGD1 produces a muon that passes through TPC2, leaving a clear trail of hits. 53

5.3 Left: histograms of neutrino energy before and after the preselection, area-normalised to show the minimal change in shape. Right: a plot of the efficiency of the preselection as a function of neutrino energy. Both these plots were generated using NEUT Monte Carlo. 56

5.4 Plots of the distributions of M_1 (left) and M_2 (right) for particle tracks in T2K NEUT beam Monte Carlo. The distributions for muons and non-muons are unit-normalised. 59

5.5 Plots of the ratio of non-muons to muons as a function of M_1 and M_2 for particle tracks in T2K NEUT beam Monte Carlo. The plot on the right shows the full Z-scale; the plot on the left is truncated at a ratio of 2, so that all bins where non-muons outnumber muons by a factor of 2 appear in red. 59

5.6 A selection of event displays showing contrasting examples of the reconstructed topologies created by δ-rays in simulation. (A) Small δ-ray "stub" in TPC2. (B) Large messy δ-ray in TPC1. (C) Large δ-ray with a clean spiral track in TPC3. All three are shown with the beam entering from the left. 61

5.7 Plots showing the distribution of $P \times E$ as a function of the vertex extent cut, in TPC1 (top left), TPC2 (top right) and TPC3 (bottom). 62

5.8 A selection of event displays showing some of the most problematic misreconstruction backgrounds in simulation. (A) A through-going track split into two paths where it crosses a gap between MicroMEGAS pads. (B) A through-going track with a false junction from a stubby δ-ray too small to be caught by the δ-ray cut. (C) Two strongly colinear entering tracks that form a false junction when they separate. 63

5.9 Plot showing the lowest angle distributions of signal and background in Monte Carlo surviving the δ-ray cut (in all three TPCs). 64

5.10 Plots showing the distribution of $P \times E$ as a function of the angular cut value for TPC1 (top left), TPC2 (top right) and TPC3 (bottom). 64

5.11 Plots showing the number of surviving signal and background events from preselection-skimmed Monte Carlo after each cut, in TPCs 1 (top left), 2 (top right) and 3 (bottom left). Although cuts 1-6 reduce the number of events from outside the fiducial volume by two orders of magnitude, without the final cut these events still dominate the selected sample. These plots are also given in tabular form in Table 5.6. The contribution of each of the three tests involved in the final cut is shown in Figure 5.12, in the order that they are applied. 66

5.12 Plots showing the number of surviving signal and background events from preselection-skimmed Monte Carlo at each stage of the misreconstruction veto, in TPCs 1 (top left), 2 (top right) and 3 (bottom left). 67

5.13 Plots showing the selection's efficiency (top left), purity (top right) and the product of the two (bottom left) after each cut. The decrease in $P \times E$ in TPC1 after the muon-finding cut is attributable to TPC1 seeing fewer non-leptonic entering backgrounds than TPCs 2 and 3 (which sample progressively more particles from interactions in the barrel ECals). The decrease in $P \times E$ in TPC3 after the bunch timing cut is attributable to TPC3 observing the highest proportion of these entering backgrounds (which occur inside the bunch), lessening the contribution the cut makes to the purity while keeping the efficiency loss constant. 68

5.14 Plot showing the efficiency of the selection plotted as a function of neutrino energy, in each of the three TPCs. 70

5.15 Plot showing the efficiency of the selection plotted as a function of Q^2, in each of the three TPCs. 70

5.16 Plot showing the efficiency of the selection plotted as a function of the angle between the outgoing muon and the beam axis, in each of the three TPCs. 71

5.17 Plot showing the efficiency of the selection plotted as a function of number of protons emerging from the interaction, in each of the three TPCs. . . . 71

6.1 A plot of the error sources contributing to the ND280 ν_μ flux uncertainty as a function of energy. Figure provided by the T2K beam group. 74

6.2 A plot of the track-finding efficiency in TPC2 (shown as an example - TPCs 1 and 3 show almost identical behaviour), evaluated for data and MC in both TREx and tpcRecon. In many bins the two are identical. . . 78

6.3 A plot of the muon PID efficiency in TPC2, evaluated for data and MC in both TREx and tpcRecon. 79

6.4 A plot of the proton PID efficiency in TPC2 as a function of the reconstructed proton momentum, evaluated for data and MC in TREx. 81

6.5 The flux covariance matrix, produced from 1,000 throws. The high amount of background subtraction in the zero-proton bins inflates the flux uncertainty in those bins. 83

6.6 The cross-section covariance matrix, produced from 1,000 throws varying the FSI parameters and 1,000 throws varying the other cross-section parameters. 84

6.7 The proton counting covariance matrix, produced from 1,000 throws. . . 84

6.8 The combined covariance matrix from all sources of systematic uncertainty. 85

7.1 Plot showing the cumulative selection efficiency cut-by-cut for Monte Carlo gas interactions in TPC1, as a function of the proton multiplicity. All cuts show some model dependence, leading to a $\sim 40\%$ discrepancy in the 4+ proton bin (and large discrepancies in other bins also). TPCs 2 and 3 display similar behaviour. 90

7.2 Plots showing results of unfolding GENIE fake data with the NEUT prior prediction (top) and NEUT fake data with the GENIE prior prediction (bottom) in TPC1. In both cases the unfolding can be seen to perform poorly, with the unfolded distribution matching the prior more closely than the true distribution of the fake data. Error bars are statistical only. 91

7.3 A signal event failing the preselection in GENIE Monte Carlo. The backward-going track leaves enough hits in FGD1 to trigger the upstream veto. A higher rate of these events in GENIE is consistent with GENIE's preference for vertices with higher particle multiplicity (as shown in Figure 2.7). 93

7.4 A plot showing the unfolding of a reweighted NEUT fake data distribution with the nominal NEUT prior. To produce the fake data I applied weights of $1 + M/4$ to events in each proton multiplicity bin, where M is the proton multiplicity. Error bars are statistical only. 94

7.5 The fractional covariance matrix from Monte Carlo statistics. The largest uncertainties from this source arise in the 4+ proton multiplicity bins, where low Monte Carlo statistics in the background sample strongly affect the background subtraction in Equation 7.7. Migration from high- to low-statistics bins causes uncertainties greater than 100% in some on-diagonal bins (a feature also seen in some of the other covariance matrices). 95

7.6 The fractional covariance matrix from data statistics. The largest uncertainties from this source arise in the zero-proton bins in TPCs 1 and 3, which were among the bins with the lowest non-zero populations. 95

7.7 The fractional covariance matrix from all systematic and statistical sources of uncertainty, shown with both a linear (top) and log (bottom) scale. Extreme errors on the zero multiplicity and highest multiplicity bins arise due to the low population (in many cases zero) of the zero multiplicity and high multiplicity bins in data, showing that with the current statistics and selection it is difficult for this analysis to make strong statements about the tails of the multiplicity distribution. 96

7.8 Plots showing the number of selected events before unfolding in TPC1 (top), TPC2 (middle) and TPC3 (bottom), compared to the predicted distributions of selected events from NEUT and GENIE for the same exposure. No statistical error bars are drawn on the data points, as the full treatment of the statistical error is given in the unfolded distributions in Figures 7.11 and 7.12. 98

7.9 Plot showing the number of selected events across all three TPCs, compared to the predictions from NEUT and GENIE for the same exposure. For the same reasons as in Figure 7.8, no statistical error bars are drawn on the data. 99

7.10 Plots showing the unsmearing matrices (as defined in Equation 7.3) calculated for TPC1 (top left), TPC2 (top right) and TPC3 (bottom). 100

7.11 Plots showing the unfolded distributions for TPC1 (top), TPC2 (middle) and TPC3 (bottom). 101

7.12 Plot showing the unfolded distribution across all three TPCs. 102

8.1 Event displays showing signal events seen in the selected data sample, in the YZ(left) and XY(right) views. The Z axis is aligned with the beam direction, and the X axis with the drift direction in the TPCs. 107

8.2 Event display showing signal event with a low-energy track emerging from the neutrino interaction vertex and stopping in the gas volume. The view of the entire detector is shown at the top, in the YZ (left) and XY (right) views. A close view of the interaction is shown in the lower images, in the same views (respectively). 108

8.3 ND280 event display showing a background event seen in the selected data sample, in the YZ(left) and XY(right) views. Two particles from separate interactions outside the TPC cross over to create a false junction. 108

8.4 ND280 event display showing a background event seen in the selected data sample, in the YZ(left) and XY(right) views. A muon from an interaction in the bottom ECal scatters in the gas, forming a kink in the track that is detected as a junction. The short path segment after the scatter does not receive a muon PID and so this junction is selected as an interaction vertex. 109

8.5 ND280 event display showing a background event seen in the selected data sample, in the YZ(top left) and XY(top right) views. A close-up of the selected track is shown at the bottom. The physics behind this "hairy" track is unclear: it may be one particle travelling inside the spiral trajectory of a δ-ray. Its complicated structure leads to the finding of false junctions, one of which is erroneously selected as a signal interaction. . . 109

8.6 Orthogonal views of an event in the selected data sample with an unexpectedly thick, noisy track emerging from the vertex. Green boxes indicate MicroMEGAS hits used by the pathfinding, red boxes hits which were not used, and purple boxes hits that were assigned to a junction. Blue lines indicate fitted paths, and red lines paths for which the fitting failed. X is the drift axis and Z is the beam axis. 110

8.7 Orthogonal views of a signal event in NEUT Monte Carlo with a similar topology to the data event in Figure 8.6. All tracks emerging from the interaction vertex appear much cleaner in the pattern of hits they leave behind. To interpret the colour scheme refer to Figure 8.6. 111

8.8 Plot showing the correlation between central hit charge and lateral hit charge (in fC) in the ND280 TPC MicroMEGAS pads, as measured before installation at the near detector site. The measurement was made by irradiating the TPC with a collimated ^{55}Fe source for which almost all the energy deposited in the detector is contained in a single pad. During these tests the TPC was operated with a higher saturation threshold (240 fC) than its current setting (120 fC). Plot taken from Reference [6]. . . . 112

8.9 Y-Z view of the event shown in Figure 8.6, with a colour scale corresponding to the hit charge (in arbitrary units). No fitted tracks are drawn. A thin core of saturated (red and orange) MicroMEGAS hits can be observed following the "true" particle track; the surrounding blue hits have a very low charge. This is consistent with cross-talk from the saturated hits creating low-charge hits in neighbouring pads. 112

8.10 A histogram of the lowest angle between two tracks for both the data and the Monte Carlo prediction for the same exposure, post-selection. The colinearity cut removes all events with a lowest angle below 32°. 114

8.11 A selection of saturated waveforms in data, showing the saturation plateau. Plot provided by Lukas Koch. 115

8.12 A plot showing the ratio between background and signal in Monte Carlo as a function of the true proton multiplicity and the lowest angle between two tracks, for events for which at least two joined tracks were found. The Z scale is truncated at 2 (i.e. background = 2 × signal). 117

List of tables

1.1 The 2014 best-fit values and 3σ allowed ranges of the 3-neutrino oscillation parameters produced by the Particle Data Group [7], with the addition of recent constraints on δ from joint fits between data from T2K and reactor experiments [8]. Δm^2 is defined $\Delta m^2 = m_3^2 - (m_1^2 + m_2^2)/2$. 7

3.1 The characteristics of the J-PARC beam. 24

5.1 The fiducial volumes used in the preselection. Note that since the preselection is hit-based, there is no drift information, therefore limits on X (the drift axis) are not imposed. The fiducial volume limits were optimised to maximise the product of the preselection's purity and efficiency, stepping in from the TPC edges in increments equal to the dimensions of the MicroMEGAS pads (see Section 3.3.3). 54

5.2 The configuration of the upstream veto for each TPC. The veto in the P0D was taken unchanged from the existing reconstruction; the FGD vetoes were optimised to maximise the product of the preselection's purity and efficiency. 54

5.3 Table summarising the preselection performance on NEUT Monte Carlo (MC) and data. Efficiency is assessed on a specially generated sample of interactions occurring only in the gas; as such losses from coincidental interactions in other detectors cannot be accounted for. 55

5.4 The fiducial volumes used in the final selection. A 20 mm buffer zone is applied around the central cathode in order to exclude neutrino interactions on the cathode itself. 60

5.5 Table summarising selection performance on NEUT Monte Carlo, assessed in the same way as for the preselection in Table 5.3 with uncertainties from Monte Carlo statistics. The decrease in purity for the downstream TPCs is attributed to the same difference in backgrounds entering from the ECals as described for Table 5.3. 68

5.6 Table containing the cut-by-cut predictions for the surviving number of events shown in Figure 5.11, both for all events (top) and signal events only (bottom). The uncertainties quoted are purely statistical. Figures for the data quality cut are not shown, as on Monte Carlo this cut has no effect; the scaled statistics used correspond the amount of good-quality data available at ND280 (5.9×10^{20} POT). 69

6.1 List of the cross-section uncertainties used for this analysis, as provided by the NIWG. These uncertainties are described in detail in Reference [4]. 75

6.2 Uncertainties on the CC-inclusive ν_μ cross-section for the three most prevalent target elements in the surviving out-of-fiducial background. . . 76

6.3 Table of detector systematics assessed in the TPCs. 77

6.4 Vertex connection efficiencies for Monte Carlo samples with a normal rate of noise hits (Nominal) and twice that rate (Double Noise). 80

6.5 A comparison of background survival from sand muons and cosmic muons in data and Monte Carlo. Limits are given where no backgrounds survived the selection; the relatively high limit on the sand muon survival rate in data is due to the smaller size of the available control sample. 82

7.1 A table of the net signal efficiencies for NEUT and GENIE Monte Carlo in each of the three TPCs. 92

7.2 A table of the net background rejection from the preselection, for NEUT and GENIE Monte Carlo in each of the three TPCs. 92

7.3 Table of events surviving the preselection in NEUT, GENIE and data, with both Monte Carlo samples scaled to the POT of the data. The statistical errors reflect the size of each sample before scaling. 93

8.1 Table of the purity of the gas sample, as inferred from examining the data by eye. I make no attempt to assign an uncertainty to my judgement of whether events are signal or background; example event displays are shown in Figures 8.1 to 8.5. 105

"Everything starts somewhere, although many physicists disagree."

— Terry Pratchett

Chapter 1

Neutrino Oscillation Theory

1.1 A Brief History of Neutrino Physics

Neutrinos have been part of the landscape of particle physics for more than half a century. First postulated by Pauli in 1930 [9], their existence was confirmed by Cowan *et al.* in 1956 [10]. Though neutrinos' weak interactions with matter make them difficult to observe in the lab, in recent decades their behaviour has proved one of the richest sources of new discoveries in particle physics.

In 1968, Davis, Harmer and Hoffman measured the solar ν_e flux with a chlorine-based detector in the Homestake mine, and observed a large deficit in the neutrino flux predicted by the solar model [11]. This effect became known as the Solar Neutrino Problem, and was corroborated by later experiments such as GALLEX [12] and SAGE [13].

With the solar model thought to be well understood, one explanation of the observed neutrino deficit was an incomplete understanding of the behaviour of the neutrinos themselves. Until this point, neutrinos had been assumed to be massless. If however neutrinos had mass — and more importantly, if their mass states did not have a one-to-one mapping to their flavour states — it would be possible for neutrinos to "oscillate" from one flavour to another over extended distances, as first suggested by Pontecorvo in 1957 [14]. Some large fraction of the Sun's ν_e flux could therefore be evading detection by changing flavour along the way. In 2001, the SNO experiment provided conclusive evidence that this was the case [15].

The mounting evidence for neutrino oscillations prompted searches with other sources. The disappearance of muon neutrinos from cosmic rays in Earth's atmosphere was

observed in 1998 by Super-Kamiokande [16], while a deficit in the antineutrino flux from nuclear reactors was seen by KamLAND in 2003 [17]. These experiments (and many subsequent measurements) laid the empirical groundwork for our current understanding of the mechanism by which neutrinos change flavour: the PMNS mixing matrix.

1.2 The PMNS Mixing Matrix

The concept of neutrino mass mixing was first introduced by Pontecorvo [14], and extended by Maki, Nakagawa and Sakata[18]. It is best expressed as a mixing matrix relating the three known neutrino flavour states ν_e, ν_μ and ν_τ to three neutrino mass states ν_1, ν_2 and ν_3:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \boldsymbol{U} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}. \tag{1.1}$$

The unitary matrix $\boldsymbol{U}$ is known as the PMNS matrix, after the initials of its inventors. In the notation used by Giunti and Kim [19], it can be expressed as follows:

$$\boldsymbol{U} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{23} & \sin\theta_{23} \\ 0 & \sin\theta_{23} & \cos\theta_{23} \end{pmatrix} \begin{pmatrix} \cos\theta_{13} & 0 & \sin\theta_{13}e^{i\delta} \\ 0 & 1 & 0 \\ \sin\theta_{13}e^{i\delta} & 0 & \cos\theta_{13} \end{pmatrix} \begin{pmatrix} \cos\theta_{12} & \sin\theta_{12} & 0 \\ \sin\theta_{12} & \cos\theta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^{-i\alpha_1} & 0 & 0 \\ 0 & e^{-i\alpha_2} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1.2}$$

In this parameterisation, the matrix is constructed from six independent parameters:

- The three mixing angles θ_{ij}.
- The CP-violating phase δ.
- The Majorana phases α_k.

This causes neutrinos' flavour makeup to change as they propagate through space, according to the following mechanism.

A neutrino beginning in a pure flavour state α is made up of a superposition of mass states (indexed by j):

$$|\nu_\alpha(t=0)\rangle = \sum_j U^*_{\alpha j}|\nu_j\rangle. \tag{1.3}$$

While travelling, the mass states in the superposition evolve according to the neutrino energy.

$$|\nu_\alpha(t)\rangle = \sum_j U^*_{\alpha j}|\nu_j\rangle e^{-iE_j t}. \tag{1.4}$$

The different masses of the states $|\nu_j\rangle$ lead to slight energy differences:

$$E_j = \sqrt{p^2 + m_j^2} \approx p + \frac{m_j^2}{2E}, \tag{1.5}$$

in the approximation $m_j \ll E_j \approx E$ (and in natural units).

The neutrino therefore departs from its pure flavour state α, evolving into a superposition of flavour states. The probability of observing a different flavour β can be assessed:

$$P(\nu_\alpha \to \nu_\beta) = |\sum_j \langle\nu_\beta|\nu_\alpha(t)\rangle|^2 \tag{1.6}$$

$$= |\sum_j U^*_{\alpha j} e^{-iE_j t}\langle\nu_\beta|\nu_j\rangle|^2 \tag{1.7}$$

$$= |\sum_j U^*_{\alpha j} e^{-iE_j t} U^*_{\beta j}|^2. \tag{1.8}$$

Using equation 1.5, this can be rewritten

$$P(\nu_\alpha \to \nu_\beta) = |\sum_j U^*_{\alpha j} e^{-i(p+\frac{m_j^2}{2E})t} U^*_{\beta j}|^2 \tag{1.9}$$

$$= |\sum_j U^*_{\alpha j} e^{-i(p+\frac{m_j^2}{2E})(L/c)} U^*_{\beta j}|^2, \tag{1.10}$$

where L is the distance travelled by the neutrino in the time t.

The cross-terms between different neutrino masses prevent this expression from cancelling to zero. In the case of $\nu_\mu \to \nu_e$ oscillation, as studied by the T2K experiment, we obtain:

$$P(\nu_\mu \to \nu_e) \approx 2\sin^2(2\theta_{13})\sin^2(\theta_{23})\sin^2\left(1.27\frac{\Delta m^2_{32}}{(\mathrm{eV}^2)}\frac{L}{(\mathrm{km})}\frac{(\mathrm{GeV})}{E}\right), \tag{1.11}$$

$$P(\nu_\mu \to \nu_\mu) \approx 1 - \cos^4(\theta_{13})\sin^2(2\theta_{23})\sin^2\left(1.27\frac{\Delta m^2_{32}}{(\mathrm{eV}^2)}\frac{L}{(\mathrm{km})}\frac{(\mathrm{GeV})}{E}\right), \tag{1.12}$$

where $\Delta m^2_{ij} \equiv m^2_i - m^2_j$.

This formula suffices for neutrino oscillations in vacuum. When neutrinos propagate through matter, however, the oscillation formula is complicated by the fact that the $|\nu\rangle_e$ component of the wavefunction can interact with electrons present in the medium (while ordinary matter contains no muons or taus). This extension, developed by Mikheyev, Smirnov [20], and Wolfenstein [21], modifies the oscillation probability:

$$P(\nu_\mu \to \nu_e) \approx 2\sin^2(2\theta_{13})T_1 - \xi\sin^2(2\theta_{13})(T_2 - T_3) + \alpha^2 T_4, \tag{1.13}$$

where

$$T_1 = \sin^2\theta_{23}\Big(\frac{\sin^2[(1-x_\nu)\Delta]}{(1-x_\nu)^2}\Big), \tag{1.14}$$

$$T_2 = \sin\delta\sin^2(2\theta_{12})\sin^2(2\theta_{23})\sin\Delta\Big(\frac{\sin(x_\nu\Delta)}{x_\nu}\Big)\Big(\frac{\sin[(1-x_\nu)\Delta]}{(1-x_\nu)}\Big), \tag{1.15}$$

$$T_3 = \cos\delta\sin^2(2\theta_{12})\sin^2(2\theta_{23})\cos\Delta\Big(\frac{\sin(x_\nu\Delta)}{x_\nu}\Big)\Big(\frac{\sin[(1-x_\nu)\Delta]}{(1-x_\nu)}\Big), \tag{1.16}$$

$$T_4 = \cos^2\theta_{23}\sin^2(2\theta_{12})\Big(\frac{\sin^2(x_\nu\Delta)}{x^2_\nu}\Big), \tag{1.17}$$

with new terms defined $\xi \equiv \frac{\Delta m^2_{21}}{\Delta m^2_{31}}$, $\Delta \equiv \frac{\Delta m^2_{31}L}{4E}$, and $x_\nu = \frac{VL}{2\Delta}$. The term V in the definition of x_ν describes the effective charged-current potential 'seen' by the neutrino as it traverses the medium.

The most significant result of this modification is the appearance of the CP-violating phase δ in the second-order terms T_2 and T_3, offering an additional opportunity for the direct measurement of this parameter by experiments that are able to exploit matter effects via neutrino propagation over long subterrannean distances.

1.3 The State of the Field

At the present time, the PMNS framework appears to show excellent agreement with data from a wide variety of experiments. Complementary experiments using solar, atmospheric, reactor and accelerator neutrinos all show results consistent with three-flavour mixing. Global fits of the results of these experiments, carried out by the Particle Data Group, give values for the oscillation parameters shown in Table 1.1.

Parameter	Best fit value ($\pm 1\sigma$)
Δm^2_{21} [10^{-5}eV^2]	7.53 ± 0.18
$\lvert\Delta m^2\rvert$ [10^{-3}eV^2]	2.44 ± 0.06 ($\Delta m^2 > 0$), 2.51 ± 0.07 ($\Delta m^2 < 0$)
$\sin^2(2\theta_{12})$	0.846 ± 0.021
$\sin^2(2\theta_{23})$	$0.999^{+0.001}_{-0.018}$ ($\Delta m^2 > 0$), $1.000^{+0.000}_{-0.017}$ ($\Delta m^2 < 0$)
$\sin^2(2\theta_{13})$	0.093 ± 0.008
δ	$-\pi$ – 0.14π or 0.87π – π (90% confidence) [8]

Table 1.1: The 2014 best-fit values and 3σ allowed ranges of the 3-neutrino oscillation parameters produced by the Particle Data Group [7], with the addition of recent constraints on δ from joint fits between data from T2K and reactor experiments [8]. Δm^2 is defined $\Delta m^2 = m_3^2 - (m_1^2 + m_2^2)/2$.

While the sizes of the three mixing angles are now well-constrained, several significant questions remain unanswered.

1. **What is the value of δ?** Since θ_{13} is now confirmed to be non-zero, a non-zero value of δ would make neutrino oscillation a CP-violating process. This would have fundamental consequences for our understanding of particle physics and cosmology, as CP violation has until now only been observed in quarks. Moreover, the amount of CP violation in the quark sector is too small to explain the observed matter-to-antimatter asymmetry of the universe [22]. CP violation in the neutrino sector offers the possibility of compensating for this deficit via leptogenesis [23].
2. **What are the absolute values of m_1, m_2 and m_3? What is the sign of Δm^2?** At present, only relative differences between the neutrino masses are known. In addition, the sign of the difference between m_3 and the other two (more closely-spaced) neutrino masses is not known. This ambiguity — whether $m_3 \gg m_2 > m_1$ or $m_2 > m_1 \gg m_3$ — is known as the question of the 'mass hierarchy'.

3. **Are neutrinos Majorana or Dirac particles?** The existence of non-zero neutrino masses raises the question of how these masses are generated. They can be generated by the standard coupling to the Higgs field, making neutrinos Dirac particles and requiring the addition of right-handed, 'sterile' neutrino states that would not interact through any known channel. Alternatively, they could possess an intrinsic mass as a result of being Majorana particles (and therefore their own antiparticles). If so, they would be the first Majorana particles discovered in nature. Majorana neutrinos are a necessary condition for lepton number violation via leptogenesis, as their masses are not constrained by electroweak symmetry breaking, allowing for sufficiently massive right-handed neutrinos to decay into leptons at the required rate through couplings to a Higgs multiplet.

The question of the value of δ can be answered by direct measurements of oscillation probabilities, as long-baseline oscillation experiments such as T2K [24] and NOνA have already begun to do. As the oscillation probability is measured with increasing precision, limits will be placed on δ, particularly in combination with high-precision measurements of θ_{13} from reactor experiments such as Daya Bay [25].

In the same way, long-baseline oscillation experiments with increased precision have the potential to determine the mass hierarchy by exploiting matter effects; namely the second- and third-order terms in Equation 1.13 that are sensitive to the sign of Δm^2_{31}. Accessing the absolute neutrino masses, however, requires a different approach. From simple kinematical considerations, it is possible that a large enough neutrino mass would be visible as a cut-off in the energy spectrum of β decays. The KATRIN experiment [26] will search for this effect with sensitivity to neutrino masses as low as 0.2 eV.

Finally, if neutrinos are Majorana particles, this could be detected through the process of neutrinoless double β decay. Experiments such as SNO+ [27] and SuperNEMO [28] are under construction to tighten the limits on the decay rate for this process. These measurements are necessarily complementary with neutrino mass searches, as the decay rate is given by

$$\Gamma = G|M|^2|\sum_{i=1}^{3} m_i U_{ei}^2|^2, \tag{1.18}$$

where G is the phase-space factor, $|M|$ the nuclear matrix element, U_{ei} the elements of the PMNS matrix, and m_i the neutrino masses. Therefore limits on the decay rate cannot exclude the possibility of Majorana neutrinos without measurements of the masses —

as ever-lower limits on the decay rate can be explained by an ever-lower neutrino mass — and a direct observation of neutrinoless double β decay would make an immediate contribution to our knowledge of the neutrino mass spectrum.

Chapter 2

Nuclear Effects in Neutrino Experiments

The electrically neutral, weakly-interacting nature of neutrinos makes them a difficult subject for empirical testing. Direct measurements of quantities like the neutrino energy are not possible; instead they must be deduced kinematically from the other particles produced by neutrino interactions. Since neutrinos interact so rarely, this enforces upon accelerator neutrino experiments a need for high-intensity beams impinging on large amounts of instrumented target material. As such, all such experiments study neutrino interactions on atomic nuclei.

This introduces a significant complication in the characterisation of the incoming neutrino: are the particles emerging from the nucleus the same particles that were created in the initial neutrino interaction? Do they still have the same energy? To answer these questions requires an accurate modelling of both the initial neutrino interaction, and the process by which interaction products escape from inside the nucleus. Without such nuclear models, the incoming neutrino energy — a crucial variable in the study of neutrino oscillations, as shown in Equation 1.12 — cannot be accurately reconstructed.

Uncertainties from these models become particularly significant in the context of the increased precision needed for oscillation measurements to determine the value ofδ and the mass hierarchy, as discussed in Section 1.3. As accelerator experiments such as T2K move out of the regime in which they are statistically limited, neutrino cross-section uncertainties become one of the leading systematic errors — in the case of T2K at the level of 3 to 7%, depending on the exact value of $\sin^2(2\theta_{13})$ [24] (currently known to within approximately 10% uncertainty, as given in Table 1.1). The motivation for the work I

have carried out in this thesis is to contribute to the reduction of these uncertainties, as detailed in Section 2.3.

2.1 GeV-scale Neutrino Interactions

At the energies used in accelerator neutrino experiments, neutrinos interact with nuclei via a number of channels. At energies below 1 GeV (the peak region of the T2K beam — see Figure 3.2), the dominant process is quasi-elastic scattering, shown in Figure 2.1.

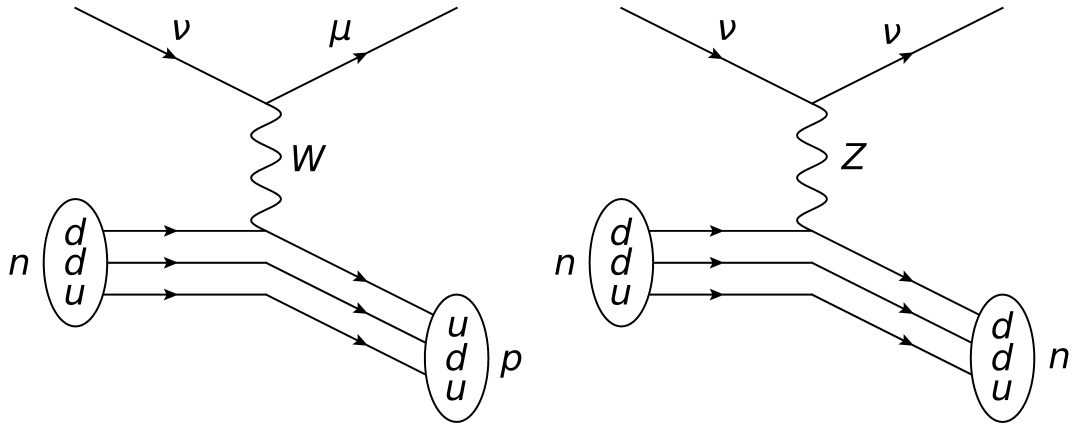


Figure 2.1: Feynman diagrams for charged-current quasi-elastic scattering ("CCQE", left) and neutral-current quasi-elastic scattering ("NCQE", right).

In this process, the neutrino has insufficient energy to remove the quark with which it interacts from the nucleon. As a result it interacts coherently with the nucleon as a whole. In the charged-current case, the neutrino is converted into a charged lepton, while a neutron is converted into a proton (or vice versa in the case of an antineutrino interaction); in the neutral-current case, the neutrino survives, scattering elastically from the nucleon. In either case, the momentum imparted to the target nucleon by the scatter determines whether the nucleon escapes or remains bound in the nucleus. At the T2K beam energy, quasi-elastic interactions are the dominant interaction mode.

At energies between 1 and 5 GeV, the dominant interaction is resonant pion production, shown in Figure 2.2.

While the neutrino still does not have enough energy to disrupt the nucleon, it does impart enough energy to raise the nucleon into an excited state, which subsequently decays (e.g. $n \rightarrow \Delta^+ \rightarrow \pi^+ + n$). The decay products then either escape or are absorbed

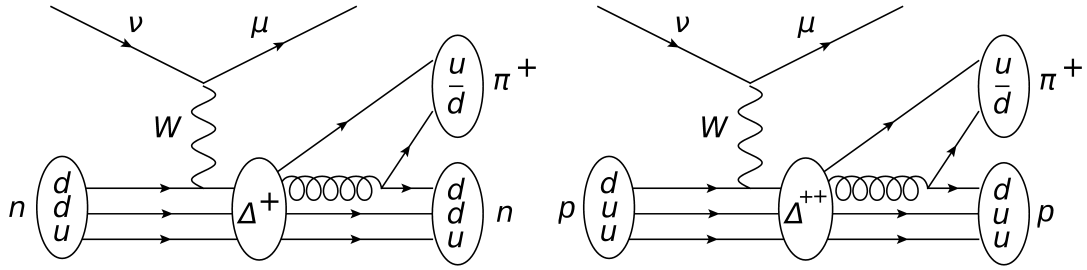


Figure 2.2: Feynman diagrams for charged-current resonant pion production ("CCRES").

by the nucleus in the same manner as the converted or scattered nucleon in quasi-elastic scattering.

Above 5 GeV, the dominant process is deep inelastic scattering, shown in Figure 2.3.

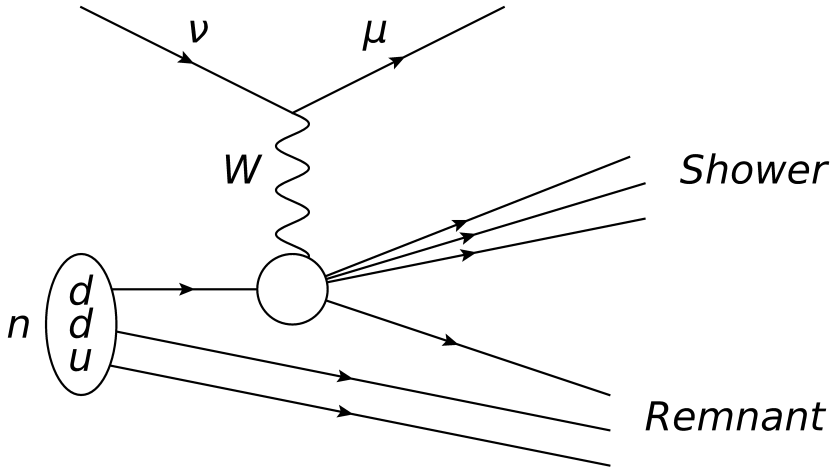


Figure 2.3: Feynman diagram for deep inelastic scattering ("DIS").

In this energy regime, the neutrino is energetic enough to disrupt the nucleon, interacting with an individual target quark. The subsequent disintegration of the nucleon remnants produces hadronic jets, which usually results in the break-up of the target nucleus.

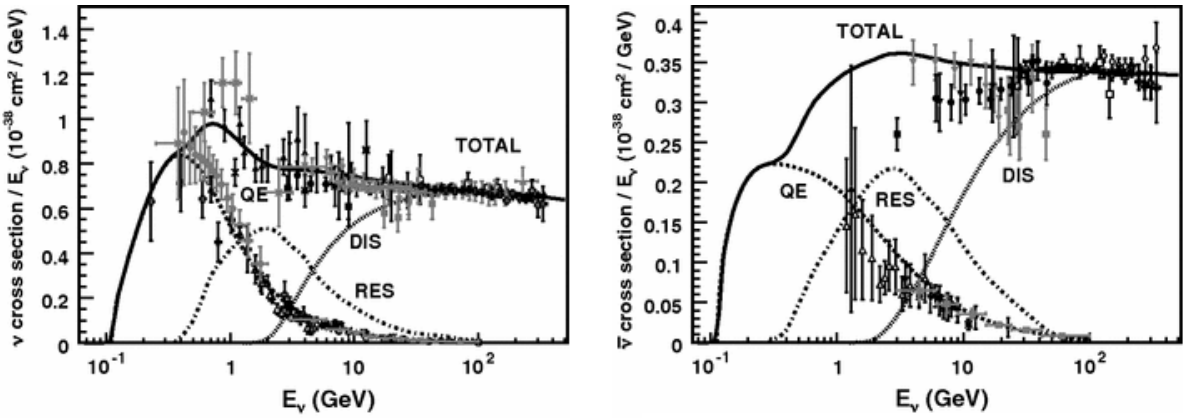


Figure 2.4: Plots showing the total neutrino (left) and antineutrino (right) per nucleon cross-section divided by neutrino energy, using data from multiple experiments (figure taken from [1]; see reference for details). Predictions for the different interaction modes are drawn from the NUANCE simulation package.

As Figure 2.4 shows, there are still significant uncertainties on the cross-section across a broad range of energies. Most recent cross-section results have focused on CCQE interactions, as these are both available at low beam energies and easy to reconstruct. In this channel, a significant tension has also been observed between measurements of the

cross-section parameter M_A^{QE} (the "axial mass" featured in the dipole form of the nucleon axial form factor, $G_A = \frac{1}{(1+\frac{Q^2}{M_A^2})^2}$) from the MinBooNE and NOMAD experiments, as shown in Figure 2.5.

