

Measurement of the charged current inclusive muon neutrino interaction cross-section on lead using the T2K ND280 electromagnetic calorimeters

Dominic Brailsford

Imperial College London
High Energy Physics

A thesis submitted to Imperial College London
for the degree of Doctor of Philosophy

Abstract

This thesis presents an analysis of neutrino interactions using the Electromagnetic Calorimeters (ECals) of the Tokai-to-Kamioka (T2K) off-axis near detector (ND280) as a target using data collected during T2K run 3C.

The analysis presented shows the development of a new set of reconstruction algorithms which are able to reconstruct multiple tracks which originate from the same neutrino interaction. The output of this reconstruction is used as the basis of a ν_μ charged current inclusive selection in the ND280 ECals. The selected events are then used in a simple χ^2 fit to extract a T2K flux-averaged ν_μ charged current inclusive cross-section on lead, which is measured as $\langle\sigma_{\text{Pb}}^{\text{CC}}\rangle_\phi = 8.13^{+1.33}_{-1.26} \times 10^{-39} \text{ cm}^2 \text{ nucleon}^{-1}$.

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.

The work presented in this thesis was made possible due to a large number of collaborators involved with and beyond the T2K experiment. Citations have been used where possible to reference work not done by the author. When this is not possible, the people responsible have been explicitly named.

The main contributions by the author to this thesis have been the development of a new set of reconstruction algorithms for the ECal which allow neutrino interactions to be studied, development of the ν_μ charged current inclusive selection in the ECal, evaluation of the ECal detector systematic uncertainties and implementation of the χ^2 fit (the original idea for this fit was by M. Scott).

The author has also made many contributions to the T2K experiment which have not been discussed in this thesis. These contributions largely involved package management of the ECal reconstruction algorithms (ecalRecon) and the final analysis file-producing software (oaAnalysis).

Acknowledgements

I would like to thank Dr. Asher Kaboth and Dr. Mark Scott for their enormous help during my time on the T2K experiment. You two have provided excellent guidance and patiently answered a great many questions during my time as a student and this analysis would not have progressed without your help.

I would also like to thank my supervisor, Dr. Yoshi Uchida. Your gentle nudges have aimed me in the correct direction numerous times. For the horrendous task of reading this thesis in its entirety, I am eternally grateful.

Thanks also go to Dr. Morgan Wascko. The many times I had to quickly come to your office to discuss an analysis problem have been such a huge help to me.

I would also like to thank Dr Per. Jonsson for the advice and many thought provoking discussions during evaluation of the ECal systematic uncertainties.

I also want to give many thanks to the coffee break group, Dr. Ben Smith, Dr. Peter Sinclair, Pip Hamilton, Luke Pickering and Wing Ma. Discussion of my analysis during this time has been a huge help.

Kelly, thank you so much for putting up with me, even more so when writing my thesis. I know I can't have been the easiest person to deal with a lot of the time, but you remaining by my side has kept me sane.

Finally, the largest thanks go to my parents. I can never repay you for all you have done for me. No matter how far I go or what I achieve in life, it is because of both of you.

Contents

List of Figures	ix
List of Tables	xxix
1 Introduction	1
1.1 The state of the field	2
1.2 The future	4
2 Neutrino interactions with atomic nuclei	7
2.1 Neutrino interactions at the GeV-scale	7
2.2 Neutrino interactions with heavy nuclei	11
3 The T2K Experiment	16
3.1 T2K beam	17
3.1.1 Accelerator complex	18
3.1.2 Neutrino beamline	18
3.1.3 Off-axis beam	20
3.2 Near detector complex	21
3.2.1 Multi-Pixel Photon Counter	21
3.2.2 INGRID	22
3.2.3 ND280	23
3.2.4 The far detector	34
4 ND280 software and existing ECal event reconstruction	36
4.1 Monte Carlo production software	37
4.1.1 Neutrino flux simulation	37
4.1.2 Neutrino interaction simulation	38
4.1.3 ND280 detector simulation	39
4.1.4 Detector response simulation	39
4.2 Real data processing software	40

4.3	Main software chain	40
4.3.1	Detector calibration	40
4.3.2	Event reconstruction	46
4.3.3	Data reduction and summarising	47
4.4	ECal event reconstruction	47
4.4.1	Hit preparation	47
4.4.2	Basic clustering	48
4.4.3	Cluster combination	48
4.4.4	Cluster expansion	49
4.4.5	3D cluster formation	50
4.4.6	3D hit reconstruction	51
4.4.7	Energy reconstruction	51
4.4.8	Event classification	52
5	Enhanced ECal reconstruction	53
5.1	The Hough transform	53
5.1.1	Line-point duality	54
5.1.2	The parameter space	56
5.1.3	Redefinition of parameters	57
5.1.4	Discretisation of the parameter space	58
5.2	ECal application of the Hough transform	59
5.2.1	Modelling the ECal bar	60
5.2.2	Parameter space generation	63
5.2.3	Parameter space analysis	64
5.2.4	2D track quality checks	66
5.2.5	3D track reconstruction	66
5.2.6	Track splitting	70
5.2.7	Track pairwise crossing reconstruction	73
5.3	Output of the reconstruction	73
5.4	Validation of the reconstruction	73
5.4.1	Validation of algorithm performance	74
5.4.2	Data-motivated validation of the algorithms	75
6	Magnetic field simulation in ND280	79
6.1	Magnetic field model in the ND280 flux return	80
6.2	Effect of magnetic field on the ECal	84

7 Selection of neutrino interactions in the ECal	87
7.1 Signal definition	89
7.2 Monte Carlo sample	89
7.3 Background particles	90
7.4 Vertex reconstruction	92
7.4.1 Reconstructed track rejection	95
7.4.2 The vertex reconstruction algorithm	96
7.5 Track merging	101
7.6 Monte Carlo selection	108
7.6.1 Selection cuts	110
7.6.2 Performance of the selection	125
8 Measurement of the ν_μ charged current inclusive cross-section on Pb	140
8.1 Measurement method	140
8.2 Input samples to the fit	141
8.3 The ECal rate fit	143
8.4 Systematic uncertainties	145
8.4.1 Flux systematic evaluation	146
8.4.2 Cross-section systematic evaluation	150
8.4.3 ECal detector systematic evaluation	158
8.4.4 FGD detector uncertainty evaluation	179
8.4.5 The measured cross-section uncertainty	184
8.5 Validation of method	185
8.5.1 Validation of the fit machinery	185
8.5.2 Physics validation of the fitter	190
8.6 Applying the fit to ND280 data	195
8.7 The ν_μ charged current inclusive cross-section on lead	208
9 Discussion and conclusions	210
9.1 Discussion	210
9.2 Conclusions	211
9.3 The future	212
Bibliography	215

List of Figures

1.1	The pattern of neutrino masses for the normal and inverted hierarchies with the atmospheric (Δm_{atm}^2) and solar (Δm_{sol}^2) mass splittings labelled. The flavour composition of the mass eigenstates as a function of the unknown CP phase (labelled δ_{CP}) is also shown [26].	5
2.1	Quasi-Elastic (QE) interactions of a ν_μ with a nucleon. The small ellipse represents the neutrino interacting with the nucleon as a whole, rather than with an individual parton.	8
2.2	A Charged Current RESonant pion (CCRES) interaction of a ν_μ with a neutron. The Δ resonance decays to a neutron and a π^+	9
2.3	A Charged Current Deep Inelastic Scattering (CCDIS) interaction of a ν_μ with a neutron. X represents the leftover nuclear remnant.	10
2.4	ν_μ CC cross-section measurements per nucleon and divided by neutrino energy for a range of energies, showing the QE, RES and DIS contributions [37]. The data points are provided by a range of experiments (from 1979 to 2010) including BEBC (1979) [38], NuTeV (2006) [39], MINOS (2010) [40] and others. The example predictions are provided by the NUANCE generator [41]. Many measurements have a large associated uncertainty, particularly in the lower energy (less than 1 GeV) regime. The heaviest target nucleus probed in the region of interest, near 1 GeV, is carbon.	11

2.5	The CCQE cross-sections measured by the MiniBooNE (2010) [43], LSND (2002) [45] and NOMAD (2009) [46] experiments. The solid and dashed lines represent predictions from the NUANCE generator with different values of M_A [43]. Each prediction does not well model all of the data shown in the figure, indicating tension between the data collected by each of the experiments.	12
2.6	Sketch of the potential wells for protons and neutrons in the Relativistic Fermi-Gas (RFG) model. The difference between the Fermi level and the potential well edge, E_b , is approximately constant for different nuclei. The energy between the proton and neutron ground states and their maximum filled state, E_F^p and E_F^n respectively, are also shown. Due to Coulomb repulsion, the proton potential well is shallower than the neutron potential well.	13
2.7	Measured values of the ν_μ CC relative cross-section for several elements, as measured by the CHORUS experiment (2003) [48]. The black and white points are the collected data and prediction respectively. The solid and dashed lines are the linear best fit lines to the data and prediction respectively. Going from left to right, the points represent data and prediction for lead, iron marble and polyethylene. The prediction is taken from a set of quark distribution functions [49] provided by PDFLIB. 14	
2.8	Ratio of the measured ν_μ CC inclusive cross-section on lead to plastic scintillator as a function of neutrino energy, as measured by the MINER ν A experiment (2014) [50]. The simulation is based on the GENIE generator [51]. The error bars on the simulation (data) are statistical (systematic) uncertainties.	15
3.1	Schematic of the T2K experiment showing the near and far sites, separated by the 295 km baseline.	16
3.2	The 68% and 90% confidence level allowed regions for $\sin^2 2\theta_{13}$ as a function of δ_{CP} for normal hierarchy (top) and inverted hierarchy (bottom). The solid line represents the best fit $\sin^2 2\theta_{13}$ for a given δ_{CP} . The shaded region shows the average θ_{13} provided by the reactor constraint [22]. .	17
3.3	A schematic of the T2K neutrino beamline (left) and a side view of the secondary beamline (right) [52].	18

3.4	The POT recorded by CT5 as a function of time (blue line) and the recorded beam power in ν running mode (red dots) and $\bar{\nu}$ running mode (purple dots). The recorded POT as a function of time shows T2K has been successful in greatly increasing the size of its dataset during each data collection run.	20
3.5	Muon neutrino survival probability at 295 km (top) and neutrino fluxes for different off-axis angles (bottom) [55]. The peak neutrino lowers and the energy distribution narrows as the off-axis angle increases. By carefully selecting the off-axis angle (2.5°), the peak energy of the neutrino beam can be aligned with the first maximum of the ν_μ disappearance probability spectrum.	21
3.6	A schematic of INGRID, showing the horizontal, vertical and off-axis modules [52].	22
3.7	An exploded view of ND280 [52], showing the locations of its subdetectors.	23
3.8	The π^0 reconstruction efficiency in initial ECal simulation as a function of the thickness of the lead layers. The blue and black lines show π^0 events which occurred in FGD1 and FGD2 respectively. The thicker lead layers lead to a large energy absorption while the thin layers lead to poor shower containment. Both extremes lead to a poorer π^0 reconstruction efficiency. A bar thickness of 1 cm was used in all of the ECal simulations shown [63].	29
3.9	The π^0 reconstruction efficiency in initial ECal simulation as a function of the scintillator bar width. The blue and black lines show π^0 events which occurred in FGD1 and FGD2 respectively. The reconstruction efficiency is very compromised when the bars are allowed to become wide. A bar thickness of 1 cm was used in all of the ECal simulations shown [63].	30
3.10	A DS ECal layer under construction. The two sheets of lead have been placed on top of the layer of scintillator bars. The aluminium frame and locator pins, which secure the scintillator bars, are visible [60].	31

3.11 An example attenuation profile of a DS ECal scintillator bar, formed using the 2D scanner. Each plot refers to the readout from a MPPC, connected to either end of the scintillator bar. The x -axis of both plots is the position of the 2D scanner along the bar, recorded in cm. The y -axis of both plots is the ratio of the integrated light yield recorded by the MPPC with the 2D scanner source present, to the integrated light yield without the source. The points are the recorded data and the solid curves are exponential fits to the central portions of the data [60].	32
3.12 Example of reconstructed T2K events in SK for a muon-like ring (left) and an electron-like ring (right) [52]. Because of its MIP-like nature, the muon Cherenkov ring appears sharper than the electron ring.	35
4.1 Example pedestals in the high (top) and low (bottom) channels for an MPPC. The photo-electron peaks are clearly visible in the high gain channel. The pedestal value is calculated by fitting a Gaussian function to a 9.9 ADC count range, centered on the mode of the gain distribution [76].	41
4.2 The effect of the pedestal subtraction on the charge readout by MPPC channels in the DS ECal. The top distribution shows the raw ADC collected by the MPPC channels while the bottom distribution shows the same ADC readout after the pedestal subtraction has been applied [76].	42
4.3 Example of the method used to calculate the gains for an MPPC channel. The shaded regions, defined by calculating the mode and trough positions, are the ranges used to the Gaussian fits to the pedestal and first photo-electron peak [76].	43
4.4 Charge as a function of time for three identical capacitor with different final charging values [76]. The black dashed line represents the threshold imposed by a Trip-T discriminator. A capacitor which has a higher final charge value passes the threshold faster than a capacitor with a lower final charge value.	45

4.5 The effect of the timewalk correction on MPPC charges and timestamps in DS ECal data [76]. The left and right distributions show the data before and after the fibre timewalk correction has been applied. There is a distinct flattening of the charge-vs-timestamp dependence after application of the correction. All data shown are recorded during Trip-T cosmic triggers.	46
4.6 The track-like/shower-like discriminator for the DS ECal which is uses an Artificial Neural Network (ANN) [60]. Solid lines show the control sample Monte Carlo (based on NEUT simulation of beam events) and the points show the control sample data (taken from collected beam data). For the muon samples, exactly one track is required in each of the TPCs and the DS ECal with an additional requirement that the TPC tracks are muon-like. For the electron samples, photon pair production is checked in the FGD, requiring tracks with opposite charge with at least one TPC track which must be electron-like. The muon and electron simulated distributions show good agreement with the collected data.	52
5.1 Representations of a 2D line in Cartesian space.	54
5.2 Representations of a 2D point in Cartesian space.	55
5.3 The three points defined in equation 5.4 and their representation in the parameter space. The colour coding matches the Cartesian points to their respective parameter lines.	57
5.4 The line represented by the parameter space intersection in Fig. 5.3b with the Cartesian points it intercepts.	58
5.5 θ - ρ parameterisation of a 2D line.	59
5.6 The discrete θ - ρ space. The plotted lines are those defined in equation 5.5 and reparameterised using equation 5.8.	60
5.7 Simulated neutrino interaction with 3 final states in the side-right ECal. The green line entering from the left is the ν_μ . The green line exiting to the right is a μ^- , the brown line is a π^+ and the blue line is a proton. The purple rectangles represent the hit ECal bars.	61

5.8	The grid representation of an ECal bar and its representation in parameter space.	62
5.9	The cross representation of an ECal bar and its representation in parameter space.	63
5.10	The full parameter space of the 2D cluster shown in Fig. 5.7. The height of each point in the parameter space corresponds to the number of hit scintillator bars intersected by the Cartesian line which the parameter space point represents.	64
5.11	The reduced parameter space of the 2D cluster shown in Fig. 5.7. The reduced parameter space was formed by removing all of the hit scintillator bar representations which contributed to the full parameter space's (see Fig. 5.10) maximum.	65
5.12	Q_{ratio} and its probability density distribution in the barrel ECal for the two track case. NEUT-based simulation of ND280 beam events were used to produce the distributions.	67
5.13	$\Delta_{\text{layer, first}}$ and its probability density distribution in the barrel ECal for the two track case. NEUT-based simulation of ND280 beam events were used to produce the distributions.	68
5.14	$\Delta_{\text{layer, last}}$ and its probability density distribution in the barrel ECal for the two track case. NEUT-based simulation of ND280 beam events were used to produce the distributions.	69
5.15	Event display (XY view) of a problematic (for the reconstruction) neutrino interaction in the ECal. The solid green track is the muon, the solid blue tracks are protons and the pink tracks are neutrons. The neutrino is the short dashed green line. The proton travelling vertically downwards and the muon are created almost back to back leaving a line of hit scintillator bars. The 2D hough transform would register this as a single straight line.	71

5.16	The angular resolution as a function of trajectory length in the DS ECal when applying the enhanced reconstruction to muon MC. The colour coding refers to the true entry angle range of the muons. The figure was generated using single muon particle gun fired into the front face of the DS ECal with a controlled range of entry angles. The angular resolution distributions were created by L. Pickering.	74
5.17	The summed charge contained on all tracks reconstructed in each ECal cluster. The red histograms and black points are Monte Carlo and data respectively.	76
5.18	The summed number of hits contained on all tracks reconstructed in each ECal cluster. The red histograms and black points are Monte Carlo and data respectively.	77
5.19	The number of DS ECal events in the bskmu sample after relaxing the 2D track quality check. The red histograms and black points are the Monte Carlo and data respectively. The track quality fix has corrected the data excesses in both distributions, leaving only the offset between the data and Monte Carlo peaks.	78
6.1	Graphical display of the X-Y ND280 cross-section with the flux return, magnetic field coils, ECals and the Basket (Tracker) labelled.	80
6.2	Simple model of the magnetic field (blue lines) in the Tracker (orange) and flux return (red) regions. The flux return yoke has been separated into ten sections: four vertical B field sections, four corner B field sections and two horizontal B field sections. The height of the Tracker region, h , is also shown.	81
6.3	A horizontal section of the flux return with its corresponding magnetic field lines (blue). The corner sections, vertical sections and the Tracker are also visible. This figure is a zoomed in cut-out of Fig. 6.2.	82
6.4	A vertical section of the flux return with its corresponding magnetic field lines (blue). In this vertical section, the magnetic field has exited the Tracker and is travelling vertically upwards through the flux return. A corner section, another vertical section and the Tracker are also visible. This figure is a zoomed in cut-out of Fig. 6.2.	83

6.5	A corner section of the flux return with its corresponding magnetic field lines (blue). In this corner section, the magnetic field has entered from a vertical section in which the field was travelling vertically upwards, away from the Tracker. A vertical section, a horizontal section and the Tracker are also visible. This figure is a zoomed in cut-out of Fig. 6.2.	83
6.6	Comparisons of data and Monte Carlo in the bottom-left barrel ECal for previous software productions. The red and pink histograms are Monte Carlo simulation of T2K beam neutrinos incident on ND280 and the surrounding pit respectively. The blue data points are collected data from run 3C. Both of the plots are POT normalized. The data excesses shown are especially significant because they occur in the highest population regions.	84
6.7	Comparisons of data and Monte Carlo in the bottom-left barrel ECal with the magnetic field model described in section 6.1 implemented. The red and pink histograms are Monte Carlo simulation of T2K beam neutrinos incident on ND280 and the surrounding pit respectively. The blue data points are collected data from run 3C. Both of the plots are POT normalized. Comparing the shown distributions with those in Fig. 6.6, the effect of the simple magnetic field model is clearly visible.	85
7.1	The average energy of neutrino interactions seen in the ECal against the number of reconstructed tracks in the ECal that the neutrino final states created. The uncertainty on the neutrino energy is calculated as the standard error on mean of the neutrino energy. There is a clear trend between the two variables.	88
7.2	The number of signal events in the NEUT-based beam Monte Carlo sample as a function of neutrino energy in the bottom-left (red), top-right (blue) and DS (black) ECals. The bottom-left ECal sees the most signal events and at a higher neutrino energy which is caused by the off-axis effect.	91
7.3	The number of background particles, separated out into particles species, as a function of their momentum at point of entry into the ECal.	92

7.4	The entry momentum of particles entering the barrel ECal Vs. the point of creation of the particles. For all ranges of entance momentum, the dominant point of provenance for all particles is from the magnet and pit regions.	93
7.5	The entry momentum of particles entering the DS ECal Vs. the point of creation of the particles. Unlike the barrel ECal case(Fig. 7.4), the creation point of the particles is more varied and includes significant contributions from the downstream end of the tracker (FGD 2 and TPC 3).	94
7.6	Angular separation of the reconstructed tracks in an ECal cluster with the beam-simulated particle that created them. The black, red and green histograms are the tracks that were reconstructed first, second and third respectively. All histograms are area normalised.	96
7.7	Values of ϕ^{vertices} in (d_q, d_c) space. The colour corresponds to the magnitude of ϕ^{vertices} . The distribution plateaus rather than peaks, suggesting that there are no preferred values of d_q and d_c , only disfavoured ones.	98
7.8	ϕ^{vertices} vs the marginalised vertex reconstruction parameters. The same plateau effect shown in Fig. 7.7 can be seen here. The ideal values for both reconstruction parameters are the ones which get as close to the sharp drop in ϕ^{vertices} as possible.	98
7.9	Example of the vertex position, \vec{P} for three lines: \vec{l}_1 , \vec{l}_2 and \vec{l}_3 . $ \vec{r}_1 $, $ \vec{r}_2 $ and $ \vec{r}_3 $ are the perpendicular distances of \vec{l}_1 , \vec{l}_2 and \vec{l}_3 to \vec{P} respectively.	99
7.10	Example event display showing a set of track merging candidates. The final state μ^- (in green) curves in a complicated way and is reconstructed as four tracks (in light blue). The pairwise crossings of the reconstructed tracks are the red squares. The muon is produced by a charged current interaction in the top left barrel ECal producing two π^+ (in brown) and the curving μ^-	101
7.11	The cosine of the angle subtended by the merging candidates, bounded between 0 and 1. The correct matches and incorrect matches are the blue and red histograms respectively. There is a clear build-up of events which are correctly matched as $\cos \theta \rightarrow 1$	103

7.12 Example event display showing how the parameters of the distance ratio are calculated. The distance ratio is defined as $d^{\text{small}} / d^{\text{large}}$. The blue lines are the two reconstructed tracks which form the merging candidate and the red square is the crossing location of those tracks.	103
7.13 The distance ratio of the merging candidates. The correct and incorrect matches are the blue and red histograms respectively. When no cuts are applied, the distance ratio shows little separation between the correctly matched and incorrectly matched events. However, after demanding $\cos \theta > 0.8$, the incorrectly matched peak significantly flattens without altering the correctly matched peak.	104
7.14 The distance ratio vs the cosine of the opening angle for track merging candidates. This 2D space significantly separates out the correctly matched and incorrectly matches events.	105
7.15 Values of the tuning metric, as described in equation 7.10, in distance ratio cut vs $\cos \theta$ cut space. The peak of this distribution is very broad, suggesting a range of preferred values for the distance ratio and $\cos \theta$ cuts.	105
7.16 ϕ^{merge} vs the track merging conditions. There is a clear but somewhat broad maximum of ϕ^{merge} in $\cos \theta$ space whereas there is only a plateau in distance ratio space.	106
7.17 Schematic showing the degeneracy of two merging candidate topologies which would pass the $\cos \theta$ cut. The arrows represent reconstructed tracks and θ is the opening angle measured between those tracks.	107
7.18 Example event display showing how the inputs to the swing parameter are calculated. Swing is defined as $l^{\text{long}} / d^{\text{large}}$. The blue lines are the two reconstructed tracks which form the merging candidate and the red square is the crossing location of those tracks.	107
7.19 The number of merging candidates as a function of the swing parameter. The correct matches and incorrect matches are the blue and red histograms respectively. After applying the $\cos \theta$ and distance ratio cuts, the swing parameter distribution becomes bimodal, signifying that the swing parameter has separated out the two merging candidate topologies shown in Fig. 7.17.	108

7.20 The number of reconstructed events in the Monte Carlo sample, separated into the prong topologies. Each event is categorised by the associated truth information from the simulation. The number of reconstructed events and the background contamination strongly depends on the number of reconstructed prongs. This effect indicates that selection development should be tailored towards each prong topology individually.	110
7.21 $\phi_1^{\text{selection}}$ as a function of the distance from the $-y$ face (bottom face) of the bottom-left barrel ECal for the 1 prong topology. There is a clear maximum of $\phi_1^{\text{selection}}$ which indicates that there is a preferred minimum distance of 1 prong vertices from the $-y$ face.	111
7.22 Total prong charge vs the number of prong hits for barrel ECal events in the 1 prong topology.	115
7.23 Total prong charge vs the number of prong hits for barrel ECal events in the 2 prong topology.	116
7.24 Total prong charge vs the number of prong hits for barrel ECal events in the 3 prong topology.	117
7.25 ΔN vs the number of prong hits for barrel ECal events in the 1 prong topology.	119
7.26 ΔN vs the number of prong hits for barrel ECal events in the 2 prong topology.	120
7.27 ΔN vs the number of prong hits for barrel ECal events in the 3 prong topology.	121
7.28 The opening angle subtended by the constituents prongs for the 2 prong topology. Signal and background events are the blue and red histograms respectively.	122
7.29 $\phi_2^{\text{selection}}$ as a function of the opening angle cut for the 2 prong topology in the barrel ECal. There is a clear maximum of $\phi_2^{\text{selection}}$, indicating a preferred cut on the opening angle.	123

7.30	The summed opening angle subtended by every pairwise combination of the constituents prongs for the 3 prong topology. Signal and background events are the blue and red histograms respectively.	124
7.31	Event survival as a function of the cuts for all topologies in the barrel ECal. Each bin refers to the number of events after a cut has been applied. The rejection of background events by each cut can be clearly seen.	126
7.32	Event survival as a function of the cuts for all topologies in the DS ECal. Each bin refers to the number of events after a cut has been applied. The rejection of background events by each cut can be clearly seen.	127
7.33	ECal OOFV event survival as a function of the cuts for all topologies in the barrel ECal. Each bin refers to the number of events after a cut has been applied. It is clear that each cut rejects a specific set of particle types. For example, the FV cuts strongly reject the track-like particles whereas the visible energy and ΔN cuts reject shower-like particles.	129
7.34	ECal OOFV event survival as a function of the cuts for all topologies in the DS ECal. Each bin refers to the number of events after a cut has been applied. The cuts shown behave in a similar manner to those in the barrel ECal (see Fig. 7.33).	130
7.35	The selection efficiency (black) and purity (red) as a function of the cuts in the barrel ECal. Each bin refers to the efficiency and purity after a cut has been applied. The cuts generally degrade the selection efficiency and increase the selection purity. It is clear that strength of all cuts are not equal.	131
7.36	The selection efficiency (black) and purity (red) as a function of the cuts in the DS ECal. Each bin refers to the efficiency and purity after a cut has been applied. The distributions suggest the cuts behave in a similar manner to those in the barrel ECal (see Fig. 7.35).	132
7.37	$\phi^{\text{selection}}$ as a function of the cuts in the barrel ECal. Each bin refers to the value of $\phi^{\text{selection}}$ after a cut has been applied. Generally, each cut increases $\phi^{\text{selection}}$ with the exception of the 4+ prong topology. However, it should be noted that the selection was not tuned for this topology.	133

-
- 7.38 $\phi^{\text{selection}}$ as a function of the cuts in the DS ECal. Each bin refers to the value of $\phi^{\text{selection}}$ after a cut has been applied. Unlike the barrel ECal (see Fig. 7.37), the cuts do have a positive effect on the 4+ prong topology. 134
- 7.39 The selection efficiency (black) and purity (red) for the combined prong topologies as a function of the selection cuts. After combination of all topologies, each cut has an expected effect; an increase in sample purity at the expense of selection efficiency. 135
- 7.40 $\phi^{\text{selection}}$ for the combined prong topologies as a function of the selection cuts. Generally speaking, $\phi^{\text{selection}}$ increases as each cut is applied to either the barrel ECal or DS ECal samples. The only exception to this is when the entering background cut is applied in the barrel ECal which slightly decreases $\phi^{\text{selection}}$ 136
- 7.41 The selection efficiency (black) and purity (red) for the combined prong topologies as a function of neutrino energy. The selection efficiency is largely flat for both the barrel ECal and DS ECal showing the selections are not biased towards any particular energy range. This is not the case for the selection purities but this can be explained by lower energy backgrounds having less penetrating power, resulting in a higher purity for lower neutrino energies. 136
- 7.42 The number of selected events in the Monte Carlo sample, separated out into the prong topologies. Each event is categorised by the associated truth information from the simulation. The effect of the developed selection can be clearly seen in all prong topologies; each topology is now signal dominated. 137
- 8.1 The neutrino energy of events seen in the bottom-left ECal module broken down by the neutrino interaction mode. The plots correspond to 3.949×10^{20} POT of simulated data. The reverse sample contains a large population of high energy (5 GeV) events which is not seen in the selected sample. This can be explained by the higher energy events creating a surplus of reconstructed clusters which are cut away in the selection, causing such clusters to fall into the reverse sample. 142

8.2	The number of reconstructed prongs per vertex for events seen in the bottom-left ECal module broken down by the neutrino interaction mode. The plots correspond to 3.949×10^{20} POT of simulated data. The higher energy events seen in Fig. 8.1b (which are mostly DIS events) are spread evenly between each prong topology in the reverse sample.	143
8.3	The number of events in each input sample separated into the lead (black), carbon (red) and other (blue) templates. "ECal" has been omitted in all but the DS ECal and FGD sample names. Not only does the event rate vary significantly between each sample, but the composition of those events also significantly varies.	144
8.4	The flux prediction fractional covariance matrix. The matrix is binned primarily in detectors (ND280 and SK). There is then a subsequent binning scheme for each detector in neutrino flavour (ν_μ , $\bar{\nu}_\mu$, ν_e and $\bar{\nu}_e$) and neutrino energy.	147
8.5	The sample fractional covariance induced by the neutrino flux uncertainties. The same binning for sample-based covariance matrix elements is used in subsequent plots in this chapter.	149
8.6	The variation in the selection efficiency of ν_μ CC interactions on lead in the DS ECal caused by the neutrino flux systematic. The width of the distribution is 0.18% of the size of the nominal efficiency value (0.539).	150
8.7	The neutrino cross-section model fractional covariance matrix binned in values of the NEUT cross-section parameters.	151
8.8	The sample fractional covariance induced by the cross-section model uncertainties.	152
8.9	The variation in the selection efficiency of ν_μ CC interactions on lead in the DS ECal caused by the cross-section model systematic. The width of the distributions is 0.2 of the size of the nominal efficiency (0.539).	153

8.10 The ratio of the $\pm \sigma$ cross-section to the nominal cross-section for lead as a function of neutrino energy. Only variation in P_F^{Pb} shows a significant deviation of the cross-section from its nominal value and this only occurs as the cross-section approaches the sub-GeV energy range. The inputs to the distributions were created by simulating ν_μ CC interactions on lead with a uniform energy using NEUT. Events and distributions produced by W. Ma.	154
8.11 The sample fractional covariance induced by variation of P_F^{Pb}	155
8.12 The sample fractional covariance induced by variation of of the FSI parameters outlined in table 8.1.	157
8.13 The layer-by-layer efficiency for the ND280 ECals, measured using cosmic data [60]. The efficiencies were calculated by studying the rate of hits in each bar and the rate of hits in the bar's nearest neighbours in the surrounding layers.	159
8.14 The charge distribution of hit scintillator bars created by MC cosmic muons. The distribution is truncated at 5.5 PEU as that is the minimum charge threshold for hits to be reconstructed in the ECal reconstruction. The peak of the charge distribution is located near to the 5.5 PEU threshold.	160
8.15 The sample fractional covariance induced by the ECal hit inefficiency.	161
8.16 The charge of ECal hits after applying the enhanced reconstruction to run 3C cosmic events. The distributions only show information for reconstructed objects in the top-left barrel ECal module. The blue and black histograms are Monte Carlo and data respectively. All distributions are area normalised.	162
8.17 The sample fractional covariance induced by the ECal charge uncertainties.	165
8.18 The estimation of the number of noise hits (solid blue line) and the number of true noise hits (solid red line) per event in Monte Carlo. The coloured dashed lines are the Gaussian fits to the respective distributions. The similarity of the estimated and true noise distributions shows that the estimator is accurate.	166

8.19 The estimation of the number of noise hits in Monte Carlo (solid blue line) and data (solid black line) per event. The coloured dashed lines are the Gaussian fits to the respective distributions.	166
8.20 The sample fractional covariance induced by the ECal noise uncertainties.	168
8.21 The fractional covariance matrix for the mass of the contributing elements in the DS ECal, with all contributing elements considered.	172
8.22 The total mass of the DS ECal active volume with (black) and without (blue) nitrogen and hydrogen as contributing elements. The smooth lines are the Gaussian fits to the distributions. Using the Gaussian mean and width as the mass central value and uncertainty respectively, the DS ECal active mass with (without) hydrogen and nitrogen is 4060 ± 20 kg (3960 ± 20 kg).	172
8.23 The fractional covariance matrices for the mass of each contributing element to an ECal module (hydrogen and nitrogen omitted).	173
8.24 The sample fractional covariance induced by the ECal mass uncertainties.	175
8.25 The sample fractional covariance induced by the uncertainty in the ECal OOFV events.	177
8.26 The sample fractional covariance induced by the bug fix to the reconstruction.	178
8.27 The number of selected FGD events for each systematic throw using PSYCHE. The smooth line is a Gaussian fit to the number of events.	179
8.28 The sample fractional covariance induced by all of the detector-based uncertainties.	181
8.29 The sample fractional covariance induced by all of the systematic uncertainties.	182
8.30 The number of selected events in the GENIE-based Monte Carlo sample, separated out into the prong topologies. Each event is categorised by the associated truth information from the simulation.	183

8.31 Distributions of R^{Pb} related information returned from the fitter for the 10,000 fake data throws. The red dashed lines are Gaussian fits to the distributions. The distributions show that the fitter is correctly returning the value of R^{Pb} and is unbiased.	187
8.32 Performance of the fit for the 10,000 fake data sets when using normalisation set 1. For all parameters, the fitter returns the correct values and is unbiased.	187
8.33 Performance of the fit for the 10,000 fake data sets when using normalisation set 2. For all parameters, the fitter returns the correct values and is unbiased.	188
8.34 Performance of the fit for the 10,000 fake data sets when using normalisation set 3. For all parameters, the fitter returns the correct values and is unbiased.	188
8.35 Performance of the fit for the 10,000 fake data sets when using normalisation set 4. Despite correct errors being returned, the fitter is returning slightly biased values of the fit parameters. It is important to note that normalisation set 4 completely removes the carbon template and even in this extreme situation, the bias on the fitted parameters is small. . .	189
8.36 Performance of the fit for the 10,000 fake data sets when using normalisation set 5.	189
8.37 Performance of the fit with the reverse ECal samples removed for the 10,000 fake data sets when using normalisation set 1.	190
8.38 The pre-fit number of NEUT Monte Carlo events compared with GENIE fake data. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the GENIE fake data. In all of the ECal selected bins (the first 7 bins) there is a systematic deficit of GENIE events. This effect is not seen in the reverse samples, because the reverse samples contain a significant fraction of sand muons which are taken from the NEUT Monte Carlo.	191

8.44 The shape-only correlation matrix for the samples used in the fit. There is a large degree of (anti-)correlation between all samples. This gives the fitter less freedom to change the normalisations of the MC samples.	200
8.45 An area-normalised comparison of a sub-set of the MC templates with data for the barrel ECal samples. The coloured histograms and data points are the MC Templates and data respectively. The red, blue and green histograms are the lead, carbon and <i>other</i> templates respectively. There is significant disagreement in the shape of the <i>other</i> template with the data.	200
8.46 An area-normalised comparison of a sub-set of the MC templates with data for the barrel-reverse ECal samples. The coloured histograms and data points are the MC Templates and data respectively. The red, blue and green histograms are the lead, carbon and <i>other</i> templates respectively. Similarly to Fig. 8.45, there is disagreement in the shape of the <i>other</i> templates with the data.	201
8.47 The number of events in the <i>other</i> template as a function of true neutrino energy for the barrel ECal samples. The distributions are broken down by true neutrino interaction location. For all distributions, the dominating location is either the magnet or the sand in the surrounding pit.	202
8.48 The number of events in the <i>other</i> template as a function of true neutrino energy for the barrel reverse ECal samples. The distributions are broken down by true neutrino interaction location. For all distributions, the dominating location is either the magnet or the sand in the surrounding pit.	203
8.49 The number of events which pass the fiducial volume cut as a function of the number of hits in the reconstructed cluster. The red and blue histograms show the number of beam MC events and sand MC events respectively. The black points show the run 3C data. In all distributions there is an excess of low number of hit clusters in data. For the ECal modules closest to the beam (bottom-left, bottom-right and side-left) there is a significant MC excess at approximately 31 hits which aligns with the peak of the sand MC events.	205

- 8.50 The post-fit number of NEUT Monte Carlo events compared with T2K data when including the ad hoc uncertainty shown in table 8.16. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic and ad hoc uncertainties on the number of NEUT MC events. The black points are the T2K data. The inflated errors has reduced the level of correlations, allowing the fitter to make finer changes to the MC normalisation. 207
- 8.51 The simulated flux for T2K run 3C as a function of neutrino energy. . . 209

List of Tables

1.1	The 2014 best-fit values of the 3-neutrino oscillation parameters provided by the Particle Data Group [12]. $\Delta m^2 \equiv m_3^2 - (m_2^2 - m_1^2) / 2$. The values (values in brackets) correspond to $m_1 < m_2 < m_3$ ($m_3 < m_1 < m_2$). The values come from a global fit to solar, atmospheric, reactor and accelerator (both short and long baseline) experiments [13]. The most recent contributing data is a precision measurement of ν_μ disappearance from T2K (2014) [14].	3
3.1	Summary of the ECal design showing the overall dimensions, numbers of layers, length and orientation of the scintillator bars, numbers of bars, and lead thickness for each module [60].	26
5.1	The percentage number of correct matches in the 3D matching separated by which track matching likelihood was used. NEUT-based Monte Carlo simulation of the T2K beam was used to calculate the matching percentages.	75
7.1	The number of signal events in the NEUT-based beam Monte Carlo sample separated by ECal module.	90
7.2	Parameters for the vertex reconstruction in the ECal.	97
7.3	Cut values for the track merging in the ECal.	108
7.4	The fiducial volume definitions for each ECal module for the 1 prong topology. The strongest cuts are for the $\pm y$ ECal faces (barrel modules only) which rejects 1 prong vertices which occur in the inner or outer two layers. For the DS ECal, the $\pm z$ face cut is the strongest which rejects 1 prong vertices which occur in the inner or outer three layers. .	112

7.5	The fiducial volume definitions for each ECal module for the 2, 3 and 4+ prong topologies. The only strong cuts for the 2+ prong topologies are the $+z$ ECal face cuts (all modules). The power of the $+z$ face cuts is due to secondary interactions having a preference for occurring downstream of the neutrino interactions, leading to background-tagged, reconstructed vertices positioned towards the $+z$ of each ECal module.	113
7.6	The definitions of the 2 dimensional cut lines involving the total prong charge and the number of associated scintillator hits. N refers to the total number of scintillator hits associated to the constituent prongs in the vertex.	118
7.7	The definitions of the 2 dimensional cut lines involving ΔN and the number of scintillator hits associated with the constituent prongs. N refers to the total number of scintillator hits associated to the constituent prongs in the vertex.	122
7.8	The definitions of the opening angle cuts.	124
7.9	The selection efficiencies for each prong topology and ECal module.	138
7.10	The selection purities for each prong topology and ECal module.	138
7.11	The topology combined efficiency and purity.	139
8.1	FSI parameters, provided by the NIWG, representing the 1σ contour in the FSI parameter space [89].	156
8.2	Summary of the lead absorbers for the ND280 Tracker ECals [60].	163
8.3	Summary of the study to estimate the number of noise hits per event in the ECals.	164
8.4	Summary of the lead absorbers for the ND280 Tracker ECals [60].	169
8.5	Summary of the scintillator bars for the ND280 Tracker ECals [60].	170
8.6	Summary of the WLS fibres for the ND280 Tracker ECals [60].	171
8.7	Summary of the ECal OOFV systematic study when using the FGD cosmic control sample.	176

8.8	The selection efficiencies from the GENIE-based Monte Carlo sample for each prong topology and ECal module.	183
8.9	The topology combined efficiency for the GENIE-based Monte Carlo sample. The efficiency values include simulated events which were not reconstructed.	183
8.10	The normalisation sets used in the machinery validation of the ECal rate fit.	185
8.11	The GENIE to NEUT cross-section ratios, calculated from the truth information provided by the ND280 simulation.	192
8.12	The values returned from the fitter when passing the NEUT MC templates and GENIE fake data shown in Fig. 8.38.	192
8.13	The values returned from the fitter when passing the NEUT MC templates and GENIE fake data which contained a re-weighted FGD sample.	193
8.14	The number of selected events in each sample.	196
8.15	The values returned from the fitter when processing the NEUT MC templates with T2K data.	196
8.16	Comparison of the number of events between the barrel and barrel-reverse samples. The second and third column show the ratio of the number of events in the barrel samples to the number of events in the barrel-reverse samples for Monte Carlo and data respectively. The fourth column shows the percentage difference between the Monte Carlo and data ratios.	206
8.17	The values returned from the fitter modified with the ad hoc uncertainty when processing the NEUT MC templates with T2K data.	206

“These chickens jackin’ my style.”

— Fergie

Chapter 1

Introduction

The field of neutrino physics is currently evolving very rapidly. With its tenuous postulation [1] acting as a future omen, the neutrino's mark on history would not become apparent from its discovery [2–4], but rather from a spate of surprising discoveries at the end of the 20th century [5–7] which conclusively proved that the Standard Model, while very successful, was incomplete. This revelation was experimental proof of Maki, Nagakawa and Sakata's extension [8] to Pontecorvo's theory of neutrino oscillation [9] with the inclusion of the Mikheyev-Smirnov-Wolfenstein (MSW) effect which is the alteration of the oscillation effect due to differences in the coherent forward scattering of the three neutrino flavours with electrons in matter [10, 11]. The findings were ground breaking as the underlying theory requires neutrinos to be massive, which is in direct contradiction to the Standard Model. The, now, standard theory of neutrino oscillation defines three neutrino flavours and three neutrino masses. However, the map between flavour and mass is not one-to-one, but rather a rotation of mass space onto flavour space. The main consequence of this rotation is that the flavour eigenstates are a superposition of mass eigenstates, namely

$$|\nu_\alpha\rangle = \sum_{i=k}^3 U_{\alpha k}^* |\nu_k\rangle, \quad (1.1)$$

where $\alpha \in \{e, \mu, \tau\}$, ν_k are the neutrino mass eigenstates and $U_{\alpha k}^*$ is an element of a unitary rotation matrix which is known as the PMNS mixing matrix. As there are three mass and flavour eigenstates, the PMNS matrix is a 3×3 matrix and is often

parameterised as

$$U \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (1.2)$$

where $c_{ij} \equiv \cos \theta_{ij}$ and $s_{ij} \equiv \sin \theta_{ij}$. θ_{ij} are known as the mixing angles and parametrise how strong mixings between the flavour and mass eigenstates are, and δ is a CP violating phase. The most surprising observable feature of this mechanism is the non-zero probability to detect a neutrino of specific flavour which was created at source in a different flavour state. By propagating the mass eigenstates through time, one can arrive at this probability which has the following form:

$$P(\nu_\alpha \rightarrow \nu_\beta) = |\langle \nu_\beta | \nu(t) \rangle|^2 = |U_{\beta k} e^{-iE_k t} U_{\alpha k}^*|^2, \quad (1.3)$$

where $\nu(t)$ is the time-dependent neutrino mass eigenstate and E_k is the energy of the ν_k . For an accelerator-based neutrino oscillation experiment, the beam will be ν_μ -dominated. So, the ν_μ survival probability, $P(\nu_\mu \rightarrow \nu_\mu)$, and ν_e appearance probability, $P(\nu_\mu \rightarrow \nu_e)$, which are typically of interest, can be approximated in the following forms:

$$P(\nu_\mu \rightarrow \nu_\mu) \approx 1 - \cos^4 \theta_{13} \sin^2 2\theta_{23} \sin^2 \left(1.27 \frac{\Delta m_{23}^2}{[\text{eV}^2]} \frac{L}{[\text{km}]} \frac{[\text{GeV}]}{E} \right) \quad (1.4)$$

$$P(\nu_\mu \rightarrow \nu_e) \approx \sin^2 2\theta_{13} \sin^2 \theta_{23} \sin^2 \left(1.27 \frac{\Delta m_{23}^2}{[\text{eV}^2]} \frac{L}{[\text{km}]} \frac{[\text{GeV}]}{E} \right), \quad (1.5)$$

where $\Delta m_{ij}^2 \equiv m_i^2 - m_j^2$, L is the distance the neutrino propagates and E is the energy of the neutrino.

1.1 The state of the field

Data provided from a wide range of experiments show excellent agreement with the theory of neutrino oscillation and with a three-flavour neutrino picture. Global fits applied to the data provided by these experiments gives best fit values for the

Parameter	best-fit ($\pm 1\sigma$)
Δm_{12}^2 [10 $^{-5}$ eV 2]	7.54 $^{+0.26}_{-0.22}$
$ \Delta m^2 $ [10 $^{-3}$ eV 2]	2.43 ± 0.06 (2.36 ± 0.06)
$\sin^2 \theta_{12}$	0.308 ± 0.017
$\sin^2 \theta_{23}, \Delta m^2 > 0$	$0.437^{+0.033}_{-0.023}$
$\sin^2 \theta_{23}, \Delta m^2 < 0$	$0.455^{+0.039}_{-0.031}$
$\sin^2 \theta_{13}, \Delta m^2 > 0$	$0.0234^{+0.0020}_{-0.0019}$
$\sin^2 \theta_{13}, \Delta m^2 < 0$	$0.0240^{+0.019}_{-0.022}$
$\sin^2 \theta_{13}, \Delta m^2 < 0$	$0.0240^{+0.019}_{-0.022}$
δ/π (2 σ range quoted)	$1.39^{+0.38}_{-0.27}$ (1.31 $^{+0.29}_{-0.33}$)

Table 1.1: The 2014 best-fit values of the 3-neutrino oscillation parameters provided by the Particle Data Group [12]. $\Delta m^2 \equiv m_3^2 - (m_2^2 - m_1^2)/2$. The values (values in brackets) correspond to $m_1 < m_2 < m_3$ ($m_3 < m_1 < m_2$). The values come from a global fit to solar, atmospheric, reactor and accelerator (both short and long baseline) experiments [13]. The most recent contributing data is a precision measurement of ν_μ disappearance from T2K (2014) [14].

oscillation parameters, which are summarised in table 1.1 [12]. The experiments which provided the data inputs to the global fit generally fall into one of four categories, with each category sensitive to a different subset of the neutrino oscillation parameters.

Reactor neutrino experiments measure $\bar{\nu}_e$ disappearance provided by inverse β decay in nuclear reactors with an average neutrino energy of 3 MeV. The baseline for oscillations varies between experiments, but a baseline of around 1 km provides excellent sensitivity to θ_{13} . Examples of reactor experiments are Double CHOOZ [15], Daya Bay [16] and RENO [17].

Solar neutrino experiments detect neutrinos generated in the core of the Sun as a result of nuclear fusion reaction chains. Such experiments are primarily sensitive to θ_{12} and Δm_{12}^2 , which are often referred to as the solar mixing parameters. The final-state neutrinos created in the Sun's core are MeV-scale ν_e but, because of propagation through the core's surrounding matter, the MSW effect results in a highly-pure state of ν_2 at the Sun's surface. As ν_2 is a mass eigenstate, no oscillation occurs between the surface of the Sun and the Earth. Homestake [18], SAGE [19] and SNO [6] are examples of such experiments. It should be noted here that the KamLAND experiment [7], while

a reactor neutrino experiment, was sensitive to the solar mixing parameters due to its 180 km baseline.

Atmospheric neutrino experiments detect neutrinos which are produced when π and K mesons, created by cosmic rays interactions with the upper atmosphere of the Earth, decay. The neutrinos produced are a mixture of ν_μ , $\bar{\nu}_\mu$, ν_e and $\bar{\nu}_e$. Because the cosmic ray flux is fairly uniform, atmospheric neutrino experiments are exposed to neutrinos from all directions, which results in a very wide range of oscillation baselines. The oscillation parameters that such experiments are sensitive to are θ_{23} and Δm_{13}^2 . Super-Kamiokande [5] is an example of an atmospheric neutrino experiment.

Long-baseline accelerator neutrino experiments produce beams of high purity ν_μ (or $\bar{\nu}_\mu$) at GeV-scale energy with wide-ranging baselines which are generally $\mathcal{O}(100 \text{ km})$. The highly man-made nature of such experiments allows almost complete control over L/E allowing careful tuning of parameter sensitivity. Accelerator neutrino experiments are generally sensitive to θ_{13} , θ_{23} , Δm_{13}^2 and δ . K2K [20], MINOS [21], T2K [22] and NO ν A [23] are examples of such experiments.

1.2 The future

It should be clear that an immense amount of progress has been made in the field, with remarkable contributions to the picture coming in only the last 20 years. However, there are several key questions which remain unanswered.

By far the most sought-after answer is whether CP violation occurs in the lepton sector. The magnitude of CP violation is encapsulated in the CP violating phase δ and so it is this parameter which current and future long-baseline experiments are aiming towards. Currently, only T2K and NO ν A can provide hints for values of δ , with the possibility of future constraints. The future long-baseline experiments, Hyper-Kamiokande [24] and DUNE (formerly LBNE) [25] are being designed with a possible measurement of δ as a primary goal.

The second question still to be answered is the ordering of the mass eigenstates. It is not known whether ν_1 , which is dominated by the electron neutrino, or ν_3 is the lightest mass eigenstate. Written more succinctly, is $m_3 \gg m_2 > m_1$ (the normal mass

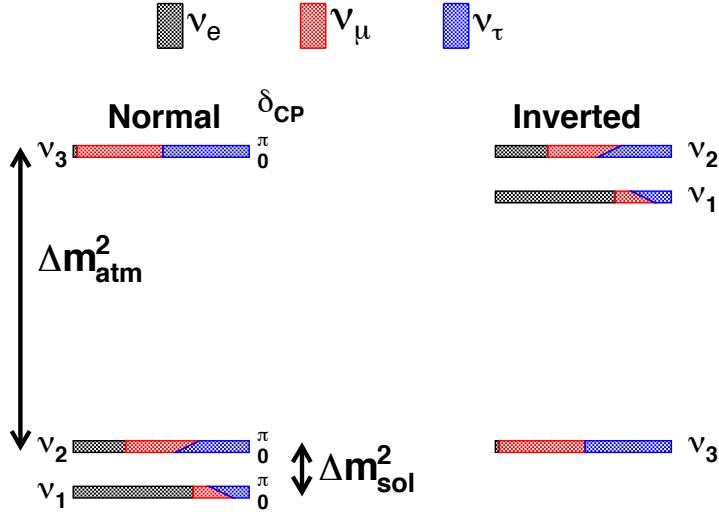


Figure 1.1: The pattern of neutrino masses for the normal and inverted hierarchies with the atmospheric (Δm_{atm}^2) and solar (Δm_{sol}^2) mass splittings labelled. The flavour composition of the mass eigenstates as a function of the unknown CP phase (labelled δ_{CP}) is also shown [26].

hierarchy) or $m_2 > m_1 \gg m_3$ (the inverted mass hierarchy)? This is known as the mass hierarchy problem and its two possible solutions are shown in Fig. 1.1. The matter effects introduced by the MSW effect are mass hierarchy-dependent. So, for very long-baseline experiments, there is mass hierarchy sensitivity. Currently NO ν A has the potential to resolve the mass hierarchy problem. However, both Hyper-Kamiokande (via atmospheric measurements) and DUNE have measurement of the mass hierarchy as a primary goal.

Oscillation experiments only have the capability to measure the differences between the square of the neutrino masses. This means that all oscillation experiments have no sensitivity to the absolute neutrino mass scale and an entirely different type of neutrino experiment is required. Neutrinos are one of the final states associated with β decay and the mass of the neutrino should appear as a cut off in the β spectrum. The visibility of the cut-off entirely depends on the mass scale. So, the KATRIN experiment [27] will attempt to utilise this feature of β decay, with a neutrino mass sensitivity of 0.2 eV.

It is not currently known whether neutrinos are their own anti-particle, which are referred to as Majorana neutrinos. A widely accepted method of studying the Majorana neutrino hypothesis is neutrinoless double β decay in which a pair of neutrons in a nucleus undergo β decay with the final state neutrinos pair-annihilating due to their

Majorana nature. A large neutrinoless double β decay experiment effort is ongoing, including EXO [28], SuperNEMO [29] and SNO+ [30].

The short-baseline neutrino oscillation programme has seen several anomalies [31] which remain unexplained. The LSND experiment found evidence of $\bar{\nu}_e$ in a $\bar{\nu}_\mu$ beam, which was consistent with neutrino oscillations [32]. However the data suggested a mass-squared splitting of 0.2–10 eV². This large splitting is consistent with a fourth species of neutrino. Because the data suggesting three flavours of weakly-interacting neutrino is strong [33], this postulated fourth species must not interact through the weak force, and are generally referred to as sterile neutrinos. More recently, the Mini-BooNE experiment observed a separate short-baseline excess of $\bar{\nu}_e$ in a $\bar{\nu}_\mu$ beam with a mass-squared splitting of 0.01–1.0 eV² [34], further suggesting the sterile hypothesis. New experiments are now under development which aim to test this hypothesis, which include MicroBooNE [35] and SBND (formally LAr1-ND) [36].

Chapter 2

Neutrino interactions with atomic nuclei

The neutrino is a strictly weakly-interacting particle. This has difficult implications for any experiment aiming to study neutrinos as particle detectors generally rely on the electromagnetic force. In fact, the only proven method of neutrino detection is to utilise a high mass target in which the neutrinos can interact. Generally speaking, charged particles are produced by this interaction which can be detected by the usual means. The collected information from these charged final states can then be used to infer information about the incident neutrino. Many modern neutrino experiments rely on this method and so, generally speaking, attempted measurements (e.g. a measurement of δ) rely on our understanding of neutrino interactions with atomic nuclei. Our understanding of such processes is encompassed in the models we use to simulate the interactions.

2.1 Neutrino interactions at the GeV-scale

As the neutrino is weakly interacting, there are two channels available to a neutrino interacting with a nucleon: the Charge Current (CC) interaction in which a W boson is exchanged and the Neutral Current (NC) interaction in which a Z boson is exchanged. For neutrino energies below ~ 1 GeV, the neutrino-hadron interactions are largely Quasi-Elastic (QE) [37]. In such an interaction, the incident neutrino scatters off the nucleon as if it were a single particle, rather than with one of the nucleon's constituent partons. In the case of a CCQE interaction, the neutrino is converted into its charged

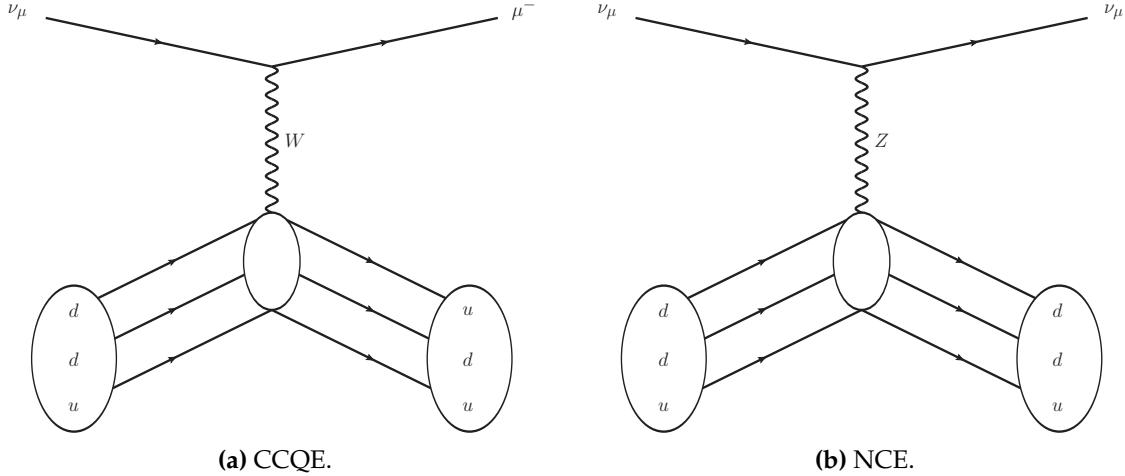


Figure 2.1: Quasi-Elastic (QE) interactions of a ν_μ with a nucleon. The small ellipse represents the neutrino interacting with the nucleon as a whole, rather than with an individual parton.

lepton equivalent and the target neutron converted to a proton. In the specific case of an incident ν_μ , the interaction takes the following form

$$\nu_\mu n \rightarrow \mu^- p. \quad (2.1)$$

For Neutral Current Elastic (NCE) interactions, the incident neutrino remains after the interaction has occurred and no nucleon conversion takes place. Because of this fact, the target nucleon in a NCE interaction need not be a neutron. So, for ν_μ NCE interactions, there are two channels available

$$\nu_\mu n \rightarrow \nu_\mu n, \quad (2.2)$$

$$\nu_\mu p \rightarrow \nu_\mu p. \quad (2.3)$$

The two kinds of QE interaction are shown in Fig. 2.1.

For higher energy neutrinos, there is sufficient energy to promote the target nucleon to an excited state. A quick after-effect of this promotion is that the excited state decays, resulting in further particle emission. This interaction topology, which dominates in the 1 GeV to 5 GeV energy range, is known as RESonant pion (RES) production as the neutrino interaction produces Δ resonance which typically decays to a nucleon and a

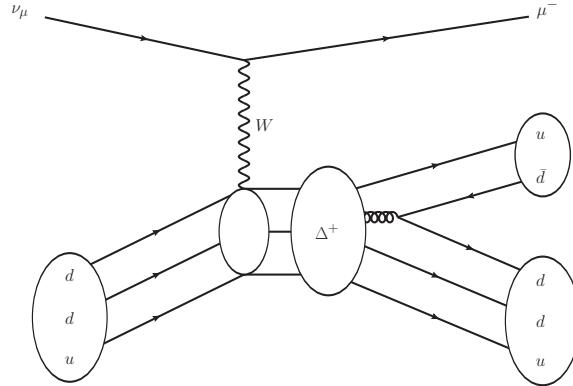


Figure 2.2: A Charged Current RESonant pion (CCRES) interaction of a ν_μ with a neutron. The Δ resonance decays to a neutron and a π^+ .

single pion in the final state. In the case of ν_μ CCRES, the interaction generally takes the following form

$$\nu_\mu N \rightarrow \mu^- N^*, \quad (2.4)$$

$$N^* \rightarrow \pi N', \quad (2.5)$$

where $N, N' = n, p$. An example diagram of a ν_μ -CCRES interaction with a π^+ in the final state is shown in Fig. 2.2.

For neutrinos with energy above the RES-dominant region, the neutrino has enough energy to penetrate the nucleon and scatter off an individual quark. Because of the nature of the strong force, the scattered quark and the nucleon remnant produce a hadronic shower in the final state. This process is known as Deep Inelastic Scattering (DIS). For ν_μ -CCDIS, the interaction takes the following form

$$\nu_\mu N \rightarrow \mu^- X, \quad (2.6)$$

where X is the remnant of the nucleus after the interaction occurs. An example diagram of a ν_μ -CCDIS interaction is shown in Fig. 2.3.

While the value of a particular interaction cross-section should depend on the nuclear environment, it is possible to make comparisons of the measured cross-section per nucleon. Fig. 2.4 shows a comparison of ν_μ CC cross-section measurements per nucleon

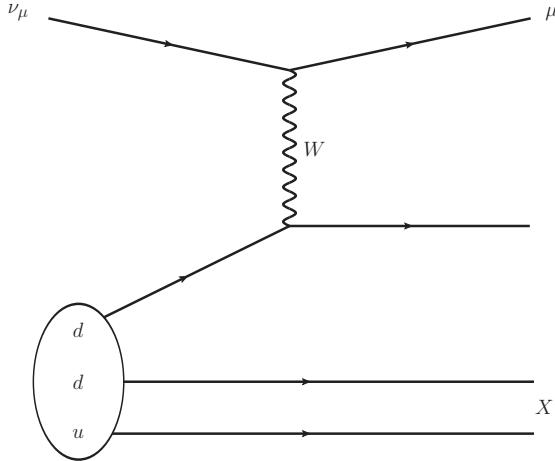


Figure 2.3: A Charged Current Deep Inelastic Scattering (CCDIS) interaction of a ν_μ with a neutron. X represents the leftover nuclear remnant.

from different experiments, all of which sample a different neutrino energy range. There are large uncertainties for many of the cross-section measurements, particularly for the ones sampling the lower energy ranges. The T2K beam energy is ~ 700 MeV, which sits in the region of higher uncertainty.

CCQE interactions are experimentally the most interesting and this is the interaction region where most recent measurements have been focused. Because of the simplicity of the CCQE topology, the interaction can be treated as a two-body scatter. So, by applying simple conservation rules, the neutrino energy can be kinematically reconstructed. In such interactions, the nucleon structure is parameterised using a set of form factors, the most interesting of which is the axial-vector form factor, $F_A(Q^2)$. $F_A(Q^2)$ has been, and still is, assumed to take a dipole form

$$F_A(Q^2) = \frac{F_A(0)}{(1 + Q^2/M_A^2)^2}, \quad (2.7)$$

where Q^2 is the negative of the squared four-momentum transfer of the lepton to the hadron, $F_A(0) = 1.2694 \pm 0.0028$ [42], and M_A is known as the axial mass. Recent measurements of the CCQE cross-section by the MiniBooNE [43] and NOMAD [44] experiments have sparked interest by reporting measured cross-sections which are in tension with one another, the results of which are shown in Fig. 2.5. A popular explanation for this discrepancy is a lack of understanding of the nuclear environment. Because the neutrino is not scattering of a free nucleon, but rather a nucleon in a strongly contained

system, experiments actually measure an effective M_A . It is possible that the nuclear effects cause a modification to the effective M_A that the experiments measure. This possible explanation for the discrepancy has placed a heavier emphasis on nuclear modelling in neutrino interaction experiments.

2.2 Neutrino interactions with heavy nuclei

As introduced above, consideration of nuclear effects in cross-section measurements is important. This is especially true for neutrino interactions on heavy target nuclei. As one can imagine, the presence of a nucleus can dramatically effect the interactions that are observed in a detector. A popular model for the nucleus is the Relativistic Fermi-Gas (RFG) model [47]. At its heart, the RFG represents the target nucleus as an ideal gas of weakly interacting fermions. All nucleons in the nucleus model con-

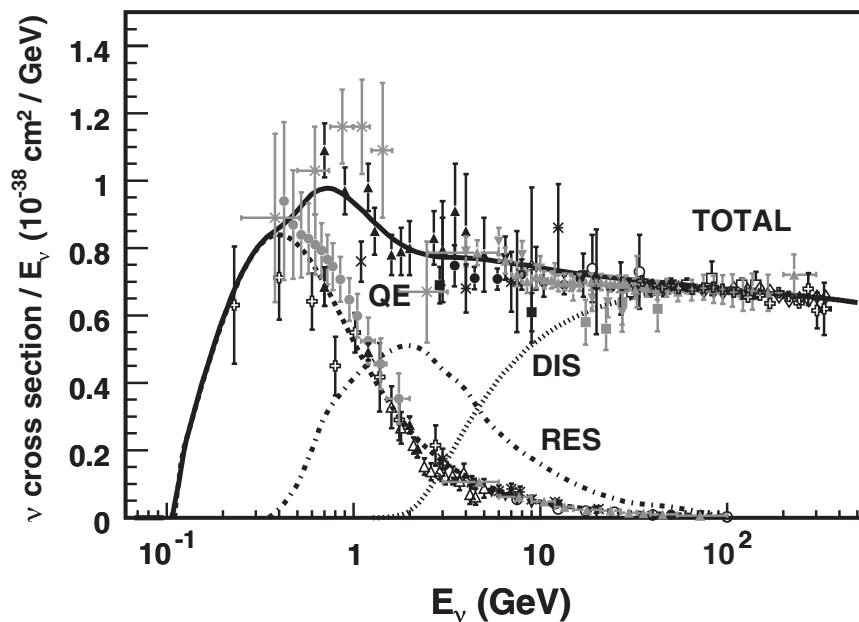


Figure 2.4: ν_μ CC cross-section measurements per nucleon and divided by neutrino energy for a range of energies, showing the QE, RES and DIS contributions [37]. The data points are provided by a range of experiments (from 1979 to 2010) including BEBC (1979) [38], NuTeV (2006) [39], MINOS (2010) [40] and others. The example predictions are provided by the NUANCE generator [41]. Many measurements have a large associated uncertainty, particularly in the lower energy (less than 1 GeV) regime. The heaviest target nucleus probed in the region of interest, near 1 GeV, is carbon.

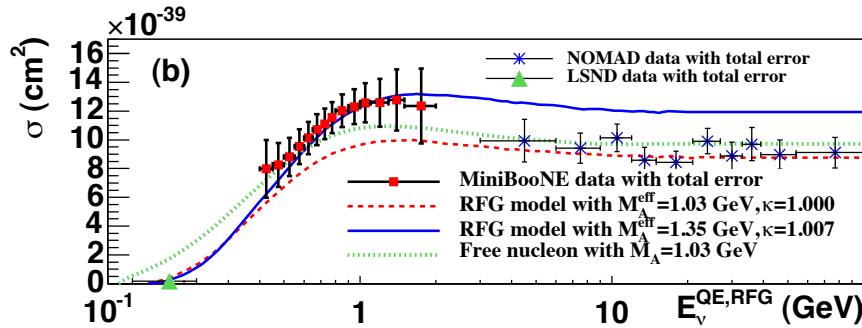


Figure 2.5: The CCQE cross-sections measured by the MiniBooNE (2010) [43], LSND (2002) [45] and NOMAD (2009) [46] experiments. The solid and dashed lines represent predictions from the NUANCE generator with different values of M_A [43]. Each prediction does not well model all of the data shown in the figure, indicating tension between the data collected by each of the experiments.

tribute to the nuclear potential and are considered as moving freely within the nuclear volume. In such a model, the constituent neutrons and protons are distinguishable, which results in two nuclear potential wells within the nucleus. An example schematic of the RFG model is shown in Fig. 2.6. Because the system obeys Fermi-Dirac statistics, each nucleon energy level can be occupied by two nucleons, corresponding to each of the possible spin states. It should be noted that the Fermi-level for both the proton and neutron potential wells must be equal as an imbalance would lead to the nucleus decaying to a more energetically favourable state. As heavier nuclei generally contain more constituent neutrons than protons, it follows from the previous arguments that the depth of the neutron potential well must be greater than that of the protons, as shown in Fig. 2.6. The conclusion is that constituent protons are less strongly bound than the constituent neutrons. When the nucleus sits in its ground state (at $T = 0$), all nuclear energy levels are filled from the bottom up to a maximum momentum, called the Fermi momentum, p_F .

Because the nuclear system in the RFG model obeys Fermi-Dirac statistics, only two states per nuclear level can be occupied. This means that, at $T = 0$, it is only possible to excite a nucleon to an energy level which is above the Fermi momentum of the nuclear potential well. This process is known as Pauli blocking and has a direct effect on the neutrino interaction cross-section. Specifically, because of the requirement that the energy transferred from the neutrino to the nucleon must promote the nucleon momentum beyond the Fermi-momentum, the available phase-space of the interaction is reduced, leading to a reduction in the neutrino interaction cross-section. This effect

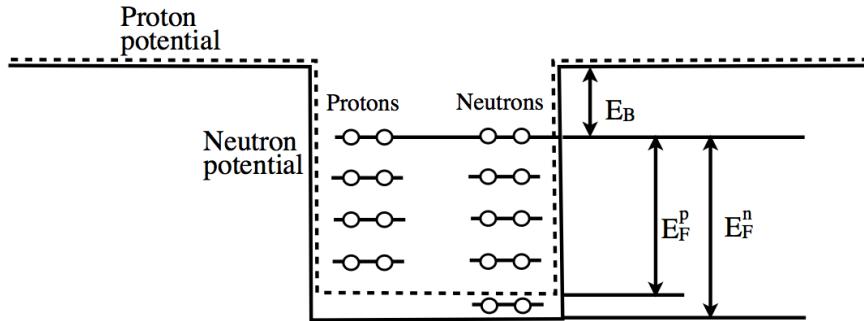


Figure 2.6: Sketch of the potential wells for protons and neutrons in the Relativistic Fermi-Gas (RFG) model. The difference between the Fermi level and the potential well edge, E_B , is approximately constant for different nuclei. The energy between the proton and neutron ground states and their maximum filled state, E_F^p and E_F^n respectively, are also shown. Due to Coulomb repulsion, the proton potential well is shallower than the neutron potential well.

is strongest when the neutrino energy is approximately the Fermi-momentum of the nucleus.

The RFG can only model the effect of the nucleus on the initial neutrino interaction which creates the final states. However, these final states are created within the nucleus and so additional interactions of the final states with the nucleus can occur. The Final-State Interactions (FSI) can significantly alter the momentum and direction of the final-state particles. As the final-state particles are used to infer neutrino properties, the FSI effects can alter the interpretation of the reconstructed events. In simulation, variations of the cascade model are typically used. This involves pushing the final-state particles through the nucleus in discrete steps and, at each step, probabilistically updating the particle properties. If at any point a final-state particle knocks out another nucleon, the additional nucleon is also pushed through the nucleus in parallel. The discreet stepping occurs until all relevant particles have escaped the nucleus.

To test such cross-section models, including nuclear effects, it is necessary to compare prediction with collected data. However, collected cross-section data for heavy nuclei is relatively sparse. In the case of lead, only two experiments have performed cross-section measurements. The first measurement was performed by the CHORUS [48] experiment in 2003. The CHORUS detector, exposed to the CERN SPS beam with a wide-band ν_μ beam of 27 GeV average energy, measured a cross-section for lead, iron, marble and polyethylene. However, because the absolute flux was not measured in the experiment, all of the cross-section measurements were normalised to a common

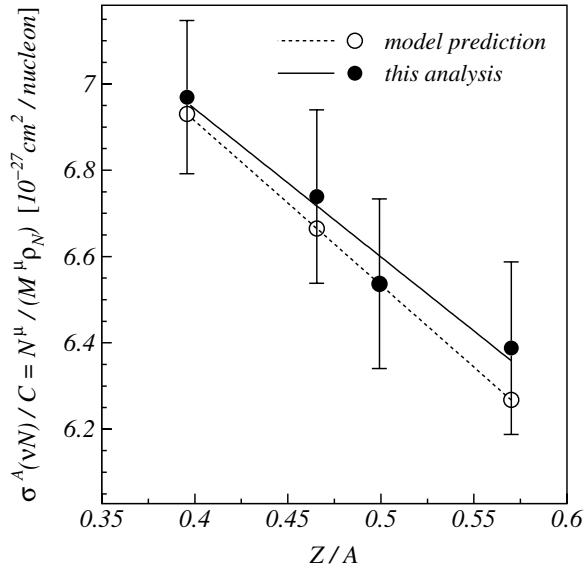


Figure 2.7: Measured values of the ν_μ CC relative cross-section for several elements, as measured by the CHORUS experiment (2003) [48]. The black and white points are the collected data and prediction respectively. The solid and dashed lines are the linear best fit lines to the data and prediction respectively. Going from left to right, the points represent data and prediction for lead, iron marble and polyethylene. The prediction is taken from a set of quark distribution functions [49] provided by PDFLIB.

constant. Their results are summarised in Fig. 2.7. The second measurement was made by the MINERνA experiment (2014) [50], which used the Fermilab NuMI beam with a 8 GeV average energy, to measure the relative ν_μ CC cross-section on lead to that of plastic scintillator as a function of neutrino energy. Their results, shown in Fig. 2.8, largely agreed with the prediction. The energy sampled is above the 1 GeV region of interest.

As neutrino oscillation physics has entered the precision era, it has become very important that our understanding of neutrino cross-sections improves. To achieve this goal, more cross-section measurements across a range of nuclear targets are needed. Lead is a widely used material in particle detectors. The short radiation length (5.6 mm) allows for excellent containment of electromagnetic showers while its high density allows for a large mass target with a relatively small volume of material. Despite this, very little information has been collected on neutrino interactions on lead. At the time of writing, only the experiments discussed above have undertaken such studies and both provided cross-section ratio measurements only (no measurement of the

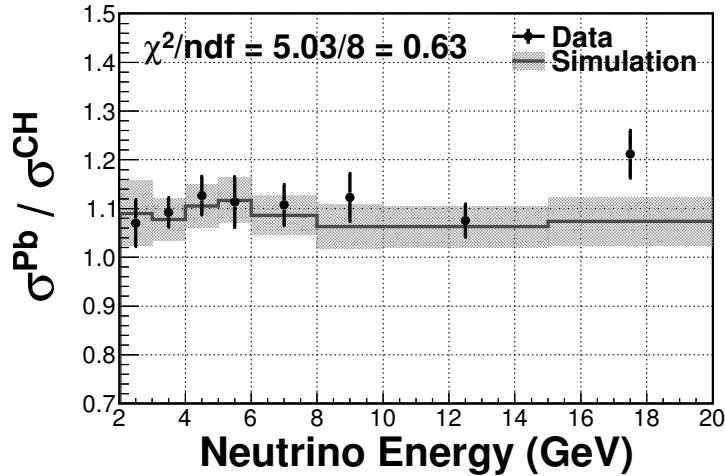


Figure 2.8: Ratio of the measured ν_μ CC inclusive cross-section on lead to plastic scintillator as a function of neutrino energy, as measured by the MINERνA experiment (2014) [50]. The simulation is based on the GENIE generator [51]. The error bars on the simulation (data) are statistical (systematic) uncertainties.

absolute cross-section). Additionally, both experiments used a neutrino beam well above the T2K neutrino beam energy. This leads to the possibility of making two world first measurements using data collected in the T2K experiment. Therefore, this thesis presents a measurement of the absolute ν_μ CC inclusive cross-section on lead using the electromagnetic calorimeters contained in the near detector of the T2K experiment.

