
Constraint of Systematic Uncertainties in
an Electron Neutrino Search Using Muon
Neutrinos at MicroBooNE

Adam Lister

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Lancaster University
2019

Abstract

MicroBooNE is a liquid argon time projection chamber which has been running in the Booster Neutrino Beam at Fermilab since 2015. The primary goal of MicroBooNE is investigation of the excess of electromagnetic events observed by the MiniBooNE collaboration. Due to limitations of the Cherenkov-based particle identification of MiniBooNE, this excess could be interpreted as either photon-like or electron-like. A photon-like excess would indicate that there are processes which are not well understood which could act as a background in neutrino oscillation measurements, while an electron-like excess could indicate the presence of sterile neutrinos, the existence of which is one of the most hotly debated questions in the field.

This work will outline the MicroBooNE strategy for investigation of this *low-energy excess*, with particular attention given to the role of the muon neutrino sideband which is used as an important constraint on systematic uncertainties. A procedure has been developed in order to apply this constraint to an electron neutrino dataset, and it has been shown that the constraint results in an improvement to the sensitivity.

In order to perform this constraint, an exclusive-state ν_μ CC selection has been developed, which results in 804 selected events from on-beam data. The ratio of the data with respect to simulation is $R = 0.78 \pm 0.04$ (stat.) ± 0.12 (syst.).

In addition, this thesis presents a first measurement of the longitudinal ionisation electron diffusion coefficient from the MicroBooNE data, which is determined to be $3.73_{-0.68}^{+0.70}$ cm²/s.

Declaration

This thesis is a presentation of original research performed by the author. It has not been submitted in support of an application for another degree at this or any other university.

The work presented in this thesis was ultimately made possible by a large number of MicroBooNE collaborators. References to work not performed by the author have been used where possible.

The primary contributions of the author to the MicroBooNE experiment are through the development of a method to measure the longitudinal electron diffusion component as outlined in Chapter 5, which was undertaken with Mr Andrew Mogan and Prof. Sowjanya Gollapinni, the development and implementation of a new particle identification algorithm as described in Chapter 6, which was done in collaboration with Dr Kirsty Duffy, and the development of an exclusive-state muon neutrino selection as described in Chapter 7.

In addition to these analyses, several service tasks were undertaken. Validation of the MicroBooNE software used in this work was performed, which involved development of a continuous integration suite, described in Chapter 4.8, which is now used as standard in MicroBooNE. This was also developed in collaboration with Dr Kirsty Duffy. An extension to this work meant acting as a continuous integration expert to liaise between experiment shifters and software developers in order to minimise errors in releases. In addition to this, the author was responsible for implementation of the data-driven noise model, which is outlined in Chapter 4.5.1.

Adam Lister

For Olivia.

Acknowledgements

First and foremost, to Mam, Dad, LJ, and my grandparents: I owe you a monumental debt of gratitude. I would not be half of the person I am today without you having my back day in, day out. You have been my biggest supporters for as long as I can remember and have been unfailingly encouraging (well, maybe not you, LJ, “*so, when are you getting a real job?*”). Thank you for telling me loudly and repeatedly how important it is to be myself. This would not have been possible without each and every one of you. I love you.

Thanks to my advisors, Andrew Blake and Jaroslaw Nowak, for providing me with this incredible opportunity, for the occasional course correction when I lost sight of the bigger picture, and for actually reading this thing. It kind of just kept growing.

I also owe an absurd amount to Dom Brailsford, I would probably still be struggling to write my first ROOT script if it wasn't for you. Thanks for always providing a critical eye: it's made me a better physicist.

There's a lot of people on MicroBooNE I need to thank. Let me start by first saying that I would not have made it to the end without an amazing amount of input from Kirsty Duffy and Sowjanya Gollapinni. I think I owe you everything. Thanks also go out to Corey Adams, for showing me that I'm not alone in my cynicism, and to Joseph Zennamo for keeping me enthused enough that I could overlook the cynicism most of the time: I needed both of you. Thanks Andy

Furmanski, for always being available to provide advice (but mostly to rant about the differences between the UK and the US, or actually, most things). Thanks to David Caratelli, for seemingly knowing everything (and for ruining my dreams by telling me every pretty event display is a cosmic). Thanks also to Lorena Escudero Sanchez and Raquel Castillo Fernández, for providing me with a very Spanish outlook when I needed it. Also huge props to Mark Ross-Lonergan and Davio Cianci for talking stats, SBNFit, and generally reassuring me that I wasn't crazy in the last few months. If I keep going then this will go for pages, but suffice to say thank you to all of the MicroBooNE grad students for keeping me sane (especially the Butterfield guys!), I don't think I could have asked for a better lot. This goes especially for Colton Hill, who's had to deal with me twice as much as the rest of you by virtue of living 5 meters from me.

I could write *many* paragraphs for the friends who should take as much ownership of this as I do, but I'll stop here and just say that to friends near and far: I love you very much, thank you for everything, you've been the best group of people to see this through with.

Glossary

ACPT Anode-or-Cathode Piercing Tracks

ADC Analogue-to-Digital Converter

APA Anode Plane Array

art The event-processing framework made use of by LArSoft (art is not an acronym)

ASIC Application Specific Integrated Circuit

BNB Booster Neutrino Beam

BNB+Cosmic Simulated neutrino interaction from the GENIE generator overlaying a simulated cosmic background from CORSIKA

CC Charged-Current

CI Continuous Integration

CORSIKA Cosmic Ray Simulations for Kascade, the primary package used in the simulation of cosmic rays

CRT Cosmic Ray Tagger, scintillator panels surrounding the MicroBooNE cryostat

DAQ Data Acquisition System

DIC Dynamic Induced Charge, a simulated dataset which makes use of a preliminary induced charge model

DIS Deep Inelastic Scattering

EXTBNB Dataset external to the beam window, but with the same conditions placed on the software trigger as in the on-beam dataset (BNB trigger conditions)

FEM Front End Motherboard

GENIE Generates Events for Neutrino Interaction Experiments is used to generate neutrino interactions

ICARUS Imaging Cosmic And Rare Underground Signals experiment, the far detector of SBN

LArSoft Liquid Argon Software, the software package used by LArTPCs, which is built upon the art event-processing framework

LArTPC Liquid Argon Time Projection Chamber

LINAC Linear Accelerator

MicroBooNE Micro Booster Neutrino Experiment, the middle detector of SBN

MiniBooNE Mini Booster Neutrino Experiment, the predecessor to MicroBooNE

NC Neutral-Current

Pandora The multi-algorithm pattern recognition reconstruction framework used in analyses in this work

PFO Particle Flow Object, a collection of PFPs

PFP Particle Flow Particle, a reconstructed object from Pandora which corresponds to one reconstructed particle in a PFO

PID Particle Identification

PMT Photomultiplier Tube

POT Protons On Target

QE Quasi-Elastic

RES Resonant Pion Production

SBN Short Baseline Neutrino Program

SBND Short Baseline Near Detector Experiment, the near detector of SBN

TPC Time Projection Chamber

Contents

1	Introduction	1
1.1	Liquid Argon Time Projection Chambers: Precise and Scalable	2
1.2	Sterile Neutrinos: A Known Unknown	3
1.3	Probing the Sterile Neutrino Question with LArTPCS	4
1.4	Thesis Outline	4
2	Neutrino Physics	7
2.1	Dear Radioactive Ladies and Gentlemen...	7
2.2	The Standard Model Neutrino	8
2.3	Neutrino Interactions	10
2.4	Neutrino Oscillations	12
2.5	The Global Sterile Neutrino Picture	17
2.5.1	Results from Accelerator Experiments	18
2.5.2	Results from Gallium Experiments	23
2.5.3	Results from Reactor Neutrino Experiments	25
2.5.4	Global Fits	26
3	The MicroBooNE Experiment	31
3.1	The Booster Neutrino Beam	32
3.1.1	Beam Composition	33
3.2	The MicroBooNE Detector	35

3.2.1	Signals in LArTPCs	36
3.2.2	The Liquid Argon Time Projection Chamber	38
3.2.3	The Light Collection System	46
3.2.4	The Cosmic Ray Tagger	48
3.3	Data-Taking Triggers	49
3.3.1	Accelerator Trigger	49
3.3.2	PMT Trigger	49
3.4	Detector Operations	50
4	MicroBooNE Simulation and Reconstruction	53
4.1	Overview	53
4.2	Neutrino Flux Simulation	54
4.2.1	Fits to World Data	54
4.3	Neutrino Interaction Simulation	56
4.4	Cosmic Particle Simulation	59
4.5	MicroBooNE Detector Simulation	60
4.5.1	TPC Simulation	60
4.5.2	Optical Simulation	63
4.6	Reconstruction	65
4.6.1	Signal Processing	65
4.6.2	The Pandora Pattern Recognition Software	66
4.6.3	Optical Reconstruction	69
4.7	Energy Scale Calibration	70
4.7.1	Misconfigured Region Correction	71
4.7.2	dQ/dx calibration	71
4.7.3	dE/dx calibration	71
4.8	Continuous Integration Validation	72

5	Longitudinal Electron Diffusion in MicroBooNE	75
5.1	Diffusion in LArTPCs	75
5.2	Method for Effective Longitudinal Diffusion Coefficient Extraction	79
5.2.1	Selection of Tracks	82
5.2.2	Selection of high quality waveforms	82
5.2.3	Waveform Averaging Technique	85
5.2.4	Extracting the Diffusion Value	86
5.2.5	Estimation of Error	88
5.3	Method Validation	88
5.3.1	Using a Reduced Simulation Sample	88
5.3.2	Including Noise in the Simulation	93
5.3.3	Including Physics Effects in the Simulation	94
5.3.4	Moving to a Simulated Cosmic Sample	95
5.3.5	Performing the Diffusion Measurement on a Full Monte Carlo Simulation	102
5.4	Measuring the Effective Longitudinal Diffusion Coefficient in Data	105
5.5	Systematic Studies	106
5.5.1	Transverse Diffusion	106
5.5.2	Response Function	106
5.5.3	Drift Velocity	106
5.5.4	Space Charge Effect	107
5.5.5	Summary of Uncertainties	108
5.6	Summary and Future Work	109
5.6.1	Detector Systematic Uncertainties to Address	109
5.6.2	Geant4-level Physics Uncertainties to Address	109
5.6.3	Further Measurements	109
6	Particle Identification	111
6.1	Energy Deposition in Liquid Argon	112

6.2	Simulated Data dE/dx Tuning	114
6.3	Algorithm Comparisons	119
6.3.1	Sample Information and Scaling Factors	120
6.3.2	Deposited Energy Versus Energy By Range Algorithm	121
6.3.3	PIDa Algorithm	124
6.3.4	χ^2 Algorithm	126
6.3.5	Bragg Likelihood Algorithm	128
6.4	Discussion	136
7	Selection of ν_μ CC0πNp Events	145
7.1	Signal Definition and Backgrounds	145
7.2	Data Samples	147
7.2.1	Run 1 Data	147
7.2.2	Simulated Data	147
7.2.3	Out-of-TPC Simulation	148
7.2.4	Detector Variation Samples	148
7.3	The ν_μ CC-inclusive selection	152
7.3.1	Rejection of Cosmic-Ray Backgrounds	152
7.3.2	Flash to TPCObject Matching	155
7.3.3	Fiducial Volume	156
7.3.4	Performance of the ν_μ CC-inclusive Selection	157
7.4	Selection of ν_μ CC 0 π N p Events	158
7.4.1	Topology Cuts	158
7.4.2	Particle Identification	166
7.4.3	Quality Cuts	167
7.4.4	Selection Performance	168
7.5	Distributions of Selected Events	174
7.6	Neutrino Energy Reconstruction	176