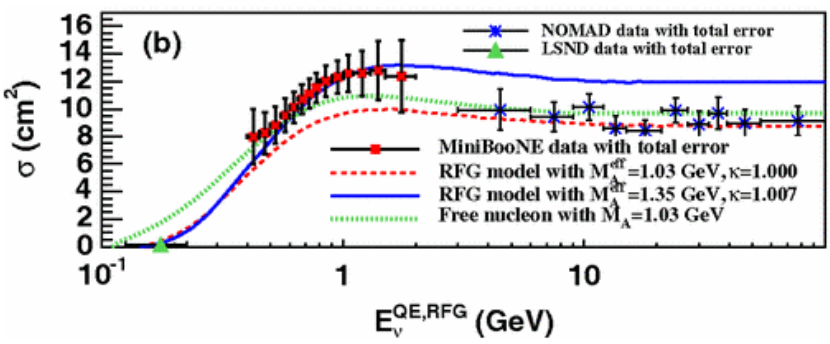


Figure 2.5: Data from MiniBooNE and NOMAD compared to model predictions for the values of M_A^{QE} that best fit each dataset. Figure taken from [2]

One possible explanation for this discrepancy is that it arises from an incomplete understanding of the nuclear environment, which modifies the effective M_A^{QE} measured by both experiments. The nuclear environment carries significant complications for each interaction mode, namely:

- Modelling the momentum of the target nucleon.
- Modelling the interactions of the target nucleon with its surroundings (e.g. correlated nucleon pairs, also referred to as 'MEC' or meson exchange currents).
- Modelling the propagation of interaction and decay products through the nucleus (known as final state interactions, 'FSI').

A sketch of the models used to make these predictions is given in Section 2.2

2.2 Modelling the Nucleus

The models used to describe neutrino-nucleus scattering are mostly based on results from electron scattering experiments, as electrons also offer a low-mass leptonic probe of the nucleus. The accuracy of these models has been a subject of contention among neutrino oscillation experiments — in particular, the validity of the 'impulse approximation', where the incoming neutrino interacts with a single bound nucleon within the nucleus.

A variety of implementations exist for each component of the nuclear model, and as yet there is no consensus on which best represent nature. The essential ingredients of a nuclear model are described below; the exact implementation in different neutrino generators is discussed in later chapters.

2.2.1 Elements of Nuclear Models

In the impulse approximation, the nucleus is treated as being composed of a collection of quasi-free nucleons. To calculate the results of a neutrino interaction with one of the nucleons, the nuclear model must first provide the momentum of the target nucleon. The two main approaches used to model this are the relativistic Fermi gas [29], and spectral functions.

In the relativistic Fermi gas treatment, the nucleus is modelled as a gas of non-interacting nucleons with plane-wave wavefunctions in a potential well. Nucleons stack up through energy levels according to the Pauli exclusion principle, leading to a uniform momentum distribution up to the Fermi momentum p_F.

Spectral functions are an alternative approach widely used in the field of electron scattering. The spectral function itself is a function that describes the four-momentum of the nucleons within the nucleus. The calculation of these functions is analytically tractable for nuclei with $A < 4$ [30] or in the limit $A \to \infty$ [31]; in the intermediate region they must be calculated with numerical methods.

Regardless of whether a Fermi gas model or a spectral function is used, the next step is to model the propagation of the interaction products out of the nucleus, in order to calculate their momentum once they have exited it — if they exit at all. Pions and protons have a high cross-section for re-interaction inside the nucleus, leading most generators to use intranuclear cascade models to represent this process.

While each generator has its own cascade model, they share a common approach:

- Choose a vertex location within the nucleus, based on a nuclear density distribution.
- Propagate the hadron through the nuclear medium in short steps.
- At the end of each step, assess the probability of interaction, and decide whether the hadron interacts via a throw from a random number generator.
- Propagate any interaction products as before until all mobile particles have left the nucleus.

The interaction probabilities are assessed differently from generator to generator, but are usually validated on hadron scattering data.

A recent extension to the impulse approximation is to also consider neutrino interactions on bound pairs of coupled nucleons. These "two-particle, two-hole" (2p2h) interactions, also known as meson exchange currents (MEC), provide another interaction channel with different kinematics to single-nucleon interactions. Recent observations from ArgoNeuT [32] (as well as earlier electron scattering measurements) of pairs of protons with highly correlated momenta being ejected from the nucleus after a neutrino interaction offer an empirical indication that these interactions do occur in nature. Recent efforts have been made to integrate these into neutrino generators, e.g. the implementation of the Nieves model [33] in NEUT.

2.2.2 Neutrino Generators

There are a variety of neutrino generators currently available to oscillation experiments. To give an impression of the breadth of models currently in use, I outline below the four most prominent.

- **NEUT [34]** was originally developed for use in measurements of atmospheric neutrinos by the Kamiokande detector. It was designed for use with a water target, and so is tuned to provide the best performance on hydrogen and oxygen targets, although it has been extended to cover other nuclei. It has continued in use by other experiments such as Super-Kamiokande, K2K and SciBooNE, and is the primary generator used by T2K.
- **GENIE [35]** is a large-scale neutrino generator supported and developed by an international collaboration of scientists with the intention of creating a 'canonical' generator usable by all neutrino experiments at all experimental energies. It is the second generator used by T2K.
- **NuWro [36]** is a generator developed by the University of Wroclaw neutrino group for the main purpose of investigating the impact of final state interactions on cross-section measurements.
- **GiBUU [37]** aims to provide an unified nuclear transport framework at MeV–GeV energies for elementary reactions on nuclei and heavy ion collisions. It has applications outside the neutrino sector and is the only generator not to use a cascade model for calculating final state interactions.

NEUT and GENIE are of the most relevance to this analysis, as the two generators used by T2K. The component models used by these generators are listed in Section 4.2.2

2.3 Model Testing Through Observations of Neutrino Interactions in Gas

It is impossible to directly observe what happens inside a nucleus struck by a neutrino. Instead, tests of our nuclear models must depend on observing the final-state particles produced by the interaction — in the case of charged-current particles, the charged lepton, and whatever other charged ejecta are produced by the interaction.

It is an unfortunate limitation of the technology used for all neutrino detectors that only particles above a certain threshold energy — determined primarily by the density of the material in the active volume — will be visible to the detector. This limits our ability to test the complete prediction of the final state from a given nuclear model: instead, focus is usually given to the charged lepton and other significant, high-energy particles.

The energy threshold of reconstruction can be reduced by reducing the density of the medium through which the particles produced by the interaction propagate, allowing them to travel further and leave more distinct tracks. The ideal low-density medium is obviously a gas, and gas-based time projection chambers (TPCs) are already in use in the T2K near detector, ND280 (see Section 3.3.3). As shown in Figure 2.6, the minimum proton kinetic energy for reconstruction in these detectors (as determined by the minimum path length for track reconstruction being 5cm) is below 0.5 MeV. This is far lower than liquid argon TPCs, the nearest competitor with a proton energy threshold of ~ 21 MeV [38].

As I will show in Section 2.3.1, this region of extra sensitivity is one in which the tensions between model predictions are most pronounced.

2.3.1 Proton Studies

In order to assess the value of potential proton measurements with the ND280 TPCs, I looked at distributions of the proton momentum and multiplicity from interactions on argon from both NEUT and GENIE, the two neutrino generators used by T2K. These

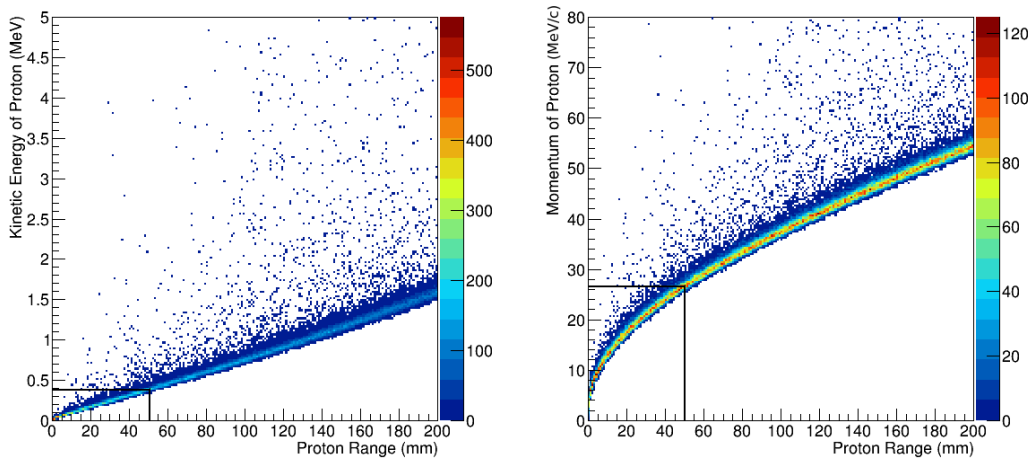


Figure 2.6: Plots the predicted range of protons in the drift gas of the ND280 TPCs as a function of kinetic energy (left) and momentum (right), from particle gun simulation. Limits are drawn on to show the anticipated minimum track length for successful reconstruction, showing that in this medium there is the potential to reconstruct protons with as little kinetic energy as 0.4 MeV.

simulation packages use different physics models, as detailed in Section 4.2.2. Figure 2.7 shows that these differences lead to large discrepancies between the two generators when it comes to the number and momentum of protons emerging from the neutrino interaction.

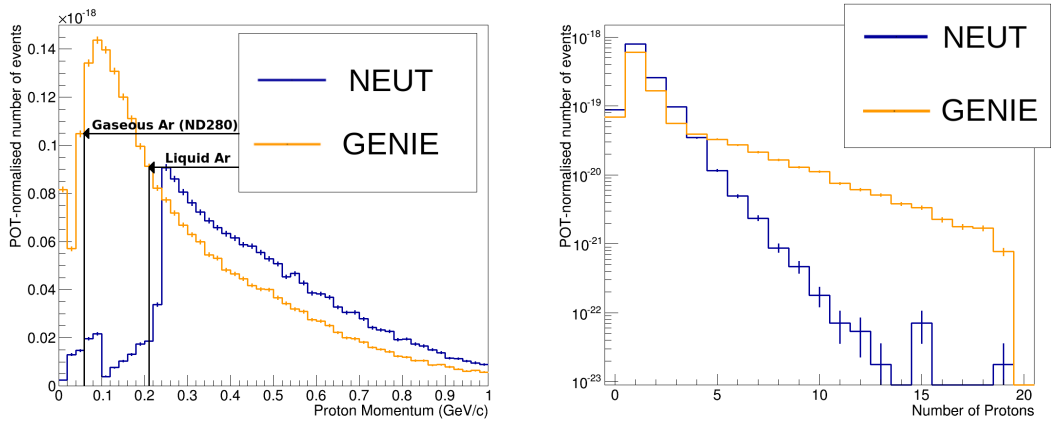


Figure 2.7: Plots showing the predictions from NEUT and GENIE for the momentum (left) and multiplicity (right) of protons leaving the target nucleus for charged-current ν_μ interactions on Ar in the T2K beam. Of particular interest are the model disagreements at high multiplicity and low momentum. The momentum distributions are annotated to show how the extra sensitivity of a gas detector covers the region of maximum tension between the two generators.

The momentum and multiplicity distributions present a consistent picture. GENIE favours many protons, carrying a commensurately smaller share of the neutrino momentum; NEUT favours fewer protons with higher momentum. In addition, the momentum distribution from NEUT displays a sharp "cliff" at the Fermi momentum (~ 250 MeV/c), while GENIE shows a smoother distribution at low momenta.

No data exist for the tuning of these predictions in this low-momentum region, and so it is not necessarily surprising if they should differ significantly from each other — or from nature. The high-multiplicity tail from GENIE seems particularly implausible, as it shows a significant fraction of events with 19 emerging protons. This indicates that in addition to the conversion of a neutron to a proton by the charged-current interaction itself, every single proton in the argon nucleus (atomic number 18) has been ejected.

Using the ND280 TPCs to analyse data from this region therefore represents a unique opportunity to observe which — if any — of the models currently in use correspond most closely to nature, and to improve our understanding and modelling of the processes that take place inside the struck nucleus.

Since many of the tools to make these measurements needed to be developed from scratch, I decided on a measurement of the proton multiplicity as the goal for this thesis, as the more simple observable to assess: measuring the proton multiplicity only relies on correct particle identification, whereas measuring momentum would also depend on correct reconstruction of the track curvature. The T2K gas interactions group will continue to develop the tools built for this analysis in order to produce a future measurement of the proton momentum.

Chapter 3

T2K and ND280

3.1 An Introduction to T2K

T2K ("Tokai to Kamioka") is an accelerator-based neutrino oscillation experiment in Japan. Constructed to look for ν_e appearance in a ν_μ beam, it fires neutrinos across a baseline of 295 km — from the JPARC accelerator complex in Tokaimura on the east coast of Japan, to the Kamioka Observatory in the Hida mountains. T2K discovered $\nu_\mu \rightarrow \nu_e$ appearance in 2013, and ruled out $\theta_{13} = 0$ with a significance of more than 7σ [24].

T2K uses an off-axis beam (see 3.2) to achieve a narrow neutrino energy spectrum. It has two near detectors to measure the beam characteristics near the point of origin: INGRID, the on-axis near detector (Section 3.3.2) and ND280, the off-axis near detector (Section 3.3.3). Super-Kamiokande ("Super-K") is used as a far detector to measure the oscillated neutrino spectrum (Section 3.4). Of these, ND280 has been the focus of my work, and will be covered in the most detail.

Since passing the 5σ threshold for ν_e appearance, T2K has been pursuing an expanded physics programme, including antineutrino beam running in order to access the complex phase of the PMNS matrix, δ_{CP}. With only 13% of its projected data gathered so far, T2K has much more yet to contribute to the field of neutrino physics.

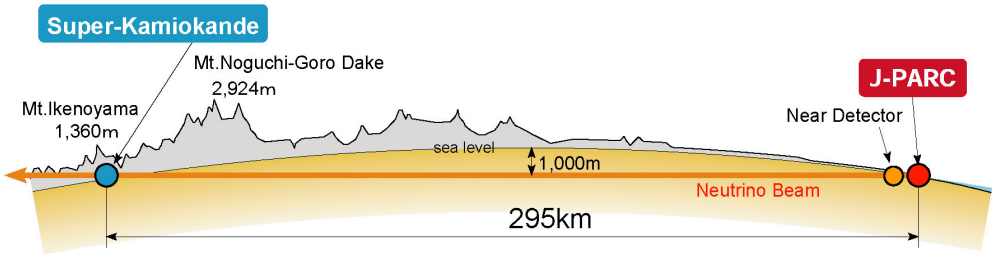


Figure 3.1: A diagram illustrating the layout of the T2K experiment [3]. Modified for clarity by T. Ishida.

3.2 Creating the Neutrino Beam

3.2.1 The Off-Axis Method

T2K uses the off-axis method to generate a narrow-band neutrino beam. Both the far detector (3.4) and one near detector (3.3.3) are placed 2.5° away from the beam axis, where they receive a narrow spectrum of neutrino energies with a peak at ~ 0.6 GeV.

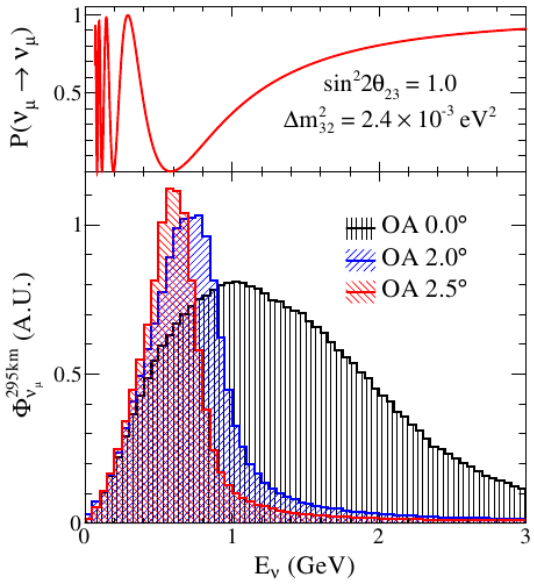


Figure 3.2: A plot showing the neutrino beam spectra for various off-axis (OA) angles [4]. The ν_μ survival probability is drawn along the top, showing that the peak of the 2.5° OA spectrum coincides with the first oscillation maximum.

This approach maximises ν_e appearance at the far detector while minimising backgrounds from the intrinsic ν_e contamination of the beam [3], at the minor cost of a

reduction in the total neutrino flux received — not a problem in T2K's high-intensity neutrino beam.

3.2.2 Proton Accelerator

The J-PARC (Japan Proton Accelerator Research Centre) accelerator complex consists of three accelerators:

- A linear accelerator (LINAC)
- A rapid-cycling synchrotron (RCS)
- The main ring synchrotron (MR)



Figure 3.3: An aerial photograph of the J-PARC accelerator complex.

H^- ions are accelerated by the LINAC to 181 MeV, then injected into the RCS. Charge-stripping foils at the point of injection convert the H^- beam to a proton beam. The RCS then accelerates these protons up to 3 GeV.

The MR takes 5% of the proton bunches from the RCS, with the rest being supplied to other experiments at the J-PARC site. Protons in the MR are accelerated up to 30 GeV before being directed down the neutrino beamline to the target, providing a proton beam with the characteristics given in Table 3.1 [3].

Successive upgrades are planned to raise the beam power to 750 kW before the end of the experiment.

Beam power	~ 220 kW
Beam kinetic energy	30 GeV
Beam intensity	$\sim 3 \times 10^{14}$ protons per spill
Spill cycle	~ 0.5 Hz
Number of bunches	8 per spill
Spill width	~ 5 μsec

Table 3.1: The characteristics of the J-PARC beam.

3.2.3 Neutrino Beamline

The neutrino beamline is broken into two sections: primary and secondary. In the primary beamline, the proton beam is deflected towards Kamioka; in the secondary beamline, the proton beam impinges on the target to produce the neutrino beam.

Primary Beamline

The primary beamline consists of three sections: the primary section, the arc section and the final focusing section. The preparation section tunes the proton beam for acceptance by the arc section; the arc section bends the beam through 80.7° towards Kamioka, and the final focusing section guides and focuses the beam onto the target, while bending it 3.6° downwards to pass through the Earth's crust.

The intensity, position, profile and loss of the proton beam in the primary sections are measured by 96 separate instruments, including five current transformers (CTs) which monitor the beam intensity. Each CT is a 50-turn toroidal coil, which measures the induced current as each bunch passes through the coil. The CTs measure the absolute beam intensity with a 2% uncertainty and the beam timing to a precision better than 10 ns [3].

Measurements from the last CT before the protons reach the target (CT5) are used to calculate the total protons on target received by T2K, as a direct measure of the data collected by the experiment. This is shown in Figure 3.4 below.

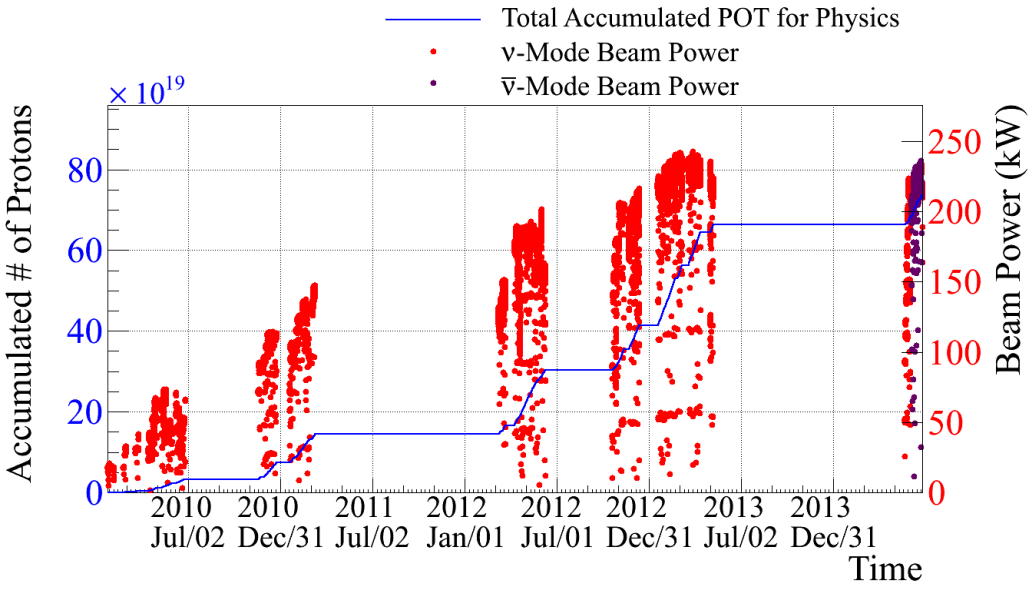


Figure 3.4: Official plot of the POT delivered by the T2K beam up to the present time, in both ν and anti-ν mode.

Secondary Beamline

In the secondary beamline, the proton beam impacts on a cylindrical graphite target, 91.4 cm ($\sim$ 2 interaction lengths) long and with a diameter of 2.6 cm. This produces a spray of predominantly pions, with some kaon contamination.

The charged particles produced are focused by a series of three magnetic horns, which re-collimate the beam and sweep out wrong-sign particles. By reversing the current direction in the horns, it is possible to invert the charge selection at this stage and create an antineutrino beam instead of a neutrino beam. Antineutrino running on T2K began in May 2014.

The focused beam of pions is then allowed to decay in flight in a $\sim$ 96 m long steel tunnel with water-cooled walls. At the end of this decay volume sits the beam dump, which absorbs the muons from pion decays while allowing the neutrinos to stream through. The beam dump is made from 75 tonnes of graphite, and is $\sim$ 3 m deep.

The entire secondary beamline is filled with helium to reduce pion absorption and suppress tritium and NO_x production by the beam.

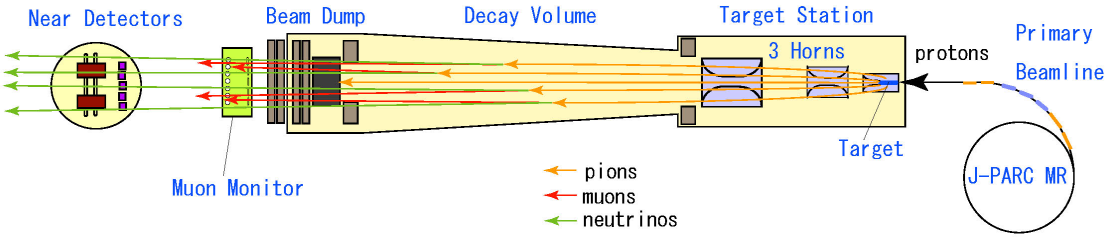


Figure 3.5: A schematic of the secondary neutrino beamline. Official T2K figure provided by T. Ishida.

3.3 Near Detectors

3.3.1 The ND280 Complex

Both near detectors sit in a cylindrical pit, located 280 m from the beam target. The pit is 37 m deep and lined with concrete, while due to its location (see Figure 3.3) the surrounding ground is mostly sand.

3.3.2 INGRID

INGRID (Interactive Neutrino GRID) is T2K's on-axis near detector. Positioned in the dead centre of the J-PARC neutrino beam, it is used to monitor the beam direction and intensity through measurements of neutrino interactions on iron.

INGRID consists of 16 identical modules. 14 of these are laid out in a cross, with seven in each "arm". Two additional modules are mounted at off-axis positions from the main cross, as shown in Figure 3.6. These two modules serve to measure the axial symmetry of the beam.

Each INGRID module consists of eleven layers of plastic scintillator interleaved with nine iron plates, and surrounded by a veto layer of scintillator to reject entering particles. With the iron plates being 1.24 metres square and 6.5 cm thick, each module contains 7.1 tonnes of iron, providing a large, dense target for neutrino interactions.

The internal scintillator layers are then used for tracking, to identify the muons from these interactions in the iron. Each scintillator plane is segmented into bars and read out by wavelength-shifting fibres. A sample event display from one INGRID module is shown in Figure 3.7.

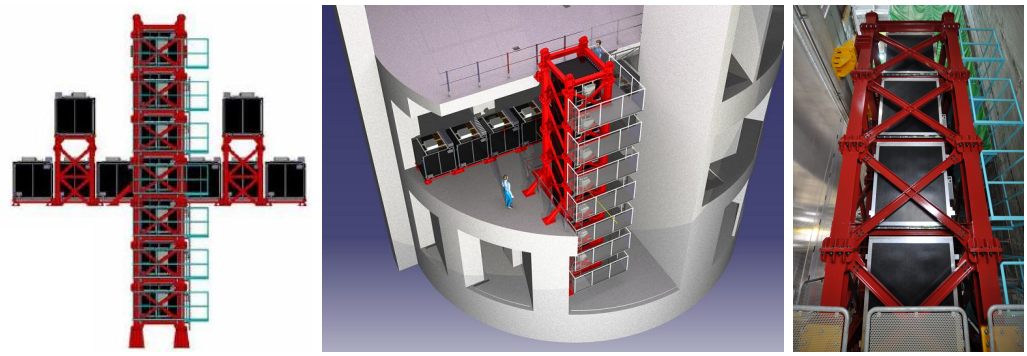


Figure 3.6: A diagram showing the layout of the INGRID detector as seen from the beam direction [3] (left), a diagram showing the placement of INGRID in the pit (centre), and a photograph of INGRID's vertical modules (right).

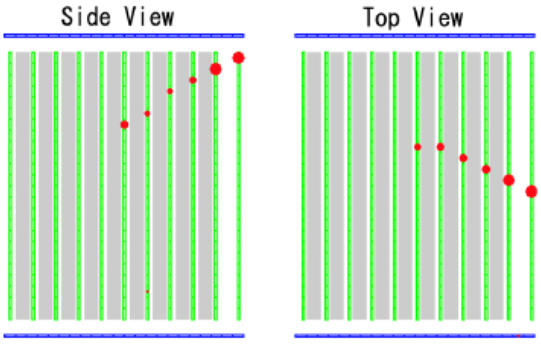


Figure 3.7: A muon track in an INGRID module, seen from both the side and top views [3].

INGRID successfully constrains the beam direction to within 0.4 mrad [3].

3.3.3 ND280

ND280 (Near Detector at 280 m) is T2K's off-axis near detector. It is used to characterise the flux, energy spectrum and electron neutrino contamination of the unoscillated beam, as well as making measurements of neutrino cross-sections. These observations are used to make predictions of signal and background at Super-K, which sits at the same off-axis angle of 2.5°.

ND280 consists of many different subdetectors, which can be split into three main groups:

- The π^0 detector (P0D), designed to measure the neutral-current production of π^0s on a water target in the same beam conditions as Super-K.

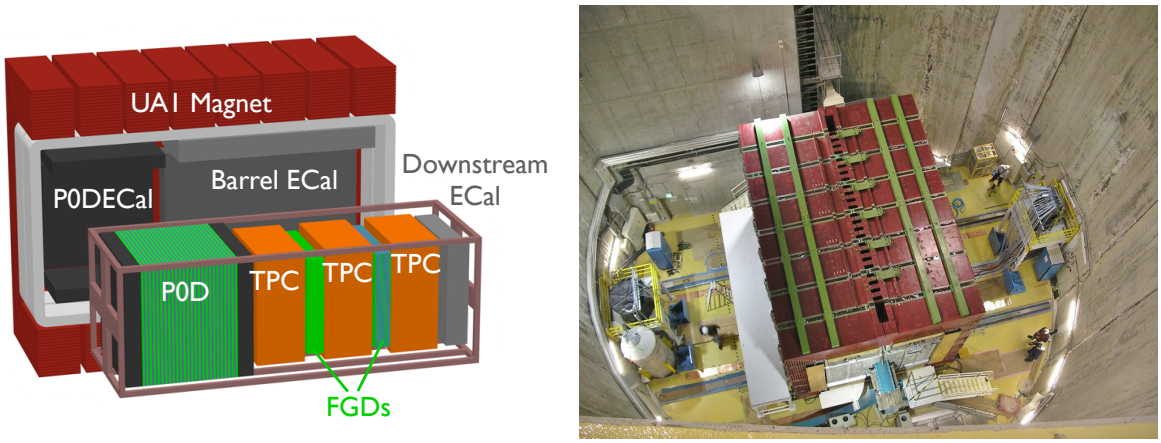


Figure 3.8: An exploded cross-section of ND280 [3] (left), and a photograph of ND280 with the magnet closed (right).

- The tracker, comprised of three time projection chambers (TPCs) and two fine-grained detectors (FGDs). These are used to perform detailed spatial reconstruction and particle identification on neutrino interactions in the tracker volume.
- Electromagnetic calorimeters (ECals) surrounding both the P0D and tracker, giving full containment for measuring the energy of neutrino interaction products created inside.

In addition there is the side muon range detector (SMRD), which measures muons leaving the detector at high angles and provides a trigger for incoming cosmic muons. The SMRD does not feature in this analysis, and so will not be described in detail.

The entire detector sits inside the repurposed UA1 magnet, which provides a uniform magnetic field of 0.2 T perpendicular to the beam direction.

P0D

The P0D is built to measure the process $\nu_\mu + N \rightarrow \nu_\mu + N + \pi^0 + X$. Neutral pion decays with highly colinear photons can emulate the Cherenkov rings formed by electrons in water, making them a significant component of the background to ν_e appearance at Super-K; therefore an accurate parameterisation of this rate is very important for T2K oscillation analyses.

The P0D is built from planes of scintillator bars, interleaved with layers of target material: fillable water bags, and lead and brass sheets. It is broken into four regions:

the upstream ECal, the upstream water target, the central water target, and the central ECal.

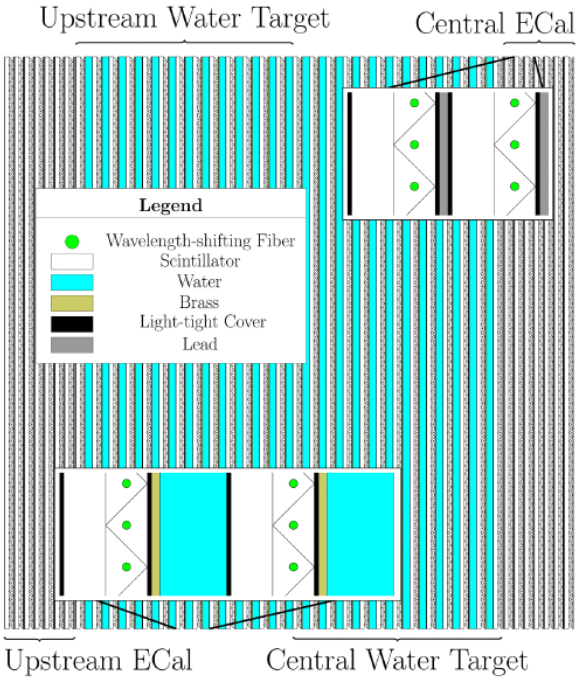


Figure 3.9: A diagram showing the construction of the P0D [3], with the beam entering from the left. The triangular segmentation of the scintillator bars is shown in close-up in the insets.

Each triangular bar is threaded through its centre with a wavelength-shifting fibre to carry the scintillation light produced to its sensor. The fibres are mirrored on one end, and read out at the other by a multi-pixel photon counter (MPPC).

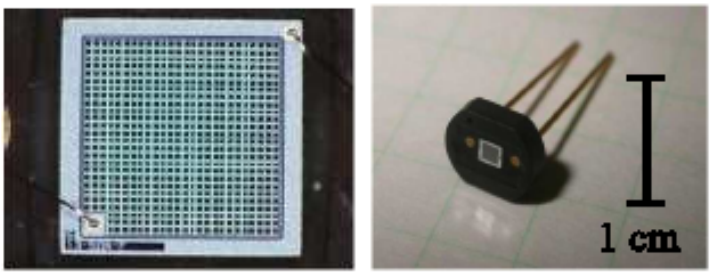


Figure 3.10: An MPPC of the type used in the ND280 P0D, FGDs and ECals. Left: magnified view of MPPC face, showing 26×26 pixel array. Right: ceramic package of the same MPPC.

MPPCs are avalanche photodiodes, which function in the 0.2 T magnetic field of ND280 where photomultiplier tubes would not. Each MPPC consists of 667 pixels

(each in Geiger mode) over an area 1.3×1.3 mm^2 (as shown in Figure 3.10), each of which counts photons incident on its surface with a gain of the same order as a vacuum photomultiplier [3]. More details on MPPCs can be found in the 2009 paper of Renker and Lorenz [39].

In this analysis, only the central ECal of the P0D is used (to provide a veto for TPC1). I will therefore not go into detail on the P0D water system.

FGDs

The FGDs are ND280's primary targets for neutrino interaction measurements. FGD1 consists purely of plastic scintillator, while FGD2 also contains water to allow cross-section measurements on water for use at Super-K. Each FGD contains 1.1 tonnes of target material.

The FGD scintillator bars are rectangular, with dimensions 9.61 mm $\times$ 9.61 mm $\times$ 1861 mm (hence "fine-grained"). They are arranged perpendicular to the beam direction, in alternating x and y orientations for 3D reconstruction (as in the P0D). This arrangement is illustrated in Figure 3.11.

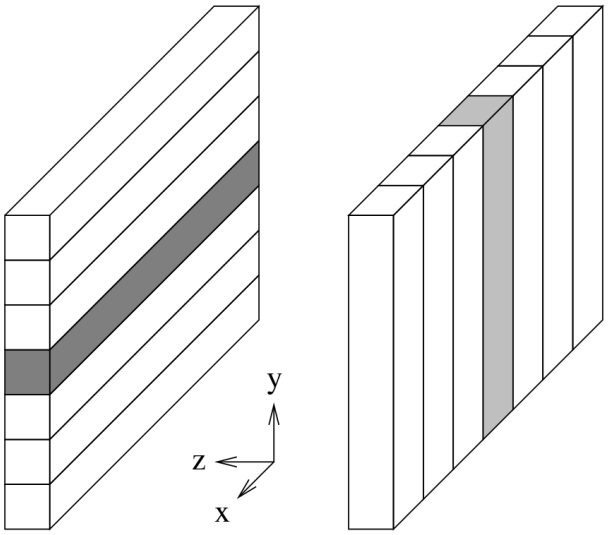


Figure 3.11: Illustration of the perpendicular layer configuration for the FGDs and ECals.The shaded bar on the left has its y coordinate defined by the bar and its z coordinate defined by its layer; the shaded bar on the right has its x coordinate defined by the bar and the z coordinate defined by its layer. Combining the 2D information from successive alternating layers allows for the reconstruction of a complete 3D track. Figure drawn by Luke Pickering.

In FGD2, paired layers of scintillator bars alternate with water layers of depth 2.5 cm. In both FGDs, scintillation light is read out through wavelength-shifting fibres threaded through the centre of each scintillator bar and connected to MPPC sensors, in the same manner as the P0D.

While the FGDs are the target of choice for most ND280 analyses, in this analysis they are used only to provide a veto for entering backgrounds in TPCs 2 and 3.

TPCs

ND280 has three identical time projection chambers, designed for tracking of interaction products from the FGDs, and measurements of their momentum and particle type. The search for neutrino interactions in the TPC drift gas is the focus of this analysis.

Each TPC consists of a copper-clad inner box containing argon-based drift gas, insulated by an aluminium-clad outer box filled with CO_2. This structure is illustrated in Figure 3.12.