Chapter 3

The T2K Experiment

The Tokai-to-Kamioka (T2K) experiment [52] is a long baseline neutrino oscillation experiment located in two sites across Japan and is designed to study the parameters governing the PMNS matrix. The first site is the J-PARC facility in Tokai-mura on Japan's east coast which houses a 30 GeV proton accelerator complex that is used to generate a highly pure ν_μ beam. J-PARC also contains a suite of detectors designed to measure the neutrino beam's unoscillated characteristics. Super-Kamiokande (SK) is located 295 km (see Fig. 3.1) and measures the contents of the neutrino beam post-oscillation.

T2K was the first experiment to observe the $\nu_\mu \rightarrow \nu_e$ appearance channel [22] which excluded $\theta_{13} = 0$ at 7.3σ significance. By comparing this result with precise θ_{13} measurements from reactor experiments, δ_{CP} regions can be excluded at 90% confidence level (see Fig. 3.2). T2K's precision analysis of the ν_μ disappearance channel provide world leading measurements of θ_{23} and Δm_{23}^2 . Independently of the oscillation analyses performed by the experiment, T2K's near detectors, ND280 and INGRID, are used to measure a range of neutrino cross-sections [53, 54]. While this is not the primary aim of T2K, such measurements are still extremely important as T2K systematic uncer-

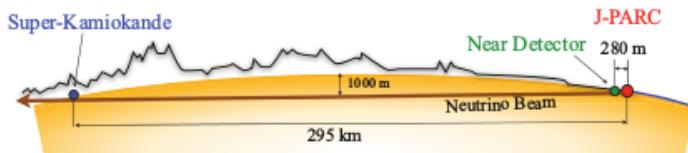


Figure 3.1: Schematic of the T2K experiment showing the near and far sites, separated by the 295 km baseline.

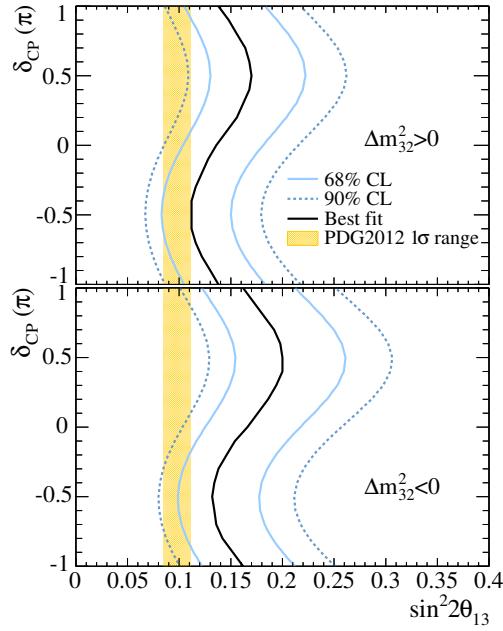


Figure 3.2: The 68% and 90% confidence level allowed regions for $\sin^2 2\theta_{13}$ as a function of δ_{CP} for normal hierarchy (top) and inverted hierarchy (bottom). The solid line represents the best fit $\sin^2 2\theta_{13}$ for a given δ_{CP} . The shaded region shows the average θ_{13} provided by the reactor constraint [22].

tainties can be constrained with additional cross-section knowledge as well as helping to understand the general neutrino interaction picture.

3.1 T2K beam

The T2K neutrino beam is generated by J-PARC's accelerator complex which produces a 30 GeV proton beam which is fired at a fixed graphite target. The final-state particles of interactions with the target are predominately charged pions which decay to produce the neutrino beam. Surrounding and behind the graphite target are a set of magnetic horns which focus the pions into a beam, resulting in a focused neutrino beam after the hadrons have decayed.

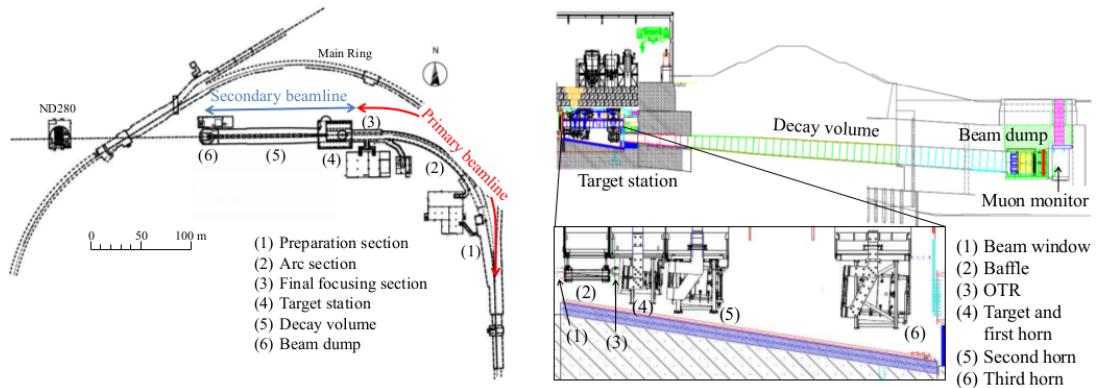


Figure 3.3: A schematic of the T2K neutrino beamline (left) and a side view of the secondary beamline (right) [52].

3.1.1 Accelerator complex

The J-PARC accelerator complex consists of three sections: the LINnear ACcelerator (LINAC), the Rapid-Cycling Synchrotron (RCS) and the Main Ring synchrotron (MR). Production of the proton beam starts at the LINAC where H^- anions are accelerated to 181 MeV which are subsequently converted to H^+ ions via charge-stripping foils at the RCS injection point. With a 25 Hz cycle, the ions are further accelerated by the RCS to 3 GeV with two bunches per cycle. Roughly 5% of the proton bunches are fed into the MR where the final acceleration to 30 GeV occurs in bunches of eight. Extraction of the bunches occurs at two points for different experiments. For T2K, all eight bunches are extracted in a single turn by five kicker magnets and aimed down the neutrino beamline to the graphite target. The extraction of all eight bunches forms a single beam spill with a width of $5 \mu\text{sec}$. The tight structure of the beam spills is vital for background discrimination in the downstream detectors.

3.1.2 Neutrino beamline

The neutrino beamline (NU) is split into a primary and secondary beamline, a schematic of which is shown in Fig. 3.3. The primary beamline consists of a preparation section, an arc section and a focusing section. The preparation section uses 11 normal conducting magnets to tune the proton beam for entry into the arc section where the proton beam is bent to its intended direction. As will be discussed in more detail in section 3.1.3, the axis of the beam is 2.5° away from SK. The final focusing sections then guides the proton beam into the secondary beamline and the graphite target.

Sound performance of the proton beam is vital for stability of the T2K neutrino beam. To ensure such performance, the primary beamline is equipped with a suite of monitors to measure the position, profile, loss and intensity of the proton beam. The beam position is measured by 21 ElectroStatic Monitors (ESMs) which consists of four cylindrical electrodes surrounding the beam. The asymmetry of the beam is measured by the induced current in the electrodes which is used to infer the position in a non-destructive manner. Segmented Secondary Emission Monitors (SSEM) are used to measure the beam loss. Each SSEM has an anode foil sandwiched between two titanium foil strips. Protons interact with the strips causing an emission of electrons which electrically drift inducing a current in the strips. The charge distribution is used to reconstruct the profile. The Beam Loss Monitors (BLMs) are Ar-CO₂ filled wire proportional counters and are used to quantify the beam loss. The intensity of the beam is measured by five Current Transformers (CTs) which consist of a 50-turn toroidal coil around a ferromagnetic coil. Passage of the beam induces a current in the coil which is used to infer the number of protons in the spill. The final CT (CT5) is positioned at the end of the focusing section of the primary beamline and is used to count the number of protons incident on the graphite target. The accumulated number of protons on target (POT) is used as a metric for the data collected by T2K. The total POT accumulated so far by T2K is shown in Fig. 3.4.

The secondary beamline consists of the graphite target, a set of magnetic focusing horns, a decay pipe and a beam dump. A schematic for the secondary beamline is shown in Fig. 3.3. The graphite target is a 2.6 cm diameter and 91.4 cm long rod which is surrounded by a 2 mm thick graphite tube and a 0.3 mm titanium case. The proton-graphite interactions produce charged pions and kaons which are focused by three magnetic horns, one of which surrounds the target. The magnetic horns consist of two coaxial conductors which generate a magnetic field with a strength inversely proportional to the distance from the beam axis. The current direction in the magnetic horns causes the induced field to focus or deflect particles depending on their charge sign. This simple control allows T2K to operate in ν or $\bar{\nu}$ beam mode. The focused mesons then travel down a decay pipe filled with Helium to reduce pion absorption. It is here that the mesons decay to produce the neutrinos which form the T2K beam. To stop measurement contamination, other decay products must be stopped before reaching the downstream detectors. So, a 75 ton graphite beam dump is positioned at the end of the decay volume. The beam dump stops almost all non-wanted decay products, with only 5 GeV or above muons successfully propagating through. As the

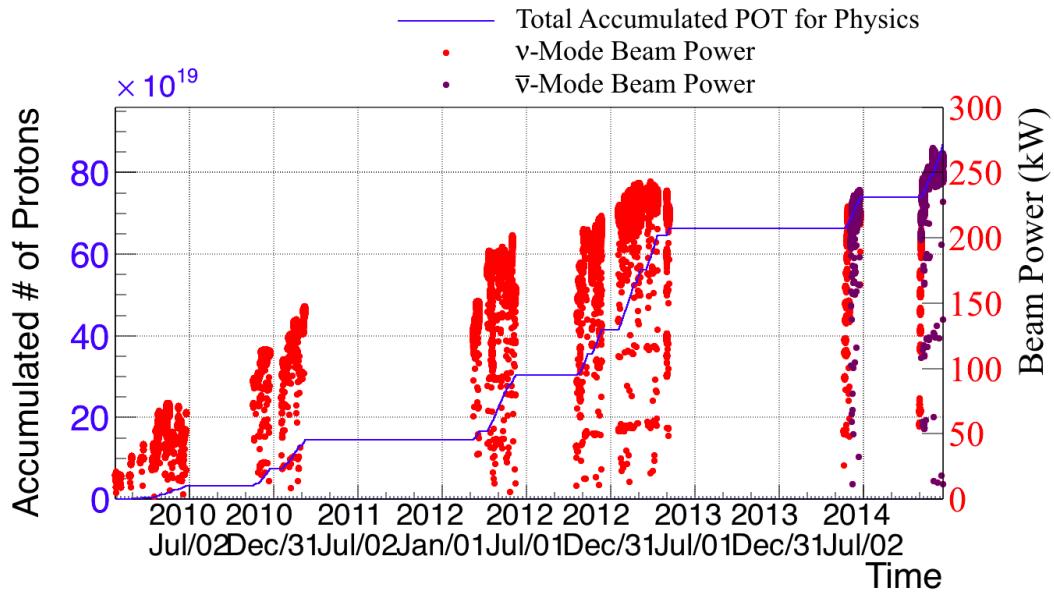


Figure 3.4: The POT recorded by CT5 as a function of time (blue line) and the recorded beam power in ν running mode (red dots) and $\bar{\nu}$ running mode (purple dots). The recorded POT as a function of time shows T2K has been successful in greatly increasing the size of its dataset during each data collection run.

muons are generally simultaneously produced with the beam neutrinos, measurements of the muons can be used to monitor the direction of the neutrino beam. To do this, a MUon MONitor (MUMON) is installed at the downstream end of the beam dump.

3.1.3 Off-axis beam

The kinematics of the pion decays dictate the energy spectrum shape of the neutrinos. Specifically, the peak width of the neutrino energy narrows and shifts as an observer moves off-axis from the pions trajectory. As the pions are the neutrino parents in the T2K beam, the same effect can be seen by moving off-axis from the neutrino beam. This effect is illustrated in Fig. 3.5. By positioning T2K's baseline detectors at 2.5° off-axis, it is possible to align the neutrino beam's peak energy with the first oscillation maximum for the ν_μ disappearance channel. Separately, an off-axis configuration reduces the beam's unwanted high energy tail, improving sensitivity to ν_e appearance and ν_μ disappearance.

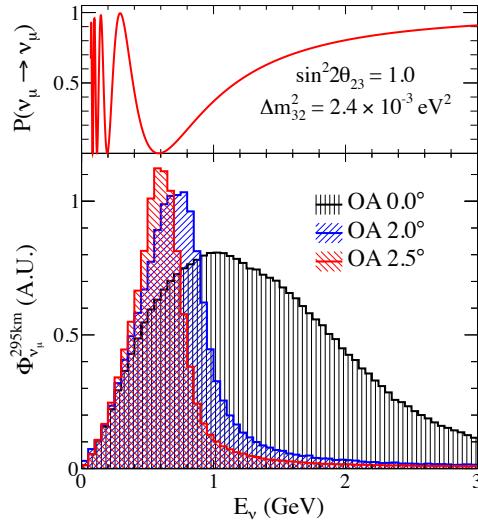


Figure 3.5: Muon neutrino survival probability at 295 km (top) and neutrino fluxes for different off-axis angles (bottom) [55]. The peak neutrino lowers and the energy distribution narrows as the off-axis angle increases. By carefully selecting the off-axis angle (2.5°), the peak energy of the neutrino beam can be aligned with the first maximum of the ν_μ disappearance probability spectrum.

3.2 Near detector complex

Located 280 m downstream of the beam target is the near detector complex which houses a pair of detectors which measure the unoscillated characteristics of the beam. The two detectors, named INGRID and ND280, sit in a 37 m deep, open air pit lined with concrete which is surrounded by sand.

3.2.1 Multi-Pixel Photon Counter

Multi-Pixel Photon Counters (MPPCs) [56] are used extensively in both INGRID and ND280 for detection of scintillation light during particle energy deposition in plastic scintillator. The choice of MPPCs, rather than more traditional photomultiplier tubes, was largely due to their ability to operate in a magnetic field. A MPPC is a multi-pixel avalanche photodiode which consists of 667 pixels over an area of $1.3 \times 1.3 \text{ mm}^2$. In terms of operation, each MPPC is held at 0.8-1.5 V above their breakdown voltage, resulting in a gain of 1×10^6 , which is consistent with the gain of a vacuum photomultiplier [52]. When light is incident on a MPPC, each pixel acts as a detector for a single photon which means the total signal collected is simply the sum of the MPPC's fired

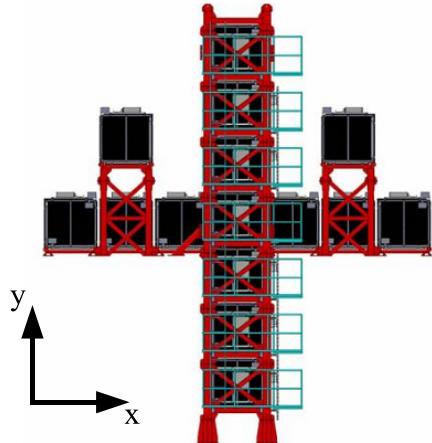


Figure 3.6: A schematic of INGRID, showing the horizontal, vertical and off-axis modules [52].

pixels.

The scintillator bars, which the MPPCs collect the light from, consist of plastic scintillator bars with a wavelength shifting (WLS) fibre threaded through the centre. During energy deposition, the plastic scintillator emits photons which are collected and transported by the WLS fibre to the bar's end where the MPPC is located. The WLS fibre has a twofold purpose: carry the light to the MPPC and shift the spectrum of the light to a region where MPPC detection is optimised.

3.2.2 INGRID

INGRID (Interactive Neutrino GRID) is one of T2K's near detectors. With its centre positioned on the beam axis, INGRID is designed to directly monitor the beam's direction and intensity. INGRID consists of 14 identical modules arranged in a cross formation with two additional modules positioned off the cross axis towards the end of each horizontal arm (see Fig. 3.6). The cross arrangement allows INGRID to sample the beam in a $10\text{ m} \times 10\text{ m}$ transverse section.

Each module consists of nine iron plates and 11 tracking scintillator planes in a sandwich structure. The iron plates are $124\text{ cm} \times 124\text{ cm} \times 6.5\text{ cm}$ and provide 7.1 tons of target per module for the neutrino beam. The scintillator planes provide tracking for the neutrino final-states and consist of scintillator bars threaded by WLS fibres and readout by MPPCs (see section 3.2.1).

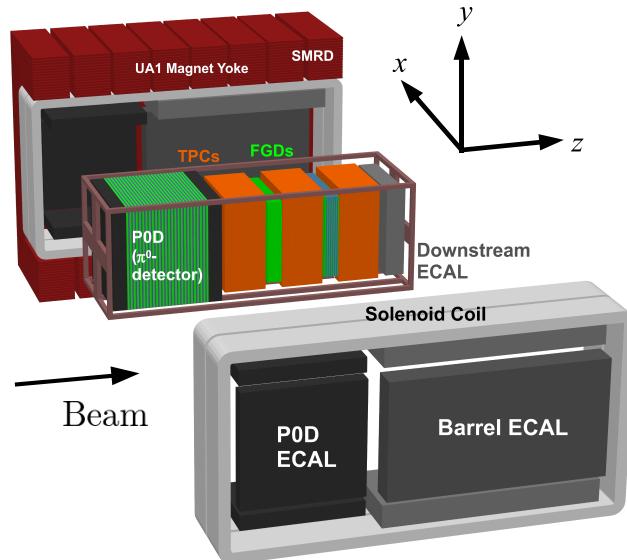


Figure 3.7: An exploded view of ND280 [52], showing the locations of its subdetectors.

This overall design allows INGRID to measure the beam centre to a 10 cm precision which corresponds to 0.4 mrad precision at the near detector complex [52].

3.2.3 ND280

ND280 (Near Detector at 280 m) is T2K’s other near detector. However, unlike INGRID, ND280 is positioned 2.5° off-axis to the neutrino beam. ND280 is a heavy, fine-grained detector which characterises the flux, energy spectrum and ν_e contamination of the ν_μ beam and additionally makes neutrino cross-section measurements. Because ND280 sits at the same off-axis angle as T2K’s far detector, ND280’s beam characterisation can be used to make signal and background predictions at the far detector.

Fig. 3.7 shows an exploded view of the detector, which reveals the many subdetectors that form ND280. The central tracking region comprises two Fine-Grained Detectors (FGDs [57]) sandwiched between three Time Projection Chambers (TPCs [58]). The FGDs, which are composed of layers of plastic scintillator bars, provide the primary target for the neutrinos to interact with and the TPCs, filled with a gaseous mixture, allow for tracking of the charged final-states. Essentially, the detectors in this region are complimentary and their combined information is used to reconstruct the major-

ity of beam events relevant for oscillation analyses. Upstream of the Tracker region lies the π^0 detector (PØD [59]), whose design is optimised for studying neutrino interactions with π^0 in the final-state and consists of layers of scintillator, water and brass. Surrounding the Tracker and PØD are a set of Electromagnetic Calorimeters (ECals [60]), primarily designed to detect particles originating from ND280’s inner region. Because particle identification is paramount to ND280’s physics goals, all of the above detectors sit in a constant 0.2 T magnetic field, which is aligned with the x -axis in Fig. 3.7. The magnetic return yoke and coils used to generate this field encompass the entire detector, allowing a constant field to be maintained within the detector, but greatly minimising the field’s outside extent. To maximise ND280’s physics capability, the magnetic return yoke is instrumented with layers of plastic scintillator, which form the Side Muon Range Detectors (SMRDs [61]).

3.2.3.1 The electromagnetic calorimeters

The ND280 ECal is a lead-scintillator sampling calorimeter which consists of 13 modules separated into three distinct regions: the PØD ECal which consists of six modules surrounding the PØD, the barrel-ECal which is separated into six modules surrounding the Tracker region and the DownStream-ECal (DS ECal) which is a single module located downstream of the inner detectors. Motivated by their physics goals, the barrel-ECal and DS ECal are often considered together and will be referred to as the Tracker ECal, while the PØD ECal is considered a separate detector which is not used in this analysis and so will not be discussed further. The primary physics goal of the ECal is to aide particle identification for final-states originating in the central region of ND280. This is particularly important for interactions with π^0 in the final-state as the decay photons are difficult to identify using the Tracker alone.

The barrel-ECals are separated into six modules: two modules above the Tracker, two either side of the Tracker and two below the Tracker. Each barrel-ECal module consists of layers of scintillating polystyrene bars with a 40 mm \times 10 mm cross-section bonded to 1.75 mm lead sheets. The scintillator bars provide a means of tracking the final-states while the lead sheets act as a radiator to produce electromagnetic showers and additionally provide a heavy mass neutrino target. To measure the light readout, each scintillator bar has a 2 mm hole running through the centre in which a WLS fibre is inserted. The fibre carries the light to the end of the bar where it can be collected by an MPPC. The size of the gap between the Tracker and the magnet placed

a strong constraint on the size of barrel modules and so a scintillator bar thickness of 10 mm was chosen to minimise the ECal depth while maintaining a bar thickness which could provide sufficient light for signal capture. The other key features of the active barrel-ECal volume (bar thickness, lead thickness and number of layers) were chosen to optimise particle identification and tracking. A smaller bar width results in a detector with a higher resolution and studies investigating this found that the π^0 reconstruction efficiency was greatly compromised for >50 mm bar widths. So, to facilitate costings, a compromise bar width of 40 mm was chosen. The thickness of the lead absorber was also optimised based on the π^0 reconstruction efficiency. The number of scintillator-lead layers was chosen such that electromagnetic showers of energy up to 3 GeV were adequately contained. It was found that 10 electron radiation lengths (X_0) were required to ensure containment of at least 50% of π_0 decay photon showers. So, this motivated a choice of 31 layers for all barrel-ECal modules which is equivalent to $9.7X_0$. For the purpose of 3D reconstruction, each ECal layer is oriented at 90° to the previous layer. This means that in all barrel modules, there are 16 layers perpendicular and 15 layers parallel to the beam direction. The perpendicular and parallel layers are slightly different, in that the scintillator bars are a different length. In all barrel modules, the parallel bars are 3840 mm and are read out at both ends by separate MPPCs. Because of the geometry, this is not the case for the perpendicular bars; the top and bottom module bar lengths are 1520 mm whereas the side module bar lengths are 2280 mm. For all perpendicular bars, the signal is read out at one end only with the other end mirrored with aluminium to reflect the light. Each barrel module is sandwiched between two carbon fibre plates and held in an aluminium frame which provides secure, structural support.

The DS ECal has almost identical features to that of the barrel in that it has scintillator bars with an identical chemical composition and cross-section while the lead absorbers are equally thick. However, unlike the barrel-ECals, the DS ECal consists of 34 lead-scintillator layers which are, again, orientated at 90° to the previous layer. Because of this, the DS ECal has a radiation length of $10.6X_0$. Additionally, every scintillator bar is 2000 mm and is read out at both ends by MPPCs. Also, because of its position in the geometry, all layers are perpendicular to the beam direction. The carbon fibre plates and aluminium frame are identical to those used in the barrel-ECal.

A summary of the Tracker ECal design is shown in table 3.1.

	DS ECal	Barrel ECal
Length (mm)	2300	4140
Width (mm)	2300	1676 top/bottom 2500 side
Depth (mm)	500	462
Weight (kg)	6500	8000 top/bottom 10000 side
Num. of layers	34	31
Bar orientation	x/y	Para. and Perp.
Num. of bars	1700	2280 Para. top/bottom 1710 Para. sides 6144 Perp. top/bottom 3072 Perp. sides
Bars per layer	50	38 Para. top/bottom 57 Para. side 96 Perp. top/bottom/sides
Bar length (mm)	2000	3840 Para. 1520 Perp. top/bottom 2280 Perp. sides
Pb thickness (mm)	1.75	1.75

Table 3.1: Summary of the ECal design showing the overall dimensions, numbers of layers, length and orientation of the scintillator bars, numbers of bars, and lead thickness for each module [60].

The scintillator bars consist of polystyrene doped with 1% PPO and 0.03% POPOP and were extruded at a dedicated Fermilab facility. During charge deposition, the PPO works as the primary scintillator and its output photons result in secondary scintillation of the POPOP. This process acts as a wavelength shifter to produce an emission peak of 420 nm which matches the absorption peak of the 1 mm diameter WLS fibre threaded through the centre of the bar. Every Tracker ECal bar contains a 0.25 mm coating of polystyrene co-extruded with TiO₂ which provides reflection of the scintillation light.

The lead absorber layers consist of naturally occurring lead and stiffened with 2% antimony. Traces of other metals are present but are below 0.15%. During construction, each lead layer was coated with a black, quick drying metal-conditioning primer to protect personnel from the toxic effects of the lead and to prevent leaching into the scintillator bars. The lead layers themselves are actually constructed from multiple sheets rather than a single sheet largely due to ease of transportation. In the case of the DS ECal, each layer consists of two 1008 mm × 2016 mm sheets. For the top and bottom ECal modules, each layer consists of two 765 mm × 3858 mm sheets and the side module absorbers are constructed from four 2330 mm × 964.5 mm sheets.

The Tracker ECal electronics system consists of several different readout boards. All MPPCs are connected to a set of bespoke Trip-T [62] Frontend Boards (TFBs). Each TFB comes with 64 channels to read out MPPCs which means that there are multiple TFBs associated with each ECal module. All TFBs subsequently connect to Readout Merger Modules (RMMs) which act as the interface between the Data AcQuisition system (DAQ) and its associated TFBs.

3.2.3.1.1 ECal design

The design process of the ECals was driven by optimising the physics capability of the subdetector whilst considering the space limitations between the tracker and the magnet. It was decided early on that the ECals would consist of plastic scintillator bars interleaved with sheets of lead and so the optimisation process consisted of tuning the dimensions of each component to suit the needs of the experiment. One of the

major physics goals of ND280 is to provide an estimation of the rate of interactions at the far detector which produce a π^0 in the final state as such interactions can fake the signature of the oscillated ν_e events. Therefore, the π^0 production channel can be considered a signal ND280. One π^0 topology that was considered during ND280 design was interactions in the FGD involving a final-state π^0 whose decay photons entered and showered in the ECal. In addition to this important topology, the ECal could provide additional Particle identification for other topologies, such as CCQE and CCRes interactions in which the final state particles (such as μ^\pm , π^\pm and protons) enter and stop in the ECal. The desired characteristics of the ECal were driven by the radiation length (X_0) of the lead absorbers which is 5.6 mm [63]. In the specific case of photons, the mean free path is 7.2 mm. Separately, because of the large mass provided by the magnetic yoke, the number of background events was expected to be many orders of magnitude more than the number of interactions in the tracker region which meant the ECal must be designed with the ability to provide sound direction and pointing information. The considerations outlined resulted in a highly-segmented sampling calorimeter choice which required component dimension tuning using the above considerations as metrics.

As the lead absorbers drive the shower containment but are also destructive, a thin lead layering scheme would result in very poor containment of the electromagnetic showers which would greatly hinder the ability to reconstruct the π^0 decay photons, while a thick lead layering scheme would absorb the vast amount of a particle's energy, resulting in little information recorded from the scintillator bars. Fig. 3.8 shows the π^0 reconstruction efficiency in the initial ECal simulations as a function of the thickness of the lead layers. As both extreme values for the thickness result in a relatively low reconstruction efficiency, a lead layer thickness of 1.75 mm was chosen as a compromise between sound reconstruction and shower containment.

The width of the scintillator bars was found to also affect the ability to reconstruct the photons from π^0 decays. Specifically, if the width of the bar became too large, the amount of charge deposited by the electromagnetic showers would only be contained on a small number of scintillator bars which would not provide the adequate level of information necessary for reconstruction. Fig. 3.9 shows the π^0 reconstruction efficiency as a function of bar width. As can be seen from the figure, as the bar width passes 50 mm, the reconstruction efficiency is severely effected. Using narrower bars in the ECal also allows for other benefits such as better particle identification. It

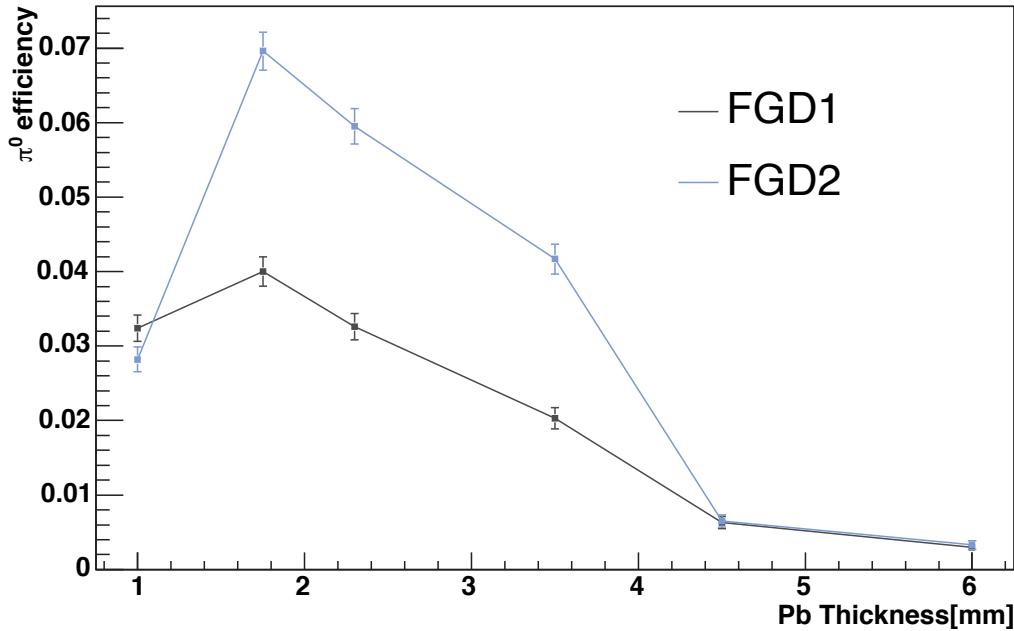


Figure 3.8: The π^0 reconstruction efficiency in initial ECal simulation as a function of the thickness of the lead layers. The blue and black lines show π^0 events which occurred in FGD1 and FGD2 respectively. The thicker lead layers lead to a large energy absorption while the thin layers lead to poor shower containment. Both extremes lead to a poorer π^0 reconstruction efficiency. A bar thickness of 1 cm was used in all of the ECal simulations shown [63].

was found that the number of T2K beam-induced muons misidentified as electrons worsened from 10% to 35% as the bar width was increased from 20 mm to 40 mm. Unfortunately, cost was also a constraint in the ECal design and so a bar width of 40 mm was selected for the design.

3.2.3.1.2 ECal construction

The construction of the ECal followed a distributed model in which all contributing institutions provided a set of services which optimised available space, efficiency and personnel use. Roughly speaking, the ECal module layer was constructed and then lowered inside of an already prepared bulkhead. The MPPCs, cooling pipes, TFBs and cooling panels were then attached, followed by the cover panel, sealing the layers. The RMMs were then attached to the outside of the constructed ECal module with the

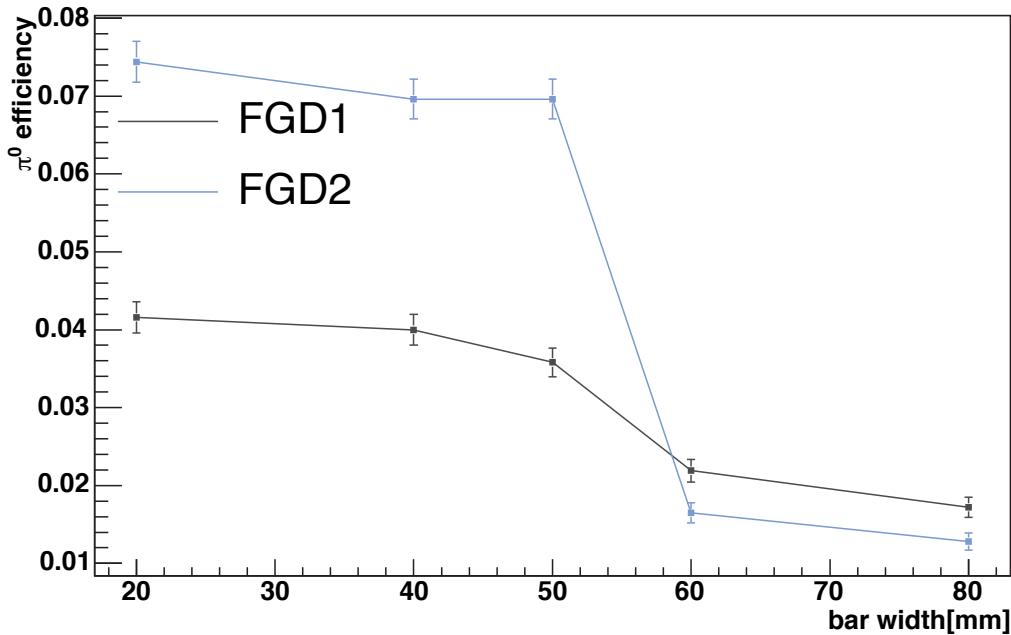


Figure 3.9: The π^0 reconstruction efficiency in initial ECal simulation as a function of the scintillator bar width. The blue and black lines show π^0 events which occurred in FGD1 and FGD2 respectively. The reconstruction efficiency is very compromised when the bars are allowed to become wide. A bar thickness of 1 cm was used in all of the ECal simulations shown [63].

connections to the TFB being set up after ECal installation in ND280.

Each ECal layer was framed by aluminium bars with an L-shaped cross-section. Initial construction of the layers consisted of screwing the aluminium frame into an assembly table. After applying a two-part epoxy to the scintillator bars, the bars were laid inside the aluminium frame such that the WLS fibre hole was aligned with a 2 mm hole in the frame. To ensure misalignment of the holes did not occur during layer construction, a Teflon-coated locator pin was temporarily inserted through the frame and scintillator bar holes. Once the layer of scintillator bars were secured in the aluminium frame, a thin layer of epoxy was applied to the bars and the lead sheets were then hoisted on top via a vacuum lifting rig attached to a crane to evenly distribute the lead sheet's weight. An example of a completed DS ECal frame is shown in Fig. 3.10. After layer construction, the entire layer was vacuum wrapped with plastic and the trapped air subsequently evacuated using a vacuum pump. This procedure allowed the epoxy to cure under vacuum compression conditions. After 12 hours of vacuum curing time, the entire layer was unwrapped, and the WLS fibre holes were studied to check for epoxy blockage. The completed layers were then stored for use.

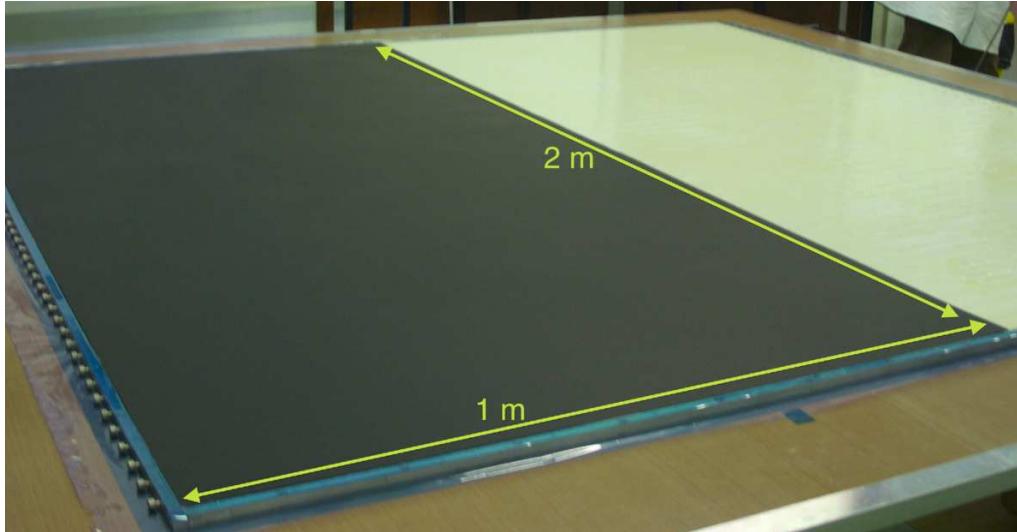


Figure 3.10: A DS ECal layer under construction. The two sheets of lead have been placed on top of the layer of scintillator bars. The aluminium frame and locator pins, which secure the scintillator bars, are visible [60].

The initial construction phase of an ECal module was to take an aluminium bulk-head (which would later hold the module layers) and attach a carbon fibre panel on one end to form an open box. The open box was placed on a construction table, with the opening facing upwards, and the first module layer was lowered inside. To ensure an even 1 cm gap occurred between the bulkhead and the module layer, grub screws were inserted through holes in the bulkhead and pressed against the layer's aluminium frame. WLS fibres were then inserted into each scintillator bar followed by the bonding of a MPPC-fibre connection ferrule to the inserted WLS fibre. A set of test MPPCs with a well understood response were then coupled to the WLS fibre using a connection sheath and also connected to TFBs via a mini-coaxial cable. The inserted layer was then made light tight so testing could begin. The testing device was a 2D scanner [60] which placed a ^{137}Cs radioactive source at multiple points above each constituent scintillator bar in the ECal layer under testing. The measurements taken at each point along the bar were used to obtain an attenuation profile, an example of which is shown in Fig. 3.11 for a single DS ECal bar. After the scanning checks were performed, the test MPPCs were removed and the next layer was installed and tested. This process was repeated until all layers were installed in the ECal module. Finally, after the layer-by-layer construction, the production MPPCs were attached to the ferrule with a foam spring installed in between to ensure a good connection. The cabling

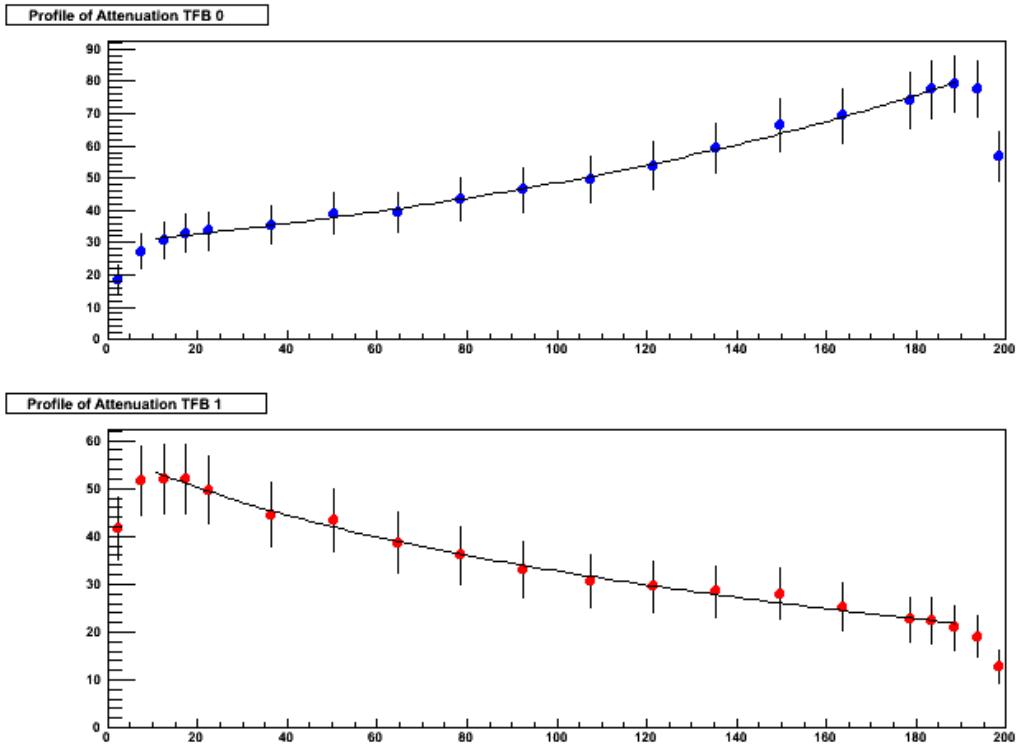


Figure 3.11: An example attenuation profile of a DS ECal scintillator bar, formed using the 2D scanner. Each plot refers to the readout from a MPPC, connected to either end of the scintillator bar. The x -axis of both plots is the position of the 2D scanner along the bar, recorded in cm. The y -axis of both plots is the ratio of the integrated light yield recorded by the MPPC with the 2D scanner source present, to the integrated light yield without the source. The points are the recorded data and the solid curves are exponential fits to the central portions of the data [60].

attached to the MPPCs was grouped together for easier connection management when attaching to the TFBs in the next step. The cooling panels were then assembled such that the TFB boards could be mounted to them. After mounting each of the TFBs via the cooling panels to the ECal module, the correct MPPC cables were connected to the mounted TFB. This procedure was repeated until all required TFBs were mounted. The outer cover panels for the ECal module were then attached with the necessary holes to allow the TFB wiring to escape. The RMMs were then fitted to the outer cover panel and subsequently connected with the corresponding TFBs.

3.2.3.1.3 ECal electronics

As briefly described in section 3.2.3.1, the ECal is equipped with a set of bespoke readout electronics. The MPPCs which collect scintillation light are connected to a set of TFBs. Each TFB contains four Application-Specific Integration Circuits (ASICs) which accept inputs from 16 MPPCs which means each TFB reads up to 64 MPPC channels. Each MPPC is split into high-gain and low-gain channels (a 1:10 split) which are read by different ASIC channels. The TFB chip integrates the collected charge in a programmable time window which comes with a programmable reset time. Once charge integration has occurred, the readout data is pushed onto ADCs for data digitisation. Readout from the MPPC high-gain channel is used to form a timestamp by routing the signal through a discriminator (contained in the TFB chip) which is then subsequently passed to a Field Programmable Gate Array (FPGA). All of the collected TFB information is then passed to the RMMs which provide the readout. Each RMM accepts signal and controls up to 48 TFBs. The controlling RMM distributes timing information and trigger signals and in return receives the TFB data after a trigger issue. RMM data is sent via ethernet to standard computer terminals where it can be processed.

3.2.3.2 Data acquisition system

ND280 comes equipped with a DAQ which is responsible for triggering the readout of information from the subdetectors and subsequent storage. Because of the low frequency of neutrino events, there are no strict trigger requirements which is in stark contrast to collider experiments. So, there are only three triggers which are:

- **Beam trigger:** When a beam spill occurs, a timing signal is sent to the DAQ which issues a command to record ND280 data.
- **Trip-T cosmic trigger:** If hits are seen on opposite sides of the outer detectors (allowed combinations are top and bottom SMRD, left and right SMRD, PØD and DS ECal) which are outside of the beam time window, then the DAQ records data as the hits were likely caused by a cosmic ray muon.
- **FGD cosmic trigger:** If hits are seen in both FGDs which are outside of the beam time window, then data is recorded as this was also likely to be caused by a cosmic ray muon.

The data is initially stored at KEK in Japan but is then replicated to TRIUMF in Canada and RAL in the UK for maximum redundancy and ease of access.

3.2.4 The far detector

The T2K experiment's far detector is Super-Kamiokande [64], which is a very large water Cherenkov detector containing 50 kton of ultra-pure water. Positioned 295 km away from the J-PARC neutrino beam, SK is located under Mt. Ikenoyama with a rock overburden of 1 km (2.7 km water-equivalent). SK consists of two detectors; the inner detector consists of 11,146 inward facing 20" PMTs which surround 35,000 ton of water while the outer detector consists of 1,885 outward facing 8" PMTs and acts as a veto for entering backgrounds.

SK detects particles via Cherenkov radiation which is produced as a result of charged particles travelling in excess of the speed of light in the local medium. This radiation is emitted at an angle of $\cos \theta = c/nv$, where n is the refractive index of the material and v is the speed of the particle. For water this equates to an angle of 42° . The medium imposes a damping effect on the velocity of the particle, which results in energy loss and there comes a point where the Cherenkov emission condition is no longer met. So, SK detects a ring of light emitted by the particles propagating through it.

The features of the Cherenkov ring are utilised for particle identification. An electron, which showers upon creation in the water, creates a fuzzy ring of light while a muon, which cleanly propagates through the water, creates a sharp ring of light. Examples of this are shown in Fig. 3.12 where the difference between muon and electron events can be clearly seen.

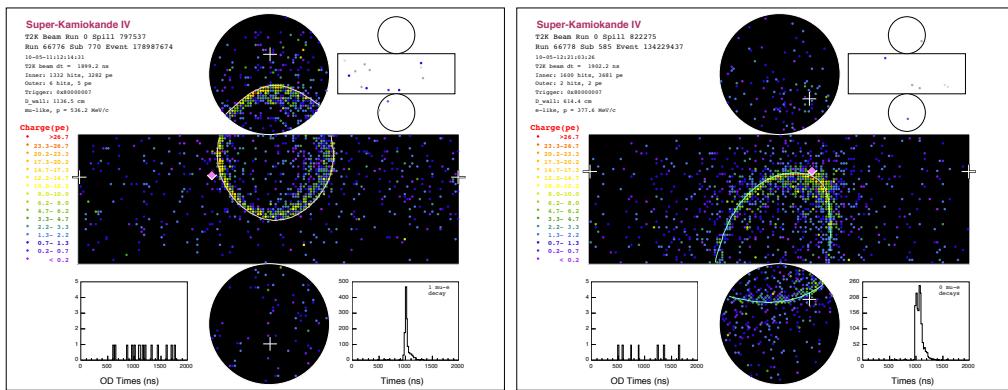


Figure 3.12: Example of reconstructed T2K events in SK for a muon-like ring (left) and an electron-like ring (right) [52]. Because of its MIP-like nature, the muon Cherenkov ring appears sharper than the electron ring.

Chapter 4

ND280 software and existing ECal event reconstruction

The T2K experiment uses a bespoke software suite for simulation and analysis of ND280 data, which is based on the ROOT framework [65]. The vast majority of the ND280 software suite utilises the oaEvent library which provides a unified framework for information manipulation and was specifically designed for this purpose. As ND280 consists of many detectors each providing a specific function, the ND280 software suite is designed to reflect this. Not only are there specific software modules for individual subdetectors, there are specific modules for each phase of the subdetector information processing e.g. Trip-T calibration, TPC reconstruction etc.

As the software suite handles both production of simulated data and the processing of collected data, there are sections of the software chain which are specific to type of data being processed. While the Monte Carlo simulation and real data do see different areas of the software chain, the general philosophy is to manipulate the Monte Carlo or the real data to a point where they can be treated as equals and them process them as such. So, the description of the software will follow the same path: the Monte Carlo and real data specifics will be discussed first and then the unified treatment will follow.

4.1 Monte Carlo production software

As described above, parts of the software chain are unique to simulated data processing. Specifically, the simulation of the beam and the detector response need to be modelled before the Monte Carlo can be treated on equal footing with the real data. This special processing is split into several steps, all of which are described below.

4.1.1 Neutrino flux simulation

The neutrino flux simulation uses Fluka2011 [66] and a software set called JNUBEAM to model the J-PARC neutrino beam. The process begins by using Fluka2011 to simulate the 30 GeV protons incident on the graphite target and their subsequent secondary interactions which produce the neutrinos parent mesons. The kinematic information of the hadrons is then passed to the JNUBEAM simulation. JNUBEAM is based on GEANT3 [67] and models the J-PARC secondary beamline. The hadrons are tracked through the decay volume and are allowed to interact or decay to produce the simulated neutrino beam. Importantly, all information associated with the daughter neutrinos and their parents are saved at this point. By storing this information, the neutrino flux can be readily re-weighted to include new information associated with beam profile measurements or external data.

The main external tuning source is NA61/SHINE which is a hadron interaction experiment that uses a 31 GeV/c proton beam colliding with changeable targets [68]. For use in the T2K flux simulation, NA61/SHINE has collected data using two graphite targets: one with a 4% nuclear interaction length thickness and a full T2K replica target. The flux simulation used for this analysis is tuned using full replica target data. Observed differences between the Fluka2011 simulation and NA61/SHINE data are used to re-weight the simulated neutrino flux.

Additionally to the external data, tuning measurements of the T2K beam profile are also used to re-weight the flux. By making such measurements on a run-by-run basis, the simulated flux can be re-weighted to better model variations of the neutrino beam in each data run.

4.1.2 Neutrino interaction simulation

After the neutrino flux has been modelled, simulation of the neutrino interactions with the T2K detectors follows. The NEUT [69] event generator is used to simulate interactions with ND280. The inputs to the interaction simulation are a neutrino vector file produced by the beam simulation and a ROOT based ND280 geometry. The used geometry includes the magnetic field return yoke and everything contained within. Using the inputs, NEUT tracks the neutrino and calculates the probability of interaction for every material it crosses. To calculate the interaction probability, the potential interaction nucleus must be modelled. For this, NEUT uses two models; the Moniz-Smith Relativistic Fermi Gas (RFG) [70] and the O. Benhar spectral function models [71]. The spectral functions are only implemented for carbon, oxygen and iron so the model used depends on the atomic number of the interaction candidate nucleus. It is important to note at this point that this thesis deals with neutrino interactions on lead, so it is the RFG model that is used for signal interactions.

The main interactions modes at T2K energies are quasi-elastic scattering (CCQE), single pion production (CC 1π) and Deep Inelastic Scattering (DIS) all of which have models in NEUT [72–74].

After the initial interactions, the final step is to simulate the final state interactions within the nucleus. Each particle involved in the interaction is pushed through the nucleus in discrete steps with the probability of a final state interaction being calculated at each step. If an interaction occurs, the final states of that interaction are also included in the subsequent steps. This interactive procedure models the particle cascade until all the final states have reached the nucleus boundary. At this point, all final state particles are recorded along with all of the information that created those particles. This information is stored in a vector file and passed on to the ND280 detector MC package which handles the detector's response to these final state particles.

While the above description provides a general overview of the NEUT-based simulation of neutrino interactions, the above only describes simulation of NEUT events within the ND280 detector itself. In reality, many interactions occur in the pit which surrounds the near detector, some of which have final state muons which enter ND280. So, a separate NEUT-based simulation of neutrino interactions from the T2K beam in the ND280 pit and the surrounding substrate are also generated. This kind simulation

will be referred to as sand MC (because the interaction target is largely sand in the surrounding pit) from now on.

4.1.3 ND280 detector simulation

The simulation of the final state particles in ND280 is handled by `nd280mc` which is based on GEANT4 [75] and ROOT. The neutrino interaction vector files are taken as input and used as seeds in the detector simulation. The neutrino vector inputs are not organised according to the J-PARC beam bunch structure so the detector simulation first groups the interactions into spills. The beam intensity to be simulated is used to define how many interactions occur in a spill with Poisson fluctuations applied to that number. The timing of the beam bunch structure is then used to group the interactions into bunches.

`nd280mc` constructs a ROOT geometry of ND280 based on the design specifications of its subdetectors and then propagates the particles given to it by the neutrino generator through the geometry, simulating energy deposition, scattering and particle decay during propagation.

The sand MC is largely treated the same at this point; the surrounding pit geometry (and ND280) is constructed in `nd280mc` and the corresponding sand MC vectors provide the final state particles which are propagated by Geant4 through the geometry. After this point, the sand MC is treated identically to the beam MC described above.

4.1.4 Detector response simulation

The next and final stage of the MC-only software chain is to model how the detector responds to the simulated particles propagating through it. The detector response software, named `elecSim`, takes the output of `nd280mc` and models how the active regions of the detector would respond given an energy deposition in that region of the detector. In the case of the calorimeters, `elecSim` handles the production of light produced by the constituent scintillator bars, how the light is propagated along the wavelength-shifting fibres and how the MPPCs would respond to the incident photons. For the TPCs, the drift of the ionisation electrons through the gas and the subsequent response of the MicroMEGAS which receive them. In all cases, the readout electronics

response is simulated to produce a data-like output format. This step concludes the section of the software chain which is specific to the MC.

4.2 Real data processing software

The real data specific section of the software chain is very short. Physics events deemed worth saving by any of the ND280 triggers are recorded by the detector and then saved for processing. The MIDAS file format is used for storing the saved ND280 events. All of the relevant information needed to process the event is stored in the MIDAS file, so the only unique step to the data processing is the conversion of the MIDAS file to the oaEvent format. After the step, the oaEvent data files are exposed to the same software as MC files outputted by elecSim (see section 4.1.4).

4.3 Main software chain

The aim of the rest of the software chain is to process the readout from the detector, be it simulated or collected data, and process it so essential information about the physics of the event can be extracted. This is separated into three steps: calibration, reconstruction and data reduction/summarising.

4.3.1 Detector calibration

The software package responsible for overseeing all aspects of the calibration stage is called oaCalib. This controlling package passes the digitised signal from ND280 to dedicated calibration packages for the various kinds of readout electronics. All of the information which can be extracted from an ECal event relies on the charge read by the MPPCs. Thus, for the Trip-T detectors, like the ECals, the main aim of the calibration is to remove all electronic effects so that an accurate estimation of the charge read by the MPPCs can be made.

All MPPCs deliberately output a non-zero ADC count in the absence of signal. This output is known as the pedestal and occurs to address low charge noise and also to avoid the non-linear response of the lowest channels on the ADC. Example pedestal

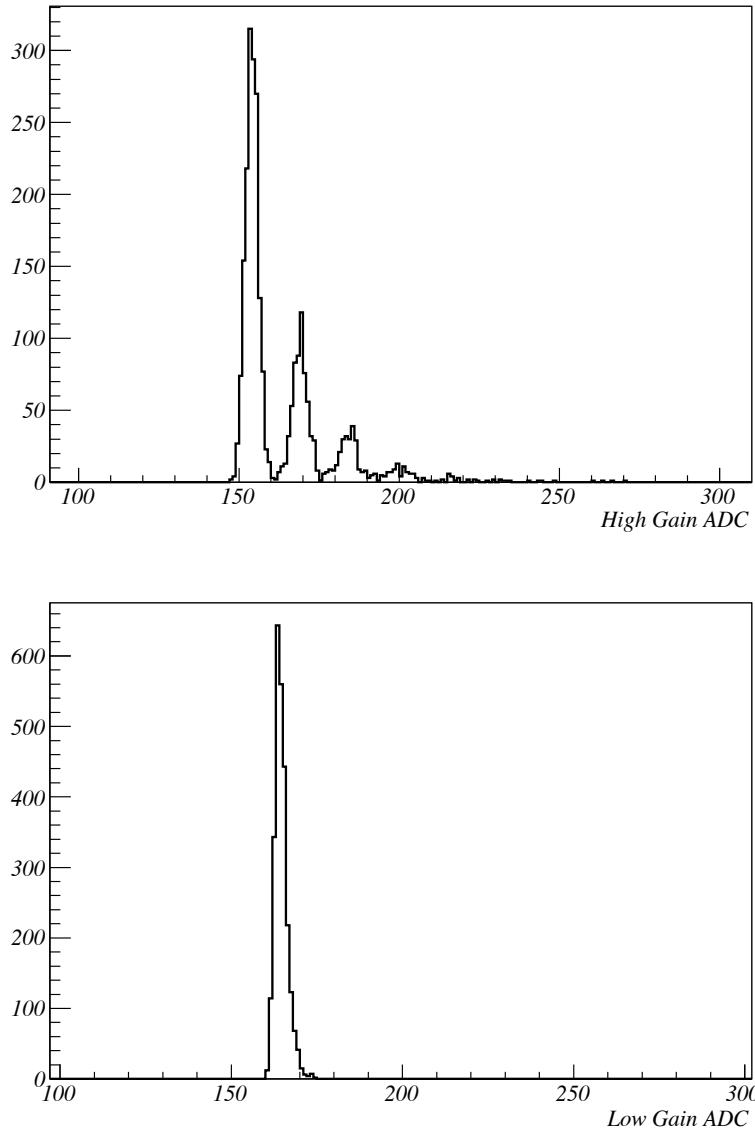


Figure 4.1: Example pedestals in the high (top) and low (bottom) channels for an MPPC. The photo-electron peaks are clearly visible in the high gain channel. The pedestal value is calculated by fitting a Gaussian function to a 9.9 ADC count range, centered on the mode of the gain distribution [76].

peaks for an MPPC are shown in Fig. 4.1. The high-gain channels shows a distinctive multi-peak structure which are photo-electron peaks caused by the background rate of individual pixels firing in the MPPC. The gain of the low-gain channels is roughly an order of magnitude lower which causes the photo-electron peaks to merge with the pedestal.

Each MPPC channel has a different pedestal, all of which vary in time. To calcu-

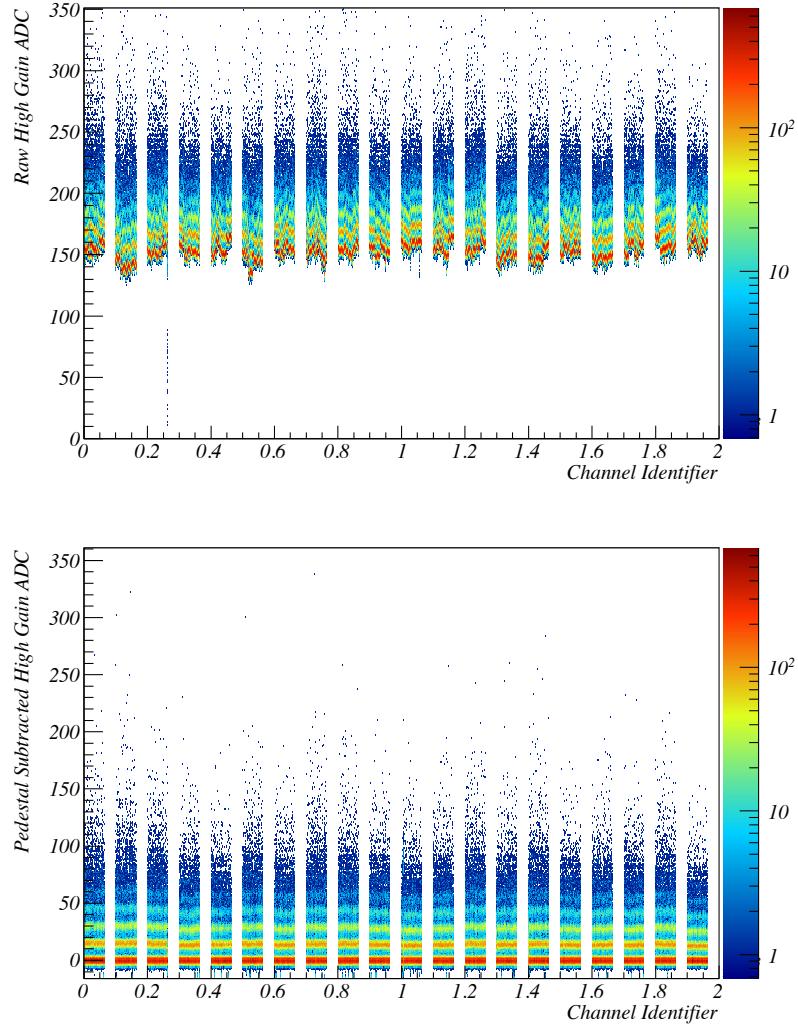


Figure 4.2: The effect of the pedestal subtraction on the charge readout by MPPC channels in the DS ECal. The top distribution shows the raw ADC collected by the MPPC channels while the bottom distribution shows the same ADC readout after the pedestal subtraction has been applied [76].

late the pedestal value for an MPPC, a Gaussian function is fit to a 9.9 ADC count range which is centered on the distribution mode; the mean value of the Gaussian fit defines the pedestal estimate and is subtracted from the raw charge collected by the MPPC. An example output of the subtraction is shown in Fig. 4.2 for the DS ECal MPPC channels.

To estimate the number of photo-electrons produced in an MPPC, The readout charge

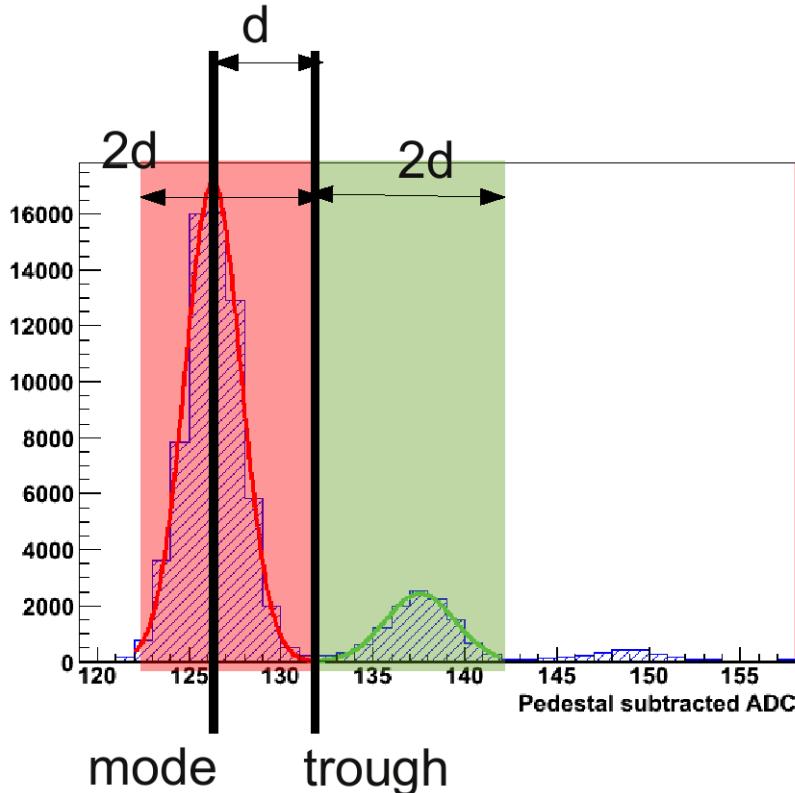


Figure 4.3: Example of the method used to calculate the gains for an MPPC channel. The shaded regions, defined by calculating the mode and trough positions, are the ranges used to the Gaussian fits to the pedestal and first photo-electron peak [76].

is divided by the MPPC gain. The breakdown voltage of an MPPC linearly varies with temperature (approximately 50 mV/°C). This translates to a percent level variation in the MPPC gain per degree. This means small variations in temperature can cause a significant variation in the MPPC gain. So, the gains for each channel are calculated every few hours during running and stored individually. As shown in Fig. 4.1, the first photo-electron peaks are clearly separated. The magnitude of this separation is defined by the gain of the MPPC channel and so a measurement of this separation provides a measurement of the channel gain. To do this, a simple peak finding algorithm is used to find both peaks by stepping upwards from the mode of the MPPC ADC distribution to find a trough by calculation of the local gradient using several nearby bins. After location of the trough, a Gaussian fit is applied to either side of the trough. The difference in the mean of the two Gaussian fits defines the MPPC gain. An example of this procedure is shown in Fig. 4.3.

The time of the signal read by the MPPC is charge dependent and so it relies on the above corrections. However, there are extra steps needed to get an accurate estimate of the timestamp. When a signal is sent from the MPPC to a TFB, the Trip-T discriminator sets the timestamp after the charging capacitor passes a defined threshold. The charge on a capacitor Q is a function of its capacitance C and the voltage V and is defined as

$$Q = CV. \quad (4.1)$$

The collected charge on the capacitor passes the Trip-T discriminator threshold Q_{Th} after time t_{Th} . Q_{Th} [76] is defined as

$$Q_{\text{Th}} = CV_0 \left(1 - e^{\frac{t_{\text{Th}}}{\tau}}\right). \quad (4.2)$$

where τ is the decay constant. Consider three identical capacitors held at different voltages which correspond to final charge values Q_1 , Q_2 and Q_3 where $Q_1 < Q_2 < Q_3$. Fig. 4.4 shows the time it takes for each of the hypothetical capacitors to reach their final charge. The black dashed line in Fig 4.4 represents the threshold imposed by a Trip-T discriminator. As can be clearly seen from the figure, the time taken to pass the threshold depends on the final charge value of the capacitor. This effect is called the electronics time-walk effect and requires correction.

Charge Injection (CI) runs are taken in the ECal to assess and correct the electronics time-walk effect. Each CI run produces several cycle of CI. For each cycle, a well known charge is injected into the Trip-T channels and the average timestamp is recorded. Afterwards, the input charge is increased and the next CI cycle starts. It is then necessary to fit the collected CI data to correct for the time-walk. To understand the fitting function, it is first necessary to re-arrange equation 4.2 such that

$$t_{\text{Th}} = -\tau \ln \left(1 - \frac{Q_{\text{Th}}}{Q}\right). \quad (4.3)$$

Considering the model defined in equation 4.3, the function to be fit to the collected CI data is

$$A + B \times \ln \left(1 - \frac{\text{ADC}_{\text{Th}}}{\text{ADC} - \text{Pedestal}}\right), \quad (4.4)$$

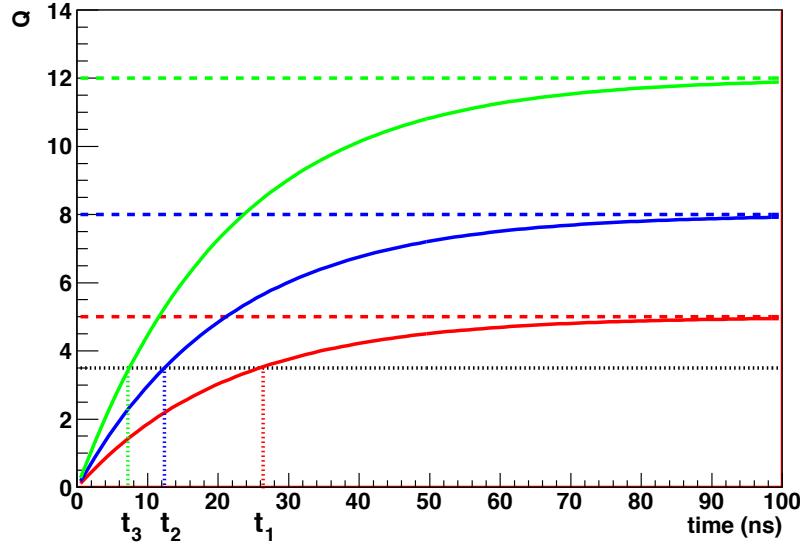


Figure 4.4: Charge as a function of time for three identical capacitor with different final charging values [76]. The black dashed line represents the threshold imposed by a Trip-T discriminator. A capacitor which has a higher final charge value passes the threshold faster than a capacitor with a lower final charge value.

where A and B are the two free parameters to be fitted, ADC_{Th} is the threshold of the discriminator in ADC counts (38 ADC counts for the ECal) and $\text{ADC} - \text{Pedestal}$ is the pedestal-subtracted ADC counts collected during the CI. The timestamp of each TFB channel is reduced according to its high-gain ADC measurement and its timestamp-vs-ADC dependency.

The scintillation light in the WLS fibre emit light according to an exponential decay function. Because the Trip-T discriminator imposes a minimum charge threshold before assigning a timestamp, the WLS fibre emission causes a separate timewalk effect, known as the fibre timewalk effect. The length of the delay before the channel passing the charge threshold depends on the fibre time constant, the discriminator threshold and the number of photo-electrons. The fibre timewalk effect is statistical which means that identical events would not see the same delay. The probability that the n^{th} photo-electron out of N total photo-electrons to be detected at time t is defined as

$$P = e^{-t \frac{(N-n-1)}{\tau}} \left(1 - e^{-\frac{t}{\tau}}\right)^{n-1} \quad (4.5)$$

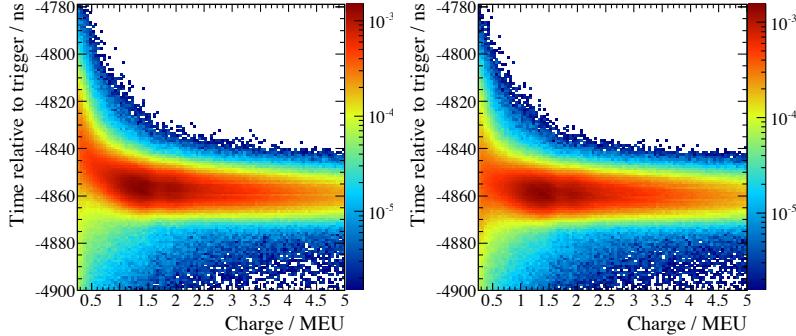


Figure 4.5: The effect of the timewalk correction on MPPC charges and timestamps in DS ECal data [76]. The left and right distributions show the data before and after the fibre timewalk correction has been applied. There is a distinct flattening of the charge-vs-timestamp dependence after application of the correction. All data shown are recorded during Trip-T cosmic triggers.

where τ is the fibre decay constant (7 ns) [76]. It is worth noting that the discriminator threshold can be measured as a number of photo-electrons and so n also defines the discriminator threshold ($n = 4$ photo-electrons in the ECal). The fibre timewalk is defined as the first moment of equation 4.5, $\langle t \rangle$. Inputting $n = 4$ photo-electrons for the ECal, $\langle t \rangle$ is

$$\langle t \rangle = \frac{2\tau(2N^3 - 9N^2 + 11N - 3)}{N(N-1)(N-2)(N-3)}. \quad (4.6)$$

Fig. 4.5 shows the effect of applying the fibre timewalk correction to cosmic data collected in the DS ECal. The left distribution in Fig 4.5 shows that lower recorded charges arrive later in the trigger window which is the key characteristic of the fibre timewalk effect. The right hand distribution in Fig. 4.5 shows the same collected charges after applying the fibre timewalk correction; there is a distinct flattening of the MPPC timestamp on the MPPC charge.

4.3.2 Event reconstruction

Once the information has been calibrated it is passed to the ND280 reconstruction software. The reconstruction algorithms are separated into two phases: the local reconstruction and the global reconstruction. The local reconstruction, which is run first, is separated into a set of algorithms specific to each subdetector. Each subdetector reconstruction attempts to form its own picture of the event which passed through it.

After this, all of the local reconstruction information is passed to the global reconstruction which attempt to match and refit the subdetector results to maximise the amount of extractable information.

4.3.3 Data reduction and summarising

The final stage of the processing chain is to reduce and summarise all of the previous steps in the chain such that they are in a suitable format for the analyser. A lot of the information, particularly from the reconstruction stage, is stripped out and the physics related objects are extracted. Most importantly, all of the information is summarised in a pure ROOT format, meaning the analyser does not rely on the oaEvent library to analyse ND280 data.

4.4 ECal event reconstruction

As mentioned in section 4.3.2, part of the reconstruction process is to run algorithms specific to the ND280 subdetectors. The ECal is no exception and is equipped with an extensive suite of reconstruction algorithms designed to reconstruct events originating from the Tracker region of ND280. The inputs to the reconstruction are the calibrated scintillators bar hits which are used to form 3D objects and attach a topology hypothesis.

4.4.1 Hit preparation

The initial ECal reconstruction stage takes the hits outputted by the calibration stage and prepares them for the downstream algorithms. The ECal hits arrive with timestamps but are not separated according to the bunch structure. So, the hits are ordered in time and then grouped into buckets where a new bucket starts when a greater than 50 ns gap appears between two time-adjacent hits. A second filter is then applied to arrange the hits according to the sensor they occurred on. For double ended bars with both sensors activated, the time of the hit is re-estimated by averaging the timestamp of the two sensors. The two hits are then merged to form a single hit.

It is then necessary to apply extra calibrations to the charges of the hits. The effect

of light attenuation along the WLS fibre is corrected for and a scaling is applied to convert the charge into MIP equivalent units (MEU) where 1 MEU is the most probable value of charge deposited by a Minimally-Ionising Particle (MIP)..

4.4.2 Basic clustering

The time ordered hits are then passed to a set of clustering algorithms which attempts to form an object out of the ECal hits. The first stage of this is called basic clustering which is a nearest neighbour algorithm designed to form 2D clusters of hits for both views of an ECal module. This is initiated by forming a 30 ns window and searching for the highest charge hit contained within it to form a seed. The seed cluster is expanded by searching for and adding candidate hits which pass the following criteria:

- Is located in the 30 ns time window
- Is at most one bar away from a hit in the cluster
- Is at most two layers away from a hit in the cluster

To qualify as a basic cluster, any formed candidate must contain at least three hits. The successfully formed clusters in both views of the ECal modules are then passed to the next stage of the clustering algorithms.

4.4.3 Cluster combination

The second stage of the clustering algorithm takes the basic clusters in a given view and attempts to form merged clusters if a set of conditions are passed. The cluster with the highest number of hits is taken as a seed and compared with all other clusters in the same view. A Principal Components Analysis (PCA) is applied to all 2D clusters as the primary axis is used in one of the merging criteria. A candidate cluster is merged with the seed cluster if:

- The distance of closest approach, taken from the primary axes, of the candidate and seed cluster is less than 80 mm
- The charge weighted average hit times of the candidate cluster is within 40 ns of the seed cluster

- The charge weighted average distance of the candidate cluster to the seed cluster is less than 400 mm

All candidate clusters which pass the conditions are merged with the seed cluster. After all comparisons have been made, a new seed cluster is found and the process is repeated until no more merges take place. All clusters in each ECal view are then passed into the next clustering algorithm.

4.4.4 Cluster expansion

The next and final stage of the 2D clustering attempts to merge any unmatched hits with the already formed clusters if the comparison passes a set of conditions. As before, a PCA is applied to all 2D clusters to calculate the primary and secondary axes as, again, these are used in one of the matching conditions. Let \vec{n}_{pri} and \vec{n}_{sec} be the primary and secondary axes of a cluster found by the PCA. The squared spread of the cluster along the primary or secondary axis is then defined as

$$\sigma_A^2 = \sum_{\alpha=x,y,z} \sigma_\alpha^2 n_{A,\alpha}^2, \quad (4.7)$$

where A refers to either the primary or secondary axis and σ_α refers to the cluster's spread along the Cartesian axes. A metric for defining how well the hit matches to the cluster can then be defined as

$$w = \sqrt{\left(\frac{\vec{H} \cdot \vec{n}_{\text{pri}}}{\sigma_{\text{pri}}}\right)^2 + \left(\frac{\vec{H} \cdot \vec{n}_{\text{sec}}}{\sigma_{\text{sec}}}\right)^2}, \quad (4.8)$$

where H is the vector joining the charge-weighted centre of the cluster and the centre of the hit. The unmatched hit is merged with the 2D cluster if:

- The unmatched hit is within 6 ns of one of the constituent hits of the 2D cluster
- The difference in bar number of the unmatched hit and at least one of the constituents hits in the 2D cluster is less than 11
- The difference in layer number of the unmatched hit and at least one of the constituents hits in the 2D cluster is less than 21
- The calculated value of w is less than 50

Unlike the previous stage where it is possible to merge every cluster in a view, an unmatched hit can be matched with one, and only one, 2D cluster. So, not only does the hit-cluster comparison have to pass the relevant criteria, the comparison has to be a better match than all other comparisons. In cases where the unmatched hit passes all of the conditions with more than one cluster, the comparison which provides the lowest w is selected as the best match.