7.7	Evaluation of Systematic Uncertainties	182
7.7.1	Flux Uncertainties	185
7.7.2	Interaction Uncertainties	186
7.7.3	Detector Uncertainties	187
7.7.4	Total Uncertainties	191
8	The ν_μ Constraint	195
8.1	Understanding the ν_μ Constraint	196
8.2	Demonstration of the ν_μ Constraint	198
8.3	Frequentist Studies	200
9	Conclusions	205
A	Additional Information for Particle Identification in MicroBooNE	219
A.1	Toy Width Study	219
A.2	Bare χ^2 Distributions	222
A.2.1	POT normalised	222
A.2.2	Template Fit	222
A.3	POT normalised Distributions	225
B	Additional Information for the ν_μ CC0πNp Selection	233
B.1	A Note on Reconstructed-To-True Object Matching	233
B.2	Kinematic Variables with the DIC Sample	234

List of Figures

1.1	Neutrino detectors presented as a function of year built and neutrino interaction vertex resolution.	3
2.1	Charged-current and neutral-current interaction Feynman diagrams.	11
2.2	Diagrams showing examples of ν_μ CC QE, RES, and DIS interactions.	11
2.3	Predicted ν_μ CC cross sections as a function of energy.	13
2.4	Neutrino oscillation probabilities for $\nu_\mu \rightarrow \nu_e$, $\nu_\mu \rightarrow \nu_\mu$, and $\nu_\mu \rightarrow \nu_\tau$ transitions.	16
2.5	Measurement of the Z width from LEP.	17
2.6	Image showing how particle identification is performed using Cherenkov light.	18
2.7	LSND beam excess (background-subtracted event count) as a function of L/E.	19
2.8	MiniBooNE excess as a function of energy.	20
2.9	The IceCube 90% and 99% confidence levels, shown with the 95% and 68% CL sensitivity bands.	22
2.10	The 90% confidence level for the combined fit of MINOS, Daya Bay, and Bugey-3 data.	23
2.11	Allowed region for MiniBooNE with best fit point, overlaying the allowed LSND region. The KARMEN and OPERA 90% confidence levels are also shown.	24
2.12	Rate results from the SAGE and GALLEX gallium experiments.	25
2.13	The reactor neutrino anomaly.	26
2.14	Observed spectra for the DANSS (left) and NEOS (right) experiments	27

2.15	Global fits to appearance and disappearance data.	28
2.16	Global allowed region for appearance results.	29
3.1	Aerial view of Fermilab beamlines.	32
3.2	Graphical representation of the Booster Neutrino Beam.	33
3.3	Predicted Booster Neutrino Beam flux at MicroBooNE.	34
3.4	Scintillation light in liquid argon.	39
3.5	MicroBooNE time projection chamber diagram.	40
3.6	Simulated baseline-subtracted TPC signals for ideal tracks.	42
3.7	MicroBooNE Run 1 event displays.	43
3.8	MicroBooNE electronics schematic.	44
3.9	Start/end points of tracks tagged by an external muon counter and reconstructed in the MicroBooNE TPC.	46
3.10	PMT schematic (3.10a) and quantum efficiency (3.10b).	47
3.11	BNB Performance.	51
4.1	HARP π^+ production data compared with the best fit Sanford-Wang prediction. . .	56
4.2	E910 π^+ production data compared with the best fit Sanford-Wang prediction . . .	57
4.3	Feynman scaling fit to 8 experiments which have been scaled to $p_T = 8.89$ GeV/c. .	58
4.4	Example of the fitting procedure used to extract the data-driven noise model. . . .	62
4.5	Time- and frequency-domain waveforms for the MicroBooNE data-driven noise model.	62
4.6	MicroBooNE noise level, measured in RMS of the ADC values, as a function of channel number.	64
4.7	Demonstration of induced charge effects.	65
4.8	Example of a one-dimensional deconvolution compared to a two-dimensional deconvolution for a single data event.	67
4.9	Example simulated ParticleFlow Object which contains a PFParticle hierarchy. . .	69

5.1	Diffusion theory plot.	77
5.2	Averaged waveforms as a function of drift distance from MicroBooNE data	78
5.3	Electron energy as a function of electric field.	79
5.4	World data for electron drift velocity and a fit to world data for electron mobility.	80
5.5	Measurement of D_L from ICARUS.	80
5.6	Single and summed deconvolved waveforms.	81
5.7	Diagram showing the necessity for t_0 -tagging	83
5.8	Hit goodness of fit.	84
5.9	Pulse height versus drift bin number for a cosmic data sample and a CORSIKA Monte Carlo sample. This contains only collection plane hits.	85
5.10	σ^2 versus bin number pre- (a) and post- (b) dynamic sigma cut for a MicroBooNE cosmic simulation	86
5.11	Illustration of the method used to merge waveforms.	87
5.12	Example collection plane event display using deconvolved signal information showing the forward-going muon sample.	90
5.13	Fractional diffusion value as a function of the number of drift bins used.	91
5.14	Fractional diffusion value as a function of the dynamic σ cut value.	92
5.15	Measured D_L values for many different input D_L values, along with associated error.	93
5.16	Measured D_L with noise turned on and with noise and Geant4 level physics processes turned on.	94
5.17	Fractional diffusion value as a function of the number of bins used, and the σ cut value for a simulated cosmic sample.	96
5.18	θ_{xz} and θ_{yz}	97
5.19	Angular distribution of reconstructed tracks with a length greater than 50 cm, and the subset of those tracks which have a reconstructed t_0	98

5.20	Plots taken from the MicroBooNE t_0 -tagging public note showing the t_0 reconstruction accuracy for anode and cathode piercing tracks.	99
5.21	Measured D_L for a cosmic sample with input $D_L = 6.36 \text{ cm}^2/\text{s}$, and $D_T = 0.0 \text{ cm}^2/\text{s}$	100
5.22	Illustration of the effects of transverse diffusion.	101
5.23	Percentage bias in measured D_L values in θ_{xz} - θ_{yz} space, and projected onto each axis.	103
5.24	Extracted diffusion value for a full cosmic MicroBooNE simulation, using the nominal input diffusion values of $D_L = 6.2 \text{ cm}^2/\text{s}$ and $D_T = 16.3 \text{ cm}^2/\text{s}$	104
5.25	Extracted diffusion value from the MicroBooNE Run 1 data. In order for a point to be considered in the linear fit each bin must have a minimum of 500 waveforms.	105
6.1	Mass stopping power ($=\langle dE/dx \rangle$) for μ^+ on copper, as a function of $\beta\gamma = p/Mc$	112
6.2	Energy deposition theory curves	113
6.3	Electron versus photon particle identification is performed by making use of the fact that photons pair produce two electrons, meaning the dE/dx is approximately double that of a single electron.	114
6.4	Measured Widths, Simulation	118
6.5	Measured Widths, Data	119
6.6	Measured Widths, Smeared Data	119
6.7	Agreement between data and simulation before and after application of the simulation tuning.	120
6.8	Template fit deposited energy versus energy by range plots, μ assumption	122
6.9	Template fit PIDa algorithm plots	125
6.10	Template fit $\chi_\mu^2 - \chi_p^2$ plots	127
6.11	Particle probability maps, Y plane	129
6.12	Track end point resolution	130
6.13	End point resolution shift	131
6.14	Template fit bare likelihoods for muons, pions, protons, kaons, and MIPs	132

6.15	Template fit bare likelihoods for muons, pions, protons, kaons, and MIPs, normalised from 0 to 1	134
6.16	Template fit $\ln(L_{MIP}/L_p)$ likelihood ratios	135
6.17	Efficiency/purity curves for muons, protons, and pions	141
6.18	Venn diagrams overlap	142
6.19	A comparison of data and simulation for tracks between 0 and 2.5 cm.	143
6.20	A comparison of data and simulation for tracks with $\theta_x < 30$ degrees.	143
6.21	Proton and muon selection efficiencies and purities for the $\ln(L_{MIP}/L_p)$ algorithm as a function of track length and θ_x	144
7.1	Primary ν_μ CC backgrounds to the selection of the signal are shown in the red box, where a blue line represents a muon, a red line represents a proton, and a green line or triangle represents a charged or neutral pion, respectively.	146
7.2	Definition of the fiducial volume for the ν_μ CC-inclusive event selection.	156
7.3	The efficiency of selecting ν_μ CC-inclusive events as a function of neutrino energy, separated by interaction type.	157
7.4	Data/simulation comparisons for the number of reconstructed PFParticles compared to the number of reconstructed tracks and showers.	159
7.5	Number of events selected by the CC-inclusive selection in the CV dataset and in the DIC dataset	161
7.6	Fraction of muons reconstructed as tracks versus showers	161
7.7	Fraction of protons reconstructed as tracks versus showers	162
7.8	Fraction of pions reconstructed as tracks versus showers	162
7.9	Fraction of electrons reconstructed as tracks versus showers	163
7.10	$\ln(L_{MIP}/L_p)$ variable for showers built as tracks.	164
7.11	Distributions of the MIP/proton separator using only tracks identified as such by Pandora (left) and including in addition the showers built as tracks (right).	165
7.12	$E_{dep}^Y - E_{range}$ for selected contained muon candidates and proton candidates.	168

7.13	PID distribution, post selection for the nominal and DIC simulations.	169
7.14	dE/dx versus residual range for uncontained and contained muons, and for leading and non-leading protons from simulation.	170
7.15	dE/dx versus residual range for uncontained and contained muons, and for leading and non-leading protons from on-beam data.	170
7.16	dE/dx versus residual range for uncontained and contained muons, and for leading and non-leading protons from off-beam data.	171
7.17	Efficiency of the selection as a function of E_v^{true} and p_μ^{true}	172
7.18	Candidate muon track ϕ distributions for the nominal simulation (left) and the DIC simulation (right) using the tight topology cuts.	173
7.19	Candidate muon track ϕ distributions for the nominal simulation (left) and the DIC simulation (right) using the loose topology cuts.	174
7.20	Reconstructed neutrino vertex positions in the x , y and z dimensions within the MicroBooNE TPC.	175
7.21	Kinematic distributions for the selected muon candidate.	177
7.22	Kinematic distributions for the selected proton candidates.	178
7.23	Fits for energy resolution of uncontained muons as a function of energy.	179
7.24	Resolution (left) and bias (right) for uncontained muons	180
7.25	Fits for energy resolution of contained muons as a function of energy.	180
7.26	Resolution (left) and bias (right) for contained muons	181
7.27	Fits for energy resolution of protons as a function of energy.	181
7.28	Resolution (left) and bias (right) for protons	182
7.29	Total deposited energy in the TPC before (left) and after (right) calibrating.	183
7.30	True neutrino energy compared with the energy estimator used in this analysis.	183
7.31	Distribution of systematic universes from the MicroBooNE flux simulation. The green histogram is the central value histogram, while the colour map shows the distribution of systematic universes in each bin.	185

7.32	Covariance matrices for flux-related systematic uncertainties.	186
7.33	Distribution of systematic universes from the GENIE generator. The blue histogram is the central value histogram, while the colour map shows the distribution of systematic universes in each bin.	187
7.34	Covariance matrices for interaction-related systematic uncertainties.	189
7.35	Effect of detector-related systematic uncertainties on the reconstructed neutrino energy spectrum.	190
7.36	Covariance matrices for detector-related systematic uncertainties.	192
7.37	Data-driven track fraction plots for on-beam data and for the CV dataset with detector-related systematic uncertainties included (green, hashed).	192
7.38	Combined covariance matrices for flux, interaction, and detector systematic uncertainties.	193
7.39	Distribution of reconstructed neutrino energy with statistical uncertainty shown in dark green and stat. + syst. uncertainty shown in light green. The black histogram includes the off-beam data as well as signal and backgrounds from the simulation. The black data points are the selected on-beam data events.	193
8.1	Selected events from the electron neutrino selection.	198
8.2	Full statistical plus systematic covariance matrix showing the full flux and cross section covariance between the muon neutrino sample and the electron neutrino sample. The first 8 bins correspond to the electron neutrino sample and the remaining 10 bins correspond to the muon neutrino sample.	200
8.3	Total stat. + syst. uncertainties on the electron neutrino selection before and after applying the muon neutrino constraint.	200
8.4	Results of performing frequentist studies for the case in which 0, 1, 2 and 5 bins of ν_μ bins are included in the fit.	203
8.5	Significance with which H_0 can be rejected in favour of	204