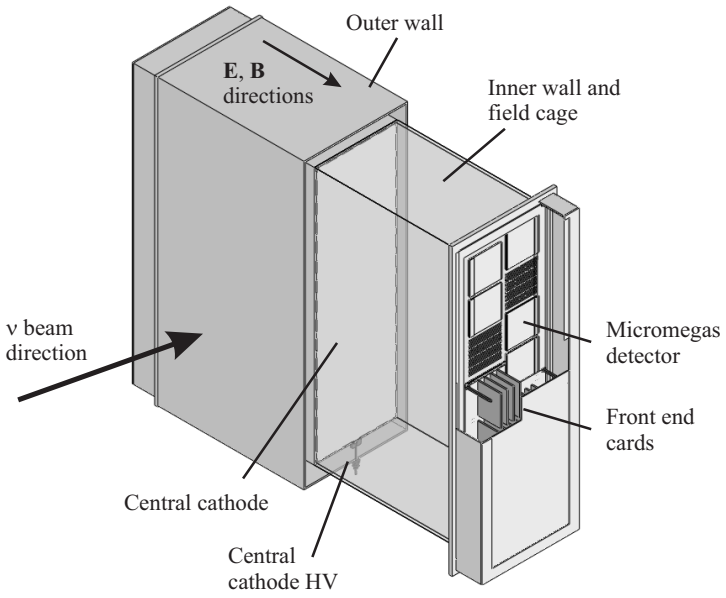


Figure 3.12: Cutaway drawing showing the TPC design [3]. The outer box has dimensions $2.30 \times 2.40 \times 0.97$ m ($x \times y \times z$, where z is the beam axis and y is normal to the surface of the earth); the active gas volume contains 3,000 L of gas mixture, with a further 3,300 L in the gap volume between the field cage and the outer wall.

A copper cathode panel runs along the z (beam) axis through the centre of each TPC inner box, and an (approximately) uniform electric field applied along x (parallel to the B-field), with a maximum field strength of 5 kV/cm (at the TPC walls) [5]. The drift gas through which the E-field is applied is composed of Ar, CF_4 (tetrafluoromethane), and iC_4H_{10} (isobutane) in a ratio of of 95:3:2 respectively.

Charged particles passing through the inner box ionise the drift gas, leaving a trail of ionisation electrons which drift along the direction of the E-field to the anodes (the TPC walls). Each wall is instrumented with 12 bulk sensor modules, allowing the net charge and arrival time of drift electrons to be read out.

The TPC sensors use the bulk MicroMEGAS technology invented in 2004 by Giomataris *et al* [40][41], in which a woven mesh is laminated onto a printed circuit board (PCB) covered by a photoimageable film. The board is then UV-irradiated and chemically treated to create a highly segmented detector, which detects incident electrons through amplification in the tiny mesh-anode gap when a voltage difference is applied. This has the advantage of creating sensor pads with very fine segmentation, while requiring no bulky infrastructure to separate them.

The ND280 MicroMEGAS modules have 9.8 mm $\times$ 7.0 mm pads ($y \times z$), with 1726 pads per module arranged in a 48 $\times$ 36 grid for a total sensitive area of 36 $\times$ 34 cm^2 per module. The dimensions of the pads were determined by cost and the design specification of less than a 2% uncertainty on the momentum scale; the spatial resolution from transverse diffusion for electrons travelling the maximum drift distance is less than 0.5 mm [5]. A close-up of one of the MicroMEGAS modules is shown in Figure 3.13.

ECals

Both the P0D and tracker of ND280 are surrounded by electromagnetic calorimeters: the P0D, Barrel and Downstream (DS) ECals, as shown in Figure 3.3.3. These are used to contain and assess the energy of interaction products produced in the interior subdetectors.

The technology used in the ND280 ECals is similar to that used in the FGDs. Thin layers of lead absorber are interleaved with planes of rectangular plastic scintillator bars in alternating orientations, read out by wavelength shifting fibres to MPPCs. As the ECals do not require the same fine granularity as the FGDs, their scintillator bars are much larger, with a 4 cm $\times$ 1 cm cross-section.



Figure 3.13: ND280 MicroMEGAS module in close-up.

The six barrel ECals surrounding the tracker each have 32 layers, for a thickness of $9.7X_0$. The DS ECal has 34 layers, for a thickness of $10.6X_0$. In both the barrel and DS ECals, the thickness of the lead layers is 1.75 mm. The P0D ECal by comparison is more lightweight, with only six layers. To compensate for this, its lead layers are 4 mm thick, giving a thickness of $3.6X_0$. All the scintillator planes in the P0D ECals are also oriented parallel to one another, giving only 2D reconstruction.

The ECals are the most massive subdetector set in ND280, with a total mass of more than 30 tonnes — most of it lead. Neutrino interactions in the lead of the ECals with particles entering the TPCs are therefore a significant background in the search for neutrino interactions in the TPC gas.

Data Acquisition

Data from ND280 is recorded by the data acquisition system (DAQ) whenever one of three trigger conditions are fulfilled:

- **Beam Trigger:** A signal is sent to the DAQ whenever a pulse is sent through the neutrino beamline, instructing the detector to record data.
- **FGD Cosmics:** When there is no beam trigger, but hits are seen in both FGDs, this is taken as the signature of a cosmic ray muon passing through the detector. These are useful for calibration and therefore the trigger fires to record the event.

- **TRIP-t Cosmics:** When hits are seen on opposite sides of the detector (in the SMRDs or in the P0D and DS ECal), this is also taken as the signature of a cosmic ray muon, and the event is recorded.

Due to the low rate of neutrino interactions in matter, the event frequency is not high enough to necessitate any more strict trigger requirements as used in collider experiments.

Data taken by ND280 is stored initially at the KEK computing centre in Japan, before being copied to TRIUMF in Canada and RAL in the UK.

3.4 Far Detector: Super-K

Super-K is, at the time of writing, the largest land-based water Cherenkov detector on the planet. Built under 1 km of rock, its active volume contains 50,000 tonnes of ultra-pure water. 13,000 inward-facing photomultiplier tubes (PMTs) image neutrino interactions in the water through the Cherenkov light produced by charged interaction products.

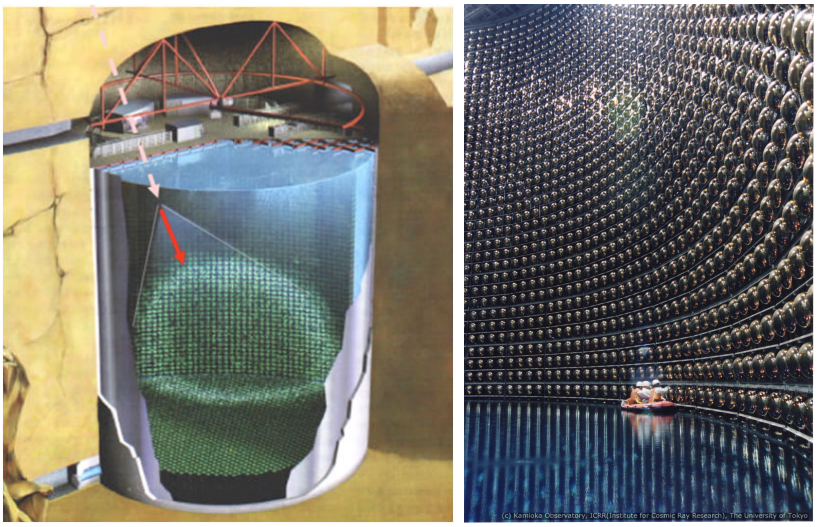


Figure 3.14: A diagram illustrating a neutrino interaction in Super-K from the Super-K website (left), and a photograph of the detector's interior wall (right).

Cherenkov radiation is the result of charged particles travelling faster than the speed of light in their local medium ($v > \frac{c}{n}$, where n is the refractive index of the medium). This produces a conical 'bow-wave' of electromagnetic radiation, with opening angle θ determined by $\cos\theta = \frac{c}{nv}$.

As charged particles travel in the water of Super-K, they lose energy until their velocity no longer satisfies the Cherenkov condition $v > \frac{c}{n}$. The PMTs therefore image a ring of light that reaches the detector walls, as shown in Figure 3.15.

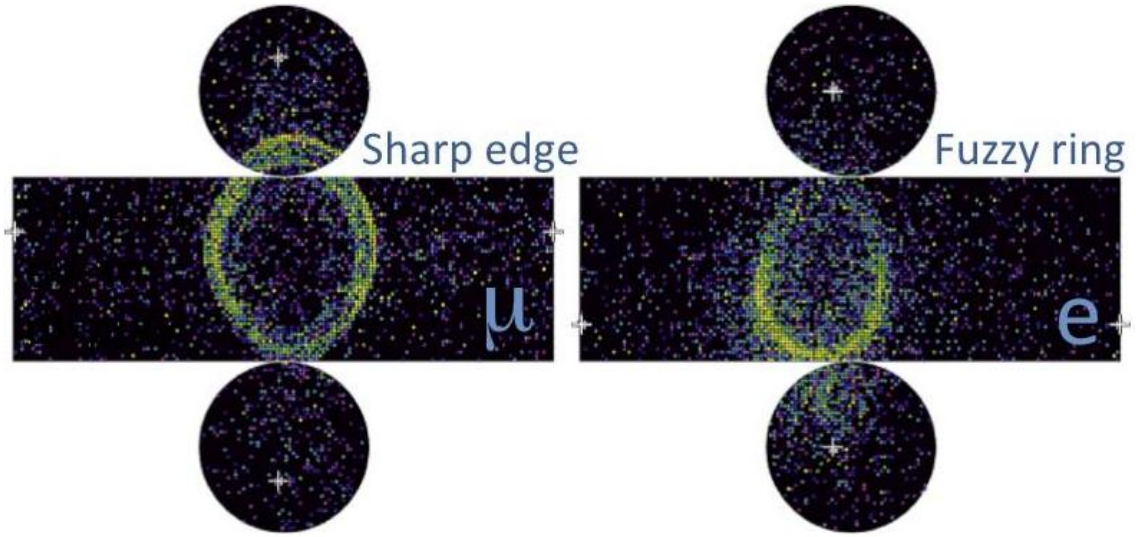


Figure 3.15: Event displays from Super-K, showing the sharp ring produced by a muon (left) and the "fuzzy" ring produced by an electron (right). Event display provided by B. Still.

Super-K can perform particle identification by examining characteristics of these rings. Muons are minimum-ionising particles in the water, and therefore travel straight through to produce a clean cone of Cherenkov light that leaves a sharp ring on the detector wall. By contrast, electrons will produce electromagnetic showers, creating a cascade of overlapping Cherenkov cones which leave a "fuzzy" ring. This difference can be used to discriminate between electrons and muons with a purity of $> 99\%$, while selecting electrons from ν_e interactions with an efficiency of 66% [42][43].

Super-K is used as T2K's far detector, looking for ν_e appearance (and ν_μ disappearance) in the ν_μ beam. This analysis uses only the ND280 near detector, but aims to constrain systematic uncertainties from nuclear effects (see Chapter 2) that affect these oscillation analyses.

Chapter 4

ND280 Software and Simulation

ND280 uses its own software to analyse data and produce Monte Carlo simulation (from here forward referred to as "Monte Carlo" or "MC"). This software is written in C++, based on ROOT [44] and uses Geant4 [45] for simulation. The software chain has several stages, and contains different packages to handle each of the different subdetector systems contained by ND280. In the context of this thesis, I will focus on the software used for reconstruction in the TPCs. This software was partially rewritten for the purposes of this analysis (see Section 4.5), but much of its underlying structure remains the same.

4.1 Overview

A top-level overview of the steps in the software chain for both data and Monte Carlo is given in Figure 4.1. Once the physics and detector response of a Monte Carlo spill have been simulated, it is processed identically to an equivalent event in data.

Most steps in the software chain require different treatments for the different subdetectors, meaning that each step is split into subdetector-specific packages — e.g. oaRecon contains ecalRecon, fgdRecon, tpcRecon etc., as well as a global reconstruction routines to connect the output of these local modules together.

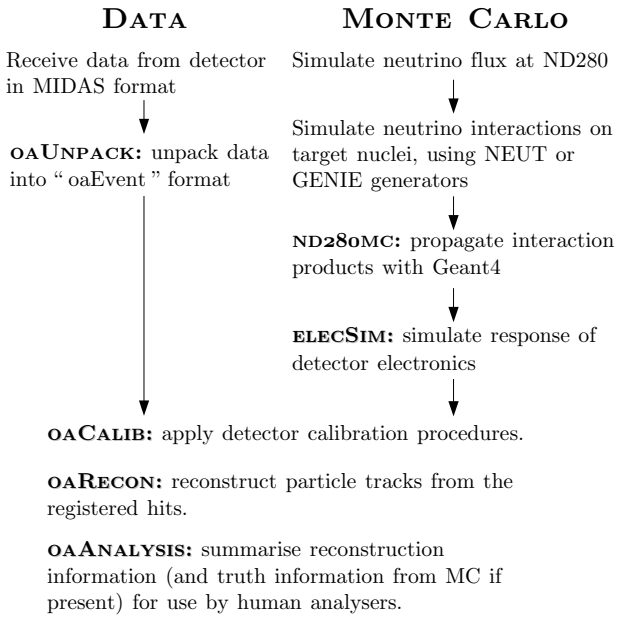


Figure 4.1: A diagrammatic overview of the steps used by the ND280 software. The prefix "oa" stands for "off-axis" and is used ubiquitously throughout the ND280 software.

4.2 Monte Carlo simulation

4.2.1 Beam simulation

The flux of neutrinos at ND280 is predicted by a full simulation of the beam's production (see Section 3.2.3), starting with 30 GeV protons in the primary beamline. The interactions of these protons with the target and baffle are modelled with FLUKA2011 [46], and the subsequent propagation of the interaction products through the secondary beamline is handled by GEANT3 [47]. These daughter particles are tracked until they either interact or decay; by saving the properties of any and all resulting neutrinos (as well as the interaction chain that produced them), a full picture of the beam is built up one particle at a time. An example simulated beam spectrum is shown in Figure 4.2.

This simulation is tuned with measurements taken directly from the beamline, using differences between the measured and simulated beam profile to reweight the flux. The T2K beam group provides reweighting histograms binned in true neutrino energy for this purpose for each run. The tuning weights are also shown in Figure 4.2.

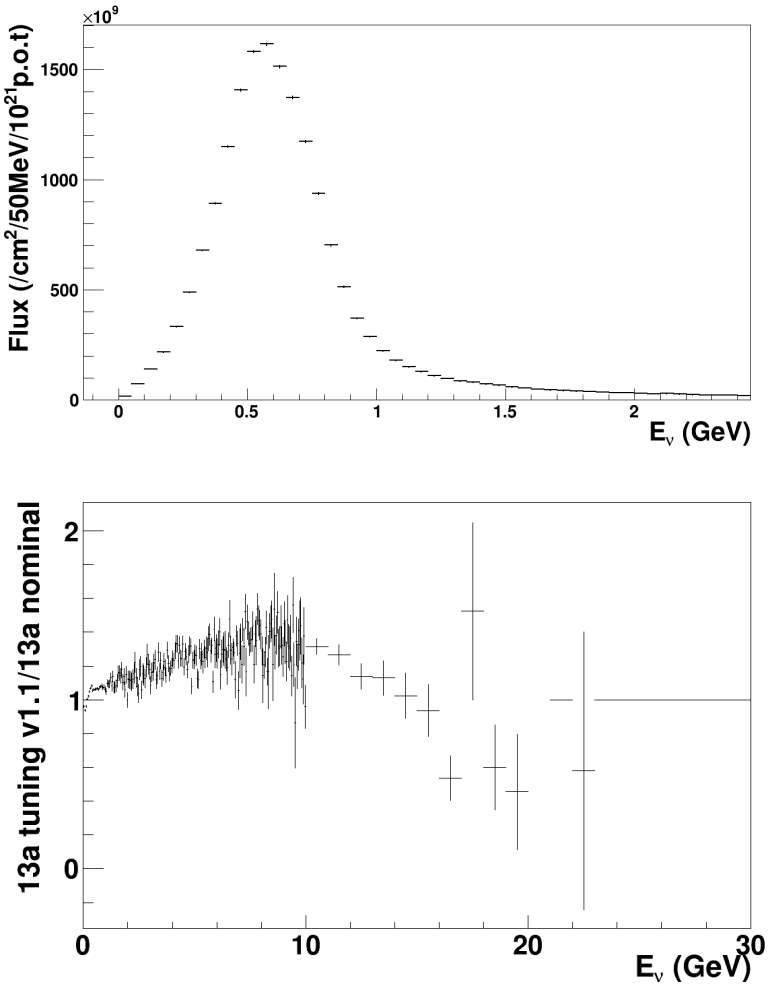


Figure 4.2: A plot of the nominal ν_μ flux from simulation (top), and a plot of the weights that are applied during the tuning (bottom). Different weights are applied for different neutrinos (i.e. $\overline{\nu}_\mu$, ν_e).

The beam simulation is also tuned with external data from the dedicated hadron interaction experiment NA61/SHINE [48]. NA61/SHINE was built to measure hadron production from protons of momentum 31 GeV/c incident on graphite targets, including a replica of the T2K target. The experiment can detect pions and kaons in 90% and 60% (respectively) of the phase-space relevant to T2K; other experiments [49] [50] are used to constrain the phase-space not covered by NA61.

4.2.2 Neutrino interaction simulation: NEUT and GENIE

The ND280 software can produce Monte Carlo using two generators: NEUT [51] (the default generator) and GENIE [52]. These generators can model the interactions of neutrinos across a broad energy range ($\sim$ 10 MeV to $\sim$ 100 TeV) and on all the nuclear targets present in ND280, including the magnet return yoke. Interactions in the surrounding structure of the pit are not included in the standard simulation, instead being handled by the production of specific "sand" Monte Carlo.

Both generators work on same principle: they track individual neutrinos provided by the beam simulation, calculate the probability of interaction with all the matter in their path, and use a pseudo-random number generator to determine whether an interaction takes place. The models and inputs used to predict these probabilities differ between the two generators, and are detailed below.

NEUT [34]

- **CCQE:** Llewellyn Smith model [53].
- **Nuclear model:** Smith and Moniz's relativistic Fermi gas model [29].
- **Pion production:** Rein-Sehgal model [54].
- **Multi-pion and DIS processes:** two processes, depending on the energy of the hadronic system. PYTHIA/JETSET [55] is used for energies above 2GeV, and internal NEUT code is used below this limit, as PYTHIA/JETSET does not reproduce experimental data well at lower energies.
- **Meson exchange currents:** Nieves "two-particle, two hole" (2p2h) model [33].
- **Final state interactions:** Handled by internal NEUT code.

GENIE [35]

- **CCQE:** Llewellyn Smith model [53]
- **Nuclear model:** Bodek-Ritchie relativistic Fermi gas model [56].
- **Pion production:** Rein-Sehgal model [54]
- **Multi-pion and DIS processes:** Code based on KNO scaling [57] is used for low invariant masses, with a gradual transition to PYTHIA/JETSET over the invariant mass window 2.3–3.0 GeV/c^2.
- **Meson exchange currents:** None.

- **Final state interactions:** Handled by the INTRANUKE sub-package.

Neutrino interactions in the walls of the pit ("sand muons") are simulated separately to interactions in the detector, and must be manually combined with detector Monte Carlo in the correct proportions.

4.2.3 Detector simulation

Beam spills in the detector are simulated by grouping together individual neutrino interactions from NEUT or GENIE, in numbers defined by the intensity of the simulated beam (with Poisson variations to account for statistical fluctuations). These spills are then grouped in time according to the beam bunch structure.

A detailed simulation of the geometry of each subdetector is generated in ROOT, taking their position, size and composition from their design specifications. Alignment constants provide the flexibility to adjust the position of the simulated modules based on measurements of the subdetectors *in situ*. Geant4 is then used to propagate the daughter particles from each neutrino interaction through this model of the detector, producing a list of Geant4 "hits" that log the time at which the particle passed them, their position, and the amount of energy the particle deposited at that point. These are then passed to the electronics simulation.

The electronics simulation software simulates the response of each subdetector depending on the technology employed. In the scintillator-based detectors (i.e. the P0D, FGDs, ECals and SMRD), it simulates the production of scintillation light, the propagation of light along the wavelength-shifting fibres and the response of the MPPCs. In the TPCs it simulates the drift of ionised electrons in the E-field and the response of the MicroMEGAS. For all subdetectors, the final step is simulating the behaviour of the readout electronics, to produce output in the same format as data.

4.3 TPC calibration

Reconstruction of tracks in the TPCs is dependent on accurate modelling of the E- and B-fields inside the TPC volume. Permutations in the fields - e.g. from edge effects - affect the speed and trajectory of drift electrons, distorting the image of the reconstructed

track. A calibration system is therefore necessary to constrain the uncertainties on the fields.

The magnetic field in the region occupied by the TPCs was mapped before installation using a Hall probe. In addition, the TPCs were constructed with thin aluminium discs affixed at regular intervals to either side of the central cathode. These discs are flashed with light from a 266 nm wavelength laser, producing photoelectrons at a known point from which the drift velocity can be checked. With 56 discs per MicroMEGAS module, this system gives good coverage over the full active volume for all three TPCs [5].

4.4 TPC reconstruction

In other ND280 analyses, the TPC reconstruction (tpcRecon) is used to provide precise tracking of particles originating from neutrino interactions in the FGDs, allowing for accurate determination of particle species and momentum. For this reason, tpcRecon focuses on reconstructing through-going particle tracks. Section 4.5 describes how the reconstruction was redesigned to look for neutrino interactions in the TPC gas itself for this analysis; many elements of the reconstruction, however, remain unchanged. Below I describe the existing reconstruction in brief, to introduce the fundamental concepts on which the enhanced reconstruction depends.

A "hit" in the TPCs corresponds to a single MicroMEGAS pad. Each MicroMEGAS pad outputs a waveform of the charge it receives as a function of time; the first stage of tpcRecon is to scan this waveform in search of peaks that rise above a noise threshold. Adjacent pads in the same horizontal row with charge peaks that overlap in time are clustered together. These clusters of hits are then combined together into particle paths by a cellular automaton, and a single-path likelihood fit is used to extract a reconstructed particle trajectory.

The distance along the drift direction (in the global coordinates of ND280, along the X axis) from the point at which the particle passed through the detector and the point at which the drift electrons from that section of its passage were registered by a MicroMEGAS pad must be deduced from the timing information. In addition to the time of the peak on the MicroMEGAS pad, tpcRecon requires a "T0" from hits in neighbouring scintillator-based subdetectors. By taking the T0 as the true time at which the particle passed through, tpcRecon can calculate the length of time for which the

electrons were drifting, and thus deduce the X-coordinates of the track. The T0 is taken from the time of the closest FGD hit to an extrapolation of the track in the Y-Z plane. If no suitable FGD hits can be found, tracks in TPC1 and TPC3 can also search the P0D and DS ECal, respectively.

Particle identification in the TPCs is based on particle's rate of energy loss as it travels through the gas, dE/dX. dE/dX is given by the Bethe-Bloch equation, and is dependent on $\beta \equiv v/c$. For a known particle momentum, dE/dX can be compared with the expected dE/dX for particle type α by defining a 'pull' δ_α, given by

$$\delta_\alpha = \frac{dE/dX_{\text{meas}} - dE/dX^{\alpha}_{\text{exp}}}{\sigma_{\text{meas}}}, \tag{4.1}$$

where dE/dX_{meas} is the measured energy loss, $dE/dX^{\alpha}_{\text{exp}}$ is the expected energy loss for particle type α, and σ_{meas} is the uncertainty on the measured energy loss. Figure 4.3 shows the excellent separation between true particle types in energy loss for most portions of the momentum spectrum, demonstrating the excellent particle identification capabilities of the TPCs (except in the small overlap regions).

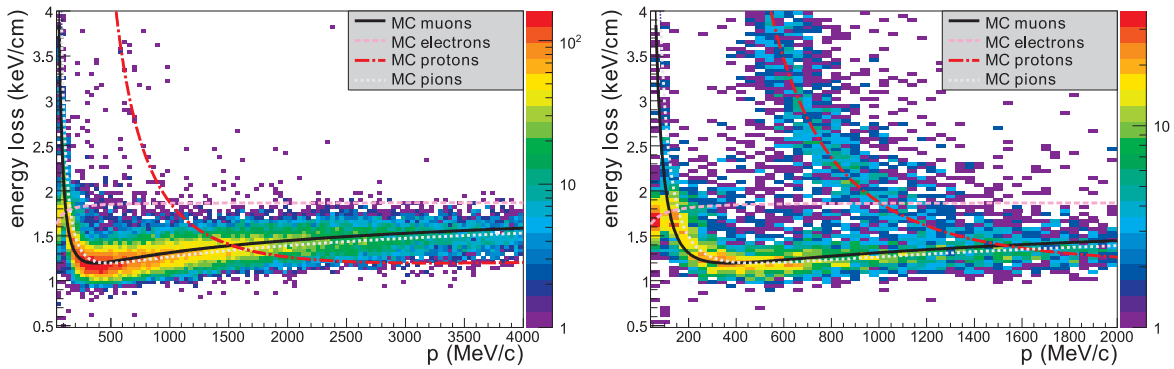


Figure 4.3: Plots showing the distribution of the energy loss as a function of the momentum for negatively (left) and positively (right) charged particles produced in neutrino interactions, compared to the expected curves from Monte Carlo for muons, electrons, protons and pions [5].

4.5 TREx

Identifying neutrino interactions in the TPCs requires entirely new reconstruction software, as gas interactions are not the event topology ND280 was designed to reconstruct. The



"ideal" configuration considered for most analyses is a neutrino interacting in one of the FGDs, with interaction products passing out through the TPCs. Examples of both topologies are shown in Figure 4.4.

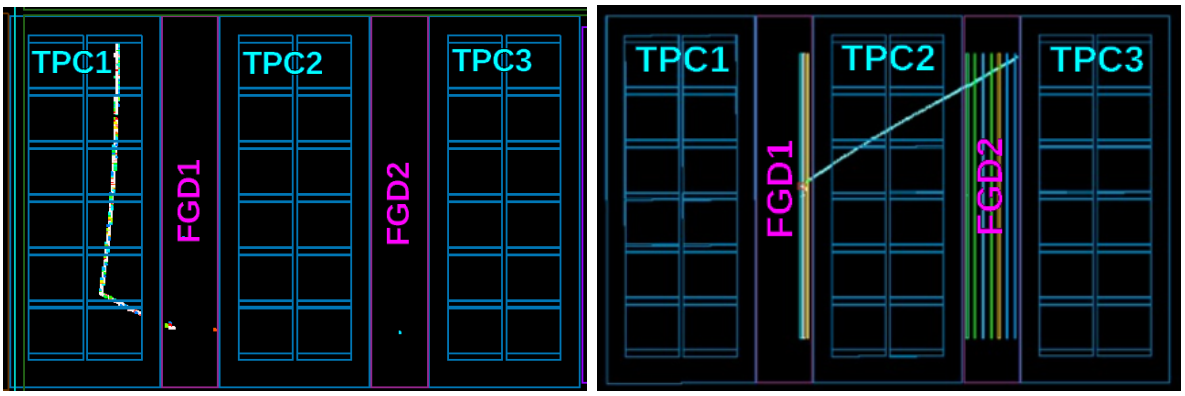


Figure 4.4: Two simulated ND280 event displays showing the event topologies looked for in this analysis (left) and for which the ND280 reconstruction software was designed (right). On the left, a neutrino interacts in the TPC drift gas, producing a vertex in the fiducial volume of the TPC. On the right, a neutrino interacts in the FGD, producing a muon track which propagates through the TPC downstream.

For this reason we developed the TPC Recon Extension — or "TREx" — in order to

- identify primary vertices in the gas,
- provide isotropic reconstruction, recognising tracks regardless of their angle from the beam axis,
- recognise complex paths, such as particles spiralling in the magnetic field,

The development of TREx was primarily the work of Dr. Anthony Hillairet, Eddy Larkin and myself, with greatly appreciated assistance from Dr. Tom Feusels, Dr. Stefania Bordoni and Lukas Koch.

The novel features of TREx are as follows:

- Isotropic hit clustering (as opposed to the horizontal hit clustering used in tpcRecon - see Section 4.4).
- A new method of pathfinding, using the well-established [58] A* algorithm to trace path between points of interest.
- Careful handling of secondary interaction products, e.g. delta rays from muon tracks.
- A new output structure of connected paths.

4.5.1 Pathfinding

The A* algorithm is used to find paths between given points of interest. The path is built up by stepping through discrete points in space (in this case, the TPC MicroMEGAS pads), always selecting the point that has the lowest cost. The cost at each step is composed of two terms: the connection cost, representing the sum of point-to-point connections made to reach the current position, and the heuristic cost, representing the estimated cost of reaching the end point from the current position.

The connection cost is defined $c_{\text{connection}} = |\vec{r}_{\text{next}}|^4$, where $\vec{r}_{\text{next}}$ is the vector of displacement between the current point and the point being considered for connection, scaled such that the shortest possible connection between adjacent MicroMEGAS pads has a connection cost of unity. Setting this cost to the fourth power has the effect of strongly favouring continuous paths of adjacent hits over paths with gaps; this in turn avoids creating paths that cut corners.

The heuristic cost is defined $c_{\text{heuristic}} = |\vec{r}_{\text{end}}|$, where $\vec{r}_{\text{end}}$ is the vector of displacement between the current point and the given end point the algorithm is trying to reach.

This procedure is illustrated in Figure 4.5.

In order to use the A* algorithm in this way, TREx must first identify the points it wishes to connect. It does this with a multi-stage approach:

1. **First Pass Edge Detection** All MicroMEGAS pads at maxima and minima in X, Y or Z are selected to identify the points of interest at which particles enter or leave the detector ("edge points"). Adjacent pads are clustered into single edge points.
2. **Second Pass Edge Detection** The pathfinding connects all existing edge points from the first pass in all possible combinations, using A*. All hits used by these connections are removed, and the first pass edge detection is repeated to identify

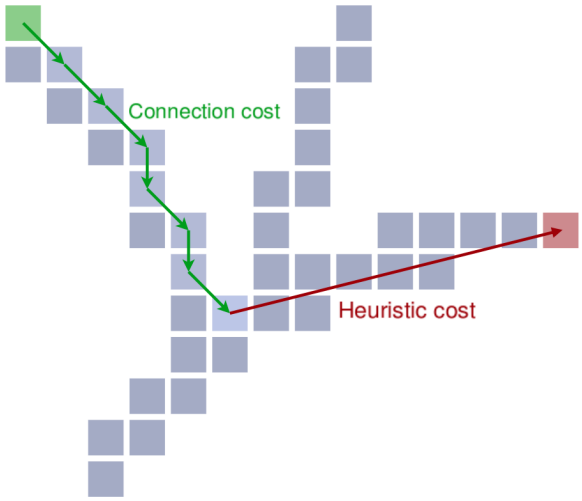


Figure 4.5: A diagram illustrating the A* pathfinding procedure, connecting activated MicroMEGAS pads to navigate between a start point (green) and an end point (red). The small green arrows represent the connection cost ($c_{\text{connection}} = |\vec{r}_{\text{next}}|^4$) at each step; the long red arrow represents the heuristic cost $c_{\text{heuristic}} = |\vec{r}_{\text{end}}|$. Diagram provided by E. Larkin.

track ends that lie inside the pattern's extrema (e.g. stopping particles). These track ends are added to the list of edge points.

3. **Redundant Edge Removal** The list of edge points may now contain some false positives, e.g. from curving tracks. All existing pairs of edge points are connected using A*, and any edge points that are found to lie on a path between two other edge points are removed.

4. **Junction Detection** If three or more edge points emerge from the first two stages, the pathfinding now searches for junctions where these paths meet. These are identified by using A* to navigate between the edge points, and finding the position at which the paths begin to diverge, as illustrated in Figure 4.6.

5. **Kink-Finding** Vertices with only two outgoing particles will not yet have been found by the edge or junction detection. To identify these, the pathfinding looks for sharp kinks in the path. These can be identified by testing whether the number of hits above a given fractional distance from the line connecting the path's two edges is less than a given fraction of the total number of hits in the path, as illustrated in Figure 4.7.

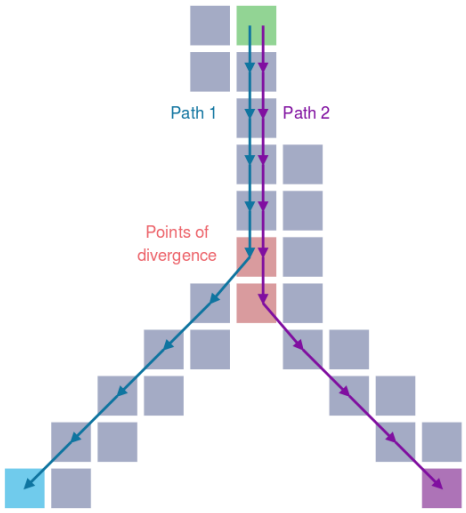


Figure 4.6: A diagram illustrating the procedure for finding junctions, provided by E. Larkin.

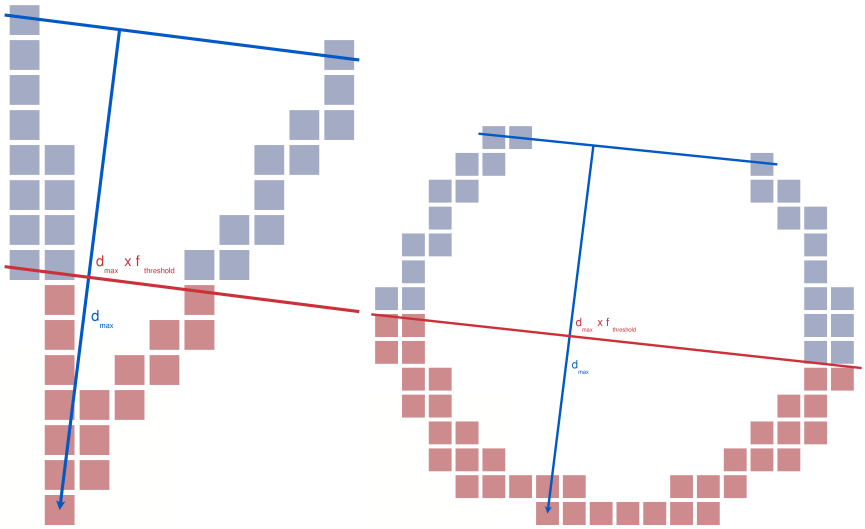


Figure 4.7: A diagram illustrating the difference in hit profile between kink-like vertices and curved tracks, provided by E. Larkin.

4.5.2 Output Structure

Once the tracks in the TPC have been traced and their meeting points identified, a new output structure is necessary for analysis of the reconstructed event. Rather than the old structure, which was designed only for through-going tracks, TREx is capable of producing a hierarchy of objects: patterns, paths and junctions.

These classes are briefly described below, as the terminology of patterns, paths and junctions is key to the cuts made in the final selection.

- **Patterns** are formed of any contiguous group of TPC hits connected together by the pathfinding. While it is possible for unconnected groups of hits to form separate patterns in the same TPC, a correctly reconstructed primary vertex in the gas should form only a single pattern in the TPC in which it occurs. Patterns can contain an arbitrary combination of paths and junctions.
- **Paths** are routes found by the pathfinding between points of interest. A single particle trajectory will ideally create only a single path in its TPC, but may break into multiple paths if it passes any points of interest before leaving the TPC (e.g. the production of a delta ray). In addition to belonging to the pattern, hits along a path are said to belong to that path.
- **Junctions** are points at which two or more paths meet, as determined by the last two stages of the pathfinding. Any correctly reconstructed primary vertex with more than one charged daughter particle should be reconstructed as a junction, but junctions can also be formed by e.g. secondary interactions or the eventual separation of strongly colinear tracks. A blob of hits around each junction are said to belong to the junction rather than its constituent paths, as it is not clear to which path they belong.

These concepts are illustrated in Figure 4.8.

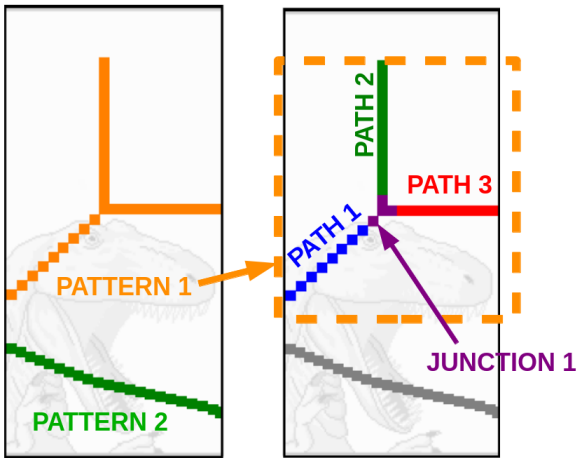


Figure 4.8: A cartoon illustrating the hierarchy of patterns, junctions and paths. In this example there are two patterns in the TPC: pattern 1 consists of three paths connected by a junction, while pattern 2 consists of a single isolated path.