The 2D clusters in each view are then passed on to the final stage of the clustering algorithms.

4.4.5 3D cluster formation

The final stage of the clustering algorithm attempts to form full 3D clusters using information from both view of an ECal module. To do this, the algorithm takes the 2D clusters from one view and matches them with the 2D clusters in the other view. All 2D clusters from one view are compared to all 2D clusters in the other view and the two that are the best match form a 3D cluster. The metric used for the matching is a simple likelihood based on three input variables. The first variable is the ratio of the total charge of the 2D matching candidate clusters, Q_{ratio} . Assuming the two 2D clusters come from the same particle, the amount of charge deposited in each view should be similar, so Q_{ratio} should be ~ 1 . The Q_{ratio} value is then compared with a probability density function generated using ND280 MC to retrieve $\mathcal{L}_{Q_{\text{ratio}}}$. The second input variable to the likelihood compares the first layer used by the two matching candidate clusters which is closest to the ND280 Tracker region, $\Delta_{\text{layer}, \text{first}}$. If the two candidate clusters are created from the same particle, the difference in starting layer number should ideally be one. As with Q_{ratio} , the $\Delta_{\text{layer}, \text{first}}$ found for the matching candidates is compared to a probability density function to retrieve $\mathcal{L}_{\Delta_{\text{layer}, \text{first}}}$. The final input to the likelihood is almost identical to the second input. The variable, called $\Delta_{\text{layer}, \text{last}}$ compares the difference in layer number furthest away from the ND280 Tracker region. Besides this difference, $\Delta_{\text{layer}, \text{last}}$ is treated in exactly the same manner as $\Delta_{\text{layer}, \text{first}}$ and so will not be described in any more detail.

Once the three inputs for the matching candidates have been calculated, the value of the matching likelihood is then

$$\mathcal{L} = \mathcal{L}_{Q_{\text{ratio}}} \times \mathcal{L}_{\Delta_{\text{layer}, \text{first}}} \times \mathcal{L}_{\Delta_{\text{layer}, \text{last}}}. \quad (4.9)$$

As mentioned above, all 2D clusters in one view are compared with all 2D clusters in the other view. The matching pair which produce the highest value of \mathcal{L} are counted as a match. The matched pair are then removed from the pool and the process is repeated until no more matches can be made or the value of \mathcal{L} falls below e^{-5} . All matched clusters are now counted as 3D objects and are passed on to the rest of the ECal reconstruction algorithms.

4.4.6 3D hit reconstruction

So far in the ECal reconstruction, it has only been possible to reconstruct 2/3 of the hit coordinates. However, now that full 3D clusters have been formed, the information from both ECal views can be used to estimate the third hit coordinate which is the coordinate along the length of the bar.

The hits are separated into their respective 2D view and then the charge-weighted average position of each layer is calculated. Then, to calculate the final hit coordinate, the nearest four layers from the other view are found and a linear least-squares fit is applied to their 2D coordinates. The line found is then extrapolated into the layer of interest which provides the unknown hit coordinate.

4.4.7 Energy reconstruction

The ECal was designed to catch electromagnetic particles originating from the Tracker region of ND280, particularly photons from π^0 decays. An integral part of particle shower reconstruction is estimation of the total energy of the incoming particle. So, the next phase of the reconstruction is particle energy estimation under the assumption that the incident particle created a full contained particle shower upon entering the ECal module. The range of the energy fitter is 25 MeV to 20 GeV which almost fully covers the range of energies seen in the ND280 ECals. The energy fitter works by minimising a likelihood function which takes the total charge, charge RMS and charge skew of the 3D cluster as inputs. For each of the three input variables, splines were generated using simple photon particle gun MC which relates the input variables to the true particle energy. The minimizer generates an estimate of the true energy and, via the splines, retrieves the expected values of the input variables. By minimising the distance between the measured input variables and those taken from the splines an

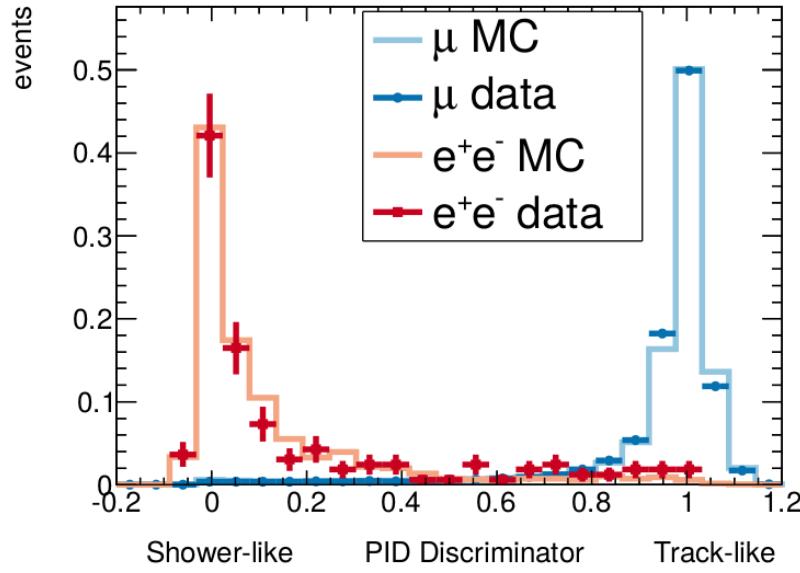


Figure 4.6: The track-like/shower-like discriminator for the DS ECal which is uses an Artificial Neural Network (ANN) [60]. Solid lines show the control sample Monte Carlo (based on NEUT simulation of beam events) and the points show the control sample data (taken from collected beam data). For the muon samples, exactly one track is required in each of the TPCs and the DS ECal with an additional requirement that the TPC tracks are muon-like. For the electron samples, photon pair production is checked in the FGD, requiring tracks with opposite charge with at least one TPC track which must be electron-like. The muon and electron simulated distributions show good agreement with the collected data.

accurate estimate of the particle energy is found. This energy estimate, along with the 3D cluster itself, is then passed on to the final section of the ECal reconstruction.

4.4.8 Event classification

The final section of the ECal reconstruction is to attach a topology hypothesis to the 3D cluster. The routines currently separate the clusters into two categories: those that are shower-like and those that are track-like using an Artificial Neural Network (ANN). This multi-variate analysis object takes multiple pieces of information about the 3D cluster which attempts to measure the energy deposition and shape of the cluster and returns a single number. The value of the returned number suggests whether the reconstructed object is track-like or shower-like. The discriminator was tested using ND280 electron and muon control samples, the results of which are shown for the DS ECal in Fig. 4.6.

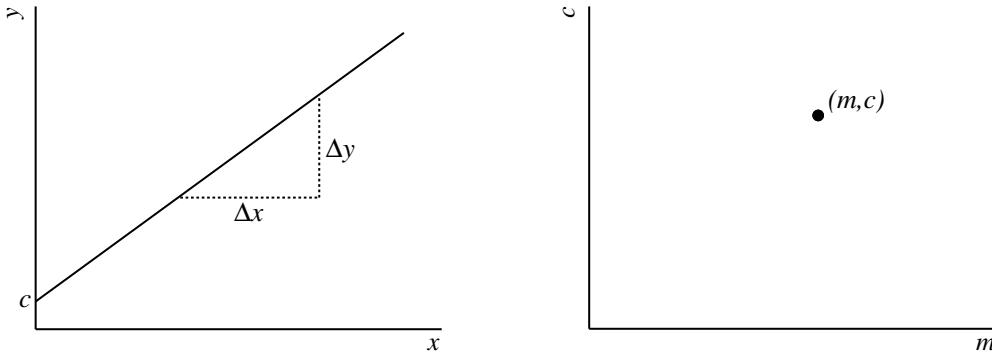
Chapter 5

Enhanced ECal reconstruction

The current implementation of the ECal reconstruction software was designed to reconstruct particles which originate from the ND280 Tracker and enter the ECal. As section 4.4 shows, this was realised by only considering reconstructed ECal clusters under the single track-like or shower-like hypothesis. It should be evident that a neutrino interaction occurring within the ECal does not well fit this topology. While it is true that there is some power in the current reconstruction to distinguish a neutrino interaction from an entering track or shower, there is little feature information available. How many final state particles propagated from the interaction? How much visible energy was deposited by each of the particles? Where in the ECal did the interaction occur? These basic questions can not be trivially answered when using the current reconstruction. To maximise the ability of distinguishing ECal neutrino interactions from entering backgrounds, the reconstruction must be revisited.

5.1 The Hough transform

The Hough transform is a popular method of machine pattern recognition used by, but is not limited to, high energy physics experiments. Originally designed for machine track recognition in bubble chamber pictures [77], the version most widely used throughout the world was developed in 1972 [78]. The Hough transform is used to isolate specific features or shapes from a digital image. The simplest implementation, which is of most interest in event reconstruction, allows the extraction of straight 2D lines from a complex pattern. This is achieved by exploitation of a simple, but remarkable, feature of 2D geometry.



(a) Line in 2D Cartesian space. The line is defined by its intercept with the y -axis, c , and its gradient, $\Delta y / \Delta x = m$ (not shown).

(b) Point representation of the line defined in Fig. 5.1a in 2D parameter space. Because the line is solely defined by its intercept, c , and its gradient, m , the line is defined as a point in this parameter space.

Figure 5.1: Representations of a 2D line in Cartesian space.

5.1.1 Line-point duality

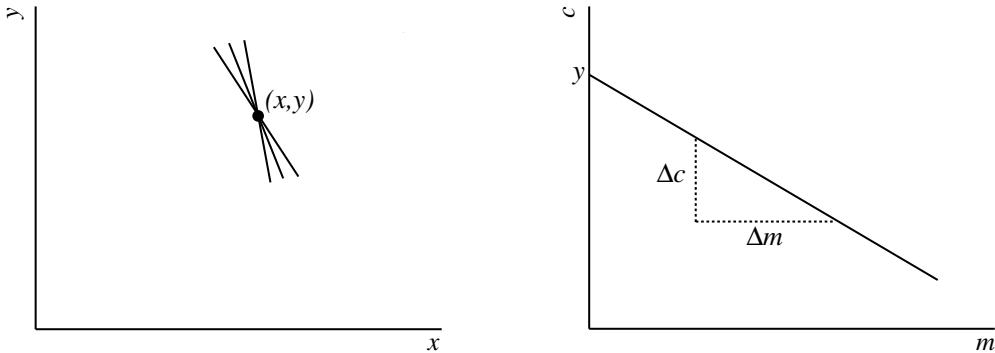
Consider a straight line formed in a 2D Cartesian space as shown in Fig. 5.1a. The line is usually described by

$$y = mx + c \quad (5.1)$$

where y and x are used as coordinates, m is the gradient of the line and c is the intercept location of the line with the y axis. While it is not necessary to analyse this simple shape in great detail, it is important to note that m and c are the only parameters necessary to completely describe the line.

Now consider a new 2D space where the axes are defined by m and c , rather than x and y (hereafter referred to as the parameter space). As this parameter space is described by the parameters of a general 2D Cartesian line, there is an underlying symmetry between the two spaces. The parameters of the 2D line shown in Fig. 5.1a can be used to form a pair of coordinates (m, c) in the parameter space as shown in Fig. 5.1b. It is important here to state clearly the general result; a straight line in Cartesian space is represented by a single point in parameter space.

Now consider the 2D Cartesian space again. Unlike before, we will define a sin-



- (a) Point in 2D Cartesian space. The point can be defined as an infinite set of lines, all of which cross at said point (three example lines are shown).
- (b) Line representation of the point defined in Fig. 5.2a in 2D parameter space. The infinite set of lines which represent the point in Fig. 5.2a follow a relation, $c = -xm + y$, which itself is a line in the parameter space.

Figure 5.2: Representations of a 2D point in Cartesian space.

gle point rather than a straight line. Such a point is traditionally described by a pair of coordinates (x, y) . However, an alternative description of the point is an infinite number of lines all of which pass through (x, y) . This is highlighted by Fig. 5.2a where three lines of the infinite set are shown along with the point they represent. As the infinite line set are used to describe a single point, all lines in the set must follow a pattern. This relationship is revealed by simple algebraic manipulation of equation 5.1 to give

$$c = -xm + y. \quad (5.2)$$

Despite the manipulation, equation 5.2 still resembles the equation of a 2D line; however, the parameters are x and y and the coordinates are m and c . Specifically, equation 5.2 is represented by a line in the parameter space defined above. The gradient of this line is

$$x = \frac{\Delta c}{\Delta m} \quad (5.3)$$

and the intercept of the line with the c axis is y as shown in Fig. 5.2b. As before, it is important to clearly state what has been shown; a point in Cartesian space is represented by a line in parameter space.

5.1.2 The parameter space

As section 5.1.1 shows, there is a clear relationship between the Cartesian space and the parameter space. Specifically, there is a symmetry between lines and points contained in the two spaces. This relationship between the Cartesian and parameter spaces is not only interesting but also very powerful. Consider again the parameter line defined by equation 5.2 and shown in Fig. 5.2b. As shown in section 5.1.1, equation 5.2 was derived by considering the infinite set of lines which pass through a Cartesian point. As this infinite set represents the parameter line, it must also be true that the parameter line represents the infinite line set. Using one of the results from section 5.1.1, any point along the parameter line represents one of the lines from our infinite set. This key feature of the parameter space is the central component of the Hough transform.

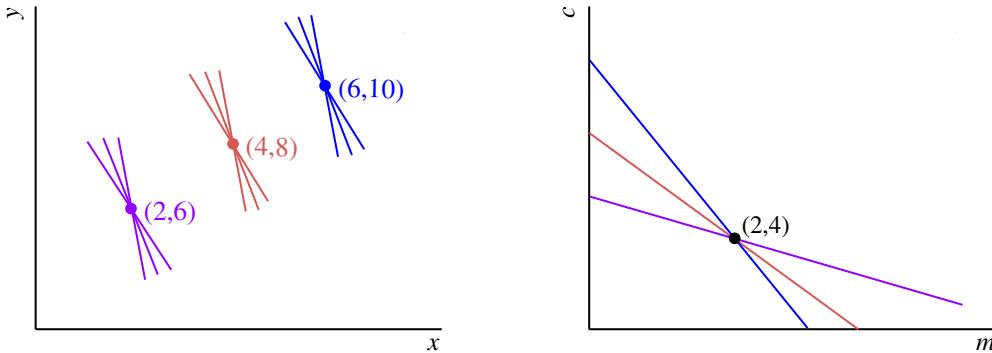
We will now return to the Cartesian space for an example of how the Hough transform works. Let us define three points in this space,

$$\begin{aligned} p_1 &: (2, 6) \\ p_2 &: (4, 8) \\ p_3 &: (6, 10). \end{aligned} \tag{5.4}$$

The three points defined in equation 5.4 are shown in Fig. 5.3a. Using one of the results from section 5.1.1 and equation 5.2, the three points can be Hough-transformed into the following parameter lines:

$$\begin{aligned} c &= -2m + 6 \\ c &= -4m + 8 \\ c &= -6m + 10. \end{aligned} \tag{5.5}$$

The three parameter lines are shown in Fig. 5.3b. From Fig. 5.3b, it is clear that the three parameter lines all cross at a common point with parameter space coordinates $(2, 4)$. Using the results from section 5.1.1, this common point in parameter space is represented by a line in Cartesian space. In addition, as $(2, 4)$ is common to all three parameter lines, the Cartesian line represented by $(2, 4)$ must also pass through all



- (a) The three Cartesian points defined in equation 5.4 along with example lines from each point's infinite line set. The coordinates of each point are also shown.
- (b) The three parameter lines defined in equation 5.5. The coordinates (2, 4) define the point of intersection of the three lines.

Figure 5.3: The three points defined in equation 5.4 and their representation in the parameter space. The colour coding matches the Cartesian points to their respective parameter lines.

three Cartesian points defined by equation 5.4. This new Cartesian line is defined by

$$y = 2x + 4 \quad (5.6)$$

and is shown in Fig. 5.4 with the original points used to generate the parameter lines. While this example is relatively simple, it demonstrates the capability of the Hough transform to recognise linear patterns in sets of points.

5.1.3 Redefinition of parameters

Unfortunately, a complication in computation arises when $m \rightarrow \infty$. However, this complication can be removed by redefining the line parameters. An alternative 2D line parameterisation is to specify a line in terms of the angle it makes with the x axis, θ , and the perpendicular distance of the line from the origin, ρ , as illustrated in Fig. 5.5. The functional form of the Cartesian line becomes

$$y = x \tan \theta + \frac{\rho}{\cos \theta}. \quad (5.7)$$

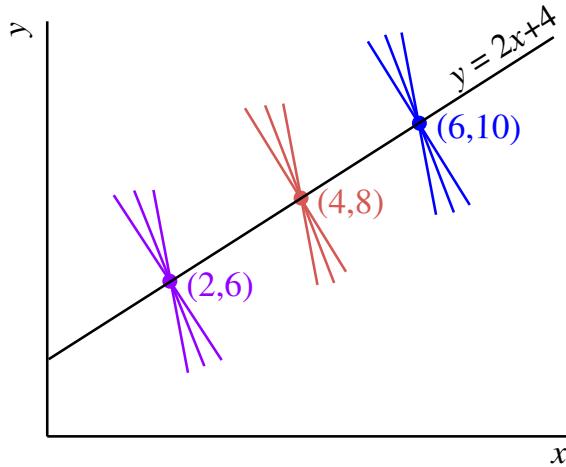


Figure 5.4: The line represented by the parameter space intersection in Fig. 5.3b with the Cartesian points it intercepts.

Using this new parameterisation, we must also define a new parameter space with axes θ and ρ . Using equation 5.7, a line in this new parameter space is defined by

$$\rho = y \cos \theta - x \sin \theta. \quad (5.8)$$

By definition, the θ axis of the parameter space must be bounded to the interval $[0, 2\pi]$. If the directionality of the line is meaningless to the analyser, then the interval can be restricted to $[0, \pi]$.

The disadvantage of this parameterisation is that parameter line generation now involves trigonometric calculations which can be computationally expensive. However, this problem is small when compared to the unbounded complication of the traditional parameterisation.

5.1.4 Discretisation of the parameter space

It must now be considered how the parameter space is analysed. When a large number of parameter lines are generated, it becomes computationally expensive to analyse the resultant parameter space. While approaches exist to analyse the parameter space with very high precision [79], it is often only necessary to extract parameters with finite resolution. In such a case, it is convenient to discretise the parameter space. Under this

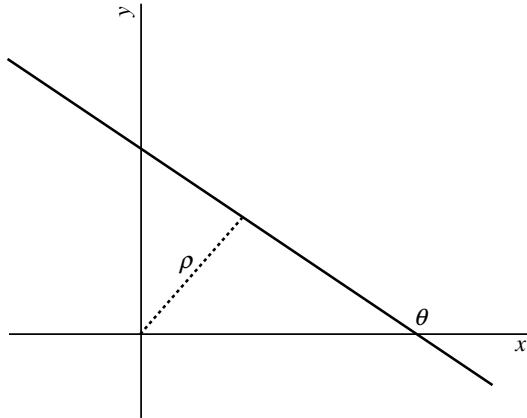


Figure 5.5: θ - ρ parameterisation of a 2D line.

regime, the parameter space is split into θ - ρ bins. Then, a parameter line is generated by incrementing the value of each θ - ρ bin it passes through. After each line has been added to the parameter space, the crossing locations can be readily found by searching for the θ - ρ bins with content larger than unity. An example of this discretisation is illustrated in Fig. 5.6, where the parameter lines defined in equation 5.5 have been re-parameterised using equation 5.8. The content of each bin in Fig. 5.6 records how many of the three parameter lines pass through each bin. The bin with value 3 is the crossing point of the three parameter lines.

If a discretised approach is acceptable, which is the case in event reconstruction, construction and analysis of the parameter space is reduced to filling a 1D array N times, where N is the number of parameter lines, followed by a 1D grid search of the array to find the bin with the highest content.

5.2 ECal application of the Hough transform

We must now address how the Hough transform can be used as a reconstruction tool in the ECal. To do this, let us consider a neutrino interaction which occurs in the ECal as illustrated in Fig. 5.7. While the propagating neutrino is invisible to the ECal, the charged final state are definitely not. To first order, the final state particles propagate in straight lines depositing energy in the scintillator bars as they go. From this, we can

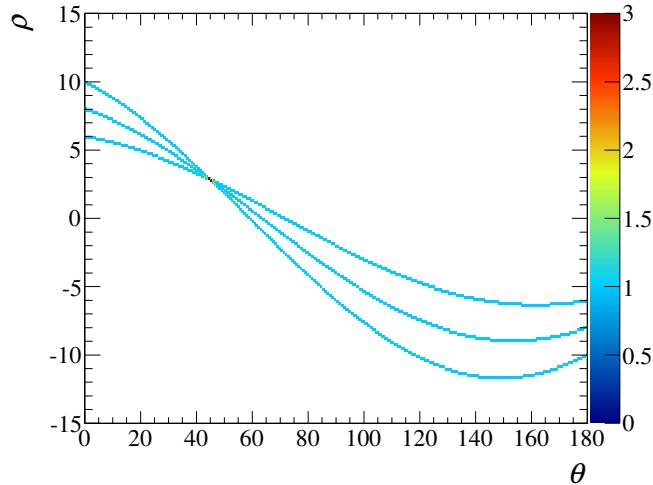


Figure 5.6: The discrete θ - ρ space. The plotted lines are those defined in equation 5.5 and reparameterised using equation 5.8.

infer that the hit bars arranged in straight lines should reveal the trajectory of the final states. As shown above, the Hough transform is capable of identifying straight lines from a set of coordinates. However, there are two complications in the ECal which the above sections have not addressed. We have only specifically discussed how to extract a single straight line from a pattern. As Fig. 5.7 shows, the number of final states can be, and is often, greater than one. This is merely a problem of computation which will be addressed in section 5.2.3. A much more severe problem is that the above demonstrations only deal with patterns constructed from infinitesimal points. While the centre of a scintillator bar can be used as a point for parameter line generation, it is unlikely that a final state particle will pass through the central point of the scintillator bars that it propagates through. If this is not addressed, the Hough transform will be of little use in trajectory reconstruction.

5.2.1 Modelling the ECal bar

To make the Hough transform viable as a reconstruction tool, the finite dimensions of the ECal bar need to be incorporated into the parameter space generation. This feature of the ECal bars would be very problematic if the parameter space was continuous. However, it is only necessary to know the line parameters with finite resolution and a discrete parameter space can be used. This means that the ECal bar can be modelled as a set of Cartesian points and each of said points can be Hough-transformed in turn

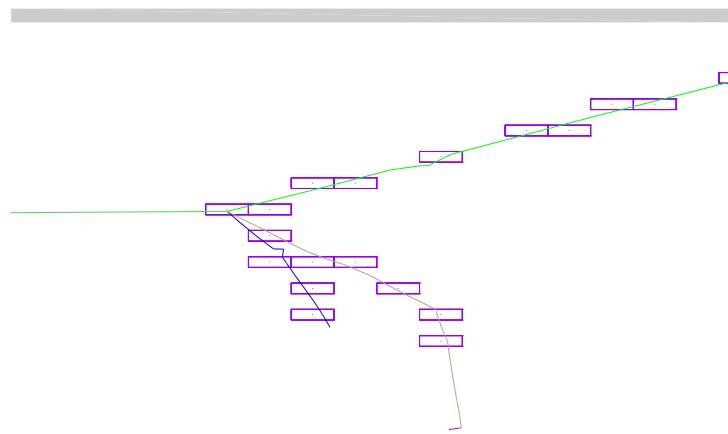


Figure 5.7: Simulated neutrino interaction with 3 final states in the side-right ECal. The green line entering from the left is the ν_μ . The green line exiting to the right is a μ^- , the brown line is a π^+ and the blue line is a proton. The purple rectangles represent the hit ECal bars.

to build up the parameter line representation of the ECal bar.

There are now two steps to consider. Firstly, how should the points be arranged? Remember that the Hough transform of a point represents all of the lines that pass through that point. So, the points should be arranged in such a fashion that any line which passes through the 2D cross-section of the ECal bar also has to pass through one of the points in the configuration. Secondly, the spacings between the points should be small enough that no gaps appear in the generated parameter line.

Assuming that every bin of the parameter space will have a $1^\circ \times 1$ mm area, an obvious choice would be to use a rectangular grid of points with 1 mm spacing superimposed over the 2D cross-section of an ECal bar. The total number of points used to model the ECal bar is 451. To Hough transform the ECal bar, every point in the grid array can be Hough-transformed individually with care taken to ensure that each $\theta - \rho$ bin is filled exactly once. The result is illustrated in Fig. 5.8b. The finite size of the ECal bar is evident by the finite size of the resultant parameter line.

While the generated parameter line accurately represents every line which passes through the ECal bar, there are two problems with this approach. Firstly, the large number of points to be Hough-transformed is very large which results in a long CPU time. Secondly, there is a very high number of redundant calculations involved in

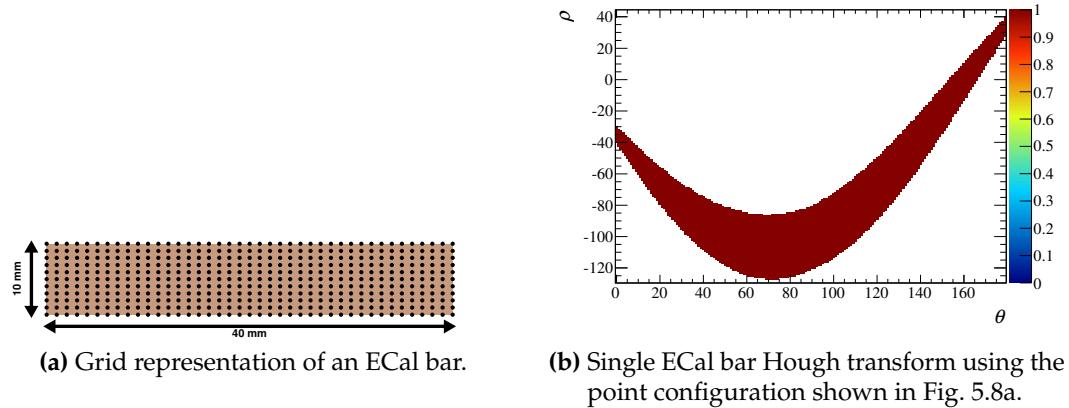


Figure 5.8: The grid representation of an ECal bar and its representation in parameter space.

the parameter line generation. Consider an exactly vertical line which passes through one of the points in the grid array. This line also passes through 10 other points in the same column of the grid. This means that when the parameter line is being generated, this vertical line is calculated 11 times for each column. Bearing this in mind, there are many points along the parameter line which are repeatedly calculated and provide no extra information. This would mean that any algorithm which uses this approach would be very CPU inefficient.

An alternative is to model the ECal bar as a set of points arranged in a cross as shown in Fig. 5.9a. Assuming that the spacing between the points on each line of the cross is infinitesimal, any line which passes through the ECal bar would also have to pass through one of the points in the configuration. As the parameter space is discrete, the spacing between the points need not be infinitesimal but only small enough to ensure that no gaps appear in the parameter line. Using 41 points on each line of the cross, the ECal bar can be Hough-transformed by Hough-transforming each point in the cross configuration. An example of this result is shown in Fig. 5.9b using the same ECal bar used to generate Fig. 5.8b. Clearly, Fig. 5.8b and Fig. 5.9b are identical showing that the cross model achieves the same result as the grid model. Comparing the two, the cross model uses a 90 point representation whereas the grid model uses a 451 point representation. This should mean that an algorithm utilising the cross model would be a factor of five faster than one using a grid model.

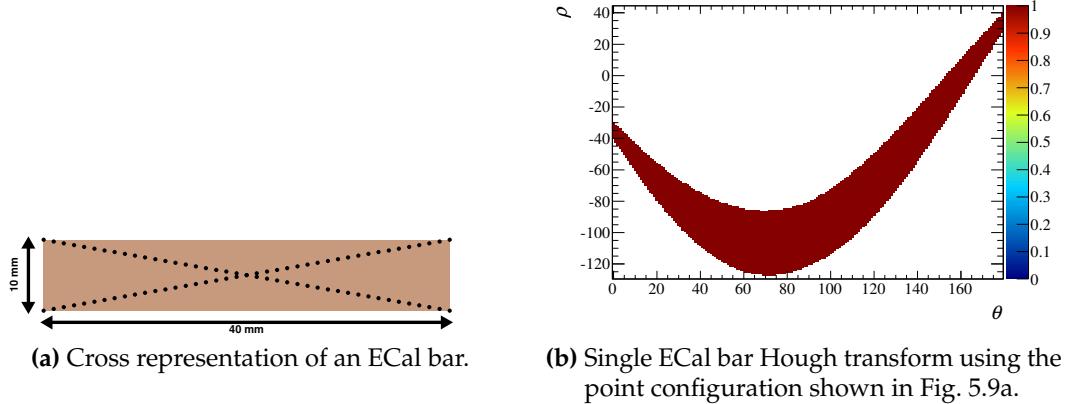


Figure 5.9: The cross representation of an ECal bar and its representation in parameter space.

5.2.2 Parameter space generation

As we have addressed how to Hough-transform an ECal bar, we are now in suitable position to generate the full parameter space for an ECal cluster. As described in section 5.1.3, the parameterisation of the 2D lines requires some point in space to act as the origin. It is possible to use the origin defined by the global ND280 geometry; however, this is located in TPC 1 which would mean ρ will usually be of the order of metres. It is more convenient to define an origin in the vicinity of the ECal cluster being reconstructed. A simple option is to use the charge-weighted centre of the ECal cluster as the origin of the Hough transform. This location is simple to calculate and generally keeps ρ small.

The provided description of the Hough transform in all previous sections is strictly defined in 2D and so the ECal cluster should be split in such a way that this definition can be used. Fortunately, a 3D ECal cluster is built up using the two 2D views that the scintillator layers provide. So, it is relatively easy to split the 3D cluster into a pair of 2D clusters by collecting the cluster's hits into their respective 2D views.

We can now partly answer one of the problems raised in section 5.2 which is how to handle the track multiplicity aspect of the reconstruction. This is partly addressed by generating N parameter spaces with the same θ - ρ bin configuration where N is the number of hits in the 2D cluster. Each of the N parameter spaces will hold one parameter line generated by one of the 2D hits (in a similar fashion to Fig. 5.9b). The final

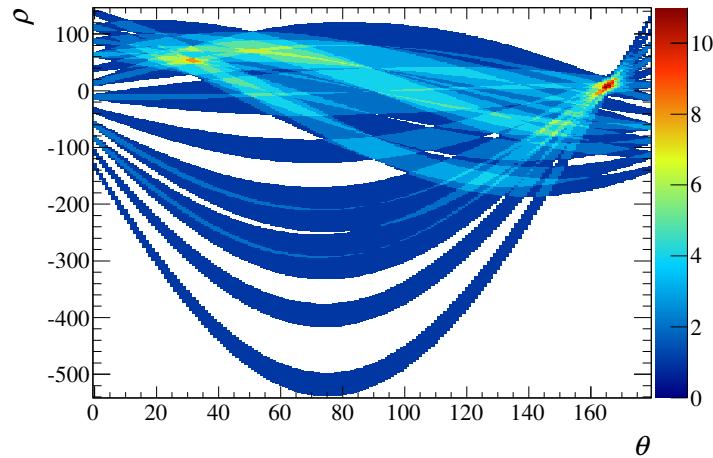


Figure 5.10: The full parameter space of the 2D cluster shown in Fig. 5.7. The height of each point in the parameter space corresponds to the number of hit scintillator bars intersected by the Cartesian line which the parameter space point represents.

parameter space can then be generated by adding together each of the N parameter spaces. The parameter space of the 2D cluster in Fig 5.7 is shown in Fig. 5.10.

5.2.3 Parameter space analysis

The full parameter space can look arbitrarily complicated. However, it contains a vast amount of trajectory related information about the cluster. Every θ - ρ bin of the parameter space describes a 2D track and the content value of said bin describes how many 2D ECal hits the track passes through. As described in section 5.2, a particle's trajectory should be straight in the ECal which means that the particles path should be revealed by finding the most hits arranged in a line. This hit arrangement can be found by finding the bin in the full parameter space with the highest value. The track candidate parameters can the be found by fetching the (θ, ρ) coordinate of the found bin.

While the preferred bin can reveal how many hits the track candidate passed through, it contains no information about which hits were contributors. However, this full parameter space was generated by summing the N parameter spaces discussed in section 5.2.2. So, the contributing hits can be readily found by looking at the same (θ, ρ) bin in each of the N parameter spaces and recording which have a non-zero value. We now have the track candidate's parameters and its contributing hits which is enough

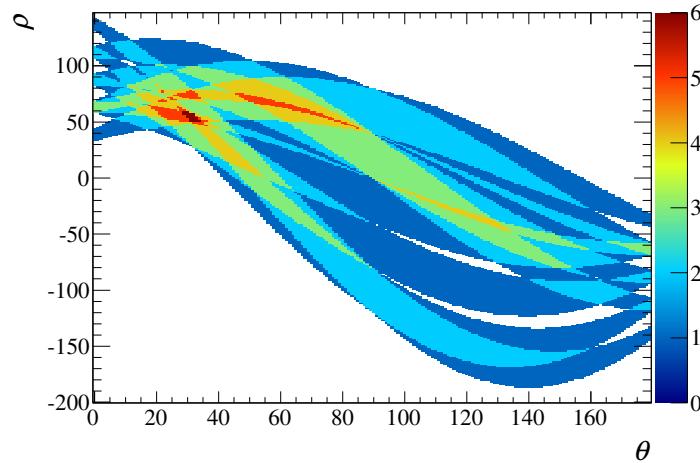


Figure 5.11: The reduced parameter space of the 2D cluster shown in Fig. 5.7. The reduced parameter space was formed by removing all of the hit scintillator bar representations which contributed to the full parameter space’s (see Fig. 5.10) maximum.

to describe the 2D trajectory.

A new search now needs to begin to find any other track candidates. However, repeating the same search of the full parameter space will return the track candidate that has already been found. To find the next track candidate, the presence of the previous track candidate must be removed. So, a reduced parameter space must be generated. The previous step found which of the N parameter lines contributed to the previous track candidate. So, this new parameter space can be formed by subtracting the contributing parameter lines from the full space. An example of this is shown in Fig. 5.11 where the reduced parameter space was formed by subtracting the contributing parameter lines to the highest bin in Fig. 5.10. The next track candidate can then be found by searching for the highest content bin of this reduced parameter space and said bin’s contributing parameter lines.

This process can be repeated until some threshold is reached. This threshold is nominally set by demanding that at least three hits are required to form a track candidate.

5.2.4 2D track quality checks

While the approach outlined above is very powerful for recognising track-like shapes in the 2D ECal clusters, it is not capable of checking whether the selected track candidates are of sound quality. So, external checks need to be done which validate each track as it is returned from the parameter space. Fortunately, the objects returned from parameter space are simple in structure and so the quality checks can be designed to reflect this. Two necessary checks were implemented in the 2D reconstruction:

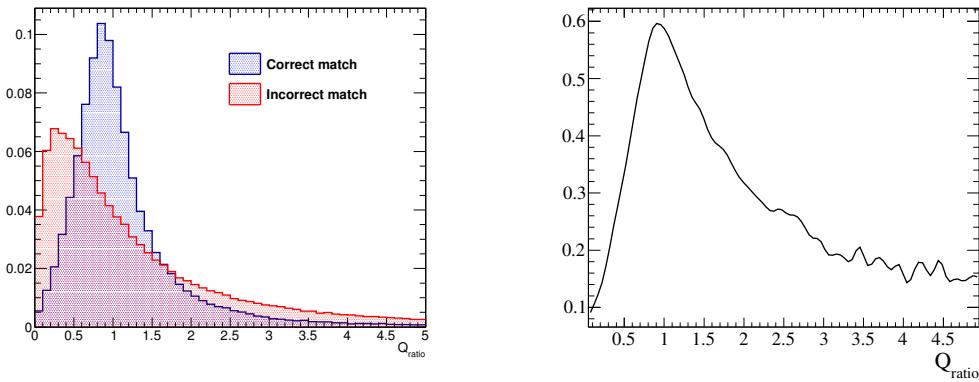
- The track can not skip a scintillator layer in a given view.
- The track can skip a maximum of 1 bar in a scintillator layer.

If a track fails either of the above conditions, the track is flagged as bad and rejected. To ensure that the same bad track is not selected in the next interaction of the parameter space analysis, the track also needs removing from the parameter space. To do this, every bin in the parameter space is checked to see if the bin is filled purely from the hits contained in the bad track. If this is the case, the bin content is set to 0.

5.2.5 3D track reconstruction

Section 5.2.3 describes the track reconstruction of a 2D ECal cluster. However, a 3D cluster consists of two sets of 2D clusters. So, the process described in section 5.2.3 must be performed on each of the 2D clusters. The result of this process is two sets of 2D tracks. To form full 3D tracks, the tracks from each view must now be matched together. This is achieved by making every pairwise comparison of the tracks from each view to find the pair which are most similar to each other. After such a pair is found, the tracks are removed from the pool and the process is repeated to find the next pair until either no tracks are left or one 2D track is left. In the latter case the single 2D track is discarded. Every pairwise combination of tracks is used to form a likelihood \mathcal{L} . The pair which produces the highest \mathcal{L} is declared the best match and removed from the pool. Three pieces of information about the matching pair are used to calculate \mathcal{L} , all of which make use of probability density distributions generated using beam Monte Carlo.

As should be expected, a vertex with one visible track will have different characteristics to a vertex with three visible tracks. So, to maximise the ability of the matcher,



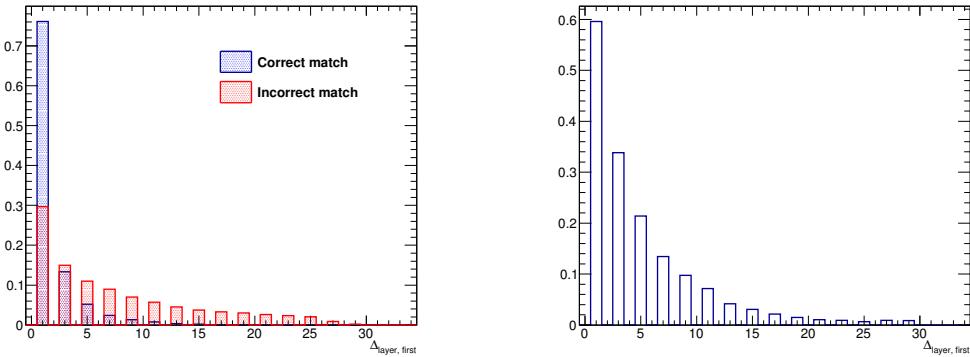
(a) Q_{ratio} distribution (area normalised). The blue and red distributions refers to matching pairs which were matched to the same true particle and different true particles respectively. The two distributions peak at different values of Q_{ratio} , which implies Q_{ratio} provides good separation of the correctly and incorrectly matched events.

(b) Q_{ratio} probability density distribution. The peak of the distribution aligns well with the correct matches peak in Fig. 5.12a which aides in maximising the number of correct matches made.

Figure 5.12: Q_{ratio} and its probability density distribution in the barrel ECal for the two track case. NEUT-based simulation of ND280 beam events were used to produce the distributions.

a different set of probability density distributions are used for the 1, 2 and 3 track cases. If the number of tracks in each view is not identical, the higher number of tracks is used to find the correct probability density distributions. Separately, due to their geometrical differences, the reconstructed shape of vertices in the DS ECal will differ to those in the barrel ECals. This leads to a separate set of probability density distributions for the barrel and DS ECal modules.

The first input to the likelihood is the ratio of the total deposited charge on each track, Q_{ratio} . The denominator is taken as the track which comes from the view with the most hits. Generally speaking, a particle propagating through an ECal module should deposit a similar amount of charge in each of the two views. So, Q_{ratio} should have a value close to 1 if the two 2D tracks are created by the same particle. An example of the Q_{ratio} distribution is shown in Fig. 5.12a, taken from beam Monte Carlo in the barrel ECals for cases where the maximum number of tracks found in a given view is 2. In Fig. 5.12a, correctly matched (in blue) shows the Q_{ratio} distribution for matching pairs which come from the same particle and incorrectly matched (in red) shows the Q_{ratio} distribution for matching pairs which were created by different parti-



(a) $\Delta_{\text{layer}, \text{first}}$ distribution (area normalised). The blue and red distributions refers to matching pairs which were matched to the same true particle and different true particles respectively. While both distributions peak in the same place, the peak for the correct matches distribution is much sharper, indicating a good level of separation between the correctly and incorrectly matches events.

(b) $\Delta_{\text{layer}, \text{first}}$ probability density distribution. The peak of the distribution aligns well with the correct matches peak in Fig. 5.13a which aides in maximising the number of correct matches made.

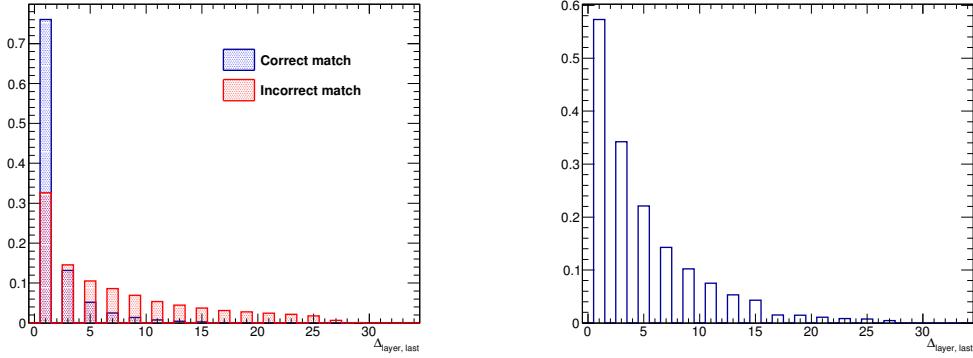
Figure 5.13: $\Delta_{\text{layer}, \text{first}}$ and its probability density distribution in the barrel ECal for the two track case. NEUT-based simulation of ND280 beam events were used to produce the distributions.

cles. As Fig. 5.12a shows, Q_{ratio} well separates the two cases. To generate a probability density distribution for Q_{ratio} , the two distributions shown in Fig. 5.12a are used, but without applying any normalisation. By comparing the bins of each distribution, the probability for correctly matching two tracks in a given bin p_i can be formed by

$$p_i = \frac{s_i}{s_i + b_i}, \quad (5.9)$$

where s_i is the number of correctly matched tracks in bin i and b_i is the number of incorrectly matched tracks in bin i . A discrete probability density distribution for Q_{ratio} can then be formed by calculating p_i for every bin. The discrete probability density distribution is then interpolated with splines to create the final probability distribution. An example of this for the two track, barrel case is shown in Fig. 5.12b. When a matching candidate pair is being considered, the value of Q_{ratio} is calculated and used in the spline to retrieve $\mathcal{L}_{Q_{\text{ratio}}}$.

The second input to the likelihood, called $\Delta_{\text{layer}, \text{first}}$, is the difference in the start-



(a) $\Delta_{\text{layer}, \text{last}}$ distribution (area normalised). The blue and red distributions refers to matching pairs which were matched to the same true particle and different true particles respectively. While both distributions peak in the same place, the peak for the correct matches distribution is much sharper, indicating a good level of separation between the correctly and incorrectly matches events.

(b) $\Delta_{\text{layer}, \text{last}}$ probability density distribution. The peak of the distribution aligns well with the correct matches peak in Fig. 5.14a which aides in maximising the number of correct matches made.

Figure 5.14: $\Delta_{\text{layer}, \text{last}}$ and its probability density distribution in the barrel ECal for the two track case. NEUT-based simulation of ND280 beam events were used to produce the distributions.

ing layer of each 2D track which forms the matching candidate pair, where starting layer refers to the layer closest to the ND280 Tracker. For 2D tracks which should be matched together, $\Delta_{\text{layer}, \text{first}}$ should be 1. The separation ability of this variable for the two track, barrel is shown in Fig. 5.13a. The discrete probability density function was created using equation 5.9. It was not necessary to interpolate using splines as $\Delta_{\text{layer}, \text{first}}$ is itself discrete. The probability density function for $\Delta_{\text{layer}, \text{first}}$ is shown in Fig. 5.13b for the two track, barrel case. For each matching candidate pair, the value of $\Delta_{\text{layer}, \text{first}}$ is calculated and the corresponding $\mathcal{L}_{\Delta_{\text{layer}, \text{first}}}$ is retrieved from the probability density function.

The third and final input to the likelihood, called $\Delta_{\text{layer}, \text{last}}$, is the difference in the ending layer of each 2D track which forms the matching candidate pair, where the ending layer refers to the layer furthest from the ND280 Tracker. Functionally, how this function is used is essentially identical to $\Delta_{\text{layer}, \text{last}}$ so it will not be described in detail. The separation ability of this variable and its corresponding probability density function for the two track, barrel case are shown in Fig. 5.14a and Fig. 5.14b

respectively.

The matching likelihood, \mathcal{L} , for a matching candidate pair is then

$$\mathcal{L} = \mathcal{L}_{Q_{\text{ratio}}} \times \mathcal{L}_{\Delta_{\text{layer, first}}} \times \mathcal{L}_{\Delta_{\text{layer, last}}}. \quad (5.10)$$

As described above, \mathcal{L} is calculated for every matching candidate pair and the pair which maximise \mathcal{L} is selected as a match and removed from the pool. The process is then repeated until no more matches can be made.

3D tracks have now been formed, but the associated directions and positions of those tracks still need to be calculated. The track fitting process for the newly formed 3D tracks is very similar to that described in section 4.4.6. The tracks are briefly separated into their constituent 2D views and a charge-weighted average position of each layer is calculated using the track's constituent hits. Then, the hits in the opposing view are used to estimate the third coordinate of a given layer using a least-squares fit. After all of the coordinates have been estimated, a full 3D least-squares fit of the positions in each layer is performed to estimate the 3D track's direction and position in that ECal layer.

5.2.6 Track splitting

During development of the reconstruction, it became clear that a certain topology had been overlooked. An example of this is shown in Fig. 5.15 which shows a MC neutrino interaction in the ECal with 3 charged final states. As can be seen from Fig. 5.15, the muon (solid green line) is emitted back-to-back with one of the final-state protons (solid blue line) in this ECal view. Because the Hough transform implementation only concerns itself with straight lines, the result is that such back-to-back trajectories are typically reconstructed as a single track. However, there are two views available for every ECal module and the back-to-back emittance typically only appears in the ECal view which is perpendicular to the beam direction (the XY view). So, the correctly reconstructed tracks in the other ECal view can be used to split the tracks in the problematic view.

Consider a neutrino interaction with two charged final-states which has been reconstructed as one track in one view (called the merged view) but reconstructed as two

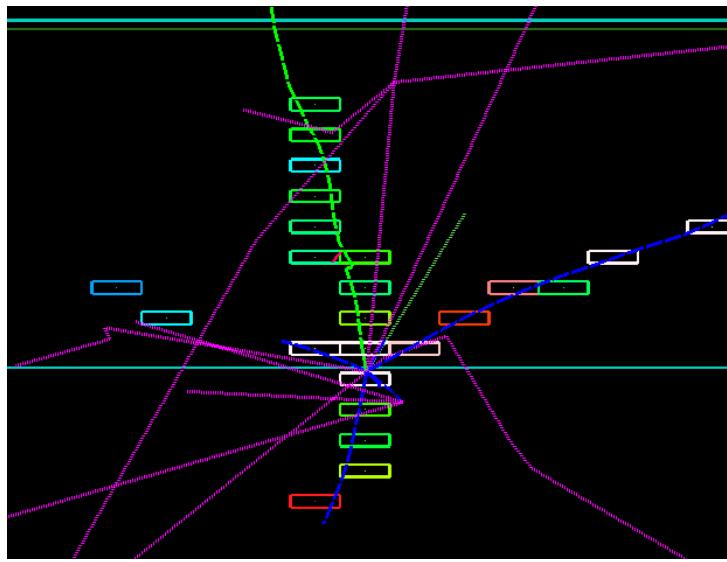


Figure 5.15: Event display (XY view) of a problematic (for the reconstruction) neutrino interaction in the ECal. The solid green track is the muon, the solid blue tracks are protons and the pink tracks are neutrons. The neutrino is the short dashed green line. The proton travelling vertically downwards and the muon are created almost back to back leaving a line of hit scintillator bars. The 2D hough transform would register this as a single straight line.

tracks in the other view (called the other view). It should be expected that the single track in the merged view is not a particularly good match for either of the tracks in the other view. However, if the two tracks in the other view were temporarily merged together and this new track was compared to the single track in the merged view, one should expect this match to return a much higher value of \mathcal{L} . This feature can be used to identify a potentially merged track in a given view.

This motivated an extension to the 3D matching aspect of the reconstruction. As a reminder, the 3D matching algorithm makes every pairwise comparison of tracks from each view to find the pair which maximise \mathcal{L} . This matching routine was modified to also include temporary mergings of every pairwise combination of tracks in a single view. These temporary mergings can then be compared to every single track in the other view. As an example, consider a situation where 3 tracks (labelled A , B and C) have been reconstructed in one view and two tracks (labelled Y and Z) have been reconstructed in the other view. With the old method, the following comparisons

would be made:

$$\begin{aligned}
 A &\longleftrightarrow Y \\
 A &\longleftrightarrow Z \\
 B &\longleftrightarrow Y \\
 B &\longleftrightarrow Z \\
 C &\longleftrightarrow Y \\
 C &\longleftrightarrow Z
 \end{aligned} \tag{5.11}$$

With the new method, which includes temporarily merged pairs of tracks in a given view, the following comparisons would be made:

$$\begin{aligned}
 A &\longleftrightarrow Y \\
 A &\longleftrightarrow Z \\
 B &\longleftrightarrow Y \\
 B &\longleftrightarrow Z \\
 C &\longleftrightarrow Y \\
 C &\longleftrightarrow Z \\
 A + B &\longleftrightarrow Y \\
 A + B &\longleftrightarrow Z \\
 A + C &\longleftrightarrow Y \\
 A + C &\longleftrightarrow Z \\
 B + C &\longleftrightarrow Y \\
 B + C &\longleftrightarrow Z \\
 A &\longleftrightarrow Y + Z \\
 B &\longleftrightarrow Y + Z \\
 C &\longleftrightarrow Y + Z
 \end{aligned} \tag{5.12}$$

In terms of the matching comparisons, the temporarily merged comparisons are treated on the same footing as the single track comparisons; \mathcal{L} is calculated for each comparison and the one which maximises \mathcal{L} is selected. However, the post-matching treatment of the best match depends on whether a temporary merge is involved. If the match which maximises \mathcal{L} involves two single tracks then the treatment is as before; they are removed from the pool and the process is repeated. If a temporarily

merged pair are involved, the 2D crossing location of the temporarily merged tracks is calculated and is then used to split the single track in the other view. The original track is removed from the pool and replaced by the two single tracks formed from the split. The matching process is then begun again and repeated until all matches have been made.

5.2.7 Track pairwise crossing reconstruction

The final step of the reconstruction is to estimate where each of the 3D track's paths cross. As each track is reconstructed as a straight line in 3D, the final step is fairly simple. Using the track direction and position information calculated at the end of section 5.2.5, the position at closest approach for every pairwise combination of 3D tracks is calculated analytically. The distance of closest approach is also calculated. Six hits, the closest three from each 3D track, are then associated to the pairwise crossing.

5.3 Output of the reconstruction

The reconstruction is run over every 3D cluster found in the ECal. By applying the steps outlined above, for each 3D cluster the following output is given:

- A set of 3D tracks
- The pairwise crossings of all 3D tracks found in the cluster

Note that no vertex formation beyond the pairwise crossings is calculated at this stage, nor is any analysis of the 3D tracks performed. While this may seem like an oversight of the reconstruction, this approach was decided as no assumptions are made about what the tracks/crossings represent at this point, making the output more generic. Any analysis which wants to make use of the reconstruction is given enough information to apply more targeted reconstruction downstream.

5.4 Validation of the reconstruction

Because of the large scope of the reconstruction and the limited time available for the presented analysis, the validation of the reconstruction was done in parallel to the rest

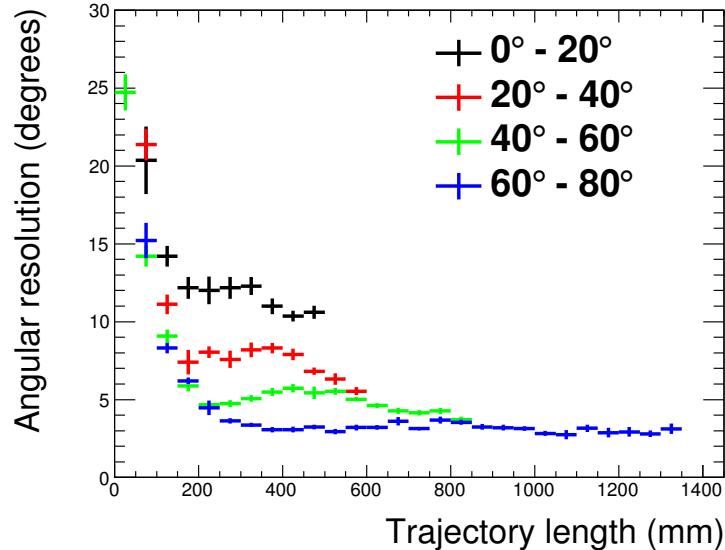


Figure 5.16: The angular resolution as a function of trajectory length in the DS ECal when applying the enhanced reconstruction to muon MC. The colour coding refers to the true entry angle range of the muons. The figure was generated using single muon particle gun fired into the front face of the DS ECal with a controlled range of entry angles. The angular resolution distributions were created by L. Pickering.

of the analysis and is still an ongoing effort. The validation that has been done can be split into two areas: validation of the performance of the algorithms purely using MC and comparisons of MC to data using control samples.

5.4.1 Validation of algorithm performance

The first performance validation investigated the angular resolution using the enhanced reconstruction (by L. Pickering [80]). This study calculated the angular resolution for MC muons fired into the side-left ECal for a range of entry angles. For each MC event, the cosine of the angular separation between the true particle angle and the reconstructed angle was calculated, $\cos \theta^{\text{Sep}}$. The values of $\cos \theta^{\text{Sep}}$ were then binned in a distribution. An outward scan from the peak of the distribution was then performed to find where the height decreased to 68% of the peak. The value of θ^{Sep} at this point was taken as the angular resolution. These results are shown in Fig. 5.16. Generally speaking, the found angular resolutions are very good. For long trajectories (200 mm), the angular resolution is within 15°. It is only for short tracks (40 mm) that the angular resolution becomes large.

	1 track likelihood	2 track likelihood	3 track likelihood
1 st track matched	96%	92%	80%
2 nd track matched		89%	79%
3 rd track matched			70%

Table 5.1: The percentage number of correct matches in the 3D matching separated by which track matching likelihood was used. NEUT-based Monte Carlo simulation of the T2K beam was used to calculate the matching percentages.

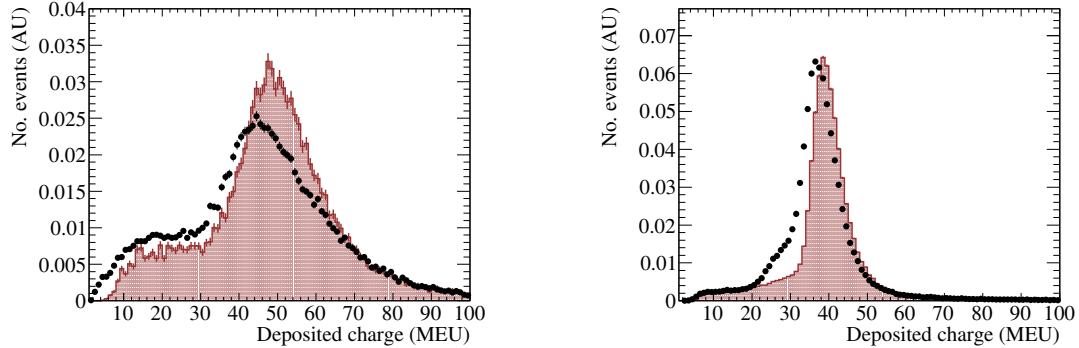
The second performance validation studied the success of the 3D matching. To do this, a sample of beam MC events in the ECals were processed through the enhanced reconstruction. After completion, the 2D components of the 3D tracks were analysed and matched to the true particles which created them. A match was counted as a success if both 2D components were matched to the same true particle. These results are shown in table 5.1, separated out by which track matching likelihood was used. For the 1 and 2 track likelihood cases, the matching performs very well. It is only in cases where the 3 track likelihood is used that the matching start to suffer slightly. However, all of the correct matching rates are well above 50% which suggests that the 3D matching is performing adequately.

5.4.2 Data-motivated validation of the algorithms

As described above, the validation is still an ongoing effort and, so far, all of the data-motivated validation has been done by L. Pickering. At time of writing, validation events measured in ND280 can be split into two groups:

1. **FGD collinear cosmics (named `fgdcol`):** A sample of clean FGD-triggered cosmic events which are scaled to have an equal distribution of angles.
2. **Through-going Tracker muons (named `bskmu`):** A sample of clean beam-triggered events which solely contain a muon depositing energy in TPC2, TPC3 and the DS ECal.

The `fgdcol` sample will naturally contain a good sample of barrel ECal events whilst having only a limited number of DS ECal events whereas the reverse is true for the `bskmu` sample. So, it follows that the `fgdcol` and `bskmu` samples should be used to



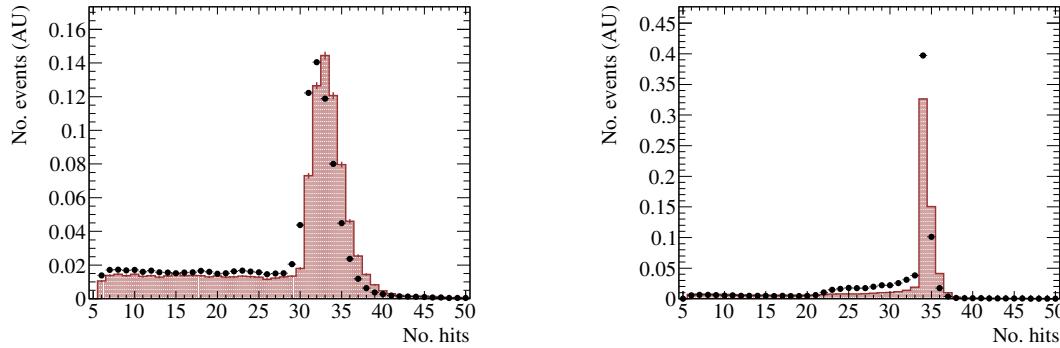
- (a) Barrel ECal events in the fgdcol sample. The data peak is offset from the Monte Carlo peak, indicating an unmodelled dead channel issue.
- (b) DS ECal events in the bskmu sample. The data peak is offset from the Monte Carlo peak, indicating an unmodelled dead channel issue. There is also a significant data excess before the peak of the distributions.

Figure 5.17: The summed charge contained on all tracks reconstructed in each ECal cluster. The red histograms and black points are Monte Carlo and data respectively.

separately assess the reconstruction in the barrel ECal and DS ECal respectively.

Fig. 5.17 shows the summed charge contained in all tracks reconstructed in each ECal cluster. To clarify what this means, if a reconstructed ECal cluster contained three reconstructed tracks where tracks 1, 2 and 3 contain 20 MEU, 50 MEU and 30 MEU of charge respectively, that event would be registered as having 100 MEU of charge in the relevant distribution in Fig. 5.17. Both the barrel ECal and DS ECal samples show a large discrepancy around the charge peak. This offset can be explained as a hit-level charge discrepancy between data and Monte Carlo and is not caused by the reconstruction. This discrepancy should, however, be considered during analysis of the systematic uncertainties. Unfortunately, the DS ECal events in the bskmu sample (Fig. 5.17b) shows an extra data excess in the 20 MEU to 30 MEU region which is just before the charge peak. This data excess can not be explained by a relative hit inefficiency.

Fig. 5.18 shows the summed number of hits contained in all tracks reconstructed in each ECal cluster. Generally speaking, the distributions in Fig. 5.18 share similar features to those in Fig. 5.17. Importantly, an excess of data events appears just before the peak in Fig. 5.18. The fact that this excess also appears in Fig. 5.18 strongly suggests that the issue is caused by the number of hits associated to the reconstructed tracks,

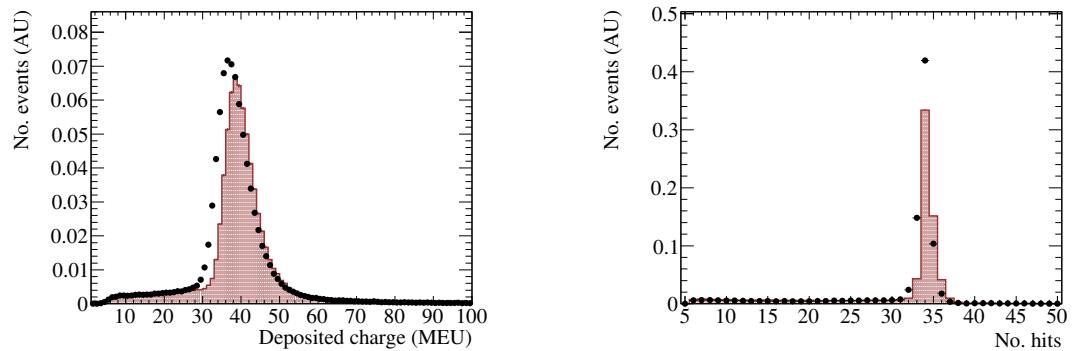


- (a) Barrel ECal events in the fgdcol sample. The only significant discrepancy is the offset of the data peak relative to the Monte Carlo peak, suggesting an unmodelled dead channel issue.
- (b) DS ECal events in the bskmu sample. There is a significant data excess before the peak in the distributions.

Figure 5.18: The summed number of hits contained on all tracks reconstructed in each ECal cluster. The red histograms and black points are Monte Carlo and data respectively.

rather than the deposited charge. Further investigation of this discrepancy revealed that a mismodelled hit inefficiency, most likely due to dead DS ECal channels, was the issue. One of the 2D track quality checks (see section 5.2.4) requires that a 2D track candidate can not skip a layer in a given view which is problematic when considering dead channels.

To test this hypothesis, this requirement was relaxed to allow a 2D track candidate to skip a single ECal layer and the Monte Carlo bskmu sample was re-processed. The re-processed DS ECal events are shown in Fig. 5.19. As can clearly be seen, the data excess before the peak in the distributions has now gone. Unfortunately, large scale processing and analysis of beam Monte Carlo events had already started by the time this issue was discovered and so it was infeasible to include this bug fix in the analysis. So, it was decided that this issue would be treated as an additional systematic uncertainty. The collected data events were processed with this fix in place.



(a) Summed charge contained on all tracks in each reconstructed ECal cluster. (b) Summed number of hits contained on all tracks in each reconstructed ECal cluster.

Figure 5.19: The number of DS ECal events in the bskmu sample after relaxing the 2D track quality check. The red histograms and black points are the Monte Carlo and data respectively. The track quality fix has corrected the data excesses in both distributions, leaving only the offset between the data and Monte Carlo peaks.

Chapter 6

Magnetic field simulation in ND280

ND280 is housed in the former UA1 magnet which provides a 0.18 T magnetic field through the Tracker. The purpose of this magnetic field is to aid particle identification and momentum measurements in ND280's TPCs. This magnetic field is accurately modelled in nd280mc by a constant 0.18 T magnetic field in the Tracker. The ND280 analyses, which are inputs to the oscillation analyses, search for a TPC track matched to an FGD track so, in a first iteration of these analyses, the magnetic field model is sufficient for its purpose.

However, a significant part of the UA1 magnet is the iron based, magnetic flux return yoke which helps to tightly contain the magnetic field outside of the Tracker region. As a result, there is a significant magnetic field contained within the magnetic flux return yoke during operation. This magnetic field in the flux return yoke is not modelled in nd280mc. For the first iterations of the ND280 oscillation input analyses, this approximation was valid. However, as the ND280 analyses become more mature, and non-Tracker based analyses are started, this approximation is no longer sufficient. Hence the magnetic field model in the simulation needs revising.

The Monte Carlo event rates seen in the ECals are directly affected by mismodelling of the magnetic field in the yokes which can have very serious consequences for any ECal-based analysis. The following chapter presents the first investigation of the magnetic field's effect in this region and an improvement to the magnetic field model. This model improvement has now been incorporated into the official ND280 MC simulations which is used by all physics analyses at ND280.

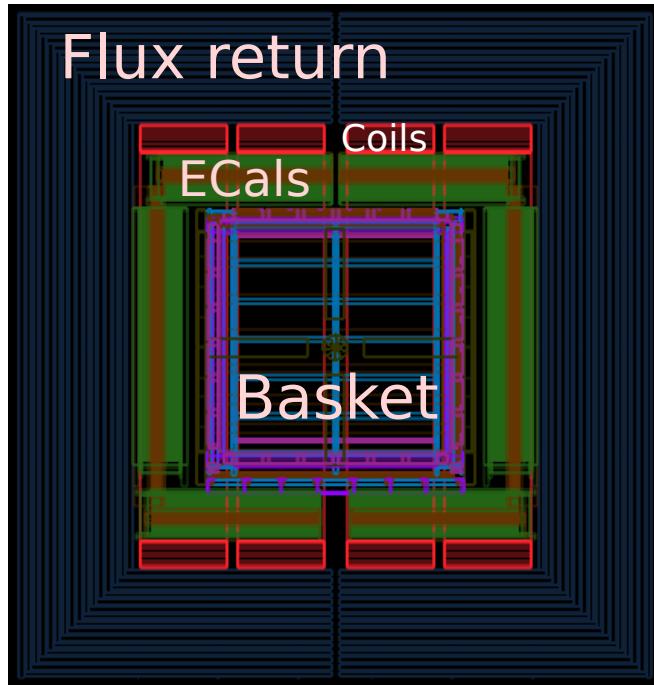


Figure 6.1: Graphical display of the X-Y ND280 cross-section with the flux return, magnetic field coils, ECals and the Basket (Tracker) labelled.

6.1 Magnetic field model in the ND280 flux return

A simple model for the magnetic field in the flux return can be found by making two assumptions. Firstly, the flux return yoke consists mostly of iron which has a much higher magnetic permeability than the air surrounding it. Therefore, it can be assumed that any magnetic flux passing through the ND280 Tracker is solely transported back around by the return yoke (no flux passes through the atmosphere in the pit). It follows that

$$\phi_b = \phi_r, \quad (6.1)$$

where ϕ_b is the magnetic flux passing through the Tracker and ϕ_r is the magnetic flux passing through the return yoke. The second assumption regards the shape of the ND280 Tracker and flux return yoke. The X-Y cross-section of ND280, as shown in Fig. 6.1, can be modelled as two distinct parts: the flux return yoke and everything contained within. It is clear from Fig. 6.1 that both areas are rectangular so the system can be modelled as a smaller rectangle (the Tracker) contained within a larger rectangle (the return yoke). Using this assumption and Eq. 6.1, the magnetic

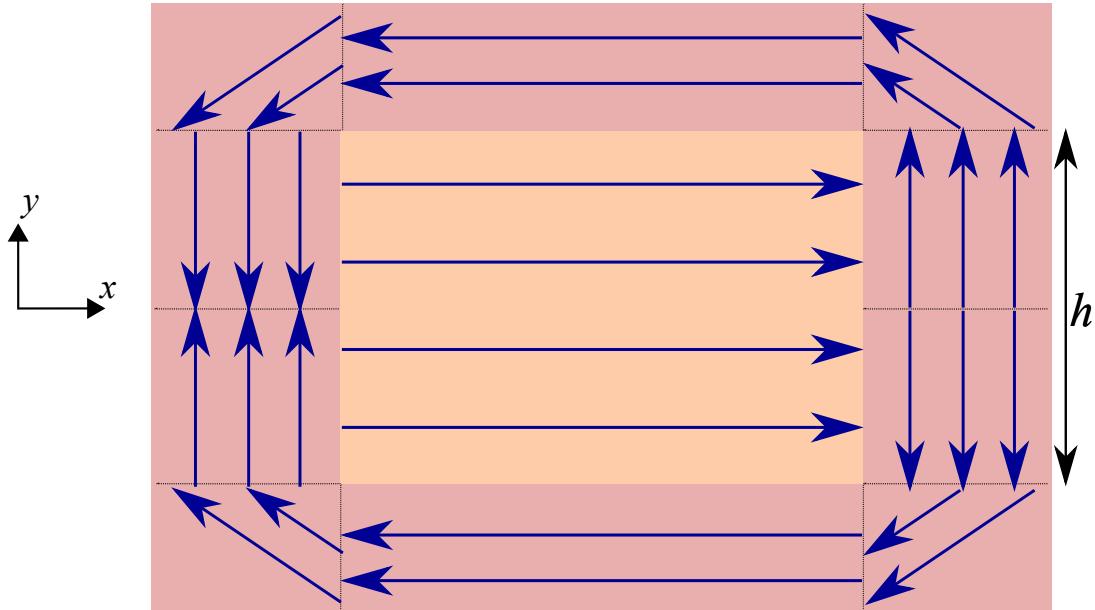


Figure 6.2: Simple model of the magnetic field (blue lines) in the Tracker (orange) and flux return (red) regions. The flux return yoke has been separated into ten sections: four vertical B field sections, four corner B field sections and two horizontal B field sections. The height of the Tracker region, h , is also shown.

field strength passing through the return yoke, B_r , is

$$B_r = \frac{B_b A_b}{A_r - A_b}, \quad (6.2)$$

where B_b is the strength of the magnetic field passing through the Tracker, A_b is the cross-sectional area of the Tracker region and A_r is the cross-sectional area of the flux return yoke if the flux return yoke was not hollow.

Eq. 6.2 is sufficient to model the strength of the magnetic field in the flux return yoke, yet Eq. 6.2 conveys no information about the direction of the magnetic field. However, the direction can be estimated by separating the flux return into ten sections as shown in Fig. 6.2. The sections are: four vertical sections where the magnetic field enters and exits the flux return region, two horizontal sections where the magnetic field in the flux return is anti-parallel to the magnetic field in the Tracker and four corner sections where the magnetic field transitions between the horizontal and vertical sections. The coordinate system used is the same as the one shown on the left hand side of Fig. 6.2 which defines the x -axis as being parallel to the magnetic field in the Tracker region and the y -axis as the vertical. By invoking the usual vector notation

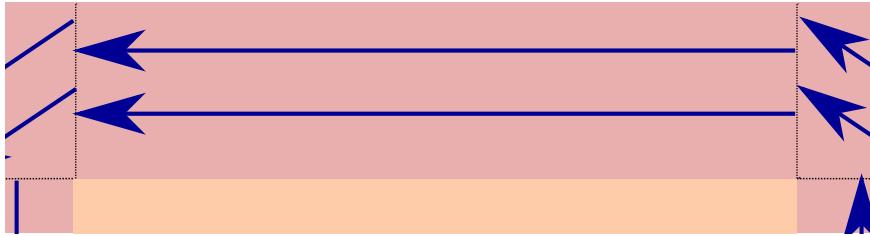


Figure 6.3: A horizontal section of the flux return with its corresponding magnetic field lines (blue). The corner sections, vertical sections and the Tracker are also visible. This figure is a zoomed in cut-out of Fig. 6.2.

\hat{i} and \hat{j} (where \hat{i} and \hat{j} are parallel to the positive x -axis and y -axis respectively), the magnetic field in the Tracker is defined as

$$\vec{B}_b = B_b \hat{i}. \quad (6.3)$$

The first magnetic field section in the flux return to be defined is the B field in the horizontal sections, an example of which is shown in Fig. 6.3. The magnetic field here should be anti-parallel to the magnetic field in the Tracker, regardless of which horizontal section is being considered. Using equation 6.3, it follows that the magnetic field in the horizontal section of the flux return is

$$\vec{B}_r^H = -B_r \hat{i}. \quad (6.4)$$

The next magnetic field section in the flux return to define is the B field in the vertical section, an example of which is shown in Fig. 6.4. The field model for the vertical section should take into account that the strength of the magnetic field is not constant, but smoothly change in strength from the point that it enters vertical section until it leaves it. The direction of the field depends on which vertical section is under consideration. For the section where the magnetic field is exiting the Tracker and moving upwards, as shown in Fig. 6.4, the relevant vector is

$$\vec{B}_r^V = \frac{y}{h/2} B_r \hat{j}, \quad (6.5)$$

where h is the height of the ND280 Tracker and y is the height at which the magnetic field is being evaluated in the ND280 coordinate system.

Finally, In the corner sections of the flux return, the magnetic field is defined to be straight, constant in strength and at 45° to the horizontal and vertical sections, an

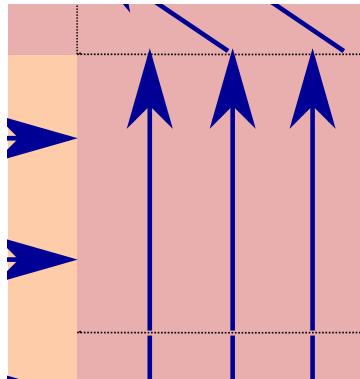


Figure 6.4: A vertical section of the flux return with its corresponding magnetic field lines (blue). In this vertical section, the magnetic field has exited the Tracker and is travelling vertically upwards through the flux return. A corner section, another vertical section and the Tracker are also visible. This figure is a zoomed in cut-out of Fig. 6.2.

example of which is shown in Fig. 6.5. As with the vertical sections, the magnetic field direction is dependent on which corner section is being considered. For the corner section where the entering magnetic field has left the vertical section in which the field was travelling vertically upwards, away from the Tracker (the same corner section as shown in Fig. 6.5), the magnetic field vector is

$$\vec{B}_r^C = \frac{B_r}{\sqrt{2}}(\hat{j} - \hat{i}). \quad (6.6)$$

The magnetic field in the other corner sections are then 45° rotations of Eq. 6.6.