A.1	dE/dx versus residual range for two-track selected events	221
A.2	dE/dx distribution for the shorter of two-track selected events	221
A.3	POT normalised bare χ^2 for muons, pions, protons, and kaons	223
A.4	Template fit bare χ^2 for muons, pions, protons, and kaons	224
A.5	POT normalised bare likelihoods for muons, pions, protons, kaons, and MIPs . . .	226
A.6	POT normalised likelihoods for muons, pions, protons, kaons, and MIPs, nor- malised from 0 to 1	227
A.7	POT normalised $\ln(L_{MIP}/L_p)$ likelihood ratio	228
A.8	POT normalised $\Delta\chi_{\mu-p}^2$ variable	229
A.9	POT normalised PIDA mean algorithm	230
A.10	POT normalised deposited energy versus energy by range, μ assumption	231
B.1	Cartoon showing the difficulties of matching reconstructed particles to true par- ticles.	234
B.2	Reconstructed neutrino vertex positions in the x, y and z dimensions within the MicroBooNE TPC.	235
B.3	Kinematic distributions for the selected muon candidate.	236
B.4	Kinematic distributions for the selected proton candidates.	237

List of Tables

2.1	Particle generations in the SM	9
2.2	The global picture of neutrino oscillation parameters under the 3 neutrino assumption from the NuFit collaboration.	15
3.1	Predicted beam composition of the BNB in neutrino running mode.	35
3.2	Important variables related to the MicroBooNE TPC, light collection system and field cage	36
3.3	Run 1 misconfigured regions.	45
5.1	Outline of the datasets used in the validation of the longitudinal electron diffusion analysis.	89
5.2	Measured D_L values for many different input D_L values.	93
5.3	Table summarising nominal cut values and other numbers of interest for this analysis.	94
5.4	This table shows the total passing rate for tracks in the Monte Carlo as a function of the θ_{xz} and θ_{yz} cut.	98
5.5	Cut values used to select out tracks for the diffusion analysis.	102
5.6	Expected maximum and minimum electric field values from space charge effect and the predicted D_L value.	108
5.7	An outline of the bounds on systematic uncertainties which have been placed so far.	108

6.1	Measured widths from MicroBooNE simulation	117
6.2	Measured widths from MicroBooNE Data	117
6.3	Measured widths from MicroBooNE Smeared Simulation	118
6.4	PID development sample information	121
6.5	Muon and proton efficiencies and purities summary	137
6.6	PID data/simulation agreement summary	138
7.1	Datasets used in the development of the ν_μ CC0 π N p selection.	151
7.2	Summary of the fiducial volume used in the geometric cosmic tagging outlined in the text.	153
7.3	Cut flow table, showing the number of selected on-beam events compared with the predicted number of events, as well as the selection efficiency and purity at each stage.	172
7.4	Neutrino interaction model parameters and associated uncertainties taken from the GENIE generator.	188
7.5	Effect of detector-related systematic uncertainties on the number of selected sig- nal candidates	191
8.1	Table showing the fractional error in each electron neutrino bin, split into sta- tistical error, systematic error, and total error before and after application of the muon neutrino constraint.	201
8.2	Binning used in the frequentist studies for the ν_μ and ν_e samples.	202

Chapter 1

Introduction

“The story so far: In the beginning the Universe was created. This has made a lot of people very angry and been widely regarded as a bad move.”

Douglas Adams

The fact that neutrinos oscillate between different flavours is one of the only pieces of evidence we have for physics beyond the Standard Model of particle physics, and this alone would make neutrino physics an area worthy of study. In addition to this, we do not know whether neutrinos and anti-neutrinos oscillate differently in vacuum; if they do then this would indicate that the charge-parity symmetry is violated in the leptonic sector, which would help to fulfill one of the Sakharov conditions for baryogenesis [1], one of the leading theories for the matter/anti-matter asymmetry of the universe. We do not know the absolute scale of neutrino masses, and further, we do not understand the mechanism by which the masses of neutrinos are generated. We do not know whether more neutrino species exist in addition to the three that we know of, and if they do then to what level they can help to explain dark matter in the universe. Neutrinos can also be used to study nuclear physics effects, which in many cases is not well understood in neutrino interactions. Suffice it to say that neutrinos are an exciting probe of fundamental physics.

1.1 Liquid Argon Time Projection Chambers: Precise and Scalable

The field of neutrino physics has expanded rapidly since the Reines-Cowan experiment of the 1950s [2, 3]. This is demonstrated in Figure 1.1, where the neutrino vertex resolution of various detectors is shown as a function of the year each detector was built. In this figure, the experiments are separated by type of detector, and the size of each circle scales with the size of the detector.

Neutrino physics has matured to a point where there is a desire for detectors to be both high-precision and large in scale. This has not always been the case: initial experiments searching for the existence of the neutrino were essentially event-counting in nature meaning that they were not required to be precise. After the discovery phase, high-precision bubble chamber detectors were used to study neutrino interactions and shortly after this, with the discovery that neutrinos oscillate from one flavour to another, precision began to be sacrificed in order to build large-scale detectors which maximised event rate.

It is clear from Figure 1.1 that Liquid Argon Time Projection Chambers (LArTPCs, such as ArgoNeuT [4], ICARUS [5], MicroBooNE [6], SBND [7], and DUNE [8]) strike an important balance: they are able to attain excellent vertex resolution, allowing precision cross section physics to take place, while remaining scalable for use in oscillation physics. Importantly the output of such detectors is digital, allowing for automated reconstruction to take place, unlike in the days of the bubble chamber.

With the Deep Underground Neutrino Experiment (DUNE) on the horizon, and LArTPCs still being a relatively new technology, they have naturally become a focus of significant research and development. Specifically, the Short Baseline Neutrino (SBN) program [7] at Fermilab has provided invaluable experience through the MicroBooNE experiment, and will continue to do so when the near and far detectors come online in the coming years. The new protoDUNE detectors [9] will provide additional input, given that these important prototypes include significant advances after incorporating what has been learned from other liquid argon detectors such as

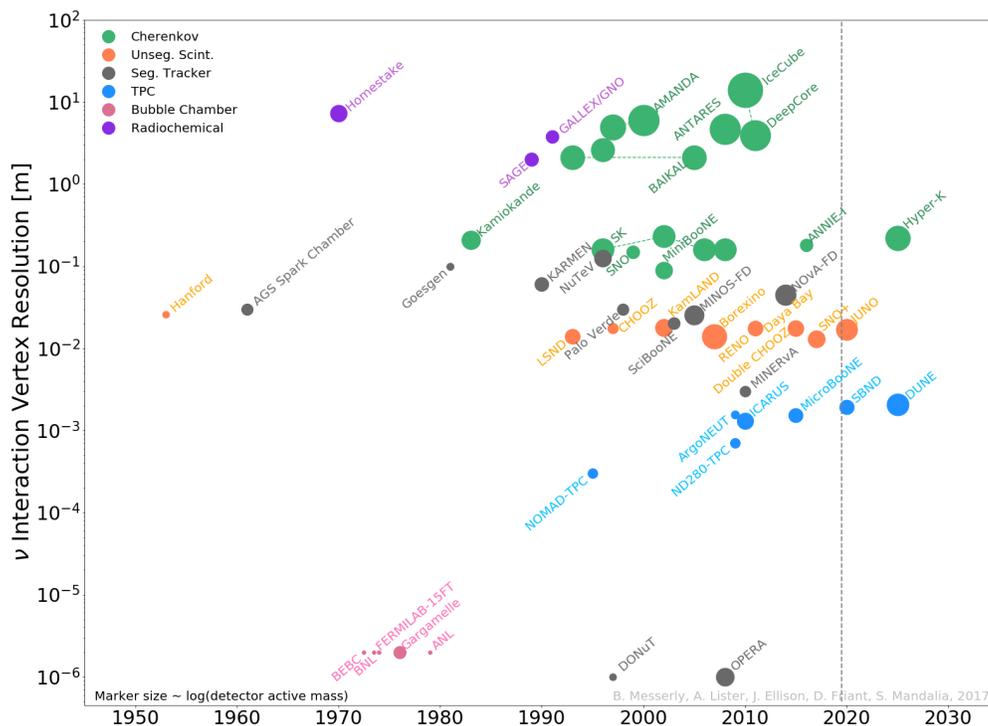


Figure 1.1: Neutrino detectors presented as a function of year built and neutrino interaction vertex resolution. The detectors are roughly separated by the technology they leveraged. In some cases detectors fall into multiple categories. Such is the case for MiniBooNE, where both scintillation light and Cherenkov light were used by the experiment. Where experiments have had multiple runs in different configurations, a dashed line connects the runs, such as in the case of IceCube and its upgrade to use more finely spaced digital optical modules in the DeepCore extension. The vertical grey line indicates the year 2019.

MicroBooNE, ICARUS, ArgoNeut, and LArIAT [10].

1.2 Sterile Neutrinos: A Known Unknown

One of the known unknowns of neutrino physics is whether additional neutrino states exist, and whether any of these additional states are *sterile* in nature: “do these hypothetical neutrinos feel the *weak nuclear force*?”. This is not a purely theoretical question. Several experiments have claimed

observations which could be described by non-standard neutrino oscillations with the inclusion of additional, heavier, sterile states. One such result is the excess of low-energy ν_e candidates observed by the MiniBooNE experiment [11].

1.3 Probing the Sterile Neutrino Question with LArTPCs

The Micro Booster Neutrino Experiment (MicroBooNE) is a LArTPC which has been built in order to directly probe the MiniBooNE result. There are myriad reasons why this technology is an excellent choice for this specific task, however it is still in its infancy compared to more traditional detection techniques. As such, the goals of this work very much mirror the goals of the MicroBooNE experiment: analysis techniques have been developed in order to constrain systematic uncertainties in a search for a MiniBooNE-like excess of ν_e events, and work has been undertaken to characterise detector effects specific to LArTPCs.

1.4 Thesis Outline

Chapter 2 outlines our current understanding of neutrino physics with a specific focus on neutrino oscillations. This chapter also includes a review of some anomalous results from the last several decades, which when taken together might act to indicate that additional neutrino species exist in addition to the three that we know about.

In Chapter 3, a review of the MicroBooNE experiment is presented. This includes information about the beamline, the detector, and the data-taking triggers which are used.

Chapter 4 describes the main parts of the simulation, reconstruction, and calibrations that are used in MicroBooNE data analyses.

Chapter 5 contains the original work performed in order to measure the longitudinal ionisation electron diffusion from the MicroBooNE data.

The entirety of Chapter 6 is dedicated to a description of particle identification in MicroBooNE. This includes a description of a tuning applied to the simulation in order to compare it

to the MicroBooNE data, and development of a new particle identification algorithm, the Bragg Likelihood algorithm. Both the tuning and Bragg Likelihood algorithm have been developed as part of this work.

Chapter 7 outlines the original work carried out to develop a selection of ν_μ charged-current events with zero pions and at least one proton in the final state.

Chapter 8 describes the constraint of systematic uncertainties in a ν_e search using the ν_μ selection developed in Chapter 7.

Finally, Chapter 9 concludes this work.

Chapter 2

Neutrino Physics

*“Oh, it’s better not to think about this at all,
like new taxes.”*

Peter Debye

2.1 Dear Radioactive Ladies and Gentlemen...

The neutrino was introduced to the scientific community as a “desperate remedy”. By 1930, it had been shown by Chadwick that electrons emitted during β decay had a spectrum of energies rather than the well defined energies seen in α and γ decay [12]. Faced with this, many were ready to abandon the principle of conservation of energy, however Wolfgang Pauli realised that if an undetectable particle was introduced, then this could carry a portion of the energy released [13], giving rise to the observed distribution in electron energies. Thus the idea of the neutrino was born.