4.5.3 δ-rays

δ-rays — secondary electrons produced by through-going particle tracks — have the potential to produce messy spiral tracks in the TPCs, creating a minefield for possible misreconstruction. In addition, even well-reconstructed δ-rays create junctions from through-going tracks, mimicking the appearance of primary vertices in the gas. For this reason, TREx attempts to recognise δ-rays at the reconstruction level.

TPC hits from δ-rays frequently have multiple peaks in their waveform. Since *delta*-rays are expected to spiral in the drift direction, when such hits are detected TREx chooses one at random and draws a cylinder around it with its base in the Y-Z plane, with initial radius r_0 equal to the width of two MicroMEGAS pads. The cylinder is then expanded by increments Δr, again equal to the width of two pads, checking at each step how many new multi-peak hits are included by the expansion. When further expansion yields no more multi-peak hits, all hits inside the cylinder are assumed to also be δ-ray hits, including those with single peaks. The process is repeated with this expanded hit selection, and all hits inside the final cylinder are flagged as originating from a δ-ray.

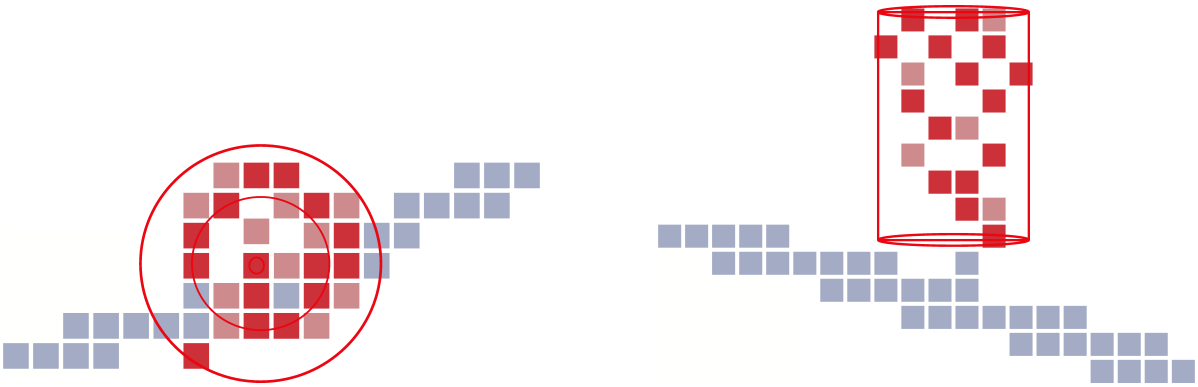


Figure 4.9: A diagram illustrating the δ-ray tagging procedure in the YZ (left) and XZ (right) views, provided by E. Larkin. Multi-peak hits are shown in deep red, while other hits tagged as belonging to the δ-ray are shaded in pink.

This flagging procedure prevents the largest and most distinctive δ-rays from causing problems for the analysis, although small or unusually clean δ-rays must still be handled by the final selection. The δ-ray tagging was developed using the ND280 Monte Carlo, which tracks δ electrons down to energies of 1 eV.

With all these features in place, TREx provides the tools necessary to recognise neutrino interactions in the TPC gas, as described in Section 5.2. Performance tests (shown in Section 6.3) indicate that its reconstruction of through-going tracks matches

tpcRecon, while its vertex-finding capability is also largely working well. The TREx group are continuing to make improvements to the reconstruction and even better performance is anticipated in future.

Chapter 5

Selection

Selecting neutrino interactions on argon gas from interactions on the other materials in the ND280 detector is the most challenging part of this analysis. The ND280 detector contains approximately 16 kg of argon gas, compared to more than 50 tonnes of other (mostly solid) materials in the active volume. To a first approximation, the entire detector sits in a uniform bath of neutrino flux, meaning that the interaction rate in any given volume of the detector scales as the target density. Combined with the weakly-interacting nature of neutrinos, this means that while the absolute number of interactions in the argon is low, an extremely high number of particles are continuously sleeting in from interactions in the surrounding subdetectors. In Monte Carlo simulation, only 0.02% of all simulated beam spills have an interaction in the argon gas.

With these challenges in mind, our selection proceeds in three stages:

- A preselection, to winnow out the most easily rejected backgrounds and reduce the large ND280 dataset to a manageable size.
- Reprocessing, which is run over the skimmed files produced by the preselection to provide the more detailed TPC reconstruction we require.
- The final selection, which uses the outputs of TREx to recognise genuine gas interactions, and produces a high-purity sample from the enriched sample produced by the preselection.

TREx has already been detailed in Section 4.5. In this chapter I will describe the preselection and the final selection. All the cuts used in the selection were tested and tuned with NEUT Monte Carlo. Where possible, each cut was tuned to maximise the product of the purity and the efficiency, in order to minimise the statistical error on the

final sample [59] (under the assumption that the surviving backgrounds are small and well-modelled).

5.1 Preselection

The goal of the preselection is to remove easily identifiable backgrounds, which make up the majority of the ND280 dataset. It searches for and rejects two main background categories with a simple cut-based procedure, while aiming to preserve signal events with minimal selection bias.

5.1.1 Preselection Background Topologies

First Background Category: Empty TPCs

The first and easiest background category to reject is the category of events with no particles passing through any TPC. An example of this event type is shown in Figure 5.1: it can be seen that the only hits registered in the TPCs come from electronic noise.

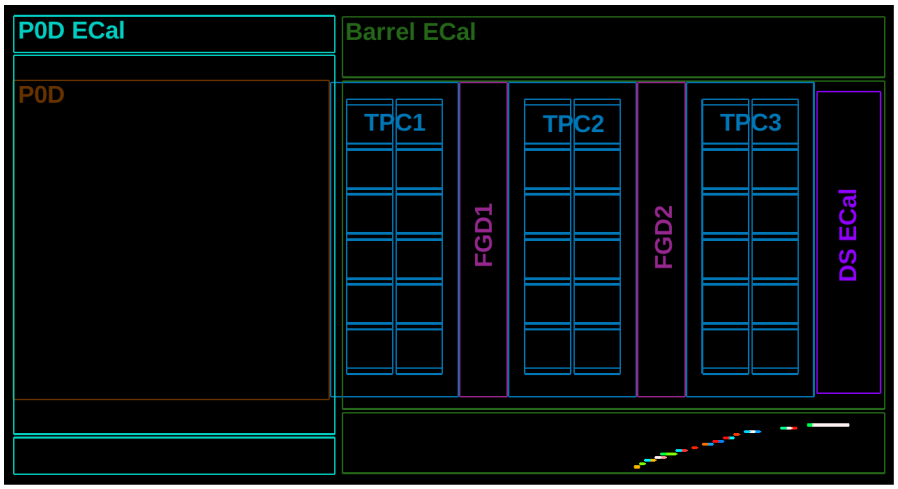


Figure 5.1: A simulated event with no activity in the TPCs. Hits in the bottom ECal show a contained neutrino interaction, with no interaction products reaching the TPCs.

This accounts for a large portion of "backgrounds", as approximately 80% of events in ND280 have no reconstructed tracks in the TPCs.

Second Background Category: Entering Particles

The second background topology the preselection aims to remove are events with no neutrino interaction in any TPC, but particles from neutrino interactions in other subdetectors passing through one or more TPCs (as illustrated in Figure 5.2). This is a very broad category of events, which makes these backgrounds more challenging to remove. In particular, harsh cuts to remove these events run the risk of causing large efficiency losses in the selection of signal events. Backwards-going particles from signal interactions can make signal interactions resemble entering backgrounds, and coincident interactions elsewhere in the detector can cause the veto to trigger even when there is also a signal interaction in one of the TPCs.

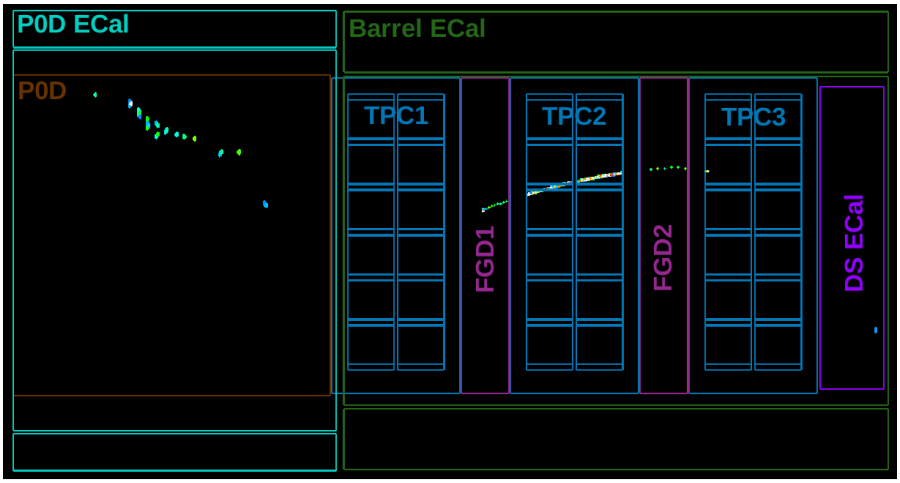


Figure 5.2: A simulated event in which a neutrino interaction in FGD1 produces a muon that passes through TPC2, leaving a clear trail of hits.

5.1.2 Preselection Cuts

I designed the preselection cuts with the aim of introducing a bare minimum of systematic uncertainties at this stage. Following this design philosophy, both cuts are purely hit-based, rather than using any information from the existing ND280 reconstruction algorithms. There are two cuts:

- **TPC Activation Criterion:** Reject event if no cluster of more than five hits is found inside the fiducial volume (defined in Table 5.1) of the relevant TPC, with gaps no larger than 240 ns in time and 50 mm in the Y-Z plane between each hit

and its nearest neighbour. This removes empty events of the type shown in Figure 5.1.

- **Upstream Veto:** Reject event if a cluster of hits above a certain size is found inside the "veto volume" of the subdetector directly upstream of the activated TPC. The cluster sizes, clustering requirements and veto volumes used are subdetector-dependent, and summarised in Table 5.2. This removes entering backgrounds of the type shown in Figure 5.2.

Module	Min Y (mm)	Max Y (mm)	Min Z (mm)	Max Z (mm)
TPC1	−930.00	1030.00	−724.85	−71.15
TPC2	−930.00	1030.00	634.15	1287.85
TPC3	−930.00	1030.00	1993.15	2646.85

Table 5.1: The fiducial volumes used in the preselection. Note that since the preselection is hit-based, there is no drift information, therefore limits on X (the drift axis) are not imposed. The fiducial volume limits were optimised to maximise the product of the preselection's purity and efficiency, stepping in from the TPC edges in increments equal to the dimensions of the MicroMEGAS pads (see Section 3.3.3).

Module	**TPC1**	**TPC2**	**TPC3**
Veto module	P0D	FGD1	FGD2
Max. hit separation (ns)	50	100	100
Min. hit charge (PEU)	7	0	0
Min. cluster size	4 hits	5 hits	5 hits
Veto volume extent (mm)	500.00	144.15	130.36

Table 5.2: The configuration of the upstream veto for each TPC. The veto in the P0D was taken unchanged from the existing reconstruction; the FGD vetoes were optimised to maximise the product of the preselection's purity and efficiency.

I also attempted to develop a veto on particles entering from the barrel ECals, but this did not prove practical due to the ECals' insensitivity to particle direction. As such, entering backgrounds that do not pass through a subdetector upstream of the TPC in question are allowed to survive the preselection.

These cuts are evaluated in turn for each of the three TPCs. If any TPC passes both conditions, the preselection preserves the event, with a flag to denote in which TPCs the preselection was passed and in which it was failed. If all three TPCs fail, the event is removed.

5.1.3 Preselection Performance

Tests on Monte Carlo indicate that the preselection rejects approximately 95% of all backgrounds, while preserving more than 99% of all signal interactions. Cross-checks on data are not possible for signal efficiency, due to the extreme rarity of signal interactions. Cross-checks on the rate of background rejection can be made by simply examining the total proportion of events in data rejected, as the proportion which are signal is to a very good approximation negligible.

These cross-checks give results consistent with the Monte Carlo figure for background rejection, as shown in Table 5.3.

Module	**TPC1**	**TPC2**	**TPC3**
Signal efficiency (MC)	99.47% ±0.05%	99.42% ±0.05%	99.73% ±0.04%
Background rejection (MC)	96.89% ±0.01%	95.36% ±0.01%	94.85% ±0.01%
Background rejection (data)	97.77% ±0.03%	94.45% ±0.05%	94.30% ±0.05%

Table 5.3: Table summarising the preselection performance on NEUT Monte Carlo (MC) and data. Efficiency is assessed on a specially generated sample of interactions occurring only in the gas; as such losses from coincidental interactions in other detectors cannot be accounted for.

The decrease in background rejection in the more downstream TPCs is expected, as these TPCs sample an increasing volume of forward-going particles from interactions in the surrounding barrel ECals, for which there is no veto. Monte Carlo tests also indicate that the preselection introduces little to no change in the energy spectrum of the selected neutrino interactions, as illustrated in Figure 5.3.

After passing through the preselection, the Monte Carlo sample achieves a purity of 0.1%, up from an initial purity before cuts of 0.04%. Increasing this purity further requires detailed analysis of the information from the activated TPC. This is achieved by the final selection.

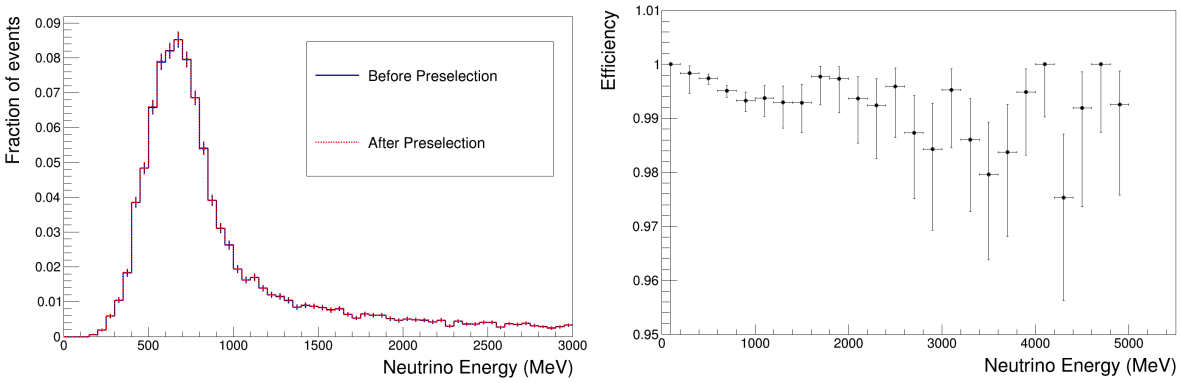


Figure 5.3: Left: histograms of neutrino energy before and after the preselection, area-normalised to show the minimal change in shape. Right: a plot of the efficiency of the preselection as a function of neutrino energy. Both these plots were generated using NEUT Monte Carlo.

5.2 Final Selection

After the skimmed files provided by the preselection have been reprocessed with TREx, the final step in the selection is to use the information TREx provides to perform a series of cuts on the patterns of hits seen by the TPCs. These cuts are ordered according to the following strategy:

- Firstly, to remove events with obvious reasons for disqualification (e.g. bad data quality, tracks outside the beam bunch time window) - "preliminary checks".
- Secondly, to identify candidate charged-current neutrino interactions in the fiducial volumes of the TPCs - "candidate identification"
- Thirdly, to reject false positives from difficult background topologies - "final background rejection"

The specific cuts used are, in order

1. **Preliminary Checks**
   a) Data quality
   b) Preselection recheck
   c) Bunch timing
2. **Candidate Identification**
   a) Muon identification
   b) Fiducial volume
3. **Final Background Rejection**

a) δ ray veto
b) Reconstruction failure veto

These are explained in Section 5.2.1.

5.2.1 Cuts

The cut sequence is evaluated in turn for each of the three TPCs, in effect forming three separate selections. The cuts used are as follows.

Data Quality Cut

The ND280 data acquisition system saves flags to record when there are any problems with data-taking. Before proceeding to any more advanced cuts, the selection checks the flag for the TPC in question to ensure that the data is valid.

Preselection Recheck Cut

The preselection (detailed in Section 5.1) saves all surviving events to the same file, regardless of which TPC passed the cuts. In order not to allow preselection-failing backgrounds in the other TPCs to survive, a flag is saved for each TPC detailing whether it would have passed or failed the preselection alone. If not, the event is rejected.

Bunch Timing Cut

The selection examines all patterns present in the TPC, and checks the associated values of T0 from the surrounding subdetectors. If the time measured does not fall within 60 ns of the central time of a beam bunch, the pattern is rejected as being either a secondary interaction or a cosmic background. If no patterns survive this cut, the event is rejected.

Muon Identification Cut

To identify a charged current interaction from a muon neutrino, the selection must first identify a muon. Each surviving pattern is examined, and its negative paths identified. Since the path charge is determined by curvature, the path direction must be decided

before the charge can be fixed. Single paths or paths with junctions at both ends are assumed to be pointing downstream, while paths attached to only a single junction are assumed to begin at the junction.

The negative paths are then iterated through from highest to lowest momentum, and tested to see if they fulfil the conditions of the muon PID. This is calculated by means of a particle "pull", as detailed in Section 4.4. The gas selection looks for muons by an inclusive muon PID that makes no attempt to distinguish muons from pions and other minimum-ionising particles, calculated as follows.

For each hypothesised particle type α, a probability $P(\alpha)$ is defined

$$P(\alpha) = e^{\frac{-\delta_\alpha^2}{2}}, \tag{5.1}$$

where δ_α is the pull for hypothesis α as given in Equation 4.1. From these probabilities, a likelihood is then defined

$$L(\alpha) = P(\alpha) / \sum_\beta P(\beta). \tag{5.2}$$

These likelihoods can then be used to define the cut variables

$$M_1 = \frac{L(\mu) + L(\pi)}{1 - L(p)} \tag{5.3}$$

$$M_2 = \frac{L(\mu) + L(\pi)}{1 - L(e)}. \tag{5.4}$$

M_1 is the ratio of the probability that the observed track comes from a muon or a pion to the probability that the observed track does not come from a proton. M_2 is the ratio of the probability that the observed track comes from a muon or a pion to the probability that the observed track does not come from an electron. Individually, each of these variables gives some separation between muons and non-muons, as shown in Figure 5.4.

Combining M_1 and M_2 gives the best separation, as M_2 breaks the degeneracy between muons and non-muons close to $M_1 = 1$. Figure 5.5 shows the ratio of non-muons to muons in the M_1-M_2 space; this distribution is used to justify a rectangular cut in which any track with

- $L(\mu) > 0.05$, and

- **either** defined values of both M_1 and M_2 with $M_1 < 0.5$ or $M_2 > 0.05$
- **or** momentum > 500 MeV/c

is considered a muon.

Paths which pass this test are treated as candidate muons. If no muon candidate is found, the event is rejected. In the case of multiple candidate muons, all candidates are retained for examination, prioritised from highest to lowest momentum.

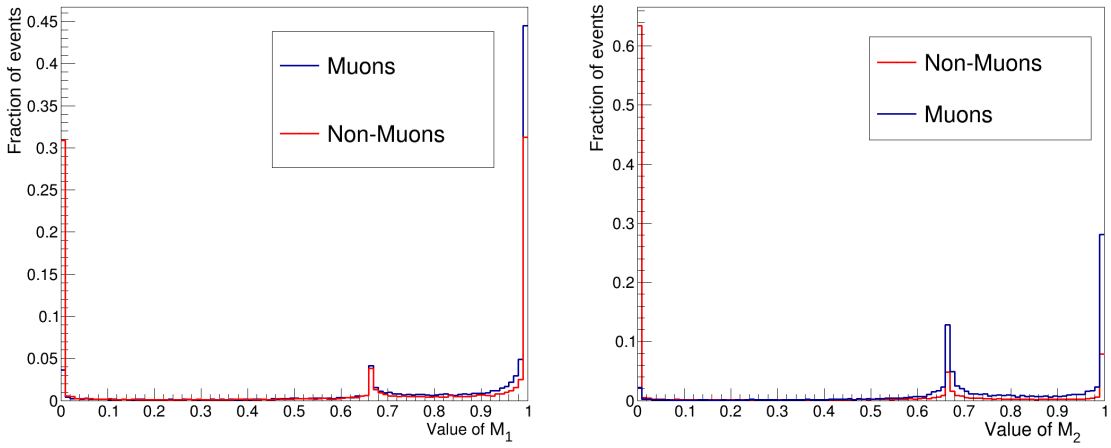


Figure 5.4: Plots of the distributions of M_1 (left) and M_2 (right) for particle tracks in T2K NEUT beam Monte Carlo. The distributions for muons and non-muons are unit-normalised.

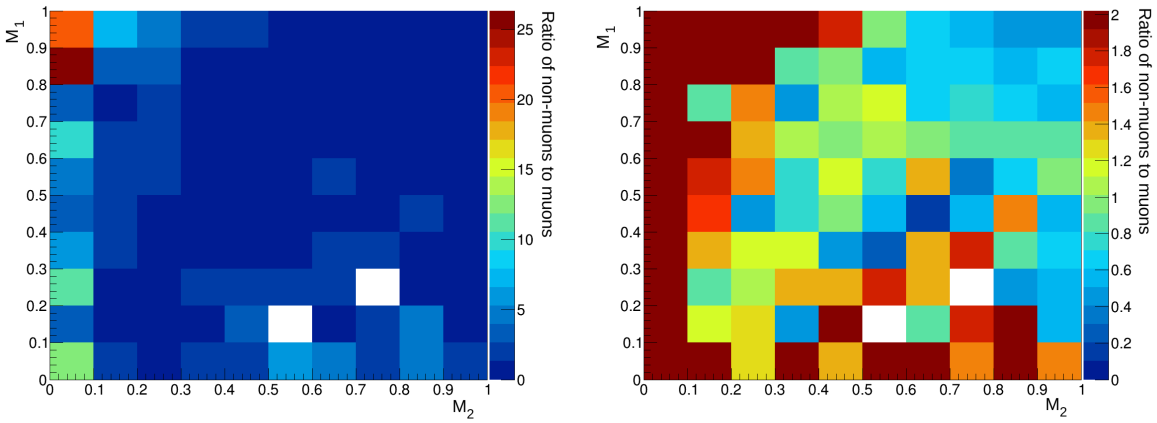


Figure 5.5: Plots of the ratio of non-muons to muons as a function of M_1 and M_2 for particle tracks in T2K NEUT beam Monte Carlo. The plot on the right shows the full Z-scale; the plot on the left is truncated at a ratio of 2, so that all bins where non-muons outnumber muons by a factor of 2 appear in red.

Fiducial Volume Cut

Muon candidates which start and end close to the TPC walls are almost certainly through-going muons from interactions in the surrounding subdetectors. The selection therefore checks all candidate muon paths — again starting with the candidate path with the highest momentum — to see if either end of the path is inside the fiducial volume of the TPC. The same fiducial volume is used as for the preselection, with the addition of constraints on the position in the X-coordinate, as drift information is now available. This allows neutrino interactions on the copper of the central cathode to be excluded.

Module	Min |X| (mm)	Max |X| (mm)	Min Y (mm)	Max Y (mm)	Min Z (mm)	Max Z (mm)
TPC1	20.00	870.00	−930.00	1030.00	−724.85	−71.15
TPC2	20.00	870.00	−930.00	1030.00	634.15	1287.85
TPC3	20.00	870.00	−930.00	1030.00	1933.55	2646.85

Table 5.4: The fiducial volumes used in the final selection. A 20 mm buffer zone is applied around the central cathode in order to exclude neutrino interactions on the cathode itself.

If no muon candidate is found with an end inside the fiducial volume, then the event is rejected. Otherwise, the selection continues by considering the origin of the highest momentum muon candidate that passes this cut as the candidate gas vertex.

δ-ray Veto

Once a suitable muon candidate has been identified by the muon identification cut, and the fiducial volume cut has established that it appears to start in the gas, the remaining cuts must check for false positives. δ-rays are the leading cause of false vertices in the gas, and can present a variety of reconstructed topologies (see Figure 5.6). Therefore, this cut uses three tests to check for δ-rays. If any test is failed, the muon vertex is considered to be a δ-ray, and the event is rejected.

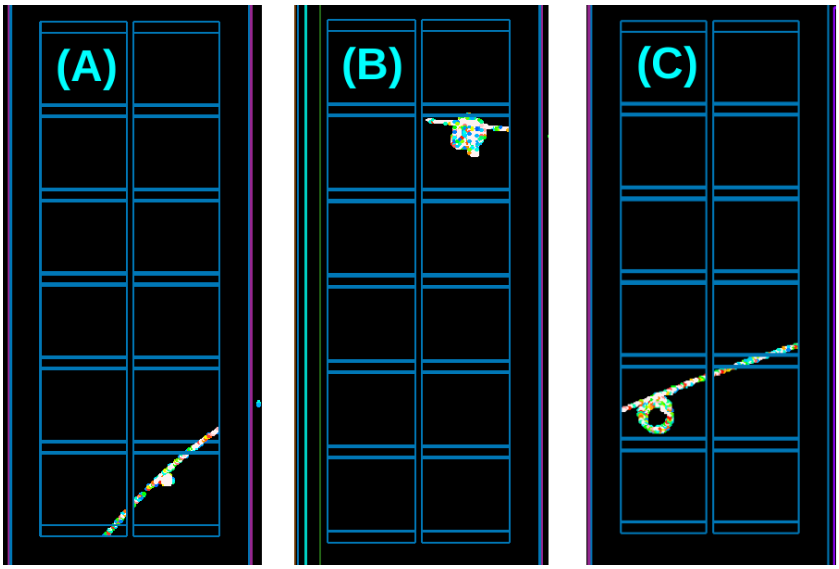


Figure 5.6: A selection of event displays showing contrasting examples of the reconstructed topologies created by δ-rays in simulation. (A) Small δ-ray "stub" in TPC2. (B) Large messy δ-ray in TPC1. (C) Large δ-ray with a clean spiral track in TPC3. All three are shown with the beam entering from the left.

The tests used are:

1. **PID test:** a correctly reconstructed δ-ray will contain two muon paths and a single electron path. The PID test looks for this topology by rejecting vertices with
   - At least one path with an electron PID, a δ-ray tag (see Section 4.5.3) or a failed PID.
   - At least one path with a muon PID.
   - No paths with other PIDs.
2. **Vertex extent test:** δ-rays, having much lower energies than the primary daughter particles from beam interactions, tend to spiral in the drift direction (the X axis in ND280's global coordinates). The vertex extent test rejects vertices that come from junctions whose constituent hits have an X extent of more than 42.0 mm. This value was optimised based on the $P \times E$ distributions shown in Figure **??**.
3. **Vertex momentum test** δ-rays are typically low-momentum. The vertex momentum test rejects vertices where the sum of the momenta of all paths emerging from the vertex is less than 100 MeV. This value was chosen to have a low impact on the energy distribution of the selected signal.

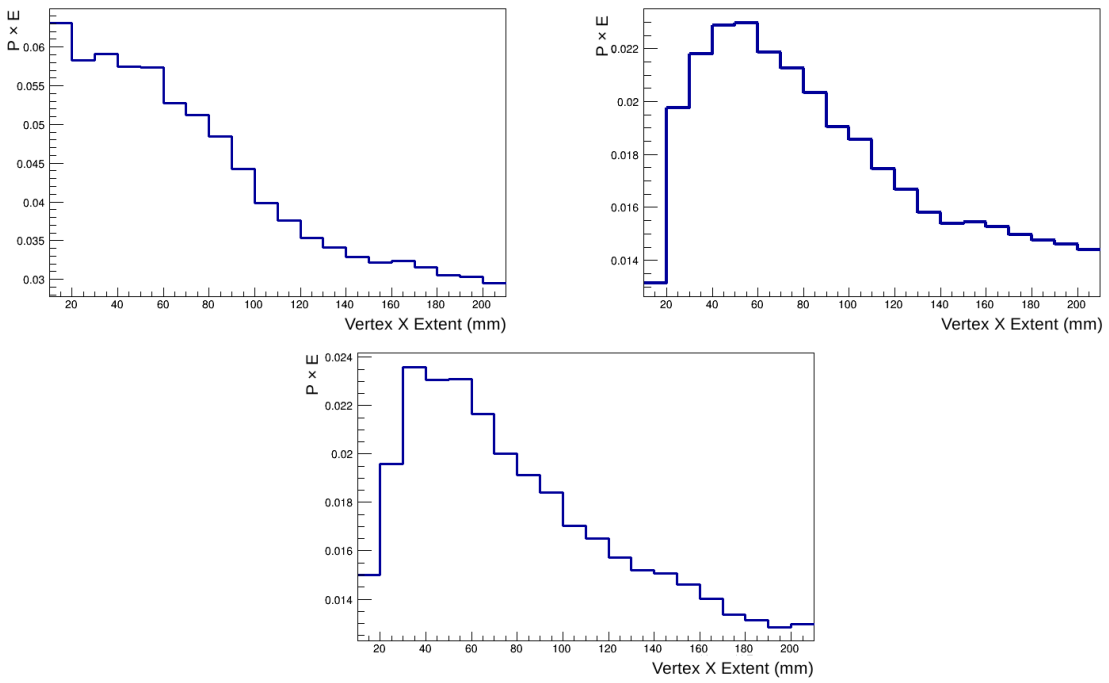


Figure 5.7: Plots showing the distribution of $P \times E$ as a function of the vertex extent cut, in TPC1 (top left), TPC2 (top right) and TPC3 (bottom).

Reconstruction Failure Veto

The other leading cause of false vertices comes from failures in the reconstruction. Like δ-rays, these come in a variety of forms, some examples of which are shown in Figure 5.8.

Like the δ-ray cut, three tests are used to identify misreconstructed events. If any test is failed, the event is rejected.

1. **Single-track veto:** if the muon candidate is not associated with a junction, it is rejected. While this causes significant signal loss, it has by far the strongest effect of any cut in reducing the number of entering backgrounds (see final bin in Figure 5.11).
2. **PID test:** if the muon candidate is associated with a junction whose only associated paths all have muon PIDs, it is assumed to be a broken track (as in Figure 5.8 (A)) and rejected.
3. **Colinearity test:** if the muon candidate is associated with a junction with two paths that are within 32° of being colinear, it is assumed to be two colinear entering backgrounds (as in Figure 5.8 (C)) and rejected. The value of 32° was optimised to

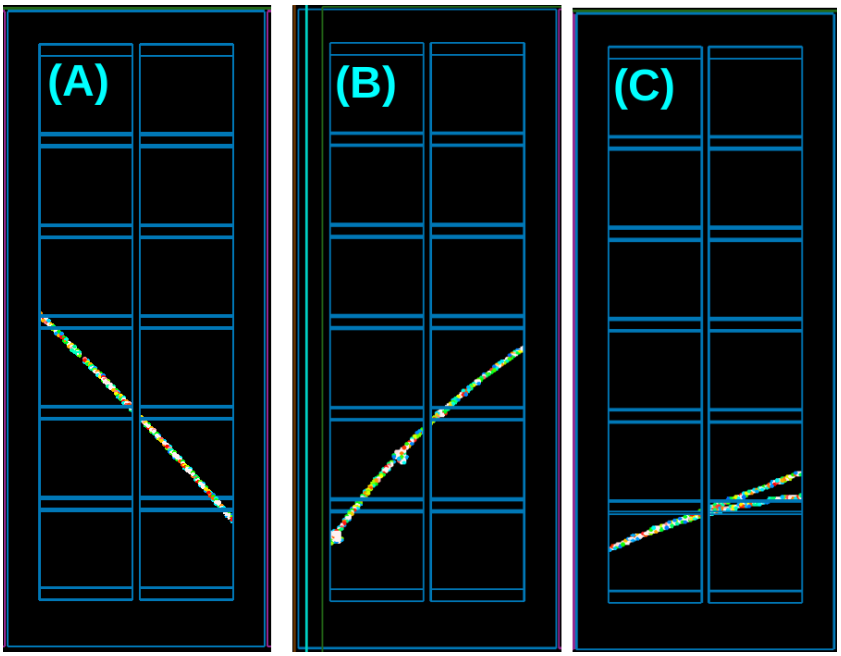


Figure 5.8: A selection of event displays showing some of the most problematic misreconstruction backgrounds in simulation. (A) A through-going track split into two paths where it crosses a gap between MicroMEGAS pads. (B) A through-going track with a false junction from a stubby δ-ray too small to be caught by the δ-ray cut. (C) Two strongly colinear entering tracks that form a false junction when they separate.

maximise the product of the selection's purity and efficiency; the distributions used for this tuning are shown in Figures 5.9 and 5.10.

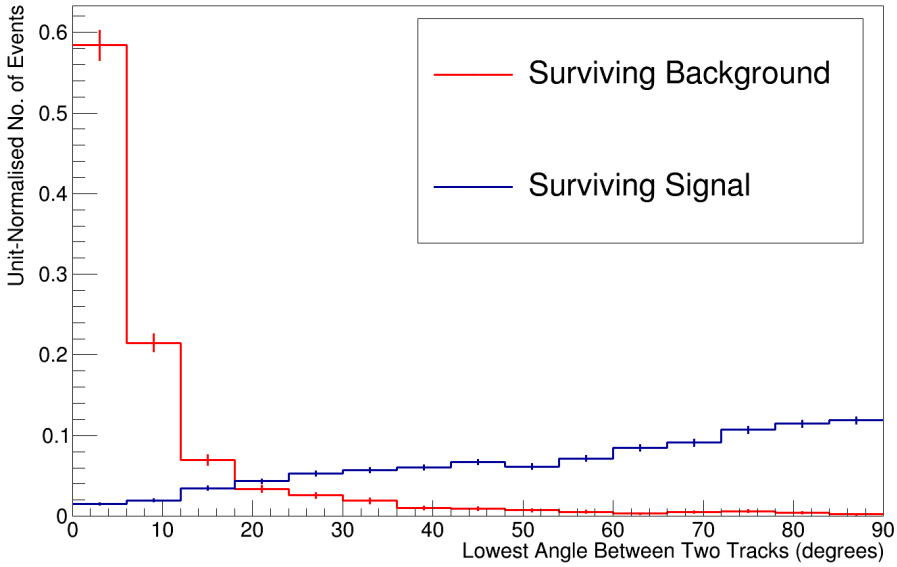


Figure 5.9: Plot showing the lowest angle distributions of signal and background in Monte Carlo surviving the δ-ray cut (in all three TPCs).

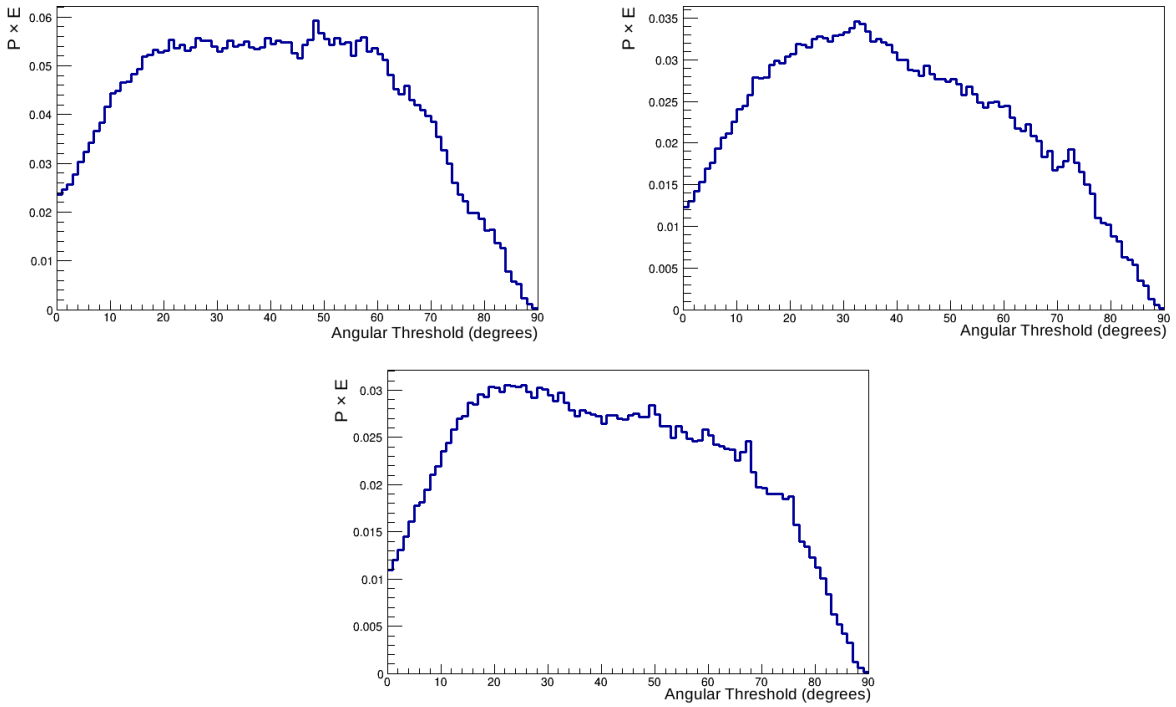


Figure 5.10: Plots showing the distribution of $P \times E$ as a function of the angular cut value for TPC1 (top left), TPC2 (top right) and TPC3 (bottom).