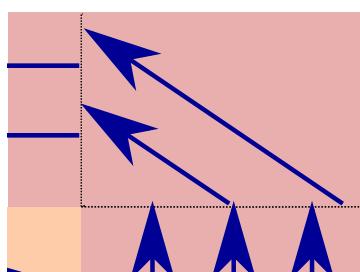


Figure 6.5: A corner section of the flux return with its corresponding magnetic field lines (blue). In this corner section, the magnetic field has entered from a vertical section in which it was travelling vertically upwards, away from the Tracker. A vertical section, a horizontal section and the Tracker are also visible. This figure is a zoomed in cut-out of Fig. 6.2.

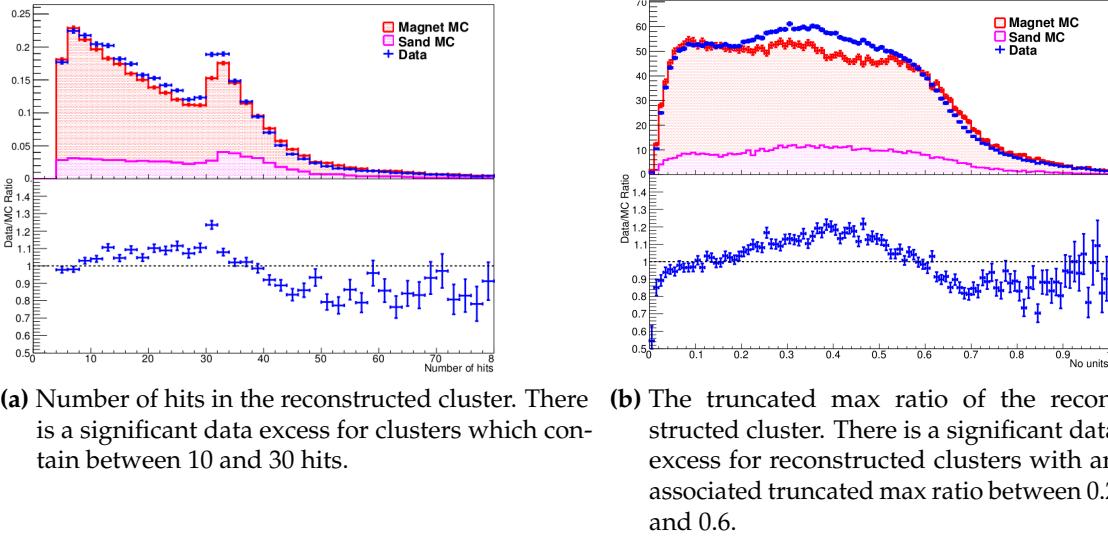
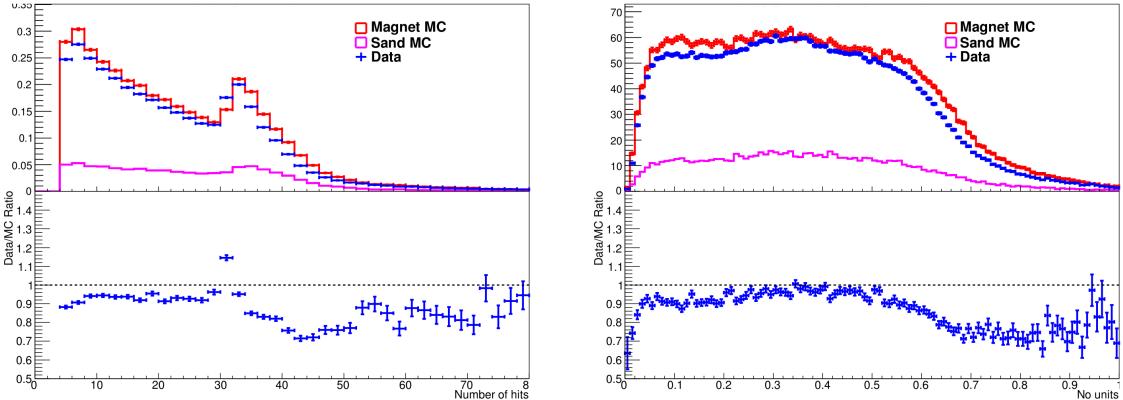


Figure 6.6: Comparisons of data and Monte Carlo in the bottom-left barrel ECal for previous software productions. The red and pink histograms are Monte Carlo simulation of T2K beam neutrinos incident on ND280 and the surrounding pit respectively. The blue data points are collected data from run 3C. Both of the plots are POT normalized. The data excesses shown are especially significant because they occur in the highest population regions.

6.2 Effect of magnetic field on the ECal

It was found in previous iterations of ND280 analyses that there were significant discrepancies between the Monte Carlo and the collected data in the ECal. The discrepancies found had an ECal module dependence where the bottom modules were affected most. The problem appeared in the low-level distributions, such as how many scintillator hits were assigned to each reconstructed object. An example of this is shown in Fig. 6.6a which shows a comparison of Monte Carlo with run 3C data for the number of constituent hits in each reconstructed object measured in the bottom-left barrel ECal. For the number of hits region with high population (between roughly 10 and 30) there is a clear excess of collected data events. As the high level variables, such as the ones used for track-shower discrimination, are based on these low-level quantities, the discrepancy propagated through causing significant disagreement at all levels. An example of this effect is shown in Fig. 6.6b which shows a data and Monte Carlo comparison, again for the bottom-left barrel ECal, for the Truncated Max Ratio (TMR) of the reconstructed objects. The TMR, which calculates the ratio of the lowest total charge found in an ECal layer to the high total charge in an ECal



(a) Number of hits in the reconstructed cluster. The data excess seen in Fig. 6.6a has become a Monte Carlo excess (approximately 10%). A similar Monte Carlo excess is seen in other ND280 subdetectors.

(b) The truncated max ratio of the reconstructed cluster. The data excess seen in Fig. 6.6b has become a Monte Carlo excess which is much smaller in magnitude.

Figure 6.7: Comparisons of data and Monte Carlo in the bottom-left barrel ECal with the magnetic field model described in section 6.1 implemented. The red and pink histograms are Monte Carlo simulation of T2K beam neutrinos incident on ND280 and the surrounding pit respectively. The blue data points are collected data from run 3C. Both of the plots are POT normalized. Comparing the shown distributions with those in Fig. 6.6, the effect of the simple magnetic field model is clearly visible.

layer after removing the top and bottom 20% of hits, is an input into the track-shower discriminator described in section 4.4.8. The track-shower discrimination is a key feature of the ECal reconstruction which is used by several ND280 analyses so such a big discrepancy is a serious problem.

As ND280 is off-axis relative to the T2K beam, there is an increase in the neutrino flux in the flux return yoke and surrounding pit regions below the bottom ECals. This means that any charged final states in this region would generally have to propagate through the flux return yoke. If a magnetic field is present in this region, as there would be during data taking, the trajectory of such particles would be bent upwards towards the ND280 Tracker region, causing an increase in the event rate in the bottom ECals. If the magnetic field in this region is missing, the above statement does not hold and a relative deficit of events would be seen which describes the situation seen in Fig. 6.6. To test this hypothesis, the simple magnetic field model in the flux return yoke described in section 6.1 was implemented in nd280mc and a batch of beam and sand Monte Carlo was produced to test its effect. To get an idea of the magnetic field's effect on the rates measured by the ECals, the same variables as shown in Fig. 6.6

are shown in Fig. 6.7, but with the magnetic field activated. The difference is very clear; Fig. 6.7a shows that the excess in the 10 to 30 hits region is now gone. It is also clear that the effect of the magnetic field has propagated through to the high level discriminators, as shown in Fig. 6.7b.

Despite this study only briefly investigating the presence of a magnetic field in the UA1 flux return yoke, the improvement provided is undeniable. It was decided that the model would be a permanent feature of the ND280 simulation and is now used in all software productions including the inputs to this analysis.

Chapter 7

Selection of neutrino interactions in the ECal

This analysis presents a measurement of the CC inclusive interaction cross-section of ν_μ with lead nuclei using the ND280 Tracker ECals. To make such a measurement, a sample of neutrino interaction vertices within the ECal must be found. The selection of events is based on the enhanced reconstruction outlined in chapter 5, which was specifically designed to be sensitive to track multiplicity. As a result of this method, vertices in the ECal are naturally separated into topologies defined by the number of reconstructed tracks. Any selection development should take advantage of this situation and tailor cuts to be specific to each topology, which should result in a higher overall sample purity. The 3D track matching aspect of the reconstruction was only tuned to handle up to three tracks simultaneously. However, this analysis aims to measure a CC-inclusive cross-section so it should not be biased against any neutrino energy range. There is a deep connection between the number of reconstructed tracks and the energy of the neutrino that created them. The number of reconstructed ECal tracks correlates with the number of final state particles involved with the neutrino interaction as shown in Fig. 7.1 and the number of final state particles correlates with the energy of the interacting neutrinos. Therefore it is important to not reject events based solely on the number of reconstructed tracks. Bearing this information in mind, the selection should separate out the events into the following topologies:

- 1 prong topology
- 2 prong topology
- 3 prong topology

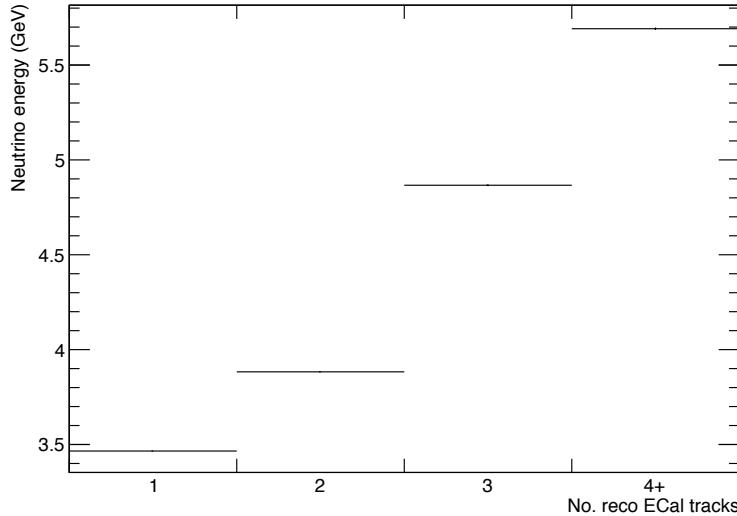


Figure 7.1: The average energy of neutrino interactions seen in the ECal against the number of reconstructed tracks in the ECal that the neutrino final states created. The uncertainty on the neutrino energy is calculated as the standard error on mean of the neutrino energy. There is a clear trend between the two variables.

- 4+ prong topology

The definition of a prong is a reconstructed track associated with a reconstructed vertex. It should also be noted that there exists a 0 prong topology in which a neutrino interacts but does not have sufficient energy to produce any recognisable signal in the ECal. This topology is automatically rejected by the reconstruction.

However, as described in section 5.3, the output of the enhanced reconstruction has been kept generic and is not specifically tailored to this task. The reconstruction outputs a set of clusters which contain a set of 3D tracks and every pairwise crossing that said tracks make. While it is true that in some situations the reconstruction will accurately represent a vertex “out of the box”—e.g., when only two tracks are reconstructed in the cluster—there will be many situations where extra reconstruction steps are needed before any further selection can take place. Hence the structure of this chapter is as follows: the definition of signal is described first along with the sample used to develop the selection, then a description of the main sources of background followed by a discussion of the vertex reconstruction. After the final reconstruction steps have been discussed, a full discussion of the neutrino selection follows.

7.1 Signal definition

The measurement that is the intended outcome of this analysis is the CC inclusive interaction cross-section of ν_μ with lead nuclei using the ND280 Tracker ECals. The active volume of the ECal consists of layers of plastic scintillator and lead absorbers. Neutrino interactions that occur in the lead absorbers are indistinguishable from those that occur in the plastic scintillator. Because of this fact, the selection does not attempt to separate the two cases out. So, the strict signal definition used by this selection is as follows:

- Interacting neutrino is flavour ν_μ
- Interaction type is charged current
- Interaction occurs within the active volumes of either the barrel ECal or DS ECal

The active volume is defined by the ROOT geometry used in the simulation. Specifically, it is a rectangular box which encompasses all of the lead and scintillator layers for each ECal module. Any target element contained within this volume is a signal target.

7.2 Monte Carlo sample

NEUT was used to generate a neutrino beam sample of Monte Carlo events which corresponds to 3.949×10^{20} POT. All of this sample was generated using a simulated beam power of 178 kW which is the average beam power used for Run III. At this intensity, there are an expected 9.46 neutrino interactions per eight bunch spill in ND280. So, there will be approximately one interaction per bunch across the entire ND280, meaning the chance of pileup is small.

The sample described above only simulated neutrino interactions within ND280 (including the outermost magnet). However, we expect many interactions in the surrounding pit (referred to as “sand interactions” from now on) to have final states which enter ND280. So, NEUT was used to generate an additional sample of sand interactions (hereafter referred to as sand Monte Carlo) which corresponds to 3.708×10^{20} POT. There is roughly a 6% difference in the POT of the beam and sand Monte Carlo. As the sand Monte Carlo has a sufficiently high level of statistics, scaling the smaller sample

ECal module	No. signal events
Bottom-right	223042
Side-right	231955
Top-right	135476
Bottom-left	262500
Side-left	293065
Top-left	140539
Downstream	134100
Total	1420677

Table 7.1: The number of signal events in the NEUT-based beam Monte Carlo sample separated by ECal module.

to match the larger is acceptable.

The number of signal interactions from the MC sample is shown in table 7.1, separated into the host ECal module they occurred in. The number of signal interactions is also shown as a function of neutrino energy for the bottom-left, top-right and DS ECals in Fig. 7.2. Clearly, there are an extremely high number of signal events in the sample, which means that the statistical error in this analysis should be small. As an example, consider a selection with a 1% signal efficiency in the DS ECal. Assuming 100% selected sample purity, the number of selected events would be 1341 which corresponds to a statistical error of 2.73%. Even in this extremely unrealistic situation, the statistical error is small.

7.3 Background particles

For any selection process to be successful, it must be able to identify the signal particles of interest while simultaneously rejecting background particles. To make this possible, it is necessary to have a solid understanding of the background particle flavour and provenance which is based solely on simulation before application of reconstruction. In the ECal, the backgrounds can be separated into two topologies. The first is the intrinsic background which are based on neutrino interactions in the ECal which are not defined as signal e.g. ν_e interactions. As the reconstruction stands, this topology is

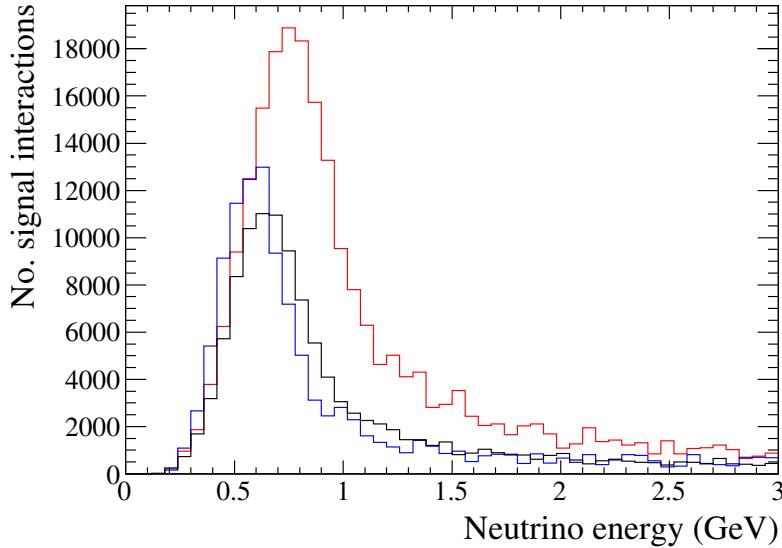


Figure 7.2: The number of signal events in the NEUT-based beam Monte Carlo sample as a function of neutrino energy in the bottom-left (red), top-right (blue) and DS (black) ECals. The bottom-left ECal sees the most signal events and at a higher neutrino energy which is caused by the off-axis effect.

difficult to reject, largely because the reconstruction has not progressed to the point where particle identification can be reliably used to tag the final state particles. As will be shown, this background topology will be reduced at a similar rate to the signal-loss during selection development. The second background topology concerns particles which are produced in another area of the geometry, but enter the ECal volume. This background topology will be referred to as the ECal Out Of Fiducial Volume (OOFV) background from now on. It is especially important to understand the ECal OOFV background as it dominates the reconstructed events seen in the ECal. To tailor a selection to reject such a topology, it is necessary to understand its characteristics. Fig. 7.3 shows the number of simulated ECal OOFV background particles, separated into particle species as a function of their momentum at point of ECal entry. At all points, muons are the dominant background. However, the magnitude of this dominance is in no way uniform. For example, in the lowest momentum regime, photons are the second largest dominant contributor, while at the momentum peak charged pions are the second largest contributor. It should be expected that the muon background and photon background will create reconstructed clusters with a very different shape and charge deposition. This promotes the idea of developing separate cuts which are tailored towards each particles species in the ECal OOFV background.

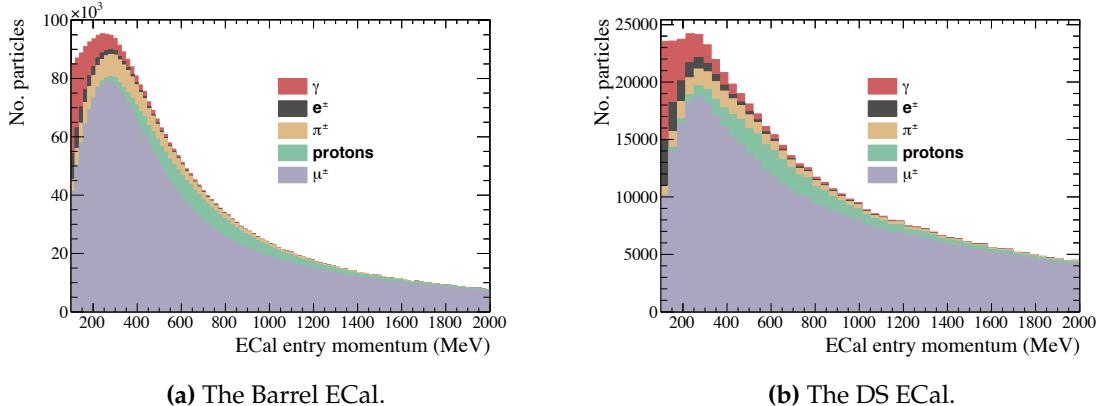


Figure 7.3: The number of background particles, separated out into particles species, as a function of their momentum at point of entry into the ECal.

It is also important to consider where the background particles originate. The barrel and DS ECals occupy very different areas of the geometry and are not exposed to the same level of neutrino flux. Fig. 7.4 and Fig. 7.5 show the ECal entrance momentum of the simulated entering background Vs. the point of creation of the OOFV background for the barrel ECal and DS ECal respectively. It is clear from the plots that the creation point of the entering background is very different for the barrel and DS ECal. For particle species in the barrel ECal, the dominant creation region is the magnet and surrounding pit. This should be expected as all of the barrel ECal is surrounded by the magnetic flux return yoke. However, in the DS ECal case, there is significantly more variation in which ND280 regions contribute to the DS ECal OOFV background. Notably, there is a significant contribution from the downstream region of the tracker (FGD 2 and TPC 3). Such differences promote the idea of separating the barrel ECal and DS ECal in the selection process.

7.4 Vertex reconstruction

The following section outlines the algorithm developed to reconstruct vertices in the ECal using the enhanced reconstruction described in chapter 5. The core of the reconstruction is a nearest neighbour algorithm which clusters together tracks based on their pairwise crossing locations. Because the output of the enhanced reconstruction was kept generic, it is necessary to first reject poorly reconstructed tracks and the

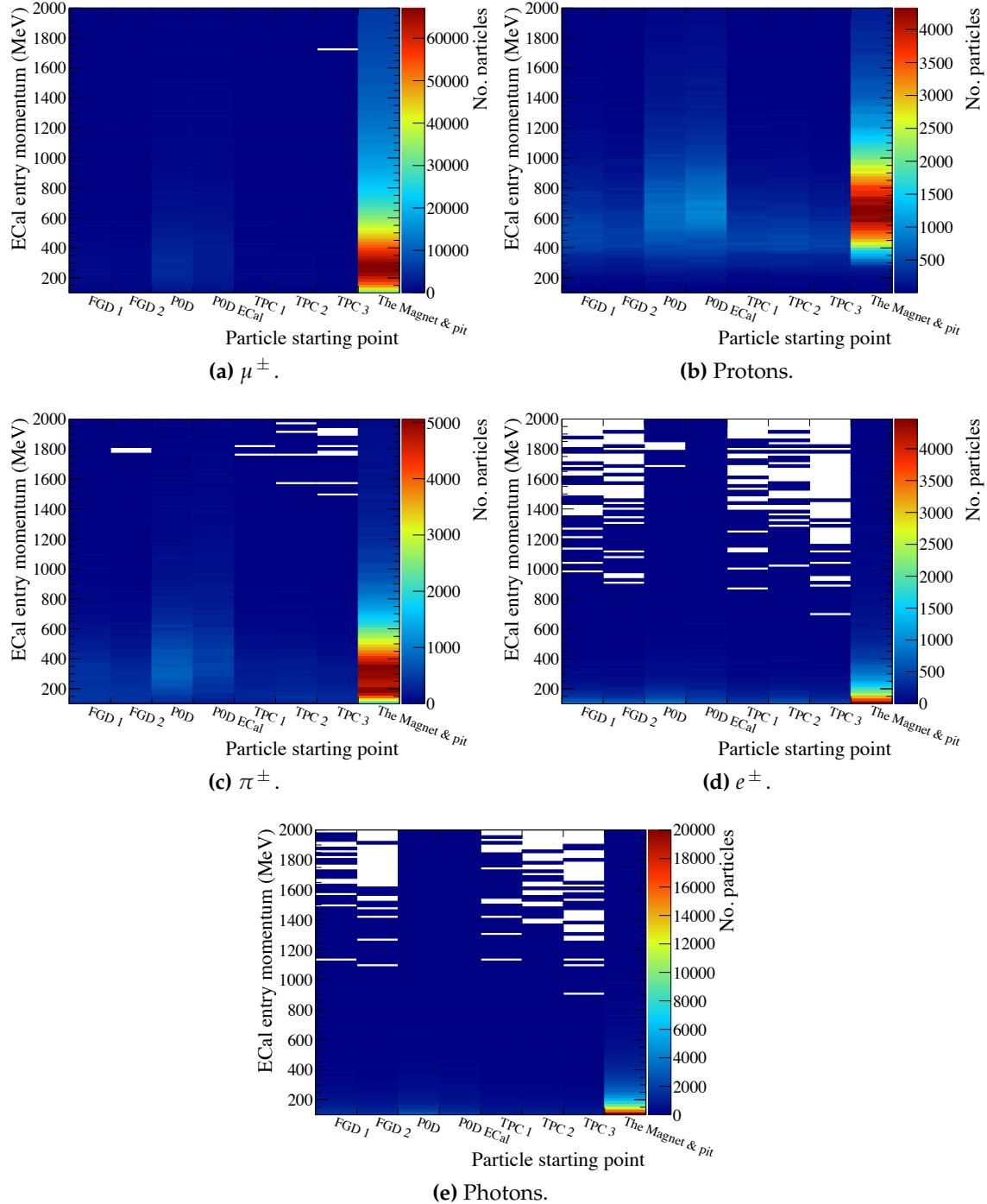


Figure 7.4: The entry momentum of particles entering the barrel ECal Vs. the point of creation of the particles. For all ranges of entance momentum, the dominant point of provenance for all particles is from the magnet and pit regions.

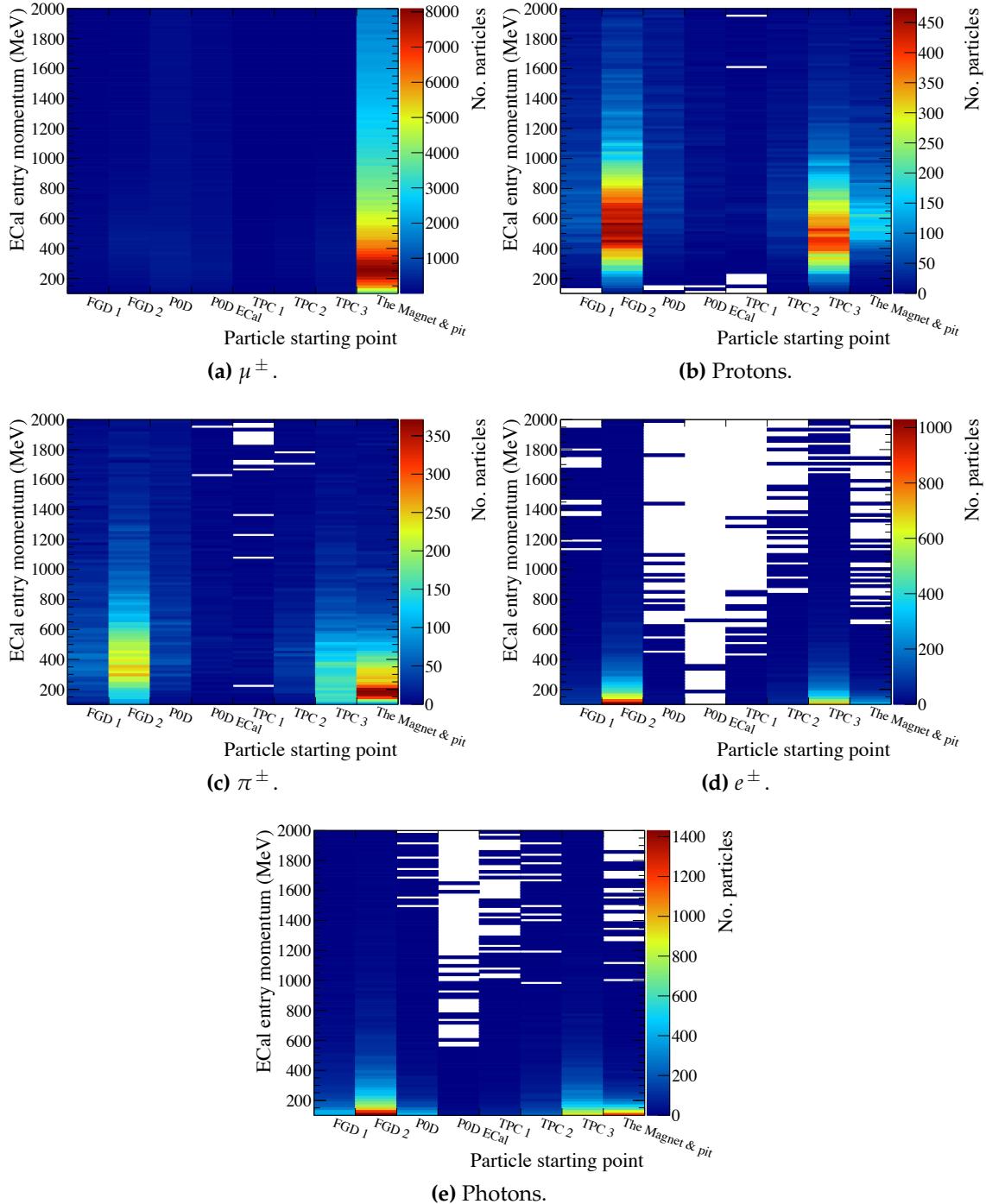


Figure 7.5: The entry momentum of particles entering the DS ECal Vs. the point of creation of the particles. Unlike the barrel ECal case(Fig. 7.4), the creation point of the particles is more varied and includes significant contributions from the downstream end of the tracker (FGD 2 and TPC 3).

spurious pairwise crossing locations that the tracks form.

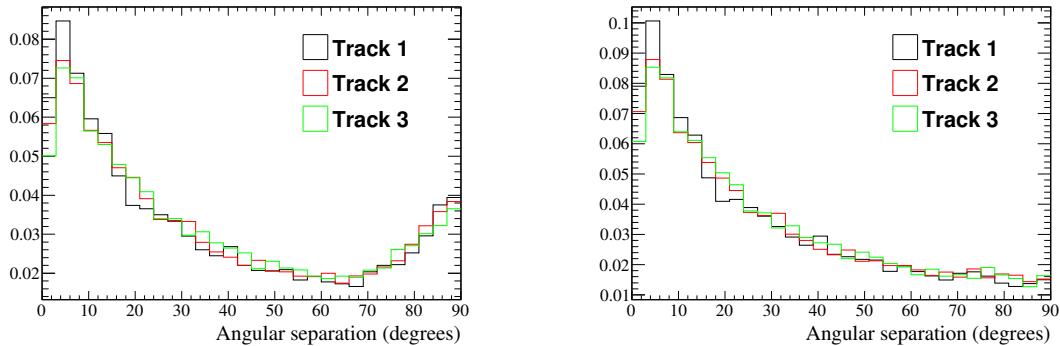
The output of the enhanced reconstruction already supplies most of the necessary information to reconstruct vertices in the ECal. As a reminder, the reconstruction outputs a set of ECal clusters which contain the following:

- A set of 3D reconstructed tracks
- The position at which every pairwise combination of 3D tracks most closely cross (pairwise crossings)

7.4.1 Reconstructed track rejection

The reconstruction makes no quality checks on the 3D tracks in each cluster. So, the first step is to remove any poorly reconstructed tracks from the cluster. The need for this step is shown in Fig. 7.6a which shows the angular separation of reconstructed tracks with the simulated particle which created them, taken from beam Monte Carlo. While the majority of tracks generally have a small angular separation, Fig. 7.6a clearly shows a build up of tracks which are offset by 90° to the simulated particles.

There are two categories of bad track reconstruction. The first is where the 3D matching matches two 2D tracks which do not have sufficient information to fully model a 3D track. Specifically, one of the 2D tracks involved in the matching only uses one ECal layer. As the 3D track reconstruction requires information from both ECal views, tracks which fall into this category do not supply enough information to reconstruct a good quality track. The second category is where the 3D matching produces a grossly incorrect match. The matcher is designed to continually match 2D tracks together until no more 2D tracks are left in the matching pool. So, there are situations where the last possible match made is in no way suitable. This situation is easily identified by matched 2D tracks which do not overlap. For example, one of the 2D tracks uses layers 16, 18 and 20 and the other 2D track uses layers 1 and 3. By removing these two types of tracks from the cluster, the 90° build up in the angular separation distribution is suppressed, as shown in Fig. 7.6b. By removing these categories of tracks, approximately 22% of tracks are rejected.



(a) No track rejection. There is a build-up of reconstructed tracks which are at right angles to the simulated particle that created them.

(b) With track rejection.

Figure 7.6: Angular separation of the reconstructed tracks in an ECal cluster with the beam-simulated particle that created them. The black, red and green histograms are the tracks that were reconstructed first, second and third respectively. All histograms are area normalised.

7.4.2 The vertex reconstruction algorithm

The final states of a neutrino interaction originate from the same point in space. So, assuming that the reconstructed tracks represent the final states, the reconstructed tracks should most closely cross at the point of interaction. As stated above, one of the outputs of the reconstruction are the pairwise crossings of the constituent tracks in the ECal cluster. Using the above assumptions, the pairwise crossings should be in close proximity to one another when the reconstructed tracks represent the final states of an interaction. This idea allows for a relatively simple method of vertex reconstruction: attempt to cluster the pairwise crossings together if the pairwise crossings are in close proximity. To quantitatively define this proximity, the quality of the pairwise crossing also has to be considered.

As with the 3D tracks, the reconstruction does not perform any quality checks on the pairwise crossings of the tracks. As the reconstruction will always find a crossing location for a pair of 3D tracks, some of the pairwise crossings will not represent anything physical. The quality definition chosen is simple: pairwise crossings are defined as bad if the crossing location is far away from either of the tracks it is associated with. However, the distance definition is not trivial to define and should be closely correlated with the vertex reconstruction method. For example, a two track vertex

would provide very little constraint on the distance which defines a bad crossing, whereas a three track vertex may provide a bad crossing distance constraint which is too strict and destroys the majority two track vertices.

The above discussion suggests that two parameters should govern the vertex reconstruction: the required proximity of two crossings to be clustered together, d_c , and the distance of a crossing from its constituents tracks to be classified as bad quality, d_q . To quantify these values, a sample of reconstructed events matched to interactions in the ECals are used which are taken from beam Monte Carlo. The reconstruction (the pairwise crossing rejection and clustering) is repeatedly run over the sample for different values of the d_c and d_q to find the optimum values. To define the optimum value, a figure of merit is necessary. After running the reconstruction for a given d_c and d_q , the number vertices are separated into one, two and three track vertices and the true neutrino interactions which created the reconstructed vertices are associated. A reconstructed vertex is tagged as correctly reconstructed if it contains the same number of reconstructed tracks as the number of charged final states in the associated neutrino interaction. By defining the number of correctly reconstructed one, two and three track vertices as N_1 , N_2 and N_3 respectively, the figure of merit, ϕ^{vertices} , is

$$\phi^{\text{vertices}} = N_1 N_2 N_3. \quad (7.1)$$

By mapping out ϕ^{vertices} in (d_q, d_c) space, information about the preferred values of d_q and d_c can be found. This space is shown in Fig. 7.7. It is clear from Fig. 7.7 that there is no clear maximum, but rather a plateau of ϕ^{vertices} for d_q and d_c greater than 140 mm. So, marginalised distributions of d_q and d_c can be produced to find where ϕ^{vertices} approaches zero which are shown in Fig. 7.8a and Fig. 7.8b respectively. The values chosen are shown in table 7.2. The reconstruction now assesses the quality

of the crossings and then attempts to cluster the good quality crossings together to form vertex candidates. The final step is to use the constituent tracks of each vertex candidate in a fit to estimate the position of the vertex. The following method was

ϕ_d	ϕ_c
$< 140 \text{ mm}$	$< 200 \text{ mm}$

Table 7.2: Parameters for the vertex reconstruction in the ECal.

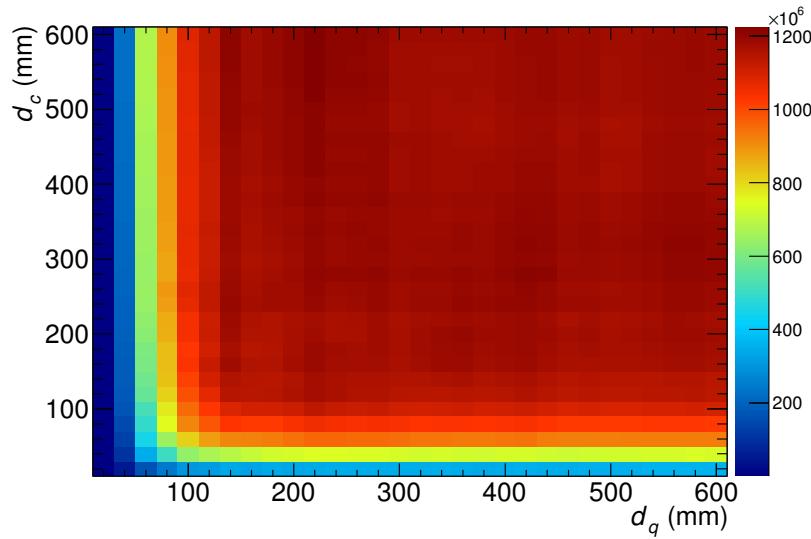


Figure 7.7: Values of ϕ^{vertices} in (d_q, d_c) space. The colour corresponds to the magnitude of ϕ^{vertices} . The distribution plateaus rather than peaks, suggesting that there are no preferred values of d_q and d_c , only disfavoured ones.

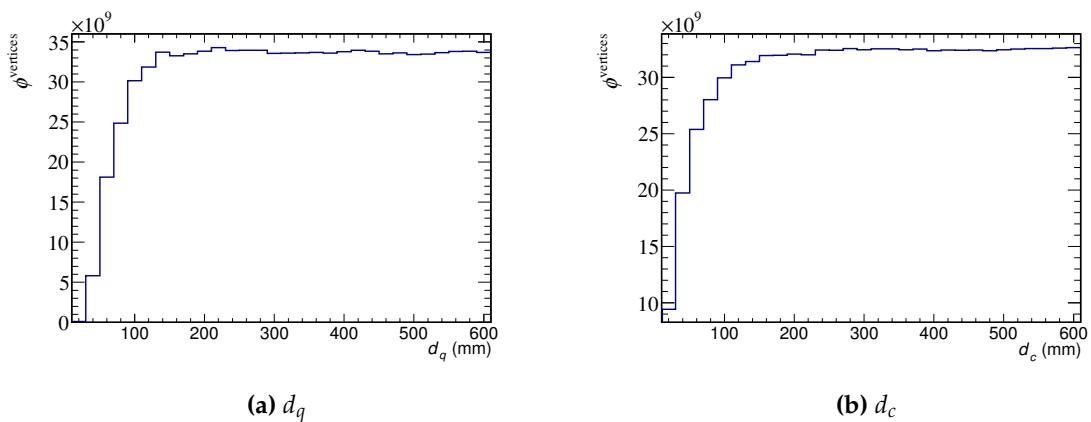


Figure 7.8: ϕ^{vertices} vs the marginalised vertex reconstruction parameters. The same plateau effect shown in Fig. 7.7 can be seen here. The ideal values for both reconstruction parameters are the ones which get as close to the sharp drop in ϕ^{vertices} as possible.

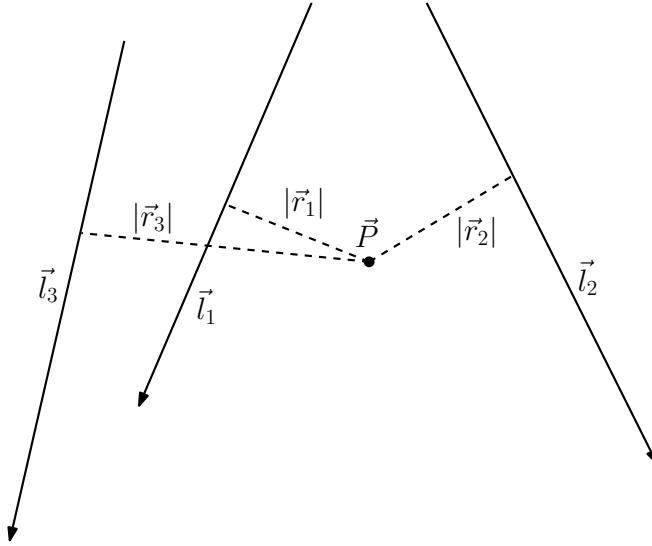


Figure 7.9: Example of the vertex position, \vec{P} for three lines: \vec{l}_1 , \vec{l}_2 and \vec{l}_3 . $|\vec{r}_1|$, $|\vec{r}_2|$ and $|\vec{r}_3|$ are the perpendicular distances of \vec{l}_1 , \vec{l}_2 and \vec{l}_3 to \vec{P} respectively.

suggested by X. Lu [81]. The position of the vertex, \vec{P} , is defined such that the sum of the squares of the distance of each track to \vec{P} is minimised. An example setup of this is shown in Fig. 7.9 for three constituent tracks. By defining the square of the distance of a line, l_i , to \vec{P} as $|\vec{r}_i|^2$, the function to minimise is

$$D = \sum_i |\vec{r}_i|^2. \quad (7.2)$$

\vec{P} is then defined as the point in space which satisfies

$$\frac{\partial D}{\partial x} = \frac{\partial D}{\partial y} = \frac{\partial D}{\partial z} = 0. \quad (7.3)$$

The value of $|\vec{r}_i|$ is trivially defined by simple vector properties as

$$|\vec{r}_i| = \frac{|(\vec{P} - \vec{a}_i) \times \vec{v}_i|}{|\vec{v}_i|} = |(\vec{P} - \vec{a}_i) \times \hat{v}_i|, \quad (7.4)$$

where \vec{v}_i is the direction vector of \vec{l}_i and \vec{a}_i is a point along \vec{l}_i . By defining \vec{P} as

$$\vec{P} = x\hat{i} + y\hat{j} + z\hat{k}, \quad (7.5)$$

equation 7.3 can be solved. The following derivation is only for $\partial D / \partial x$ as the method is identical for $\partial D / \partial x$, $\partial D / \partial y$ and $\partial D / \partial z$.

$$\begin{aligned} \frac{\partial D}{\partial x} &= \sum_i 2\vec{r}_i \cdot \frac{\partial \vec{r}_i}{\partial x} \\ &= 2 \sum_i [(\vec{P} - \vec{a}_i) \times \hat{v}_i] \cdot [\hat{i} \times \hat{v}_i] \\ &= 2x \sum_i (\hat{i} \times \hat{v}_i) \cdot (\hat{i} \times \hat{v}_i) + 2y \sum_i (\hat{j} \times \hat{v}_i) \cdot (\hat{i} \times \hat{v}_i) + 2z \sum_i (\hat{k} \times \hat{v}_i) \cdot (\hat{i} \times \hat{v}_i) \\ &\quad - 2(\vec{a}_i \times \hat{v}_i) \cdot (\hat{i} \times \hat{v}_i). \end{aligned} \quad (7.6)$$

By applying the same steps for $\partial D / \partial y$ and $\partial D / \partial z$, the matrix equation

$$\begin{pmatrix} \sum_i [\hat{i} \times \hat{v}_i] \cdot (\hat{i} \times \hat{v}_i) & \sum_i [\hat{j} \times \hat{v}_i] \cdot (\hat{i} \times \hat{v}_i) & \sum_i [\hat{k} \times \hat{v}_i] \cdot (\hat{i} \times \hat{v}_i) \\ \sum_i [\hat{i} \times \hat{v}_i] \cdot (\hat{j} \times \hat{v}_i) & \sum_i [\hat{j} \times \hat{v}_i] \cdot (\hat{j} \times \hat{v}_i) & \sum_i [\hat{k} \times \hat{v}_i] \cdot (\hat{j} \times \hat{v}_i) \\ \sum_i [\hat{i} \times \hat{v}_i] \cdot (\hat{k} \times \hat{v}_i) & \sum_i [\hat{j} \times \hat{v}_i] \cdot (\hat{k} \times \hat{v}_i) & \sum_i [\hat{k} \times \hat{v}_i] \cdot (\hat{k} \times \hat{v}_i) \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \sum_i [\vec{a}_i \times \hat{v}] \cdot (\hat{i} \times \hat{v}) \\ \sum_i [\vec{a}_i \times \hat{v}] \cdot (\hat{j} \times \hat{v}) \\ \sum_i [\vec{a}_i \times \hat{v}] \cdot (\hat{k} \times \hat{v}) \end{pmatrix} \quad (7.7)$$

can be built up. By defining equation 7.7 as

$$\underline{\underline{A}} \vec{P} = \vec{B}, \quad (7.8)$$

where $\underline{\underline{A}}$ is the matrix on the left hand side of equation 7.7 and \vec{B} is the vector on the right hand side. \vec{P} is finally

$$\vec{P} = \underline{\underline{A}}^{-1} \vec{B}. \quad (7.9)$$

While the inversion of $\underline{\underline{A}}$ is in principle analytically solvable, it was decided that the inversion would be handled numerically. So, to find \vec{P} , the vertex reconstruction builds $\underline{\underline{A}}^{-1}$ (by building and inverting $\underline{\underline{A}}$) and \vec{B} and then applies equation 7.9.

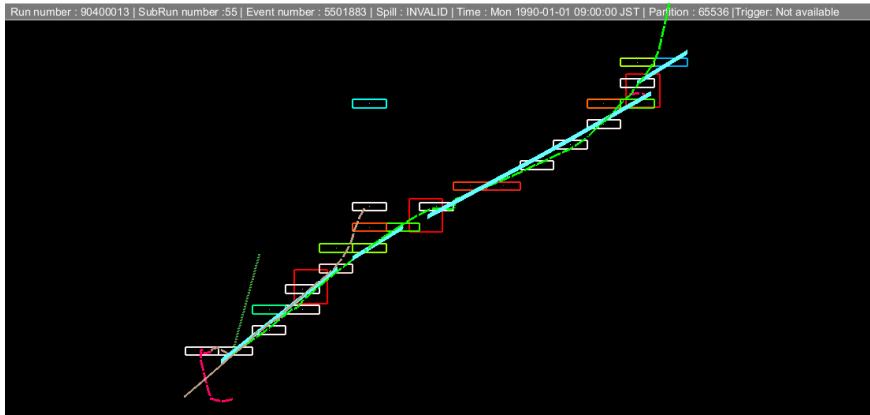


Figure 7.10: Example event display showing a set of track merging candidates. The final state μ^- (in green) curves in a complicated way and is reconstructed as four tracks (in light blue). The pairwise crossings of the reconstructed tracks are the red squares. The muon is produced by a charged current interaction in the top left barrel ECal producing two π^+ (in brown) and the curving μ^- .

7.5 Track merging

By applying the steps outlined above to every reconstructed ECal cluster, a set of candidate vertices are formed. As the enhanced reconstruction is only capable of reconstructing straight tracks, bending trajectories tend to be reconstructed as two or more tracks. The crossings associated with such tracks can, and do, pass the vertex reconstruction criteria defined in table 7.2. An extreme example of this topology is shown in Fig. 7.10 in which a curving μ^- is reconstructed as four tracks. So, the final step of the reconstruction is to attempt to merge tracks to model the curving trajectory topology. Firstly, the reconstructed crossings associated with a clean curving trajectory should pass the pairwise crossing quality cut but should be sufficiently far away from any other crossings that it is not clustered during the vertex clustering stage. Therefore, track merging candidates can be initially identified by searching for reconstructed vertices with exactly two track constituents. As an example, all three of the crossings shown in Fig. 7.10 are correctly identified by this check. The two constituent tracks of the identified vertex will form the merged track, so the reconstruction checks the shape that the merged track would make against a set of conditions to decide if the merging should be performed. To identify and tune the conditions for merging tracks, two metrics are used. The first metric is used for identification of the merging conditions and uses the truth information provided by the Geant4 simulation. For every two track vertex that was formed by the vertex clustering, the simulated particle that produced

each of the two constituent tracks was checked. If the same simulated particle is matched to both tracks, then the merging candidate is tagged as a correct match, otherwise it is tagged as an incorrect match. The second metric is used for tuning the identified merging conditions. As part of this analysis is a search for neutrino interactions using vertex-based reconstruction, there will be an expected loss of signal by track merging which must be minimised. So, the track merging condition tuning should be based on ECal signal interactions. As described above, the only tracks that are proposed as merging candidates are those which are constituents of a two track vertex. So, the track merging tuning attempts to separate signal and background events which are reconstructed as a vertex with two track constituents. It is preferable that the merging chooses quality over quantity so the tuning figure of merit is defined as

$$\phi^{\text{merge}} = \epsilon\eta^2 \quad (7.10)$$

where ϵ is the efficiency of the track merging to keep signal events reconstructed as two track vertices and η is the purity of the events that remain as two track vertices after merging has taken place. The reason for the metric defined in equation 7.10, rather than the more commonly used $\epsilon\eta$, is that the track merging should aim to only merge tracks when there is great evidence that the merge is correct. Specifically, if the merging algorithm used relaxed parameters and passed more two track events for merging, a large amount of signal events which should fall into the 2 prong topology would instead move to the 1 prong topology, where there is less reconstruction information to separate signal from background.

The first merging condition identified is the cosine of the opening angle, $\cos\theta$, subtended by the two constituent tracks bounded between 0 and 1 which is shown in Fig. 7.11. The opening angle is clearly a powerful discriminator. The distribution for incorrect matches is very flat across the full angular range whereas there is a clear build up of correct matches as $\cos\theta \rightarrow 1$.

The second merging condition identified, called ‘distance ratio’, measures the ratio of the distance between the two constituent tracks’ closest points, d^{small} to the distance between their furthest points, d^{large} . An example of how d^{small} and d^{large} are calculated is shown in Fig. 7.12. The distance ratio distribution is shown Fig. 7.13a. While it may seem that there is very little discrimination power present in the dis-

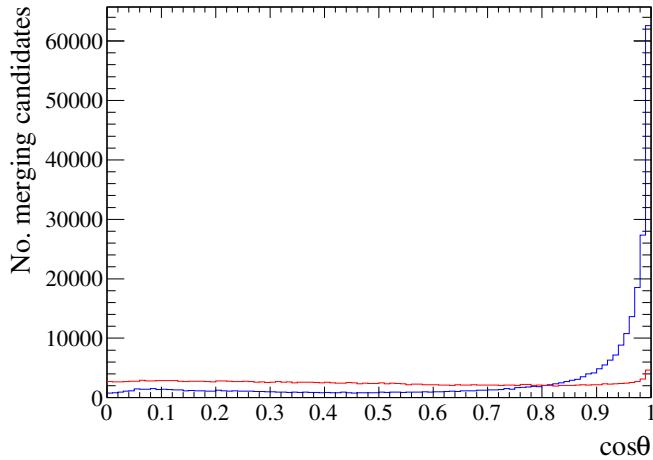


Figure 7.11: The cosine of the angle subtended by the merging candidates, bounded between 0 and 1. The correct matches and incorrect matches are the blue and red histograms respectively. There is a clear build-up of events which are correctly matched as $\cos\theta \rightarrow 1$.

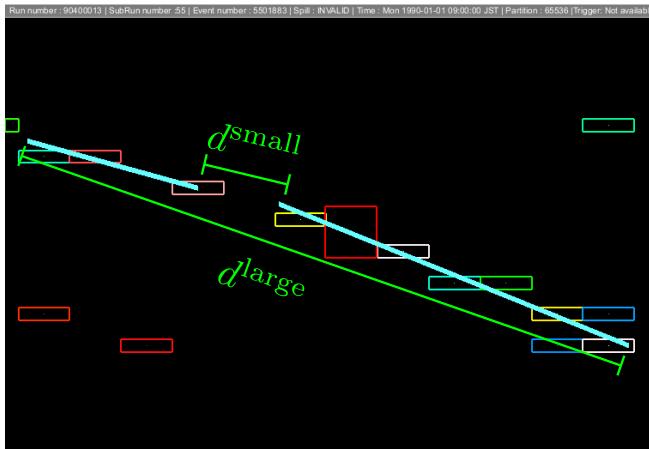


Figure 7.12: Example event display showing how the parameters of the distance ratio are calculated. The distance ratio is defined as d^{small}/d^{large} . The blue lines are the two reconstructed tracks which form the merging candidate and the red square is the crossing location of those tracks.

tance ratio distribution, there is an underlying dependency between the distance ratio and the opening angle subtended by the constituent tracks, which has already been identified as a merging condition. To illustrate this dependency, two distributions are shown in Fig. 7.14 which both show the distance ratio vs the cosine of the opening angle. Fig. 7.14a only shows the correct matches and Fig. 7.14b only shows the incorrect matches. While it is true that there is a pileup of both correct matches and

incorrect matches for low values of the distance ratio, the opening angle separates the two categories out. Specifically, the correct matches pileup occurs as $\cos \theta \rightarrow 1$ and the incorrect matches pileup occurs as $\cos \theta \rightarrow 0$. To further illustrate this point, a new distance ratio distribution is shown in Fig. 7.13b, but with a $\cos \theta > 0.8$ guess cut applied. Comparing the distributions shown in Fig. 7.13a and Fig. 7.13b, the effect of the $\cos \theta$ cut can clearly be seen. The original pileup of incorrect matches seen in Fig. 7.13a is now gone, leaving an essentially flat incorrect matches distribution in Fig. 7.13b while leaving the correct matches structure intact.

By utilising both the distance ratio and $\cos \theta$ simultaneously, a better degree of separation can be found. However care must be taken when tuning the distance ratio and $\cos \theta$ cuts to ensure optimal separation of signal and background is achieved. It is clear from Fig. 7.11 and Fig. 7.13b that the correct matches pileup for low values of the distance ratio and high values of $\cos \theta$. So events should only be tagged for merging when they have a distance ratio value lower than some threshold and a $\cos \theta$ value higher than some other threshold. To find these cut values, the track merging reconstruction was run multiple times, using different values of the thresholds for each run. To take the dependency shown in Fig. 7.14 into account, a square grid search in distance ratio and $\cos \theta$ space was used to find optimum cut values. The tuning metric, as described in equation 7.10, was used to find the optimum cut values. The tuning metric values in distance ratio cut vs $\cos \theta$ cut space are shown in Fig. 7.15. As was

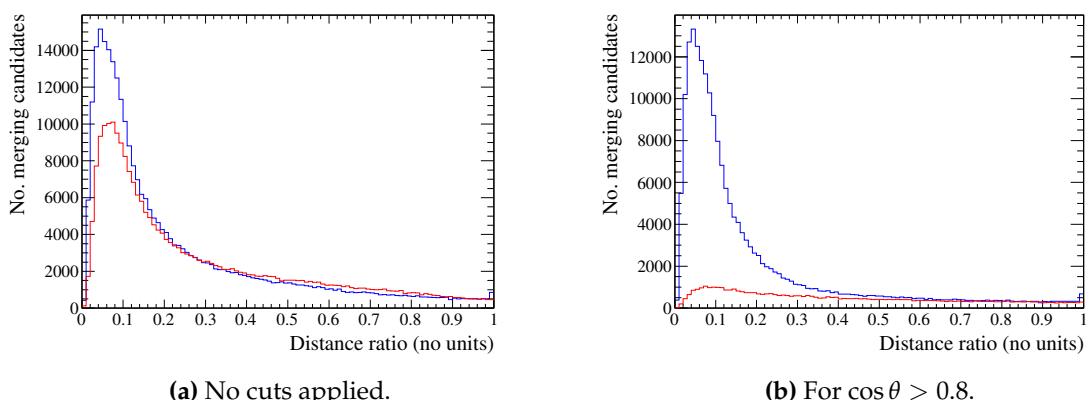


Figure 7.13: The distance ratio of the merging candidates. The correct and incorrect matches are the blue and red histograms respectively. When no cuts are applied, the distance ratio shows little separation between the correctly matched and incorrectly matched events. However, after demanding $\cos \theta > 0.8$, the incorrectly matched peak significantly flattens without altering the correctly matched peak.

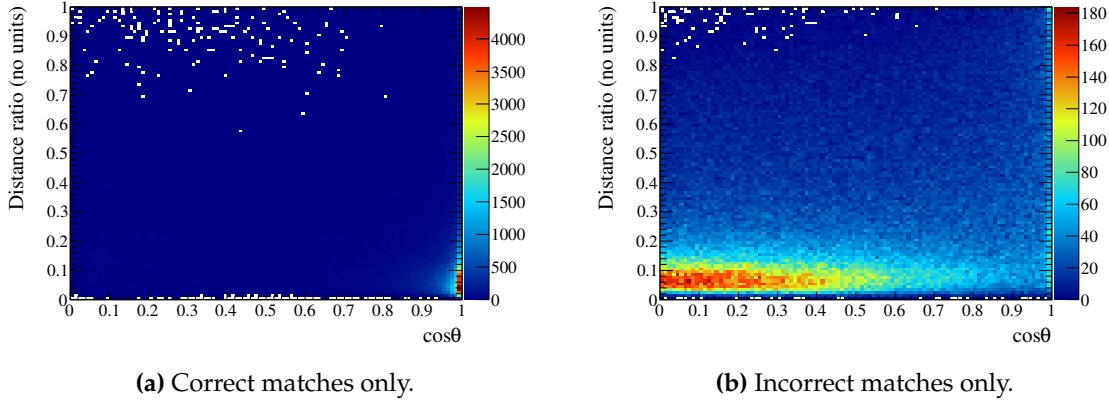


Figure 7.14: The distance ratio vs the cosine of the opening angle for track merging candidates. This 2D space significantly separates out the correctly matched and incorrectly matched events.

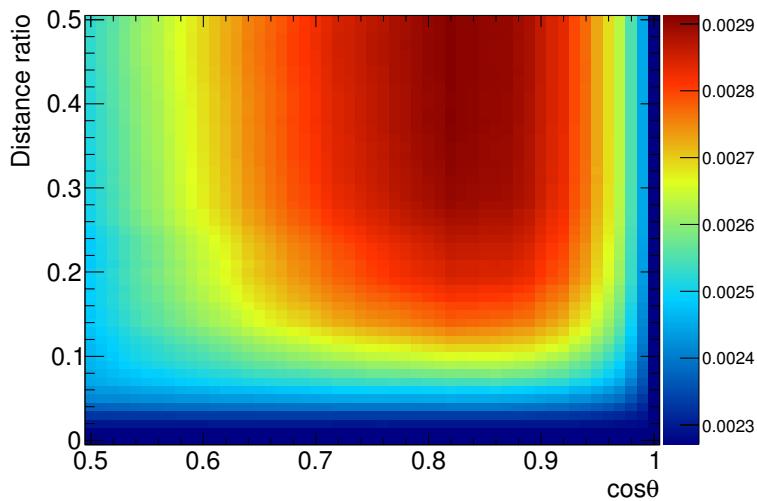


Figure 7.15: Values of the tuning metric, as described in equation 7.10, in distance ratio cut vs $\cos \theta$ cut space. The peak of this distribution is very broad, suggesting a range of preferred values for the distance ratio and $\cos \theta$ cuts.

found in the tuning of the vertex clustering parameters, there is no clear maximum value of the tuning metric, but rather a plateau. So, as was done in the vertex clustering tuning, marginalised distributions of tuning metric in distance ratio cut space and $\cos \theta$ cut space can be produced to find the optimum cut values. The marginalised distributions for $\cos \theta$ and the distance ratio cuts are shown in Fig. 7.16a and Fig. 7.16b respectively. In the case of the $\cos \theta$ cut there is a clearly preferred value, merging candidates should only be merged if $\cos \theta > 0.82$. In the case of the distance ratio cut

distribution, there is less of a clear maximum. As the reconstruction is striving for quality over quantity, the cut value should be fairly close to the drop in ϕ^{merge} . It was decided that merging candidates should only be merged if the distance ratio is less than 0.32.

While the $\cos \theta$ condition is clearly a powerful discriminator, there is a topology degeneracy which $\cos \theta$ is not capable of separating, which is shown in Fig. 7.17. The diagram on the left of Fig. 7.17 is a representation of a signal event which is reconstructed as two tracks, whereas the diagram on the right of Fig. 7.17 represents a curving trajectory reconstructed as two tracks. Importantly, the same opening angle is measured for both situations which is small enough that both topologies pass the $\cos \theta$ track merging condition. To rectify this, an extra sanity check is needed when considering merging candidates. So, the final merging condition identified, called 'swing', measures the rotation of one track relative to the other. Specifically, the swing is the ratio of the longest track length, l^{long} , to d^{large} (the same d^{large} that was used in the distance ratio calculation). An example of how these values are calculated is shown in Fig. 7.18. Provided that the opening angle of a merging candidate is not near 90° , the swing should be less than one for the topology on the left of Fig. 7.17 and should be greater than one for the topology on the right of Fig. 7.17. The separation power of the swing parameter is shown in Fig. 7.19. After applying the merging conditions discussed above, the swing parameter becomes bi-modal as shown in Fig. 7.19b. The swing parameter is introduced purely as an extra sanity cut which is physically

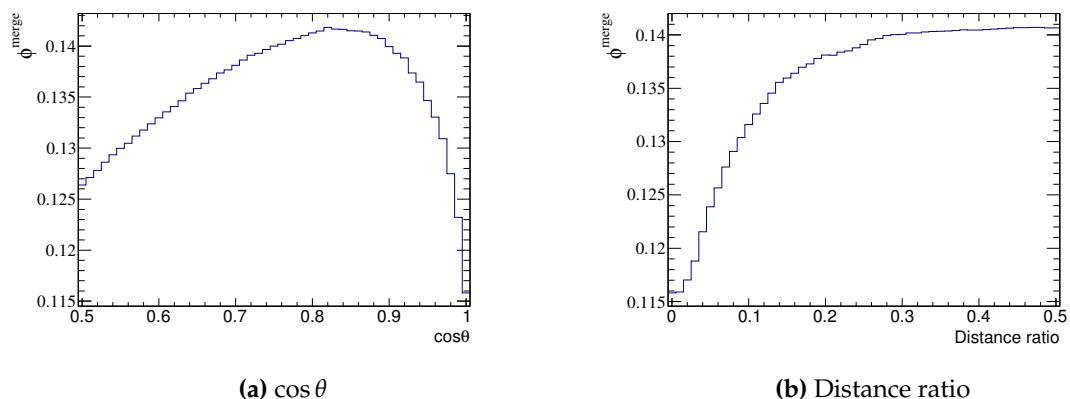


Figure 7.16: ϕ^{merge} vs the track merging conditions. There is a clear but somewhat broad maximum of ϕ^{merge} in $\cos \theta$ space whereas there is only a plateau in distance ratio space.

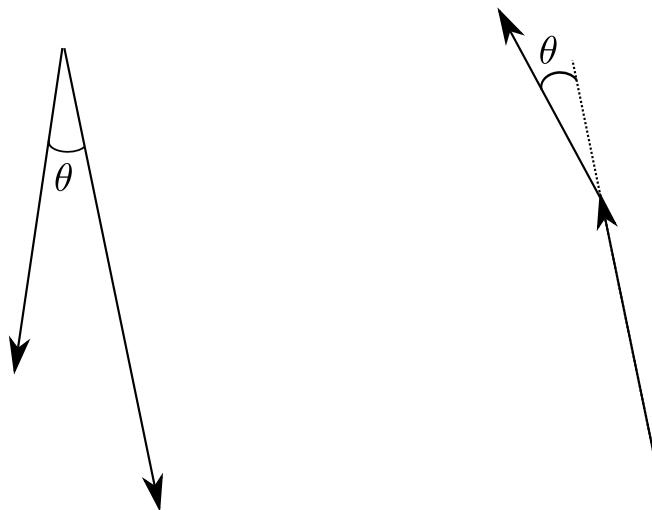


Figure 7.17: Schematic showing the degeneracy of two merging candidate topologies which would pass the $\cos \theta$ cut. The arrows represent reconstructed tracks and θ is the opening angle measured between those tracks.

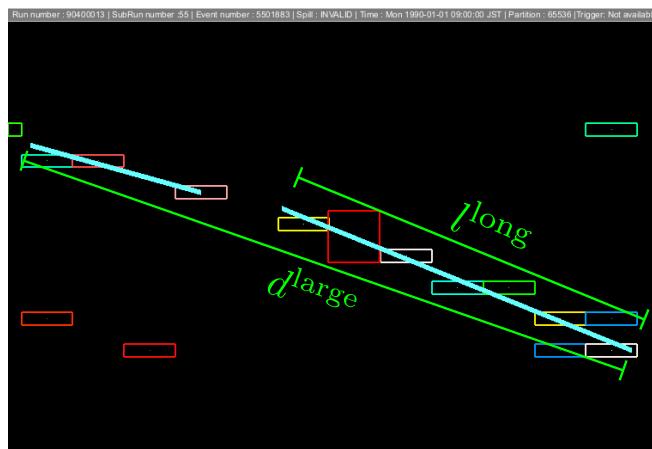


Figure 7.18: Example event display showing how the inputs to the swing parameter are calculated. Swing is defined as $l^{\text{long}} / d^{\text{large}}$. The blue lines are the two reconstructed tracks which form the merging candidate and the red square is the crossing location of those tracks.

motivated and requires no tuning. So, merging candidates are only merged if the measured swing is less than one.

All of the merging conditions have now been identified and the associated cut values are summarised in table 7.3.

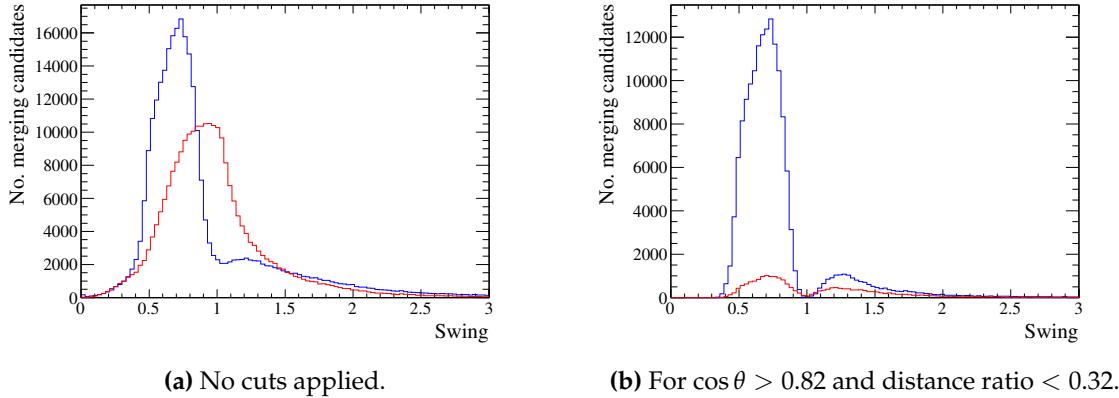


Figure 7.19: The number of merging candidates as a function of the swing parameter. The correct matches and incorrect matches are the blue and red histograms respectively. After applying the $\cos \theta$ and distance ratio cuts, the swing parameter distribution becomes bimodal, signifying that the swing parameter has separated out the two merging candidate topologies shown in Fig. 7.17.

$\cos \theta$	Distance ratio	swing
> 0.82	< 0.32	< 1

Table 7.3: Cut values for the track merging in the ECal.

7.6 Monte Carlo selection

As briefly described at the start of this chapter, the selection separates the reconstructed ECal events into a set of topologies, where each topology is defined by a number of associated prongs. In addition to this separation, the geometrical differences between the barrel and DS ECal suggest a further event separation. The DS ECal lies perpendicular to the beam. So, particles from neutrino interactions will typically travel at right angles to the DS ECal face. Conversely, all of the barrel ECal lie parallel to the beam axis which means neutrino interaction products will generally pass along the barrel scintillator planes. This motivates a separate selection for the barrel and DS ECal. This essentially means that there are eight individual selections to be made and tuned (1, 2, 3 and 4+ prong topologies in the barrel and 1, 2, 3 and 4+ prong topologies in the DS ECal). Clearly, this approach can very quickly become complicated. To mitigate this, similar discriminators are used for each topology. As the reconstruction was only tuned for three reconstructed tracks, the 4+ prong topology will use exactly the same cuts as the 3 prong topology.

Care must be taken when defining what a reconstructed vertex is. For the 2, 3 and 4+ prong topologies, it is fairly simple: the reconstructed vertex is as described in section 7.4, but it is not possible to fit for a vertex position when dealing with a single prong. In fact, there is little that can be done about this. So, beam kinematics are assumed and the ‘vertex’ for a single prong is the most upstream end of said prong.

The definition of signal has already been described in section 7.1. However, how the reconstructed events map to the signal interactions also needs discussion. The selection takes place at the reconstructed vertex level but there is not a clear 1:1 map of reconstructed vertices to signal interactions. Consider a signal interaction in an ECal module in which the final state particles are involved in one or more secondary interactions. The most likely outcome of this is the reconstruction of two or more vertices: one for the neutrino interaction and one or more for the secondary interactions. By interrogating the associated truth information, each vertex will be matched to the same signal neutrino. It is wholly incorrect to classify each vertex as a reconstruction of a signal event as this will lead to a gross overestimate of the signal rate. To alleviate this, a reconstructed vertex is only tagged as coming from a signal interaction if it obeys the following conditions:

- The reconstructed vertex is matched to a signal interaction
- The reconstructed vertex is the one closest in space to the matched signal interaction

Any reconstructed vertices which pass the first condition but fail the second are tagged as a special case of background events (hereafter referred to as the ‘split signal’ background).

To illustrate the benefit of separating events into specific prong topologies, Fig. 7.20 shows the truth makeup of events seen in the ECal after applying the reconstruction separation. It is important to note that only the vertex reconstruction and track merging has been applied to the sample at this point. There is a clear difference in the level of backgrounds seen in each topology e.g. the ECal OOFV background is almost exclusively contained in the one prong topology. By applying such a separation and then focusing cuts on each topology, the overall event purity should be higher.

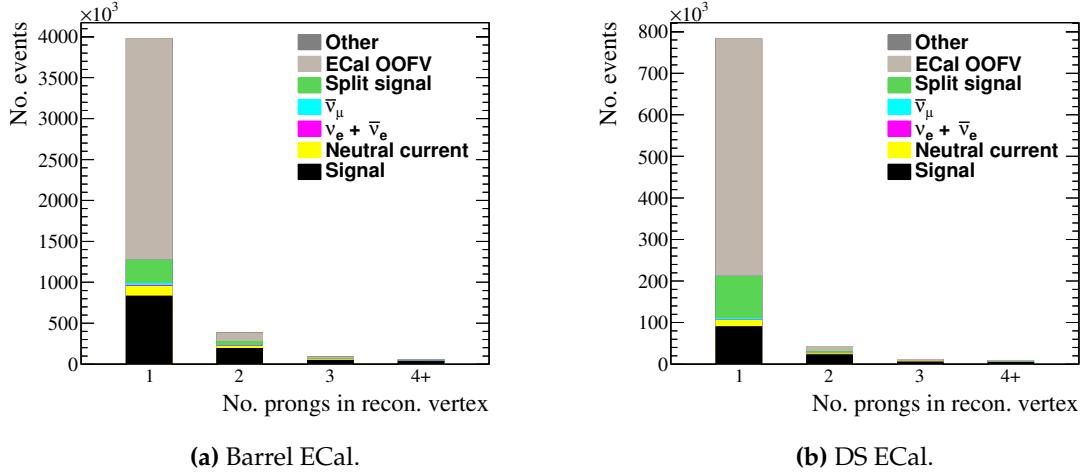


Figure 7.20: The number of reconstructed events in the Monte Carlo sample, separated into the prong topologies. Each event is categorised by the associated truth information from the simulation. The number of reconstructed events and the background contamination strongly depends on the number of reconstructed prongs. This effect indicates that selection development should be tailored towards each prong topology individually.

7.6.1 Selection cuts

The enhanced reconstruction was developed with an ECal cross-section analysis in mind. So, most of complex work should already have been handled by the reconstruction aspect of the analysis. In addition, the selection method discussed involves separating events into prong topologies and focusing individual selections on each. These points motivate a simple, cut-based, selection. Some cuts will inevitably require some tuning and a metric is often useful for this purpose. As demonstrated in section 7.2, a high level of statistics is seen in each ECal module which means the selection can safely strive for quality over quantity without inflating the final uncertainty. Bearing this in mind, the metric used for tuning for prong topology i is

$$\phi_i^{\text{selection}} = \epsilon_i \eta_i^2, \quad (7.11)$$

where ϵ_i and η_i are the selection efficiency and purity of prong topology i . Six selection cuts have been identified which will now be discussed.

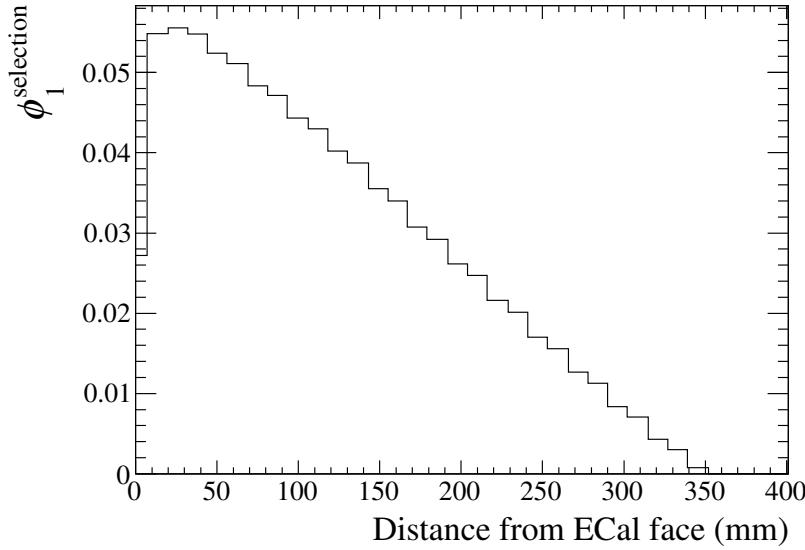


Figure 7.21: $\phi_1^{\text{selection}}$ as a function of the distance from the $-y$ face (bottom face) of the bottom-left barrel ECal for the 1 prong topology. There is a clear maximum of $\phi_1^{\text{selection}}$ which indicates that there is a preferred minimum distance of 1 prong vertices from the $-y$ face.

7.6.1.1 Fiducial volume cut

The first cut, the fiducial volume cut, removes most of the ECal OOFV backgrounds in the Monte Carlo sample. This is achieved by defining some outer veto region in which occurring vertices are rejected. There are two items to bear in mind when defining what the fiducial volumes are. Firstly, the off-axis configuration causes each ECal to be exposed to a different energy and particle rate. A good example of this was shown in Fig. 7.2 which shows a higher event rate, but also a higher peak energy of interactions, in the bottom-left ECal than in the top-right ECal. So, the optimised fiducial volume of one ECal need not be the same as another ECal module. Separately to this, Fig. 7.20 shows that the ECal OOFV is almost completely contained in the 1 prong topology bin. This means that a separate fiducial volume definition should be used for the 1 prong topology (it is sufficient for the 2, 3 and 4+ prong topologies to use the same fiducial volume definition).

The fiducial volume for a prong topology/ECal module is defined in terms of the distance from each face of the module. So, there are six numbers which define the fiducial volume: the distance from the $\pm x$, $\pm y$ and $\pm z$ faces of a module. If the

	$-x$ (mm)	$+x$ (mm)	$-y$ (mm)	$+y$ (mm)	$-z$ (mm)	$+z$ (mm)
Bottom-right	36	36	31	29	206	28
Side-right	31	29	46	44	206	28
Top-right	39	27	29	31	172	28
Bottom-left	25	50	31	29	205	28
Side-left	29	31	50	25	207	28
Top-left	21	34	29	31	207	28
Downstream	29	29	29	30	41	43

Table 7.4: The fiducial volume definitions for each ECal module for the 1 prong topology. The strongest cuts are for the $\pm y$ ECal faces (barrel modules only) which rejects 1 prong vertices which occur in the inner or outer two layers. For the DS ECal, the $\pm z$ face cut is the strongest which rejects 1 prong vertices which occur in the inner or outer three layers.

distance of a vertex from each ECal face is further than what the fiducial volume defines, the event passes the cut, otherwise it is rejected. To tune these values, each fiducial volume value was systematically increased from 0 mm to 600 mm in 1 mm increments. At each increment the number of passing/rejected events was recorded and $\phi_i^{\text{selection}}$ calculated. The fiducial volume value which maximises $\phi_i^{\text{selection}}$ is accepted as the optimised value. An example of the variation in $\phi_i^{\text{selection}}$ is shown in Fig. 7.21 which shows $\phi_i^{\text{selection}}$ as a function of the distance from the $-y$ face of the bottom-left barrel ECal for the 1 prong topology. There is a clear maximum found which shows the process is effective in finding an optimum distance from the ECal face.