Pauli was initially skeptical of the idea, despite having conceived of it, writing

I do not dare to publish anything about this idea, and trustfully turn first to you, dear radioactive people, with the question of how likely it is to find experimental evidence for such a [neutrino]

and the idea was not considered seriously until Enrico Fermi unified Pauli's neutrino and Dirac's positron into a single theory [14]. This was quickly followed by the first calculation of the neutrino cross section by Bethe and Peierls [15], confirming Pauli's fears that the neutrino would be almost undetectable with the technology of the time.

It would not be until the 1950s when Reines and Cowan were able to claim experimental evidence for the "free neutrino" [2, 3]. Time would reveal that this was in fact the discovery of the electron neutrino, ν_e , when the muon neutrino, ν_μ , was discovered by Leon M. Lederman, Melvin Schwartz and Jack Steinberger in 1962 [16], for which they were awarded the 1988 Nobel Prize¹. Given that these two particles were associated with charged leptons (e and μ , respectively), it was long hypothesised that there should be a third flavour of neutrino corresponding to the τ lepton. It was not until the year 2000 that this was experimentally verified by the DONUT collaboration [17].

The ν_τ was the final lepton of the Standard Model of particle physics (SM) to be experimentally confirmed, and the second-to-last particle in total, only being succeeded in 2012 by the discovery of the Higgs boson [18].

2.2 The Standard Model Neutrino

The SM describes within the framework of quantum field theory the unification of three of the four fundamental forces of nature: the weak nuclear force, the strong nuclear force, and electromagnetism. Efforts to include the final fundamental force, gravity, have thus far been unsuccessful. Despite this, it is a monumentally successful scientific theory and has been experimentally tested and verified throughout the twentieth and twenty-first centuries.

The SM is based upon a $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge group, with C , L , and Y denoting colour charge, left-handed chirality, and the weak hypercharge, respectively. It describes the strong, weak, and electromagnetic interactions of fermionic particles by mediation of the relevant spin-1 gauge bosons. After Electroweak Symmetry Breaking (ESB), this gauge group is modified:

¹Frederick Reines would later be given the 1995 Nobel Prize for his work on neutrino physics.

$$SU(3)_C \times SU(2)_L \times U(1)_Y \xrightarrow{\text{ESB}} SU(3)_{\text{colour}} \times SU(2)_{\text{QED}} \quad (2.1)$$

Strong interactions, mediated by one of 8 gluons (g), come from the $SU(3)_{\text{colour}}$ part of the gauge group, and affect only quarks. Weak and electromagnetic interactions are mediated by the charged or neutral weak bosons (W^\pm, Z^0), or the photon (γ), and are governed by the $SU(2)_{\text{QED}}$ term.

Weak interactions are described within the electroweak part of the SM, only act upon particles with a left-handed chirality. Mathematically this is equivalent to the statement that left-handed particles transform as a doublet under $SU(2)_L \times U(1)_Y$, while right handed particles transform as a singlet,

$$Q_L = \begin{pmatrix} q_u \\ q_d \end{pmatrix}_L, \quad L_L = \begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix}_L, \quad q_{uR}, \quad q_{dR}, \quad \ell_R$$

where q_u is up type quarks, q_d is down type quarks, ℓ is charged leptons of a given flavour, and ν_ℓ is the corresponding neutrino. These particles can be grouped into three generations, each of which have the same quantum numbers but differing masses. These are outlined in Table 2.1.

	Generation I	Generation II	Generation III
Quarks	$\begin{pmatrix} u \\ d \end{pmatrix}$	$\begin{pmatrix} c \\ s \end{pmatrix}$	$\begin{pmatrix} t \\ b \end{pmatrix}$
Leptons	$\begin{pmatrix} e \\ \nu_e \end{pmatrix}$	$\begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}$	$\begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}$

Table 2.1: Table showing the three generations of particles contained within the SM. The generations have the same quantum numbers and are only differentiated by their masses.

Of note is that any right-handed neutrinos which might be introduced would not interact via the weak nuclear force (as they are a singlet under $SU(2)_L \times U(1)_Y$) or the electromagnetic force (because they are electrically neutral), and so they would not be observable. Due to this, no

right-handed neutrino component is included in the SM. One consequence of this is that a mass term cannot be constructed in the same way as for other fermions in the SM Lagrangian²,

$$\mathcal{L}_{mass} = m(\psi_L^\dagger \psi_R - \psi_R^\dagger \psi_L).$$

For this reason, neutrinos are assumed to be massless within the SM, although neutrino masses are a common feature of most extensions to the SM.

2.3 Neutrino Interactions

There are two terms of interest for neutrino interactions in the SM Lagrangian,

$$\mathcal{L}_\nu^{CC} = -\frac{ig}{\sqrt{2}} \sum_{\alpha=e,\mu,\tau} \bar{\nu}_{\alpha L} W \ell_{\alpha L} + h.c., \quad (2.2)$$

and

$$\mathcal{L}_\nu^{NC} = -\frac{ig}{2 \cos \theta_W} \sum_{\alpha=e,\mu,\tau} \bar{\nu}_{\alpha L} Z^0 \ell_{\alpha L} + h.c., \quad (2.3)$$

where \mathcal{L}_ν^{CC} represents charged-current (CC) interactions which occur with exchange of a W boson, and \mathcal{L}_ν^{NC} represents neutral-current (NC) interactions, which instead exchange a neutral Z boson [19]. These interaction terms can be represented as Feynman diagrams, as shown in Figure 2.1.

In these Lagrangian terms, g denotes the weak coupling constant, θ_W is the Weinberg angle, $h.c.$ is shorthand for the hermitean conjugate, and the Feynman slash notation³ has been used.

In addition to separating out neutrino interactions by whether they are charged or neutral current, GeV-scale interactions are often broken into three distinct types which can be separated by the energy regime in which they operate. The three main categories of interaction are quasi-elastic (QE), resonant (RES), and deep inelastic (DIS), and examples of each of these are shown in

²The \dagger here refers to the hermitean conjugate, or conjugate transpose of the field.

³ $A = \gamma^\mu A_\mu$

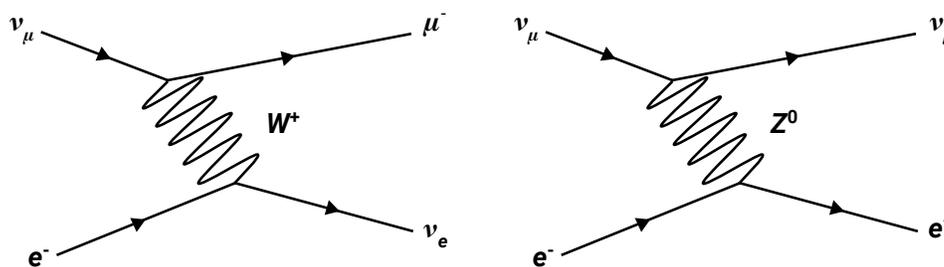


Figure 2.1: Charged-current and neutral-current interaction Feynman diagrams.

Figure 2.2.

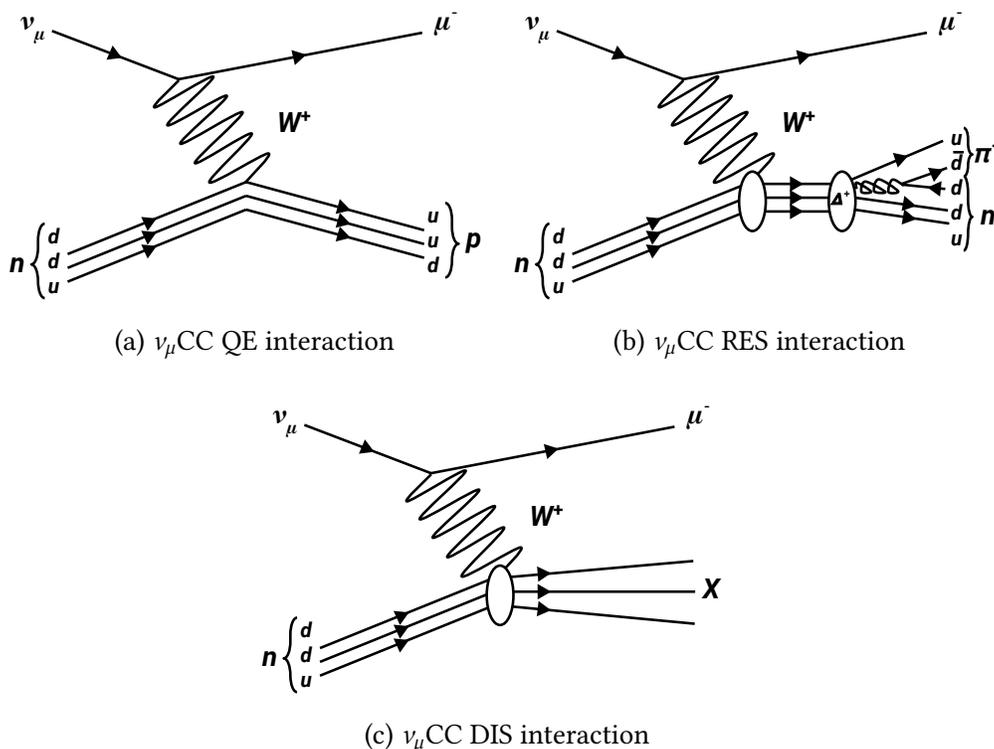


Figure 2.2: Diagrams showing examples of ν_μ CC QE, RES, and DIS interactions.

Quasi-Elastic Interactions

Quasi-elastic interactions occur when an incoming neutrino interacts with a nucleon, modifying the flavour of one of the constituent quarks, and often knocking the nucleon out of the nucleus. An example for this type of interaction is shown in Figure 2.2a, where an incoming ν_μ interacts

with a neutron, changing the flavour of one quark from d to u , and producing a μ^- and a proton in the final state.

Resonant Interactions

Resonant events occur when the incoming neutrino interacts with a nucleon, exciting it into a resonant state, which then decays into some final state, most often to a nucleon and a single pion, as shown in Figure 2.2b.

Deep Inelastic Interactions

Deep inelastic interactions happen when the neutrino imparts a lot of energy to a quark in a nucleon, causing the breakup of the nucleon and a shower of hadronic particles to emerge from the interaction point. This is shown in Figure 2.2c.

The predicted ν_μ CC neutrino cross section is shown as a function of energy in Figure 2.3. Here, the three components are shown separately along with a total predicted cross section and neutrino cross section measurements from a large number of experiments.

2.4 Neutrino Oscillations

During the 1960s, neutrino observatories with a focus of detecting neutrinos originating from the Sun stumbled upon a problem: the number of neutrinos detected were in direct contradiction to models of thermonuclear fusion in standard solar models which had been validated by results in helioseismology [21, 22]. This problem was further aggravated when a similar discrepancy was found in measurements of atmospheric neutrinos [23].