5.2.2 Performance

The limitations imposed by the extremely high starting background-to-signal ratio in this selection necessitated the imposition of harsh cuts. Certain cuts in the later stages of the selection have the consequence of selectively suppressing specific background topologies. All of these selective losses occur in the reconstruction failure veto; they are summarised briefly below.

- The single-track veto strongly suppresses signal interactions with zero protons, as shown in Figure 5.17. In combination with the low predicted rate of zero-proton events, this effectively restricts the analysis to looking at the last five bins of the proton multiplicity distribution.
- The PID test also contributes to the suppression of events with zero protons, as a genuine neutrino interaction producing a muon and a charged pion can easily pick up a muon PID for both tracks.
- The colinearity test by definition removes events with back-to-back pairs of particles, such as the "hammer" events seen by ArgoNeuT [32], which are of particular interest in the study of initial state short-range nucleon-nucleon correlations.

Many of these topologies are eminently recoverable with improvements addressed in Section 8.3. On the timescale of this analysis, however, these cuts are necessary to achieve a non-negligible purity in the final sample. The impact they have on the purity is clear in the last bin of the survival plots in Figure 5.11. The components of this last cut are broken down further in 5.12, while the evolution of the purity, efficiency, and the selection's figure of merit (the product of the two) are shown in Figure 5.13.

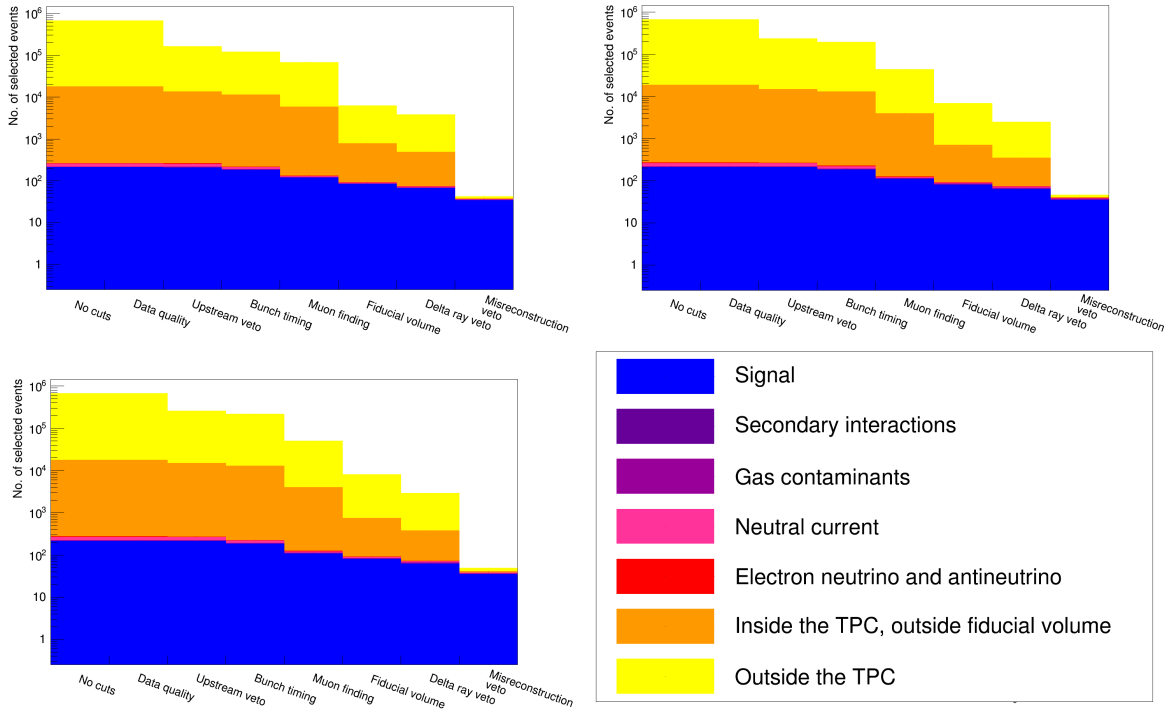


Figure 5.11: Plots showing the number of surviving signal and background events from preselection-skimmed Monte Carlo after each cut, in TPCs 1 (top left), 2 (top right) and 3 (bottom left). Although cuts 1-6 reduce the number of events from outside the fiducial volume by two orders of magnitude, without the final cut these events still dominate the selected sample. These plots are also given in tabular form in Table 5.6. The contribution of each of the three tests involved in the final cut is shown in Figure 5.12, in the order that they are applied.

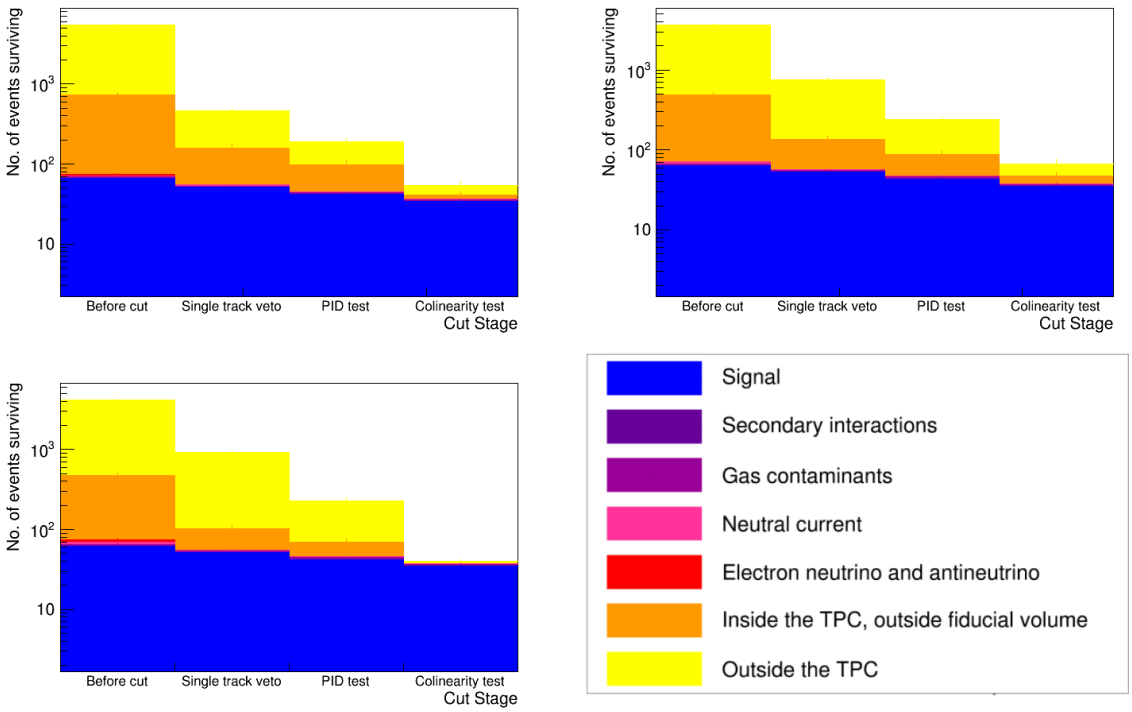


Figure 5.12: Plots showing the number of surviving signal and background events from preselection-skimmed Monte Carlo at each stage of the misreconstruction veto, in TPCs 1 (top left), 2 (top right) and 3 (bottom left).

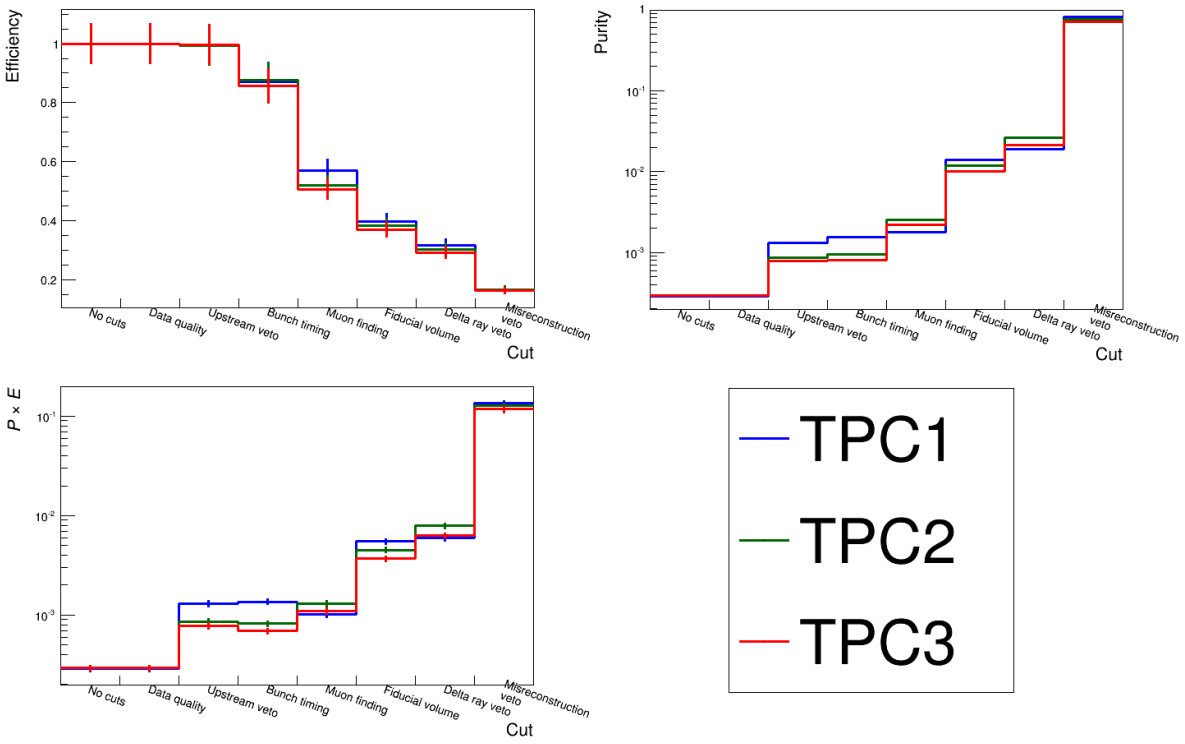


Figure 5.13: Plots showing the selection's efficiency (top left), purity (top right) and the product of the two (bottom left) after each cut. The decrease in $P \times E$ in TPC1 after the muon-finding cut is attributable to TPC1 seeing fewer non-leptonic entering backgrounds than TPCs 2 and 3 (which sample progressively more particles from interactions in the barrel ECals). The decrease in $P \times E$ in TPC3 after the bunch timing cut is attributable to TPC3 observing the highest proportion of these entering backgrounds (which occur inside the bunch), lessening the contribution the cut makes to the purity while keeping the efficiency loss constant.

The final selection is extremely pure, as summarised in Table 5.5, with a predicted selected sample of 105 signal events and 33 backgrounds across all three TPCs.

Module	Purity	Efficiency	Predicted no. of selected signal events
TPC1	$82.0 \pm 0.6\%$	$16.4 \pm 0.2\%$	33.9 ± 0.6
TPC2	$76.0 \pm 0.6\%$	$16.7 \pm 0.3\%$	35.0 ± 0.6
TPC3	$71.7 \pm 0.7\%$	$16.4 \pm 0.3\%$	34.6 ± 0.6

Table 5.5: Table summarising selection performance on NEUT Monte Carlo, assessed in the same way as for the preselection in Table 5.3 with uncertainties from Monte Carlo statistics. The decrease in purity for the downstream TPCs is attributed to the same difference in backgrounds entering from the ECals as described for Table 5.3.

The efficiency is also approximately uniform in neutrino energy (as shown in Figure 5.14), with a slight preference for events with low four-momentum transfer (as shown in Figure 5.15). Due largely to the directional assumptions made in identifying track charge (described in Section 5.2.1) the selection is insensitive to events with backward-going muons, as shown in Figure 5.16. Finally, for the reasons discussed above, the selection efficiency varies strongly as a function of the proton multiplicity. This is shown in Figure 5.17 and is addressed by use of the Bayesian unfolding technique described in Section 7.1.

Cut	No. of selected events			
	TPC1	TPC2	TPC3	Total
No cuts	$713,193 \pm 188$	$712,992 \pm 188$	$713,132 \pm 188$	$2,139,317 \pm 326$
Upstream veto	$156,039 \pm 92$	$240,518 \pm 112$	$267,809 \pm 118$	$664,366 \pm 187$
Bunch timing	$114,812 \pm 80$	$195,430 \pm 102$	$223,598 \pm 108$	$533,840 \pm 167$
Muon finding	$66,013 \pm 60$	$43,001 \pm 48$	$48,946 \pm 52$	$157,960 \pm 93$
Fiducial volume	$5,890 \pm 18$	$6,749 \pm 19$	$7,820 \pm 21$	$20,459 \pm 34$
Delta ray veto	$3,474 \pm 14$	$2,423 \pm 11$	$2,857 \pm 12$	$8,754 \pm 21$
Misrecon. veto	41 ± 1	46 ± 1	48 ± 1	135 ± 2
Cut	No. of selected signal events			
	TPC1	TPC2	TPC3	Total
No cuts	207 ± 1	209 ± 1	211 ± 1	627 ± 2
Upstream veto	205 ± 1	208 ± 1	210 ± 1	623 ± 2
Bunch timing	180 ± 1	184 ± 1	180 ± 1	544 ± 2
Muon finding	118 ± 1	109 ± 1	107 ± 1	334 ± 2
Fiducial volume	82 ± 1	80 ± 1	78 ± 1	240 ± 2
Delta ray veto	65 ± 1	64 ± 1	62 ± 1	191 ± 2
Misrecon. veto	34 ± 1	35 ± 1	35 ± 1	104 ± 2

Table 5.6: Table containing the cut-by-cut predictions for the surviving number of events shown in Figure 5.11, both for all events (top) and signal events only (bottom). The uncertainties quoted are purely statistical. Figures for the data quality cut are not shown, as on Monte Carlo this cut has no effect; the scaled statistics used correspond the amount of good-quality data available at ND280 (5.9×10^{20} POT).

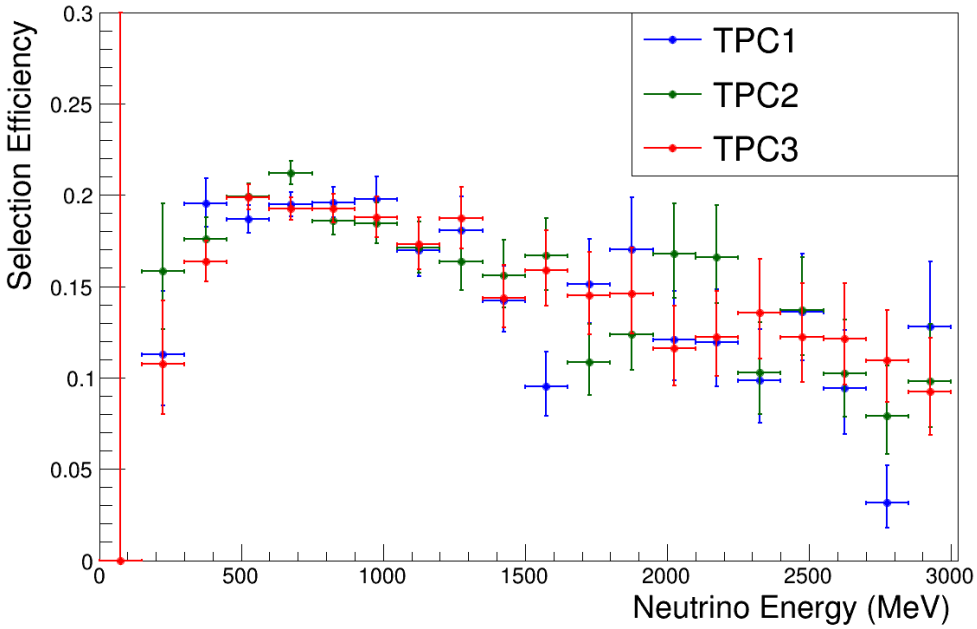


Figure 5.14: Plot showing the efficiency of the selection plotted as a function of neutrino energy, in each of the three TPCs.

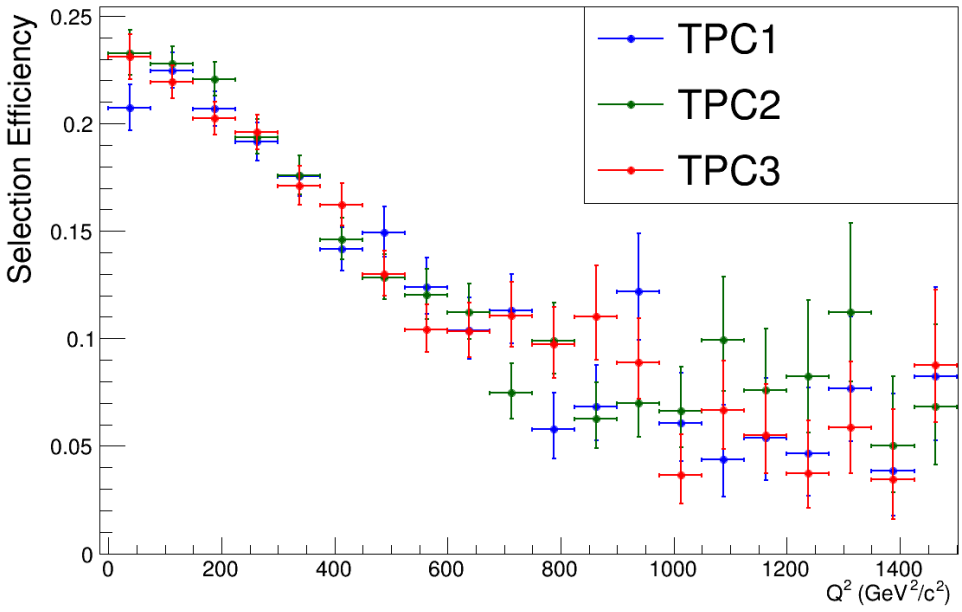


Figure 5.15: Plot showing the efficiency of the selection plotted as a function of Q^2, in each of the three TPCs.

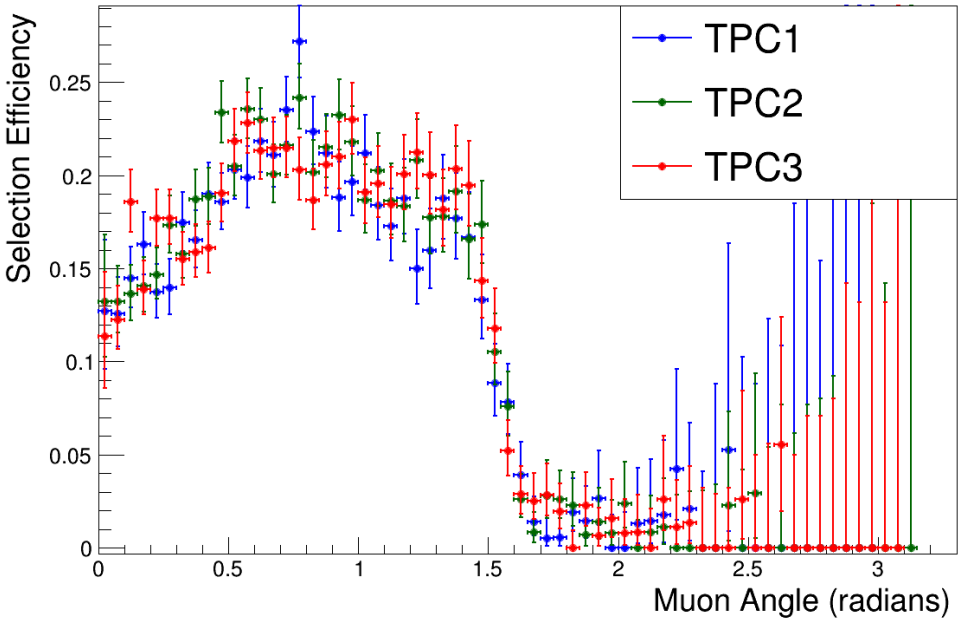


Figure 5.16: Plot showing the efficiency of the selection plotted as a function of the angle between the outgoing muon and the beam axis, in each of the three TPCs.

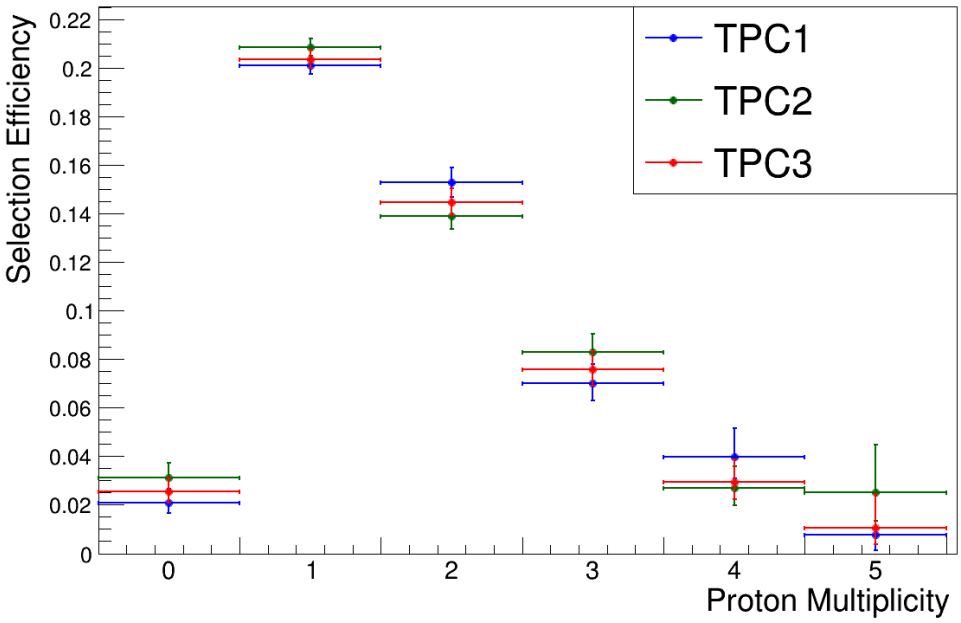


Figure 5.17: Plot showing the efficiency of the selection plotted as a function of number of protons emerging from the interaction, in each of the three TPCs.

Chapter 6

Systematic Uncertainties

The systematic uncertainties in this analysis can be broken into four categories:

- Uncertainties from variations in the power and energy spectrum of the J-PARC neutrino beam.
- Uncertainties from the limits of our knowledge of neutrino cross-sections on the materials used in ND280.
- Uncertainties from our simulation of the behaviour of the TPCs themselves.
- Uncertainties from our simulation of particles entering from outside the TPC fiducial volume - both from rescattering and detector response in the surrounding subdetectors.

Each of these categories is detailed below. I would like to thank the T2K beam group for providing the information used in calculating the flux uncertainty, the Neutrino Interactions Working Group (NIWG) for the information used in calculating the cross-section uncertainties, and the ND280 ν_μ group for their measurements of the detector systematics, which were used as benchmarks for assessing the impact of the revised detector uncertainties on this analysis.

6.1 Flux systematics

Flux uncertainties come from a variety of sources:

- Rates of hadron production in the target.
- The profile and alignment of the proton beam.
- The alignment of the target.

- The alignment of the magnetic horns, and variations in the horn B-field.

Hadron production uncertainties are constrained with external data from NA61-SHINE, as mentioned in Section 4.2.1. All the remaining sources of uncertainty, being unique to the T2K beam apparatus, are assessed by varying the relevant quantities in simulation.

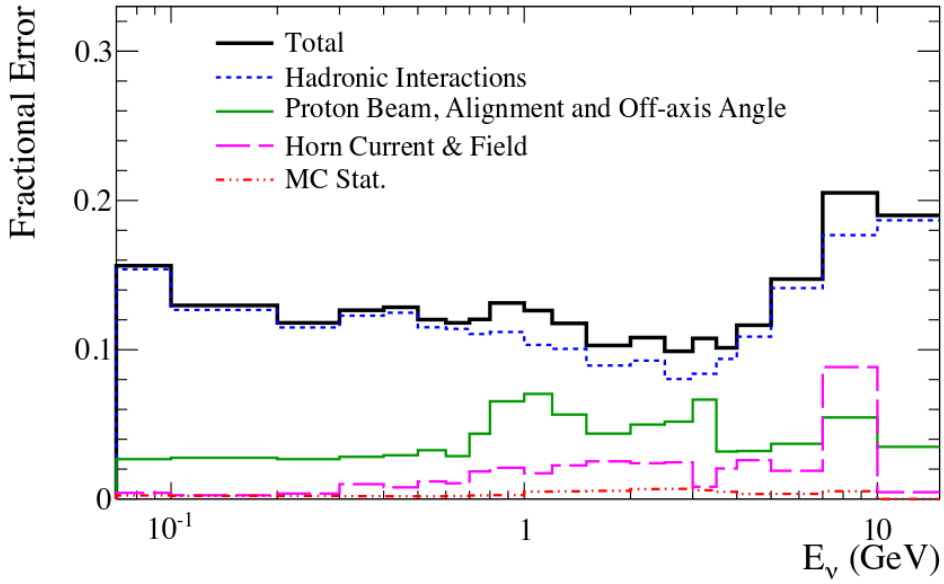


Figure 6.1: A plot of the error sources contributing to the ND280 ν_μ flux uncertainty as a function of energy. Figure provided by the T2K beam group.

Each of these error sources is assessed individually, and a covariance matrix is produced in bins of neutrino energy and flavour. These covariance matrices are then summed to produce the final flux covariance matrix.

6.2 Cross-section systematics

Cross-section uncertainties come from our modelling of multiple processes (as described in Chapter 2), and therefore break down into a number of different cross-section parameters. The T2K Neutrino Interactions Working Group (NIWG) provide constraints on these parameters using fits to external data. These constraints can be broken down into the following rough categories:

- **CCQE model:** M_A^{QE}, the effective axial mass seen by charged-current quasi-elastic interactions, is constrained by fits to MiniBooNE neutrino [2] and antineutrino [60] data, and MINERνA neutrino data [61].

- **Pion production model:** $M_A^{1\pi}$, $C_5^A(0)$ and the scalar background ($I = \frac{1}{2}$ Bkg), the parameters governing pion production, are constrained with a reanalysis of bubble chamber data from ANL and BNL [62].
- **FSI model:** The NEUT cascade model is characterised by six parameters, defining inelastic scattering and charge exchange at low and high energies as well as pion absorption and production. Constraints on the uncertainty for each of these parameters are derived from data from the DUET experiment [63], and several pion scattering experiments(e.g. [64]).

Little data is available on which to tune the other cross-section parameters. In the case of the normalisations for neutral current events, conservative uncertainties are set to be consistent with the limits observed by K2K [65], MiniBooNE [66] and SciBooNE [67]. In the case of the coherent pion production normalisation a 100% error is assigned to reflect the fact that external data is consistent with no charged-current coherent pion production in this energy region (although new data from MINERνA may give cause to revise this assessment [68]).

The uncertainties relevant to this analysis are listed in table 6.1.

Parameter	Nominal Value	Uncertainty
M_A^{QE} (GeV/c^2)	1.2	14.6%
$C_5^A(0)$	1.01	12%
$M_A^{1\pi}$ (GeV/c^2)	0.95	16%
$I = \frac{1}{2}$ Bkg	1.30	15%
CC coherent normalisation	100%	100%
NC coherent normalisation	100%	30%
NC other normalisation	100%	30%
FSI inelastic low-E	0	0.41 (absolute)
FSI inelastic high-E	0	0.34 (absolute)
FSI pion production	0	0.30 (absolute)
FSI pion absorption	0	0.41 (absolute)
FSI charge exchange low-E	0	0.57 (absolute)
FSI charge exchange high-E	0	0.28 (absolute)

Table 6.1: List of the cross-section uncertainties used for this analysis, as provided by the NIWG. These uncertainties are described in detail in Reference [4].

In addition, there are uncertainties on the total charged-current neutrino cross-section on the elements in the surrounding subdetectors that provide the majority of entering backgrounds. Taking carbon, lead and iron as the primary background sources, I took constraints on the CC-inclusive interaction rate on these elements from results published by T2K [69] [70] and MINERνA [71]. These uncertainties are summarised in Table 6.2.

Target Element	Proportional Uncertainty on Cross-Section	Experiment
C	12%	T2K [69]
Fe	19%	T2K [70]
Pb	20%	MINERνA [71]

Table 6.2: Uncertainties on the CC-inclusive ν_μ cross-section for the three most prevalent target elements in the surviving out-of-fiducial background.

The measurement of proton multiplicity is highly dependent on cross-section parameters, in particular those from FSI. As such I only apply these systematic variations to my predictions of the background; no cross-section systematics are assigned to the signal.

6.3 TPC Detector systematics

Since the selection focuses entirely on information from the TPCs, the main systematic uncertainties to consider arising from the detector are those concerning the TPCs. These uncertainties fall into two categories: those that affect the ability of the selection to detect the outgoing muon from a neutrino interaction, and those that affect the ability of the reconstruction software to correctly count the outgoing protons. The former cause events to be lost from the sample; the latter cause events to migrate between bins of proton multiplicity. A full list of the detector systematics and which category they fall into is given in Table 6.3.

The selection uses no reconstructed information from the other subdetectors, and as such their contributions to the detector systematics are minimal. The only external systematic considered is the hit-finding efficiency in the upstream subdetectors used for the preselection; this is assessed as part of the out-of-fiducial-volume systematic (see Section 6.4).

Systematic	Category
Track-finding efficiency	Muon finding
Muon identification	Muon finding
Vertex connection efficiency	Proton counting
Proton identification	Proton counting

Table 6.3: Table of detector systematics assessed in the TPCs.

The muon finding systematics have already been evaluated for official T2K analyses and found to be small; as such, my approach in assessing them has been to establish that with TREx, these uncertainties still have a similar size as with tpcRecon. They can then be safely neglected, as they will be negligible by comparison to the large anticipated statistical error.

6.3.1 Track-Finding Efficiency

The TPCs reconstruct tracks from particles crossing them with a certain efficiency. Variations in this efficiency impact the selection through the possibility that the muon track from a signal interaction is lost, in which case the signal event will be rejected by the cuts.

The established procedure for assessing this efficiency is to use a control sample of through-going "sand" muons, selected by requiring a single minimum-ionising track in a TPC with ends in each of the subdetectors upstream and downstream of the TPC in question. Using this sample, the efficiency is calculated as the fraction of events in which a track is seen in a TPC that has reference tracks in its neighbouring upstream and downstream detectors.

I calculated the track-finding efficiency for TREx on sand muon samples from both data and Monte Carlo, and compared the performance in both cases to tpcRecon. This comparison is shown in Figure 6.2.

As the figure shows, the track-finding efficiency in TREx is very similar to the efficiency in tpcRecon, as is the data-MC difference. Taking the largest difference between data and Monte Carlo in any bin as the systematic uncertainty, TREx and tpcRecon have the exact same uncertainty on the track-finding efficiency: 2.3%

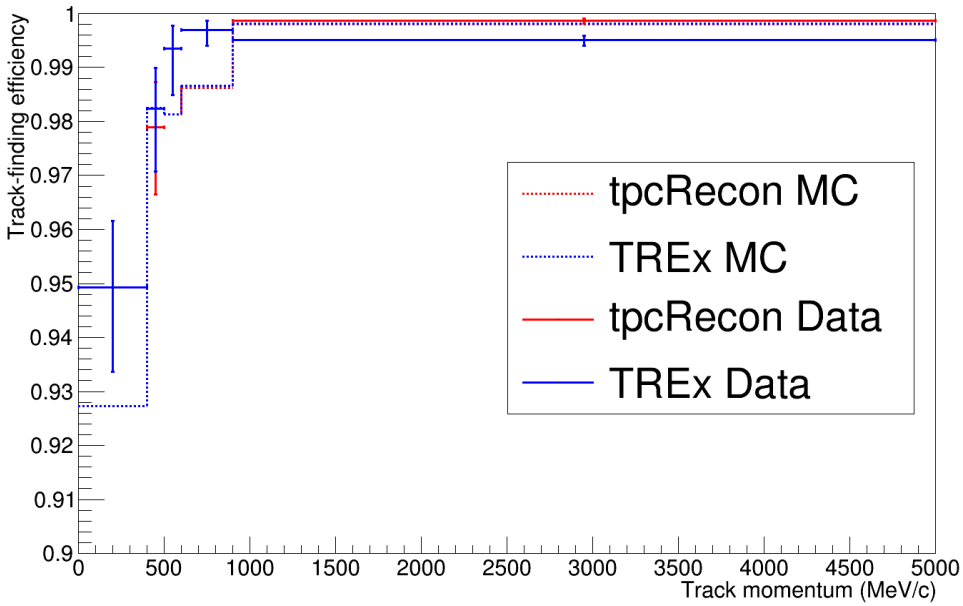


Figure 6.2: A plot of the track-finding efficiency in TPC2 (shown as an example - TPCs 1 and 3 show almost identical behaviour), evaluated for data and MC in both TREx and tpcRecon. In many bins the two are identical.

In the official CC-inclusive ν_μ selection, using tpcRecon, the propagated uncertainty from this source is 0.13%. Having observed that both the efficiency and the uncertainty on the efficiency are consistent between TREx and tpcRecon, there is no reason to believe this error would be significantly larger, and it can therefore be safely neglected.

6.3.2 Muon Identification Efficiency

The particle ID calculation described in Section 5.2.1 is capable of mis-identifying particle tracks. Variations in the success rate of the PID can impact the selection by misidentifying a muon from a signal interaction as a different particle, causing the signal event to be rejected by the cuts; they can also introduce new backgrounds by misidentifying a non-muon as a muon, erroneously tagging an event as a charged-current interaction. The primary causes of variations in the PID performance are differences between the values of the particle pulls returned by data and Monte Carlo.

To assess the impact of variation in the pulls on the PID, I used a control sample of sand muons passing the criteria used for the track-finding efficiency. Operating under the assumption that these tracks came from true muons, I calculated the PID efficiency

as the proportion of these tracks which received a positive muon PID. The results of this test are shown in Figure 6.3 for both TREx and tpcRecon, as differences in the construction of the track could conceivably lead to a difference in the pulls.

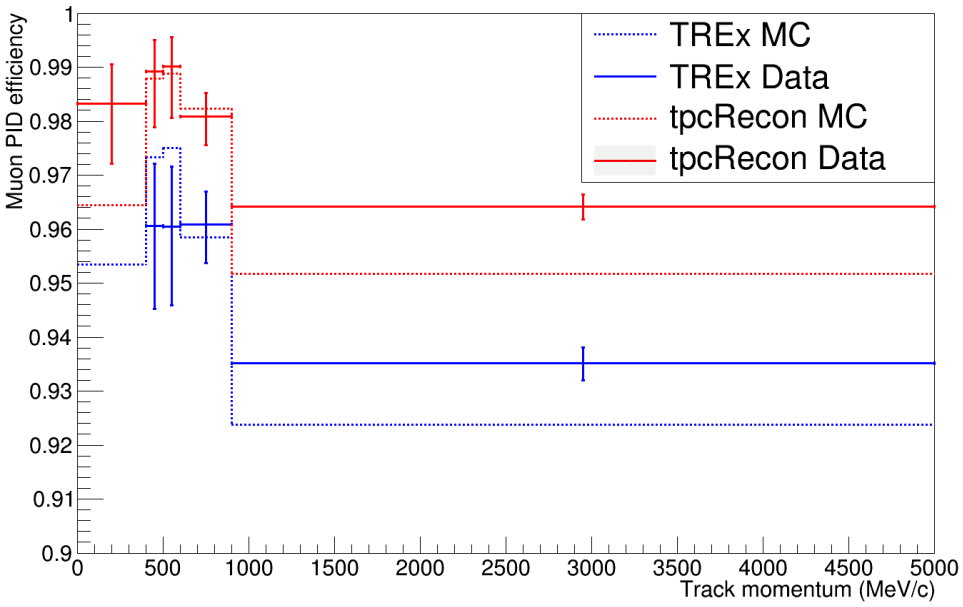


Figure 6.3: A plot of the muon PID efficiency in TPC2, evaluated for data and MC in both TREx and tpcRecon.