The found fiducial volumes for each ECal module for the 1 prong topology are shown in table 7.4. While all of the ECal face cuts have some discriminatory power, those which cut out according to layer number reject the most background events. Those cuts are the $\pm y$ cuts for the bottom/top barrel ECal, the $\pm x$ cuts for the side ECals and the $\pm z$ cuts for the DS ECal. In the barrel ECal, the inner and outer two layers are removed by the fiducial volume cut and in the DS ECal, the inner and outer three layers are removed. As described above, the same fiducial volumes are used for the

2, 3 and 4+ prong topologies. So, to tune the fiducial volume cuts, the topologies are temporarily combined and the combined sample is used to tune the cuts. Those fiducial volume cuts are shown in table 7.5. There are clear differences between the fiducial volumes shown in table 7.4 and table 7.5. The cuts suggested by the 2, 3 and

	$-x$ (mm)	$+x$ (mm)	$-y$ (mm)	$+y$ (mm)	$-z$ (mm)	$+z$ (mm)
Bottom-right	19	9	16	0	3	39
Side-right	20	0	6	9	7	65
Top-right	15	9	4	3	5	38
Bottom-left	11	1	34	0	9	69
Side-left	1	4	13	1	3	67
Top-left	11	0	0	7	4	45
Downstream	13	9	13	4	0	64

Table 7.5: The fiducial volume definitions for each ECal module for the 2, 3 and 4+ prong topologies. The only strong cuts for the 2+ prong topologies are the $+z$ ECal face cuts (all modules). The power of the $+z$ face cuts is due to secondary interactions having a preference for occurring downstream of the neutrino interactions, leading to background-tagged, reconstructed vertices positioned towards the $+z$ of each ECal module.

4+ prong topology tunings are not as strict as those suggested by the 1 prong topology. This should be expected, as ECal OOFV backgrounds reconstructed as 2, 3 or 4+ prong vertices will most likely not have their vertex reconstructed near the faces of an ECal module. The one exception to this is the downstream face of each ECal module. It is clear from table 7.4, that the 2, 3 and 4+ prong topologies suggest a stricter downstream face cut ($+z$ cut) than the 1 prong topology. The reason for this strict cut is due to the nature of the neutrino beam. J-PARC’s neutrino beam is (almost) parallel with the $+z$ axis defined by the ND280 coordinate system. Because of the very forward nature of the beam, any final state particles from ECal neutrino interactions are most likely to travel downstream. So, any subsequent secondary interactions, which are tagged as background, will also occur downstream. The $+z$ face cut attempts to remove these.

7.6.1.2 Visible energy cut

The second cut rejects events based on the amount of deposited charge which is associated to the constituent prongs. The focus of this cut is to remove particle showers (typically e^\pm and γ) which have been reconstructed as a vertex. For the one prong topology, this information is trivial to evaluate: find which scintillator hits are associated to the single prong and sum their respective charge deposit. The situation is not as clear for a vertex with multiple prongs associated. The variable chosen in this case

is the total charge associated to all of the constituent prongs. To further increase the power of this variable as a discriminator, the total prong charge is compared to the total number of scintillator hits associated to the prongs. The aim is to find a 2 dimensional cut which rejects events based on their associated charge and hit information.

The signal and background distributions for the barrel ECal, 1 prong topology are shown in Fig. 7.22a and Fig. 7.22b respectively. To illuminate the separation power, the background/signal ratio is shown in Fig. 7.22c, with the ratio truncated at 2. By applying the truncation, areas where the background contamination is at least two times as large as the signal population is easily found. To develop the 2 dimensional cuts, a test line is drawn which appears to roughly cut out the highly contaminated (red) areas in Fig. 7.22c. Then, to tune the cut line, the parameters of said line are varied in a grid search. At each point in the grid, $\phi_i^{\text{selection}}$ is recorded. The maximum value in this grid corresponds to the optimised parameters of the cut line. As is evident from Fig. 7.22c, there are two areas of high background contamination which motivated the development two cut lines which are both overlaid on Fig. 7.22c.

Fig. 7.23 and Fig. 7.24 show similar distributions for the 2 and 3 prong topology in the barrel ECal. The ratio distributions also have the cut lines overlaid. An identical process was used to generate and tune the cut lines. By using a definition for the total number of scintillator hits associated to all constituent prongs in a vertex, N , the definition of the cut lines are shown in table 7.6.

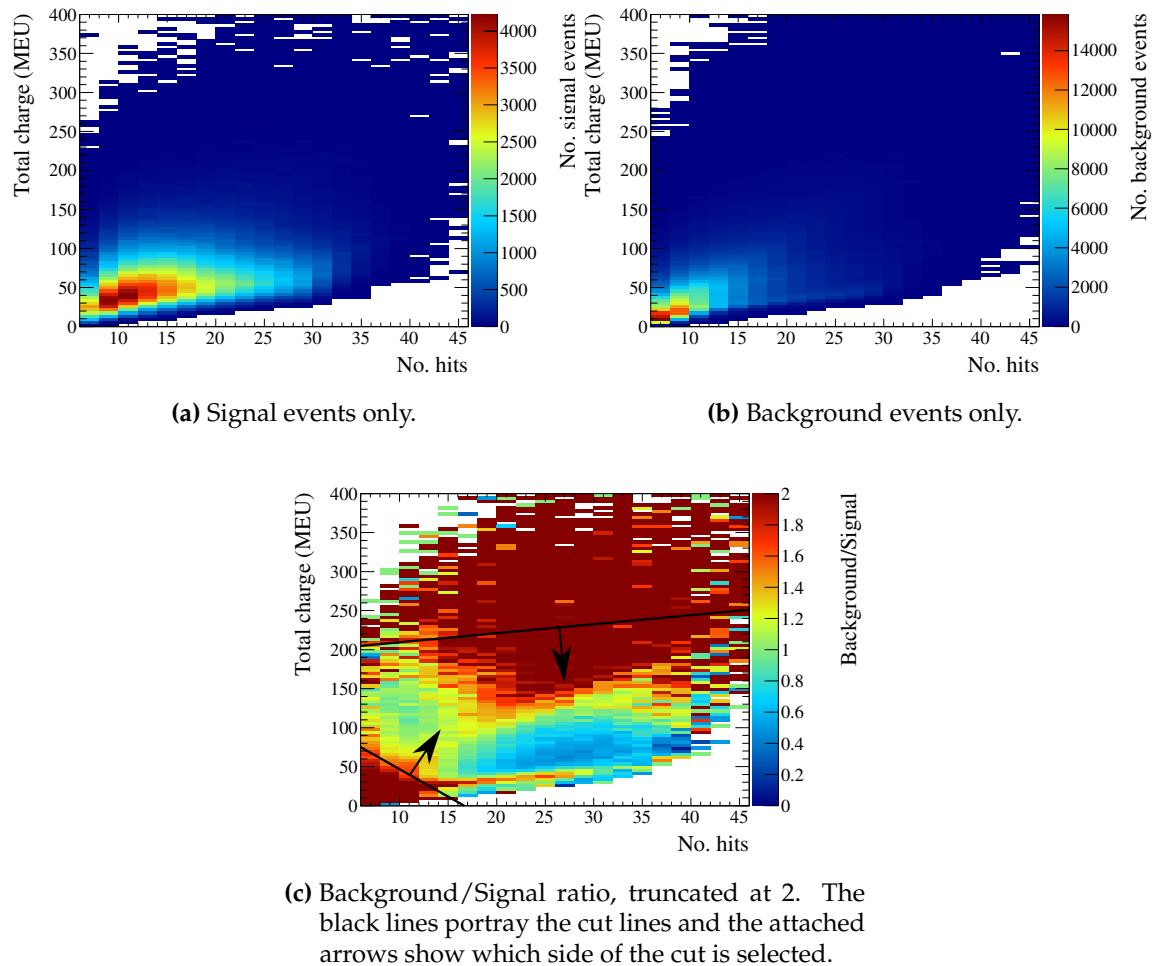


Figure 7.22: Total prong charge vs the number of prong hits for barrel ECal events in the 1 prong topology.

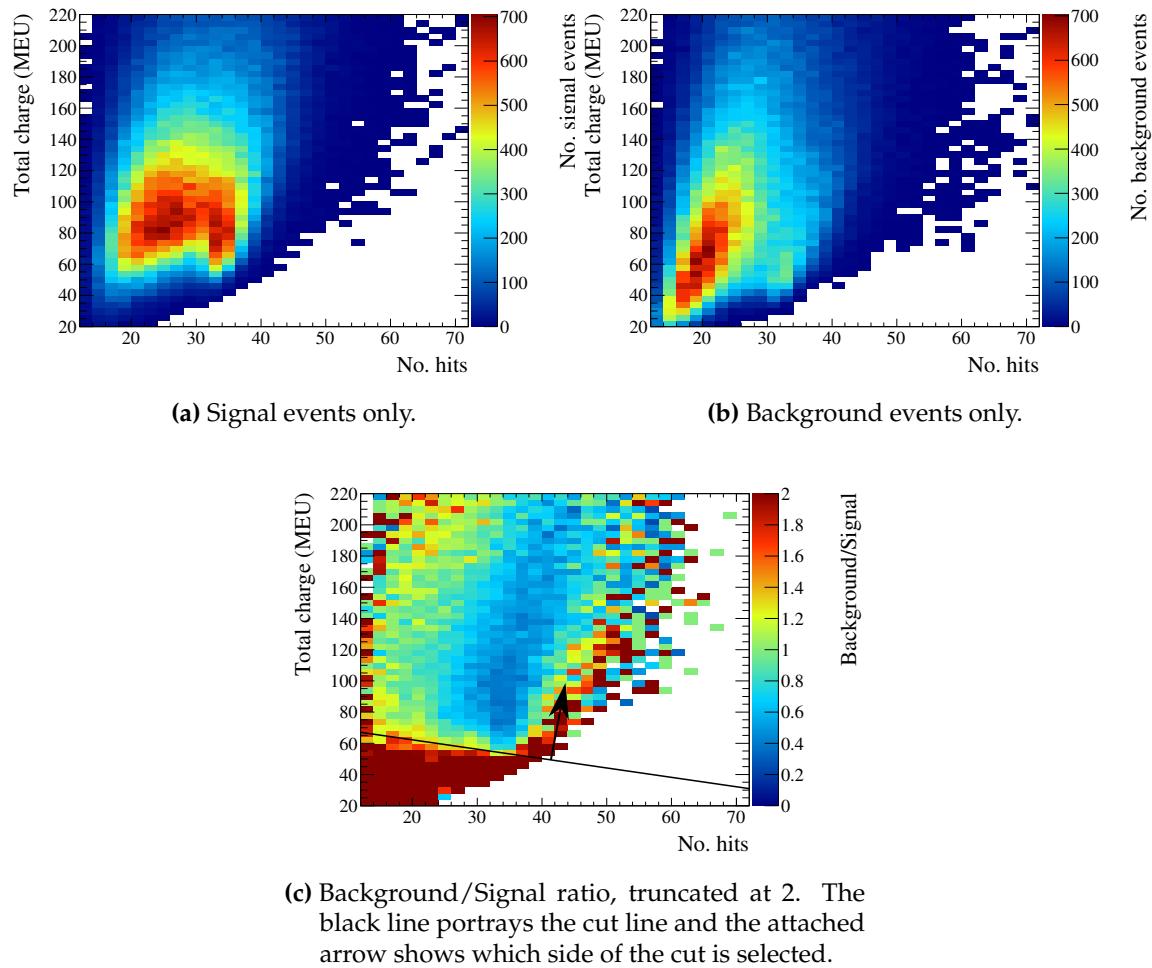


Figure 7.23: Total prong charge vs the number of prong hits for barrel ECal events in the 2 prong topology.

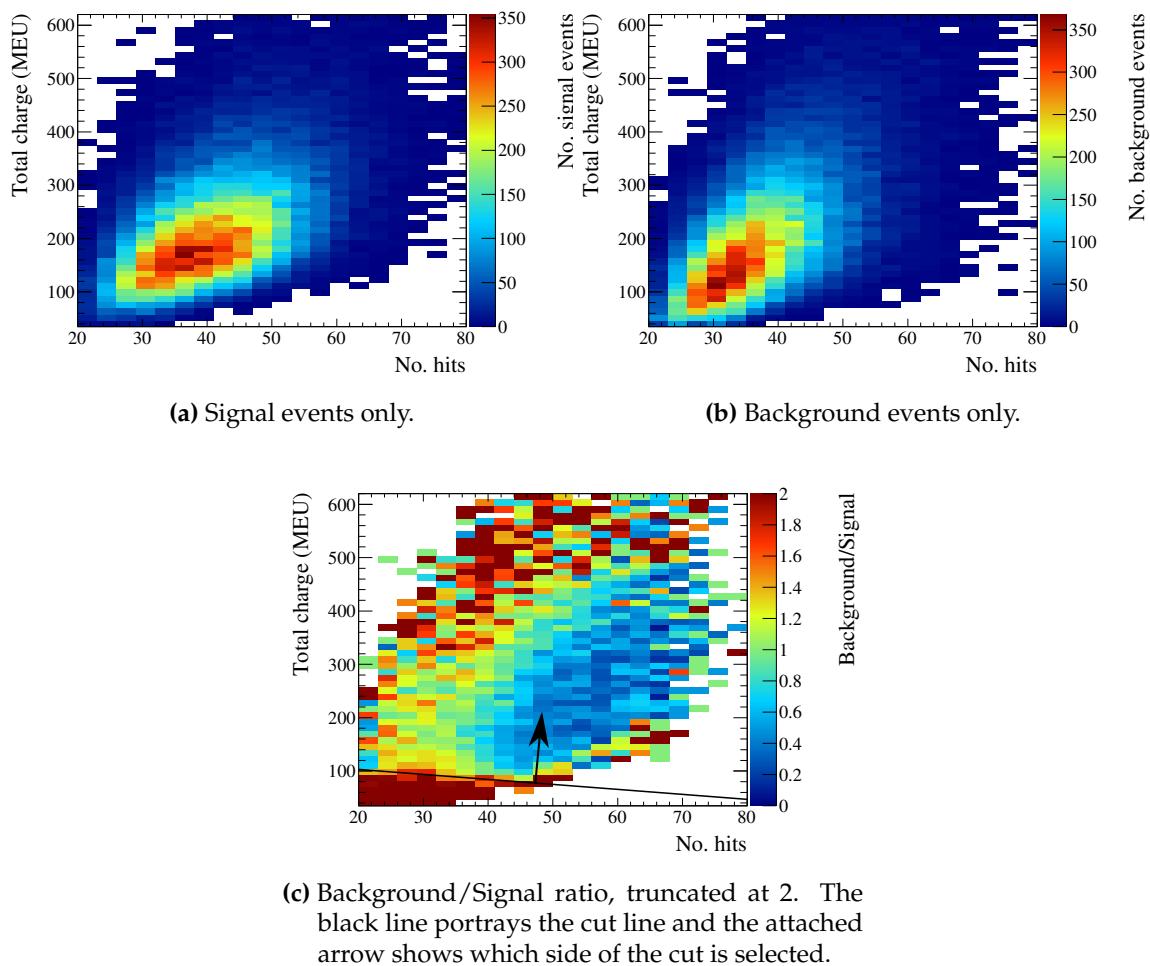


Figure 7.24: Total prong charge vs the number of prong hits for barrel ECal events in the 3 prong topology.

Prong topology	Barrel cuts (MEU)	DS cuts (MEU)
1	$> -4.5N + 75$	$> -3.4N + 62$
	$< N + 205$	
2	$> -0.5N + 67$	$> 0.2N + 51$
3	$> 0.7N + 103$	$> -1.1N + 117$

Table 7.6: The definitions of the 2 dimensional cut lines involving the total prong charge and the number of associated scintillator hits. N refers to the total number of scintillator hits associated to the constituent prongs in the vertex.

7.6.1.3 Unused hits cut

The output of the enhanced reconstruction is a set of reconstructed clusters. Within those clusters are a set of tracks and their pairwise crossings which are used to reconstruct the ECal vertices used by this analysis. The clusters which contain the tracks and crossings can also be utilised as a discriminator. Particle showers should generally produce a lot of hits and those hits are not necessarily arranged in a linear fashion. So, the third cut implemented assesses how many hits are associated with a cluster but not associated with the vertex. The vertex is defined by its prong constituents so this is really a measure of how many scintillator hits were not associated with the reconstructed prongs. By defining the number of hits in the cluster as N^{Cluster} and the number of hits associated with the i^{th} constituent prong as N_i^{prong} , the discriminator is

$$\Delta N = N^{\text{cluster}} - \sum_i N_i^{\text{prong}}. \quad (7.12)$$

The cut value really should depend on how many hits are associated to the constituent prongs. So, as was done in section 7.6.1.2, the ΔN cut should be 2 dimensional. Specifically, the ΔN cut should depend on the number of scintillator hits associated to the constituent prongs.

So, an identical approach is used to define and tune the cut lines as was used in section 7.6.1.2. The signal, background and background/signal ratio distributions for the 1, 2 and 3 prong topologies in the barrel ECal are shown in Fig. 7.25, Fig. 7.26 and Fig. 7.27 respectively. The layout of the figures is the same as was shown for the visible energy cut. In a similar manner to table 7.6, the ΔN cut definitions are shown in table 7.7.

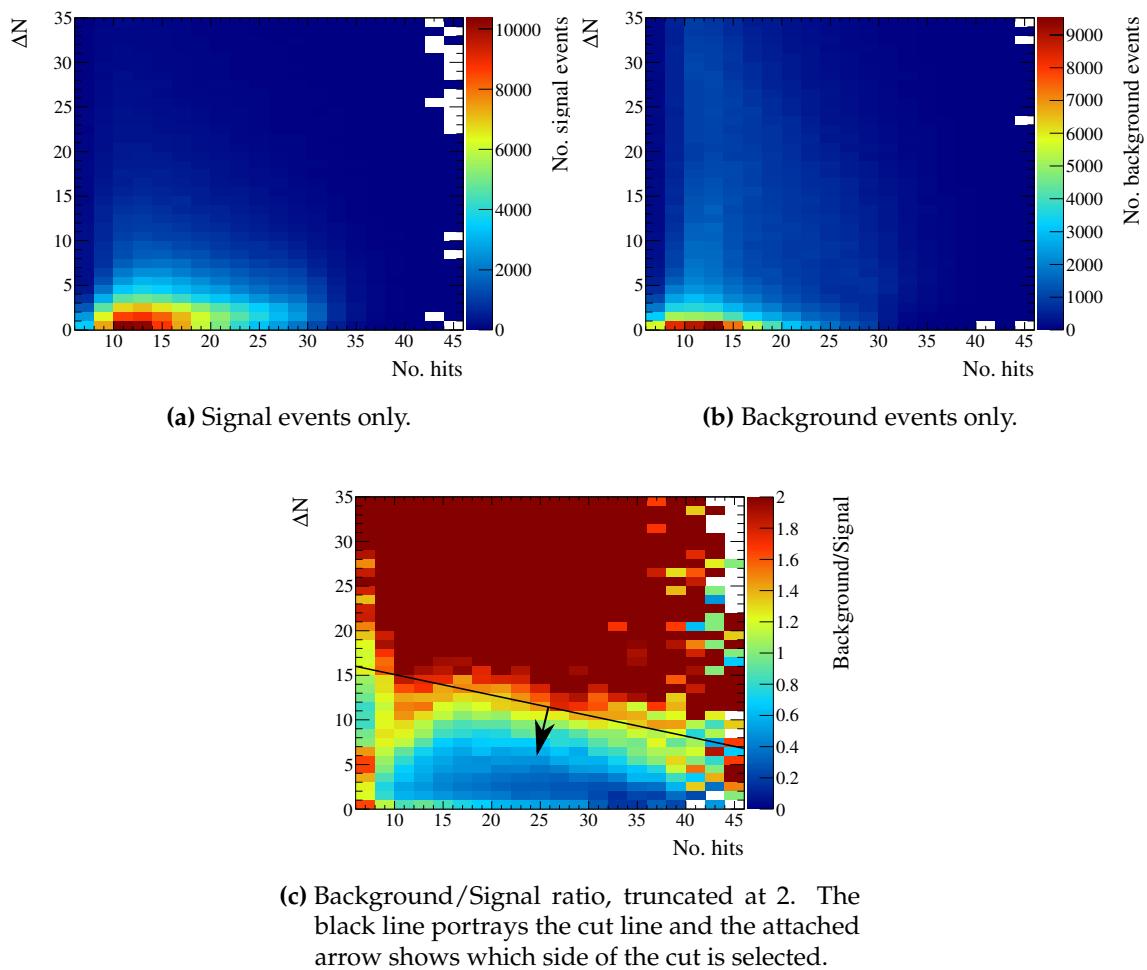


Figure 7.25: ΔN vs the number of prong hits for barrel ECal events in the 1 prong topology.

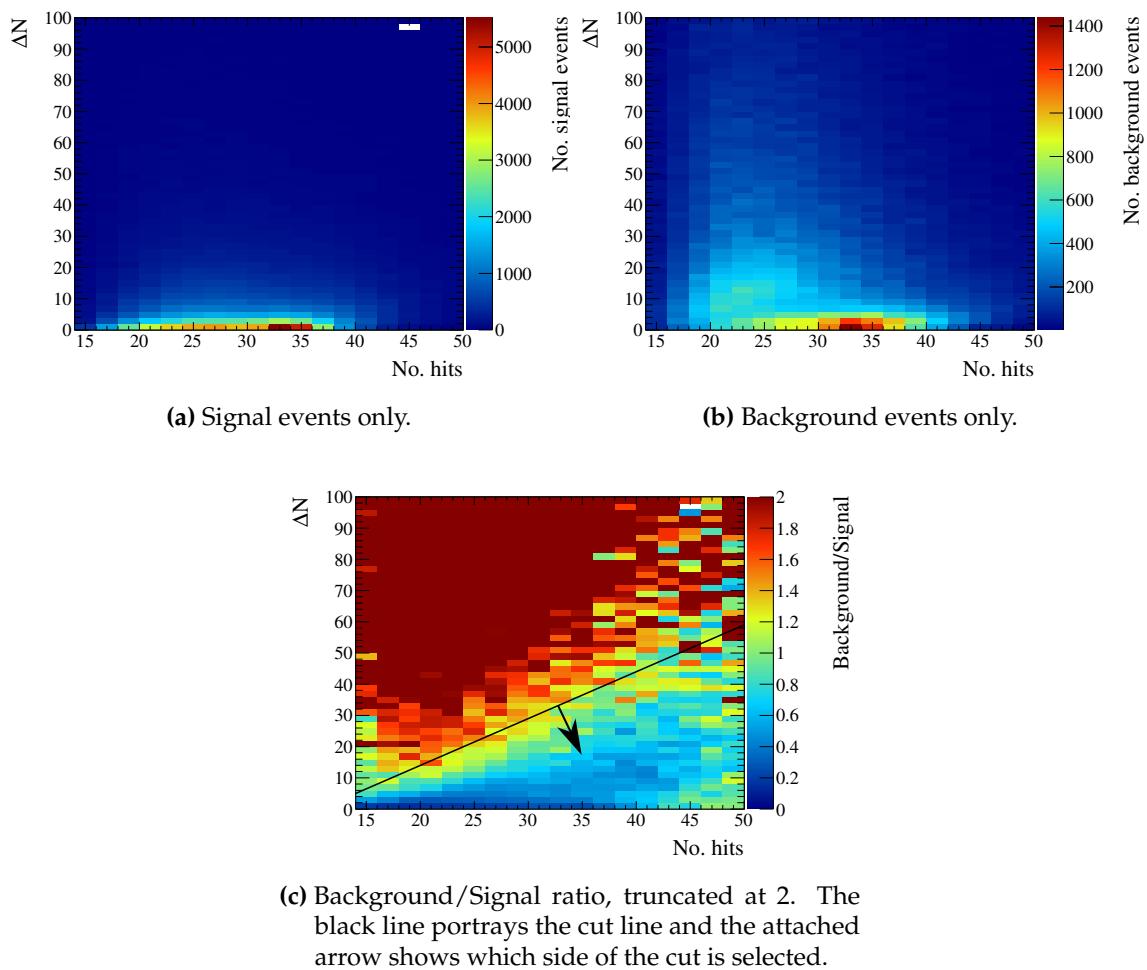


Figure 7.26: ΔN vs the number of prong hits for barrel ECal events in the 2 prong topology.

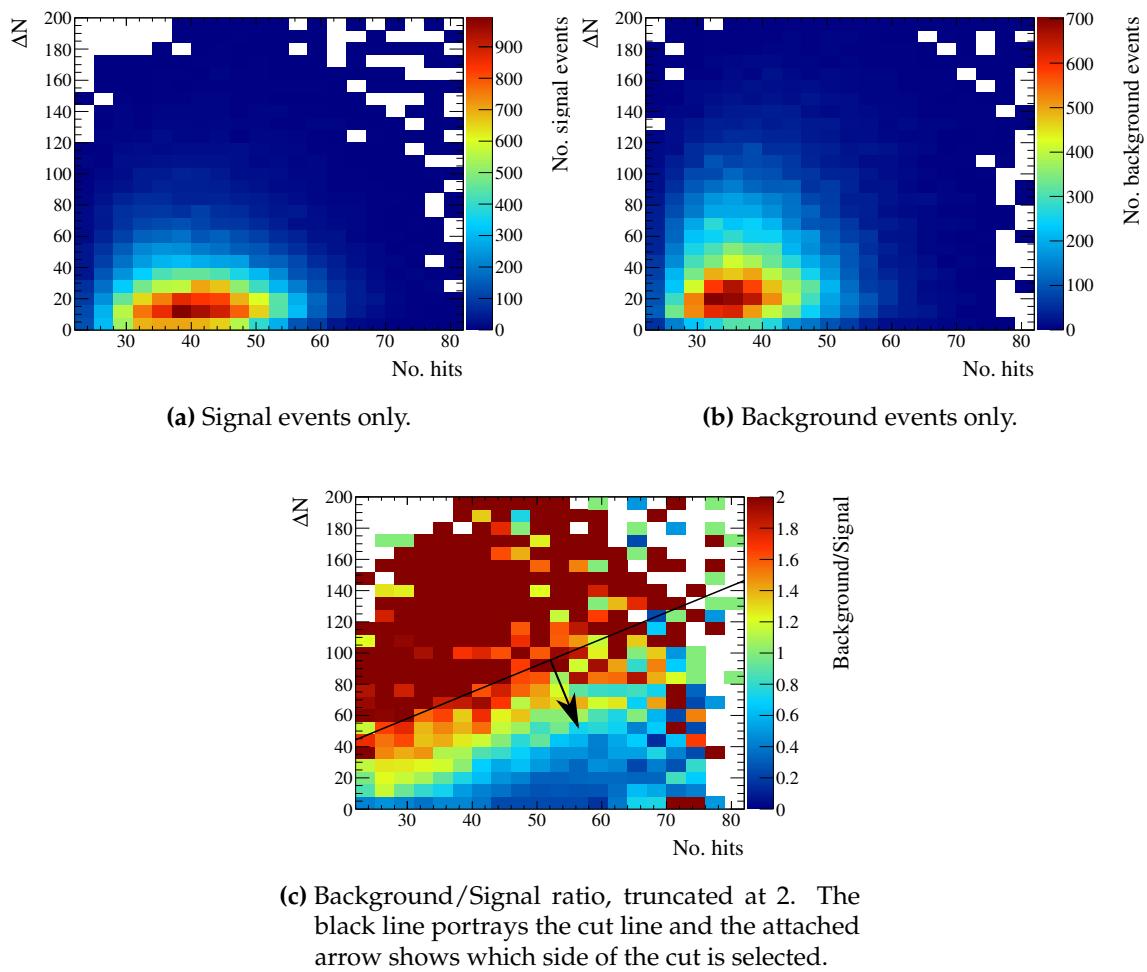


Figure 7.27: ΔN vs the number of prong hits for barrel ECal events in the 3 prong topology.

Prong topology	Barrel cuts (no units)	DS cuts (no units)
1	$< -0.2N + 16$	$< 1.2N + 3$
2	$< 1.5N - 16$	$< 0.7N - 4$
3	$< 1.7N + 7$	$< 0.8N + 1$

Table 7.7: The definitions of the 2 dimensional cut lines involving ΔN and the number of scintillator hits associated with the constituent prongs. N refers to the total number of scintillator hits associated to the constituent prongs in the vertex.

7.6.1.4 Opening angle

Despite the best efforts of the track merging, not all curving trajectories have reconstructed tracks which are successfully merged. So, there will be an inherent ECal OOFV contamination of the 2, 3 and 4+ prong topologies where the reconstruction has failed. While future iterations of the analysis should revisit the reconstruction to combat this, it is sufficient to cut these background events out in the selection. As was described in section 7.4, a key signature of a curving trajectory is two prongs with a small opening angle. This means that the opening angle can be used as a cut to remove such background events. To define an opening angle, at least two tracks are required so, unfortunately, such a method can not be implemented for the 1 prong topology.

The signal and background separation provided by the subtended opening angle for the 2 prong topology is shown in Fig. 7.28. While the barrel ECal sees more separa-

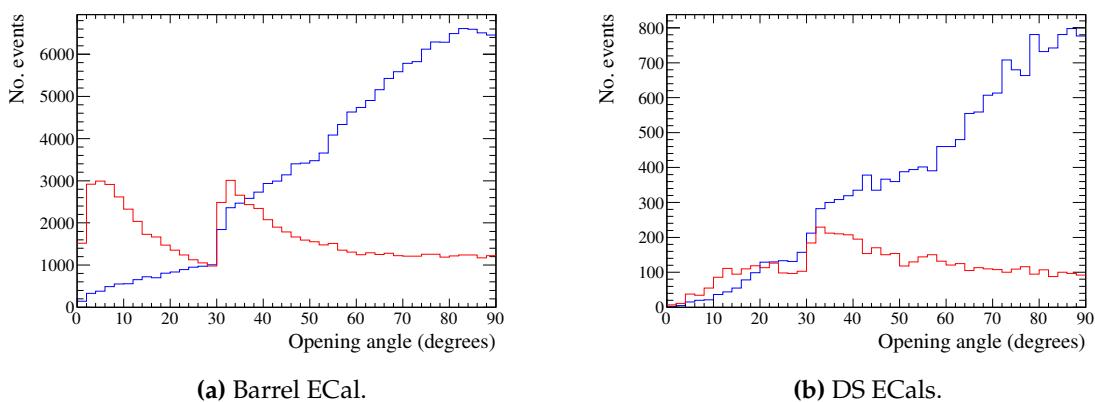


Figure 7.28: The opening angle subtended by the constituents prongs for the 2 prong topology. Signal and background events are the blue and red histograms respectively.

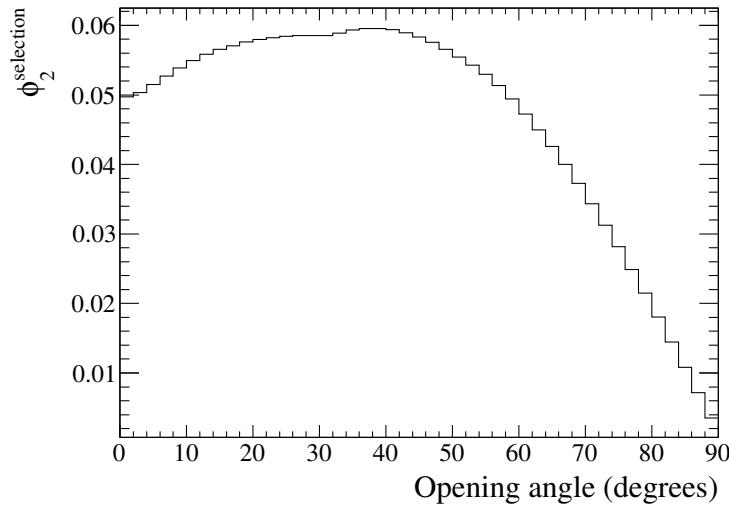


Figure 7.29: $\phi_2^{\text{selection}}$ as a function of the opening angle cut for the 2 prong topology in the barrel ECal. There is a clear maximum of $\phi_2^{\text{selection}}$, indicating a preferred cut on the opening angle.

tion power in the opening angle, it should be clear from Fig. 7.28a and Fig. 7.28b that events should only be selected if their opening angle is bigger than some threshold. To find this threshold, a test cut was systematically varied between 0° and 90° in 1° increments. At each increment, the number of events with an opening angle larger than the test cut was recorded and $\phi_2^{\text{selection}}$ subsequently calculated. An example of the variation in $\phi_2^{\text{selection}}$ is shown in Fig. 7.29 for the barrel ECal. A maximum in $\phi_2^{\text{selection}}$ is clearly found which shows that the opening angle is a good discriminator and that the method works. By applying this method to 2 prong events, events are only selected if the subtended opening angle is greater than 34° for the barrel ECal or greater than 21° for the DS ECal. Similar information should be available for the higher prong topologies. However it is not as trivial to define an opening angle in such situations. The chosen variable for the 3 prong topology is the summed opening angle for every pairwise combination of the constituent prongs. The signal and background separation for this summed opening angle is shown in Fig. 7.30. The cut development process used here was identical to the one used for the 2 prong topology. The study suggested that events should only be selected if their summed opening angle is greater than 81° for the barrel ECal and 49° for the DS ECal. The cuts found by this study are summarised in table 7.8.

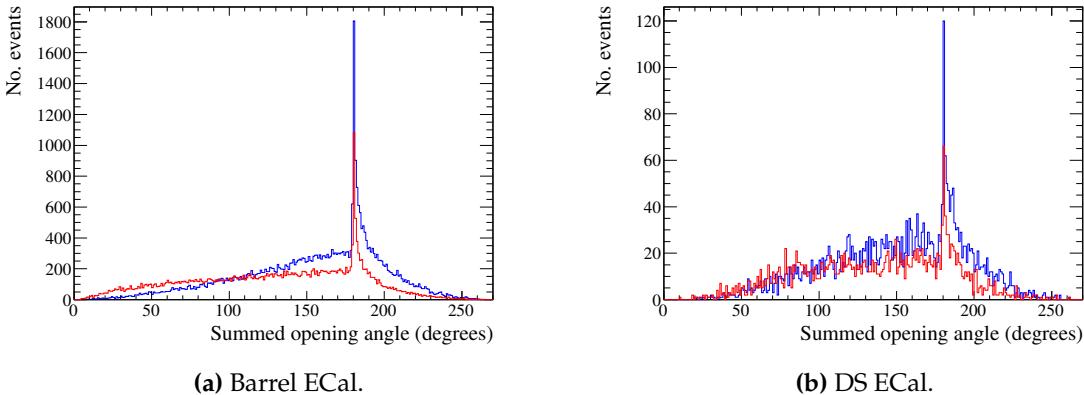


Figure 7.30: The summed opening angle subtended by every pairwise combination of the constituents prongs for the 3 prong topology. Signal and background events are the blue and red histograms respectively.

7.6.1.5 Entering background cut

The fifth cut, simply called the ‘entering background’ cut, uses the prong multiplicity aspect of a vertex to reject events that enter the ECal and then mimic a signal vertex. The cut requires at least two constituent prongs to work, so the 1 prong topology is unaffected by this. The idea is fairly simple: for each vertex, check if any of the prongs have an end which is upstream of the reconstructed vertex position. If so, it may be an entering background and could need rejecting. The position of the upstream prong end is key, as the cut should not reject every event which has an apparent backwards going track. So, the cut checks whether the upstream prong end is outside of the host ECal fiducial volume. As described in section 7.6.1.1, each ECal module uses two fiducial volumes. The 2+ prong fiducial volume would be of little use in this situation as the only strict face definition is for the downstream end of each ECal module. This motivates the use of the 1 prong fiducial volume for this cut. This cut is physically motivated and relies on parameters which have already been tuned. So, this cut requires no tuning to operate.

Prong topology	Barrel cuts (degrees)	DS cuts (degrees)
2	> 34	> 21
3	> 81	> 49

Table 7.8: The definitions of the opening angle cuts.

7.6.1.6 Most upstream vertex

The sixth and final cut implemented is motivated by the beam kinematics. As discussed in section 7.2, the expected pileup per beam bunch is small. As the reconstruction outputs events in buckets, the same philosophy can be used here; there should only be one signal event per bucket. The assumption used for this cut is any secondary interactions are likely to happen downstream of the interaction. So, the cut selects the most upstream remaining vertex in the beam bucket. As with the ‘entering background’ cut, the selection of the most upstream vertex requires no tuning to operate.

7.6.2 Performance of the selection

Now that the cuts have been identified, their performance can be assessed. To do this, the Monte Carlo sample was processed through the selection and the event composition assessed at each stage. As each cut was focused on a specific topology, it is important to check the effect of the cuts on each topology. The first check looks at the event survival as a function of the cuts for each topology. These checks are shown in Fig. 7.31 and Fig. 7.32 for the barrel and DS ECals respectively. Each bin in both figures shows how many events survive after the cut has been applied. For example, the FV bin in all of the figures shows how many events remain after applying the fiducial volume cut. As the ‘most upstream’ cut is the final cut in the selection, the final bin shows how many events remain after the full selection has been applied. The main piece of information shown is that the backgrounds are being rejected at a much higher rate than the signal, resulting in a selected sample that is mostly signal.

While Fig. 7.31 and Fig. 7.32 shows that the cuts are successfully rejecting background while retaining signal, it is not clear what particle species are rejected. It is important to check this so that the developed selection is understood. The background type which is rejected at the highest rate is the ECal OOFV background so it is sufficient to study the particle composition of this to understand the selection’s ability to reject particles. The particle type is defined by looking at which simulated particles contributed to the ECal cluster as whole. The particle which deposited the highest amount of energy is accepted as the one which created the ECal cluster. The ECal OOFV event survival as a function of the cuts is shown in Fig. 7.33 and Fig. 7.34 for the barrel and DS ECal respectively. For each prong topology, the ECal OOFV events are separated into the

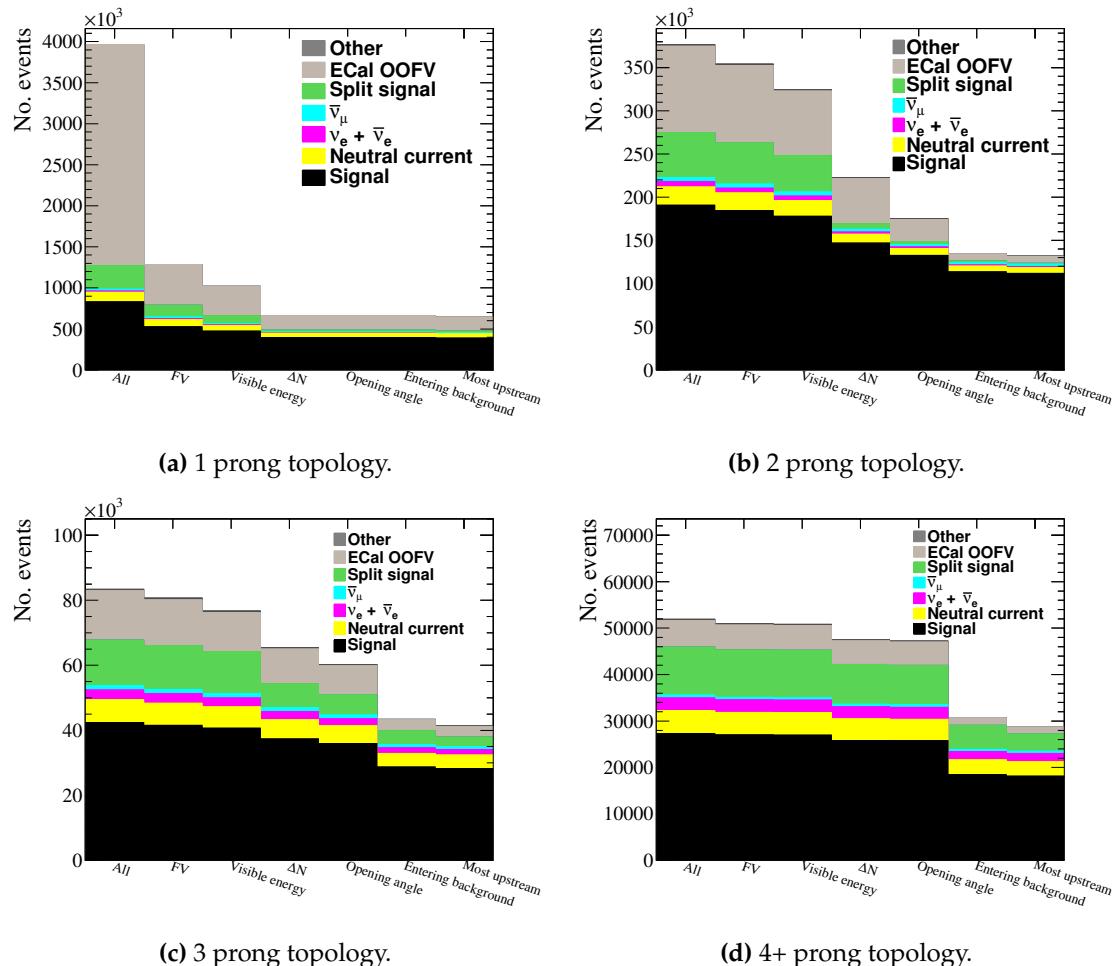


Figure 7.31: Event survival as a function of the cuts for all topologies in the barrel ECal. Each bin refers to the number of events after a cut has been applied. The rejection of background events by each cut can be clearly seen.

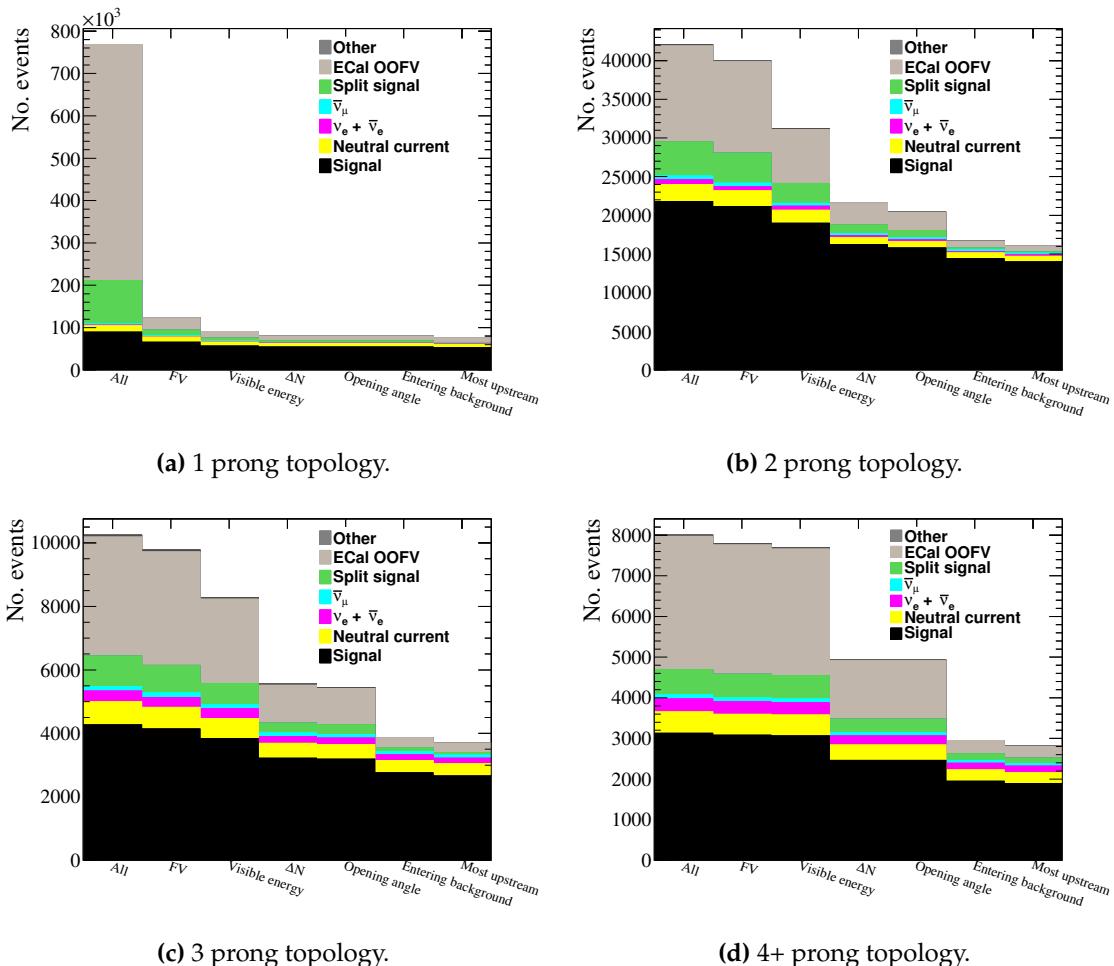


Figure 7.32: Event survival as a function of the cuts for all topologies in the DS ECal. Each bin refers to the number of events after a cut has been applied. The rejection of background events by each cut can be clearly seen.

main particle types which constitute the ECal OOFV background. It is clear that the cuts are rejecting the particle species that one would expect them to reject. For example, the fiducial volume cut in the 1 prong topology almost exclusively rejects entering MIPs. The visible energy and ΔN cuts worth together as an effective particle shower tagger to reject e^\pm and γ particles. The opening angle and entering background cuts work as an additional MIP rejector in cases where the track merging was unsuccessful. The composition of the ECal OOFV background has interesting features, regardless of the cuts applied. Specifically, the particle species varies to a great degree between each prong topology. For example, the 1 prong and 2 prong topologies in the barrel ECal (Fig. 7.33a and Fig. 7.33b) see a muon dominated ECal OOFV background. Whereas, the 3 prong and 4+ prong topologies see a very different background which is π^\pm dominated. The natural separation provided by the topologies should allow investigation of a specific particle species if required by any future analyser.

The selection efficiency and purity as a function of the cuts are shown for the barrel ECal and DS ECal in Fig. 7.35 and Fig. 7.36 respectively. Every cut which is relevant to a prong topology has an expected effect, the purity of the sample increases while the efficiency decreases. What is not so clear from the distributions is whether the tuning metric, $\phi^{\text{selection}}$, increases as a function of the cuts. So, $\phi^{\text{selection}}$ is shown in a similar manner for the barrel and DS ECal in Fig. 7.37 and Fig. 7.38 respectively. The tuning metric optimisation in both the barrel and DS ECal shows some interesting features. Firstly, in the barrel, it is clear that for the 1 prong and 2 prong topologies, every applied cut shows a clearly increasing $\phi^{\text{selection}}$. Unfortunately, this is not the case for the 3 prong and 4+ prong topologies. Specifically, the entering background cut degrades the value of $\phi^{\text{selection}}$. This is most obvious in the 4+ prong topology; however, it has already been stated that this topology is effectively an overflow bin and is being included more for completeness. What is perhaps more concerning is the degradation of $\phi^{\text{selection}}$ in the 3 prong topology. Fortunately, the decrease in $\phi^{\text{selection}}$ is small when the entering background cut is applied and the change is well within error. It is important to bear in mind that $\phi^{\text{selection}}$ is used as a tuning guide and small deviations from the optimisation process are allowed. So, it was decided that the entering background cut would remain in place for both the 3 prong and 4+ prong topologies. The $\phi^{\text{selection}}$ distribution in the DS ECal shows some similar and some different features to that of the barrel. As with the barrel ECal case, $\phi^{\text{selection}}$ clearly increases as a function of the cuts for the 1 prong and 2 prong topologies. Unlike the barrel ECal, there is also an increase in $\phi^{\text{selection}}$ when applying the entering back-

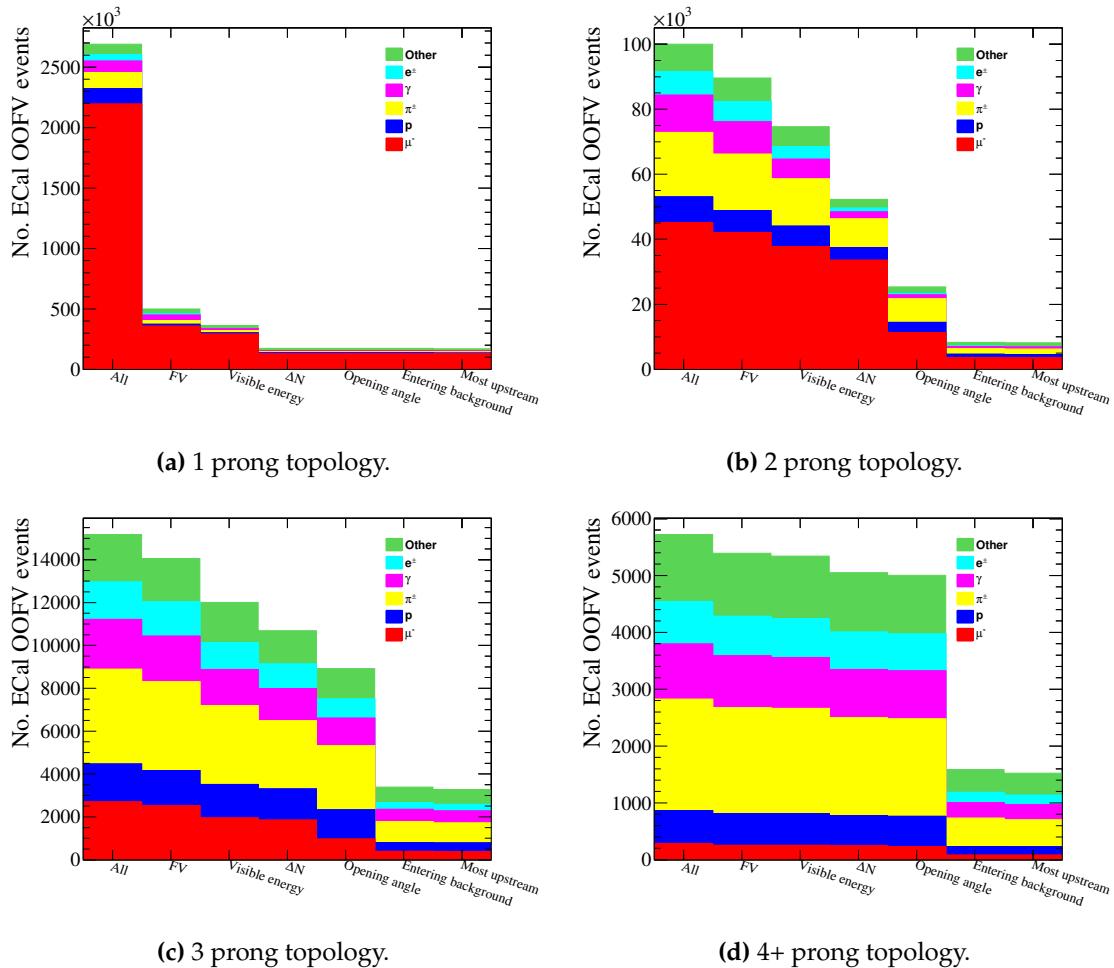


Figure 7.33: ECal OOFV event survival as a function of the cuts for all topologies in the barrel ECal. Each bin refers to the number of events after a cut has been applied. It is clear that each cut rejects a specific set of particle types. For example, the FV cuts strongly reject the track-like particles whereas the visible energy and ΔN cuts reject shower-like particles.

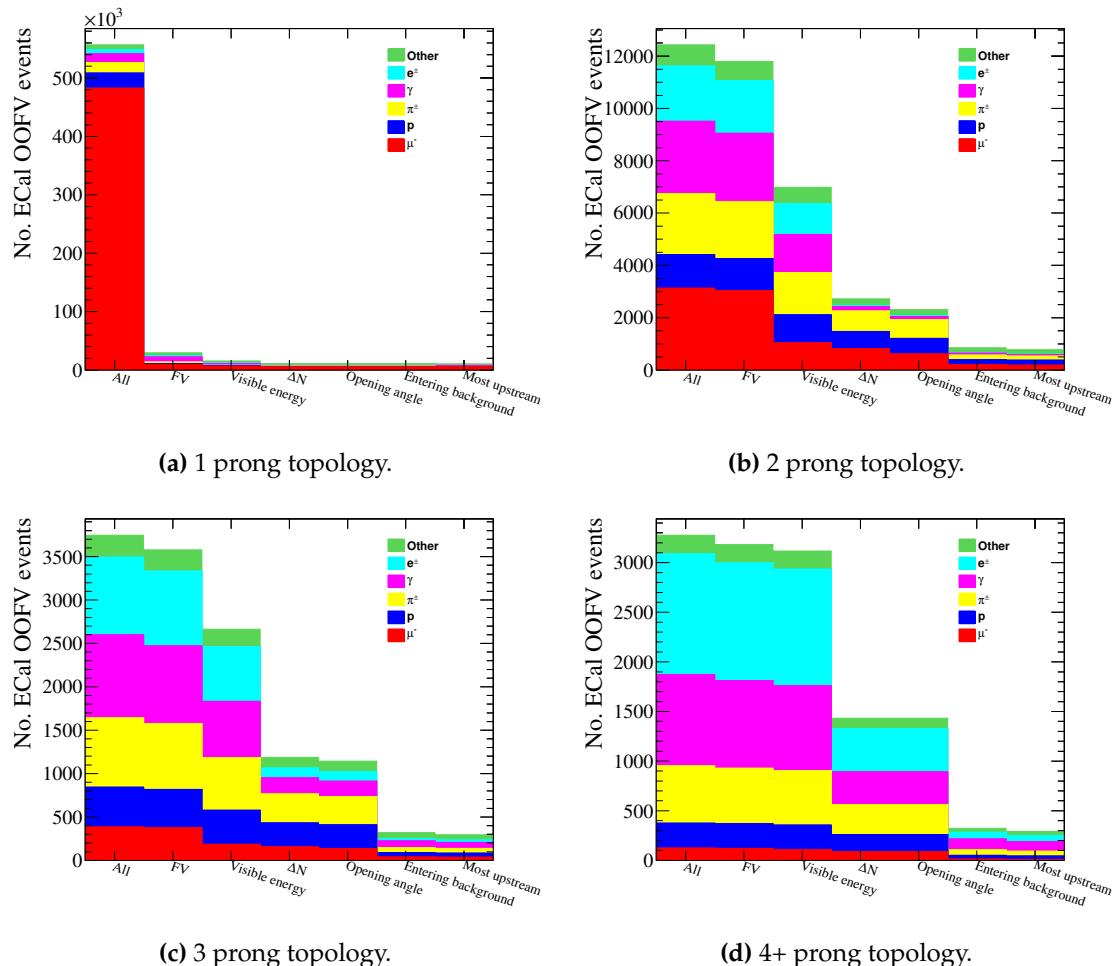


Figure 7.34: ECal OOFV event survival as a function of the cuts for all topologies in the DS ECal. Each bin refers to the number of events after a cut has been applied. The cuts shown behave in a similar manner to those in the barrel ECal (see Fig. 7.33).

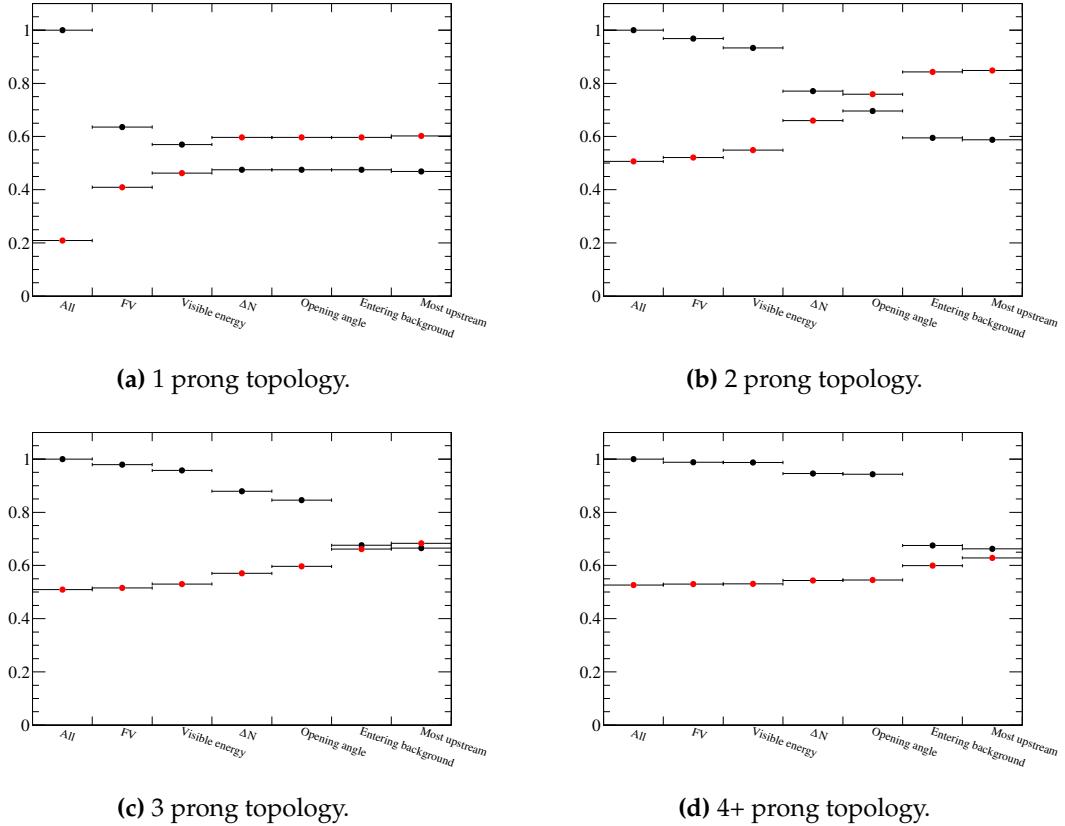


Figure 7.35: The selection efficiency (black) and purity (red) as a function of the cuts in the barrel ECal. Each bin refers to the efficiency and purity after a cut has been applied. The cuts generally degrade the selection efficiency and increase the selection purity. It is clear that strength of all cuts are not equal.

ground cut to the 3 and 4+ prong topologies. However, there is a minor decrease in $\phi_{\text{selection}}$ when selecting the most upstream vertex. It is not wholly surprising that this is the case as the DS ECal is the most downstream detector in ND280. Such a cut can easily cause a bias where events in the barrel are generally preferred because of their extent in the global Z direction. Fortunately, this effect is very minor and also well within error.

The cross-section measurement presented in this analysis is an inclusive measurement. This means that the neutrino interaction cross-section needs to be measured without bias across the J-PARC beam energy range. All of the performance checks presented so far have only assessed each prong topology individually. So, the final check is to assess how the selection performs after combining the topologies together. The selec-

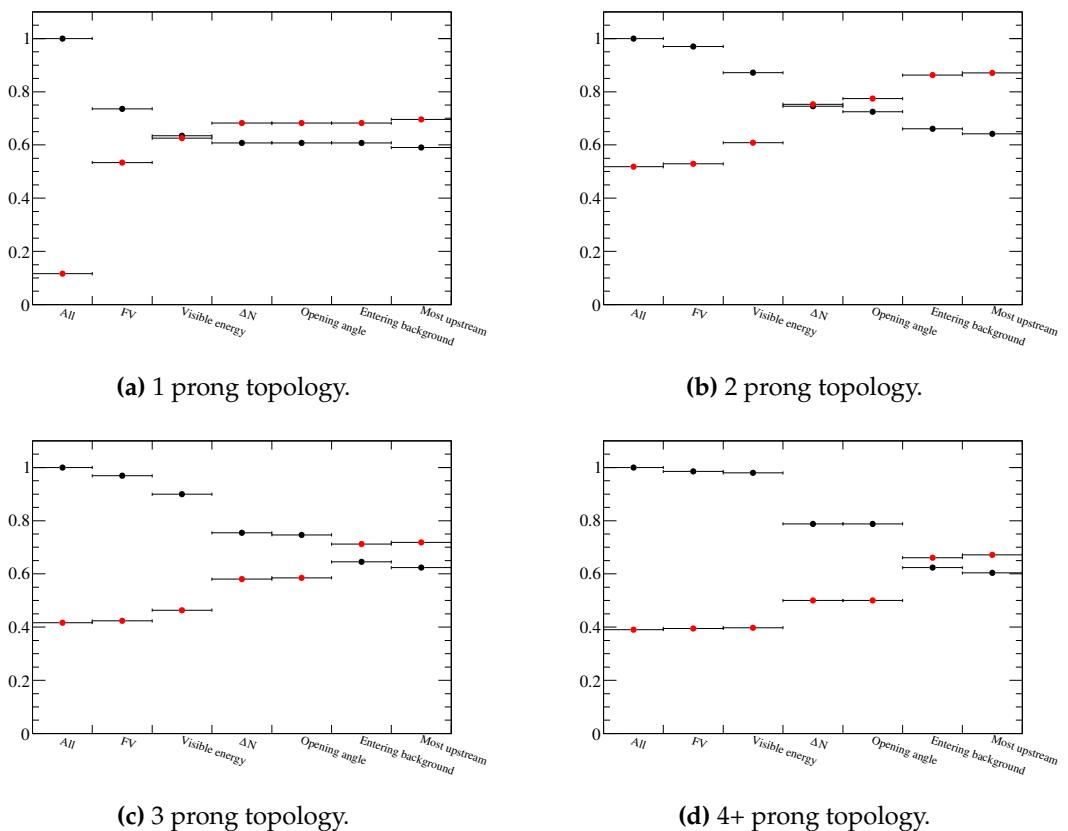


Figure 7.36: The selection efficiency (black) and purity (red) as a function of the cuts in the DS ECal. Each bin refers to the efficiency and purity after a cut has been applied. The distributions suggest the cuts behave in a similar manner to those in the barrel ECal (see Fig. 7.35).

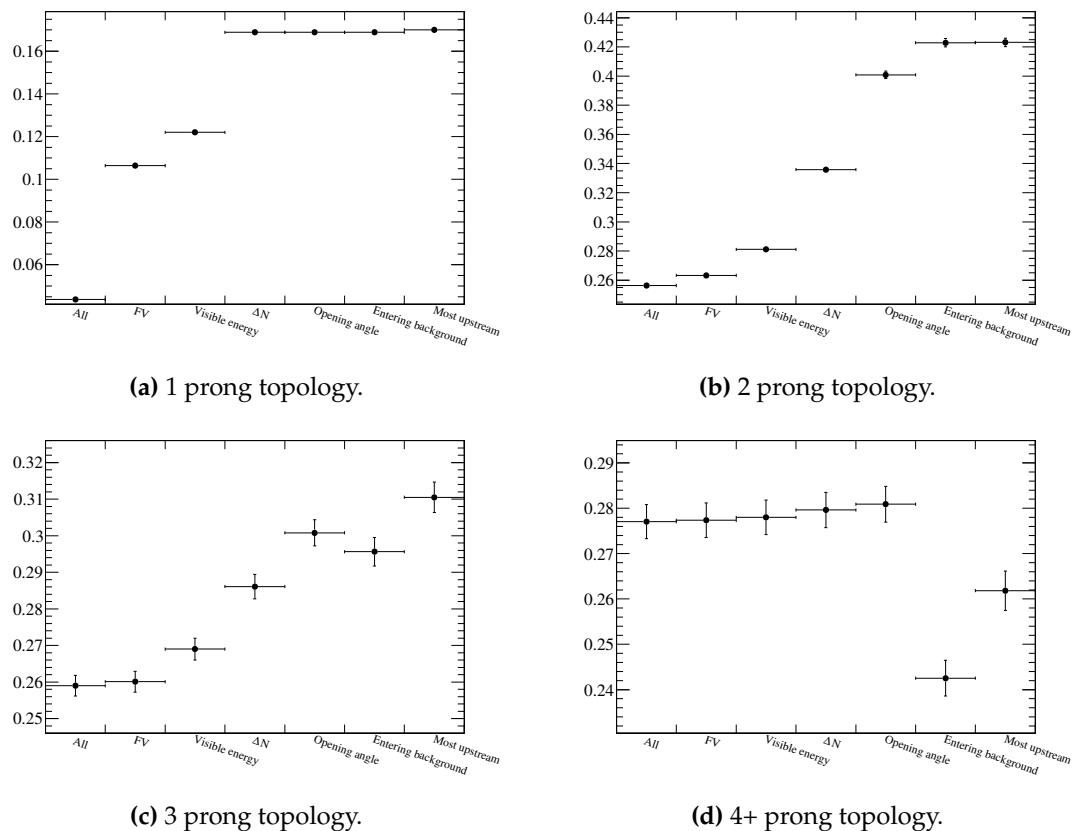


Figure 7.37: $\phi^{\text{selection}}$ as a function of the cuts in the barrel ECal. Each bin refers to the value of $\phi^{\text{selection}}$ after a cut has been applied. Generally, each cut increases $\phi^{\text{selection}}$ with the exception of the 4+ prong topology. However, it should be noted that the selection was not tuned for this topology.

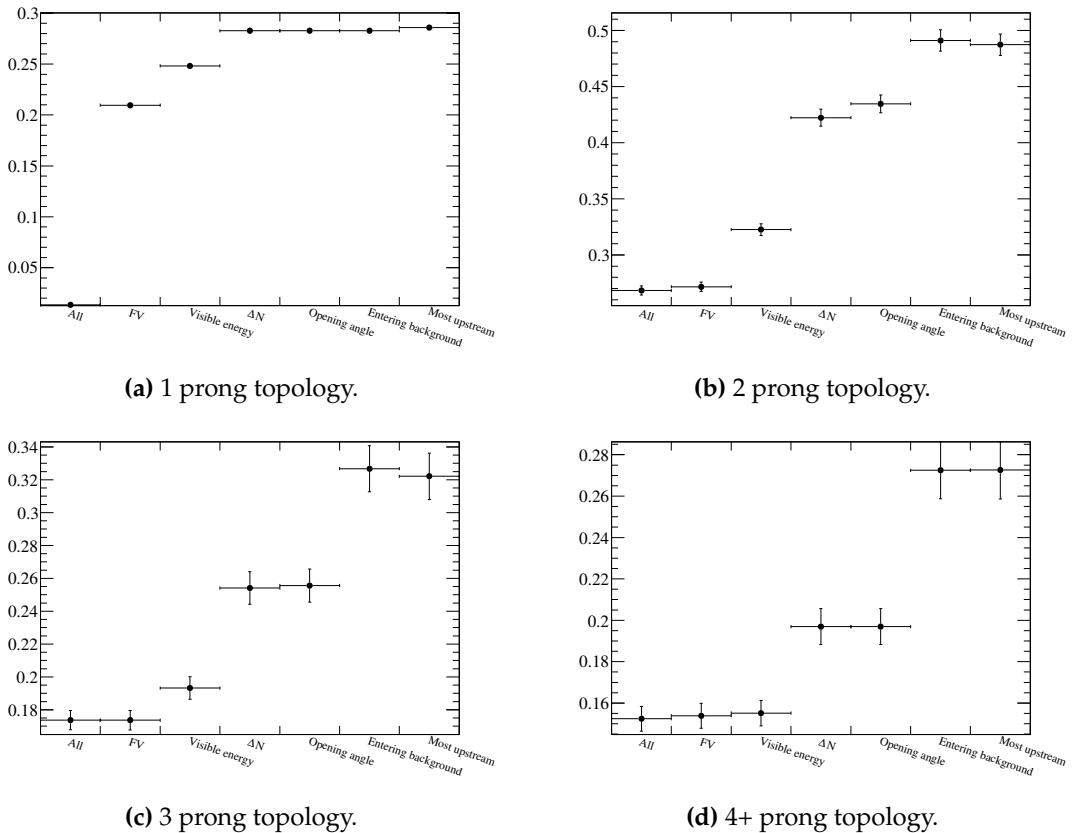


Figure 7.38: $\phi^{\text{selection}}$ as a function of the cuts in the DS ECal. Each bin refers to the value of $\phi^{\text{selection}}$ after a cut has been applied. Unlike the barrel ECal (see Fig. 7.37), the cuts do have a positive effect on the 4+ prong topology.

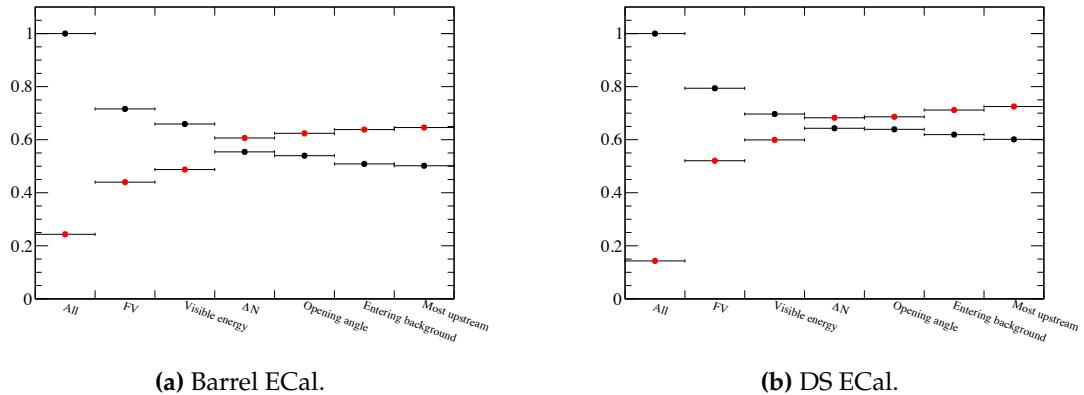


Figure 7.39: The selection efficiency (black) and purity (red) for the combined prong topologies as a function of the selection cuts. After combination of all topologies, each cut has an expected effect; an increase in sample purity at the expense of selection efficiency.

tion efficiency and purity as a function of the selection cuts for all prong topologies combined is shown in Fig. 7.39. The effect of the cuts is largely the same when treating all of the combined topologies together; the sample purity increases at the expense of selection efficiency. Fig. 7.40 shows $\phi^{\text{selection}}$ as a function of the selection cuts for the combined prong topologies. $\phi^{\text{selection}}$ increases as each selection cut is applied which suggests that the cuts still work adequately when considering all prong topologies simultaneously. The selection efficiency and purity after combining the topologies is shown in Fig. 7.41. For both the barrel ECal and DS ECal, there is a clear dependence of the purity on the neutrino energy. However, this should be expected. It is clear that the purity decreases as the neutrino energy increases which can be explained by the self-shielding effect of the ECal. In the low energy regime, the ECal OOFV background is suppressed because there is insufficient energy to either penetrate the ECal or deposit enough energy to allow cluster reconstruction. As the background energy increases, the suppression is lifted, resulting in a decrease in the selection purity. The key piece of information shown in Fig. 7.41 is the selection efficiency. As described above, it is of vital importance that the selection does not introduce a bias which measures one energy regime more than another. The flatness of the efficiency distributions for both the barrel ECal and DS ECal shows this is not the case which means the selection presented here would be suitable for a CC-inclusive measurement.

The number of selected events for each prong topology is shown in Fig. 7.42, with each

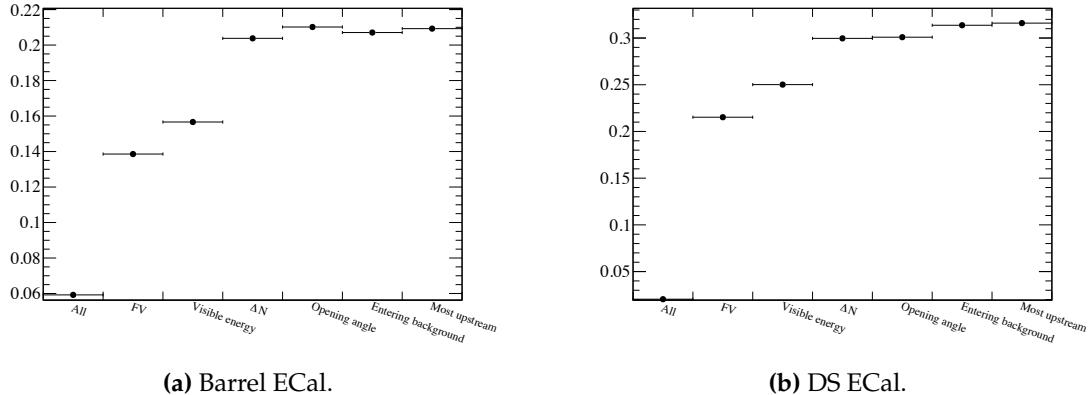


Figure 7.40: $\phi_{selection}$ for the combined prong topologies as a function of the selection cuts. Generally speaking, $\phi_{selection}$ increases as each cut is applied to either the barrel ECal or DS ECal samples. The only exception to this is when the entering background cut is applied in the barrel ECal which slightly decreases $\phi_{selection}$.

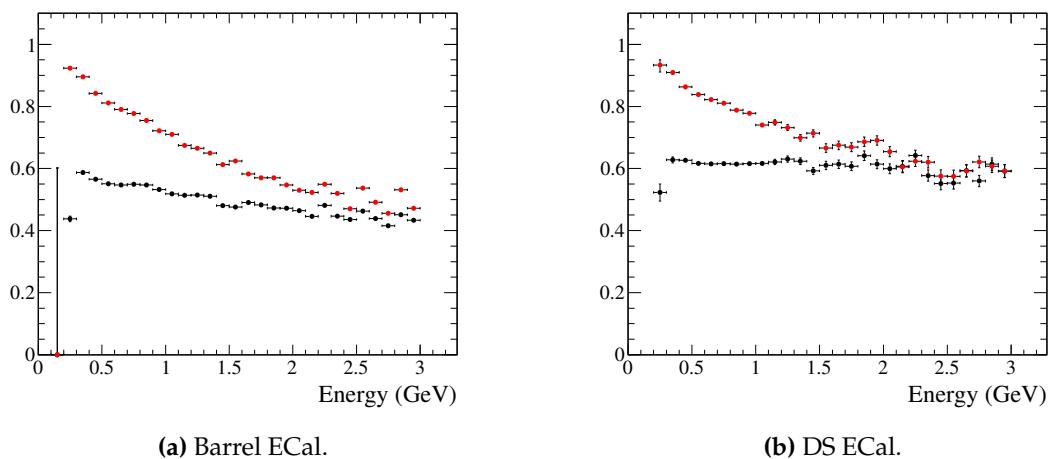


Figure 7.41: The selection efficiency (black) and purity (red) for the combined prong topologies as a function of neutrino energy. The selection efficiency is largely flat for both the barrel ECal and DS ECal showing the selections are not biased towards any particular energy range. This is not the case for the selection purities but this can be explained by lower energy backgrounds having less penetrating power, resulting in a higher purity for lower neutrino energies.