Inspired by a similar process in the neutral kaon system ($K^0 \rightleftharpoons \bar{K}^0$, see reference [24]), Bruno Pontecorvo began to formulate a solution to this problem by assuming neutrinos could change from one flavour to another⁴ through a process dubbed *neutrino oscillations*. Oscillations from

⁴Pontecorvo initially developed this idea through $\nu \rightleftharpoons \bar{\nu}$ oscillations, with flavour oscillations coming slightly

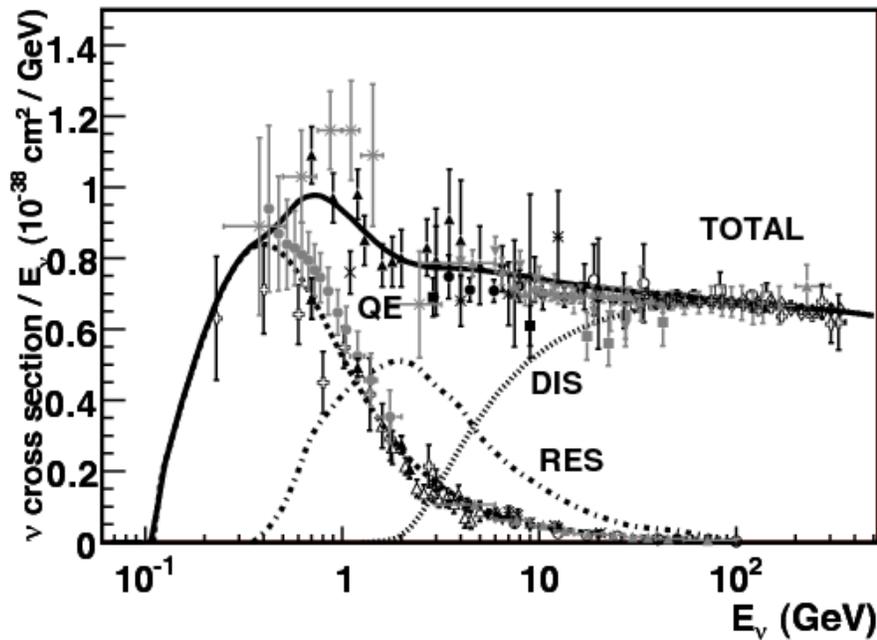


Figure 2.3: Predicted cross sections as a function of energy ν_μ CC for quasi-elastic, resonant, and deep inelastic interactions taken from [20], overlaid with data from a wide range of experiments.

electron neutrinos into either muon or tau neutrinos would then explain the deficit in observed electron neutrinos at neutrino observatories. This idea was further developed by Maki, Nakagawa and Sakata in 1962 [25], and given a standard derivation by three pairs — Fritzsche and Minkowski [26], Eliezer and Swift [27], and Bilenky and Pontecorvo [28].

In the standard formalism, the neutrino flavour states can be represented as a linear superposition of mass states, and equivalently (assuming the unitarity of the mixing matrix, U) the mass states can be thought of as a linear superposition of flavour states,

$$|v_\alpha\rangle = \sum_j U_{\alpha j}^* |v_j\rangle, \quad |v_j\rangle = \sum_\alpha U_{\alpha j} |v_\alpha\rangle. \quad (2.4)$$

Here, and in the rest of this chapter, Greek indicies (α, β) are used to represent the neutrino flavour states, while Latin indicies (i, j) are used to represent neutrino mass states, and U repre-

later.

sents the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix,

$$U_{\alpha j} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix}, \quad (2.5)$$

which can be parameterised as three two-dimensional rotation matrices. When this parameterisation is used, three *mixing angles* are introduced which describe the mixing between the three neutrino mass states. The three rotation matrices are often referred to as the atmospheric (θ_{23}), reactor (θ_{13}), and solar (θ_{12}) mixing angles. Using $s_{ij} = \sin(\theta_{ij})$, $c_{ij} = \cos(\theta_{ij})$, $U_{\alpha j}$ may be written

$$U_{\alpha j} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13} e^{i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13} e^{-i\delta_{CP}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (2.6)$$

where the δ_{CP} parameter, located in the reactor sector, is an additional Charge-Parity (CP) phase. If $\delta_{CP} \neq 0, \pi$, then this indicates that the CP symmetry is violated in the leptonic sector⁵, and so neutrinos and anti-neutrino oscillations would be expected to oscillate differently in vacuum.

By making use of the time-dependent Schrödinger equation, it can be shown that the probability for a neutrino of flavour α to oscillate into a neutrino of flavour β in a vacuum is given by

$$P_{\nu_\alpha \rightarrow \nu_\beta} = \left| \sum_j U_{\beta j}^* U_{\alpha j} \exp \left(-1.27 i \frac{\Delta m_{j1}^2 L}{E} \right) \right|^2, \quad (2.7)$$

the full derivation for which can be found in reference [29]. Importantly, the magnitude of the neutrino mass-squared difference parameters (Δm_{21}^2 and Δm_{31}^2 , where $\Delta m_{j1}^2 = m_j^2 - m_1^2$) controls the frequency of the oscillation, while the magnitudes of the mixing angles (θ_{12} , θ_{23} , and θ_{13} ,

⁵The violation of CP in the leptonic sector is a particularly interesting open question in neutrino physics due to its relevance to baryogenesis. Baryogenesis is one of the leading hypotheses for the matter/anti-matter asymmetry of the universe, however in order for it to be valid several conditions must be met; these are known as the Sakharov conditions [1]. One such condition is that CP violating processes exist. It is known that CP violation exists in the quark sector, but not in the quantities necessary for baryogenesis to be valid, meaning leptonic CP violation would be an important discovery. Neutrino oscillations are the main method currently used to probe this question.

	Normal Ordering (best fit)		Inverted Ordering ($\Delta\chi^2 = 4.7$)	
	bf $\pm 1\sigma$	3σ range	bf $\pm 1\sigma$	3σ range
θ_{12}	$33.82^{+0.78}_{-0.76}$	$31.61 \rightarrow 36.27$	$33.82^{+0.78}_{-0.76}$	$31.61 \rightarrow 36.27$
θ_{23}	$49.6^{+1.0}_{-1.2}$	$40.3 \rightarrow 52.4$	$49.8^{+1.0}_{-1.1}$	$40.6 \rightarrow 52.5$
θ_{13}	$8.61^{+0.13}_{-0.13}$	$8.22 \rightarrow 8.99$	$8.65^{+0.13}_{-0.13}$	$8.27 \rightarrow 9.03$
δ_{CP}	215^{+40}_{-29}	$125 \rightarrow 392$	284^{+27}_{-29}	$196 \rightarrow 360$
$\frac{\Delta m_{21}^2}{10^{-5}\text{eV}^2}$	$7.39^{+0.21}_{-0.20}$	$6.79 \rightarrow 8.01$	$7.39^{+0.21}_{-0.20}$	$6.79 \rightarrow 8.01$
$\frac{\Delta m_{3\ell}^2}{10^{-3}\text{eV}^2}$	$2.525^{+0.033}_{-0.032}$	$+2.427 \rightarrow +2.625$	$-2.512^{+0.034}_{-0.032}$	$-2.611 \rightarrow -2.412$

Table 2.2: The global picture of neutrino oscillation parameters under the 3 neutrino assumption from the NuFit collaboration [37]. Note that this is shown without constraints from the Super-Kamiokande atmospheric data. Results are shown under the assumption of normal ordering and of inverted ordering, although there is a slight preference for normal ordering.

contained within the elements of U), control the magnitudes of the oscillations.

Experimental evidence for neutrino flavour oscillations was provided in the late 20th century by the Super-Kamiokande collaboration [30] and the Sudbury Neutrino Observatory collaboration [31], collaborators from which were awarded the 2015 Nobel Prize in Physics. Since then, the field has rapidly worked towards constraining the values of the mixing parameters.

The three-flavour neutrino paradigm has been probed by a number of experiments. In recent years, Daya Bay [32] was able to precisely measure θ_{13} , while results on the atmospheric mixing parameters (Δm_{32}^2 and θ_{23}) and δ_{CP} have been primarily from T2K [33], NO ν A [34], IceCube [35], and MINOS [36]. A global analysis of neutrino data has been performed by the NuFit collaboration [37], giving the current global best fit of the oscillation parameters which are found in Table 2.2.

It is clear from equation 2.7 that there are two parameters which are able to be tuned in neutrino experiments, the baseline, L , and the neutrino energy, E . In practice, these parameters always appear as L/E , and so this is the characteristic variable which is discussed in many neutrino oscillation experiments. Oscillation probabilities as a function of L/E are shown for $\nu_\mu \rightarrow \nu_e$, $\nu_\mu \rightarrow \nu_\mu$, and $\nu_\mu \rightarrow \nu_\tau$ transitions in Figure 2.4.

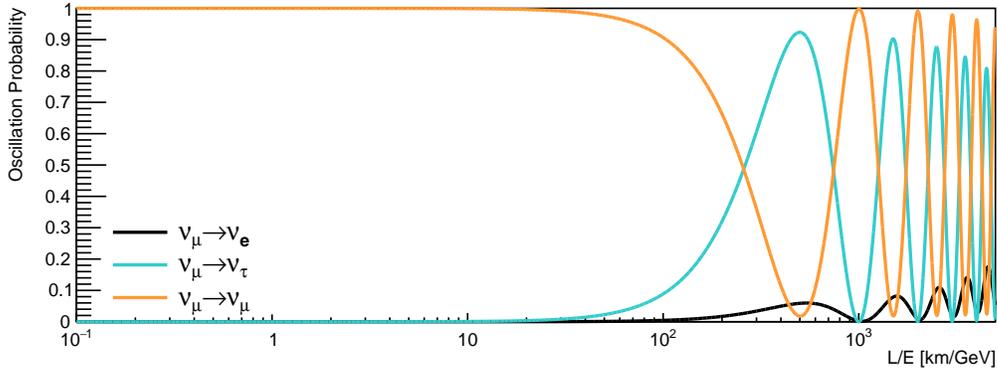


Figure 2.4: Neutrino oscillation probabilities for $\nu_\mu \rightarrow \nu_e$, $\nu_\mu \rightarrow \nu_\mu$, and $\nu_\mu \rightarrow \nu_\tau$ transitions. ν_μ is chosen as the initial flavour as experimental setups using a neutrino beam typically use an intense ν_μ source.

Figure 2.4 indicates that for L/E values relevant for neutrino beam experiments, the $\nu_\mu \rightarrow \nu_\tau$ transition is preferred to the $\nu_\mu \rightarrow \nu_e$ transition, however appearance searches⁶ mainly make use of ν_e appearance due to the high energies required for a ν_τ to produce the heavy τ lepton, and the short lifetime of such a heavy particle meaning that these interactions are often difficult to identify in a detector. To date, the ν_τ has only been observed by three collaborations: DONUT [17], OPERA [38], and more recently, IceCube [39].

An important consideration of the oscillation probability presented in Equation 2.7 is that it is unbounded from above, meaning in principle any number of neutrinos may be added. The number of light *active*⁷ neutrino flavour states is constrained by the width of the Z boson, measured at the Large Electron-Positron Collider (LEP) as described in reference [40], and shown in Figure 2.5. This, however, does not constrain neutrinos which do not undergo electroweak interactions – so-called *sterile neutrinos*.

⁶Neutrino oscillation analyses are characterised as either looking for *disappearance* of some fraction of the neutrino flavour produced by the neutrino source, or by *appearance* of a neutrino flavour different to that of the initial neutrinos produced at the neutrino source.

⁷Those neutrino flavours which undergo electroweak interactions.

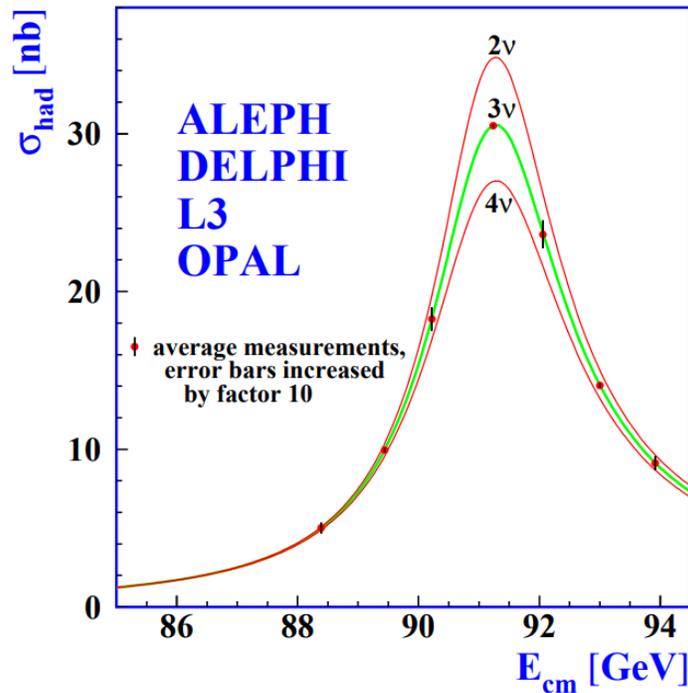


Figure 2.5: Measurements of the hadron production cross section near the Z resonance, taken from reference [40]. The three curves represent the prediction in the SM given the existence of 2, 3, or 4 light active neutrino flavours.