As the figure shows, the largest difference between data and Monte Carlo for TREx (1.1% in the high-energy bin) is no larger than in tpcRecon (1.2%, in the same bin). In the official CC-inclusive ν_μ selection, using tpcRecon, the propagated uncertainty from this source is 0.33%. For the same reason as the track-finding efficiency, I therefore neglect this uncertainty.

6.3.3 Vertex Connection Efficiency

Once a muon candidate has been selected, the number of protons emerging from the vertex must still be accurately counted. The known smearing in the multiplicity distribution is accounted for by the unfolding process (see Section 7.1), but this does not account for any differences in the performance of the reconstruction between data and Monte Carlo.

Assessing the size of this difference is difficult, as no suitable control sample of multi-proton vertices in the gas exists in data apart from the signal interactions themselves. I

first attempted to make a proxy measurement by measuring the efficiency with which TREx reconstructed junctions in a control sample of through-going muons creating delta rays in the TPC fiducial volume, with the aim of approximating the efficiency with which protons are connected to a neutrino vertex as the efficiency with which electron paths were connected to the through-going muon track.

Unfortunately, the reconstruction of delta rays proved to be uncharacteristic of the reconstruction of proton-like tracks. While protons were connected to junctions with a $\sim 97\%$ efficiency in Monte Carlo, the same efficiency for delta rays in Monte Carlo was only $\sim 50\%$. As such I abandoned the use of this control sample and carried out an evaluation of this systematic using Monte Carlo only.

Since the vertex-finding procedure is based only on hit positions (see Section 4.5.1), the main contribution to variations in performance comes from the appearance or disappearance of hits through varying levels of electronic noise in the detector. Using a Monte Carlo sample consisting only of neutrino interactions on the TPC gas, I created a reprocessed copy of the sample with double the amount of simulated noise hits. Comparing these two samples, I then defined the connection efficiency as the fraction of events with one true proton which had been reconstructed with one proton track. The results of this comparison are given in Table 6.4.

MC sample	Connection efficiency
Nominal	97.2 ± 0.3 %
Double Noise	96.8 ± 0.3 %

Table 6.4: Vertex connection efficiencies for Monte Carlo samples with a normal rate of noise hits (Nominal) and twice that rate (Double Noise).

This provides an extremely conservative estimate of the impact of noise on the vertex connection efficiency, as the observed fluctations in the gain are in reality less than 10% [5]. Nevertheless the resulting uncertainty on the efficiency is still very small, at $0.4 \pm 0.4\%$.

6.3.4 Proton Identification Efficiency

A proton that has been successfully connected to the vertex must still be identified as a proton before it can be counted. The proton PID is calculated in a similar manner to

the muon PID detailed in Section 6.3.2. A path is identified as a proton if it fulfils two conditions:

- The path does not pass the criteria to be identified as a muon.
- $L(p) > 0.5$ (see Equation 5.2).

I evaluated the uncertainty on this method by comparing data and Monte Carlo for a control sample of entering protons originating in the FGDs, defining the efficiency of the proton PID as the fraction of events in the control sample in which TREx reconstructed a path tagged as a proton. The results of this test for the two samples are shown in Figure 6.4.

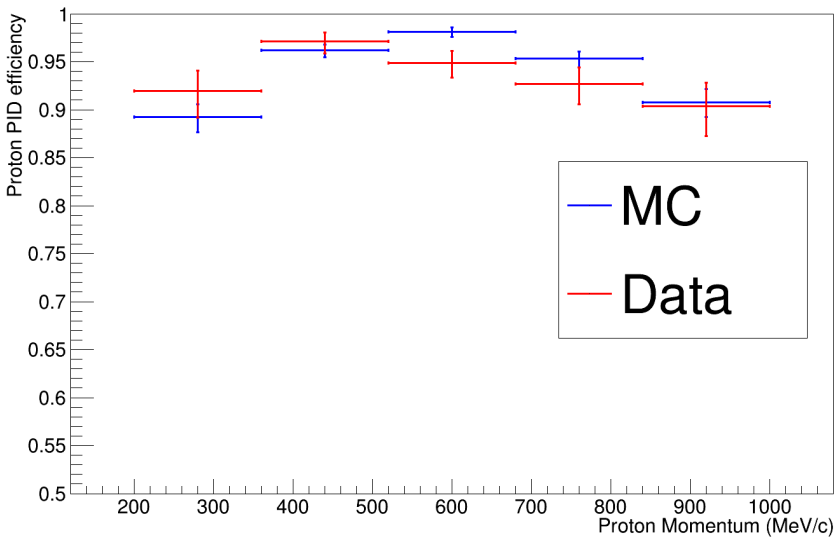


Figure 6.4: A plot of the proton PID efficiency in TPC2 as a function of the reconstructed proton momentum, evaluated for data and MC in TREx.

Taking the largest difference between data and Monte Carlo in order to give a conservative estimate, I obtain an uncertainty of $3.2 \pm 0.13\%$ on the proton PID efficiency — the largest of the detector systematics.

6.4 Out of fiducial volume systematics

Uncertainties on the number of entering background events surviving the selection arise from uncertainties on the mass and geometry of the surrounding detectors, and on the hit efficiency of the subdetectors used in the preselection to veto entering tracks. I assessed

these uncertainties simultaneously by running the selection on both data and Monte Carlo for two complementary control samples: sand muons (to represent forward-going backgrounds) and cosmic rays (to represent backgrounds at high angles to the beam axis).

The results of these tests are summarised in Table 6.5

Sample	% of backgrounds surviving in MC	% of backgrounds surviving in data
Sand (TPC1)	0.001%	$< 0.08\%$
Sand (TPC2)	$< 0.0002\%$	$< 0.08\%$
Sand (TPC3)	0.0002%	$< 0.08\%$
Cosmics (TPC1)	$< 0.0003\%$	$< 0.0003\%$
Cosmics (TPC2)	$< 0.0003\%$	$< 0.0003\%$
Cosmics (TPC3)	$< 0.0003\%$	$< 0.0003\%$

Table 6.5: A comparison of background survival from sand muons and cosmic muons in data and Monte Carlo. Limits are given where no backgrounds survived the selection; the relatively high limit on the sand muon survival rate in data is due to the smaller size of the available control sample.

The rate of survival in data is consistent with the limits observed for Monte Carlo for both the cosmic and sand control samples, with no data events surviving in any sample. As such this systematic is assumed to be small enough to neglect.

6.5 Propagation of Systematic Uncertainties

I propagate all systematic uncertainties to the final result using covariance matrices. For each category of uncertainty I produce a large number of reweighted Monte Carlo samples, in which the systematic parameters are randomly varied within their uncertainties, the events in the nominal Monte Carlo reweighted accordingly, and the unfolding procedure detailed in Section 7.1 carried out on the reweighted fake dataset. The fractional covariance matrix for each source of uncertainty can then be defined as

$$V_{ij} = \frac{1}{M} \sum_{\alpha=1}^{M} \frac{(N_i^{\alpha} - N_i^{\text{nominal}})(N_j^{\alpha} - N_j^{\text{nominal}})}{N_i^{\text{nominal}} N_j^{\text{nominal}}}, \tag{6.1}$$

where M is the number of throws, N_i^{α} is the number of events in true bin i of the unfolded distribution for throw α, and N_i^{nominal} is the number of events in true bin i of the unfolded distribution for the nominal Monte Carlo.

I performed 1,000 throws each for the flux, cross-section, FSI and path counting systematics. The covariance matrices produced by this method are shown in Figures 6.5, 6.6 and 6.7 respectively. The total covariance matrix from all sources of systematic error is simply the sum of these three; it is shown in Figure 6.8. All throws used the same Monte Carlo samples: 5.6×10^{22} POT of interactions in the gas generated in NEUT to simulate the signal (equivalent to 57,246 signal events before scaling), and 10.7×10^{21} POT of interactions throughout ND280 generated in GENIE to simulate the background. Both are scaled to the POT of the existing ND280 dataset (5.9×10^{20} POT).

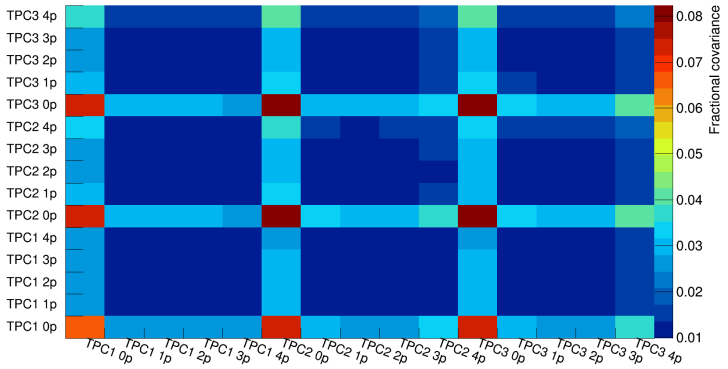


Figure 6.5: The flux covariance matrix, produced from 1,000 throws. The high amount of background subtraction in the zero-proton bins inflates the flux uncertainty in those bins.

The error from cross-section uncertainties is very small. Since the cross-section parameters are only varied for background events, the influence of this systematic is concentrated in the zero-proton bins, where the background contamination is highest. The uncertainties on the flux are the dominant error in the 1- and 2-proton bins, which are the bins with the highest predicted population of signal. In the higher-multiplicity bins the error is dominated by the uncertainty on the path counting, as for a higher number of protons there is a higher chance that one or more of them will be lost.

There is some unexpected variation between the errors on TPCs 1, 2 and 3, which I anticipated to be largely symmetric. This is most probably due to low-statistics effects in the Monte Carlo. Although a sample 3.7×10^{21} POT (approximately five times the current ND280 ν_μ data) was used to simulate background interactions in the rest of the

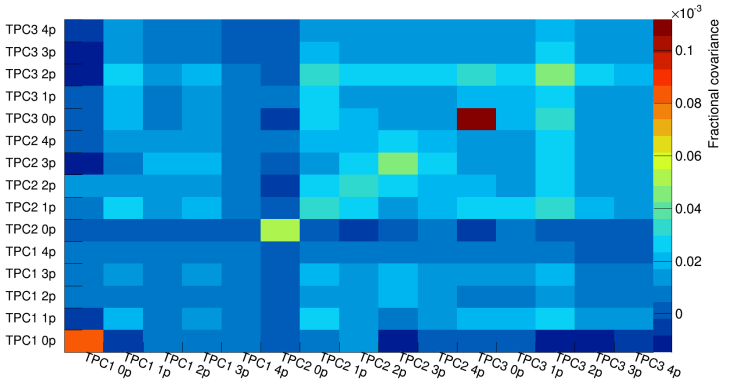


Figure 6.6: The cross-section covariance matrix, produced from 1,000 throws varying the FSI parameters and 1,000 throws varying the other cross-section parameters.

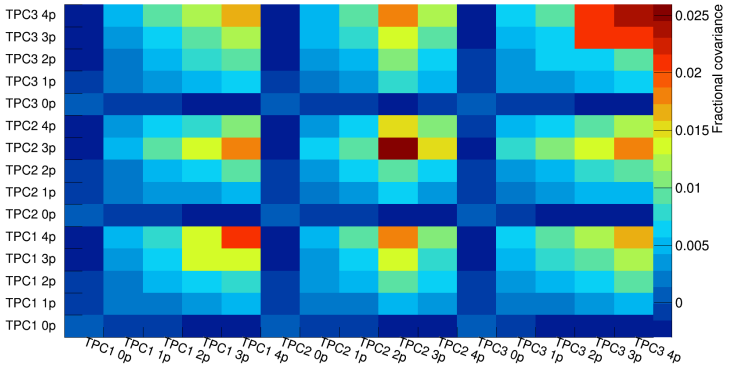


Figure 6.7: The proton counting covariance matrix, produced from 1,000 throws.

detector, the absolute number of backgrounds that survive the cuts is still low - a total of 200 in TPC1, 209 in TPC2 and 234 in TPC3. This leads to some multiplicity bins having a very low background occupancy — for example, only a single background event is selected in the 4+ protons bin.

For this analysis, the uncertainty arising from Monte Carlo statistics is quantified by the covariance matrix method, as detailed in Section 7.2. For future generations of this analysis, additional processing resources may need to be devoted to producing an extremely high-stats background sample.

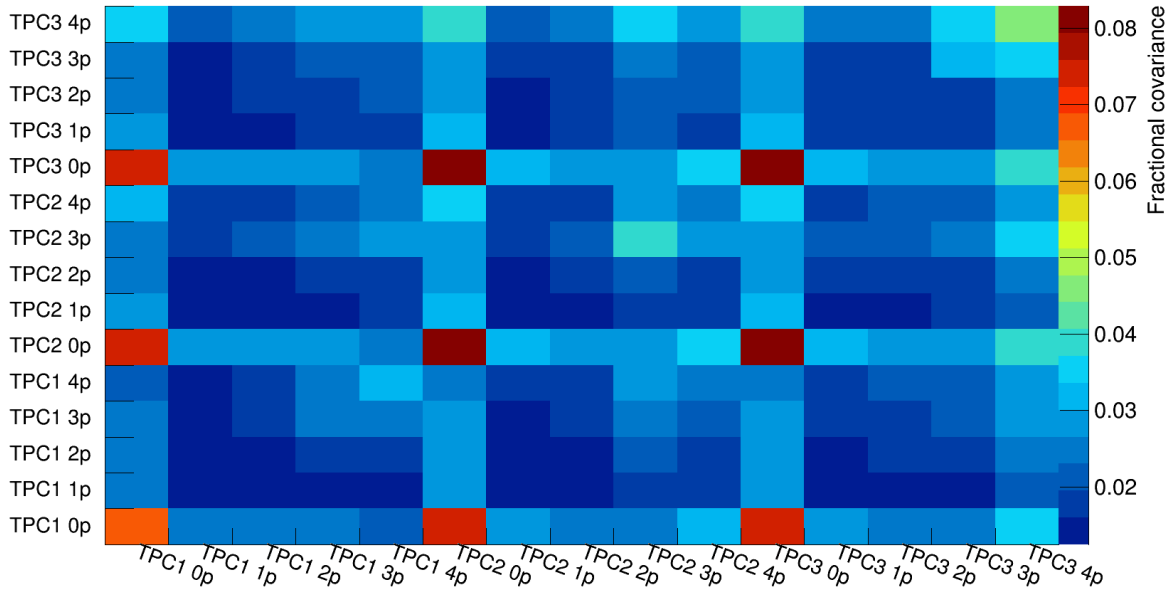


Figure 6.8: The combined covariance matrix from all sources of systematic uncertainty.

Chapter 7

Argon Cross-Section

The aim of this analysis is to produce a flux-averaged "differential" cross-section on argon as a function of the proton multiplicity of the interaction. While using a discrete variable like the multiplicity means that a true differential cross-section cannot strictly be defined, an effective cross-section can be defined as a stepped function, such that the number of events with n protons can be written

$$N_n = T\Phi\frac{\partial\langle\sigma\rangle_\Phi}{\partial n}, \tag{7.1}$$

where $\langle\sigma\rangle_\Phi$ is the effective flux-averaged differential cross-section, Φ the total flux, and T the total number of target nucleons.

In order to produce a measurement of $\frac{\partial\langle\sigma\rangle_\Phi}{\partial n}$, I use a Bayesian unfolding method to map between the reconstructed proton multiplicity and the true distribution. This gives the cross-section in the form

$$\frac{\partial\langle\sigma\rangle_\Phi}{\partial n} = \frac{\sum_j^{n_r} P(t_n|r_j)(N_j^{\text{measured}} - B_j)}{\epsilon_n T\Phi}, \tag{7.2}$$

where ϵ_n is the selection efficiency in true multiplicity bin n, N_j^{measured} is the number of selected events observed in reconstructed bin r_j, B_j is the number of background events predicted in reconstructed bin r_j, and $P(t_n|r_j)$ is the probability of an event having come from true multiplicity bin t_n given that it was found in the reconstructed bin r_j. This calculation is explained in detail in Section 7.1

7.1 Bayesian Unfolding

To make an accurate measurement of the cross-section, I employ a Bayesian unfolding technique based on D'Agostini's method of estimating the unfolding matrix [72], which uses Bayesian inference to "unfold" from the reconstructed distribution of an observable — in this case, the number of protons emerging from the nucleus — to the true distribution.

7.1.1 Unfolding Method

The unfolding technique uses Bayes' Theorem to define the probability that an event occurred in true bin i, given that it was observed in reconstructed bin j. Denoting the true bins as $t_1, t_2, ...t_{n_t}$, and similarly the reconstructed bins as $r_1, r_2, ...r_{n_r}$, this probability is defined

$$P_m(t_i|r_j) = \frac{P(r_j|t_i)P_m(t_i)}{\sum_{\alpha=1}^{n_t} P(r_j|t_\alpha)P_m(t_\alpha)}, \tag{7.3}$$

where $P(r_j|t_i)$ is the prior probability for a signal event in true bin t_i to be found in reconstructed bin r_j, and $P_m(t_i)$ is the prior probability of observing an event in true bin t_i. As denoted by the iteration index m, this formula can be applied iteratively, updating the priors each time.

To calculate the probabilities in equation 7.3, we define:

- The **signal matrix** S_{ji} — the number of selected signal interactions in true bin t_i that were reconstructed in bin r_j.
- The **missed vector** M_i — the number of signal interactions in true bin t_i that were not selected.

From these quantities we can make an initial estimate of the number of signal events in true bin t_i,

$$N_i = \sum_{j=1}^{n_r} S_{ji} + M_i, \tag{7.4}$$

which makes it simple to calculate the prior probability for a selected signal event to be found in reconstructed bin r_j:

$$P(r_j|t_i) = \frac{S_{ji}}{N_i}. \tag{7.5}$$

The prior probability to observe a true signal event in bin t_i is similarly given by

$$P_0(t_i) = \frac{N_i}{\sum_{\alpha=1}^{n_t} N_\alpha}. \tag{7.6}$$

S_{ji} and M_i therefore provide all the inputs necessary to calculate the probability $P_m(t_i|r_j)$ from Equation 7.3. We now look to apply this "unsmearing matrix" to the reconstructed dataset to obtain an estimate of the distribution of signal across the true bins t_i - known as the "unfolded distribution".

To do this, we must also account for the background surviving the selection. The **background vector** B_j is defined as the number of background events surviving the selection in reconstructed bin r_j; the estimate of the number of true signal in true bin t_i is then given by

$$N_i^{m+1} = \epsilon_i^{-1} \sum_j^{n_r} P_m(t_i|r_j)(N_j^{\text{measured}} - B_j), \tag{7.7}$$

where $N_j^{measured}$ is the reconstructed dataset, and ϵ_i the signal efficiency in true bin t_i, given by

$$\epsilon_i = \sum_{j=1}^{n_r} \frac{S_{ji}}{N_i}. \tag{7.8}$$

As previously mentioned, this estimate can then be iteratively improved by updating the prior probability $P_m(t_i)$. The initial prior given in Equation 7.6 is revised:

$$P_m(t_i) = \frac{N_i^m}{\sum_{\alpha=1}^{n_t} N_\alpha^m}. \tag{7.9}$$

Updating the priors in this way allows the unfolding to self-correct for differences between data and Monte Carlo; however, it also inflates the effect of statistical fluctuations

with each iteration, increasing the statistical error. Since this analysis is already strongly statistics-limited, only one iteration is performed.

7.1.2 Fake Data Tests

To test the capability of the unfolding method to compensate for discrepancies between data and Monte Carlo, I performed a study where the initial inputs (S_{ji}, M_i and B_j) generated from GENIE and NEUT Monte Carlo were used to unfold fake data from the other generator. As can be seen in Figure 2.7, there are tensions between the two distributions in the high-multiplicity bins, making this a robust test of the unfolding performance. The results of this study are shown in Figure 7.2.

The unfolding did not perform as expected in these tests. Investigating, I found that the poor performance was caused by significant differences in the efficiency of the selection (as defined in Equation 7.8). The ratio of the cumulative efficiency for each cut is shown in Figure 7.1.

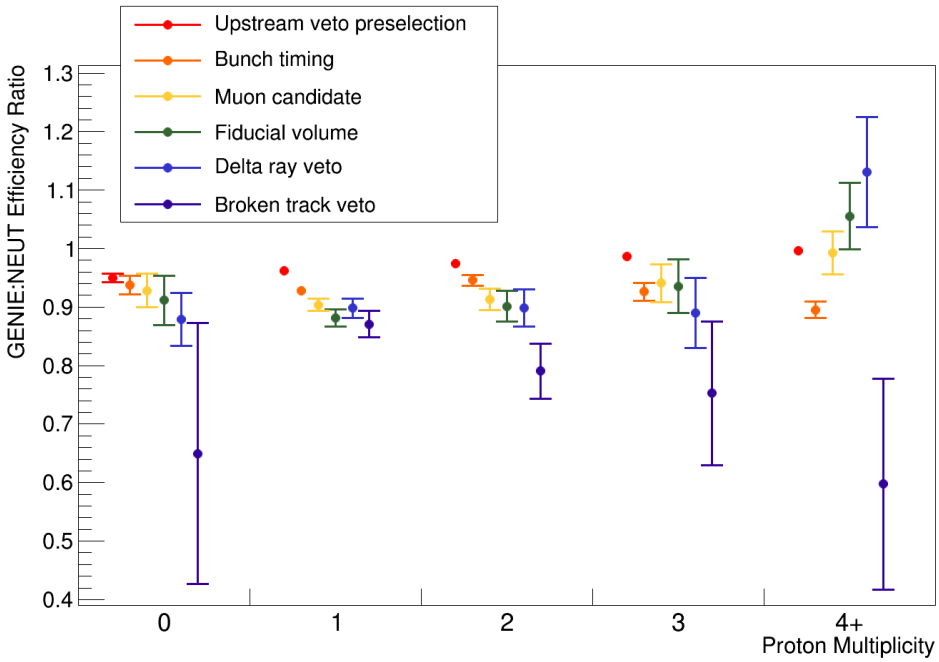


Figure 7.1: Plot showing the cumulative selection efficiency cut-by-cut for Monte Carlo gas interactions in TPC1, as a function of the proton multiplicity. All cuts show some model dependence, leading to a $\sim 40\%$ discrepancy in the 4+ proton bin (and large discrepancies in other bins also). TPCs 2 and 3 display similar behaviour.

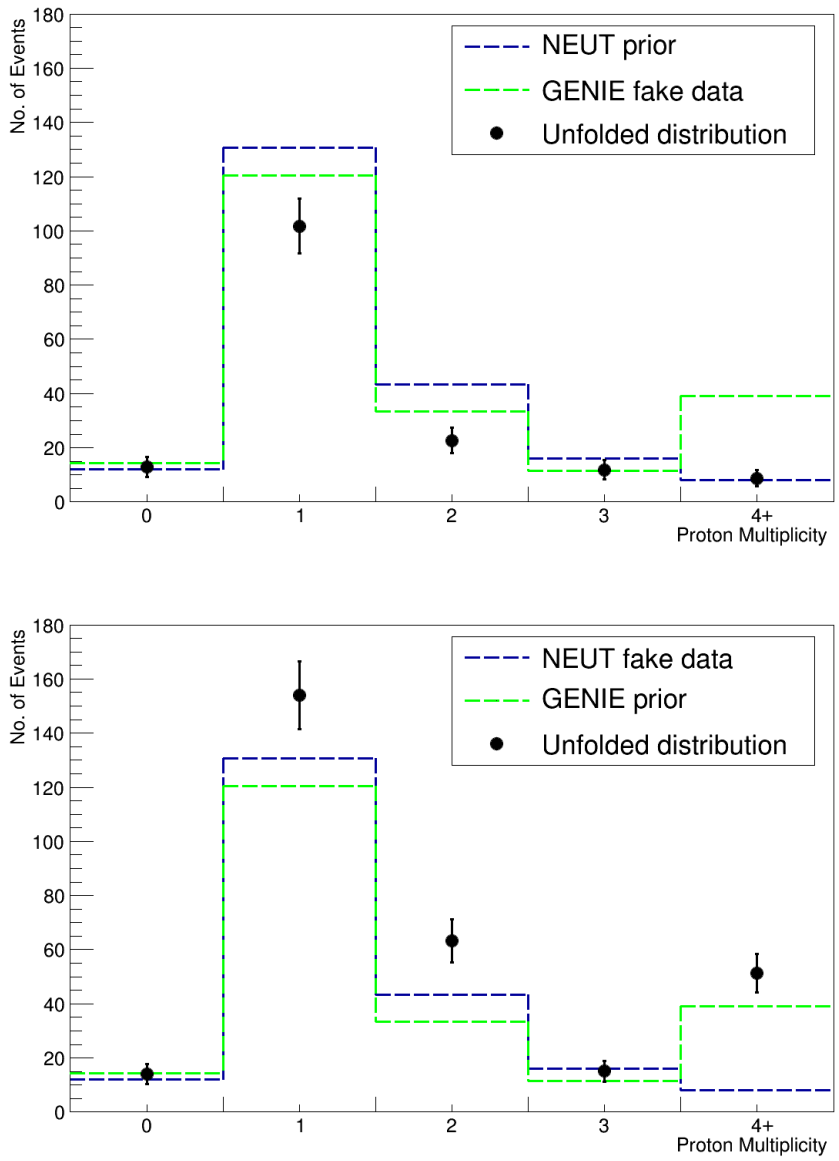


Figure 7.2: Plots showing results of unfolding GENIE fake data with the NEUT prior prediction (top) and NEUT fake data with the GENIE prior prediction (bottom) in TPC1. In both cases the unfolding can be seen to perform poorly, with the unfolded distribution matching the prior more closely than the true distribution of the fake data. Error bars are statistical only.

The strongest model dependence emerged in the final cut (the broken track veto), most probably as a result of the interaction between the colinearity cut (described in Section 5.2.1) and generator differences in kinematics. Differences are also visible in the previous cuts, even at the level of the preselection — indicating that signicantly more

GENIE signal events were triggering the upstream veto than in NEUT. The differences in the preselection efficiency are summarised in Table 7.1.

Module	Preselection efficiency (NEUT)	Preselection efficiency (GENIE)
TPC1	$99.47\% \pm 0.05\%$	$96.6\% \pm 0.1\%$
TPC2	$99.42\% \pm 0.05\%$	$96.5\% \pm 0.1\%$
TPC3	$99.73\% \pm 0.04\%$	$96.8\% \pm 0.1\%$

Table 7.1: A table of the net signal efficiencies for NEUT and GENIE Monte Carlo in each of the three TPCs.

In addition, comparisons of the rates of background rejection (given in Table 7.2)show the rejection rate to be very similar in both NEUT and GENIE, confirming that this is a phenomenon restricted to the signal.

Module	Background rejection (NEUT)	Background rejection (GENIE)
TPC1	$96.89\% \pm 0.01\%$	$97.14\% \pm 0.02\%$
TPC2	$95.36\% \pm 0.01\%$	$95.73\% \pm 0.02\%$
TPC3	$94.85\% \pm 0.01\%$	$95.21\% \pm 0.03\%$

Table 7.2: A table of the net background rejection from the preselection, for NEUT and GENIE Monte Carlo in each of the three TPCs.

Examining the lost GENIE events, the cause of this discrepancy appeared to be a higher proportion of interactions with backward-going particles leaving hits in the upstream detectors and triggering the veto - giving another indication that the discrepancies in efficiency arise from the selection's dependence on model-specific kinematics. An example of such an event is shown in Figure 7.3.

Due both to the fact that this is the first generation of this analysis, and to the low signal rate, there is no available control sample of signal or signal-like events. Therefore it is not possible to test which prediction of the selection efficiency more closely matches data. In order therefore to choose a prior informed by the data, I examined the total number of events passing the preselection in NEUT, GENIE and data, scaled to the data POT. These numbers are shown in Table 7.3.

The number of preselection-passing events in the data is 12% lower than in NEUT, and 33% higher than in GENIE. For this reason I chose to use NEUT in my unfolding

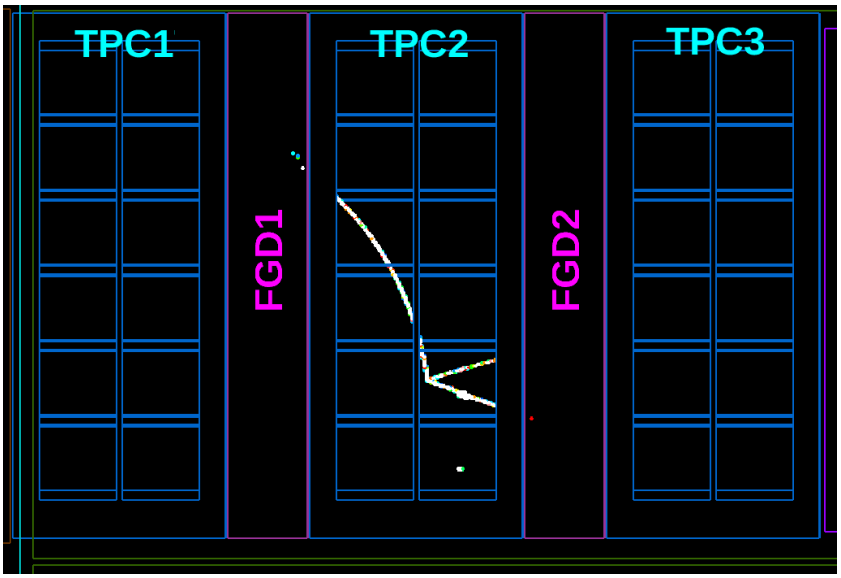


Figure 7.3: A signal event failing the preselection in GENIE Monte Carlo. The backward-going track leaves enough hits in FGD1 to trigger the upstream veto. A higher rate of these events in GENIE is consistent with GENIE's preference for vertices with higher particle multiplicity (as shown in Figure 2.7).

	No. of events in TPC1	No. of events in TPC2	No. of events in TPC3
GENIE	$151,460 \pm 152$	$224,249 \pm 185$	$247,586 \pm 194$
NEUT	$252,167 \pm 770$	$376,030 \pm 940$	$416,818 \pm 989$
Data	$197,654 \pm 445$	$360,119 \pm 600$	$372,727 \pm 611$

Table 7.3: Table of events surviving the preselection in NEUT, GENIE and data, with both Monte Carlo samples scaled to the POT of the data. The statistical errors reflect the size of each sample before scaling.

model, as the generator that is closest to the preselection rejection rate observed in data. The fact that neither generator matches the data well indicates a strong, unavoidable model dependence to my result. Further testing of the preselection's performance on data will require re-running the preselection over the complete dataset, which is not possible with the computing resources available for this cycle of the analysis. Future analyses will reappraise the cuts used in the preselection.

In addition to testing the ability of the unfolding to unsmear between generators, I also tested its ability to unfold a heavily reweighted NEUT distribution using the nominal NEUT prior. The results of this test are shown in Figure 7.4.

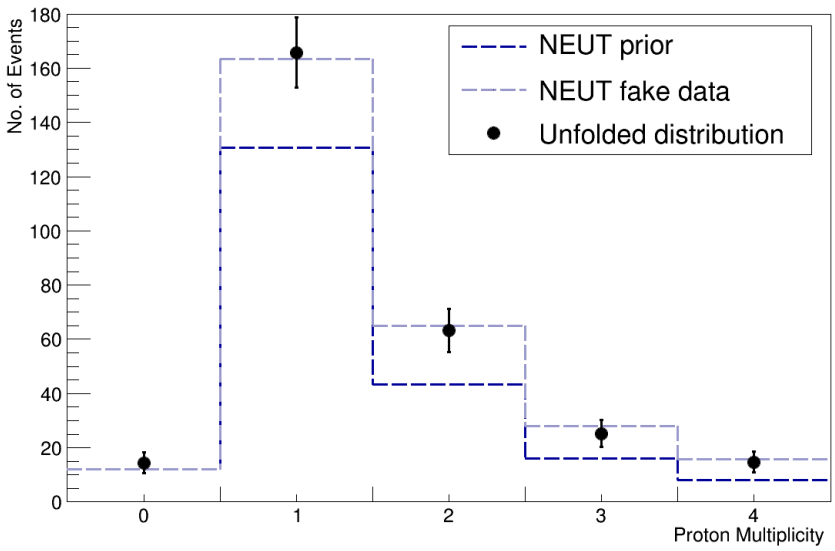


Figure 7.4: A plot showing the unfolding of a reweighted NEUT fake data distribution with the nominal NEUT prior. To produce the fake data I applied weights of $1 + M/4$ to events in each proton multiplicity bin, where M is the proton multiplicity. Error bars are statistical only.

As the figure shows, when given an accurate efficiency the unfolding has no difficulty in accurately finding the true distribution of the fake data.

7.2 Uncertainties

All uncertainties are assessed with the covariance matrix method described in Section 6.5. In addition to the systematic uncertainties described in Chapter 6, I assess a statistical uncertainty on both the data and the Monte Carlo used to generate the unsmearing matrix.

In the case of the Monte Carlo statistical uncertainty, I generated a covariance matrix from 1,000 throws in which I varied the inputs to the signal matrix, missed vector and background vector used in the unfolding according to Poisson statistics. The covariance matrix produced by this method is shown in Figure 7.5.

For the data statistical uncertainty, I perform 1,000 throws in which I varied the number of data events observed in each reconstructed bin according to Poisson statistics. The covariance matrix produced by this method is shown in Figure 7.6.

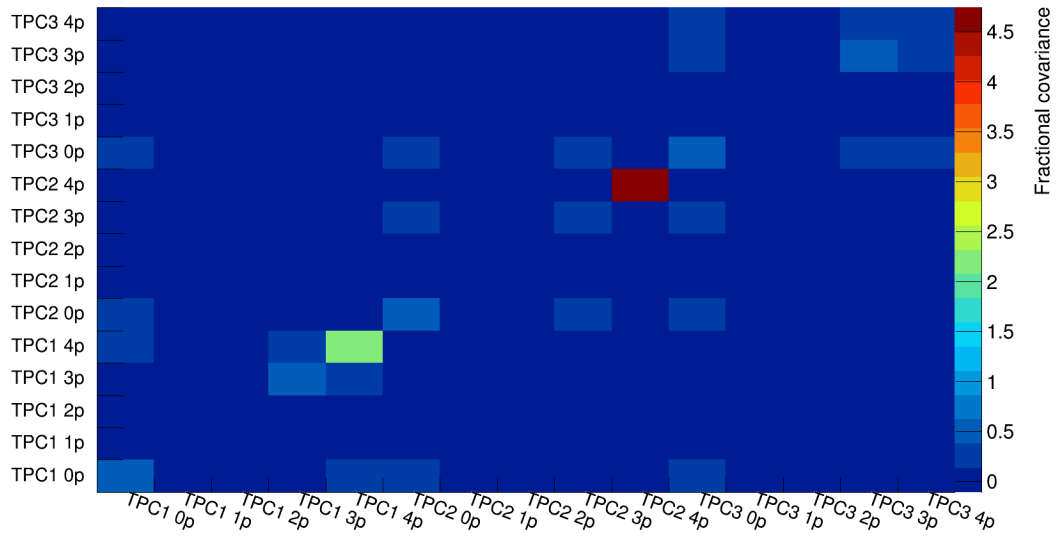


Figure 7.5: The fractional covariance matrix from Monte Carlo statistics. The largest uncertainties from this source arise in the 4+ proton multiplicity bins, where low Monte Carlo statistics in the background sample strongly affect the background subtraction in Equation 7.7. Migration from high- to low-statistics bins causes uncertainties greater than 100% in some on-diagonal bins (a feature also seen in some of the other covariance matrices).