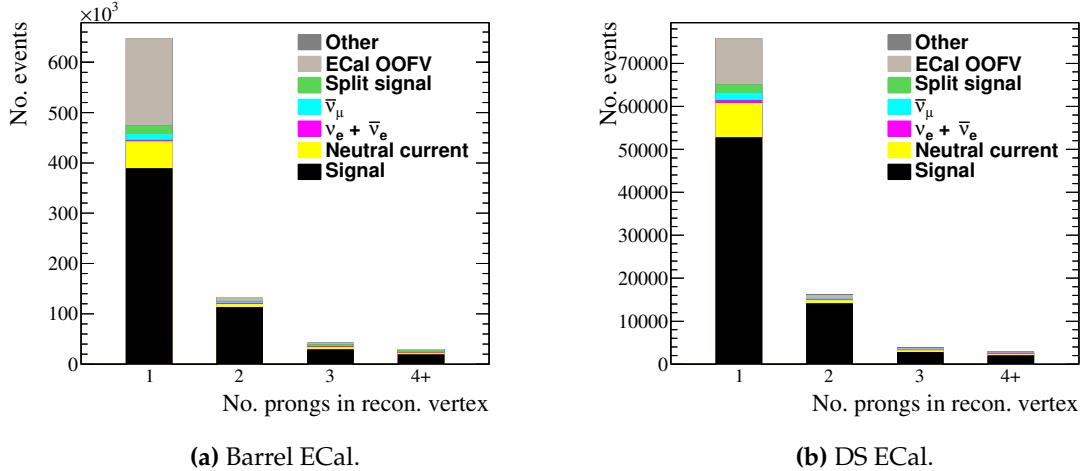


Figure 7.42: The number of selected events in the Monte Carlo sample, separated out into the prong topologies. Each event is categorised by the associated truth information from the simulation. The effect of the developed selection can be clearly seen in all prong topologies; each topology is now signal dominated.

topology broken down by truth categories. For both the barrel and DS ECal, signal events dominate each prong topology. The most impure topology for both detectors is the 1 prong topology. This is primarily due to having an insufficient number of prongs to really benefit from what the reconstruction can provide. Despite this, it is clear that the selection results in a pure sample of events. The variation in sample purities for each prong topology also highlights the benefit of prong topology separation in the MC selection. The 2 prong topology in both detectors sees a very high sample purity. This is because multi-track reconstruction information is available but the number of reconstructed tracks is sufficiently small such that the coarse detector granularity does not significantly mask the track information. As you move to the higher prong topologies, it becomes harder to discern all of the tracks clearly, resulting in less information available to precisely reject background.

The selection efficiencies and purities for each prong topology are shown in table 7.9 and table 7.10 respectively. The final purities and efficiencies are generally good.

The selection efficiency is defined to be 100% when no cuts have been made. There are inevitably signal events which are not reconstructed, primarily because the energy is below reconstruction threshold. It is non-trivial to include these events in an efficiency calculations for a specific prong topology, but it is also unnecessary to do this as the

ECal module	1 prong topology efficiency (%)	2 prong topology efficiency (%)	3 prong topology efficiency (%)	4+ prong topology efficiency (%)
Barrel	46.8	58.7	66.5	66.3
DS	59.0	64.2	62.3	60.4

Table 7.9: The selection efficiencies for each prong topology and ECal module.

prong topologies are to be summed for the CC-inclusive cross-section measurement. So, the topology combined efficiency and purity are presented in table 7.11. It is these numbers, along with the sample itself, which are the final output of the Monte Carlo selection.

ECal module	1 prong topology purity (%)	2 prong topology purity (%)	3 prong topology purity (%)	4+ prong topology purity (%)
Barrel	60.3	85.0	68.5	63.0
DS	69.6	87.2	72.0	67.3

Table 7.10: The selection purities for each prong topology and ECal module.

Ecal module	Efficiency (%)	Purity (%)
Barrel	42.4	64.4
DS	53.0	72.4

Table 7.11: The topology combined efficiency and purity.

Chapter 8

Measurement of the ν_μ charged current inclusive cross-section on Pb

This chapter presents the measurement of the ν_μ CC inclusive cross-section on Pb using the ND280 ECals. This chapter details the measurement method, the samples used in the measurement, the assessment of the systematic uncertainties, validation of the measurement method and finally the measurement itself.

8.1 Measurement method

The chosen method fits a prediction to measured data using multiple data samples [82–84]. Here, a “sample” refers to, for example, selected events in a particular ECal module. The core of the analysis method is a χ^2 fit which tries to minimise the difference between the prediction and the data. The χ^2 is defined as

$$\chi^2 = \Delta \vec{N}^T (\underline{\underline{V}}^{\text{syst}} + \underline{\underline{V}}^{\text{stat}})^{-1} \Delta \vec{N}, \quad (8.1)$$

where $\underline{\underline{V}}^{\text{syst}}$ and $\underline{\underline{V}}^{\text{stat}}$ are the systematic and statistical covariance matrices for the sample and $\Delta \vec{N}$ contains the difference between the data and the prediction for each sample. If the number of samples used is M , $\Delta \vec{N}$ is defined as

$$\Delta \vec{N} = \begin{pmatrix} N_1^{\text{data}} - N_1^{\text{pred}} \\ \vdots \\ N_M^{\text{data}} - N_M^{\text{pred}} \end{pmatrix}. \quad (8.2)$$

For sample j of the sample set M , N_j^{data} and N_j^{pred} are the number of measured data events and number of predicted events. The extractable information from the fit is located in N_j^{pred} which can be subdivided into a set of templates, each of which are assigned a normalisation parameter, namely

$$N_j^{\text{pred}} = \sum_i R_i n_{ij}, \quad (8.3)$$

where n_{ij} is the number of events in template i of sample j and R_i is the normalisation assigned to that template. The normalisation parameters are free to float in the fit and so it is in the normalisation parameters that the desired physical information is located.

8.2 Input samples to the fit

The DS ECal provides the main signal sample which will be used to extract the cross-section. As shown in table 7.11, the selected sample in the DS ECal has the highest purity and efficiency in the whole selection making the sample the natural choice to extract the cross-section.

The method outlined above allows for the simultaneous constraint of physical parameters and the background contamination of a set of input samples. The method works particularly well when the input samples do not share the same sensitivity to a particular background type. For example, in ND280, each ECal module is exposed to a different beam intensity and energy spectrum and so is exposed to a unique amount of beam-induced background. Each ECal module can be thought of as a separate input sample which fits well with the method outlined above. So, while the DS ECal will provide the target in which the cross-section will be extracted, the barrel ECals are to also be included in the fit as a background constraint. To further allow the barrel ECals to achieve this, an additional sample set of barrel ECal events are to be included. This additional sample set, called the ‘reverse’ sample set, comes from the same data set as the selected sample described in chapter 7. However, the events in the reverse sample set are events which pass the fiducial volume cuts but fail any other cut. An example of a selected sample compared with a reverse sample is shown in Fig. 8.1. There are clear differences in the shape of the energy spectrum between Fig. 8.1a and Fig. 8.1b, which suggests that the selection is biased towards lower-energy neutrino interactions,

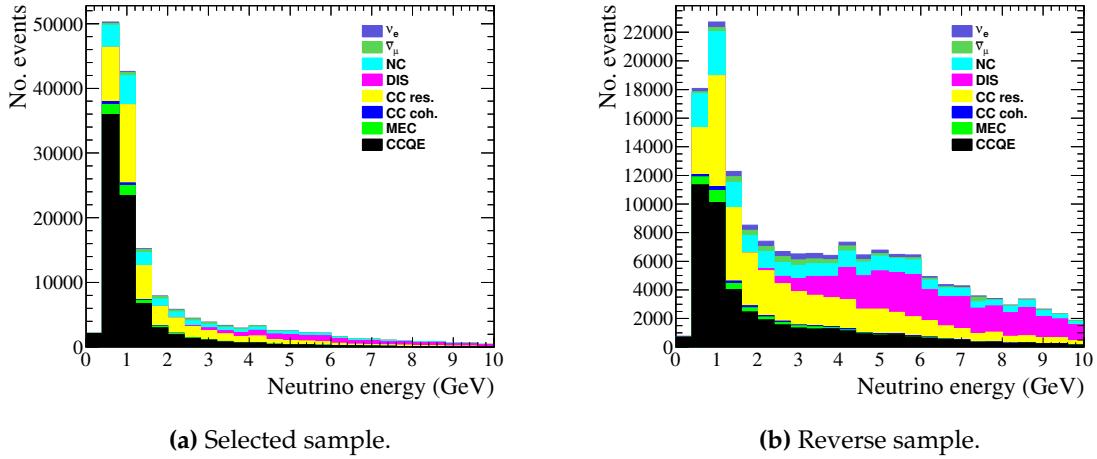


Figure 8.1: The neutrino energy of events seen in the bottom-left ECal module broken down by the neutrino interaction mode. The plots correspond to 3.949×10^{20} POT of simulated data. The reverse sample contains a large population of high energy (5 GeV) events which is not seen in the selected sample. This can be explained by the higher energy events creating a surplus of reconstructed clusters which are cut away in the selection, causing such clusters to fall into the reverse sample.

despite the flat efficiency curves shown in Fig. 7.41a. The apparent bias is actually an effect of higher-energy interactions creating a surplus of reconstructed objects, most of which are cut away by the selection. The shape of the true neutrino energy spectrum (see Fig. 7.2 for an example) is actually more like the selected sample shown in Fig. 8.1a. Because the neutrino energy is not a quantity that is currently reconstructed in the ECal, it can be difficult to understand how the information shown in Fig. 8.1 translates into detector variables. To get a better idea of this, the same events are shown as a function of the number of the reconstructed prongs per vertex in Fig. 8.2. Fig. 8.2b shows that the extra high energy events (the 5 GeV peak in Fig. 8.1b) are distributed evenly between each prong topology.

As described in chapter 7, the developed selection does not distinguish between lead and carbon interactions, partly because such events are indistinguishable from one another. It is, however, desirable to separate the two categories out, as the presented measurement purely involves lead. As the ECal is not capable of simultaneously constraining the lead and carbon events by itself, an additional constraint is needed. Therefore, a sample of CC interactions on carbon occurring in the FGD are also included, which are taken from the official ND280 oscillation input analysis [85].

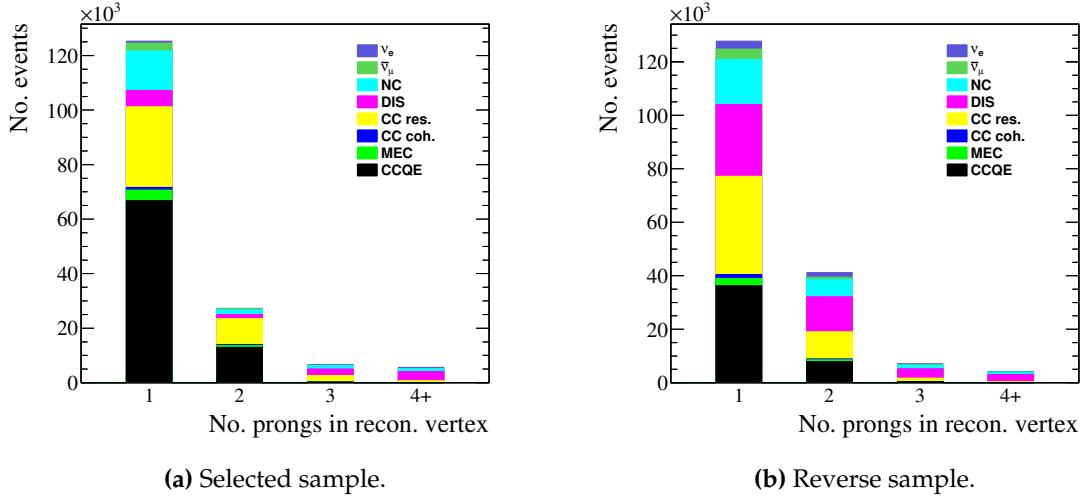


Figure 8.2: The number of reconstructed prongs per vertex for events seen in the bottom-left ECal module broken down by the neutrino interaction mode. The plots correspond to 3.949×10^{20} POT of simulated data. The higher energy events seen in Fig. 8.1b (which are mostly DIS events) are spread evenly between each prong topology in the reverse sample.

To summarise, there are many input samples to the measurement. Specifically, there are 12 barrel ECal samples (6 selected and 6 reverse), the DS ECal sample and the FGD sample which totals to 14 input samples. To get an idea on the level of statistics, the DS ECal sample, which will provide the primary interaction target events, is expected to contain 35000 events

Monte Carlo-simulated data of 3.949×10^{20} POT worth of NEUT events, which corresponds to 3 times the real data used are used for this measurement.

8.3 The ECal rate fit

Now that the general method and the input samples have been introduced, the fit used by the analysis can be introduced. The fit itself is a data-driven constraint of the event rate in each ECal module where the Monte Carlo prediction is separated into templates whose normalisation is allowed to vary. The machinery outlined in section 8.1 almost completely describes what is required to understand the fit; however, two specific definitions are needed. The first is that there are 14 input samples to the fit. The second is the definition of N_i^{pred} . To constrain the lead event rate, three Monte Carlo templates

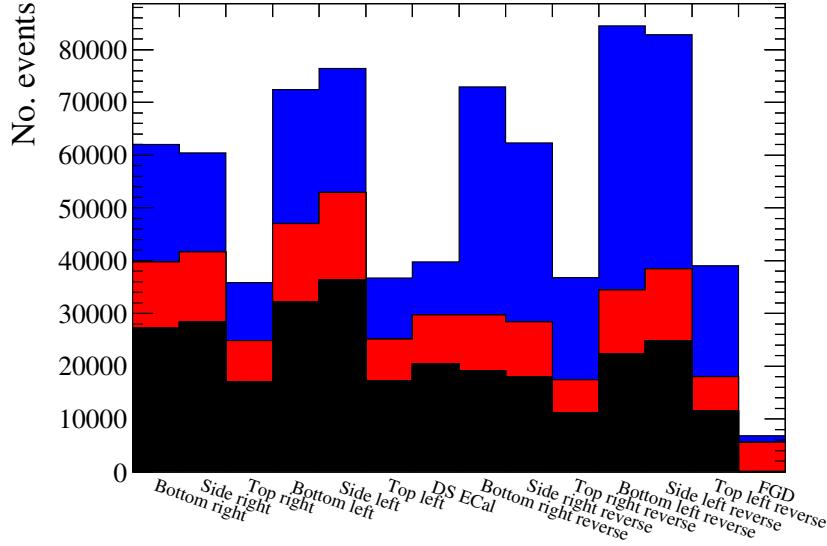


Figure 8.3: The number of events in each input sample separated into the lead (black), carbon (red) and other (blue) templates. “ECal” has been omitted in all but the DS ECal and FGD sample names. Not only does the event rate vary significantly between each sample, but the composition of those events also significantly varies.

are needed: a lead template which contains any reconstructed event associated with a ν_μ CC interaction on lead; a carbon template which contains any reconstructed event associated with a ν_μ CC interaction on carbon; and an ‘other’ template which contains any reconstructed event which does not fall into the above two categories. The number of predicted events in each sample i , N_i^{pred} is then defined as

$$N_i^{\text{pred}} = R_i^{\text{Pb}} n_i^{\text{Pb}} + R_i^{\text{C}} n_i^{\text{C}} + R_i^{\text{other}} n_i^{\text{other}}, \quad (8.4)$$

where n_i^X is the number of events in template X of sample i and R_i^X is the variable normalisation of that template.

The rate fit allows R_i^{Pb} , R_i^{C} and R_i^{other} to vary with no prior constraint in an attempt to minimise the χ^2 as defined in equation 8.1. Once the minimum has been found, the fitted R_i^{Pb} can then be used to extract the cross-section using the selected events in the DS ECal.

The number of events in each input sample, broken down by the template contributions, is shown in Fig. 8.3. As should be expected, the relative background contribution varies between each ECal module.

The underlying fit machinery uses the Minuit2 algorithm provided by the ROOT TFitter framework, chosen primarily because of its ease of use.

8.4 Systematic uncertainties

The χ^2 definition shown in equation 8.1 allows for systematic uncertainties to be directly included in the fit. As the systematic uncertainty implementation takes the form of a covariance matrix for the input samples, a good understanding of not only the systematic uncertainties, but how the samples correlate with one another, is needed.

As stated above, there are 14 input samples so the covariance matrix will be a 14×14 symmetric matrix. While it may be obvious that the contents of the matrix will be the covariances of each input sample, it may not be immediately clear what type of covariances are needed. As described in section 8.3, the algorithm attempts to fit a set of predictions, separated into templates, to a set of measured data samples simultaneously. The variation comes from the normalisations assigned to each template. This means that the overall normalisation of the Monte Carlo prediction is free to vary without prior constraint and so it is the shape difference between the prediction and data that constrains the parameters. Therefore the covariance matrix used in the fit should only contain uncertainties which account for the variations in the shape of the samples which the systematic sources introduce.

The ND280 event simulation can be broken down into three areas: Simulation of the neutrino flux; simulation of the neutrino interactions; and simulation of the detector response. No matter how sophisticated a simulation is, it is unlikely to ever simulate reality with perfect accuracy. The differences between what is simulated (be it flux, interaction or detector) and what happens in nature causes a systematic difference to be seen between collected data and Monte Carlo which must be accounted for. As the simulation can be broken down into three key areas, the systematic assessment can be broken down into the same three key areas, each of which are presented below.

The actual assessment of a systematic uncertainty can be broken down into two areas: identification of a systematic error source and the propagation of that systematic error source through the analysis. The identification stage is typically handled by an

official T2K working group or by an analyser making use of data taken outside of the analysis signal region (e.g. control samples). The propagation of the systematic error source has no strict recipe but it typically involves variation of the identified systematic uncertainty and then either a reweighting of events is applied or the analysis chain is re-run. The systematic uncertainty treatment used in this analysis uses a combination of these methods.

To propagate effects of the systematic uncertainties, a Monte Carlo sample akin to the prediction sample set used in the fit is required. Therefore, a 2.5×10^{19} POT sub sample of beam and sand Monte Carlo (see section 4.1.1 for definition) is used for this purpose. All sample systematic covariance matrices are presented as fractional covariance matrices between each of the different detector samples. For the barrel samples, the selected samples are labelled by their respective module name, and the reverse barrel samples (background-enriched samples) are labelled with their module name and ‘reverse’. The assessment of each systematic source is presented individually, resulting in a fractional covariance matrix for that source. The final systematic covariance matrix is then found by summing all of the individual covariance matrices.

8.4.1 Flux systematic evaluation

The neutrino flux is one of the core components of the ND280 event simulation and any uncertainties associated with the flux have a direct impact on the measured ND280 event rates. The associated flux uncertainties can be broken down into four categories [55]:

- Properties of the proton beam such as profile and alignment
- Alignment of the target and focusing horns
- The current in the focusing horns and the magnetic field it generates
- Hadron production induced by beam interactions with the target

As the neutrino flux prediction is so important for all T2K analyses, the flux uncertainties are constrained using a range of measurements. These include external measurements from the NA61/SHINE collaboration [68], from the proton beam monitors, measurements of a spare magnetic horn and from the INGRID detector. Measured uncertainties from each source are used to vary the flux in JNUBEAM which produces

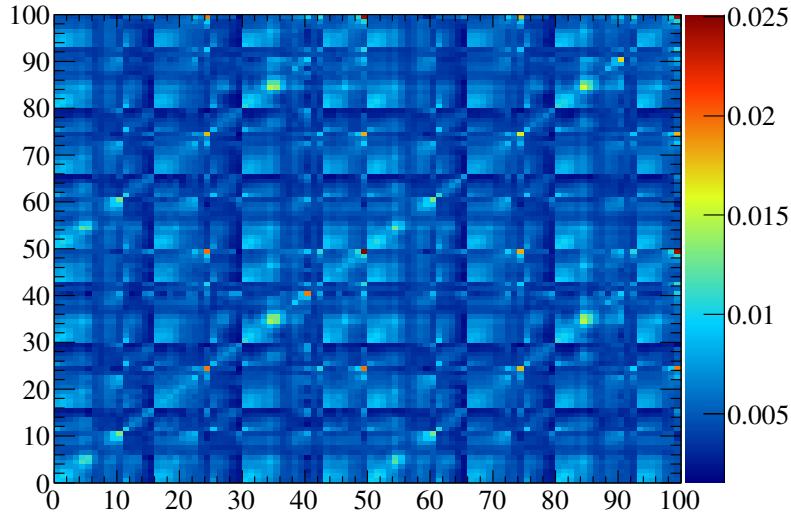


Figure 8.4: The flux prediction fractional covariance matrix. The matrix is binned primarily in detectors (ND280 and SK). There is then a subsequent binning scheme for each detector in neutrino flavour (ν_μ , $\bar{\nu}_\mu$, ν_e and $\bar{\nu}_e$) and neutrino energy.

a covariance matrix for each source. The covariance matrix to be used by the final analyser is the sum of each covariance matrix which is shown in Fig. 8.4. The covariance matrix is split into two detector sections: ND280 and SK. Each detector section is then separated into two beam running modes (ν and $\bar{\nu}$ runnings) which are also separated into four neutrino flavours (ν_μ , $\bar{\nu}_\mu$, ν_e and $\bar{\nu}_e$). Finally, each neutrino flavour is separated into a set of energy bins. The presented analysis is a ν_μ cross-section measurement using ND280, so only the ND280, ν -running mode section of the covariance matrix needs consideration. The provided flux covariance matrix is a fractional covariance matrix, binned in neutrino energy and flavour. As the neutrino energy and flavour information of events seen in this analysis is readily available, the flux covariance matrix can be used to reweight the events in the Monte Carlo sample to propagate the effect of the systematic uncertainty through the analysis. To do this, the flux covariance matrix is Cholesky-decomposed and the resulting matrix multiplied by a vector filled with random throws from a Gaussian of mean 0 and width 1. Each element in this multiplied vector constitutes a fractional change in the number of a specific neutrino flavour and energy seen in the analysis. So, the event weightings are generated by adding 1 to each element of this vector. Every event selected in the sample is then weighted by the correct event weight (in this case defined by the neutrino flavour and energy) and this number of reweighted events is then recorded.

The above description describes a single throw of the flux systematic propagation. This process is repeated 1000 times to build up the covariance matrix for the sample. The sample covariance matrix elements are calculated as

$$V_{ab} = \frac{1}{N^{\text{throws}}} \sum_{i=1}^{N^{\text{throws}}} \frac{(N_a^i - N_a^{\text{nom}})(N_b^i - N_b^{\text{nom}})}{N_a^{\text{nom}} N_b^{\text{nom}}}, \quad (8.5)$$

where N_a^{nom} is the number of events seen in the nominal Monte Carlo for sample 'a' and N^{throws} is the number of systematic throws. N_a^i is generically the number of events seen in sample 'a' for systematic throw i , but its definition depends on which kind of covariance matrix is being calculated. For a shape+normalisation covariance matrix, N_a^i is defined as

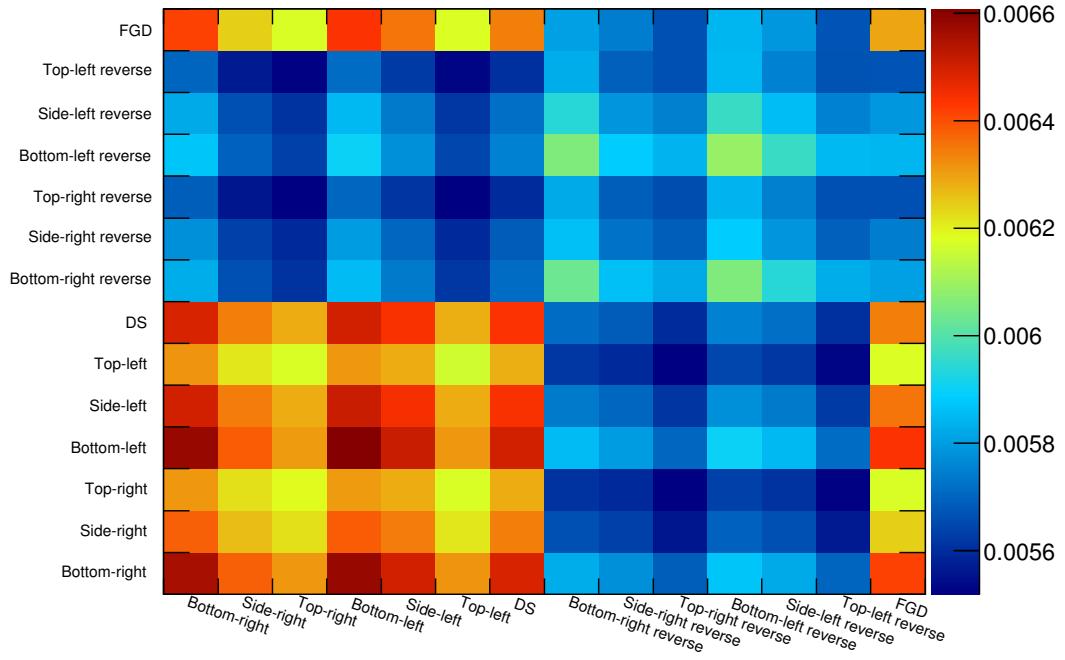
$$N_a^i = N_a^{ii}, \quad (8.6)$$

where N_a^{ii} is simply the number of events in sample 'a' for throw i after applying the systematic variation. When considering a shape-only covariance matrix, N_a^i is

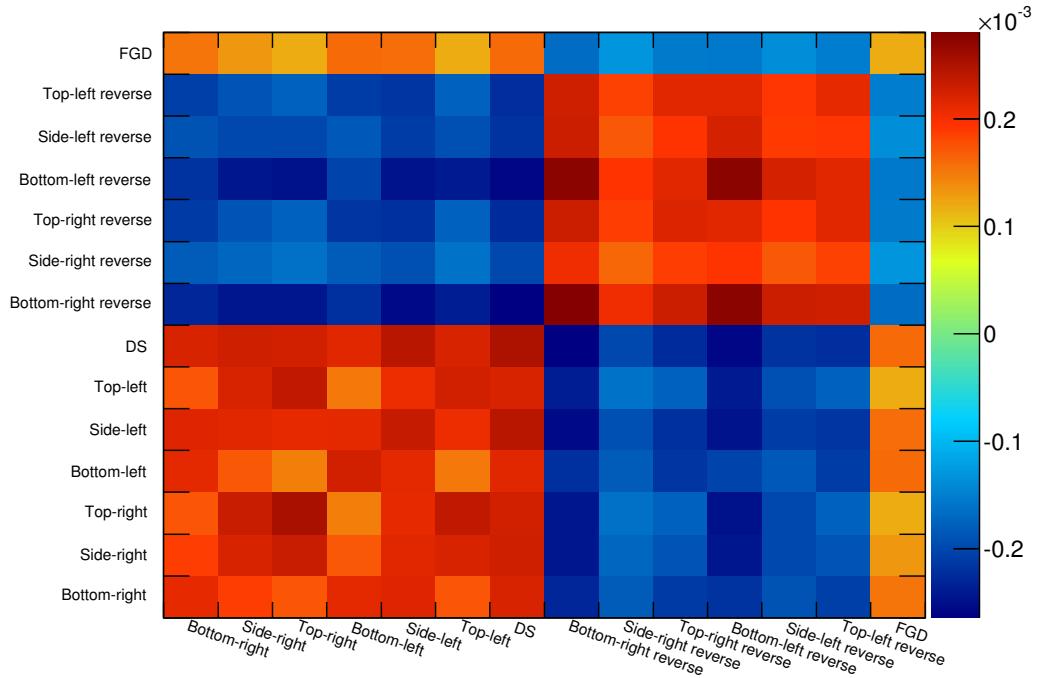
$$N_a^i = N_a^{ii} \frac{\sum_{j=1}^{14} N_j^{\text{nom}}}{\sum_{j=1}^{14} N_j^{ii}}. \quad (8.7)$$

The extra factor on the right hand side of equation 8.7 conserves the total number of events seen such that there are an equal total number of events before and after applying the systematic variation. The sample covariance matrices found by applying the above process are shown in Fig. 8.5.

The effect of the flux systematic on the selection efficiency of ν_μ CC interactions on lead was also evaluated. As the main target used in this analysis is the DS ECal, the efficiency variation was evaluated for this detector only. For each systematic throw described above, the selection efficiency was calculated and recorded. By plotting the efficiency for each throw, the width of the efficiency distribution can be found which gives the efficiency uncertainty induced by the flux systematic. This efficiency uncertainty was found to be 0.00097. The lead selection efficiency in the DS ECal is nominally 0.539, which implies a fractional change of 0.18%, which can be considered to be negligible.



(a) Shape+normalisation.



(b) Shape-only.

Figure 8.5: The sample fractional covariance induced by the neutrino flux uncertainties. The same binning for sample-based covariance matrix elements is used in subsequent plots in this chapter.

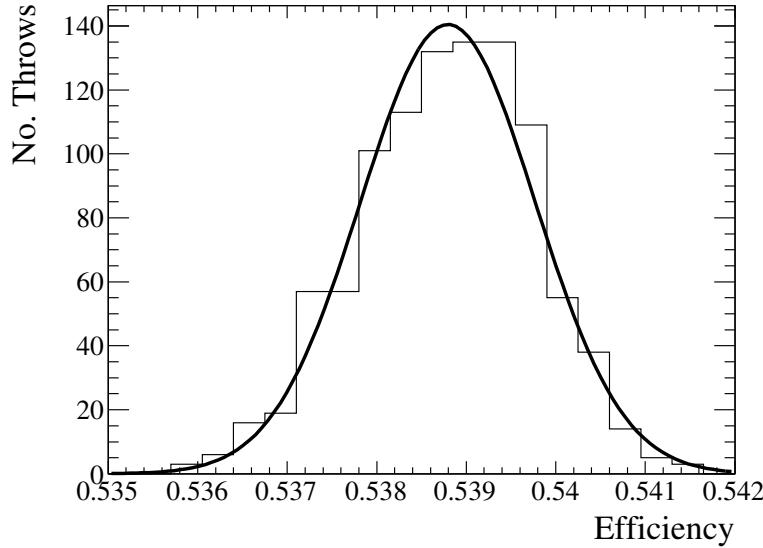


Figure 8.6: The variation in the selection efficiency of ν_μ CC interactions on lead in the DS ECal caused by the neutrino flux systematic. The width of the distribution is 0.18% of the size of the nominal efficiency value (0.539).

8.4.2 Cross-section systematic evaluation

The neutrino cross-section picture is still a very much open area of research. The interaction models used by current neutrino interaction generators use a set of empirical parameters which are tuned on experimental results. Uncertainties on any one of these empirical parameters will alter the simulated cross-section and thus the event rate seen in the ND280 simulation. A dedicated working group handles the assessment of the empirical parameters and their corresponding uncertainties. The parameter uncertainties are given as a covariance matrix along with the known correlations which is shown in Fig. 8.7. The interaction parameter assessment is chiefly undertaken with the T2K oscillation analyses in mind which means that the nuclear model parameters are only considered for carbon and oxygen. In the presented analysis, iron and aluminium are two major sources of background interactions which must also be treated in the systematic assessment. The chosen treatment is to assign a data-motivated normalisation uncertainty to such backgrounds. The iron uncertainty is taken from the iron cross-section measurement using INGRID [86]. The aluminium uncertainty is taken from the total ν_μ cross-section on aluminium measurement using the IHEP-JINR detector [87]. The IHEP-JINR detector was exposed to a neutrino beam with a 3 GeV to 30 GeV energy range which samples a higher energy than the T2K neutrino beam.

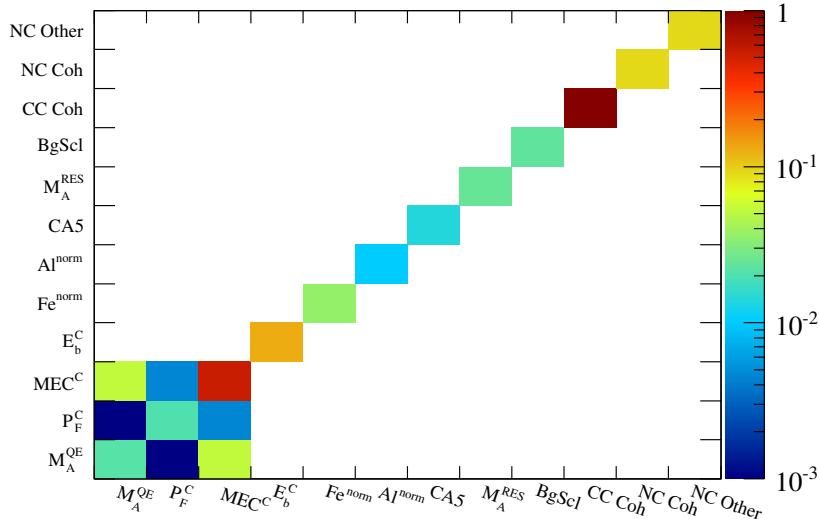
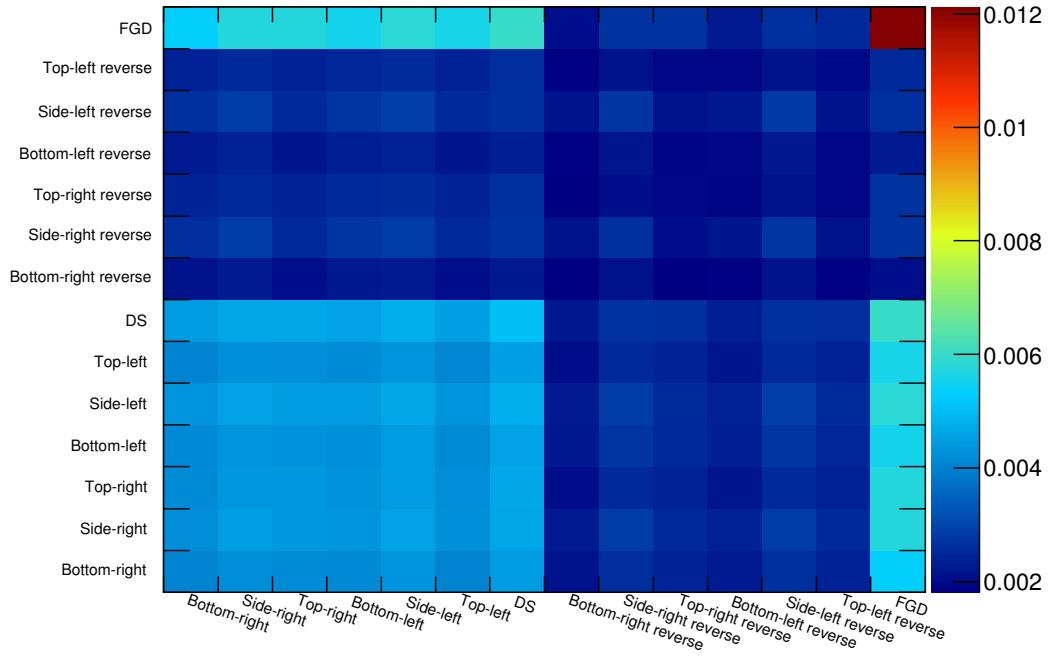


Figure 8.7: The neutrino cross-section model fractional covariance matrix binned in values of the NEUT cross-section parameters.

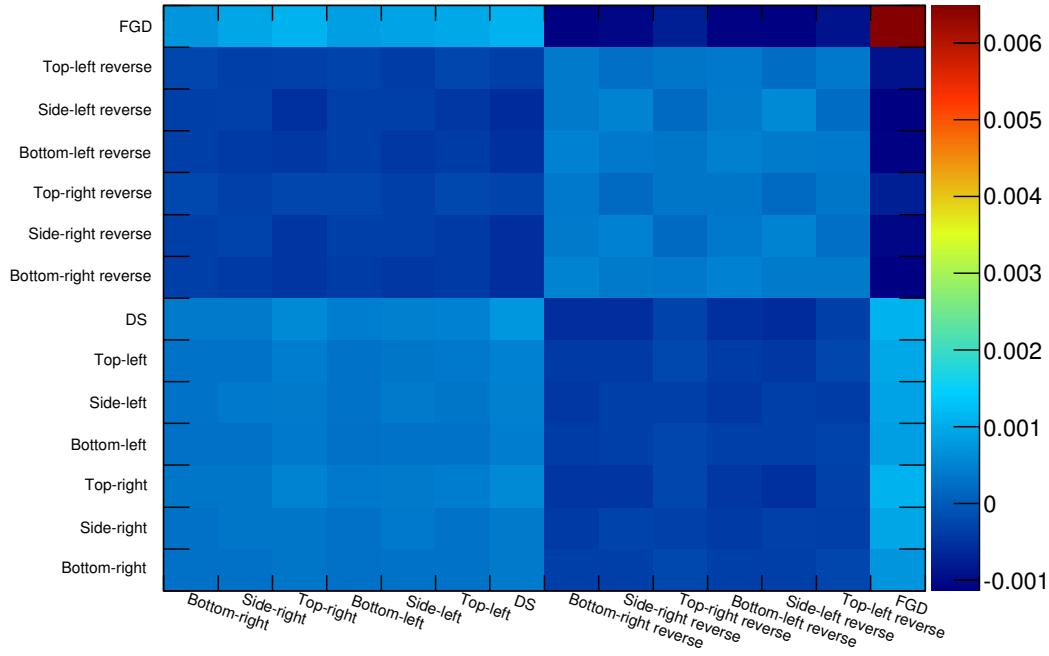
Because of this fact, it was decided to double the quoted aluminium cross-section uncertainty. As element specific parameters for lead do not appear in Fig. 8.7, such parameters are additionally assessed separately using NEUT.

To propagate the effect of the cross-section systematic through the analysis, the covariance matrix shown in Fig. 8.7 was Cholesky-decomposed and the resultant matrix multiplied by a vector of random numbers thrown from a Gaussian (the same approach as described in section 8.4.1). However, the re-weighting in this case is more complicated. The T2KReWeight [88] software takes the parameters errors from the throw along with the simulated NEUT vertex which is matched to the event being varied. T2KReWeight then calculates a weight based on the inputted errors and the NEUT vertex in question. Using the T2KReWeight machinery, 1000 systematic throws from the covariance matrix shown in Fig. 8.7 were made and sample covariance matrices were calculated, which are shown in Fig. 8.8.

As with the flux systematic, the uncertainty in the selection efficiency of ν_μ CC interactions on lead in the DS ECal due to the cross-section model systematic was found by recording the efficiency for each throw. The variation in the efficiency is shown in Fig. 8.9. The efficiency uncertainty was found to be 0.001 which corresponds to a 0.2% error which is, again, negligible.



(a) Shape+normalisation.



(b) Shape-only.

Figure 8.8: The sample fractional covariance induced by the cross-section model uncertainties.

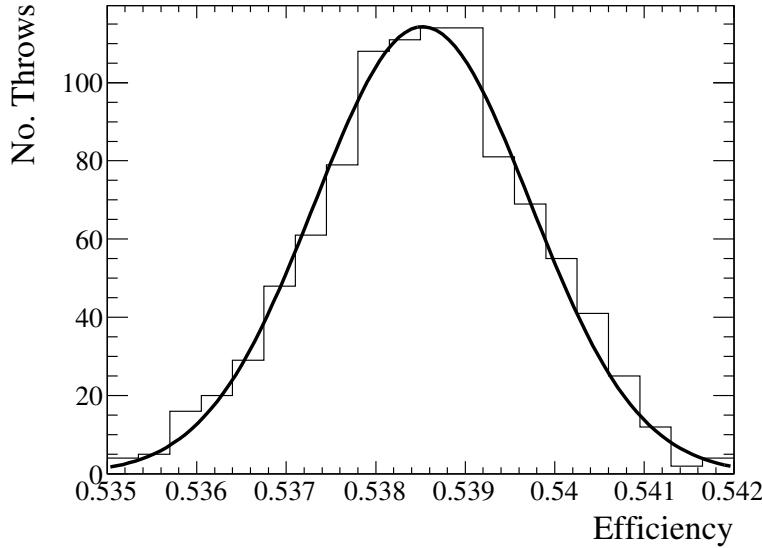


Figure 8.9: The variation in the selection efficiency of ν_μ CC interactions on lead in the DS ECal caused by the cross-section model systematic. The width of the distributions is 0.2 of the size of the nominal efficiency (0.539).

As discussed above, the re-weighting of element specific parameters for events matched to lead interactions was not possible using T2KReWeight. As meson-exchange current interactions for lead interactions are not implemented in NEUT, the only parameters necessary to vary are P_F^{Pb} and E_b^{Pb} which are the Fermi-momentum and binding energy for lead respectively. So, NEUT was used to produce 10,000 neutrino interactions on lead with uniform energy between 0 GeV and 10 GeV for the nominal parameters and $\pm \sigma$ variations of the parameters. The associated uncertainties were not readily available in the literature so 1.5 times the carbon parameter uncertainties were used as a conservative estimate. The NEUT events were used to plot the ratio of the varied cross-section to the nominal cross-section as a function of energy. The ratios for P_F^{Pb} and E_b^{Pb} are shown in Fig. 8.10a and Fig. 8.10b respectively. The variation in the cross-section caused by variation in E_b^{Pb} is less than one percent so there should be a negligible variation in the event rate which means this can be ignored. However, there is significant variation in the cross-section at low energy due to variations in P_F^{Pb} so this effect must be propagated through the analysis. To do this, the ratio distributions shown in Fig. 8.10a were treated as a set of events weights. The $+\sigma$ and $-\sigma$ weights were applied to the sample separately, creating two systematic throws. The results of the systematic throws were used to generate the sample covariances, which are shown

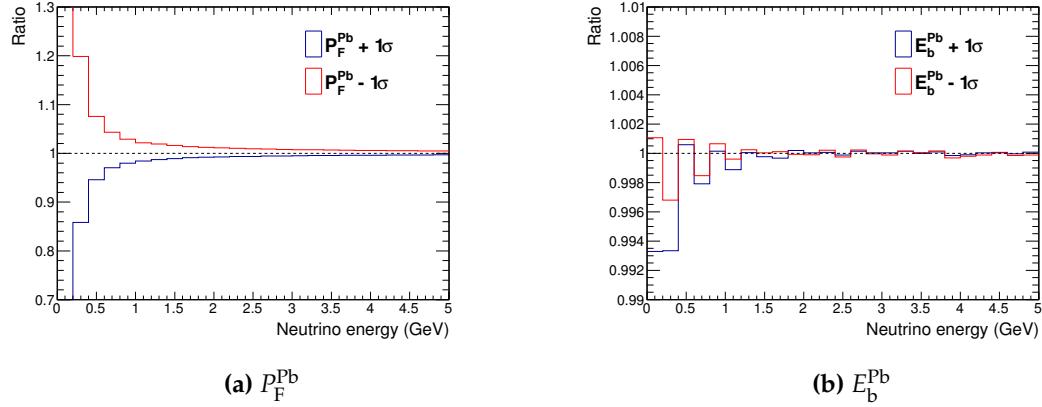


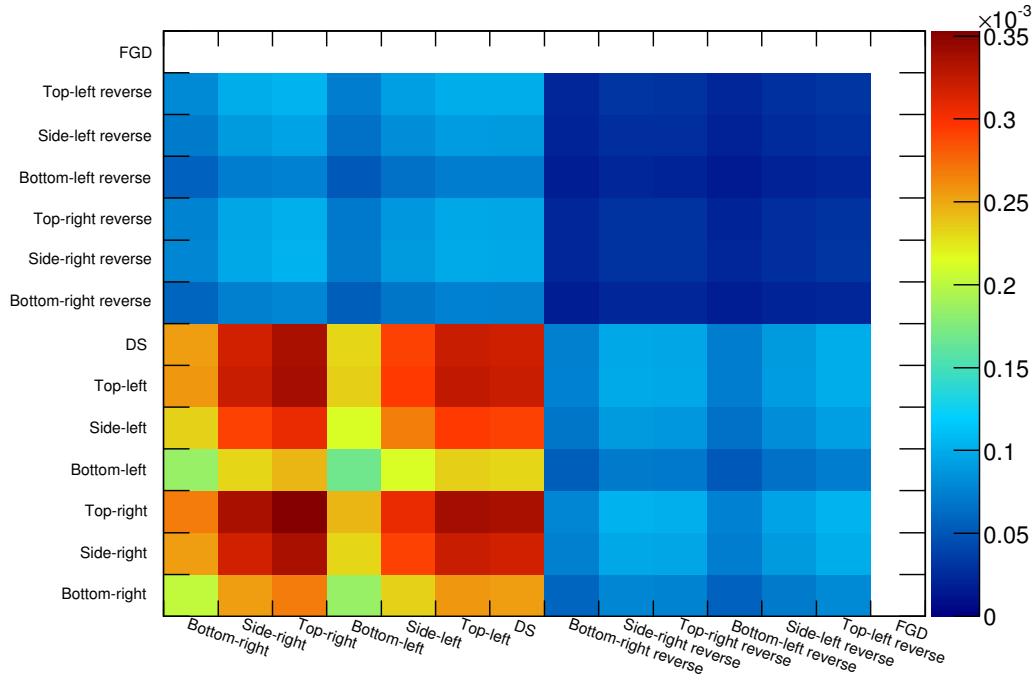
Figure 8.10: The ratio of the $\pm \sigma$ cross-section to the nominal cross-section for lead as a function of neutrino energy. Only variation in P_F^{Pb} shows a significant deviation of the cross-section from its nominal value and this only occurs as the cross-section approaches the sub-GeV energy range. The inputs to the distributions were created by simulating ν_μ CC interactions on lead with a uniform energy using NEUT. Events and distributions produced by W. Ma.

in Fig. 8.11.

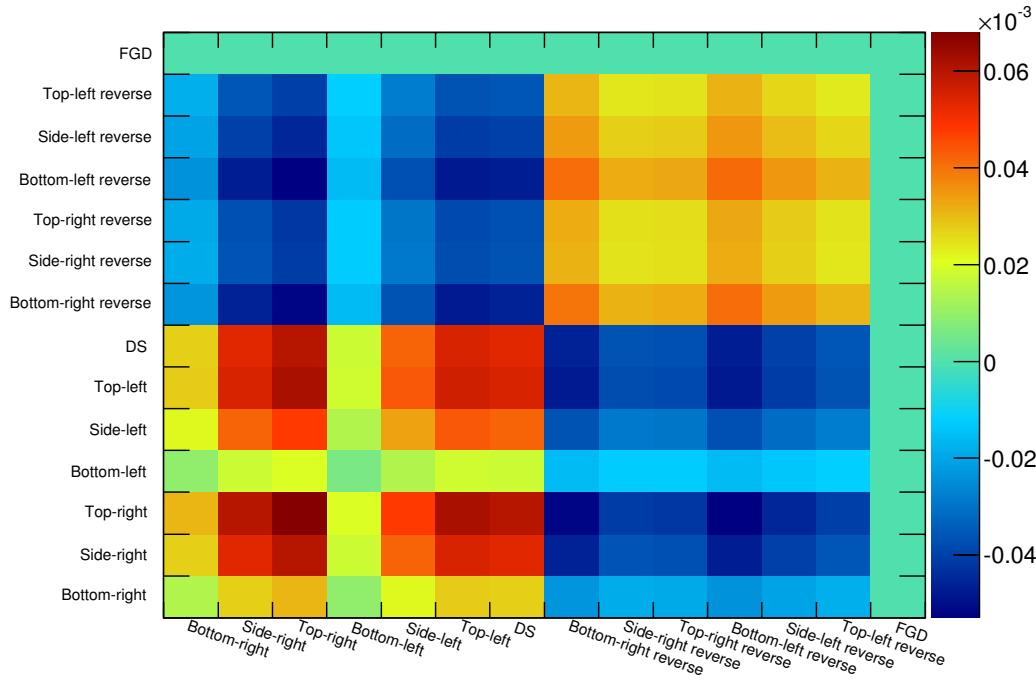
Final-state interactions (FSI) within the nucleus after a neutrino interaction has occurred could also cause a variation in the event rate. While the selection used in this analysis is CC inclusive, the method used separates events out into prong topologies. Any uncertainty in the FSI could cause migration of events from one prong topology to another. The Neutrino Interaction Working Group (NIWG) assesses these uncertainties and provides a set of parameters for an analyser to vary, all of which are associated with π^\pm interactions. There are six parameters associated with FSI:

- Low energy quasi-elastic scattering (FSIQEL)
- High energy quasi-elastic scattering (FSIQEH)
- Pion production (FSIINEL)
- Pion absorption (FSIABS)
- Low energy single charge exchange (FSICXL)
- High energy single charge exchange (FSICXH)

The NIWG provides combinations of the 1σ errors for the parameters highlighted above which span the whole FSI parameter space. The errors are provided as weights



(a) Shape+normalisation.



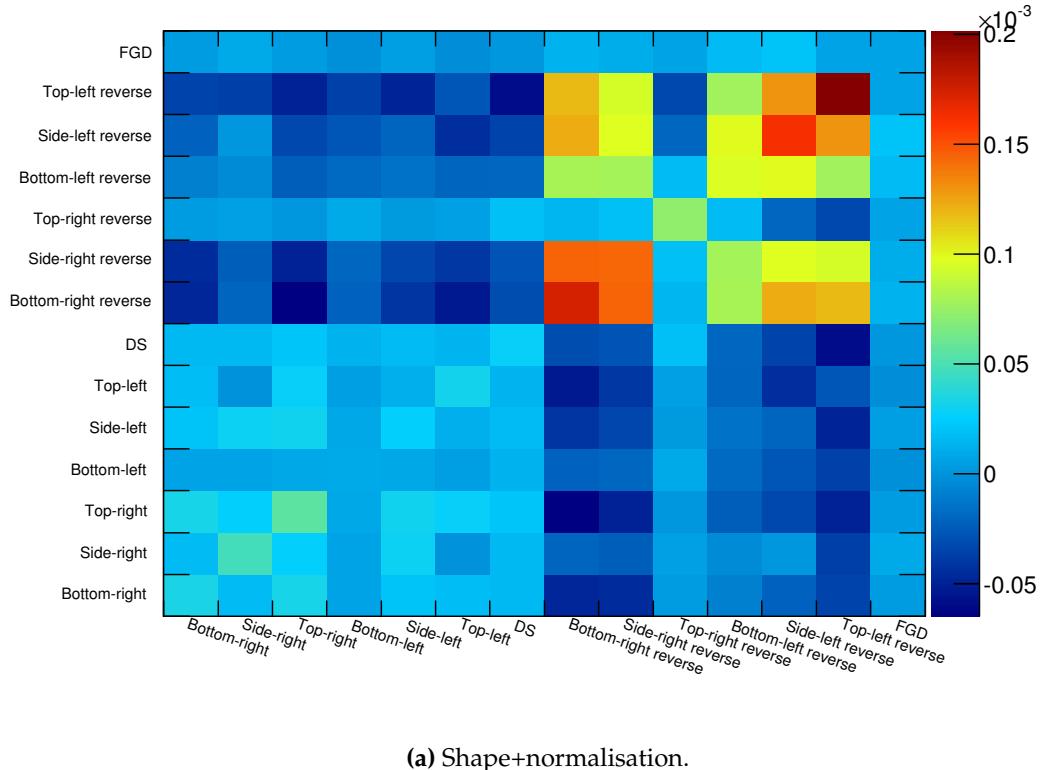
(b) Shape-only.

Figure 8.11: The sample fractional covariance induced by variation of P_F^{Pb} .

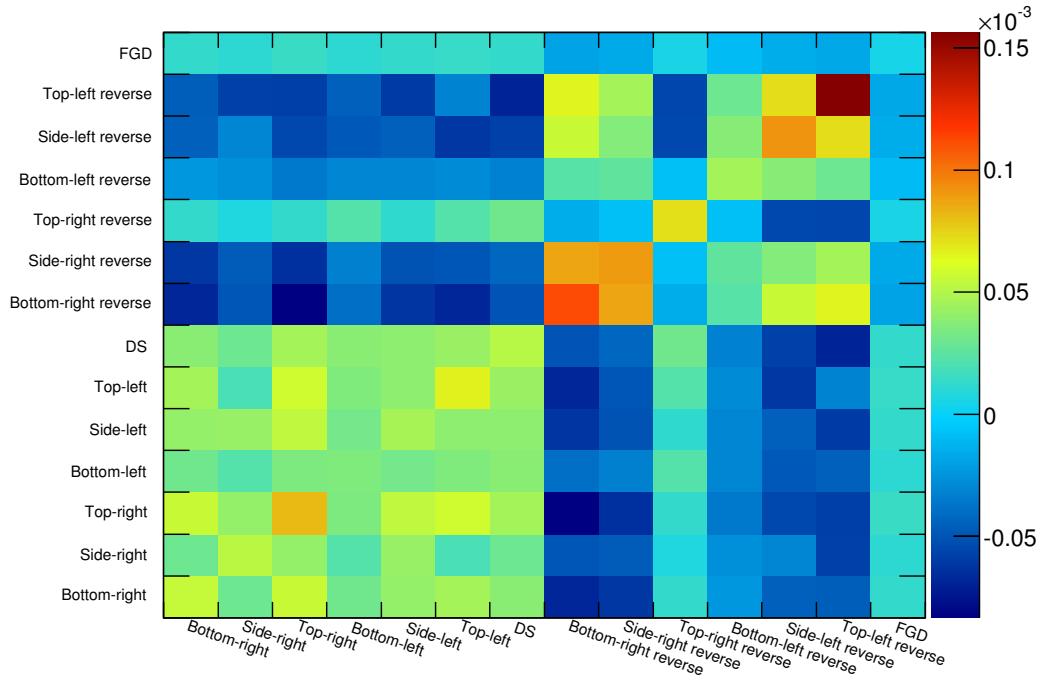
Parameter set	FSIQEL	FSIQEH	FSIINEL	FSIABS	FSICXL	FSICXH
Nominal	1.0	1.8	1.0	1.1	1.0	1.8
15	0.6	1.1	1.5	0.7	0.5	2.3
16	0.6	1.1	1.5	0.7	1.6	2.3
17	0.7	1.1	1.5	1.6	0.4	2.3
18	0.7	1.1	1.5	1.6	1.6	2.3
19	1.4	1.1	1.5	0.6	0.6	2.3
20	1.3	1.1	1.5	0.7	1.6	2.3
21	1.5	1.1	1.5	1.5	0.4	2.3
22	1.6	1.1	1.5	1.6	1.6	2.3
23	0.6	2.3	0.5	0.7	0.5	1.3
24	0.6	2.3	0.5	0.7	1.6	1.3
25	0.7	2.3	0.5	1.6	0.4	1.3
26	0.7	2.3	0.5	1.6	1.6	1.3
27	1.4	2.3	0.5	0.6	0.6	1.3
28	1.3	2.3	0.5	0.7	1.6	1.3
29	1.5	2.3	0.5	1.5	0.4	1.3
30	1.6	2.3	0.5	1.6	1.6	1.3

Table 8.1: FSI parameters, provided by the NIWG, representing the 1σ contour in the FSI parameter space [89].

in 16 parameters sets which represent the 1σ contour for the FSI parameter space and are shown in table 8.1. Each parameter set can be used to generate one systematic throw, by passing the parameters to T2KReWeight and subsequently re-weighting the events. As with the cross-section model systematics described above, the parameter variations only affect carbon and oxygen. The FSI systematic throws were used to generate the sample covariance matrices which are shown in Fig. 8.12.



(a) Shape+normalisation.



(b) Shape-only.

Figure 8.12: The sample fractional covariance induced by variation of the FSI parameters outlined in table 8.1.

8.4.3 ECal detector systematic evaluation

The analysis presented makes heavy use of a computational model of the ND280 ECal. Any differences between this model and the real ND280 ECal could cause a systematic difference between the analysed Monte Carlo and the collected data used in this study. So, several key areas have been identified which could cause such a systematic discrepancy, under the assumption that the differences exist.

The implemented selections uses the enhanced reconstruction, outlined in chapter 5. The enhanced reconstruction depends on a set of clustering algorithms which are outlined in section 4.4 and those clustering algorithms depend on the hit scintillator bars themselves. Following this trail back, it can be argued that the vast majority of the ECal detector systematics can be covered by assessing the ability of the ECal to reconstruct individual hits. Specifically, the vast majority of the ECal's systematic uncertainties can be assessed by investigating the hit efficiency of the scintillator bars.

8.4.3.1 The hit efficiency of the ECal scintillator bars

The aim of this study is to quantify the relative ECal hit inefficiency of collected data when compared to MC and use that information to invoke a similar ECal hit inefficiency in the MC. The effect of the hit inefficiency can then be propagated through the reconstruction and analysis to study its effect on the number of selected events. Fig. 8.13 shows the layer-by-layer efficiency of the ND280 ECals, measured using cosmic data [60]. Cosmic rays, which are largely composed of MIPs, are typically through going and only hit one bar per layer. So, the efficiencies shown in Fig. 8.13, can be used as a proxy for the hit efficiency of the scintillator bars. Fig. 8.13 contains almost all of the necessary information required to study this systematic uncertainty with the only missing information being the hit efficiency of the MC ECal. Due to time constraints, it was necessary to make a couple of conservative assumptions:

- The MC hit efficiency is 100%.
- The data hit efficiency for all bars is the lowest efficiency shown in Fig. 8.13, which is 96.2%.

Bearing the assumptions in mind, the hit inefficiency which must be applied to the MC is 3.8%. Care must be taken when attempting to apply the hit inefficiency to the MC.

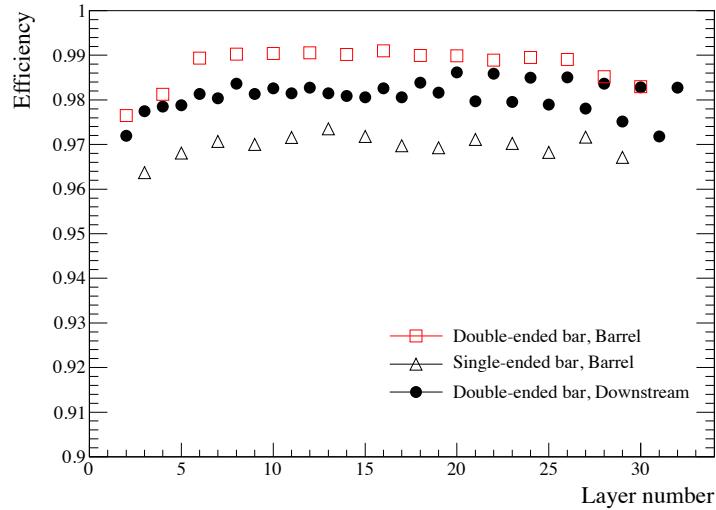


Figure 8.13: The layer-by-layer efficiency for the ND280 ECals, measured using cosmic data [60]. The efficiencies were calculated by studying the rate of hits in each bar and the rate of hits in the bar's nearest neighbours in the surrounding layers.

Naively, one might attempt to randomly remove 3.8% of hits from the reconstruction; however, this would not best represent the hit inefficiency. There are many reasons why a scintillator bar would not be 100% efficient; however, the efficiency should correlate with the deposited charge. So, to best model this systematic effect, the lowest 3.8% of hits should be removed from the MIP peak in charge space. Fig. 8.14 shows the deposited charge for each hit created by MC cosmic ray muons. The distribution is cut off at 5.5 PEU as this is a minimum threshold for output hits in the calibration stage of the ND280 software. The clearly visible MIP peak is very close to this threshold and is actually partially cut off by this limit. So, to apply the hit inefficiency, a new threshold needs to be calculated which removes an extra 3.8% of the lowest charge hits. By applying this logic to the hits used in Fig 8.14, the new threshold was calculated to be 7.55 PEU. This new threshold was applied as an initial stage of the reconstruction. The reconstruction and selection was then reapplied to the sample. This new sample was then compared to the nominal sample to construct the covariance matrices, which are shown in Fig. 8.15. The largest uncertainty for the samples was found to be 3.74%.

8.4.3.2 The charge resolution of the ECal scintillator bars

The hit efficiency systematic uncertainty actually covers several areas of uncertainty. Firstly, it addresses how a change in the core input to the reconstruction changes its output and so, by extension, addresses the efficiency of the reconstruction. However, because the inefficiency is invoked by removing low charge hits, it also addresses how low charge hits, which could migrate below the minimum reconstruction threshold, could affect the analysis. Because treatment of the charge by the hit efficiency systematic is partially covered, it is only necessary to additionally address how variation in the charge of the already reconstructed objects could alter the number of objects which pass the selection cuts. To assess this, a well understood side-band sample is needed such that the hit charge in MC and data can be compared. Any shift or smearing of the charge in data, relative to Monte Carlo must then be added to the reconstructed objects, which can then be passed through the selection, allowing the covariance matrices to be constructed. To study the variation in the hit charge, a sample of cosmic MC and data was used. The two samples were passed through the enhanced reconstruction and the charge of the hits associated to the reconstructed events were then analysed. An example of the reconstructed hit charges is shown in Fig. 8.16a. There are clear differences in the shape of the two distributions; the data charge peak is wider than the MC peak. To quantify this difference in the peak of the distributions, a Gaussian

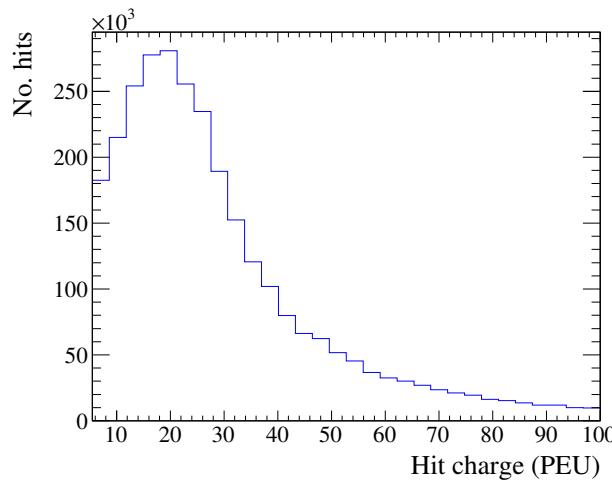
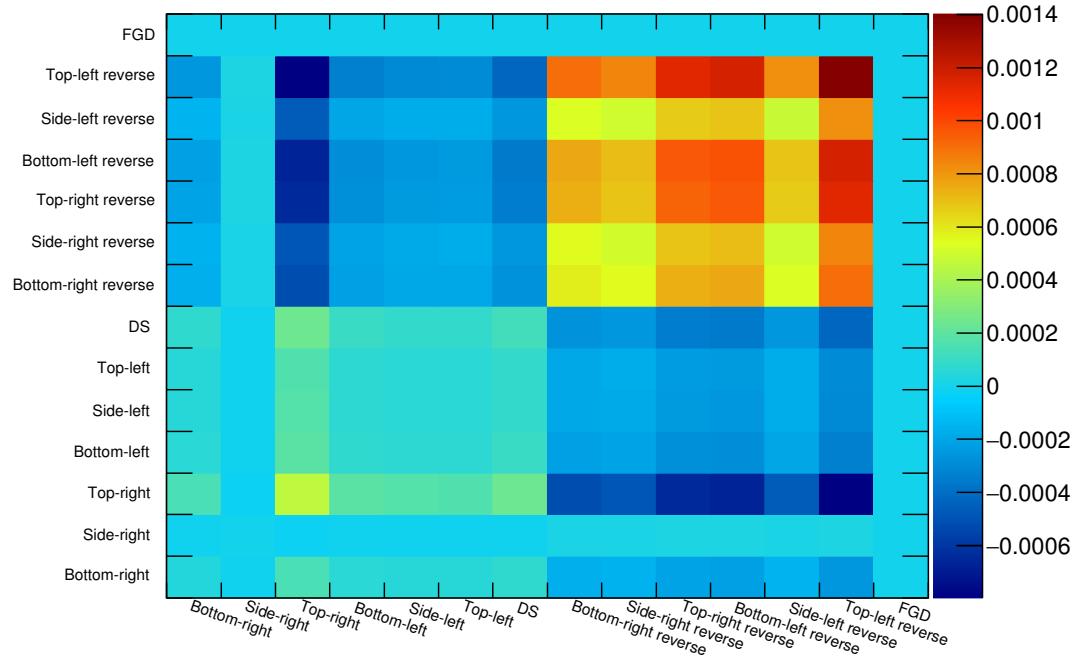
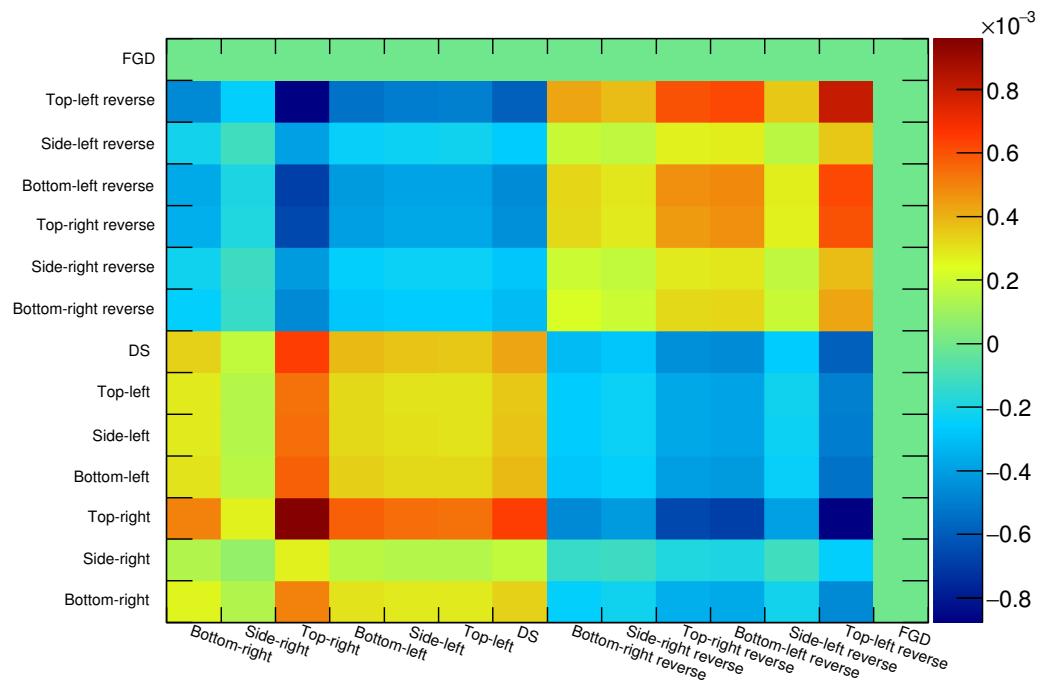


Figure 8.14: The charge distribution of hit scintillator bars created by MC cosmic muons. The distribution is truncated at 5.5 PEU as that is the minimum charge threshold for hits to be reconstructed in the ECal reconstruction. The peak of the charge distribution is located near to the 5.5 PEU threshold.



(a) Shape+normalisation.



(b) Shape-only.

Figure 8.15: The sample fractional covariance induced by the ECal hit inefficiency.

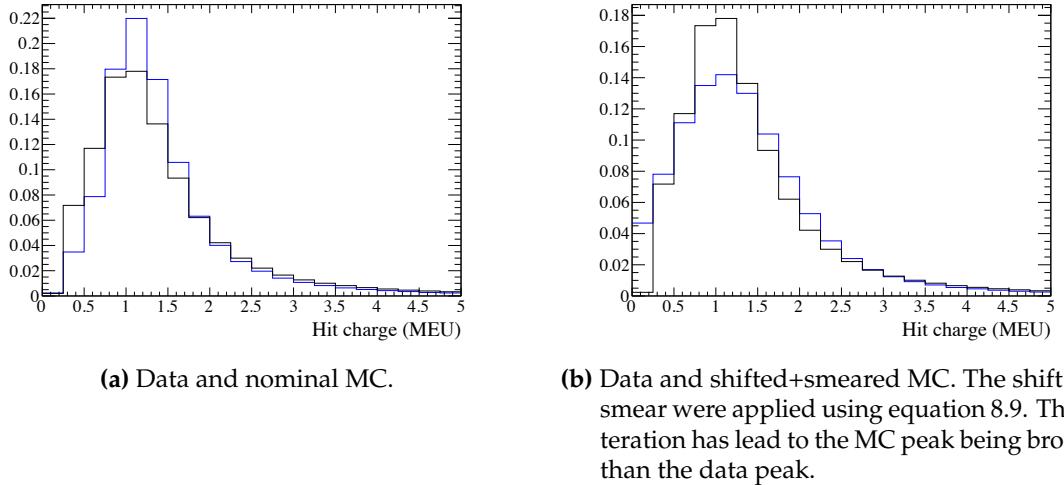


Figure 8.16: The charge of ECal hits after applying the enhanced reconstruction to run 3C cosmic events. The distributions only show information for reconstructed objects in the top-left barrel ECal module. The blue and black histograms are Monte Carlo and data respectively. All distributions are area normalised.

was fit to the top 66% of the two peaks which were then compared. The relevant parameters for the comparison are the width of the data and MC fits, σ_{Data}^Q and σ_{MC}^Q respectively, and

$$\Delta^Q = \mu_{\text{Data}}^Q - \mu_{\text{MC}}^Q, \quad (8.8)$$

where μ_{Data}^Q and μ_{MC}^Q are the mean of the data and MC fits respectively. Δ^Q measures the shift of the MC peak relative to data, while σ_{Data}^Q and σ_{MC}^Q can be used to quantify the relative width. These values are shown in table 8.2. To truly quantify the difference in the charge distributions, not only is a well understood sample required, but also a clean set of reconstructed events to remove any additional effects which may mask the charge difference. As the cosmic events will generally be almost parallel to the vertical axis, only the top and bottom ECal modules will be able to cleanly reconstruct the cosmic events, so only those events should be considered. Ideally, the values shown in table 8.2, would be used to correct and smear the MC peak to match data. However,

due to time constraints, it was only possible to use said values to over-correct and over-smear the MC. In terms of the individual hits, this meant using Δ^Q and σ_{Data}^Q for the top-left ECal to apply a correction and smearing to the hit charges, regardless of which module the hit occurred in. The top-left ECal module was chosen because it has

the largest value of Δ^Q . The adjusted hit charge is

$$Q' = Q + X \quad (8.9)$$

where Q is the nominal hit charge and X is a random variable which is defined as

$$X \sim N(\Delta^Q, \sigma_{\text{Data}}^Q). \quad (8.10)$$

It is important to note here that the width of the Gaussian distribution in which the adjustment is drawn from is σ_{Data}^Q . By using this width, the MC hit charge is over-smearred relative to data. An example application of equation 8.9 is shown in Fig. 8.16b which uses the same hit information as that in Fig. 8.16a, but with the correction applied to the MC hit charge. As can be seen by comparing the two figures, the application of the over-smearing has caused the MC charge peak to be wider than the data peak. To propagate the effect of this systematic, this smearing and correction needs to be applied to the reconstructed prongs before any selection takes place. Unfortunately, hit level information is not available at the level where the selection is applied. So, a slightly modified charge adjustment is needed. If the total charge contained on a prong is Q_{prong} , then the total adjusted charge is

$$Q'_{\text{prong}} = Q_{\text{prong}} + X_{\text{prong}} \quad (8.11)$$

where X_{prong} is a random variable and is defined as

$$X \sim N(\Delta^Q N_{\text{prong}}, \sigma_{\text{Data}}^Q), \quad (8.12)$$

ECal module	Δ^Q (MEU)	σ_{Data}^Q (MEU)	σ_{MC}^Q (MEU)
Bottom-right	-0.095	0.37	0.51
Bottom-left	+0.045	0.37	0.61
Top-right	-0.003	0.38	0.58
Top-left	-0.099	0.38	0.50
Side-right	-0.225	0.83	0.64
Side-left	-0.315	0.85	0.55
DS	-0.472	0.87	0.46

Table 8.2: Summary of the lead absorbers for the ND280 Tracker ECals [60].

ECal module(s)	True (MC)	Estimated (MC)	Estimated (data)
Barrel	107 ± 10	108 ± 11	169 ± 16
DS	104 ± 10	106 ± 11	122 ± 18

Table 8.3: Summary of the study to estimate the number of noise hits per event in the ECals.

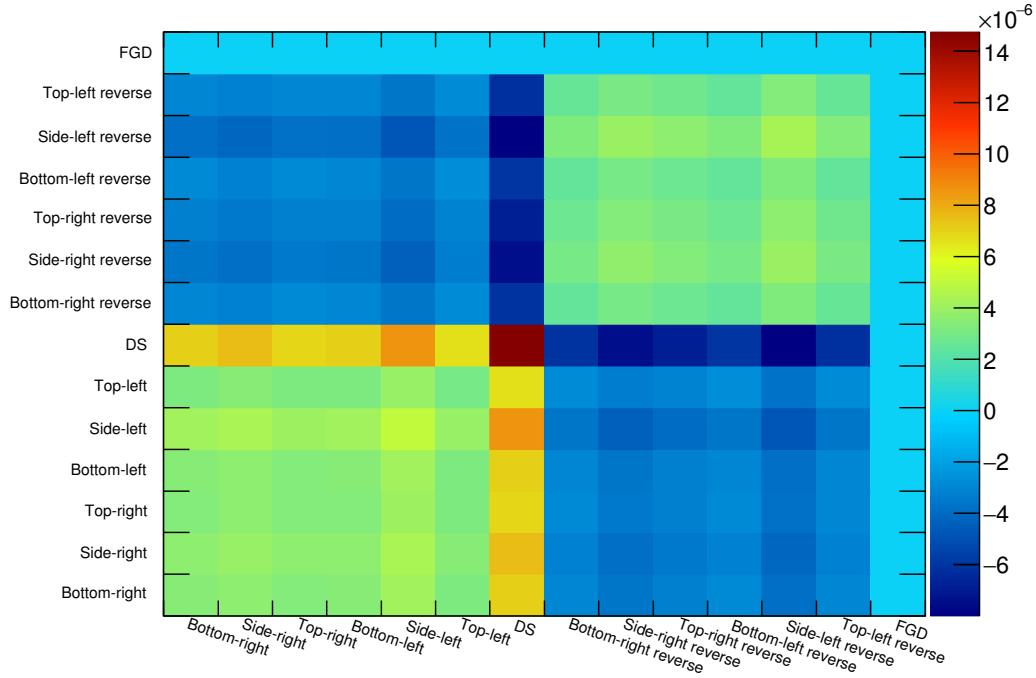
where N_{prong} is the number of hits contained on the prong. Following the above description, the prong charge variation was applied to every reconstructed prong in the systematic sample and the selection was then applied. This process was repeated 1000 times to build up the covariance matrices, which are shown in Fig. 8.17.