2.5 The Global Sterile Neutrino Picture

Sterile neutrinos are introduced in many extensions to the SM, as they are necessarily right-handed, meaning that a standard Dirac mass term can then be included, giving rise to neutrino masses. As a particle which would only interact with SM particles gravitationally and through neutrino oscillations, they are also often cited as a dark matter candidate.

Experimental results in the search for sterile neutrinos come from a variety of experimental setups, with different signal channels. Searches have been performed with reactor, accelerator, and atmospheric experiments, and for both charged-current and neutral-current interactions. These results are fraught with tension, with several collaborations reporting results consistent with sterile oscillations, and others reporting constraints with varying confidence limits.

This section will outline a non-exhaustive list of sterile neutrino search results from the late 20th and early 21st century.

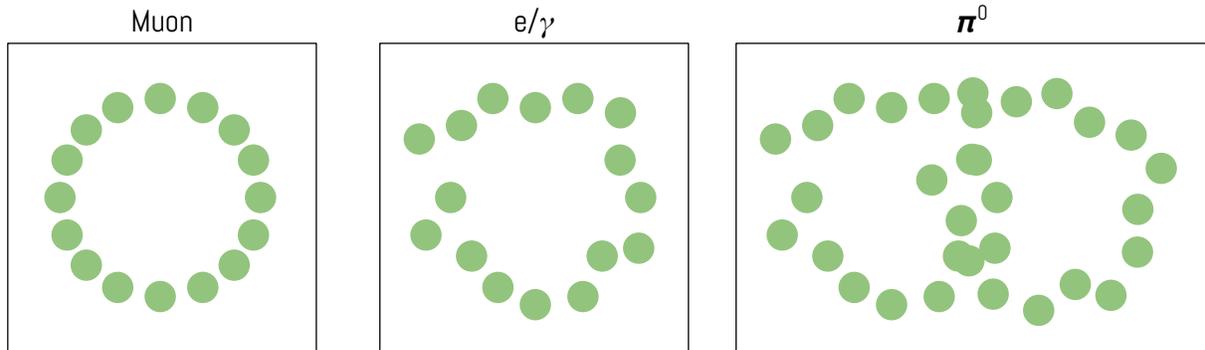


Figure 2.6: Image showing how particle identification is performed using Cherenkov light.

2.5.1 Results from Accelerator Experiments

LSND

The Liquid Scintillator Neutrino Detector (LSND) was a combined scintillation/Cherenkov detector which ran from 1993-1998 in the neutrino beam from the LAMPF proton accelerator at Los Alamos National Laboratory. For the main $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ analysis, LSND used neutrinos from μ^+ decay-at-rest (DAR), resulting in a well-understood spectrum of neutrino energies.

The Particle Identification (PID) in the detector was performed by making use of the shape of the cone of Cherenkov light produced by charged particles moving faster than the speed of light in the detector material. A muon passing through the detector will produce a stable ring, an electron or a photon will produce a more fuzzy ring, and a π^0 will produce two photon-like rings. This is shown in Figure 2.6. Note here that separating out electrons from photons is a very difficult task in detectors which use Cherenkov light for their PID. This is particularly a problem in the case that one photon from a π^0 decay exits the detector, leaving a single photon which may be mistaken for a ν_e candidate.

LSND observed an excess of $\bar{\nu}_e$ candidates in the data over the expected background at a level of 3.8σ [41, 42, 43], as shown in Figure 2.7. If this is interpreted as an excess from neutrino oscillations, then this corresponds to a Δm^2 of approximately 1 eV^2 . Given that, as shown in Table 2.2, global fits place the two known mass-squared splittings in the 10^{-5} eV^2 and 10^{-3} eV^2 regions,

this interpretation necessitates an additional neutrino mass state.

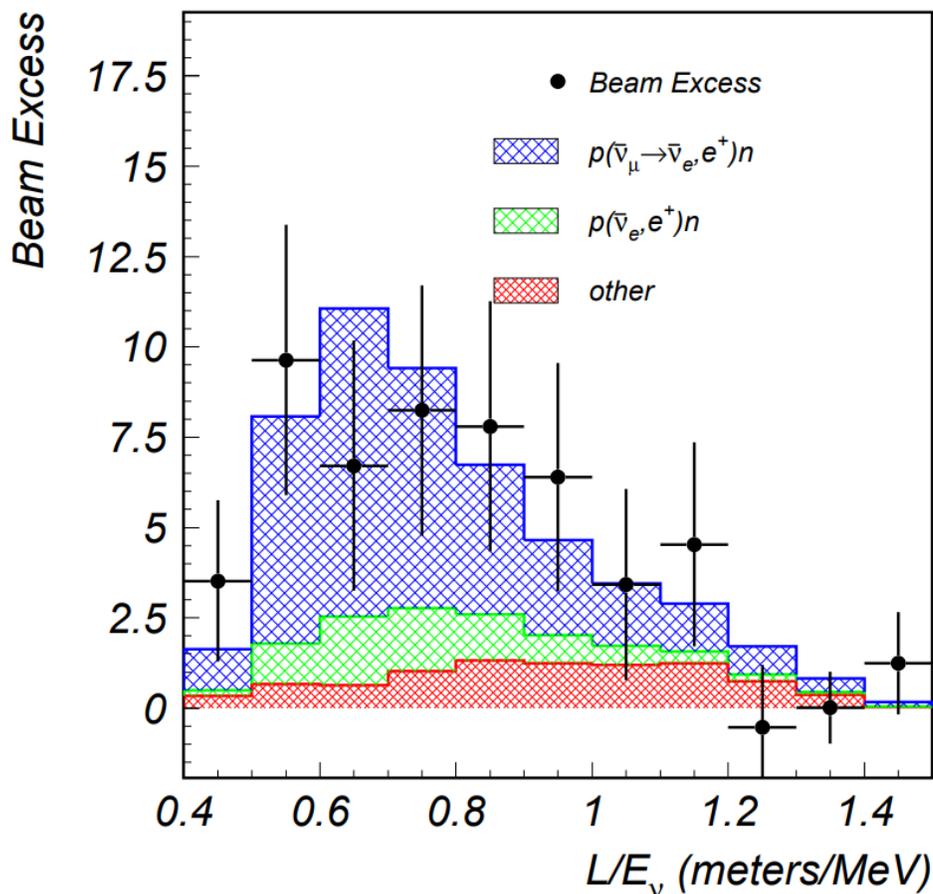


Figure 2.7: Beam excess (beam-off subtracted event count) as a function of L/E from the LSND collaboration [42]. Here the green and red shaded regions represent the expected backgrounds, while the blue shaded region represents assumed neutrino oscillations with a Δm^2 of $\sim 1 \text{ eV}^2$.

MiniBooNE

The Mini Booster Neutrino Experiment (MiniBooNE) experiment was designed to explicitly test the LSND result, and has been running in the Booster Neutrino Beam (BNB) at Fermi National Accelerator Laboratory (FNAL) since 2002. Importantly, the baseline and beam energy place this near the same L/E value as LSND, and so it is able to explore the same phase space as its predecessor. It, like LSND, is a combined scintillator/Cherenkov detector, allowing for the same

style PID method.

MiniBooNE also saw an excess of data over the predicted backgrounds (4.7σ , [44, 11]), although at a slightly lower energy, meaning that there is slight tension between the results of the two experiments (although this tension is somewhat reduced with the expanded dataset presented in reference [11]). Figure 2.8 shows the excess of events as a function of reconstructed neutrino energy in both neutrino (2.8a) and anti-neutrino mode (2.8b).

The best fit points in Figure 2.8 are independent fits to the neutrino and anti-neutrino data under the 2ν oscillation assumption, and it is clear that the anti-neutrino fit is significantly better than the neutrino mode fit.

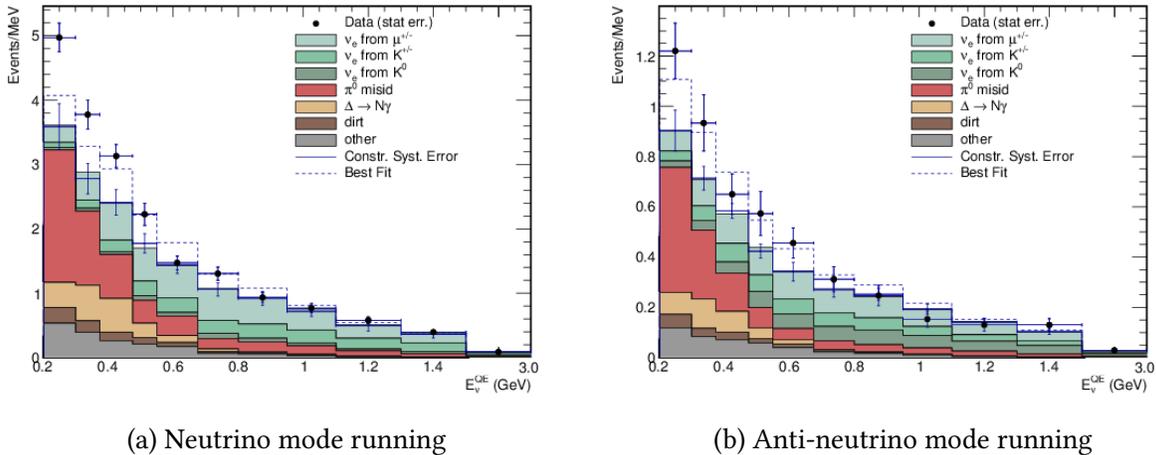


Figure 2.8: MiniBooNE results with expanded running (12.84×10^{20} POT in neutrino running mode, and 11.27×10^{20} POT in anti-neutrino running mode). Taken from reference [11].

Both neutrino and anti-neutrino mode data have significant backgrounds at low energy mostly due to π^0 mis-identification and $\Delta \rightarrow N\gamma$ decay. Both backgrounds here are photon-like. Given that the Cherenkov-style PID cannot easily differentiate these from electrons, this is something which raises suspicion. An open question from MiniBooNE is whether the observed excess is electron-like or photon-like.

KARMEN

The KARMEN experiment performed a $\bar{\nu}_e$ search in a $\bar{\nu}_\mu$ beam produced by the ISIS synchrotron at the Rutherford Appleton Laboratory in the United Kingdom. Like LSND, it was a scintillator detector, but unlike LSND, it was segmented. KARMEN is particularly interesting because it, like LSND, used muon DAR as a source of muon neutrinos.

KARMEN observed 15 candidate $\bar{\nu}_e$ events, which is consistent with the 15.8 events expected from background [45]. Performing a likelihood fit to the data results in a 90% confidence interval which is shown in Figure 2.11. It is notable that the two data sets are compatible assuming a Δm^2 of between 0.2 and 1.0 eV^2 , or a Δm^2 near to 7 eV^2 .

Null Results

In addition to the KARMEN null results, several other experiments have reported null results. ICARUS [46] and NOMAD [47] have set bounds by looking for ν_e appearance in a ν_μ beam, while null results in the ν_μ disappearance channel have been found by MiniBooNE and SciBooNE [48, 49], MINOS [50], MINOS and Daya Bay [51], CCFR [52], and IceCube [35].

Results from IceCube and MINOS/Daya Bay are presented in Figures 2.9 and 2.10, respectively.