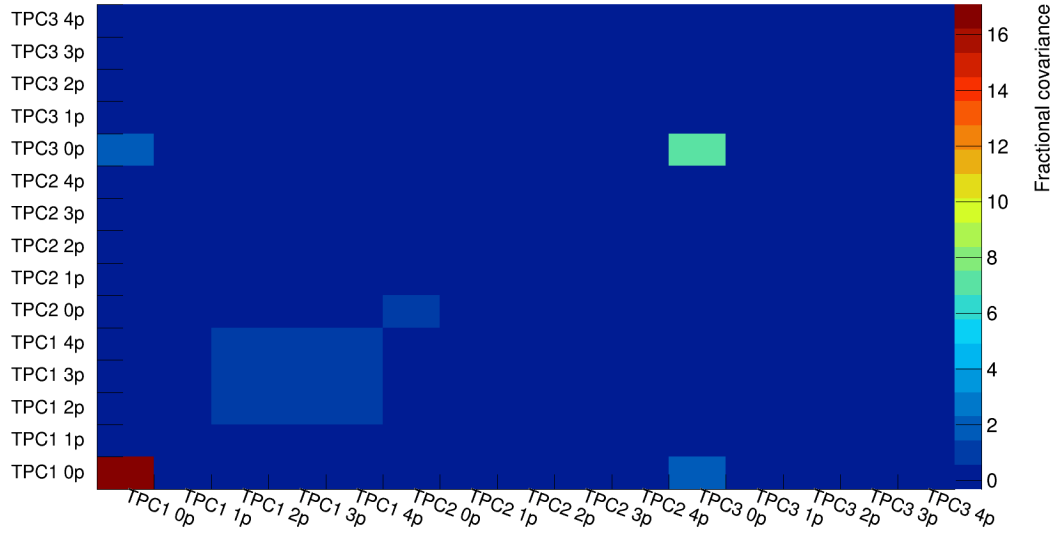


Figure 7.6: The fractional covariance matrix from data statistics. The largest uncertainties from this source arise in the zero-proton bins in TPCs 1 and 3, which were among the bins with the lowest non-zero populations.

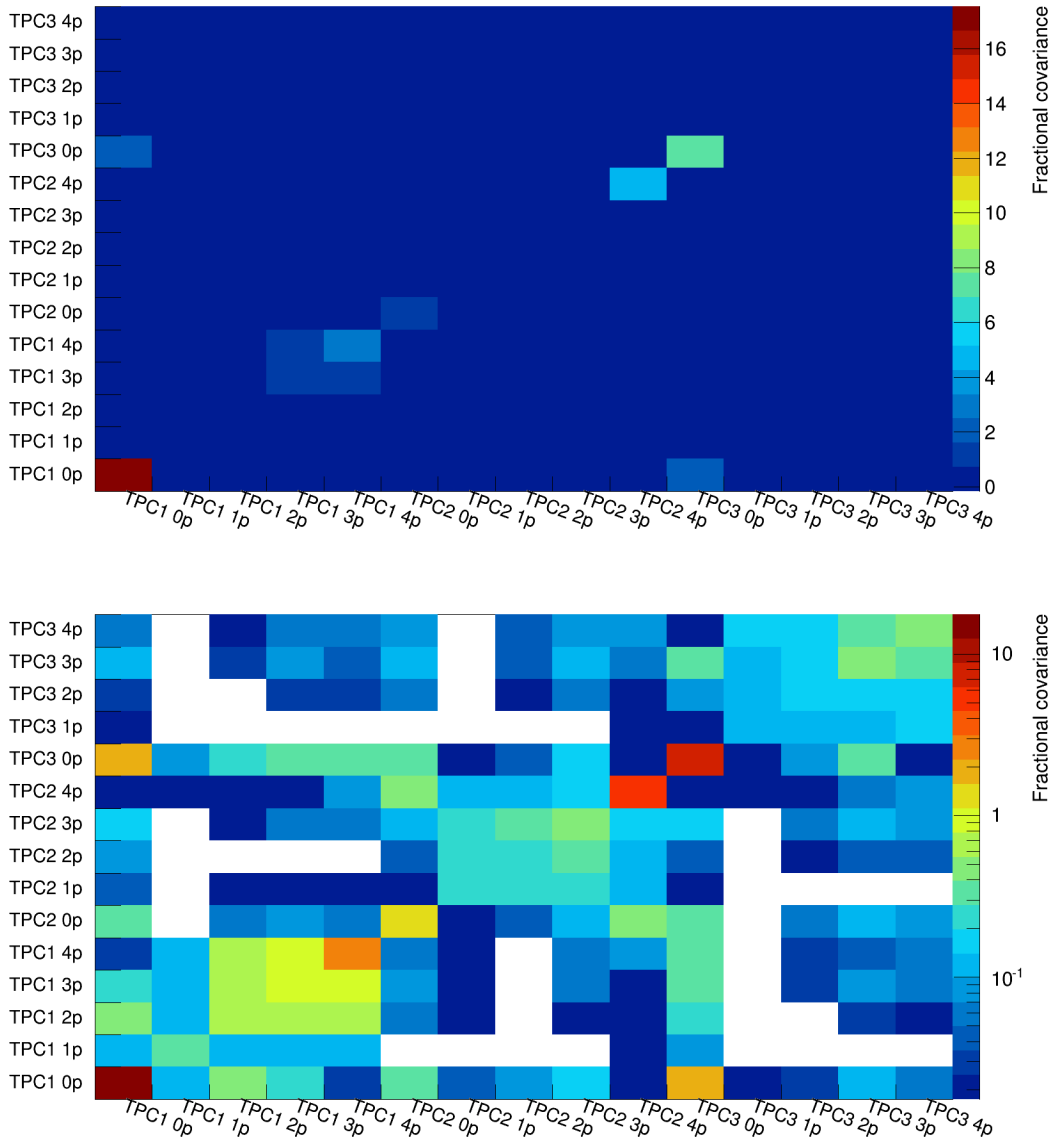


Figure 7.7: The fractional covariance matrix from all systematic and statistical sources of uncertainty, shown with both a linear (top) and log (bottom) scale. Extreme errors on the zero multiplicity and highest multiplicity bins arise due to the low population (in many cases zero) of the zero multiplicity and high multiplicity bins in data, showing that with the current statistics and selection it is difficult for this analysis to make strong statements about the tails of the multiplicity distribution.

Combining these covariance matrices with the covariance matrix from all systematic sources shown in Figure 6.8 gives the total covariance matrix for all sources of uncertainty, shown in Figure 7.7

As expected, the largest errors emerge from data statistics in the poorly-populated bins, with Monte Carlo statistics a close second. This analysis requires much more data before it will become systematically limited.

7.3 Results

Running the selection over the full available ND280 dataset — 5.9×10^{20} POT — produced 63 events that passed all the cuts. This was significantly lower than the prediction from Monte Carlo of 135 ± 2. The distribution of these events in reconstructed proton multiplicity is shown in Figures 7.8 and 7.9.

I unfolded these distributions using unsmearing matrices calculated from the same Monte Carlo samples used in assessing the uncertainties: 5.6×10^{22} POT of interactions in the gas (generated in NEUT) to simulate the signal, and 10.7×10^{21} POT of interactions throughout ND280 (generated in GENIE) to simulate the background, both scaled to the POT of the existing ND280 dataset (5.9×10^{20} POT). The unsmearing matrices for each TPC are shown in Figure 7.10.

Applying these unsmearing matrices through one iteration of the unfolding produces the unfolded distributions shown in Figures 7.11 and 7.12.

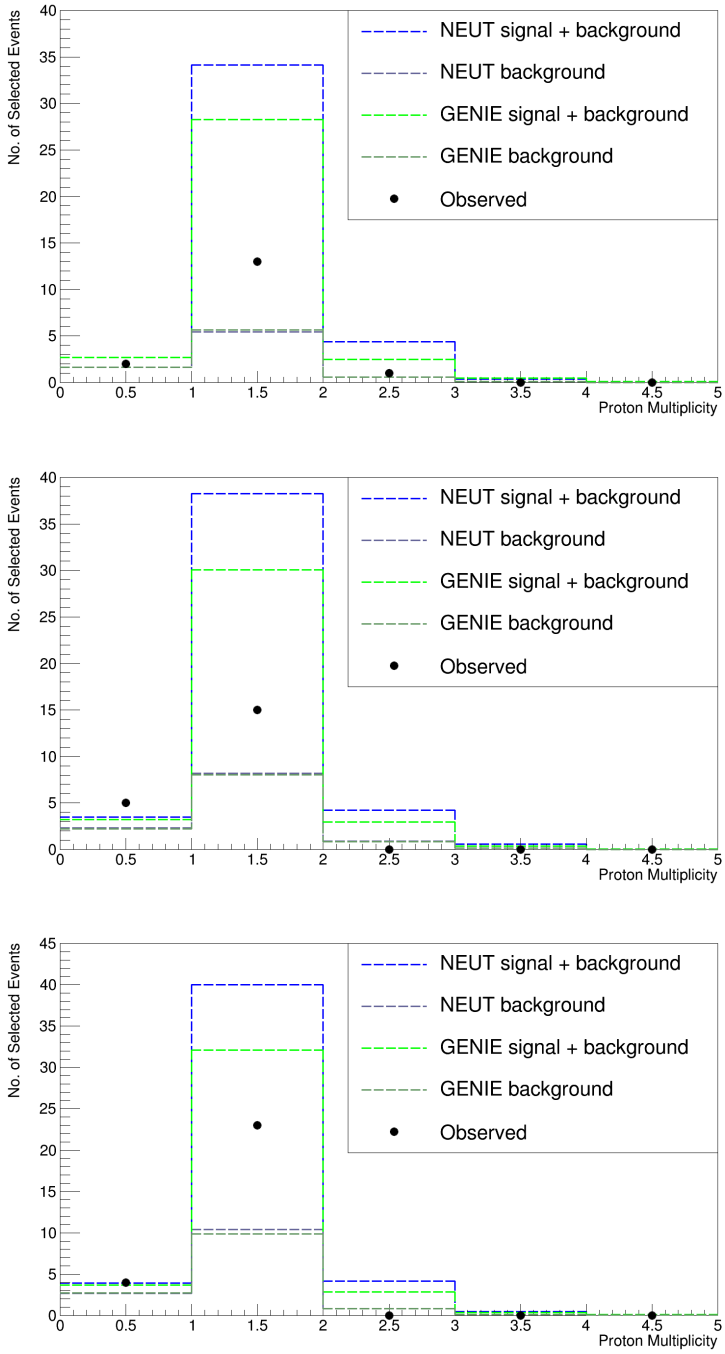


Figure 7.8: Plots showing the number of selected events before unfolding in TPC1 (top), TPC2 (middle) and TPC3 (bottom), compared to the predicted distributions of selected events from NEUT and GENIE for the same exposure. No statistical error bars are drawn on the data points, as the full treatment of the statistical error is given in the unfolded distributions in Figures 7.11 and 7.12.

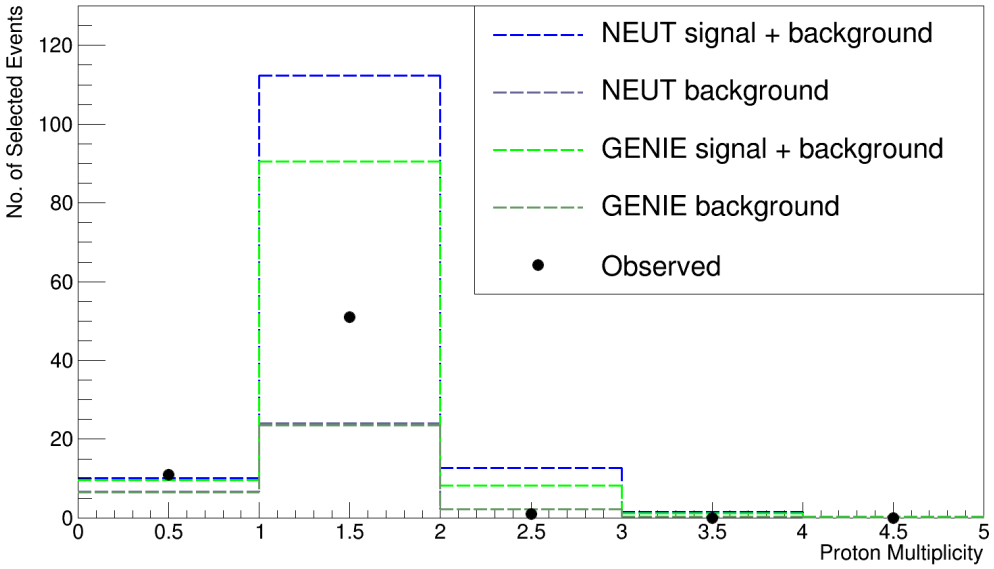


Figure 7.9: Plot showing the number of selected events across all three TPCs, compared to the predictions from NEUT and GENIE for the same exposure. For the same reasons as in Figure 7.8, no statistical error bars are drawn on the data.

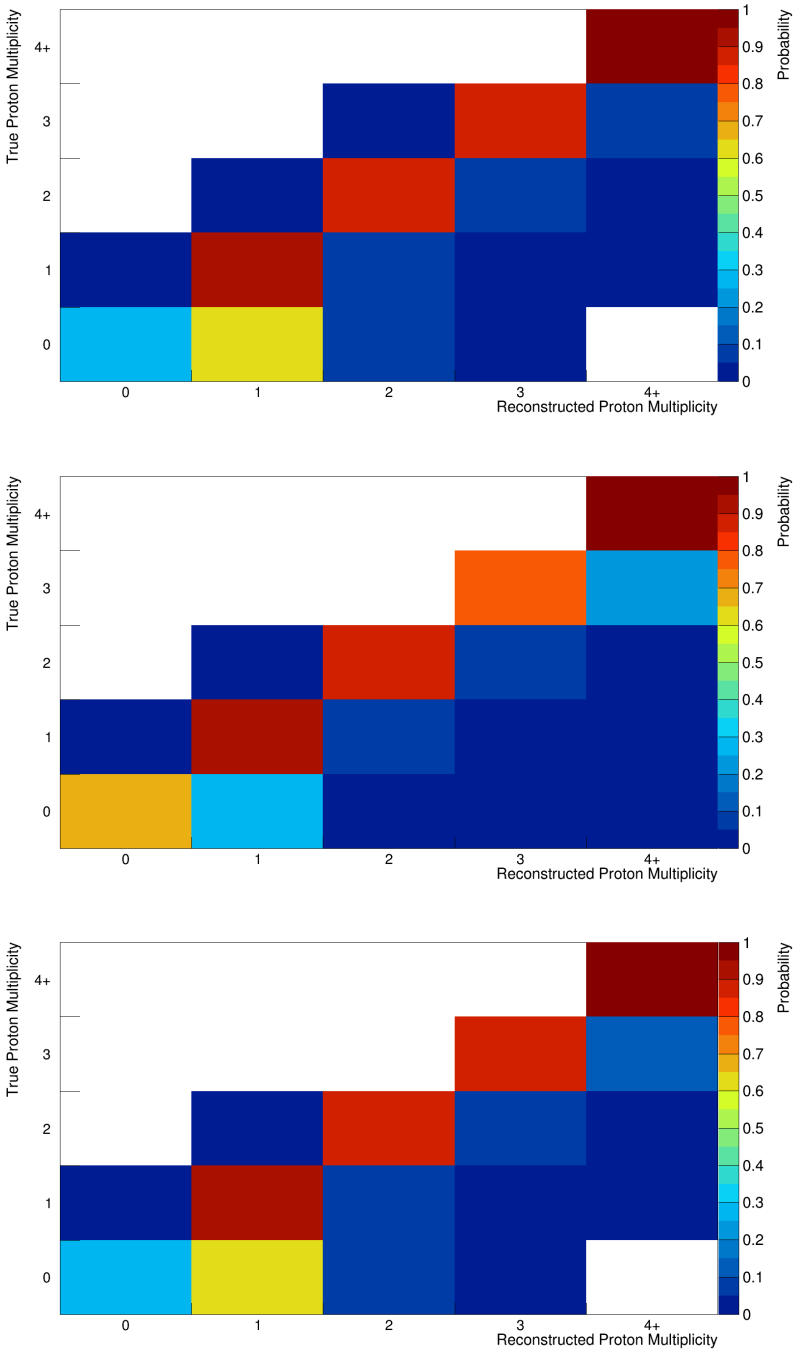


Figure 7.10: Plots showing the unsmearing matrices (as defined in Equation 7.3) calculated for TPC1 (top left), TPC2 (top right) and TPC3 (bottom).

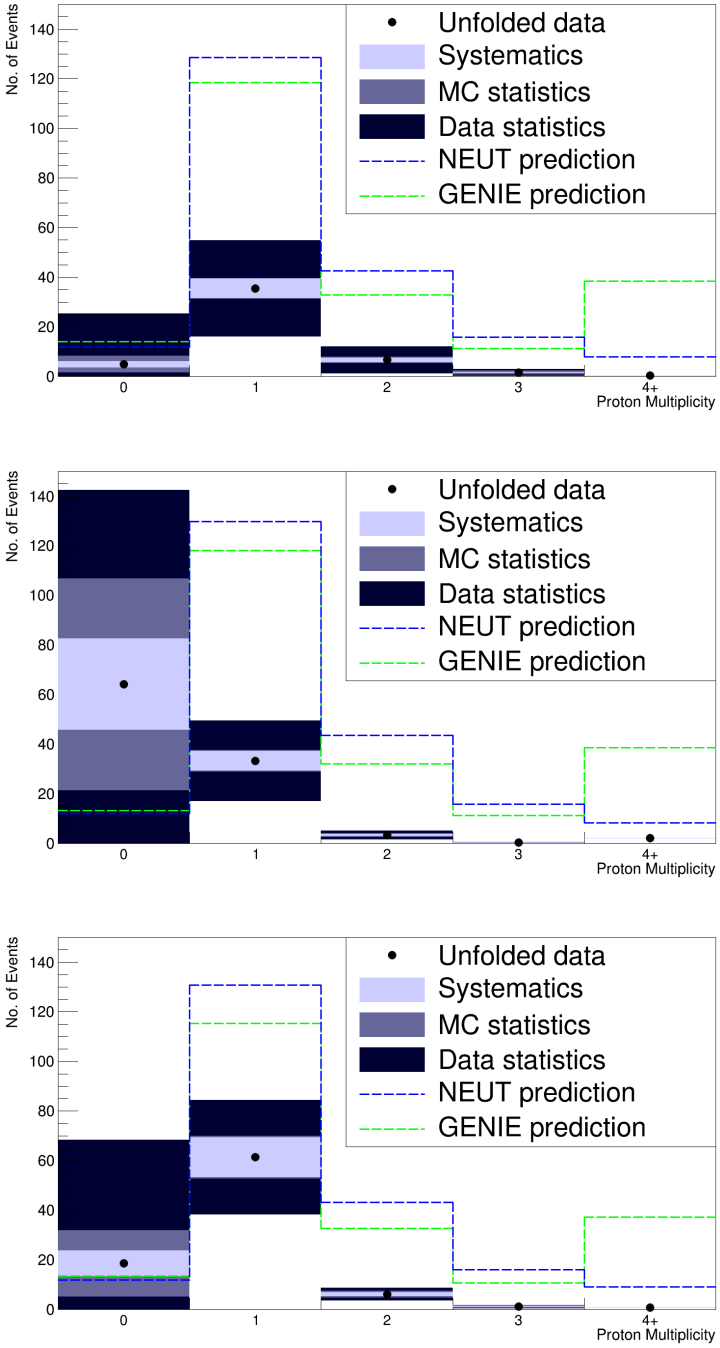


Figure 7.11: Plots showing the unfolded distributions for TPC1 (top), TPC2 (middle) and TPC3 (bottom).

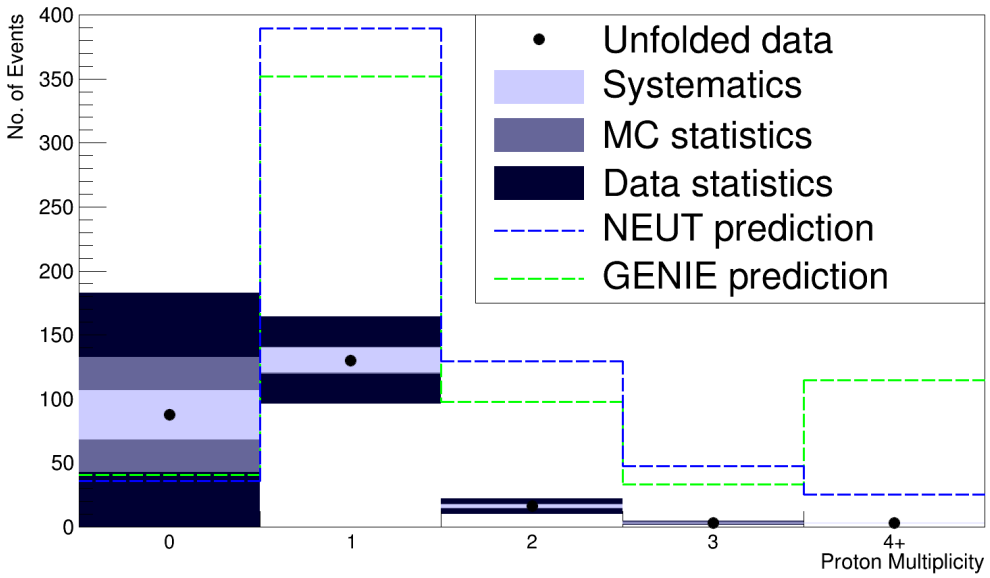


Figure 7.12: Plot showing the unfolded distribution across all three TPCs.

From the total differential cross-section shown in Figure 7.12, the integrated cross-section can be calculated:

$$\langle\sigma\rangle_{\Phi} = T\Phi \sum_{n=0}^{n=4+} \frac{\partial\langle\sigma\rangle_{\Phi}}{\partial n} \tag{7.10}$$

$$= \frac{1}{T\Phi} \sum_{n=0}^{n=4+} N_n \tag{7.11}$$

$$= 2.4 \times 10^{-39} \pm 2.6 \times 10^{-39} \text{ cm}^2 \text{ per nucleon.} \tag{7.12}$$

This does not resemble the NEUT prediction of 6.4×10^{-39} cm^2 per nucleon or the GENIE prediction of 6.5×10^{-39} cm^2 per nucleon, although it is consistent within 2σ. Measurements of the muon neutrino charged-current cross-section on argon have been made by ArgoNeuT collaboration [73], who found a value of $6.6 \times 10^{-39} \pm 0.9 \times 10^{-39}$ cm^{2}GeV^{-1} per nucleon. Following a naive extrapolation by comparing the heights of the distribution at each energy in Figure 2.4, at the T2K beam energy of 0.6 GeV this would translate to a cross-section of 4.8×10^{-39} cm^2 per nucleon, which is lower than the generator predictions. However, this extrapolation treats the T2K beam as monochromatic (as well as extrapolating across a large energy difference), making the extrapolated figure questionable. No more similar data exists with which to make a better comparison.

While the motivation for this study was that the generators differ strongly on the shape of the proton multiplicity distribution, they should be able to predict the total cross-section with reasonable accuracy. A failure to agree with either simulation or ArgoNeuT data on this quantity indicates differences between data and Monte Carlo at the level of the selection performance. In particular, the unfolding procedure can return an incorrect result if the selection purity or efficiency are incorrectly characterised, as these determine the signal matrix, missed vector and background vector that make up the key inputs to the unfolding. My investigation of possible discrepancies in these values is described in Chapter 8

Chapter 8

Investigation of Results and Future Improvements

8.1 Investigation of Data-Monte Carlo Discrepancies

As described in Chapter 7, the value of the cross-section returned by the Bayesian unfolding is clearly erroneous. Unforeseen differences between data and the Monte Carlo on which the performance of the selection was parameterised are the most likely explanation for this discrepancy. To investigate, I examined the 63 events that had passed the selection by hand, using the ND280 event display. Most were clearly either signal or backgrounds, and so I was able to make an estimate of the "true" purity from a simple count. The results of this study are given in Table 8.1.

Module	No. of events resembling signal	Total no. of events	Observed Purity
TPC1	13	16	$81\% \pm 10\%$
TPC2	11	20	$55\% \pm 11\%$
TPC3	15	27	$56\% \pm 10\%$
Total	39	63	$62\% \pm 6\%$

Table 8.1: Table of the purity of the gas sample, as inferred from examining the data by eye. I make no attempt to assign an uncertainty to my judgement of whether events are signal or background; example event displays are shown in Figures 8.1 to 8.5.

It is interesting to note (within the limitations of the low sample statistics) that the observed purity of the TPC1 selection closely matches the predicted purity in that

module, while the purity in both TPC2 and TPC3 undershoots the prediction. This implies that the additional backgrounds may be coming from the barrel ECals.

The mean observed "true" purity of 62% is significantly lower ($> 2\sigma$ deviation) than the anticipated 76% from the figures in Table 5.5. For a given efficiency, this would lead to an under-estimate of the background term B_j in Equation 7.7, resulting in an over-estimate of the signal rate and inflating the cross-section. Since the measured rate significantly undershoots the predictions from Monte Carlo, this implies a stronger effect working to reduce the number of selected signal events — i.e. a lower efficiency.

Since it is not possible to examine the signal events that have been cut, an estimate of the "true" efficiency cannot be made in the same way. However, the unexpectedly low number of selected events would be consistent with an efficiency much lower than expected, and the differences between NEUT and GENIE in Figure 7.1 already indicate a strong model dependence in the predicted efficiency.

It is not entirely surprising that there should be large unforeseen differences between data and Monte Carlo in the selected sample, as this analysis is sensitive to Monte Carlo discrepancies that would not strongly affect the official T2K analyses using the FGDs (and therefore may not have been brought to the attention of the collaboration). Discovering the source of the unexpected efficiency losses will require an intensive investigation to be carried out by the next analyser to carry this work forwards.

The success of the selection is extremely useful in providing us with actual signal events to study in data. In Section 8.2, I present some of the events seen in data and discuss the leads these events offer for investigating the discrepancy between data and Monte Carlo.

8.2 Sample Selected Data Events

The observation of 39 signal-like events in the selected sample proves that it is possible to select neutrino interactions on the TPC gas from the ND280 data with the tools we have developed - the primary purpose of this exploratory analysis. Three example data events are shown in Figure 8.1 to illustrate the quality of the selected data.

In addition, I observed one event with clear evidence of a low-energy particle from the interaction vertex stopping in the gas. This shows that daughter particles with energies in the unique sensitivity range for gas shown in Figure 2.7 are occurring at an appreciable

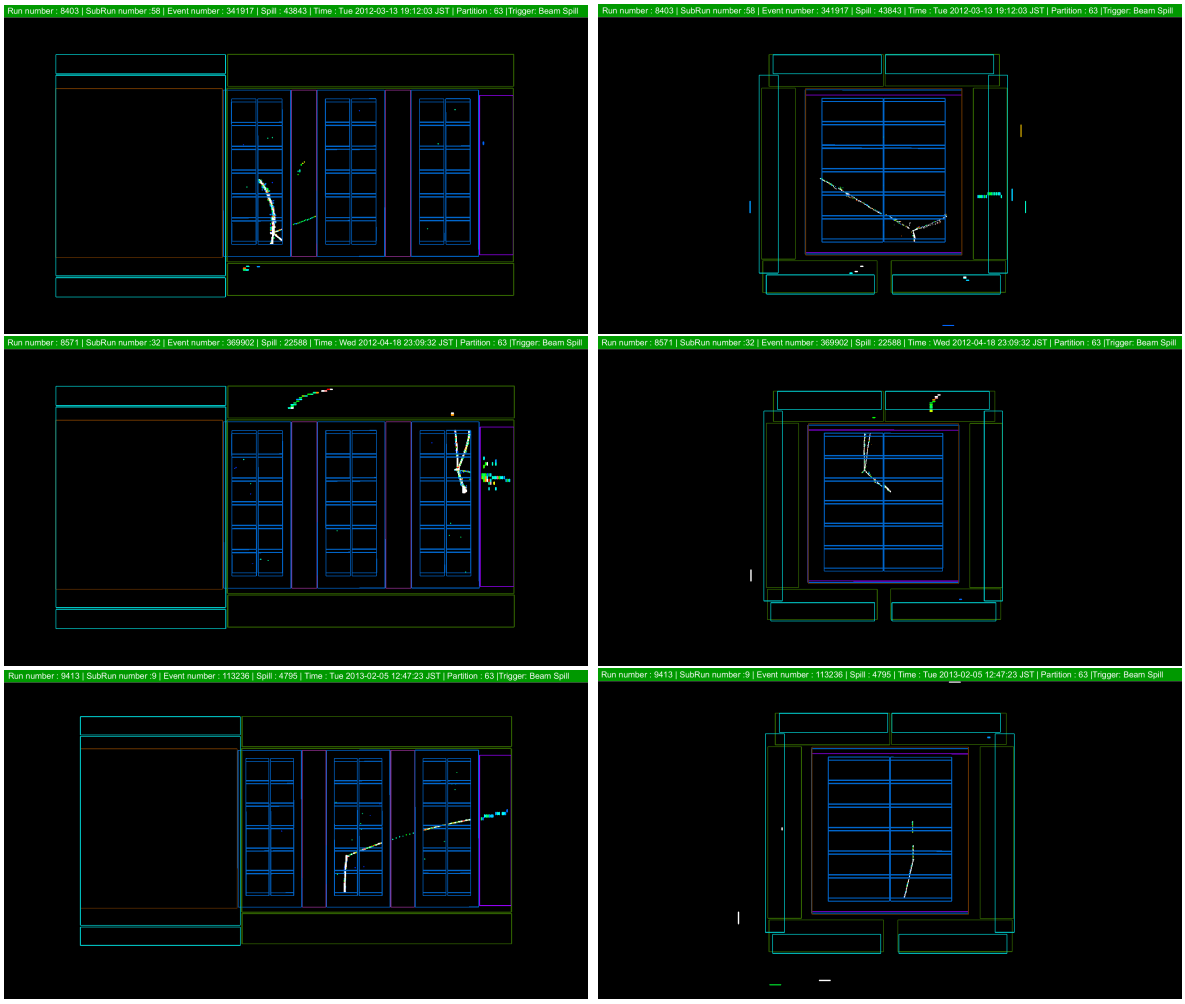


Figure 8.1: Event displays showing signal events seen in the selected data sample, in the YZ(left) and XY(right) views. The Z axis is aligned with the beam direction, and the X axis with the drift direction in the TPCs.

rate in the ND280 data, showing the clear value of this sample. The event in question is shown in Figure 8.2.

The selected sample is also useful for understanding the types of background that survive in data, and comparing these to the dominant background topologies in Monte Carlo. Three interesting examples of apparent backgrounds in the selected data are shown in Figures 8.3, 8.4 and 8.5.

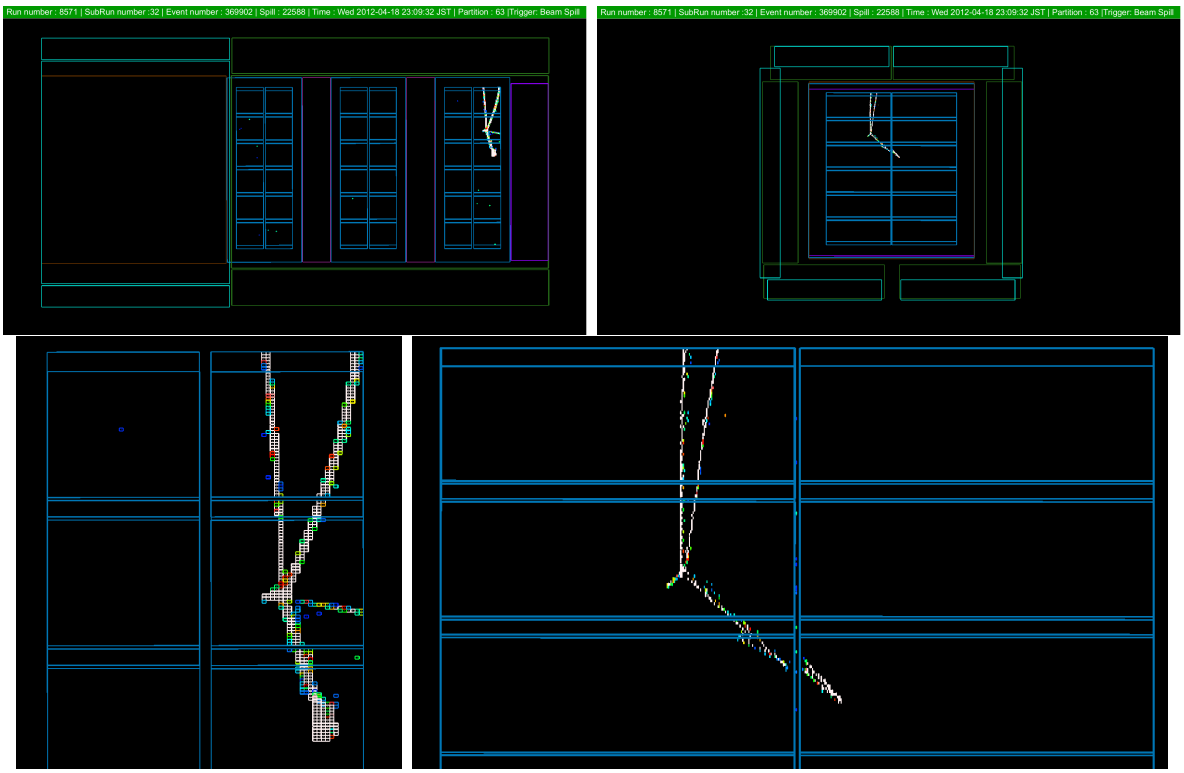


Figure 8.2: Event display showing signal event with a low-energy track emerging from the neutrino interaction vertex and stopping in the gas volume. The view of the entire detector is shown at the top, in the YZ (left) and XY (right) views. A close view of the interaction is shown in the lower images, in the same views (respectively).

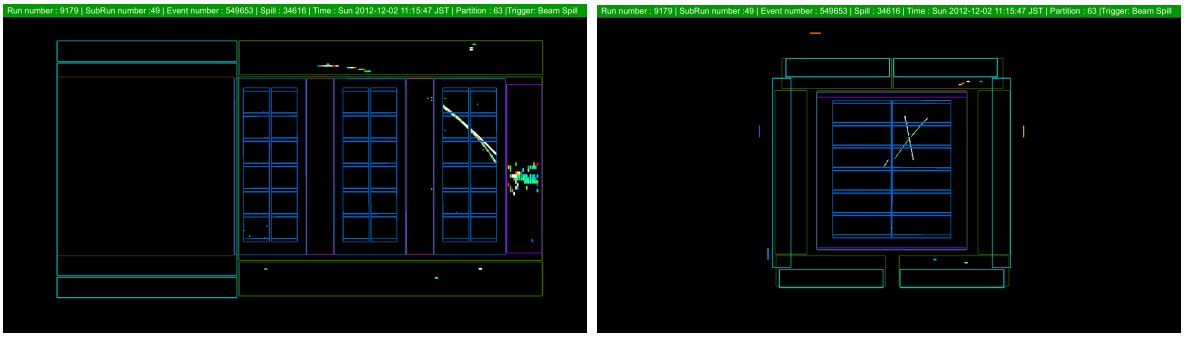


Figure 8.3: ND280 event display showing a background event seen in the selected data sample, in the YZ(left) and XY(right) views. Two particles from separate interactions outside the TPC cross over to create a false junction.