8.4.3.3 Inherent noise in the ECal

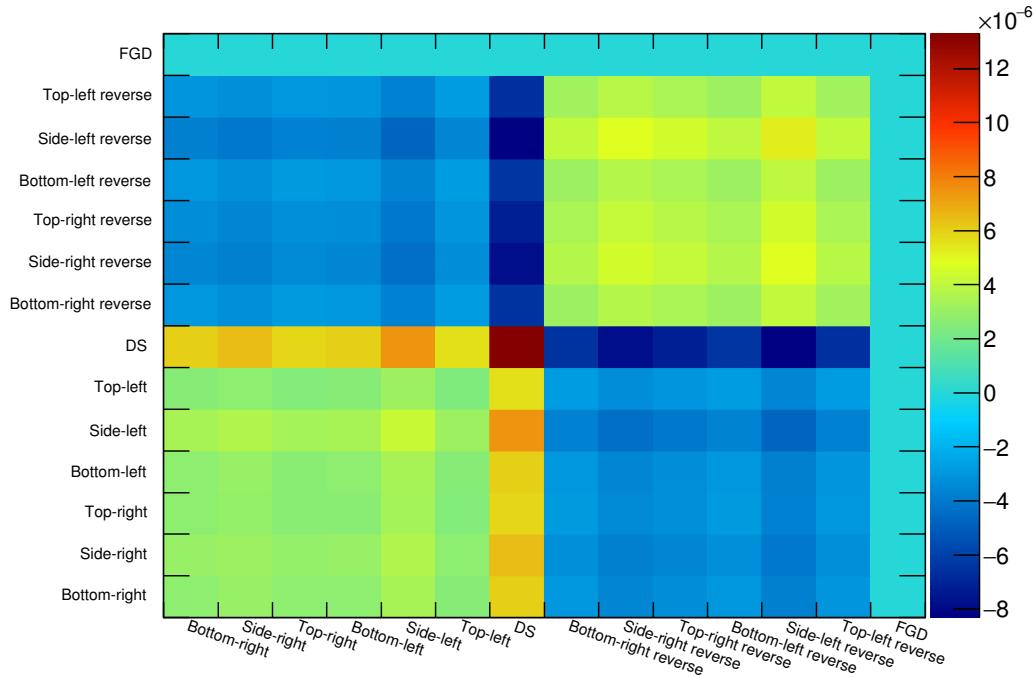
While the hit efficiency and charge systematic assessments are fairly all-encompassing in terms of ECal detector uncertainties, they can not address how the inherent noise in the ECal can affect the reconstruction and, by extension, the selected number of events in this analysis. There is only a need to address a noise systematic uncertainty if the simulated noise rate in the Monte Carlo is different to what actually happens in the ECal. The method chosen is taken from [84].

To measure the noise rate, a control sample of cosmic rays (for the barrel) and through-going muons (for the DS ECal) were passed through the reconstruction. Before the clustering stages were initiated, the number of hits in the relevant ECal module were recorded. Then, after the reconstruction chain was completed, the number of hits which formed the final reconstructed objects were also counted. The difference between these two numbers forms the noise hit estimate. To test this estimator, the true number of noise hits in the Monte Carlo were also counted. The results of this test are shown in Fig. 8.18. The Gaussian fits to the distribution are used to quantify the noise levels per event (by using the mean and width of the fits). These values are shown in table 8.3. The estimate is clearly in good agreement with the true noise levels in the Monte Carlo. So, it was decided to apply this method to estimate the noise levels in the data samples, the results of which are shown in Fig. 8.19 and also shown in table 8.3.

It is clear that there are differences between the noise levels in data and Monte Carlo. So, to propagate the effect of this systematic uncertainty, elecSim was retuned to



(a) Shape+normalisation.



(b) Shape-only.

Figure 8.17: The sample fractional covariance induced by the ECal charge uncertainties.

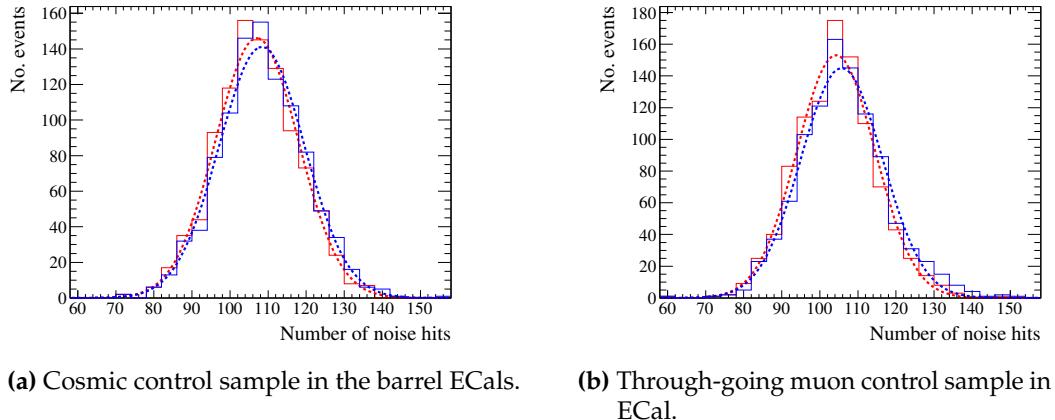


Figure 8.18: The estimation of the number of noise hits (solid blue line) and the number of true noise hits (solid red line) per event in Monte Carlo. The coloured dashed lines are the Gaussian fits to the respective distributions. The similarity of the estimated and true noise distributions shows that the estimator is accurate.

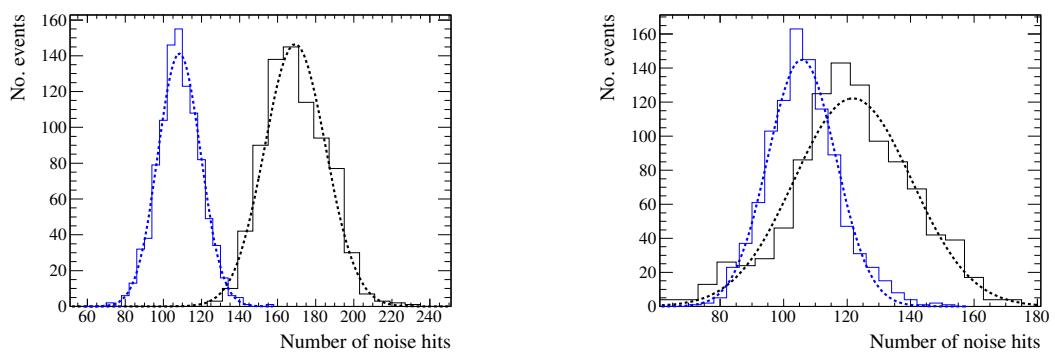


Figure 8.19: The estimation of the number of noise hits in Monte Carlo (solid blue line) and data (solid black line) per event. The coloured dashed lines are the Gaussian fits to the respective distributions.

produce the data-like noise levels in both the barrel and DS ECal. The systematic sample was then passed through the modified elecSim and all pieces of the software chain downstream. After this, the selection was applied to the modified sample. This sample was then compared to nominal and the difference was used to construct the covariance matrices, which are shown in Fig. 8.20.

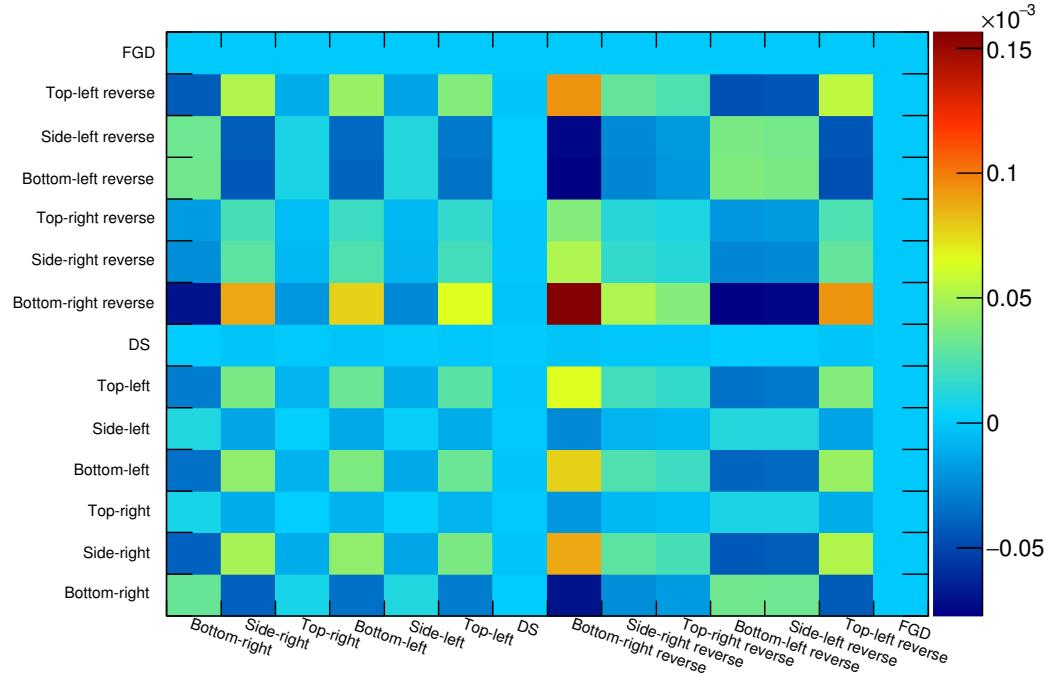
8.4.3.4 The mass of the ECal active volume

The ECal model used in the simulation was based on the as built drawings and so should be an accurate model. However, the components of each ECal module do have associated uncertainties which are not, and can not easily be, taken into account in the simulation, but could easily cause a systematic difference between data and Monte Carlo. For example, if the DS ECal lead absorbers were thicker in the simulated model, one should expect to see a relative deficit of beam triggered events in the DS ECal collected data. Provided that the component dimensions, compositions and uncertainties are known, toy Monte Carlo can be used to construct a set of toy ECal modules which model the variation in the module mass. This can be expanded further, to construct a covariance matrix for the total mass of each contributing element. This covariance matrix can then be used to re-weight the events, based on target element, in the Monte Carlo sample to model the effect of the ECal mass systematic.

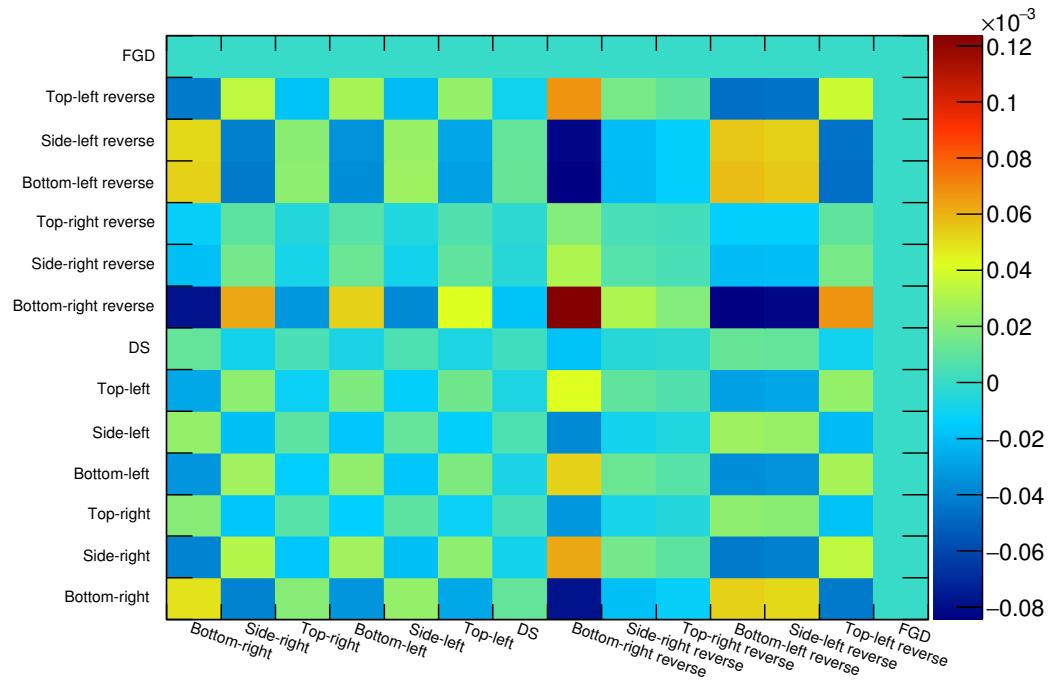
The ECal module designs are split into three types: side, top/bottom and DS. It is only necessary to generate a mass covariance matrix for each of the three types. All values used for the component dimensions/uncertainties are taken from [60]. Precise component dimensions are only quoted for the components in the active regions of each module, namely the lead absorbers, the scintillator bars and the wavelength-shifting fibres. So, only these components are considered in the toy Monte Carlo study.

8.4.3.4.1 The lead absorbers

Reference [60] provided a clear description of the composition of the lead absorbers for the ECal module designs, which are summarised in table 8.4. All component uncertainties (both dimension and composition) were given as symmetric errors. To construct an ECal lead layer, each layer dimension was randomly taken from



(a) Shape+normalisation.



(b) Shape-only.

Figure 8.20: The sample fractional covariance induced by the ECal noise uncertainties.

	Top/Bottom	Side	DS
No. per layer	2	4	2
Length (mm)	3858 ± 4	964.5 ± 4	2016 ± 1
Width (mm)	765 ± 4	2330 ± 4	1008 ± 4
Height (mm)	1.75 ± 0.1	1.75 ± 0.1	1.75 ± 0.1
Sb doping	$2.0 \pm 0.2\%$	$2.0 \pm 0.2\%$	$2.0 \pm 0.2\%$

Table 8.4: Summary of the lead absorbers for the ND280 Tracker ECals [60].

a Gaussian with a mean equal to the quoted dimension and a width equal to the dimension uncertainty. After the layer volume was defined, the Sb contamination was calculated, also using a Gaussian throw, but using the quoted Sb contamination and its uncertainty as the Gaussian's mean and width. The described information could then be used to retrieve the mass of the Pb and Sb in an ECal lead layer.

8.4.3.4.2 The scintillator bars

The information used to construct the toy scintillator bars is summarised in table 8.5, most of which was found in reference [60]. The length of the bars were quoted with symmetric uncertainties (like the lead absorbers), so the length was treated in the same way as the lead absorber dimensions. The hole and coating dimensions were also treated in the same manner. However, the width and height were quoted with asymmetric errors and so could not be treated in the same way. Only a negative uncertainty was quoted, which meant the width and height uncertainties could be treated in a somewhat similar manner, but with random numbers being drawn from a half-Gaussian. The bar composition was reported as polystyrene doped with 1% PPO ($C_{15}H_{11}NO$) and 0.03% POPOP. For the toy bar construction, the POPOP was considered negligible and so was not included. The result was that the toy bar composition (excluding the coating) was 99% CH and 1% $C_{15}H_{11}NO$ by mass. Unfortunately, the full composition of the TiO_2 coating was not extensively described. However, a similar systematic study has already been undertaken for the FGD [90] which included an investigation of the coating composition. A representative for the Fermilab facility which extruded the scintillator bars reported that the final composition of the coating was 15% rutile-form TiO_2 and 85% CH with other contributing compounds considered negligible. This information was also included in the modelling. The collected

	Top/Bottom	Side	DS
No. per layer (para)	38	57	-
No. per layer (perp)	96	96	50
Length (para) (mm)	3840 ± 0.1	3840 ± 0.1	-
Length (perp) (mm)	1520 ± 0.1	2280 ± 0.1	2000 ± 0.1
Width (mm)	$40^{+0.0}_{-0.4}$	$40^{+0.0}_{-0.4}$	$40^{+0.0}_{-0.4}$
Height (mm)	$10^{+0.0}_{-0.4}$	$10^{+0.0}_{-0.4}$	$10^{+0.0}_{-0.4}$
Hole diameter (mm)	1.75 ± 0.1	1.75 ± 0.1	1.75 ± 0.1
Composition (CH:PPO)	99:1	99:1	99:1
Coating thickness (mm)	0.25 ± 0.13	0.25 ± 0.13	0.25 ± 0.13
Coating composition (CH:TiO ₂)	85:15	85:15	85:15

Table 8.5: Summary of the scintillator bars for the ND280 Tracker ECals [60].

information could then be used to retrieve the contributing mass of each element in a scintillator bar.

8.4.3.4.3 The WLS fibres

The information used to constructed the WLS fibres is shown in table 8.6. As with the other toy components, reference [60] provided most of the necessary information for the construction and could be largely treated using the same methods as above. The major difference is that the WLS fibre diameter had an asymmetric error in which both components were not zero. So, two toy Gaussians were constructed to draw from (one representing each uncertainty) and a random number was first drawn from a uniform distribution to decide which Gaussian would be used. Very little information was provided about the composition of the WLS fibre; however, the mass study in the FGD [90] assumed the WLS fibre was composed entirely of scintillator, so the same assumption was used for this study. As with the other components, the provided information could then be used to retrieve the contributing mass of the elements in the WLS fibres.

8.4.3.4.4 ECal mass covariance

	Top/Bottom	Side	DS
No. per layer (para)	38	57	-
No. per layer (perp)	96	96	50
Length (para) (mm)	3840 ± 0.05	3840 ± 0.05	-
Length (perp) (mm)	1520 ± 0.05	2280 ± 0.05	2000 ± 0.05
Diameter (mm)	$1^{+0.02}_{-0.03}$	$1^{+0.02}_{-0.03}$	$1^{+0.02}_{-0.03}$
Composition	CH	CH	CH

Table 8.6: Summary of the WLS fibres for the ND280 Tracker ECals [60].

Toy simulation was used to construct each component described in the above sections. Each component was repeatedly built and was subsequently used to construct an ECal module. The contributing mass of each element to the toy ECal module was then recorded. 10,000 toy ECal modules were constructed for each type, totalling 30,000 constructed modules. The recorded contributing elemental mass for each construction was used to build up a mass covariance matrix using equation. 8.5 and equation. 8.6, an example of which is shown in Fig. 8.21 for the DS ECal. Unfortunately, it was found that the covariance matrices for all three module types were singular, which is of immediate concern as the matrices need to be Cholesky-decomposed as part of the systematic uncertainty evaluation. It was found that the singular nature of the covariance matrices was caused by the contributing elements which have small covariance values (both on and off diagonal) and the problem could be solved by not considering some of them in the covariance matrix. The three problematic elements are carbon, nitrogen and hydrogen. Both nitrogen and hydrogen only constitute a relatively small amount of the mass and an example of this is shown in Fig. 8.22 which shows the mass of the 10,000 toy DS ECal modules, with and without nitrogen and hydrogen. There is a 2.45% difference between the two constructions which is small but not necessarily negligible. While ideally the hydrogen and nitrogen terms should be included in the systematic evaluation, due to time constraints and that all covariance terms involving them were very small, it was decided that they would be excluded from the covariance matrix. With the extra elements excluded, a non-singular covariance matrix for each module type could be constructed, all of which are shown in Fig. 8.23.

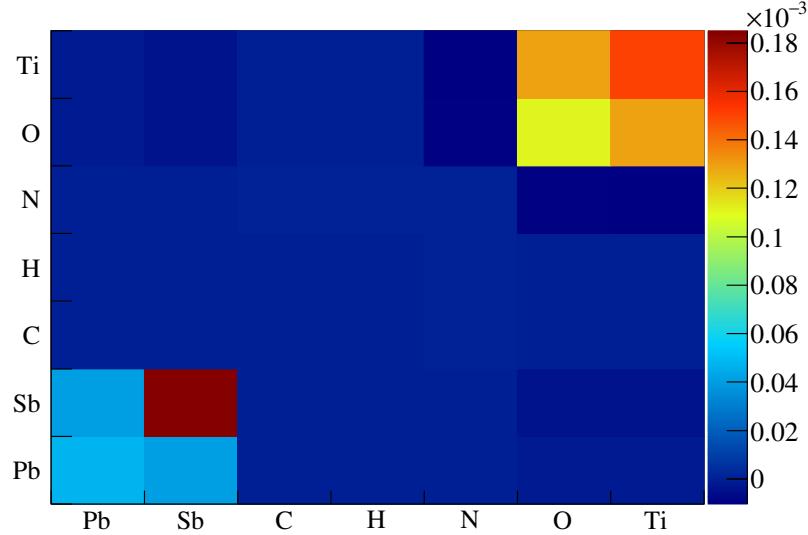


Figure 8.21: The fractional covariance matrix for the mass of the contributing elements in the DS ECal, with all contributing elements considered.

8.4.3.4.5 Propagation of the ECal mass systematic uncertainty

The covariance matrices shown in Fig. 8.23 were then used to generate systematic

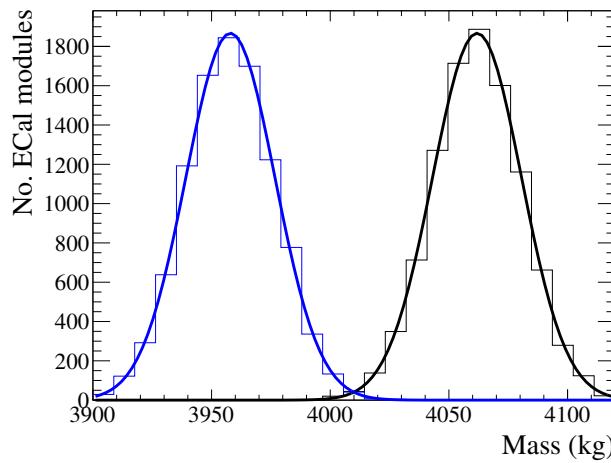


Figure 8.22: The total mass of the DS ECal active volume with (black) and without (blue) nitrogen and hydrogen as contributing elements. The smooth lines are the Gaussian fits to the distributions. Using the Gaussian mean and width as the mass central value and uncertainty respectively, the DS ECal active mass with (without) hydrogen and nitrogen is 4060 ± 20 kg (3960 ± 20 kg).

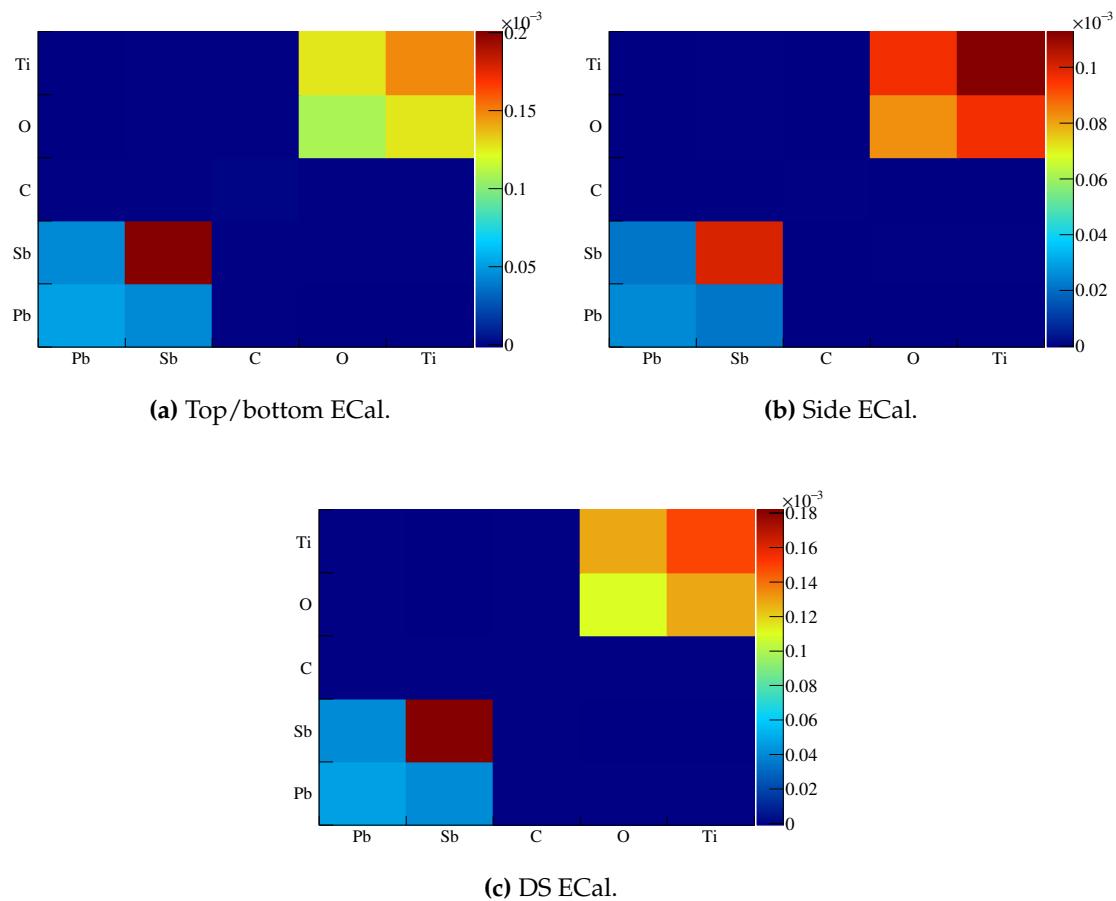


Figure 8.23: The fractional covariance matrices for the mass of each contributing element to an ECal module (hydrogen and nitrogen omitted).

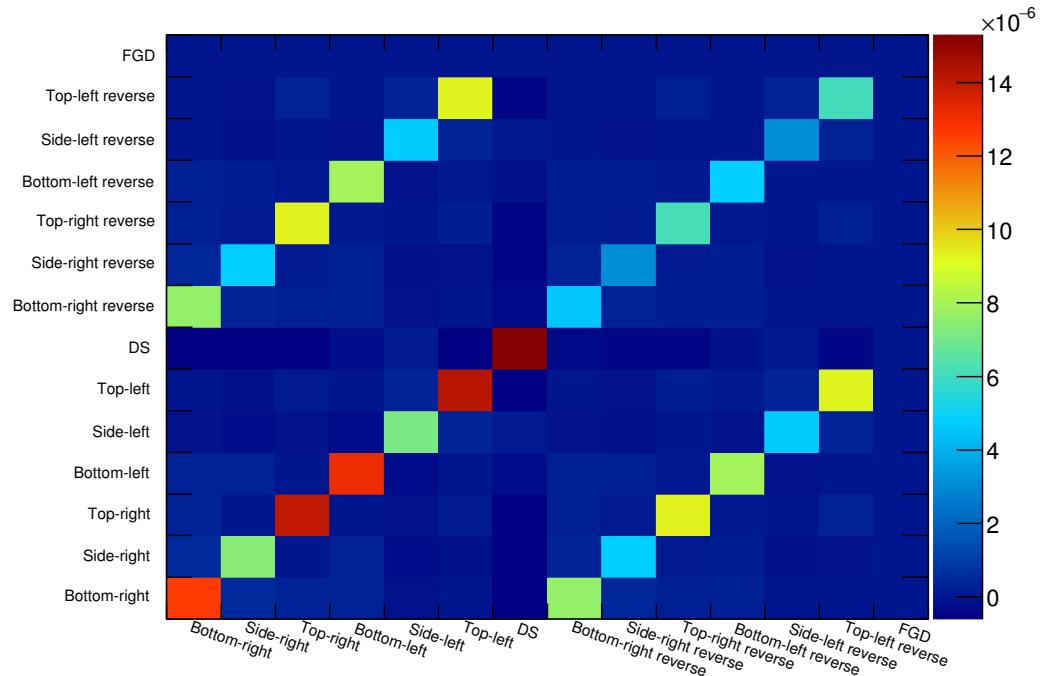
throws, in the same manner as the flux systematic uncertainty evaluation. For the input into each systematic throw i.e. one run through of the selected Monte Carlo events, a set of event weights were generated for each of the seven ECal modules using the above mass covariance matrices. Every neutrino interaction in the ECal active volume was weighted according to its target element. As with the other systematic uncertainty evaluations, 1,000 systematic throws were generated and the subsequent covariance matrices were constructed which are shown in Fig. 8.24. The maximum uncertainty, which was for the DS ECal events, was found to be 0.39%.

An additional source of systematic uncertainty could come from the ECal's ability to reject OOFV backgrounds. As fig. 7.42 shows, the dominant background selected, particularly for the barrel ECals, is the OOFV background. If the simulation of the ECal had a different OOFV background rejection capability to that of the actual ECal, the selected event composition between data and MC would differ, resulting in a systematic difference. The only feasible way to assess this systematic uncertainty is to use control samples. Unfortunately, only μ based control samples were available during the analysis. However, Fig. 7.42 shows that the one prong topology, which contains the largest number of selected events, has the biggest OOFV background contamination and as Fig. 7.33 and Fig. 7.34 show, μ s are the dominant particle species in the surviving OOFV background. To assess the systematic uncertainty, the μ based control samples, both data and MC, were passed through the selection and the number of reconstructed events which pass the fiducial volume cut were recorded, $N_{\text{CS}}^{\text{pass}}$. The efficiency of the fiducial volume cut was then defined as

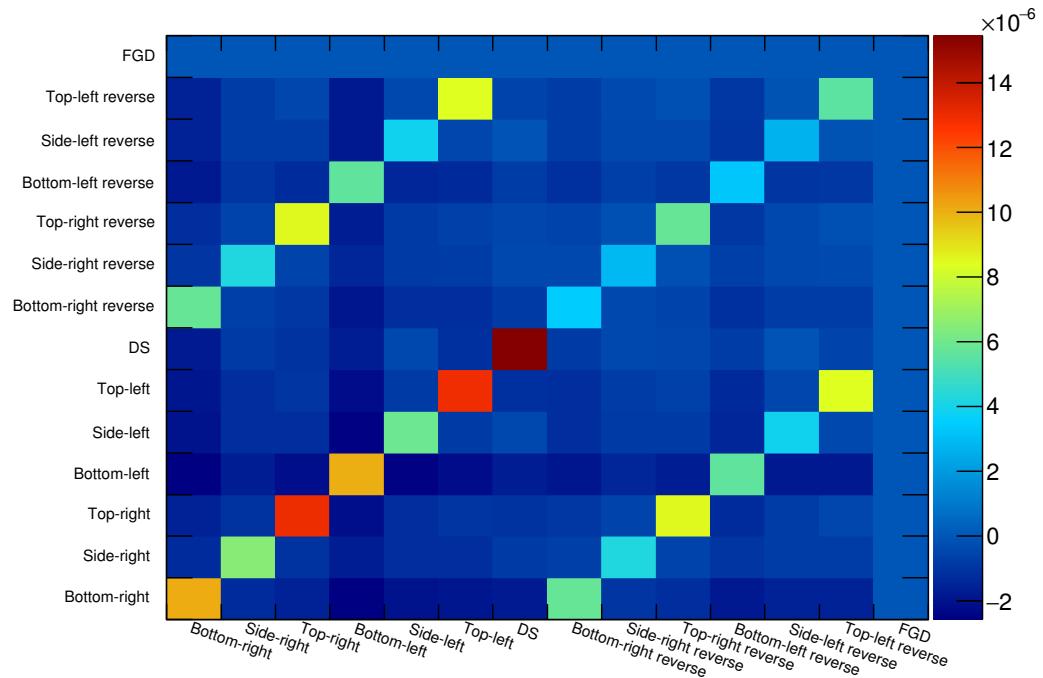
$$\epsilon_{\text{CS}} = \frac{N_{\text{CS}}^{\text{pass}}}{N_{\text{CS}}}, \quad (8.13)$$

where N_{CS} is the number of reconstructed control sample events which enter the selection. It is important to note why $N_{\text{CS}}^{\text{pass}}$ counts the number of events which pass the fiducial volume cut, rather than as the number of events which pass all of the selection cuts. The reason for this is the inclusion of the reverse sample (background-enriched sample) in the rate fit which includes any event which passes the fiducial volume cut, but fails the most upstream cut. By calculating ϵ_{CS} separately for data and MC, the systematic uncertainty can then be defined as

$$\alpha_{\text{OOFV}} = \epsilon_{\text{CS}}^{\text{data}} - \epsilon_{\text{CS}}^{\text{MC}}. \quad (8.14)$$



(a) Shape+normalisation.



(b) Shape-only.

Figure 8.24: The sample fractional covariance induced by the ECal mass uncertainties.

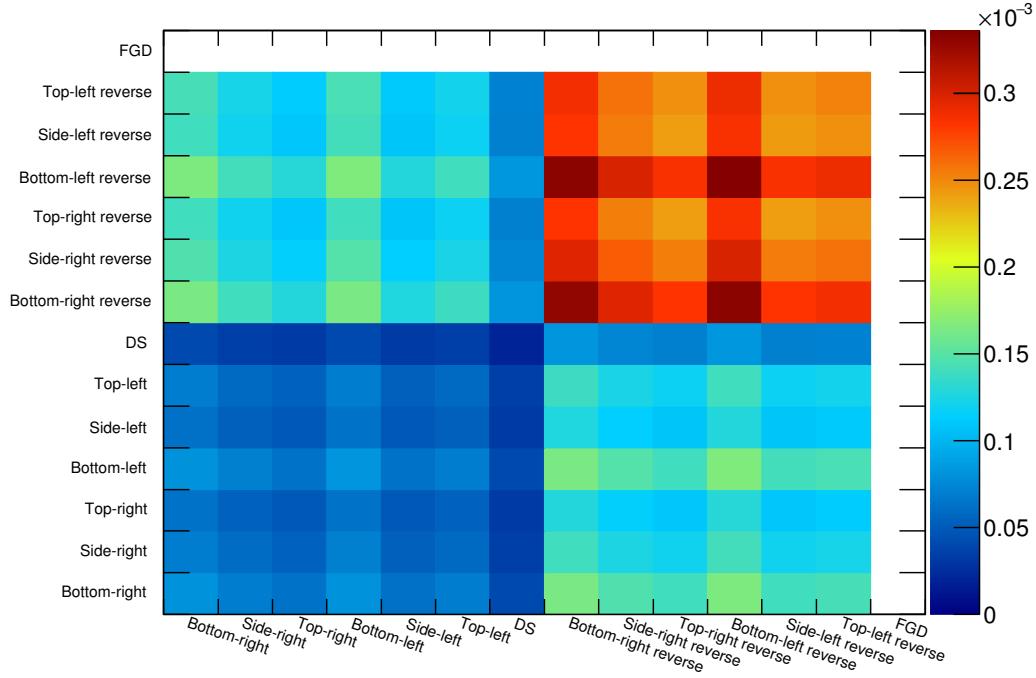
ECal module(s)	N_{CS} (MC)	N_{CS} (data)	ϵ_{CS} (MC)	ϵ_{CS} (data)	α_{OOFV}
Barrel	26560	42961	2.12%	5.92%	3.80%
DS	6225	22555	2.27%	5.97%	3.70%

Table 8.7: Summary of the ECal OOFV systematic study when using the FGD cosmic control sample.

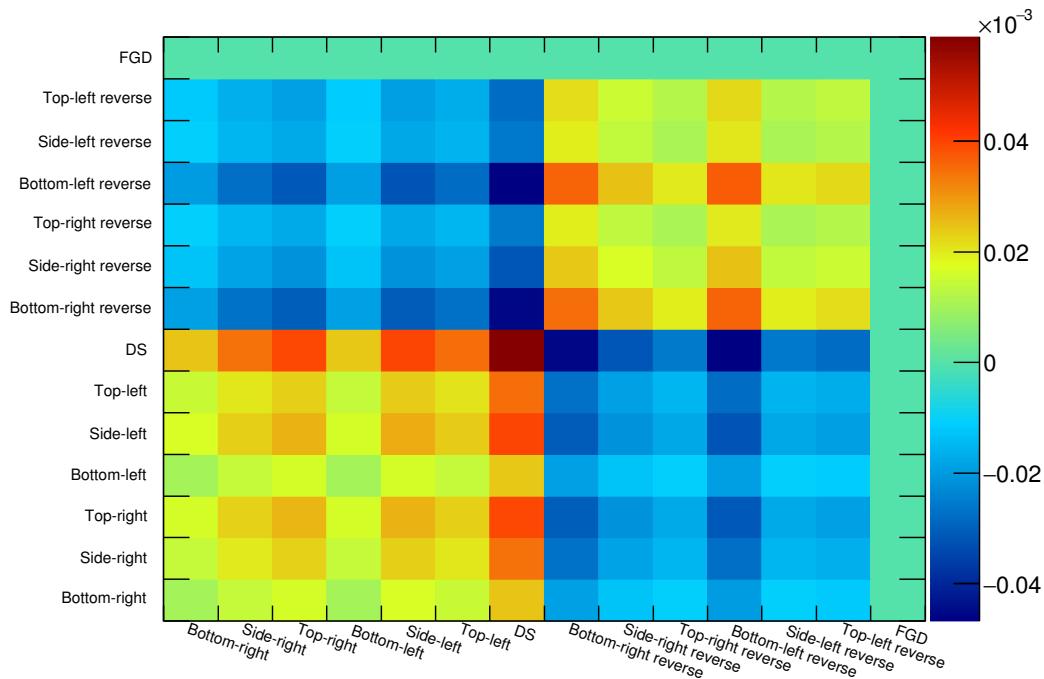
The FGD cosmic control samples were used for this study, which are events with coincident reconstructed tracks in both FGDs and which occur outside of a beam spill window. A summary of the information relevant for systematic uncertainty is located in table 8.7.

To propagate the effect of the systematic uncertainty, a similar approach to the flux systematic uncertainty propagation. However, as the OOFV systematic uncertainty is a pair of single numbers rather than a covariance matrix, only two event weights (one for the barrel ECal and one for the DS ECal) were generated per systematic throw. The event weights were used to reweight only the OOFV events in the sample and the subsequent sample covariance matrices were constructed, which are shown in Fig. 8.25. The largest uncertainty is in the bottom-left reverse sample and is 1.82%.

As described in section 5.4, a significant data/MC difference was found in the DS ECal reconstructed events when looking at the through-going muon control sample. The 2D reconstructed track quality check required that tracks contained no layer gaps. There was an unforeseen hit inefficiency in the DS ECal which was not modelled in MC, which would cause tracks to be rejected because of the 2D track quality check (see section 5.4.2). Because this problem was found after significant progress had been made on the MC selection, it was decided that this bug would be treated as a systematic uncertainty in the analysis. To assess this ad hoc systematic uncertainty, the quality check was relaxed to allow 2D tracks to skip one layer in a given view. The reconstruction and selection was then applied to the sample and was compared to the nominal sample to construct the covariance matrices, which are shown in Fig. 8.26.

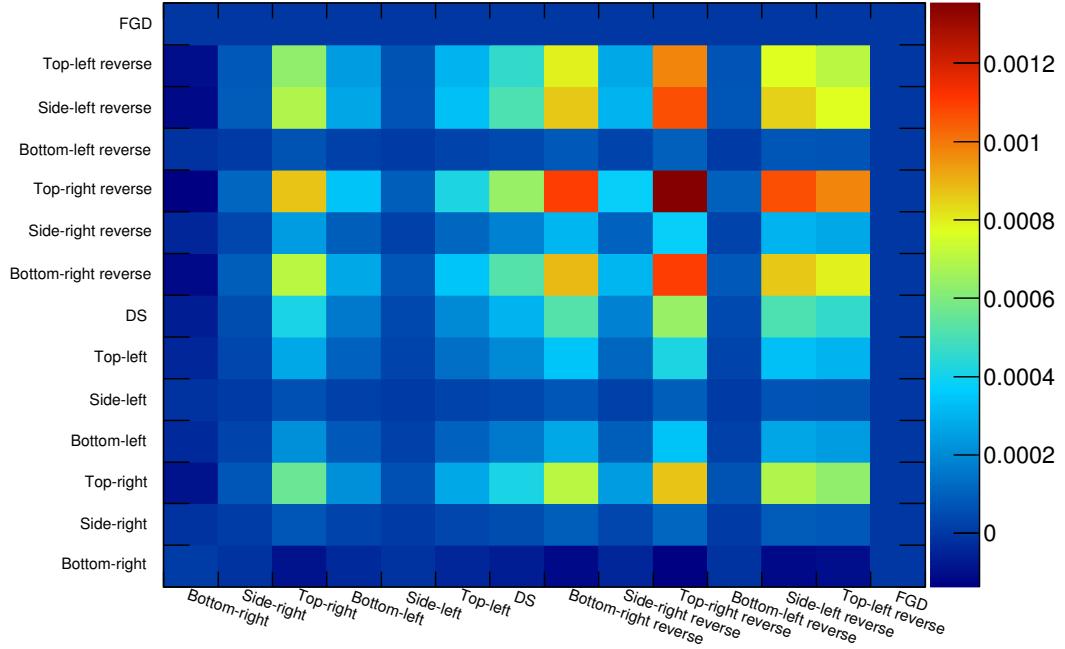


(a) Shape+normalisation.

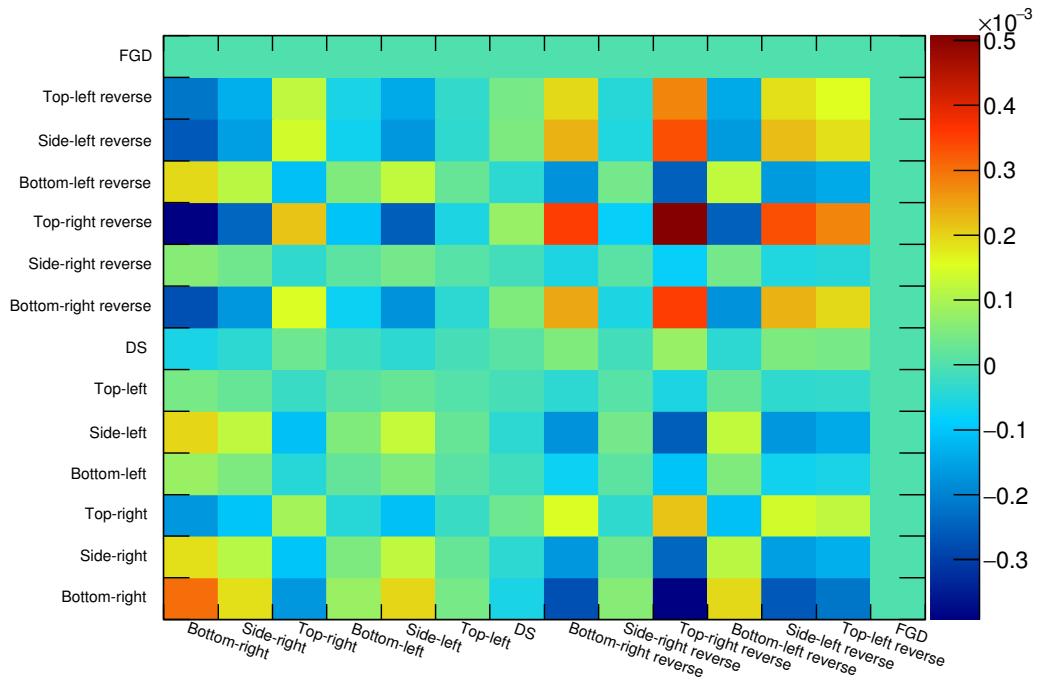


(b) Shape-only.

Figure 8.25: The sample fractional covariance induced by the uncertainty in the ECal OOFV events.



(a) Shape+normalisation.



(b) Shape-only.

Figure 8.26: The sample fractional covariance induced by the bug fix to the reconstruction.

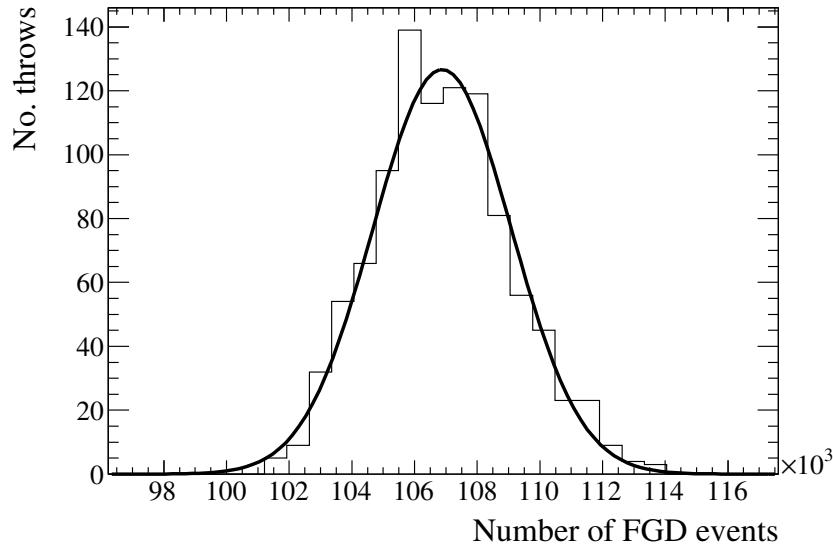


Figure 8.27: The number of selected FGD events for each systematic throw using PSYCHE. The smooth line is a Gaussian fit to the number of events.

8.4.4 FGD detector uncertainty evaluation

The systematic uncertainties associated with the FGD are described extensively in [85]. In recent times, there has been a significant push to develop a framework in which to analyse and propagate systematic uncertainties associated with ND280 analyses, which is called the PSYCHE framework. Even though outer detector analyses have not been incorporated into PSYCHE, this is not true for Tracker analyses and so PSYCHE could be partly utilised in this analysis. The design of PSYCHE allows the user to enable/disable relevant systematic uncertainties and then simultaneously propagate them through an analysis. PSYCHE was set up to apply the FGD detector systematics and produce 1000 throws, in which every systematic uncertainty was simultaneously applied. The output, in the form of a distribution of a varied number of selected CC events in the FGD, was then used to define the systematic uncertainty. This distribution is shown in Fig. 8.27. The Gaussian fit to the thrown number of events implies that the number of FGD events is $(1.069 \pm 0.022) \times 10^5$ events i.e. a 2.01% uncertainty. However, the number of nominal selected events is 1.054×10^5 and so the shift in the number of events suggests an additional 1.41% error. Combining these in quadrature, the FGD detector systematic uncertainty was found to be 2.51%. This error purely relates to the number of CC events in the FGD and so does not correlate with any of the events in the ECal.

The total detector covariance matrices are shown in Fig. 8.28. These matrices were found by summing all of the detector-based covariance matrices. The exception to this is the FGD detector uncertainty in which the single variance was added to the FGD on-diagonal bin. The largest uncertainty was found to be 5.04%. The total covariance matrices, combining all of the covariance matrices discussed above, are shown in Fig. 8.29. The largest uncertainty is 13.8%.

The selection presented in chapter 7 heavily used Monte Carlo simulation of neutrino interactions to estimate the efficiency of the selection. At the core of the simulation, NEUT was used to initially generate the simulated neutrino interactions. Ideally, the selection efficiency should be a property of the selection itself. However, because of the use of a neutrino generator in the simulation, it is possible that the selection could contain an element of generator dependency. It is non-trivial to use a data control sample to assess a systematic uncertainty on the selection efficiency. In fact, the only feasible way to address this analysis issue would be to use multiple neutrino generators as inputs to the same ND280 simulation and compare the differences. Unfortunately, a large set of generators were not available for use in the time frame of the analysis. In ND280 software productions, a large amount of neutrino events based on the GENIE neutrino generator [51] are produced, separately to the NEUT events. So 2.5×10^{19} POT worth of GENIE events were passed through the ND280 software chain (including the enhanced reconstruction) and the selection presented in chapter 7. No GENIE-based sand MC has ever been produced in ND280 software productions. However, as this check only concerns changes to the selection efficiency, an absence of sand MC should not be a problem.

The selected events distributions, separated into the prong topologies, are shown in Fig. 8.30. The selection efficiencies for each prong topology are shown in table 8.8. As described in section 7.6.2, the selection efficiencies do not include neutrino events which were not reconstructed. So, the topology combined efficiencies which do include said events are shown in table 8.9.

Overall, there are differences in the selection efficiencies between NEUT and GENIE, particularly for the DS ECal. However, when combining the topologies and including the events below reconstruction threshold, the efficiency differences are

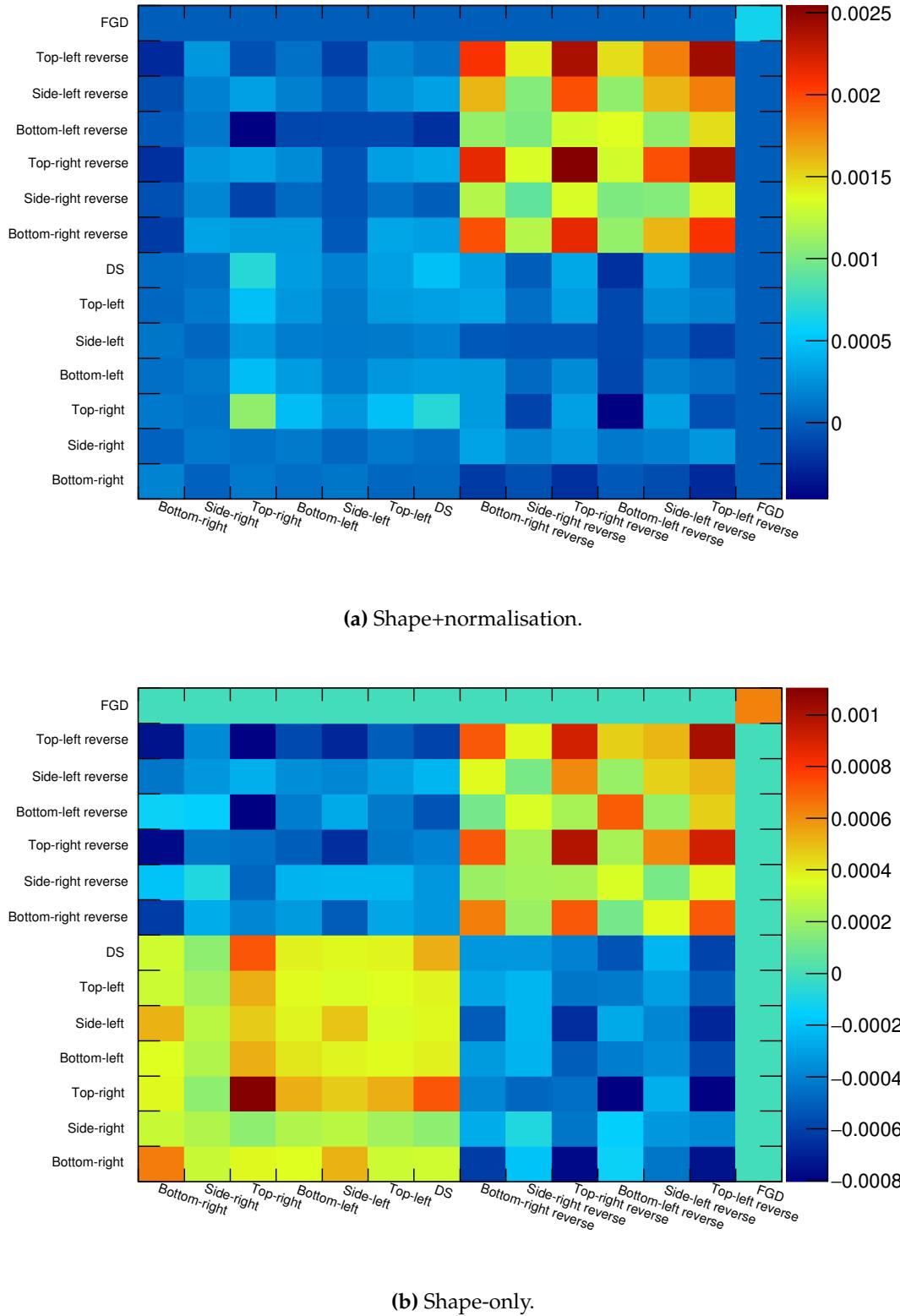
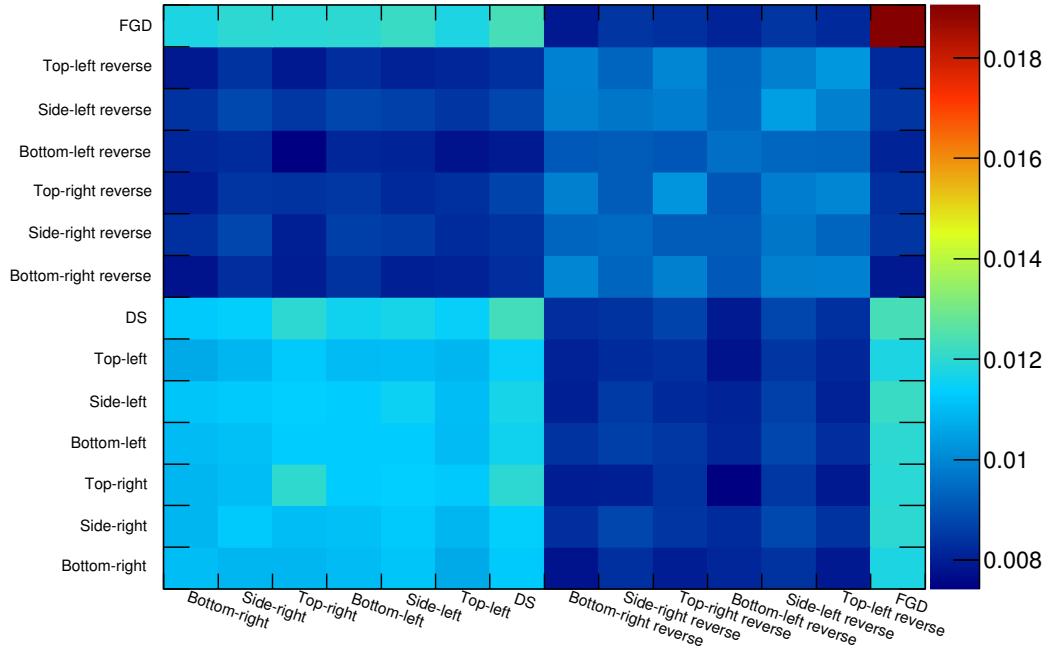
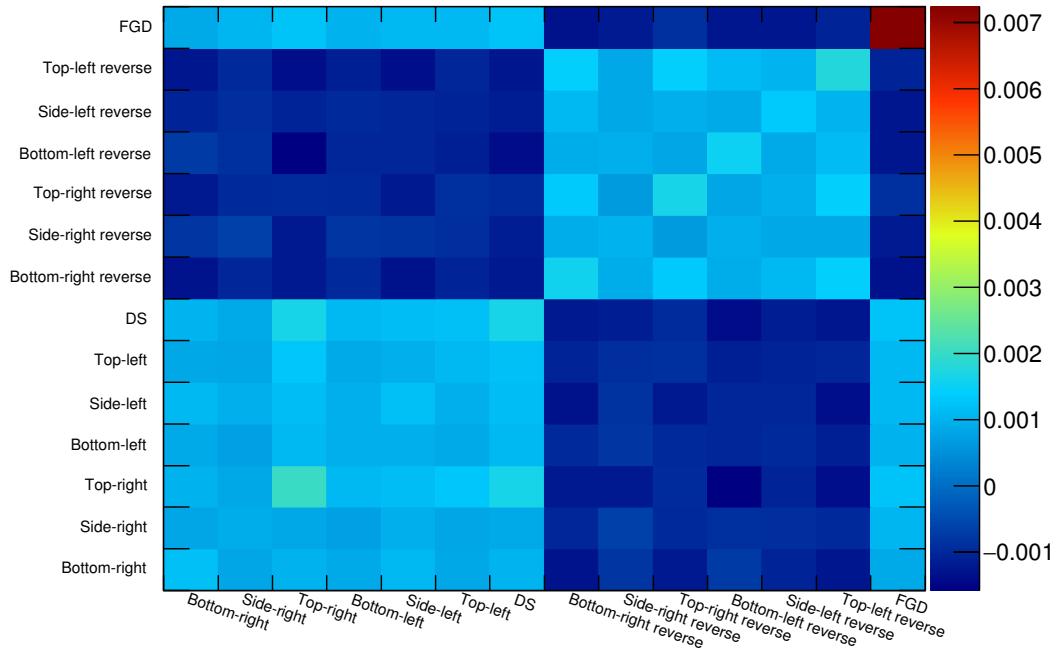


Figure 8.28: The sample fractional covariance induced by all of the detector-based uncertainties.



(a) Shape+normalisation.



(b) Shape-only.

Figure 8.29: The sample fractional covariance induced by all of the systematic uncertainties.

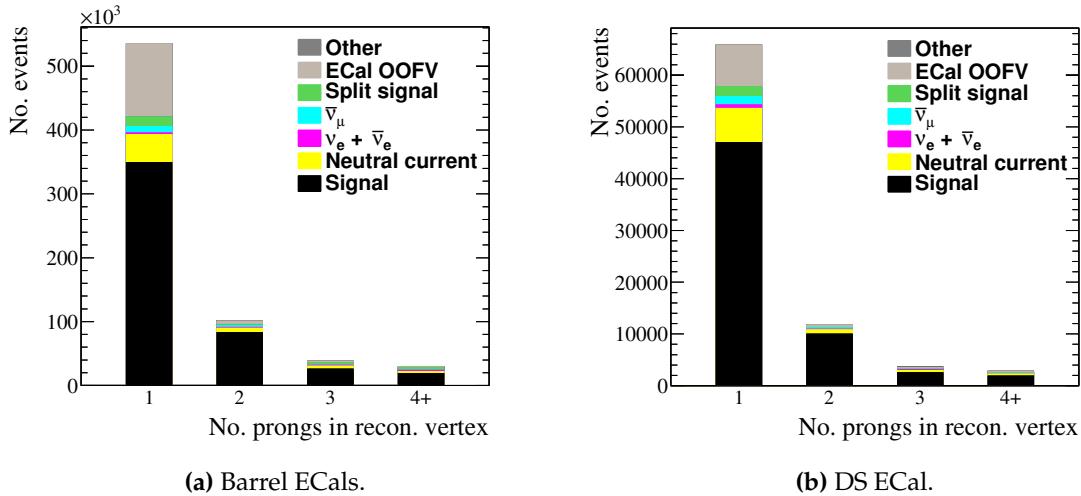


Figure 8.30: The number of selected events in the GENIE-based Monte Carlo sample, separated out into the prong topologies. Each event is categorised by the associated truth information from the simulation.

ECal module	1 prong topology efficiency (%)	2 prong topology efficiency (%)	3 prong topology efficiency (%)	4+ prong topology efficiency (%)
Barrel	46.3	54.2	66.0	65.4
DS	55.9	58.3	59.5	59.9

Table 8.8: The selection efficiencies from the GENIE-based Monte Carlo sample for each prong topology and ECal module.

ECal module	Efficiency (%)
Barrel	40.7
DS	49.9

Table 8.9: The topology combined efficiency for the GENIE-based Monte Carlo sample. The efficiency values include simulated events which were not reconstructed.

somewhat minor. As described at the start of this chapter, the DS ECal will form the primary target and so it is the DS ECal efficiency which is the most important. The DS ECal efficiency uncertainty is simply defined as the difference between the NEUT and GENIE efficiencies, which is $53.0\% - 49.9\% = 3.1\%$. This error as a fraction of the nominal uncertainty is 5.85%.

8.4.5 The measured cross-section uncertainty

Using all of the information described in chapter 7 and chapter 8, the ν_μ CC-inclusive cross-section on lead will be calculated as

$$\sigma_{R_{\text{Pb}}}^{\text{CC}} = \frac{R_{\text{Pb}} N_{\text{Pb}}^{\text{MC}} \eta_{\text{Pb}}}{T_{\text{Pb}} \Phi^{\text{MC}} \epsilon_{\text{Pb}}}, \quad (8.15)$$

where R_{Pb} is the Pb normalisation returned from the rate fit, $N_{\text{Pb}}^{\text{MC}}$ is the number of selected events in the DS ECal, T_{Pb} is the number of Pb nuclei present in the DS ECal, Φ^{MC} is the total flux used to create the MC and η_{Pb} and ϵ_{Pb} are the purity and efficiency of the DS ECal selection. The shape covariance matrix (shown in Fig. 8.29b) will be used as an input to the fit. This covariance matrix includes uncertainties on the flux and (partly) on the number of Pb nuclei. Therefore, the error on R_{Pb} will provide the error from the shape systematic uncertainties which means it is only necessary to include a couple of extra uncertainties on the actual measurement of the cross-section. It is necessary to include the normalisation uncertainty on the number of selected DS ECal events. This value, taken from Fig. 8.29a is 11.1%. The error on the efficiency, as discussed above, is 5.85%. Despite having investigated systematic differences in the ECal mass, the simulation does not include antimony in the absorber sheets. The absorber sheets are quoted as having a 0.2% uncertainty on the antimony doping [60] which also needs to be included in the final uncertainty. Combining all of this information, the final uncertainty on the measured cross-section will be

$$\sigma_{R_{\text{Pb}}}^{\text{CC}} = \sigma_{R_{\text{Pb}}} \pm 12.5\%, \quad (8.16)$$

where $\sigma_{R_{\text{Pb}}}$ is the error on R_{Pb} returned from the rate fit.

Normalisation set	R^{Pb}	R^C	R^{Other}
1	1.0	1.0	1.0
2	3.0	1.0	1.0
3	0.7	1.5	0.8
4	1.0	0	1.0
5	-0.3	1.0	1.0

Table 8.10: The normalisation sets used in the machinery validation of the ECal rate fit.

8.5 Validation of method

The ECal rate fit described in section 8.3 requires validation to ensure that not only the machinery works as intended, but also to check that the fit is suitable for such an analysis. The first set of validation checks performed deals with the former case.

8.5.1 Validation of the fit machinery

The initial validation method used the same Monte Carlo described in section 8.2 to create toy datasets which were passed through the fitter. To create said datasets, new normalisation parameters, R^{Val} , were defined which are shown in table 8.10. These new normalisation parameters scale the relevant Monte Carlo templates.

For the MC population X in each sub-detector sample i , the normalisation parameter is applied to create the new population as

$$n_i^{X \text{ Val}} = R_i^{X \text{ Val}} n_i^{X \text{ Nom}}, \quad (8.17)$$

where $n_i^{X \text{ Nom}}$ is the number of events in population X of sub-detector sample i for the nominal MC. The total number of events in each sub-detector sample is then constructed as

$$N_i^{\text{Val}} = \sum_{i=1}^X n_i^{X \text{ Val}}. \quad (8.18)$$

The final step in creation of the toy datasets was to vary the total number of events in the sub-detector samples. The method chosen was the same as that used to generate

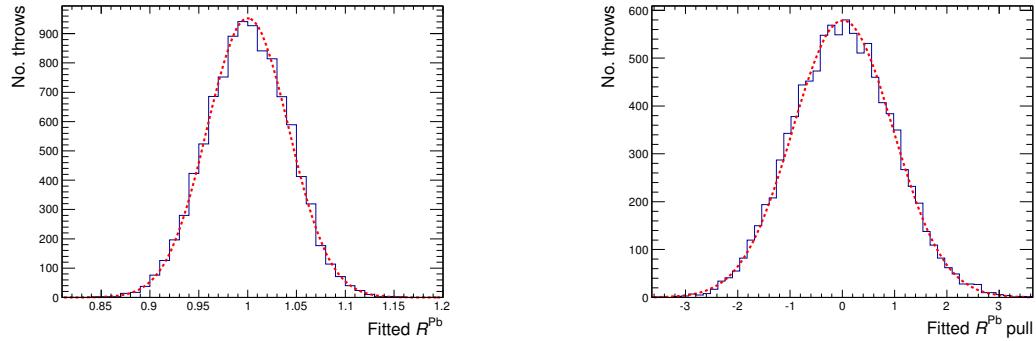
event weights for the systematic uncertainty studies described in section 8.4. The covariance matrix shown in Fig. 8.29b was Cholesky-decomposed and multiplied by a vector of random numbers, all drawn from a Gaussian of mean 0 and width 1. The sub-detector sample weights were then created by adding 1 to each element of the multiplied vector. Each scaled sub-detector sample, N_i^{Val} , was then reweighted by the relevant sample weight. This toy dataset was then passed to the fitter and its output recorded. The steps outlined above describe one pass of the fitter. This process was repeated 10,000 times for each validation normalisation set.

The value of R^{Pb} returned for each throw in parameter set 1 is shown in Fig. 8.31a. The Gaussian fit quantifies the performance of the fit: the mean specifies the final value and the width specifies a pseudo-error. To further quantify how well the machinery works, the parameter pull, P_i^X , is defined as

$$P_i^X = \frac{R_i^X \text{Fit} - R_i^X \text{Val}}{\sigma_i^X \text{Fit}}, \quad (8.19)$$

where $R_i^X \text{Fit}$ and $\sigma_i^X \text{Fit}$ are the value and error of parameter X returned from the fit for throw i respectively. The MINOS routine used in the fitter returns asymmetric errors and so care must be taken when deciding which error to use in equation 8.19. It was decided that the positive error would be used when the fitter returned a value higher than the input. P^{Pb} for parameter set 1 is shown in Fig. 8.31b. The Gaussian fit to this distribution has a width close to, but slightly below, 1. This suggests that the errors returned by the MINOS routine very slightly overestimate the error, but are still a very good representation of the uncertainty. The performance of the fit for all of the normalisations in normalisation set 1 are shown in Fig. 8.32. In each case, the fitter returns the correct normalisation and the pull characteristics suggest the fit returns the correct error and is unbiased. The performance of the fit for normalisation sets 2, 3, 4 and 5 are shown in Fig. 8.33, Fig. 8.34, Fig. 8.35 and Fig. 8.36 respectively. Generally speaking, the fitter performs well for all of the situations. It is clear that the output is sensible even when extreme situations are presented (normalisation set 2) as well as when minor shifts are applied to all three parameters (normalisation set 3). Even in un-physical situations the fitter returns the input correctly (normalisation set 5). It is only when a template is removed (normalisation set 4) that the fitter becomes biased. However, as Fig. 8.35b shows, the bias is minor.

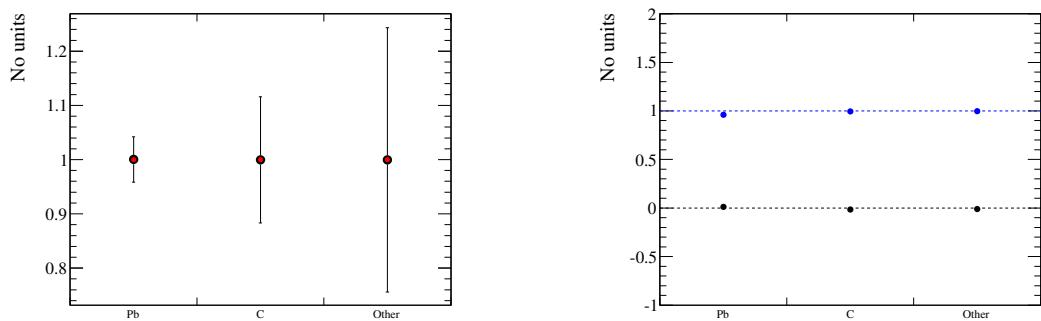
As discussed in section 8.2, a set of barrel ECal samples which contain events that



(a) The returned parameter for each throw. The fitted Gaussian has a mean of 1.00 and a width of 0.04.

(b) The parameter pull for each throw. The fitted Gaussian has a mean of 0.01 and a width of 0.96.

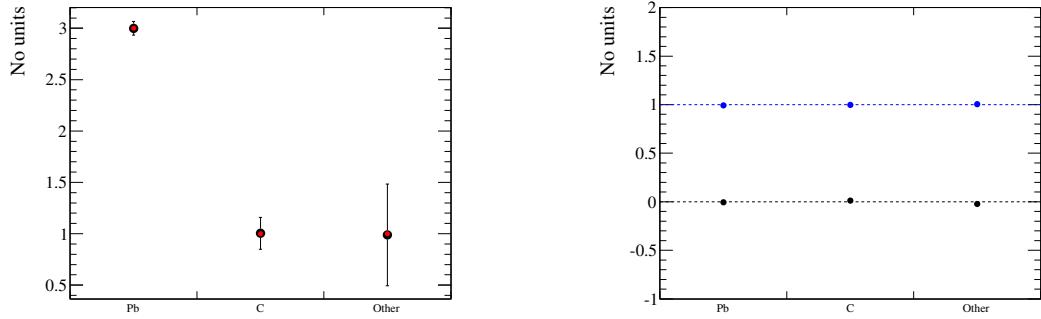
Figure 8.31: Distributions of R^{Pb} related information returned from the fitter for the 10,000 fake data throws. The red dashed lines are Gaussian fits to the distributions. The distributions show that the fitter is correctly returning the value of R^{Pb} and is unbiased.



(a) The mean fitted values for each normalisation parameter (black) compared with the input normalisation values (red).

(b) The means (black) and widths (blue) of the pull distributions after fitting with a Gaussian.

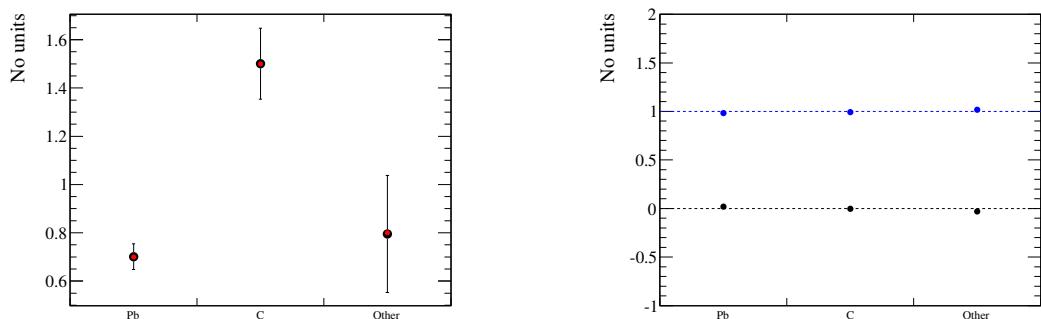
Figure 8.32: Performance of the fit for the 10,000 fake data sets when using normalisation set 1. For all parameters, the fitter returns the correct values and is unbiased.



(a) The mean fitted values for each normalisation parameter (black) compared with the input normalisation values (red).

(b) The means (black) and widths (blue) of the pull distributions after fitting with a Gaussian.

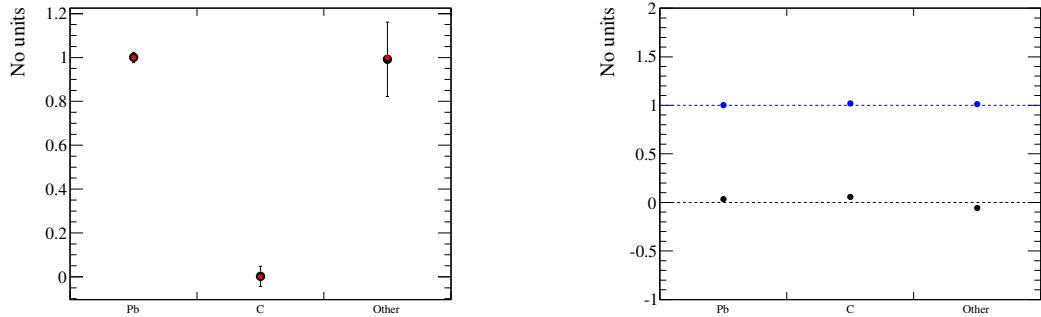
Figure 8.33: Performance of the fit for the 10,000 fake data sets when using normalisation set 2. For all parameters, the fitter returns the correct values and is unbiased.



(a) The mean fitted values for each normalisation parameter (black) compared with the input normalisation values (red).

(b) The means (black) and widths (blue) of the pull distributions after fitting with a Gaussian.

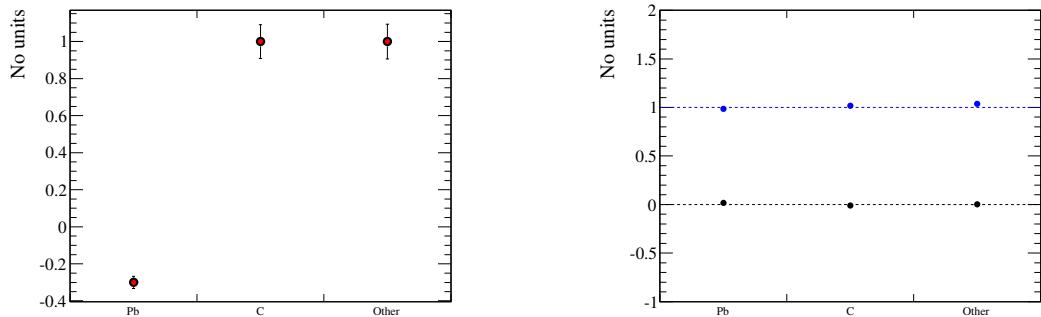
Figure 8.34: Performance of the fit for the 10,000 fake data sets when using normalisation set 3. For all parameters, the fitter returns the correct values and is unbiased.



(a) The mean fitted values for each normalisation parameter (black) compared with the input normalisation values (red).

(b) The means (black) and widths (blue) of the pull distributions after fitting with a Gaussian.