Allowed Phase Space

If the LSND and MiniBooNE results are to be interpreted as neutrino oscillations, then the $\Delta m^2 \gg \Delta m_{31}^2, \Delta m_{21}^2$, and so the neutrino oscillation can, to a first approximation, be reduced to a two neutrino scenario:

$$P_{\nu_\alpha \rightarrow \nu_\beta} = \sin^2(2\theta) \sin^2\left(1.27 \times \Delta m^2 \times \frac{L}{E}\right), \quad (2.8)$$

and the data can be fit to extract the mixing angle, θ , and mass-squared splitting, Δm^2 . Because the baselines of these experiments are relatively short, matter effects can be neglected to first order. The resulting allowed region from this procedure is shown in Figure 2.11.

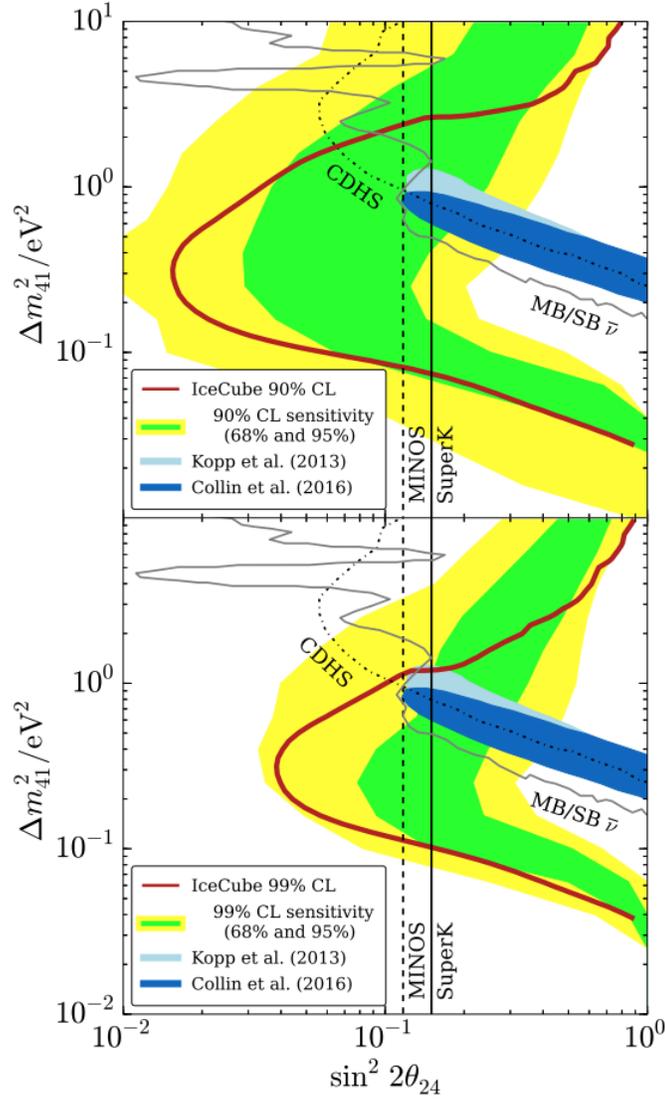


Figure 2.9: The IceCube 90% and 99% confidence levels, shown with the 99% and 68% CL sensitivity bands. The region to the right of the curve is excluded, meaning that IceCube is able to rule out the allowed MiniBooNE/SciBooNE region to 99% confidence. Figure taken from reference [35].

Here, the LSND 90% and 99% confidence level (CL) contours are shown shaded in blue and grey respectively, with the MiniBooNE allowed regions shown as solid lines overlaying this. In addition to this, the KARMEN 90% CL constraint is shown as a dashed black line, and the OPERA exclusion curve is shown as a solid grey line. Of note is that the MiniBooNE best fit point (black dot) is ruled out by OPERA at 90% CL, although it is still allowed by KARMEN.

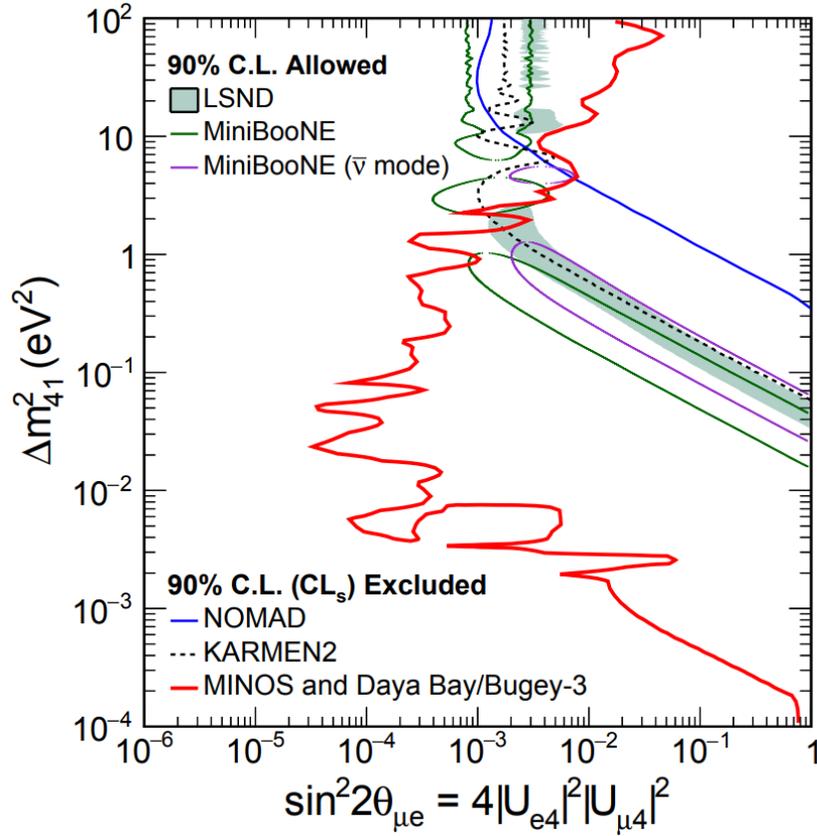


Figure 2.10: The 90% confidence level for the combined fit of MINOS, Daya Bay, and Bugey-3 data. Figure taken from reference [51]

2.5.2 Results from Gallium Experiments

As a part of their calibrations campaigns, the SAGE and GALLEX Gallium experiments had artificial radioactive sources (^{37}Ar , and ^{51}Cr) placed within them [53]. These sources decay via electron capture,



producing neutrinos of specific energies which can then be detected via



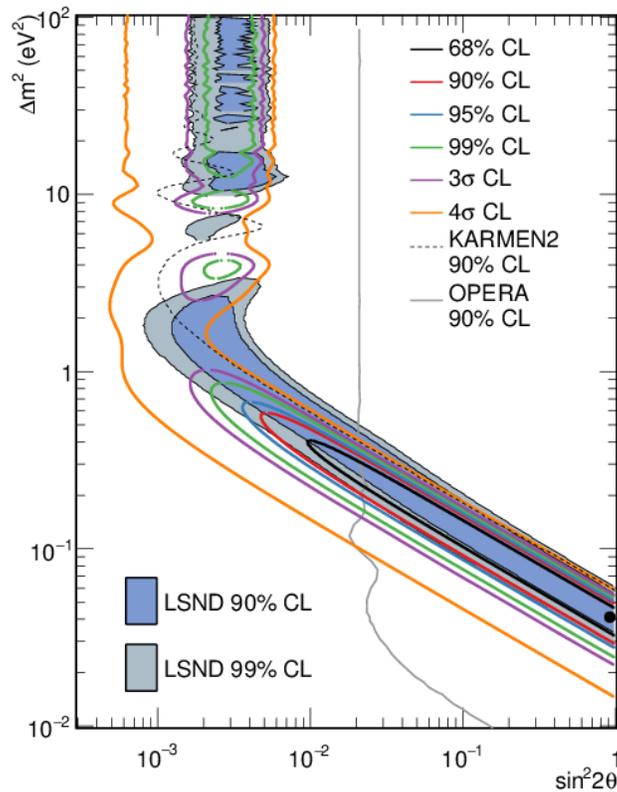


Figure 2.11: Allowed region for MiniBooNE with best fit point, overlaying the allowed LSND region. The KARMEN and OPERA 90% confidence levels are also shown.

The measured rate over the predicted rate is shown in Figure 2.12. Taking the weighted average of the four measurements indicates that there is a deficit in the ratio, equivalent to $R = 0.85 \pm 0.05$, meaning the number of events is approximately 2.8σ smaller than the prediction. One possible explanation for this discrepancy is short baseline active-to-sterile oscillations, with a Δm^2 of approximately 1eV^2 .

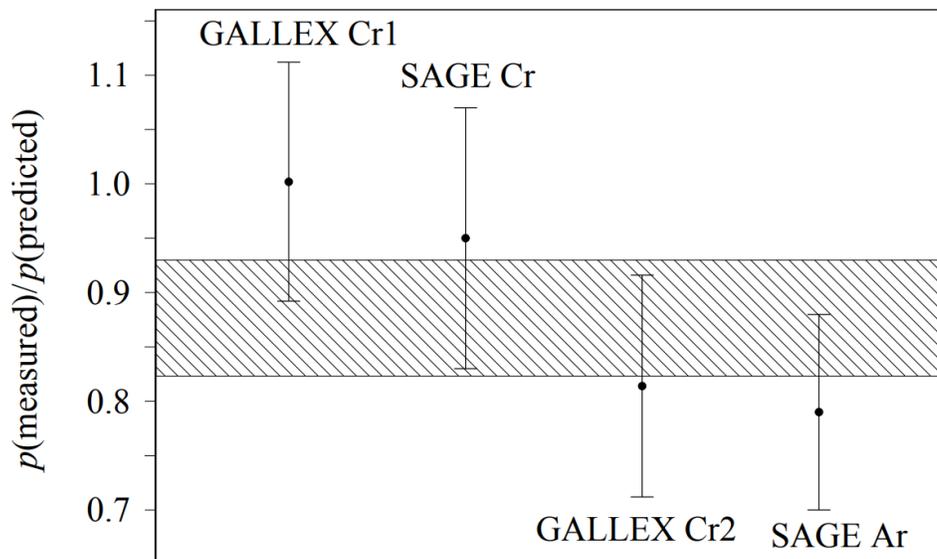


Figure 2.12: Rate results from the SAGE and GALLEX experiments, taken from reference [53]. The shaded region shows the weighted average of the four measurements.

2.5.3 Results from Reactor Neutrino Experiments

Many reactor neutrino experiments have operated over the past several decades, with some prominent examples being CHOOZ [54], KamLAND [55] and Daya Bay [56].

Initial analysis of reactor neutrino data found no evidence for sterile neutrino oscillations, however a re-analysis of reactor neutrino data in 2011 revealed that there was a deficit in the data with respect to an updated prediction at the level of $\sim 6\%$ [57]. This is shown in Figure 2.13. Here, the data points are the ratio of observed events compared to prediction, the red line is the predicted number of events under a three-neutrino oscillation paradigm, and the blue line is the predicted number of events under the assumption of an additional mass state for which $\Delta m^2 \gg 1$ and $\sin^2(2\theta) = 0.12$, shown for illustration purposes only.