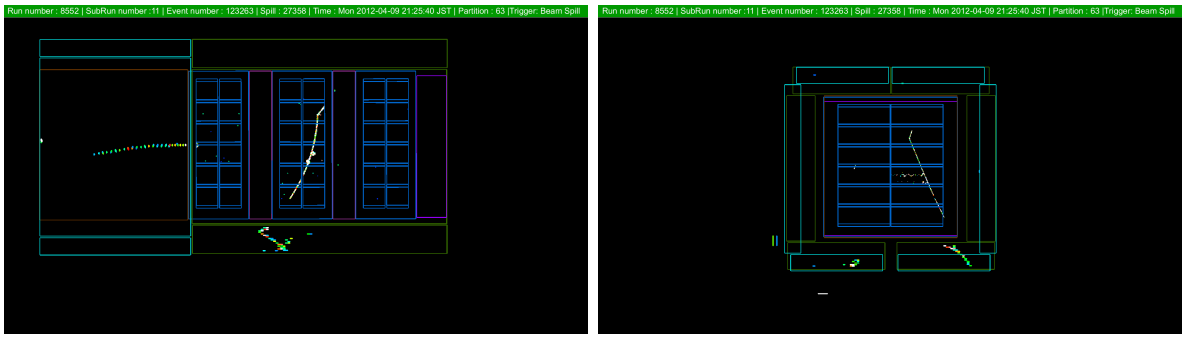


Figure 8.4: ND280 event display showing a background event seen in the selected data sample, in the YZ(left) and XY(right) views. A muon from an interaction in the bottom ECal scatters in the gas, forming a kink in the track that is detected as a junction. The short path segment after the scatter does not receive a muon PID and so this junction is selected as an interaction vertex.

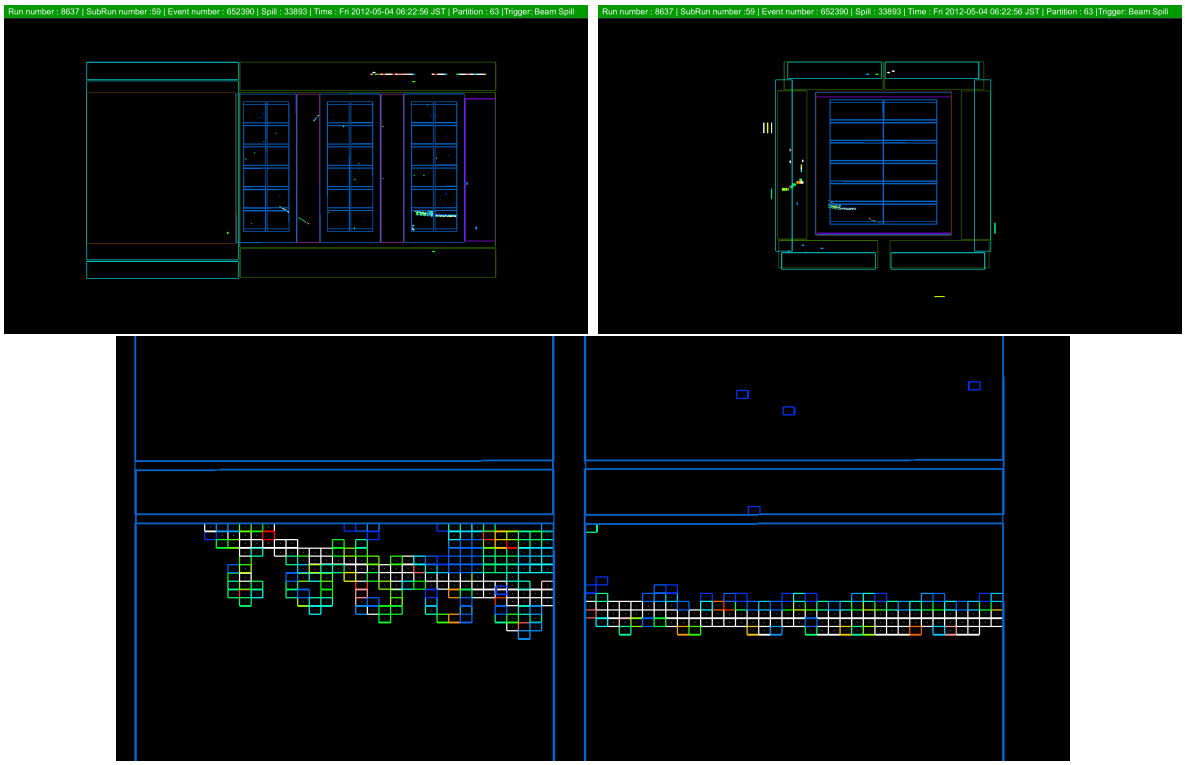


Figure 8.5: ND280 event display showing a background event seen in the selected data sample, in the YZ(top left) and XY(top right) views. A close-up of the selected track is shown at the bottom. The physics behind this "hairy" track is unclear: it may be one particle travelling inside the spiral trajectory of a δ-ray. Its complicated structure leads to the finding of false junctions, one of which is erroneously selected as a signal interaction.

From examination of these events and the others in the selected sample, I noticed three possible contributing effects to the discrepancies seen in the cross-section measurement. I list these effects below; the list is not intended to be exhaustive.

8.2.1 Noisy Tracks

I observed several events during the visual scan which showed tracks quite unlike those ordinarily seen in Monte Carlo. These tracks were typically extremely broad, with spurs of scattered hits that could be mistaken for additional tracks and large numbers of hits unused by the TREx pathfinding. An example event is shown in Figure 8.6.

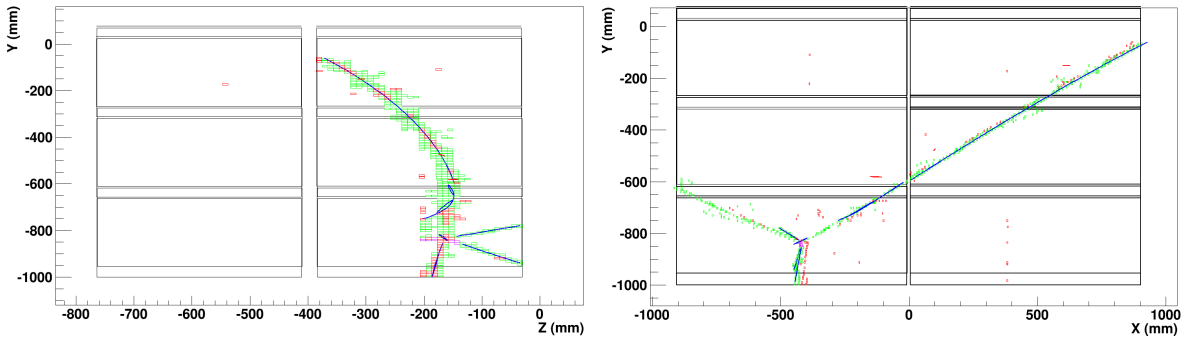


Figure 8.6: Orthogonal views of an event in the selected data sample with an unexpectedly thick, noisy track emerging from the vertex. Green boxes indicate MicroMEGAS hits used by the pathfinding, red boxes hits which were not used, and purple boxes hits that were assigned to a junction. Blue lines indicate fitted paths, and red lines paths for which the fitting failed. X is the drift axis and Z is the beam axis.

Similar topologies in Monte Carlo create patterns of hits with very clean tracks, as shown in Figure 8.7.

In total I observed thirteen data events with unexpectedly noisy tracks: six appearing to be signal events, and seven appearing to be backgrounds. This corresponds to approximately 20% of the total selected events. It is also interesting to note that of the three selected data events that appeared to the visual scan to have four or more particle tracks emerging from the vertex, all had at least one track that displayed these features — suggesting that they may be originating most frequently from highly ionising particles such as low-energy protons.

These events may indicate a need to update our noise model in the TPCs. The Monte Carlo study conducted in Section 6.3.3 provided a conservative estimate of the impact of

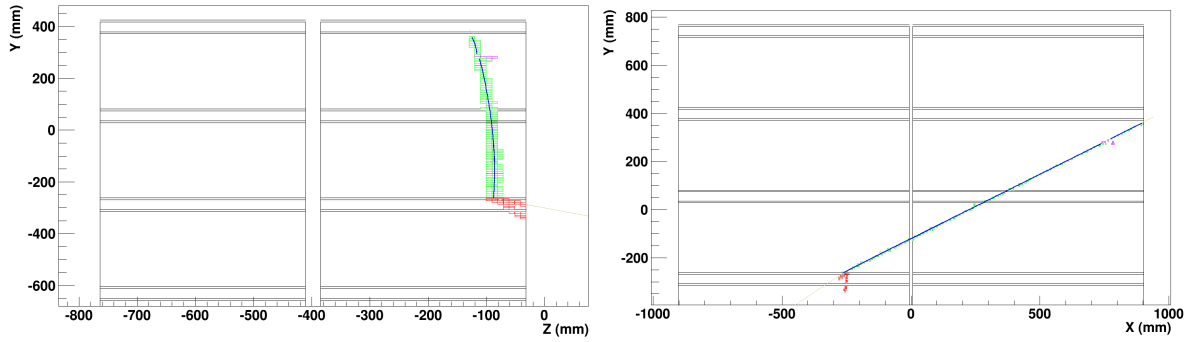


Figure 8.7: Orthogonal views of a signal event in NEUT Monte Carlo with a similar topology to the data event in Figure 8.6. All tracks emerging from the interaction vertex appear much cleaner in the pattern of hits they leave behind. To interpret the colour scheme refer to Figure 8.6.

MicroMEGAS pad noise using the existing simulation, which assesses the probability of a noise hit on any given pad independently from the other pads in the detector. Correlated noise from cross-talk between pads may have a much stronger effect on the performance of the reconstruction (as shown by the broken tracks in Figure 8.6, and be creating the thick, noisy tracks seen in data.

Cross-talk in the ND280 MicroMEGAS arises from fluctuations in the mesh voltage caused by large amounts of ionisation arriving at the MicroMEGAS at once. The cross-talk between adjacent pads has been measured at the level of 1.2% of the central hit charge [6], as shown in Figure 8.8. Charge from cross-talk would therefore begin to exceed the noise threshold for hits to be registered in adjacent pads at a central hit charge of approximately 43.9 fC — well below the saturation level of 120 fC.

Most tracks with hit charges of 30 fC and above come from protons, and many protons have been observed saturating the MicroMEGAS pads, indicating that cross-talk effects for proton tracking may be much stronger than for muons, pions and electrons. This hypothesis is supported by the charge distribution of the noisy tracks, as shown in Figure 8.9

The absence of cross-talk effects in the simple model used by the electronics simulation is therefore a very likely cause for us to mischaracterise the reconstruction of protons in Monte Carlo. Implementing this behaviour in our simulation will require significant effort to improve the model.

It is unknown how many signal events with these features may have been cut due to the poor performance of the reconstruction on these tracks. If noisy tracks increase the

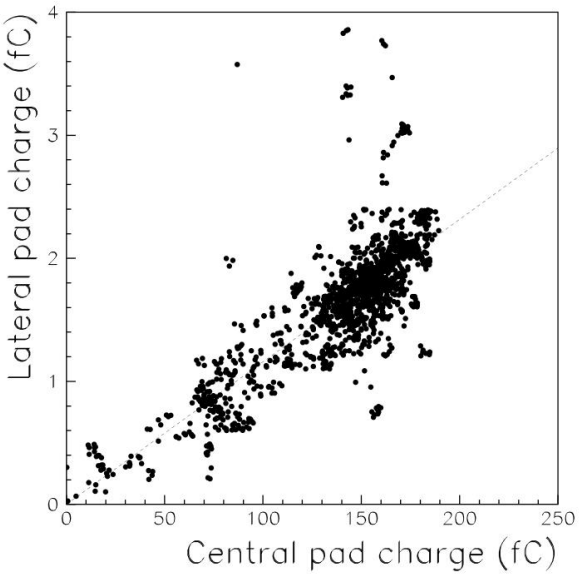


Figure 8.8: Plot showing the correlation between central hit charge and lateral hit charge (in fC) in the ND280 TPC MicroMEGAS pads, as measured before installation at the near detector site. The measurement was made by irradiating the TPC with a collimated ^{55}Fe source for which almost all the energy deposited in the detector is contained in a single pad. During these tests the TPC was operated with a higher saturation threshold (240 fC) than its current setting (120 fC). Plot taken from Reference [6].

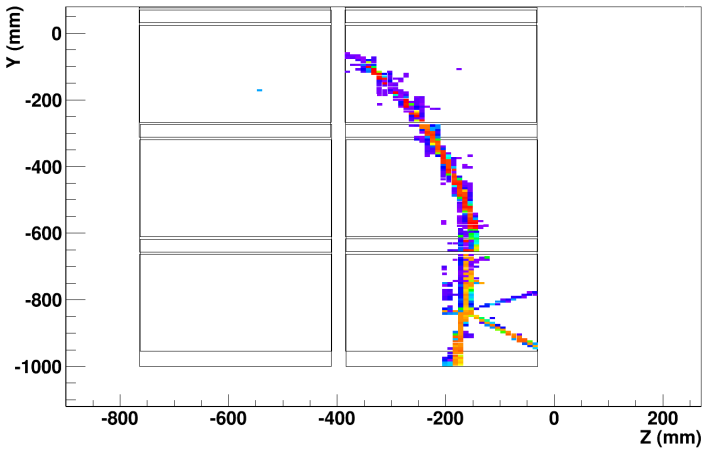


Figure 8.9: Y-Z view of the event shown in Figure 8.6, with a colour scale corresponding to the hit charge (in arbitrary units). No fitted tracks are drawn. A thin core of saturated (red and orange) MicroMEGAS hits can be observed following the "true" particle track; the surrounding blue hits have a very low charge. This is consistent with cross-talk from the saturated hits creating low-charge hits in neighbouring pads.

probability that an event is rejected, then this would reduce the selection efficiency, most probably disproportionately for events with higher proton multiplicity.

8.2.2 Incorrect Vertex Finding

In the course of the visual scan, I observed a total of eleven signal-like events with three or more particle tracks emerging from the interaction vertex. From NEUT simulation, one would expect 59% of these events to have at least two protons emerging from the vertex. According to the predicted smearing matrices in Figure 7.10, the majority of those events should be reconstructed with the correct number of protons (i.e. two or more). However, from the reconstructed distributions shown in Figure 7.8, only a single event was reconstructed with two protons, and none were reconstructed with three or more.

This discrepancy admits more than one possible explanation. It is possible that our reconstruction of the neutrino vertex is worse in data than in Monte Carlo for reasons not explored in Sections 6.3.3 and 6.3.4, e.g. the noisy tracks discussed above. It is also possible, however, that correctly reconstructed vertices are being lost due to "false vertices" elsewhere in the pattern - particularly if they occur upstream of the true vertex. This would affect not just the shape of the multiplicity distribution as misreconstructing the true vertex would do, but also reduce the overall normalisation as these false vertices would be rejected at a much higher rate than the genuine neutrino vertex (as they might easily not include a muon trajectory, or occur outside the fiducial volume). A more detailed visual scan, comparing the position of the reconstructed vertex selected to the coordinate of the visible interaction in the event display would be an excellent way of investigating this, albeit limited by the statistics of the selected sample.

8.2.3 Differing Angular Distributions

The majority of signal events appearing in the selected sample appear to be neutrino interactions with two daughter particles with a broad opening angle between them. This is the preferred topology forced by the colinearity cut, and physics differences between the true angular distribution of daughter particles and the distributions used in Monte Carlo may affect the prediction of the selection efficiency.

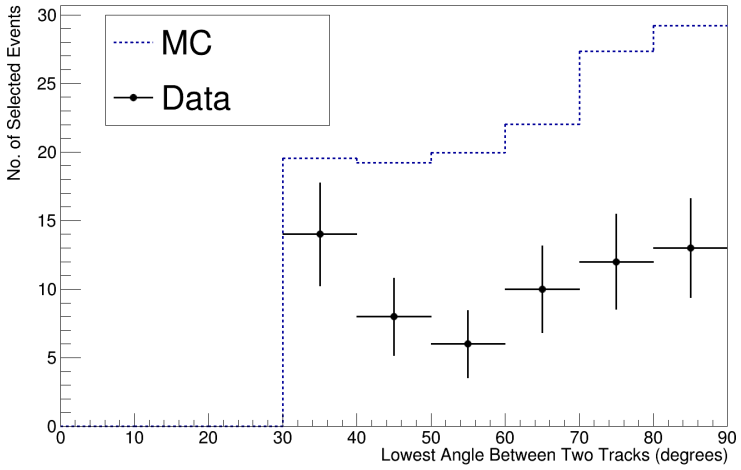


Figure 8.10: A histogram of the lowest angle between two tracks for both the data and the Monte Carlo prediction for the same exposure, post-selection. The colinearity cut removes all events with a lowest angle below 32°.

The parameter used by the colinearity cut is the lowest angle between any two particle tracks emerging from the interaction vertex. Figure 8.10 shows the distribution of this parameter for both the selected data and the Monte Carlo prediction post-selection. While the difference in normalisation is simply consistent with the lower efficiency observed previously, there is a noticeable difference in shape; this in turn implies that the shape of the data distribution may also differ from Monte Carlo in the cut region (below 32°). This could be the cause — or partial cause — of the overall efficiency loss.

These effects will be addressed as part of the ongoing validation of TREx and development of the selection. Plans for these future improvements are discussed in Section 8.3.

8.3 Future Improvements

8.3.1 Improved Preselection

The fake data study in Chapter 7.1 uncovered a significant difference in the predicted efficiency of the preselection between NEUT and GENIE, based on the different angular distribution of daughter particles from neutrino interactions in the two generators. This property of the preselection is highly undesirable as the angular distribution in data is

unknown. With more processing time to run the enhanced reconstruction, the preselection conditions could be relaxed, or the preselection could even be discarded altogether in order to avoid this model-dependent effect.

8.3.2 Improved Reconstruction

This analysis was carried out with the first stable version of TREx. At the time of writing, two improvements have already been made: extensions to the δ-ray recognition and improved merging over the MicroMEGAS gaps. The improved merging should lead to fewer broken tracks that appear to start in the fiducial volume, while the new handling of δ-rays should greatly reduce the number of false junctions found from messy δ tracks.

In addition, new methods of proton PID are being trialled to allow the identification of very low-energy protons. These protons have such a high rate of energy loss that they saturate the MicroMEGAS pads. A new technique of extrapolating a gaussian waveform based on the gradient to either side of the saturation plateau (as shown in Figure 8.11) shows much promise in identifying these tracks as protons.

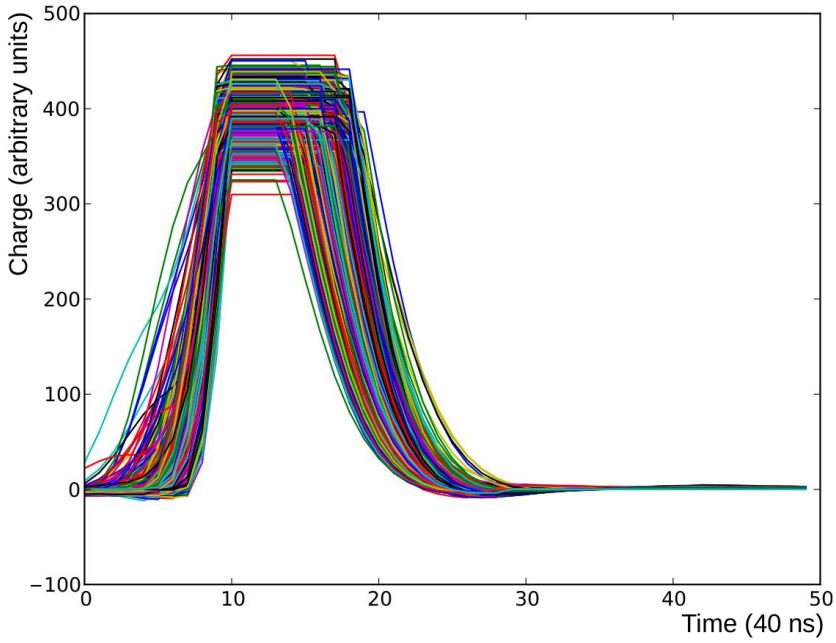


Figure 8.11: A selection of saturated waveforms in data, showing the saturation plateau. Plot provided by Lukas Koch.

The output of TREx is also not presently used by the ND280 global reconstruction, which joins together tracks from neighbouring subdetectors to build up a complete

picture of each neutrino interaction. This limited both the preselection and the main selection to treat the TPCs and their surrounding subdetectors as separate systems. Connecting tracks in other subdetectors to their corresponding TPC paths in TREx offers the opportunity to:

- Make the preselection more efficient, by preserving signal events where particles in upstream detectors miss the TPCs.
- Make the main selection more pure, by rejecting entering backgrounds that can be matched back to a vertex in another detector.
- Make the main selection more efficient in TPCs 1 and 2, by preserving signal events where the muon track in the TPC containing the vertex is too short to be identified, but can be picked up in the downstream TPCs.

TREx is designed as a fully working replacement for tpcRecon, as well as a tool for the gas interactions analysis; it is planned to be integrated into the global reconstruction in the near future.

Finally, in designing the reconstruction, we could only test its performance on signal in simulation. The success of the selection has provided us with a sample of 39 signal interactions in data, which should help us to better understand the reconstruction performance on gas vertices and make improvements where necessary.

8.3.3 Improved Selection

While it succeeded in selecting a reasonably pure sample of gas interactions, there are points on which the selection could be improved. Most importantly, as can be seen in Figure 7.1, the final cut — the misreconstruction veto — strongly suppresses events with a proton multiplicity of zero or higher than two. Making the efficiency of the selection less multiplicity-dependent may also circumvent some of the differences between data and Monte Carlo that seem to have affected the unfolding.

A simple improvement to the misreconstruction veto would be to refine the colinearity cut. This one-dimensional cut is responsible for the suppression of high-multiplicity signal events, and could be refined into a stepped two-dimensional cut with little difficulty. A suggested revised cut is shown in Figure 8.12

The improvements to the reconstruction detailed in Section 8.3.2 could also allow the other components of the misreconstruction veto to be relaxed. These cuts — the

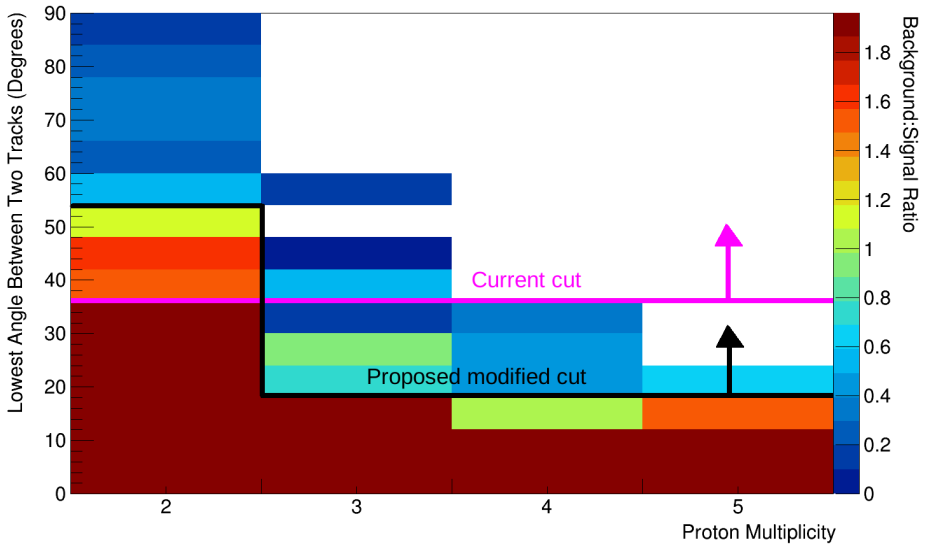


Figure 8.12: A plot showing the ratio between background and signal in Monte Carlo as a function of the true proton multiplicity and the lowest angle between two tracks, for events for which at least two joined tracks were found. The Z scale is truncated at 2 (i.e. background = 2 × signal).

single track veto and the PID test described in Section 5.2.1 — are responsible for the suppression of signal in the zero-proton bin and are largely necessary due to the frequency of track splitting on the MicroMEGAS gaps. The rate of gap-split backgrounds with the improved reconstruction will have to be reassessed before it can be determined whether these cuts can be made less stringent.

The use of successive cuts is also in itself inherently inefficient. A boosted decision tree working on the same variables used by the cuts would be likely to produce a more efficient selection. However, use of such techniques should wait until any differences between data and simulation in all of the cut variables are fully understood.

8.3.4 Improved Monte Carlo Modelling

The availability of the selected signal sample in data allows us to make comparisons with our signal Monte Carlo and check for any differences in behaviour between the two. In addition, the problems arising from the Monte Carlo statistics used in generating the unfolding matrix can be resolved by dedicating more processing time to the production of T2K beam Monte Carlo to simulate the background. Such large quantities of Monte

Carlo have not been required for previous T2K analyses, as no official selection has imposed such a steep reduction in the event rate.

In particular, the cross-talk between MicroMEGAS pads discussed in Section 8.2.1 is not included in our simulation of the electronics. Charge from cross-talk only approaches the threshold for creating new hits at central hit charges above approximately 44 fC, typical of highly-ionising tracks such as protons. Muon and pion tracks tend to create hits with lower charge and so this deficiency in simulation was of little interest to previous T2K analyses; for this analysis it may be of high importance in correctly simulating noisy proton tracks. A simple model of the cross-talk could be implemented at the level of the electronics simulation to bring our Monte Carlo closer to data.

By implementing all of these improvements, we expect to obtain a much more accurate understanding of the performance of the selection, as well as improve the overall efficiency. We also expect to discover new opportunities for improvement as the reconstruction software develops.

Chapter 9

Conclusions

Observations of neutrino interactions in gaseous detectors offer a unique window into the nuclear effects affecting neutrino cross-sections, as described in Chapter 2. This thesis aimed to achieve two goals:

- To develop the necessary tools and demonstrate the feasibility of selecting neutrino interactions on gas in an existing gaseous detector.
- To make a first measurement of the neutrino cross-section as a function of proton multiplicity from these interactions.

I succeeded in achieving the first goal. The cut-based selection has successfully selected 63 data events, of which 39 appear as genuine gas interactions to visual inspection. This is the world's first observation of neutrino-nucleus interactions in a gas target, and proves that even with a non-optimal detector, a neutrino interaction sample of reasonable size can be extracted with high purity. Planned improvements to both reconstruction and selection techniques, as well as ongoing data-taking at ND280, give us confidence that the size of our sample will greatly increase.

I only partially succeeded in achieving the second goal, as the extracted value of our flux-integrated cross-section was $2.3 \times 10^{-39} \pm 2.4 \times 10^{-39}$ cm^2 per nucleon., which appears inconsistent at the 1σ level with both the NEUT generator prediction of 6.4×10^{-39} cm^2 per nucleon and the GENIE prediction of 6.5×10^{-39} cm^2 per nucleon. This strongly implied the influence of unanticipated detector systematics.

Subsequent investigation identified what I believe to be the primary causes of the failure of the unfolding. Much of the work done for this thesis in developing the selection and evaluating the systematics remains valid and will be usable for the next cycle of the analysis. Furthermore, as the cross-talk effects encountered were not accounted for in

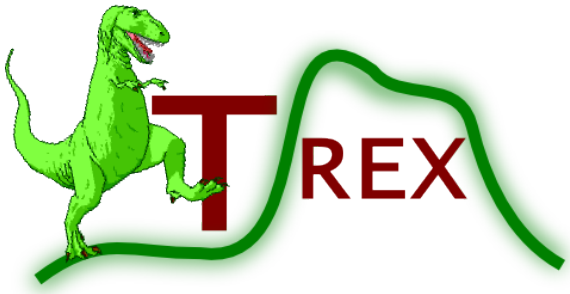
the ND280 simulation, a first analysis focused on the TPCs was critical for identifying the problem.

This thesis demonstrates the value — and challenges — of gaseous targets for future neutrino detectors, and paves the way for an official measurement at ND280 of the charged-current ν_μ cross-section on argon gas. Future work from the T2K collaboration will exploit this unique dataset to provide invaluable input to our models of nuclear effects in neutrino interactions.

Appendix A

Copyright Waiver

The TREx logo is derived from the character "T-Rex" used in Ryan North's Dinosaur Comics, found at www.qwantz.com. To use this image in our logo, I obtained a copyright waiver from Mr North, which I reproduce in this appendix. In the text of the waiver, "the extract above" refers to the TREx logo shown below.



I confirm that I am the copyright holder of the extract above and hereby give permission to include it in the print and electronic version of your thesis. I understand that the electronic version of the thesis will be made available, via the internet, for non-commercial purposes under the terms of the user licence.

Signed,



Name: Ryan North
Organisation: Dinosaur Comics
Job title: Chief Making The Dinosaurs Talk Guy

Bibliography

[1] J. A. Formaggio and G. Zeller, Reviews of Modern Physics **84**, 1307 (2012).

[2] A. Aguilar-Arevalo *et al.*, Physical Review D **81**, 092005 (2010).

[3] T2K Collaboration *et al.*, Nuclear Instruments and Methods in Physics Research A **659**, 106 (2011), 1106.1238.

[4] K. Abe *et al.*, Physical Review D **88**, 032002 (2013).

[5] T2K ND280 TPC collaboration, ArXiv e-prints (2010), 1012.0865.

[6] P. Baron *et al.*, Nuclear Science Symposium Conference Record, 2007. NSS '07. IEEE **6**, 4640 (2007).

[7] Particle Data Group, K. Olive *et al.*, Chin.Phys. **C38**, 090001 (2014).

[8] T2K, K. Abe *et al.*, Phys.Rev. **D91**, 072010 (2015), 1502.01550.

[9] W. Pauli, Open letter to the Gauverein meeting in Tubingen, 1930.

[10] C. Cowan, F. Reines, F. Harrison, H. Kruse, and A. McGuire, Science **124**, 103 (1956).

[11] R. Davis Jr, D. S. Harmer, and K. C. Hoffman, Physical Review Letters **20**, 1205 (1968).

[12] P. Anselmann *et al.*, Physics Letters B **357**, 237 (1995).

[13] J. Abdurashitov *et al.*, Physical Review C **60**, 055801 (1999).

[14] B. Pontecorvo, Zhur. Eksptl'. i Teoret. Fiz. **33** (1957).

[15] (SNO Collaboration), Q. R. Ahmad *et al.*, Phys. Rev. Lett. **87**, 071301 (2001).

[16] Y. Fukuda *et al.*, Physical Review Letters **81**, 1562 (1998).

[17] K. Eguchi *et al.*, Physical Review Letters **90**, 021802 (2003).

[18] Z. Maki, M. Nakagawa, and S. Sakata, Progress of Theoretical Physics **28**, 870 (1962).

[19] C. Giunti and C. W. Kim, *Fundamentals of neutrino physics and astrophysics* (Oxford university press, 2007).

[20] S. Mikheyev and A. Y. Smirnov, Yadernaya Fizika **42**, 1441 (1985).

[21] L. Wolfenstein, Physical Review D **17**, 2369 (1978).

[22] D. Larson *et al.*, The Astrophysical Journal Supplement Series **192**, 16 (2011).

[23] M. A. Luty, Physical Review D **45**, 455 (1992).

[24] T2K Collaboration, K. Abe *et al.*, Phys.Rev.Lett. **112**, 061802 (2014), 1311.4750.

[25] F. An *et al.*, Physical Review Letters **108**, 171803 (2012).

[26] J. Wolf *et al.*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **623**, 442 (2010).

[27] M. C. Chen *et al.*, arXiv preprint arXiv:0810.3694 (2008).

[28] R. Arnold *et al.*, The European Physical Journal C-Particles and Fields **70**, 927 (2010).

[29] R. Smith and E. Moniz, Nuclear Physics B **43**, 605 (1972).

[30] H. Meier-Hajduk, C. Hajduk, P. Sauer, and W. Theis, Nuclear Physics A **395**, 332 (1983).

[31] O. Benhar, A. Fabrocini, and S. Fantoni, Nuclear Physics A **505**, 267 (1989).

[32] ArgoNeuT Collaboration, R. Acciarri *et al.*, Phys.Rev. **D90**, 012008 (2014), 1405.4261.

[33] J. Nieves, I. Ruiz Simo, and M. Vicente Vacas, Phys.Lett. **B707**, 72 (2012), 1106.5374.

[34] Y. Hayato, Acta Phys.Polon. **B40**, 2477 (2009).

[35] C. Andreopoulos *et al.*, Nucl. Instrum. Meth. **A614**, 87 (2010), 0905.2517.

[36] C. Juszczak, J. A. Nowak, and J. T. Sobczyk, Nuclear Physics B - Proceedings Supplements **159**, 211 (2006), Proceedings of the 4th International Workshop on Neutrino-Nucleus Interactions in the Few-GeV Region Proceedings of the 4th International Workshop on Neutrino-Nucleus Interactions in the Few-GeV Region.

[37] O. Buss *et al.*, Physics Reports **512**, 1 (2012), Transport-theoretical Description of Nuclear Reactions.

[38] T. Y. for the ArgoNeuT Collaboration, Aspen Winter Workshop (2013).

[39] D. Renker and E. Lorenz, Journal of Instrumentation **4**, P04004 (2009).

[40] Y. Giomataris, P. Rebourgeard, J. P. Robert, and G. Charpak, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **376**, 29 (1996).

[41] I. Giomataris *et al.*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **560**, 405 (2006).

[42] K. Abe *et al.*, Physical Review Letters **107**, 041801 (2011).

[43] S. Kasuga *et al.*, Physics Letters B **374**, 238 (1996).

[44] R. Brun and F. Rademakers, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **389**, 81 (1997), New Computing Techniques in Physics Research V.

[45] S. Agostinelli *et al.*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506**, 250 (2003).

[46] G. Battistoni *et al.*, **896**, 31 (2007).

[47] R. Brun *et al.*, Version **3**, 14 (2002).

[48] N. Abgrall *et al.*, Physical Review C **84**, 034604 (2011).

[49] T. Eichten *et al.*, Nuclear Physics B **44**, 333 (1972).

[50] J. V. Allaby, F. Binon, and A. Diddens, (1970).

[51] Y. Hayato, Nuclear Physics B-Proceedings Supplements **112**, 171 (2002).

[52] C. Andreopoulos *et al.*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **614**, 87 (2010).

[53] C. L. Smith, Physics Reports **3**, 261 (1972).

[54] D. Rein and L. M. Sehgal, Annals of Physics **133**, 79 (1981).

[55] T. Sjöstrand, S. Mrenna, and P. Skands, Journal of High Energy Physics **2006**, 026 (2006).

[56] A. Bodek and J. Ritchie, Physical Review D **24**, 1400 (1981).

[57] Z. Koba, H. B. Nielsen, and P. Olesen, Nuclear physics B **40**, 317 (1972).

[58] R. Dechter and J. Pearl, J. ACM **32**, 505 (1985).

[59] B. List, ETH Zurich internal note (2002).

[60] A. A. Aguilar-Arevalo *et al.*, Phys. Rev. D **88**, 032001 (2013).

[61] G. Fiorentini *et al.*, Physical review letters **111**, 022502 (2013).

[62] C. Wilkinson, P. Rodrigues, S. Cartwright, L. Thompson, and K. McFarland, Phys.Rev. **D90**, 112017 (2014), 1411.4482.

[63] K. Ieki, *Observation of ν_μ to ν_e oscillation in the T2K experiment*, PhD thesis, Kyoto University, 2014.

[64] D. Ashery *et al.*, Phys. Rev. C **23**, 2173 (1981).

[65] K2K, S. Nakayama *et al.*, Phys.Lett. **B619**, 255 (2005), hep-ex/0408134.

[66] MiniBooNE, A. A. Aguilar-Arevalo *et al.*, Phys.Rev. **D81**, 013005 (2010), 0911.2063.

[67] SciBooNE, Y. Kurimoto *et al.*, Phys.Rev. **D81**, 033004 (2010), 0910.5768.

[68] MINERvA, A. Higuera *et al.*, Phys.Rev.Lett. **113**, 261802 (2014), 1409.3835.

[69] T2K Collaboration, K. Abe *et al.*, Phys.Rev. **D87**, 092003 (2013), 1302.4908.

[70] T2K Collaboration, K. Abe *et al.*, Phys.Rev. **D90**, 052010 (2014), 1407.4256.

[71] MINERvA Collaboration, B. Tice *et al.*, Phys.Rev.Lett. **112**, 231801 (2014), 1403.2103.

[72] G. D'Agostini, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **362**, 487 (1995).

[73] ArgoNeuT, R. Acciarri *et al.*, Phys.Rev. **D89**, 112003 (2014), 1404.4809.