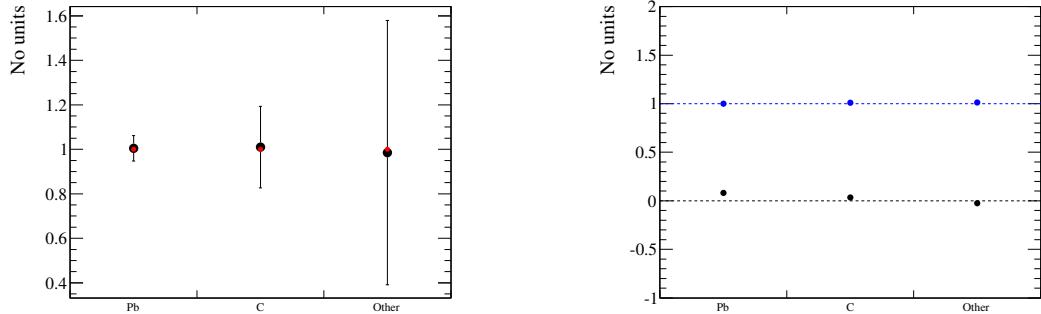
Figure 8.35: Performance of the fit for the 10,000 fake data sets when using normalisation set 4. Despite correct errors being returned, the fitter is returning slightly biased values of the fit parameters. It is important to note that normalisation set 4 completely removes the carbon template and even in this extreme situation, the bias on the fitted parameters is small.



(a) The mean fitted values for each normalisation parameter (black) compared with the input normalisation values (red).

(b) The means (black) and widths (blue) of the pull distributions after fitting with a Gaussian.

Figure 8.36: Performance of the fit for the 10,000 fake data sets when using normalisation set 5.



- (a) The mean fitted values for each normalisation parameter (black) compared with the input normalisation values (red).
- (b) The means (black) and widths (blue) of the pull distributions after fitting with a Gaussian.

Figure 8.37: Performance of the fit with the reverse ECal samples removed for the 10,000 fake data sets when using normalisation set 1.

failed the selection cuts are included. To test that the inclusion of this sample is not degrading the performance, an extra set of 10,000 fake data throws were performed for normalisation set 1. However, the reverse ECal samples were not included in the fitter. The results of this test are shown in Fig. 8.37. Comparing this with the performance of the fitter with the reverse sample included (Fig. 8.32), it is clear that the absence of the reverse sample causes a wider range of normalisation values to be returned from the fitter. Additionally, a small bias is introduced to all three normalisations. While the issues present in the reverse-absent fitter are minor, It should be clear that the inclusion of the reverse ECal samples leads to more accurate and precise results.

8.5.2 Physics validation of the fitter

If the fitter truly works as intended, then it should be able to successfully fit one cross-section model to another. As described in section 8.4, T2K produces a set of GENIE-based ND280 events during software production in addition to the NEUT-based events. As the NEUT events are being used for the MC templates in the fitter, the GENIE samples make an excellent candidate for a fake data test.

The same GENIE sample used in section 8.4 was passed through the selection and binned into the ECal samples. Unfortunately, no GENIE-based sand Monte Carlo or FGD sample were available for this study. So, the same NEUT-based sand MC and

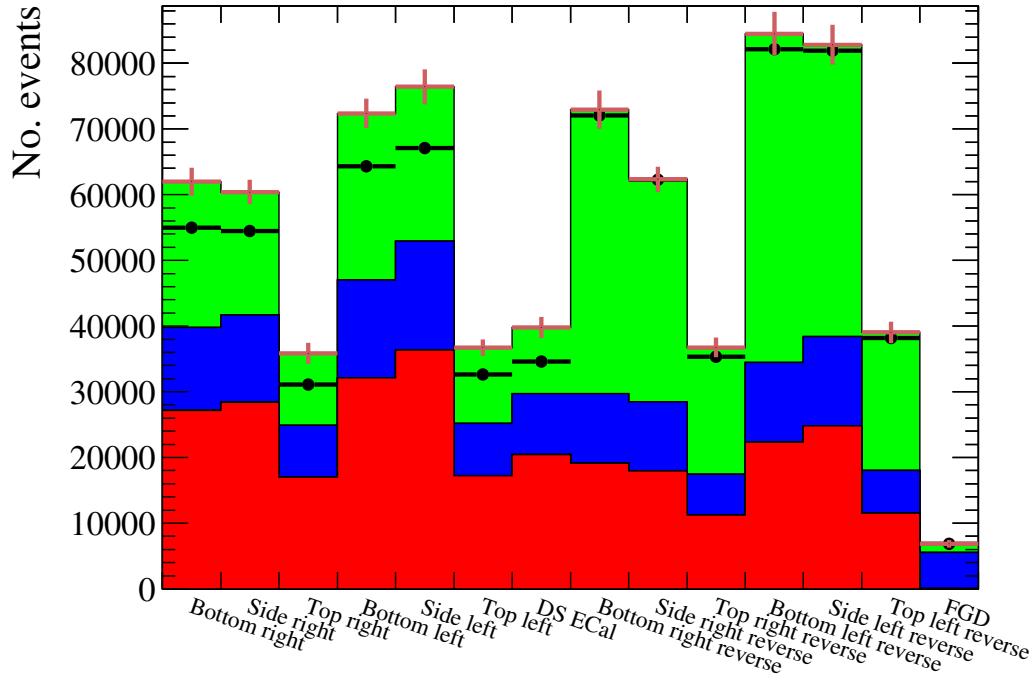


Figure 8.38: The pre-fit number of NEUT Monte Carlo events compared with GENIE fake data. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the GENIE fake data. In all of the ECal selected bins (the first 7 bins) there is a systematic deficit of GENIE events. This effect is not seen in the reverse samples, because the reverse samples contain a significant fraction of sand muons which are taken from the NEUT Monte Carlo.

FGD sample used in the MC template generation were used to fill this gap. The MC templates were normalised to the same POT as the fake data. The pre-fit samples are shown in Fig. 8.38.

Before the NEUT MC templates were actually fit to the genie fake data, it was necessary to calculate what the fit should be predicting. To calculate what the true GENIE to NEUT cross-section ratios are, the truth information provided by the MC simulation was used. For both the GENIE and NEUT samples, the number of true interactions in the DS ECal were recorded, including the target element. These numbers were weighted to the same POT and the ratio taken. These values are shown in table 8.11.

$\frac{\sigma_{\text{GENIE}}^{\text{Pb}}}{\sigma_{\text{NEUT}}^{\text{Pb}}}$	$\frac{\sigma_{\text{GENIE}}^{\text{C}}}{\sigma_{\text{NEUT}}^{\text{C}}}$
0.908	0.926

Table 8.11: The GENIE to NEUT cross-section ratios, calculated from the truth information provided by the ND280 simulation.

The MC templates and fake data shown in Fig. 8.38 were then passed to the fitter, with the fitted normalisations shown in table 8.12 and a comparison of the post-fit MC templates with the GENIE fake data shown in Fig. 8.39. The reduced χ^2 of the fit was found to be 1.71.

The values returned from the fit do not agree with the true values when considering only the errors returned from the fit. If the extra 12.5% uncertainty defined in equation 8.16 is added in quadrature with the positive error on R^{Pb} , the fitted normalisation is in agreement with its corresponding true value. It should be noted at this point that both the sand muon component and the FGD sample in the fake data are in fact NEUT, rather than GENIE. So, this test is somewhat unrealistic. While it is not trivial to produce a GENIE-based sand muon Monte Carlo, the FGD sample in the genie fake data can be re-weighted to better represent what the FGD sample should be. Using the carbon cross-section ratio shown in table 8.11, the carbon component of the FGD sample in the GENIE fake data was scaled. This modified set of fake data was then passed to the fitter, along with the NEUT MC templates. The returned normalisations are shown in table 8.13 and the comparison of the fitted MC templates with the fake data is shown in Fig. 8.40. It is clear that the situation is much improved as both the lead and carbon normalisation have moved towards their true values. The carbon normalisation still disagrees with the simulated value; however, it has already been mentioned that the sand muon component of the GENIE fake data is actually

Normalisation parameter	Fitted value
R^{Pb}	$0.79^{+0.08}_{-0.08}$
R^{C}	$1.19^{+0.15}_{-0.13}$
R^{Other}	$0.94^{+0.08}_{-0.08}$

Table 8.12: The values returned from the fitter when passing the NEUT MC templates and GENIE fake data shown in Fig. 8.38.

Normalisation parameter	Fitted value
R^{Pb}	$0.83^{+0.08}_{-0.08}$
R^C	$1.09^{+0.14}_{-0.12}$
R^{Other}	$0.95^{+0.08}_{-0.09}$

Table 8.13: The values returned from the fitter when passing the NEUT MC templates and GENIE fake data which contained a re-weighted FGD sample.

NEUT. Ideally, a GENIE-based set of sand muon MC events would be used in the fake data but due to time constraints this was not possible. As the lead normalisation agrees with the generator prediction, this test can be considered a success.

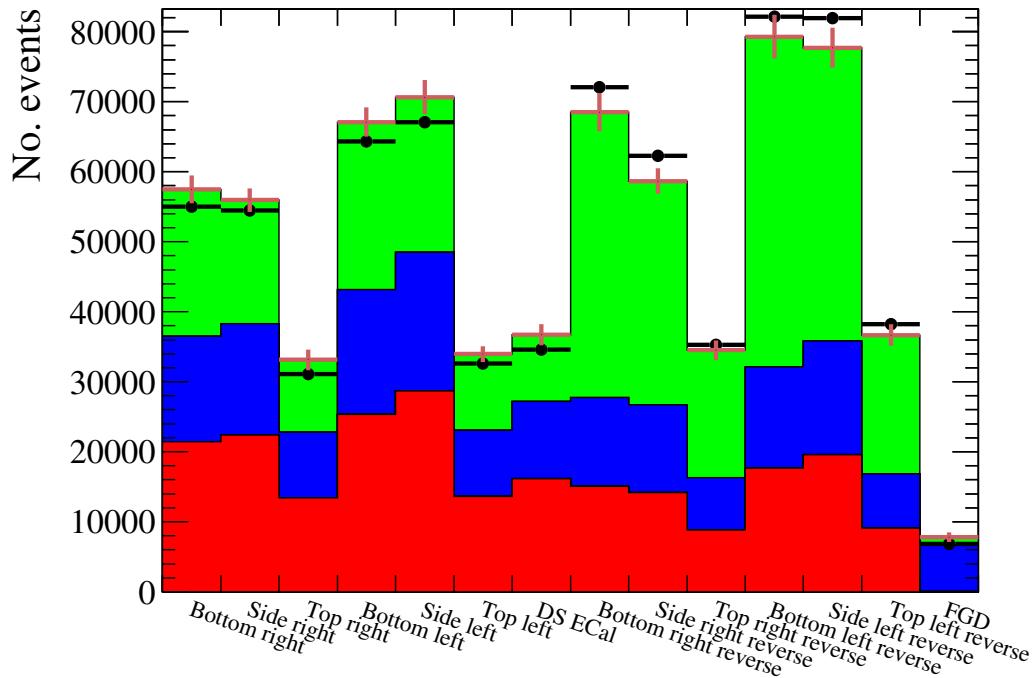


Figure 8.39: The post-fit number of NEUT Monte Carlo events compared with GENIE fake data. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the GENIE fake data. To account for the original deficit of events in the first 7 bins, the fitter raises their normalisation. But, to account for the correlations, the reverse sample bins have to have their normalisations lowered, causing a NEUT MC deficit.

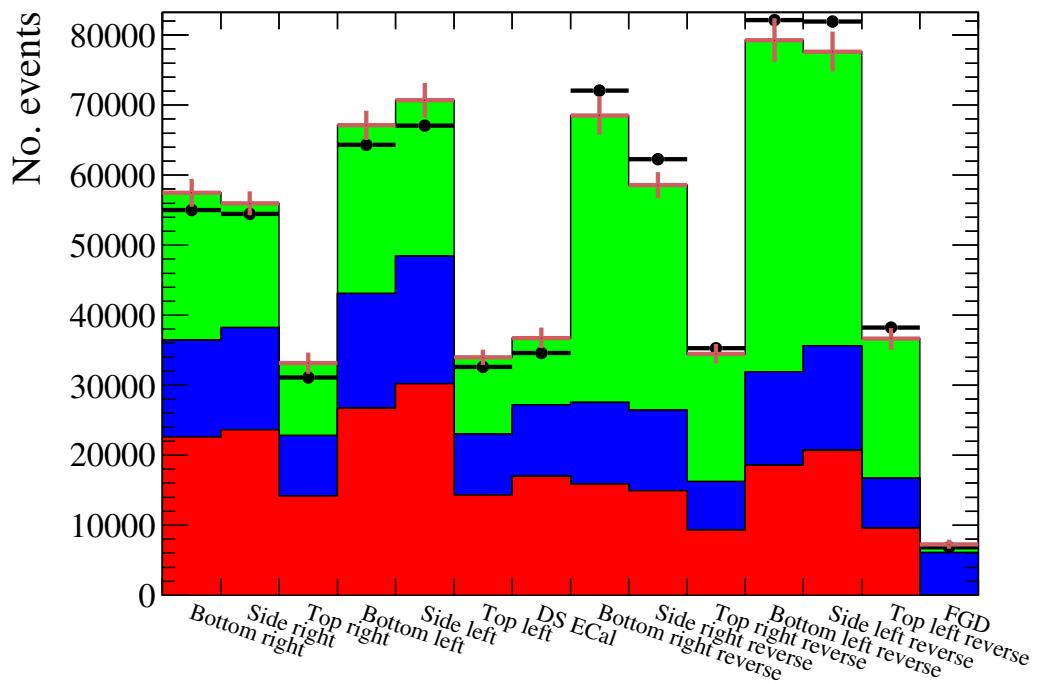


Figure 8.40: The post-fit number of NEUT Monte Carlo events compared with GENIE fake data which has a re-weighted FGD sample. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the GENIE fake data. The post-fit total normalisation of the MC is comparable to that of the original fit shown in Fig. 8.39. However, because the pre-fit MC normalisation of the FGD has been lowered, the fitter was able to make less extreme alterations to the lead and carbon normalisations.

8.6 Applying the fit to ND280 data

To measure the ν_μ CC inclusive cross-section on lead, a sample of ECal data events from the neutrino beam must be passed through the selection presented in chapter 7. However, an extra data quality check must be applied to the data events before the selection takes place. These data quality checks are separated into two parts. The first check ensures that the components of the beam-line were nominally operating when the data was collected. This involves basic hardware checks, that the horn currents were operating within 5 kA of their mean value, that the beam angle was within 1 mrad of its mean value and that the muon rate was within 5% of its mean value. The second check ensures that ND280 itself was operating optimally during data collection. A dedicated group studies the sub-detector outputs on a per spill basis, ensuring that the sub-detectors (including the ECals) were operating correctly. The output of these two checks are a pair of boolean flags which state whether the data that was provided by the beam and collected by ND280 were sound. So, before data events are passed to the selection, these flags are queried and if either flag reports bad data, the event is rejected.

The data quality checks and the event selection described in chapter 7 were applied to 1.31×10^{20} POT of neutrino beam data collected from T2K run 3C. The FGD sample was also taken from run 3C, corresponding to 1.35×10^{20} POT. The number of selected events in each sample is summarised in table 8.14.

The samples were then passed to the fitter described in section 8.3 and the parameters found are shown in table 8.15. The minimum χ^2 per degree of freedom was found to be 13.4. The MC template comparison with the data before and after the χ^2 fit are shown in Fig. 8.41 and Fig. 8.42 respectively. After the normalisation parameters had been returned from the fit, the χ^2/NDOF as a function of the normalisation parameters could be found. This was achieved by varying a single normalisation parameter whilst keeping all other normalisation parameters fixed at their χ^2 minimum value. At each step in the single normalisation parameter variation, the χ^2/NDOF was calculated. These results are shown in Fig. 8.43.

It should be clear from Fig. 8.42 that the Monte Carlo is not a good fit to the data. Before the fitting, it is clear that there are shape differences between the data and

Sample	No. Selected (MC)	No. Selected (data)
Bottom right ECal	55749	54396
Side right ECal	54011	56583
Top right ECal	32006	32233
Bottom left ECal	65169	65448
Side left ECal	68031	66095
Top left ECal	32760	33470
DS ECal	35420	34372
Bottom right ECal (Reverse)	72112	70885
Side right ECal (Reverse)	61159	64445
Top right ECal (Reverse)	35997	36191
Bottom left ECal (Reverse)	83059	80967
Side left ECal (Reverse)	81240	84145
Top left ECal (Reverse)	38282	40293
FGD	5857	5992

Table 8.14: The number of selected events in each sample.

Normalisation parameter	Fitted value
R^{Pb}	$1.58^{+0.10}_{-0.11}$
R^{C}	$1.87^{+0.21}_{-0.18}$
R^{Other}	$0.26^{+0.06}_{-0.05}$

Table 8.15: The values returned from the fitter when processing the NEUT MC templates with T2K data.

Monte Carlo templates. For example, there is a minor data deficit in the bottom-right ECal module and a larger data excess in the side-right ECal. A similar situation occurs between the bottom-left reverse and side-left reverse samples. Fig. 8.44 shows the shape-only correlation matrix for the samples used in the fit. The correlation matrix shows that all of the barrel samples are highly correlated with each other, as are all of the barrel-reverse samples. This suggests that the fit has little freedom to easily resolve the shape differences. This fact is enforced by the pre-fit information shown in Fig. 8.42, where the only way to resolve such shape differences with highly correlated constraints is to dramatically alter the normalisation of the templates. The pre-fit information shown in Fig. 8.41 can help to illuminate the underlying issue. An area-normalised comparison of a sub-set of the MC templates with the run 3C data is shown in Fig. 8.45 for the barrel samples and Fig. 8.46 for the barrel-reverse samples. As

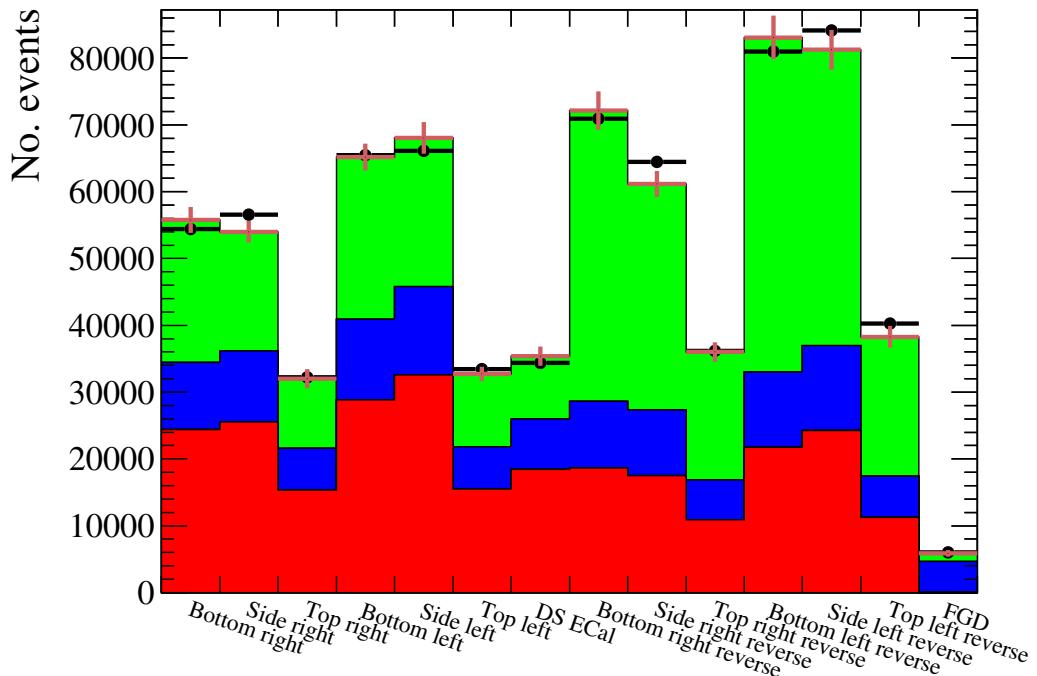


Figure 8.41: The pre-fit number of NEUT Monte Carlo events compared with T2K data. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the T2K data. While there is generally agreement between the MC and data within shape-only errors, there are shape differences between the data and MC, the most notable occur when transitioning between the bottom and side modules (e.g. bottom-right to side-right).

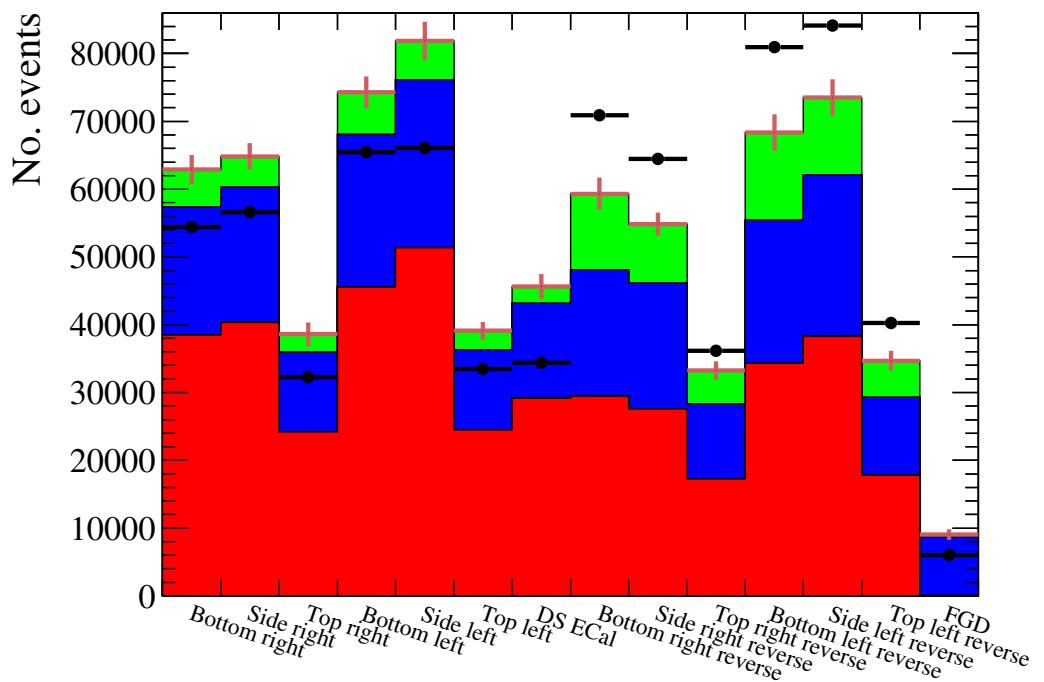


Figure 8.42: The post-fit number of NEUT Monte Carlo events compared with T2K data. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the T2K data. To correct for the shape differences between the data and MC, the fitter has applied extreme variation to the sample normalisations.

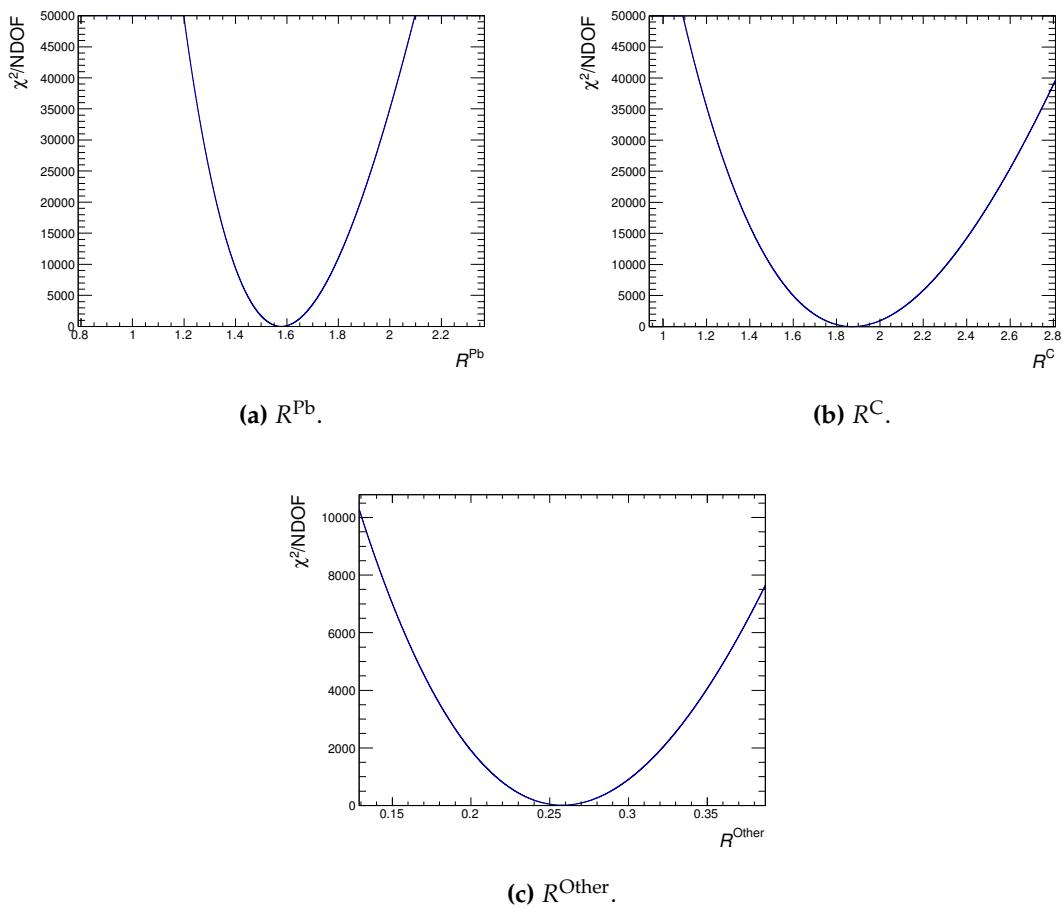


Figure 8.43: The χ^2/NDOF as a function of the normalisation parameters. The functions are all smooth and parabolic with a clear minimum.

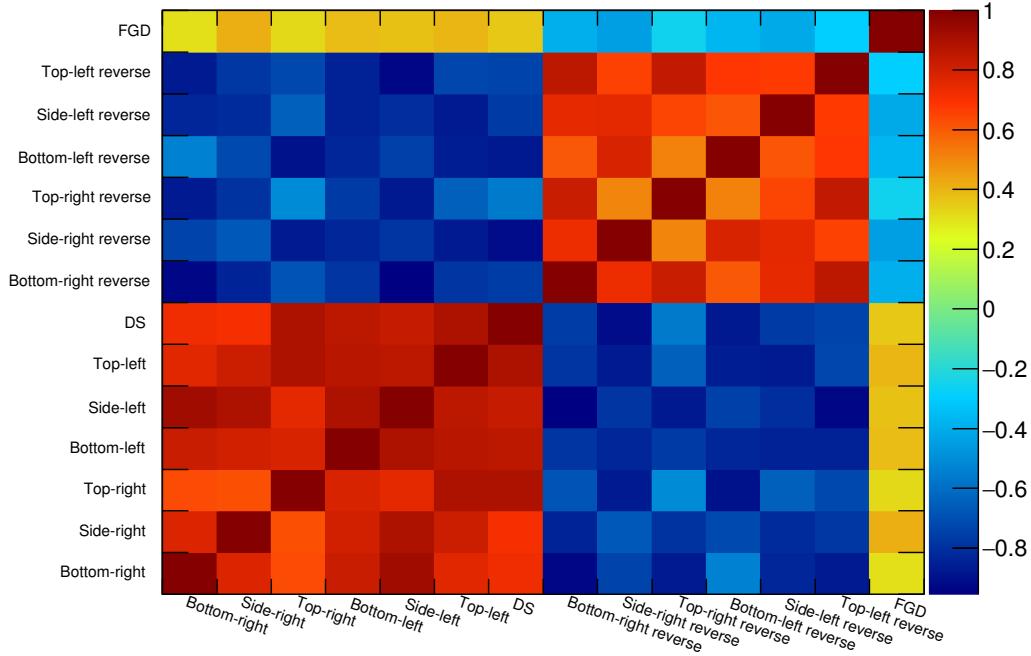


Figure 8.44: The shape-only correlation matrix for the samples used in the fit. There is a large degree of (anti)-correlation between all samples. This gives the fitter less freedom to change the normalisations of the MC samples.

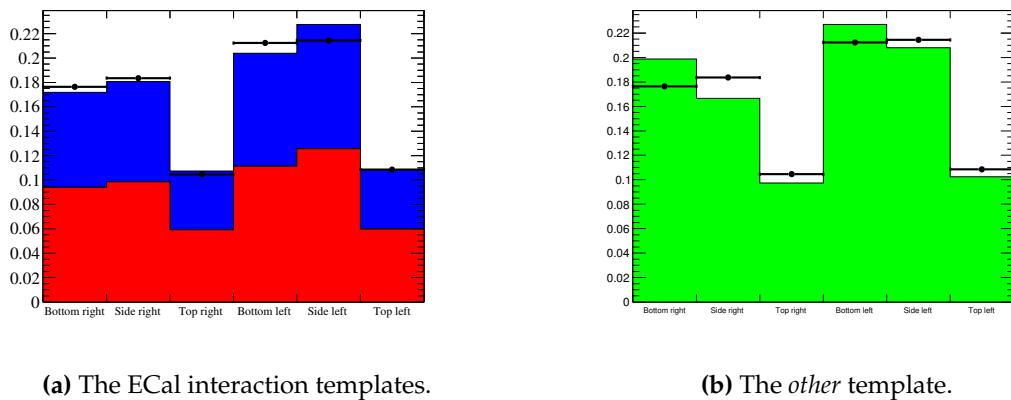


Figure 8.45: An area-normalised comparison of a sub-set of the MC templates with data for the barrel ECal samples. The coloured histograms and data points are the MC Templates and data respectively. The red, blue and green histograms are the lead, carbon and *other* templates respectively. There is significant disagreement in the shape of the *other* template with the data.

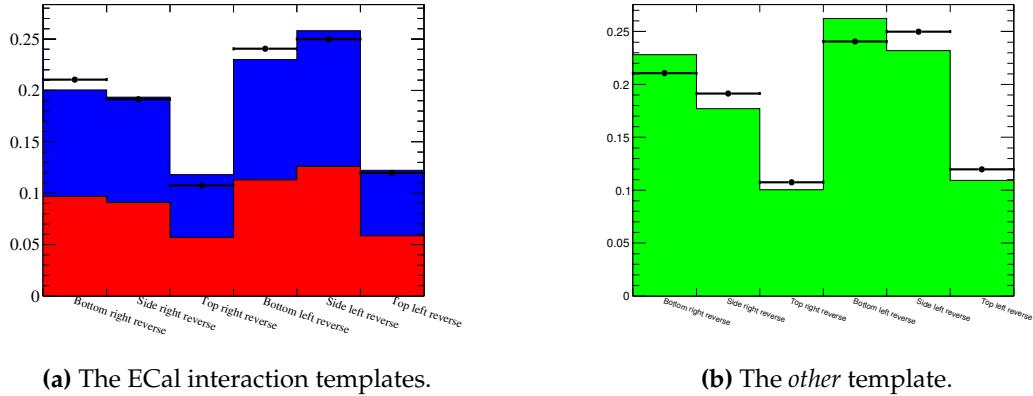


Figure 8.46: An area-normalised comparison of a sub-set of the MC templates with data for the barrel-reverse ECal samples. The coloured histograms and data points are the MC Templates and data respectively. The red, blue and green histograms are the lead, carbon and *other* templates respectively. Similarly to Fig. 8.45, there is disagreement in the shape of the *other* templates with the data.

Fig. 8.45a and Fig. 8.46a show, the shape of the ECal interaction templates in both the barrel and barrel-reverse samples generally follow the data shape. This is not the case for the shape of the *other* template (Fig. 8.45b and Fig. 8.46b). This finding suggests that it is the composition of the *other* template that is the cause of the fit failure. The *other* template contains all events which are not CC-carbon or CC-lead interactions. As was found in chapter 7 and shown in Fig. 7.42a, the dominant background is the ECal OOFV background which is the primary event category which falls into the *other* template. The number of events in the *other* template as a function of true neutrino energy is shown in Fig. 8.47 and Fig. 8.48 for the barrel and barrel-reverse samples respectively where the distributions have been separated into the true neutrino interaction location. In all cases, entering backgrounds which originate from the magnet and sand are the dominant contributor. These events sit in a region far away from the ND280 Tracker which is the only place where the neutrino flux has been studied in great detail.

Comparisons of data and MC can help to illuminate what the underlying cause is. Fig. 8.49 shows the number of events which pass the fiducial volume cut as a function of the number of hits in the reconstructed ECal cluster for both data and MC. Firstly, it is important to state why this particular set of events are being shown; the definition of the reverse selection is any event which passes the fiducial volume cut i.e. all events in Fig. 8.49 have to enter the ECal rate fit in one of the samples (selected or reverse).

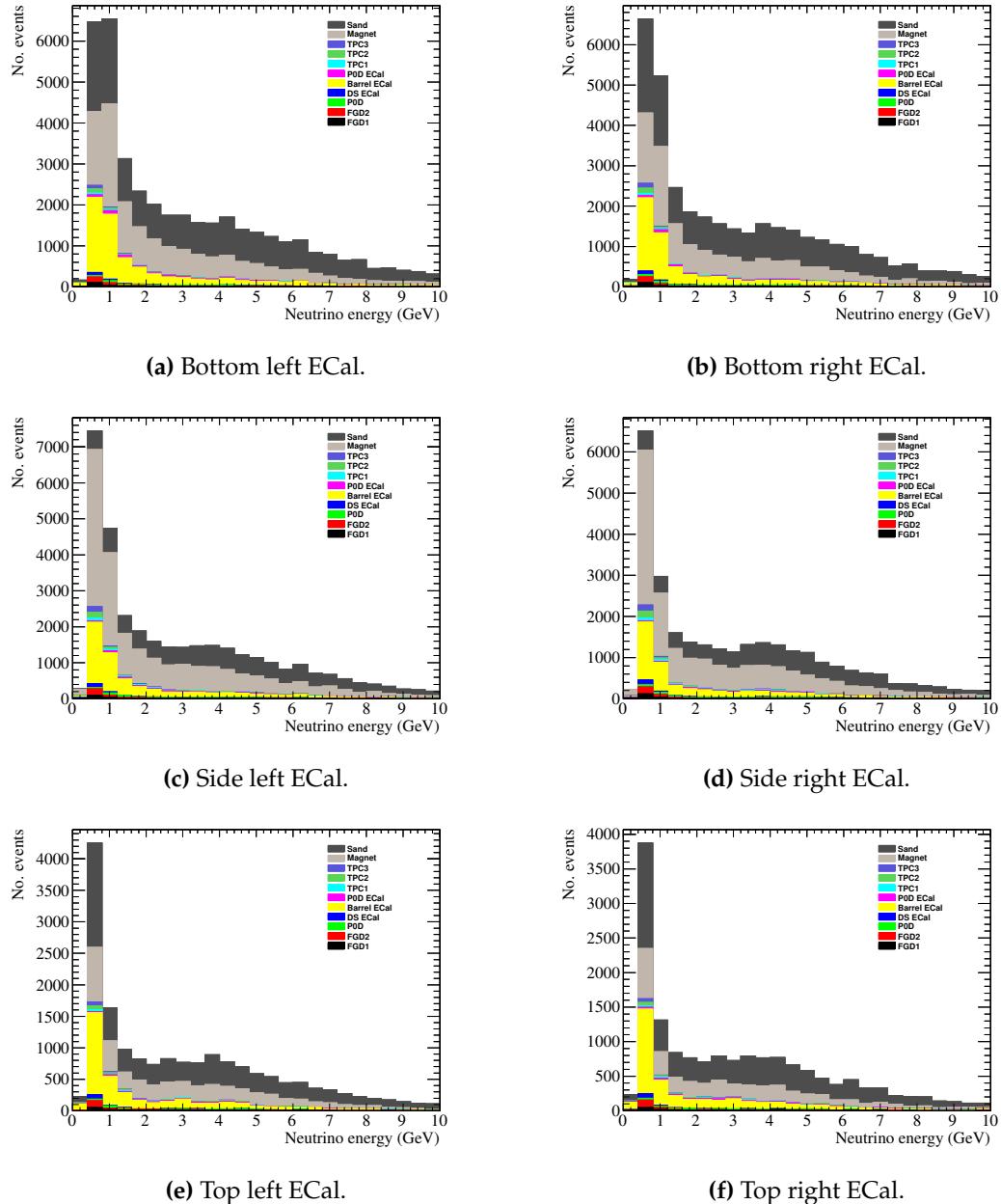


Figure 8.47: The number of events in the *other* template as a function of true neutrino energy for the barrel ECal samples. The distributions are broken down by true neutrino interaction location. For all distributions, the dominating location is either the magnet or the sand in the surrounding pit.

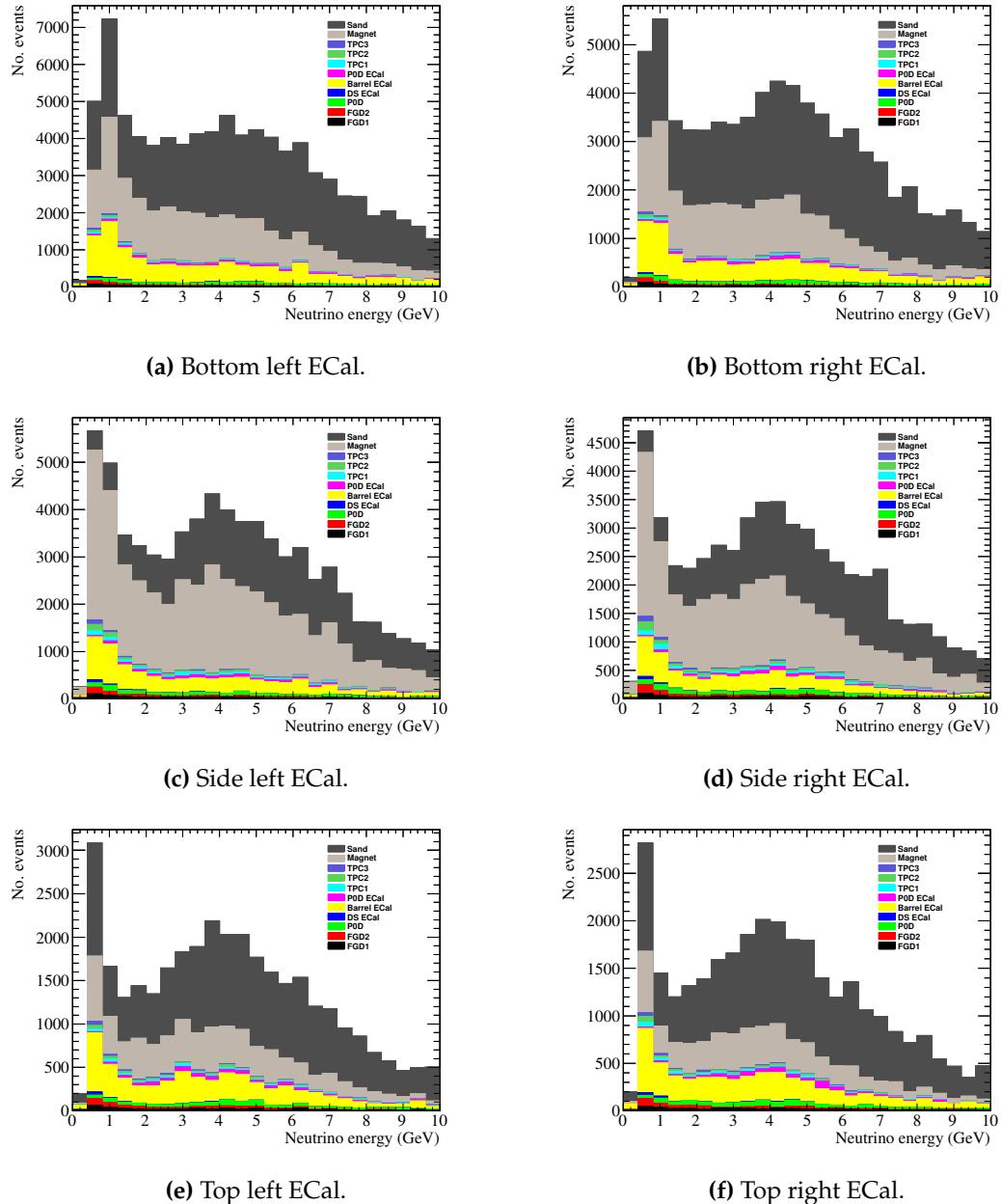


Figure 8.48: The number of events in the *other* template as a function of true neutrino energy for the barrel reverse ECal samples. The distributions are broken down by true neutrino interaction location. For all distributions, the dominating location is either the magnet or the sand in the surrounding pit.

What is particularly interesting about these distributions is that there appears to be two significant areas of discrepancy between data and MC. The first discrepancy is an excess of data events with a low number of hits in the reconstructed clusters and is common to all barrel-ECal modules. A hand-scan of these events in the event display showed well reconstructed, 1-prong like events in each ECal module which appear to aim forwards, towards the magnet region. While an in depth study of these events is needed, it is possible that a mismodelling of the ν_μ cross-section on iron has caused such events to not appear in the MC. The second significant area of discrepancy, also shown in Fig. 8.49, is where there the number of hits in the reconstructed clusters is between 30 and 40. Unlike the first discrepancy, this problem is only an issue for all left ECals and the bottom-right ECal. What is notable about this discrepancy is that it sits in a region where the sand MC reaches its maximum. It has already been discussed that the sand MC is not well modelled and so it is not surprising that a discrepancy between data and MC would occur here. It is, however, troubling that these discrepancies exist after the fiducial volume cut has been applied which means that this undefined behaviour enters the analysis in some way. The efficiency of these events under the cut streams seem to have a module dependence and, as they have to appear in the analysis somewhere, this causes a non-trivial migration of such events between the selected samples and the reverse samples.

Under normal circumstances, the background modelling needs extra study. However, due to time constraints, it is sufficient to introduce an ad hoc uncertainty which encompasses the lack of background understanding. The background modelling has already been discussed as a possible cause of the fit failure and the barrel-reverse samples, which contain the largest background, can be used to quantify this lack of understanding. Table 8.16 shows the ratio of the number of events in each barrel sample to the number of events in each barrel-reverse sample for both data and MC. Also included in table 8.16 is the percentage difference between the data and MC ratio and it should be noted that this percentage difference has a geometric dependence. This strong dependence provides an extra hint that this background uncertainty is perhaps correlated with the neutrino beam flux. Regardless of the provenance of the underlying issue, the ratios presented in table 8.16 provide a good candidate to describe this ad hoc uncertainty. The ratio itself essentially describes the size of the signal enhanced regions to the background enhanced regions. By quantifying this data and Monte Carlo difference, a crude uncertainty on the migration of events between the signal and background regions is assessed. To include this extra information in

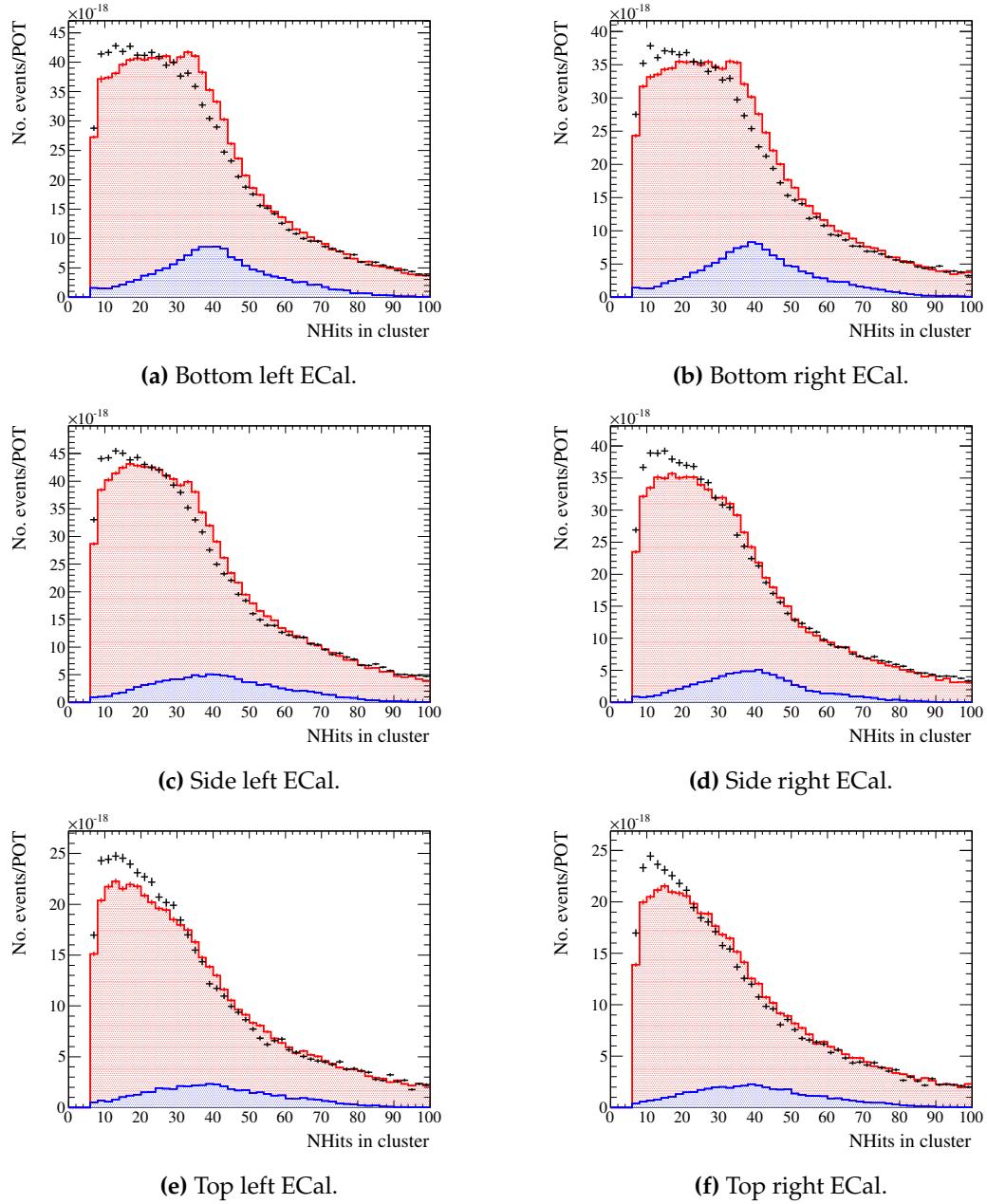


Figure 8.49: The number of events which pass the fiducial volume cut as a function of the number of hits in the reconstructed cluster. The red and blue histograms show the number of beam MC events and sand MC events respectively. The black points show the run 3C data. In all distributions there is an excess of low number of hit clusters in data. For the ECal modules closest to the beam (bottom-left, bottom-right and side-left) there is a significant MC excess at approximately 31 hits which aligns with the peak of the sand MC events.

Barrel module	NSel/NRev (MC)	NSel/NRev (data)	% diff
Bottom right	0.773	0.767	0.7%
Side right	0.883	0.878	0.6%
Top right	0.889	0.891	0.2%
Bottom left	0.785	0.808	2.9%
Side left	0.837	0.785	6.2%
Top left	0.856	0.831	2.9%

Table 8.16: Comparison of the number of events between the barrel and barrel-reverse samples. The second and third column show the ratio of the number of events in the barrel samples to the number of events in the barrel-reverse samples for Monte Carlo and data respectively. The fourth column shows the percentage difference between the Monte Carlo and data ratios.

the fit, the percentage difference shown in the fourth column of table 8.16 needs to be included in the shape-only covariance matrix. Specifically, for each barrel sample and its reverse counterpart, the corresponding percentage difference is added to the diagonal element of the covariance matrix. Using the modified covariance matrix as an input, the fit was then re-run using the NEUT MC templates and run 3C data. The number of events (post-fit) for each sample are shown in Fig. 8.50 and the normalisation parameters found by the fit are shown in table 8.17. The minimum χ^2/NDOF returned by the fit was 2.54.

The output of the fitter is more reassuring. The reduced χ^2 , while still a little high, suggests that the Monte Carlo is a much better fit to the data. Notably, all of the fitted normalisation parameters agree with the NEUT prediction within their shape-only error.

Normalisation parameter	Fitted value
R^{Pb}	$1.10^{+0.11}_{-0.10}$
R^{C}	$1.10^{+0.14}_{-0.12}$
R^{Other}	$0.90^{+0.10}_{-0.11}$

Table 8.17: The values returned from the fitter modified with the ad hoc uncertainty when processing the NEUT MC templates with T2K data.

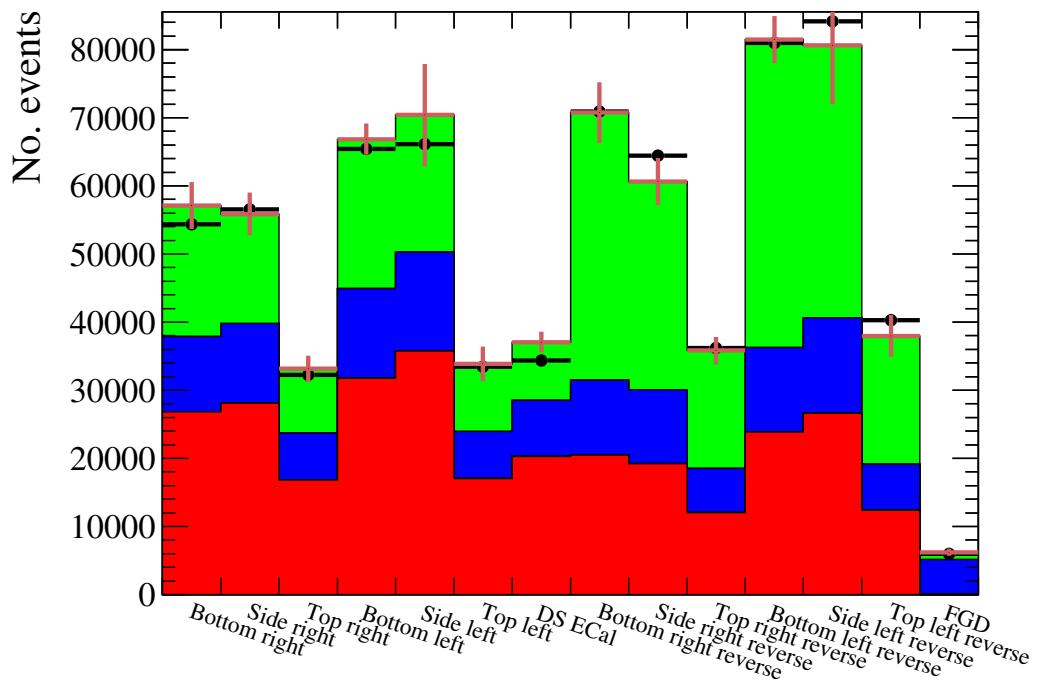


Figure 8.50: The post-fit number of NEUT Monte Carlo events compared with T2K data when including the ad hoc uncertainty shown in table 8.16. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic and ad hoc uncertainties on the number of NEUT MC events. The black points are the T2K data. The inflated errors has reduced the level of correlations, allowing the fitter to make finer changes to the MC normalisation.

8.7 The ν_μ charged current inclusive cross-section on lead

Using the fitted value of R^{Pb} , the cross-section can be calculated using equation 8.15 and event information from the DS ECal. As described in section 8.4, the errors on R^{Pb} returned from the fitter provides the shape systematic uncertainties. To calculate the full error, the extra error shown in equation 8.16 (12.5%) can be added in quadrature with the shape-only errors for R^{Pb} . Following this route, the value of R^{Pb} and its percentage uncertainty, which is the same percentage uncertainty value for the CC inclusive cross-section, was found to be

$$R^{\text{Pb}} = 1.10_{-0.17}^{+0.18}. \quad (8.20)$$

where the positive and negative errors are equivalent to +16.4% and -15.5% of the fitted value respectively. The dimensions and quantity of the lead absorbers in the DS ECal are described in chapter 3 and section 8.4. Taking the density of lead to be 11.34 g/cm³, the total mass of lead in the DS ECal is 2661 kg. The atomic mass of lead is 207.2 ± 0.1 atomic units [91], which infers there are 7.74×10^{27} lead nuclei present in the Monte Carlo simulation of the DS ECal. The lack of antimony in the simulation has already been quantified in the final uncertainty above.

The selection presented in chapter 7 had a 53.0% efficiency and 72.4% purity for selecting charged current neutrino interactions in the active volume of the DS ECal. Considering interactions on lead alone, the efficiency and purity are 53.8% and 49.7% respectively with 32427 reconstructed events selected when scaled to the collected data POT.

Fig. 8.51 shows the simulated neutrino flux (per 10^{21} POT) for T2K run 3C. The POT-scaled integral of this distribution provides the total neutrino flux used in this study. As the analysis ran over 1.31×10^{20} POT, the integral of Fig. 8.51 multiplied by the collected POT gives a total neutrino flux of 2.53×10^{12} cm⁻².

Collating all of this information together and applying equation 8.15, a T2K flux-

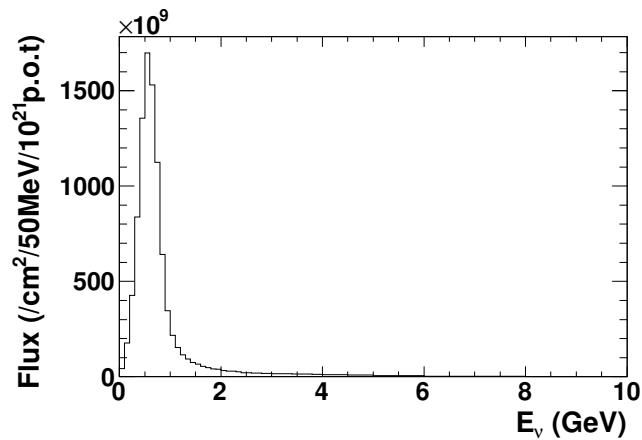


Figure 8.51: The simulated flux for T2K run 3C as a function of neutrino energy.

averaged ν_μ charged current cross-section on lead is measured as

$$\langle\sigma_{\text{Pb}}^{\text{CC}}\rangle_\phi = 8.13_{-1.26}^{+1.33} \times 10^{-39} \text{ cm}^2 \text{ nucleon}^{-1}. \quad (8.21)$$

Chapter 9

Discussion and conclusions

9.1 Discussion

The results presented in chapter 8 have potentially highlighted areas for improvement in the ND280 simulation. Fig. 8.41 showed discrepancies between data and MC for each ECal sample which ultimately led to the failure of the ECal rate fit when applied to T2K data. Upon further inspection of the fit, it was noted that there were two significant areas of discrepancy between data and MC (shown in Fig. 8.49).

As already discussed in chapter 8, one of the discrepancies appears to be correlated with the sand MC. The sand MC is a relatively new addition to the ND280 simulation and has been tuned to Tracker region data. In most Tracker analyses, the number of sand events which make it into the final selection generally number below 10 and so it is an area which has not required in depth study. This is obviously not the case for ECal based analyses which see a much higher number of events originating from the surrounding pit. The second discrepancy identified showed an excess of data events with a low number of hits in the reconstructed ECal clusters. A brief study of such events suggests that they originate from the surrounding magnet region. What is most troubling about these two issues is that they appear in the analysis at all. Both topologies appear to be an entering background which strongly suggests that further study of the ECal fiducial volume is needed.

As a separate problem, the reverse selection itself should be revisited. This selection is currently defined as any event which passes the fiducial volume cut but fails the final (most upstream) cut. While this allowed a large number of events to be included

in the analysis, it also open up the possibility of data and MC mismodelling which either increases the systematic uncertainty or, like in this analysis, causes major issues in the final measurement.

However, this analysis has provided several important contributions to the T2K experiment. The first is that the ECal is capable of reconstructing topologies outside of the single track or single shower hypotheses. This is a potentially important milestone for the ECal as a neutrino detector as it paves the way for analysing neutrino events where the ECal is the target. The second contribution is that it is possible to apply a selection to the enhanced reconstruction to actually select neutrino interactions in the ECal. While these developments have been used in a cross-section measurement, they were in no way developed solely with this analysis in mind. So, the cross-section measurement itself can essentially be thought of as a by-product of the development of the reconstruction and selection.

9.2 Conclusions

As highlighted in chapter 1, the field of neutrino physics has rapidly evolved since the neutrino’s discovery. At time of writing, current and future generation neutrino oscillation experiments are aiming towards a measurement of the mass-hierarchy and the CP violating phase δ . To do this, a solid understanding of neutrino cross-section physics with atomic nuclei is vital.

The overall aim of this analysis was to study neutrino interactions in the ND280 ECals, with a measurement of the ν_μ CC inclusive cross-section on lead as the final goal. To realise this, a new set of reconstruction algorithms were developed to allow the reconstruction of the charged final-states of the neutrino interactions in the ECal. The output of this reconstruction was then used as the foundation of a ν_μ CC inclusive selection which achieved a good efficiency and purity for both the barrel ECals and DS ECal. The final stage of the analysis was to use the selected events in a χ^2 fit to extract the data-constrained rate of neutrino interactions on lead. An important input to the fit itself was the evaluation of systematic uncertainties. This involved a set of ECal systematic uncertainties which were evaluated for the first time and should be extremely useful to future analysers.

Unfortunately, there were unforeseen discrepancies between data and MC which caused major problems in the final fit. These were overcome by introducing an ad hoc uncertainty which encompassed the migration of problematic background events between the signal enhanced and background enhanced samples. With the inclusion of this extra uncertainty, a T2K flux-averaged ν_μ charged current cross-section on lead was measured to be

$$\langle\sigma_{\text{Pb}}^{\text{CC}}\rangle_\phi = 8.13_{-1.26}^{+1.33} \times 10^{-39} \text{ cm}^2 \text{ nucleon}^{-1}. \quad (9.1)$$

9.3 The future

The highest priority for future iterations of this analysis should be an in depth analysis of the problematic background events shown in Fig. 8.49. The ideal approach would be to study this unforeseen background in great detail. It should then be possible to improve the ND280 simulation to more accurately model this background and enhance the selection cuts to remove the background from the analysis.

There is an immense amount of scope for the reconstruction algorithms. While the Hough transform has shown itself to be a very powerful method of reconstruction, analysis of the generated parameter space is actually very rudimentary. As a reminder, the current method looks for the maxima in the parameter space which actually represent the 2D lines that pass through the most hits. This analysis method needs improvement. For example, the hits which are selected as constituents of the track are removed from the parameter space which means any subsequent track candidate does not have that information available. A more advanced method of hit masking would allow hit sharing between tracks to take place.

Improvements can also be made to the selection. As it stands, the selection does not fully harness the power of the enhanced reconstruction. For example, particle identification could be applied to the individual tracks which would potentially allow removal of intrinsic background.

Most importantly, the reconstruction and selection provide much scope for expanding this analysis far beyond what is presented. At the time of writing, a dedicated group of analysers is being set up with a common aim of studying neutrino interactions in

the ECal and its potential for physics contributions. For example, using the ν_μ interaction rates in the barrel ECals to help constrain the flux in the ND280 Tracker region and matching reconstructed ECal tracks to TPC tracks to make a ν_μ CC differential cross-section measurement on lead.

Despite the large issues raised here, the main contribution of this analysis is that the many tools developed here pave the way for a wide range of neutrino interaction studies in which the ECal is the target.

Bibliography

- [1] W. Pauli, Open letter to the Gauverein meeting in Tübingen, 1930.
- [2] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, *Detection of the Free Neutrino: a Confirmation*, Science **124**, 103 (1956).
- [3] BNL Collaboration, G. Danby et al., *Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos*, Phys. Rev. Lett. **9**, 36 (1962).
- [4] DONUT Collaboration, K. Kodama et al., *Observation of tau neutrino interactions*, Physics Letters B **504**, 218 (2001).
- [5] Super-Kamiokande Collaboration, Y. Fukuda et al., *Evidence for Oscillation of Atmospheric Neutrinos*, Phys. Rev. Lett. **81**, 1562 (1998).
- [6] SNO Collaboration, Q. R. Ahmad et al., *Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by 8B Solar Neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **87**, 071301 (2001).
- [7] KamLAND Collaboration, K. Eguchi et al., *First Results from KamLAND: Evidence for Reactor Antineutrino Disappearance*, Phys. Rev. Lett. **90**, 021802 (2003).
- [8] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the Unified Model of Elementary Particles*, Progress of Theoretical Physics **28**, 870 (1962).
- [9] B. Pontecorvo, *Mesonium and antimesonium*, JETP Lett. **33**, 549 (1957).
- [10] L. Wolfenstein, *Neutrino oscillations in matter*, Phys. Rev. D **17**, 2369 (1978).
- [11] S. Mikheev and A. Y. Smirnov, *Resonance Amplification of Oscillations in Matter and Spectroscopy of Solar Neutrinos*, Sov.J.Nucl.Phys. **42**, 913 (1985).
- [12] Particle Data Group, K. Olive et al., *Review of Particle Physics*, Chin.Phys. **C38**, 090001 (2014).

- [13] F. Capozzi et al., *Status of three-neutrino oscillation parameters, circa 2013*, Phys. Rev. D **89**, 093018 (2014).
- [14] T2K Collaboration, K. Abe et al., *Precise Measurement of the Neutrino Mixing Parameter θ_{23} from Muon Neutrino Disappearance in an Off-Axis Beam*, Phys. Rev. Lett. **112**, 181801 (2014).
- [15] Double Chooz Collaboration, Y. Abe et al., *First measurement of delayed neutron capture on hydrogen in the Double Chooz experiment*, Physics Letters B **723**, 66 (2013).
- [16] Daya Bay Collaboration, F. P. An et al., *Observation of Electron-Antineutrino Disappearance at Daya Bay*, Phys. Rev. Lett. **108**, 171803 (2012).
- [17] RENO Collaboration, J. K. Ahn et al., *Observation of Reactor Electron Antineutrinos Disappearance in the RENO Experiment*, Phys. Rev. Lett. **108**, 191802 (2012).
- [18] B. T. Cleveland et al., *Measurement of the Solar Electron Neutrino Flux with the Homestake Chlorine Detector*, The Astrophysical Journal **496**, 505 (1998).
- [19] SAGE Collaboration, J. N. Abdurashitov et al., *Measurement of the solar neutrino capture rate with gallium metal. III. Results for the 2002~2007 data-taking period*, Phys. Rev. C **80**, 015807 (2009).
- [20] K2K Collaboration, M. H. Ahn et al., *Measurement of neutrino oscillation by the K2K experiment*, Phys. Rev. D **74**, 072003 (2006).
- [21] MINOS Collaboration, D. G. Michael et al., *Observation of Muon Neutrino Disappearance with the MINOS Detectors in the NuMI Neutrino Beam*, Phys. Rev. Lett. **97**, 191801 (2006).
- [22] T2K Collaboration, K. Abe et al., *Observation of Electron Neutrino Appearance in a Muon Neutrino Beam*, Phys. Rev. Lett. **112**, 061802 (2014).
- [23] NOvA Collaboration, D. Ayres et al., *NOvA: Proposal to build a 30 kiloton off-axis detector to study $\nu_\mu \rightarrow \nu_e$ oscillations in the NuMI beamline*, (2004), hep-ex/0503053.
- [24] Hyper-Kamiokande Working Group, K. Abe et al., *A Long Baseline Neutrino Oscillation Experiment Using J-PARC Neutrino Beam and Hyper-Kamiokande*, (2014), 1412.4673.
- [25] LBNE Collaboration, C. Adams et al., *The Long-Baseline Neutrino Experiment: Exploring Fundamental Symmetries of the Universe*, (2013), 1307.7335.

- [26] X. Qian and P. Vogel, *Neutrino mass hierarchy*, Progress in Particle and Nuclear Physics **83**, 1 (2015).
- [27] C. Weinheimer, *KATRIN, a next generation tritium decay experiment in search for the absolute neutrino mass scale*, Progress in Particle and Nuclear Physics **48**, 141 (2002).
- [28] EXO Collaboration, M. Auger et al., *Search for Neutrinoless Double-Beta Decay in ^{136}Xe with EXO-200*, Phys. Rev. Lett. **109**, 032505 (2012).
- [29] A. Barabash and the SuperNemo Collaboration, *SuperNEMO double beta decay experiment*, Journal of Physics: Conference Series **375**, 042012 (2012).
- [30] SNO+ Collaboration, M. C. Chen, *The SNO+ Experiment*, (2008), 0810.3694.
- [31] J. Fan and P. Langacker, *Light sterile neutrinos and short baseline neutrino oscillation anomalies*, Journal of High Energy Physics **2012** (2012).
- [32] LSND Collaboration, A. Aguilar-Arevalo et al., *Evidence for neutrino oscillations from the observation of anti-neutrino(electron) appearance in a anti-neutrino(muon) beam*, Phys.Rev. **D64**, 112007 (2001), hep-ex/0104049.
- [33] Various LEP Collaborations, T. S. Electroweak and H. F. Groups, *Precision electroweak measurements on the Z resonance*, Physics Reports **427**, 257 (2006).
- [34] MiniBooNE Collaboration, A. Aguilar-Arevalo et al., *Improved Search for $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ Oscillations in the MiniBooNE Experiment*, Phys.Rev.Lett. **110**, 161801 (2013), 1207.4809.
- [35] C. M. Ignarra, *MicroBooNE*, ArXiv e-prints (2011), 1110.1604.
- [36] LAr1-ND Collaboration, C. Adams et al., *LAr1-ND: Testing Neutrino Anomalies with Multiple LArTPC Detectors at Fermilab*, (2013), 1309.7987.
- [37] J. A. Formaggio and G. P. Zeller, *From eV to EeV: Neutrino cross sections across energy scales*, Rev. Mod. Phys. **84**, 1307 (2012).
- [38] D. C. Colley et al., *Cross-sections for Charged Current Neutrino and Anti-neutrino Interactions in the Energy Range 10-GeV to 50-GeV*, Z. Phys. **C2**, 187 (1979).
- [39] NuTeV Collaboration, M. Tzanov et al., *Precise measurement of neutrino and antineutrino differential cross sections*, Phys. Rev. D **74**, 012008 (2006).

- [40] MINOS Collaboration, P. Adamson et al., *Neutrino and antineutrino inclusive charged-current cross section measurements with the MINOS near detector*, Phys. Rev. D **81**, 072002 (2010).
- [41] D. Casper, *The Nuance neutrino physics simulation, and the future*, Nucl. Phys. Proc. Suppl. **112**, 161 (2002), hep-ph/0208030, [,161(2002)].
- [42] K. Nakamura and P. D. Group, *Review of Particle Physics*, Journal of Physics G: Nuclear and Particle Physics **37**, 075021 (2010).
- [43] MiniBooNE Collaboration, A. A. Aguilar-Arevalo et al., *First measurement of the muon neutrino charged current quasielastic double differential cross section*, Phys. Rev. D **81**, 092005 (2010).
- [44] V. Lyubushkin et al., *A study of quasi-elastic muon neutrino and antineutrino scattering in the NOMAD experiment*, The European Physical Journal C **63**, 355 (2009).
- [45] LSND, L. B. Auerbach et al., *Measurements of charged current reactions of muon neutrinos on C-12*, Phys. Rev. C**66**, 015501 (2002), nucl-ex/0203011.
- [46] NOMAD Collaboration, V. Lyubushkin et al., *A Study of quasi-elastic muon neutrino and antineutrino scattering in the NOMAD experiment*, Eur. Phys. J. C**63**, 355 (2009), 0812.4543.
- [47] R. Smith and E. Moniz, *NEUTRINO REACTIONS ON NUCLEAR TARGETS*, Nucl.Phys. **B43**, 605 (1972).
- [48] CHORUS Collaboration, *Measurement of the Z/A dependence of neutrino charged-current total cross-sections*, The European Physical Journal C - Particles and Fields **30**, 159 (2003).
- [49] M. Gluck, E. Reya, and A. Vogt, *Dynamical parton distributions revisited*, Eur. Phys. J. C**5**, 461 (1998), hep-ph/9806404.
- [50] MINERvA Collaboration, B. G. Tice et al., *Measurement of Ratios of ν_μ Charged-Current Cross Sections on C, Fe, and Pb to CH at Neutrino Energies 2^20 GeV*, Phys. Rev. Lett. **112**, 231801 (2014).
- [51] C. Andreopoulos et al., *The {GENIE} neutrino Monte Carlo generator*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **614**, 87 (2010).

- [52] T2K Collaboration, K. Abe et al., *The T2K experiment*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **659**, 106 (2011).
- [53] T2K Collaboration, K. Abe et al., *Measurement of the Inclusive Electron Neutrino Charged Current Cross Section on Carbon with the T2K Near Detector*, Phys. Rev. Lett. **113**, 241803 (2014).
- [54] T2K Collaboration, K. Abe et al., *Measurement of the inclusive ν_μ charged current cross section on carbon in the near detector of the T2K experiment*, Phys. Rev. D **87**, 092003 (2013).
- [55] T2K Collaboration, K. Abe et al., *T2K neutrino flux prediction*, Phys. Rev. D **87**, 012001 (2013).
- [56] D. Renker and E. Lorenz, *Advances in solid state photon detectors*, Journal of Instrumentation **4**, P04004 (2009).
- [57] P.-A. Amaudruz et al., *The T2K fine-grained detectors*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **696**, 1 (2012).
- [58] N. Abgrall et al., *Time projection chambers for the T2K near detectors*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **637**, 25 (2011).
- [59] S. Assylbekov et al., *The T2K ND280 off-axis pizero detector*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **686**, 48 (2012).
- [60] D. Allan et al., *The electromagnetic calorimeter for the T2K near detector ND280*, Journal of Instrumentation **8**, P10019 (2013).
- [61] S. Aoki et al., *The T2K Side Muon Range Detector (SMRD)*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **698**, 135 (2013).
- [62] J. Estrada, C. Garcia, B. Hoeneisen, and P. Rubinov, *MCM II and the trip chip*, (2003).
- [63] T2K UK, *The T2K UK proposal*, (2009).

- [64] Super-Kamiokande Collaboration, S. Fukuda et al., *The Super-Kamiokande detector*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **501**, 418 (2003).
- [65] R. Brun and F. Rademakers, {ROOT} *An object oriented data analysis framework*, 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.
- [66] A. Ferrari et al., Fluka: a multi-particle transport code, in *CERN 2005-10* (2005), INFN/TC 05/11, SLAC-R-773.
- [67] R. Brun, F. Bruyant, M. Maire, A. McPherson, and P. Zanarini, *GEANT3*, (1987).
- [68] NA61/SHINE Collaboration, N. Abgrall et al., *Measurements of cross sections and charged pion spectra in proton-carbon interactions at 31 GeV/c*, Phys. Rev. C **84**, 034604 (2011).
- [69] Y. Hayato, *Neut*, Nuclear Physics B - Proceedings Supplements **112**, 171 (2002).
- [70] G. A. Miller, *Fermi Gas Model*, Nuclear Physics B - Proceedings Supplements **112**, 223 (2002).
- [71] O. Benhar, A. Fabrocini, S. Fantoni, and I. Sick, *Spectral function of finite nuclei and scattering of GeV electrons*, Nuclear Physics A **579**, 493 (1994).
- [72] C. L. Smith, *Neutrino reactions at accelerator energies*, Physics Reports **3**, 261 (1972).
- [73] D. Rein and L. M. Sehgal, *Neutrino-excitation of baryon resonances and single pion production*, Annals of Physics **133**, 79 (1981).
- [74] T. Sjstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, Journal of High Energy Physics **2006**, 026 (2006).
- [75] S. Agostinelli et al., *Geant4a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506**, 250 (2003).
- [76] The ND280 calibration group, *Calibration of the ND280 scintillator detectors for Prod 5*, T2K Technical note **130** (2012).
- [77] P. Hough, *Machine Analysis Of Bubble Chamber Pictures*, Conf.Proc. **C590914**, 554 (1959).

- [78] R. O. Duda and P. E. Hart, *Use of the Hough Transformation to Detect Lines and Curves in Pictures*, Commun. ACM **15**, 11 (1972).
- [79] J. Vuillemin, Fast linear hough transform, in *Application Specific Array Processors, 1994. Proceedings. International Conference on*, pp. 1–9, 1994.
- [80] L. Pickering, Private communication.
- [81] X. Lu, Private communication.
- [82] K2K Collaboration, A. Rodriguez et al., *Measurement of single charged pion production in the charged-current interactions of neutrinos in a 1.3 GeV wide band beam*, Phys. Rev. D **78**, 032003 (2008).
- [83] SciBooNE Collaboration, Y. Nakajima et al., *Measurement of inclusive charged current interactions on carbon in a few-GeV neutrino beam*, Phys. Rev. D **83**, 012005 (2011).
- [84] M. Scott, *Measuring charged current neutrino interactions in the electromagnetic calorimeters of the ND280 detector*, PhD thesis, Imperial College London, 2013.
- [85] M. Ravonel, J. Lagoda, and A. Hillairet, ν_μ CC event selections in the ND280 tracker using Run 2+3+4 data, T2K Technical note **212** (2015).
- [86] T2K Collaboration, K. Abe et al., *Measurement of the inclusive ν_μ charged current cross section on iron and hydrocarbon in the T2K on-axis neutrino beam*, Phys. Rev. D **90**, 052010 (2014).
- [87] IHEP-JINR Collaboration, V. B. Anikeev et al., *Total cross-section measurements for muon-neutrino, anti-muon-neutrino interactions in 3-GeV - 30-GeV energy range with IHEP-JINR neutrino detector*, Z. Phys. **C70**, 39 (1996).
- [88] C. Andreopoulos, J. Dobson, and S. Dytman, *Handling Neutrino Interaction Uncertainties using Event Reweighting*, T2K Technical note **7** (2009).
- [89] P. de Perio, K. Mahn, and P. Sinclair, *Implementation of the NIWG cross section parameterization*, T2K Technical note **113** (2012).
- [90] S. Oser, *Elemental composition and masses of FGD XY modules*, T2K Technical note **91** (2011).
- [91] J. R. de Laeter et al., *Atomic weights of the elements*, Pure and Applied Chemistry **75**, 683 (2003).