More recently, the NEOS [58] and DANSS [59] collaborations have performed measurements of the ratios of the number of events as a function of baseline, rendering them insensitive to uncertainty around the $\bar{\nu}_e$ flux from reactors. Both experiments observe a spectral distortion which is oscillation-like (see Figure 2.14) and a combined fit of the data from the two experiments results in the no-oscillation hypothesis being disfavoured with respect to the sterile neutrino

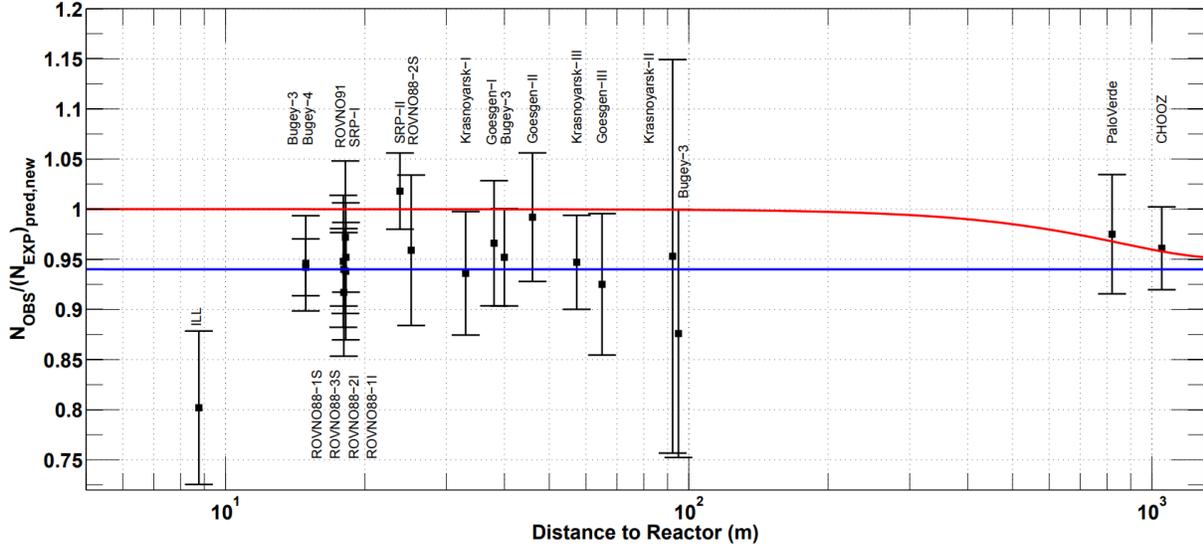


Figure 2.13: Results from reactor neutrino experiments, taken from reference [57]. The data points represent the ratio of observed to predicted events. The red line represents the predicted number of events under the assumption of a three-neutrino model. The blue line includes an additional mass state with $\Delta m^2 \gg 1$ and $\sin^2(2\theta) = 0.12$ for illustrative purposes.

oscillation hypothesis at 3.3σ [60].

The Daya Bay experiment has also performed a sterile neutrino search [61], but found no evidence for a signal.

2.5.4 Global Fits

The picture for sterile neutrinos is nothing if not murky. Several groups have undertaken the task of producing global fits to assess the compatibility of the different data sets. Results from one such global fit are shown in Figure 2.15. These results are separated into $\nu_e/\bar{\nu}_e$ appearance (2.15a), which are dominated by LSND and MiniBooNE, $\nu_\mu/\bar{\nu}_\mu$ disappearance (2.15b), the strongest limits for which come from IceCube and MINOS/MINOS+, and $\bar{\nu}_e$ disappearance, for which NEOS and DANSS contribute the strongest constraints.

If the disappearance datasets are combined and compared to the allowed region from appearance searches, as shown in Figure 2.16, then it is clear that the the allowed regions are incompatible. Here, the reactor neutrino datasets have been treated in two different ways: the first assumes

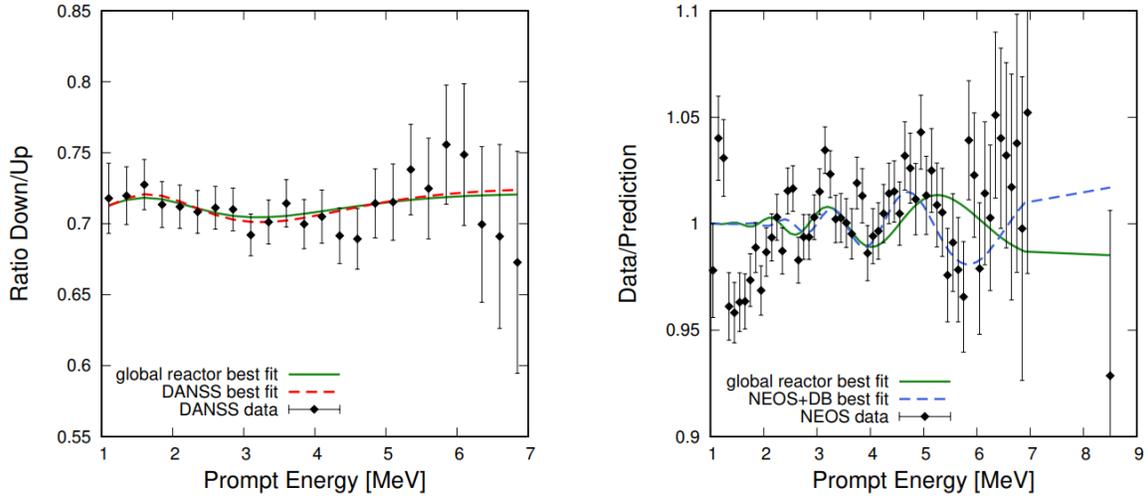


Figure 2.14: Observed spectra for the DANSS (left) and NEOS (right) experiments taken from reference [60]. The green solid line shows the prediction for a $\Delta m_{41}^2, \theta_{14}$ global reactor best fit, while the dashes curves show the best fit predictions for the respective experiments.

the published $\bar{\nu}_e$ fluxes to be correct, while the second allows the flux normalisation from ^{235}U , ^{238}U , ^{239}Pu , and ^{241}Pu to float (the sub-dominant ^{238}U and ^{241}Pu fluxes have a 1σ bound of 20% to stop the relative sizes becoming unphysical). The appearance results are presented both with and without the LSND decay-in-flight data.

The conclusions of the global fit presented in reference [60] state that they rule out sterile oscillations as an explanation of the LSND and MiniBooNE anomalies, but that they remain a viable option for the reactor and gallium anomalies. This is mostly due to the increasing power of the ν_μ disappearance results in recent years.

Despite the conclusion of this global fit there are many proponents of the sterile neutrino explanation for both LSND and MiniBooNE, and it is important that the field characterise the source of the anomalous results seen in these experiments.

In order to conclusively address the question of eV-scale sterile neutrinos a three-detector experiment is under development at Fermi National Accelerator Laboratory. This first detector of this *Short-Baseline Neutrino Program*, MicroBooNE, has been running in the Booster Neutrino Beam since 2015.

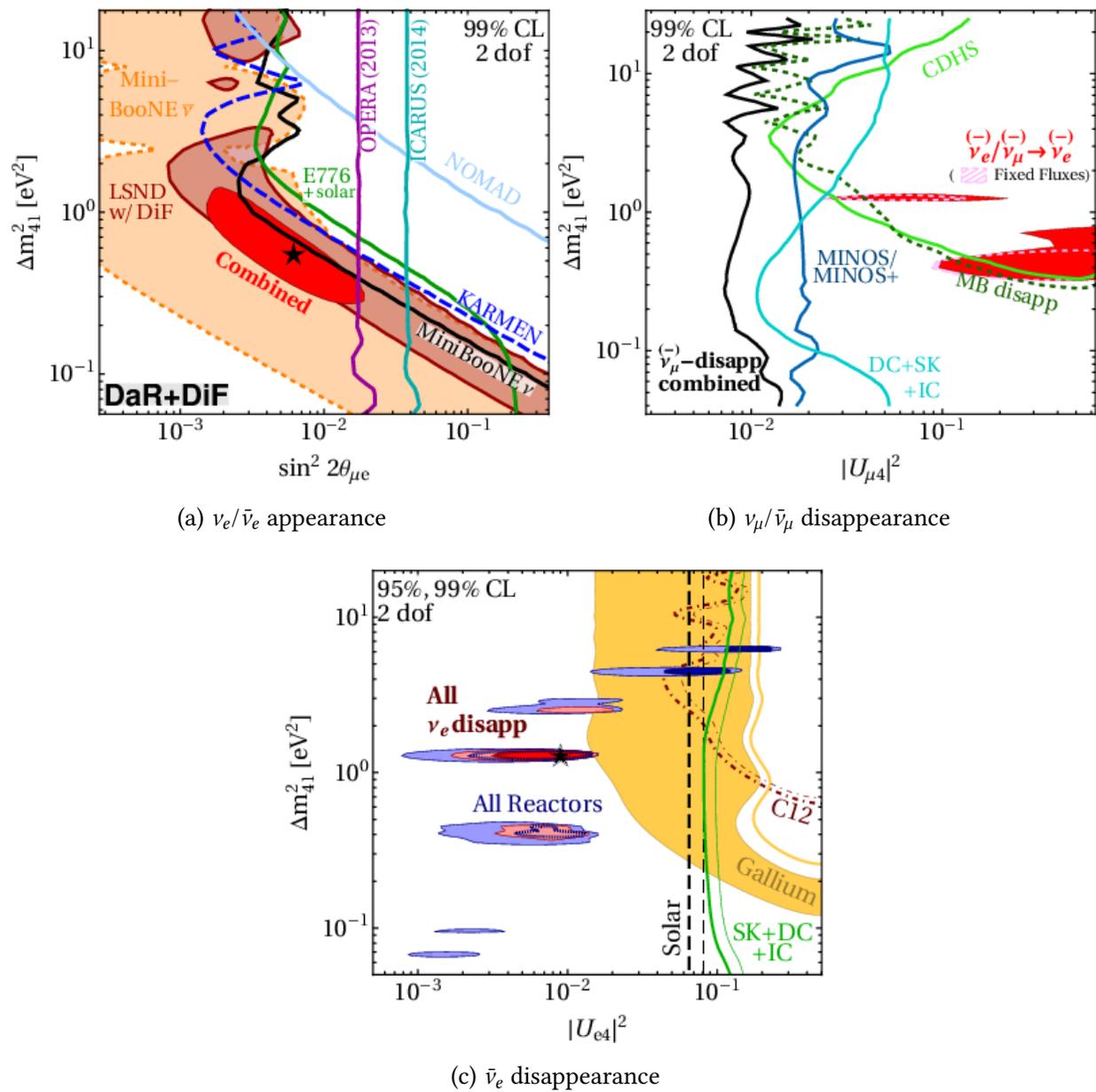


Figure 2.15: Global fits to appearance and disappearance data taken from reference [60].

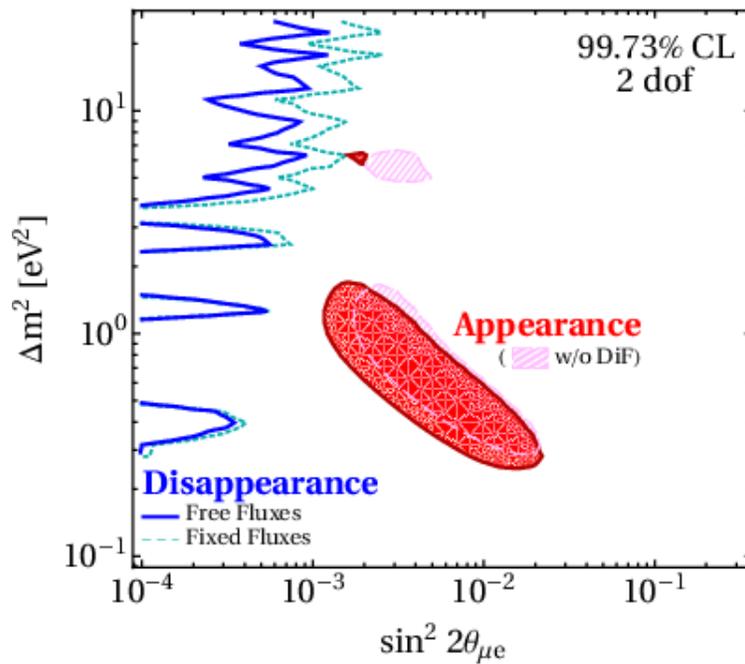


Figure 2.16: Global allowed region for appearance results including (excluding) LSND decay-in-flight are shown in red (pink, hashed). The limits set from disappearance datasets ($\bar{\nu}_e$, and $\nu_\mu/\bar{\nu}_\mu$) are shown for the free-fluxes assumption in blue, and fixed fluxes in cyan, dashed. Taken from reference [60].

Chapter 3

The MicroBooNE Experiment

*“It doesn’t matter how beautiful your theory is... if it
doesn’t agree with experiment, it’s wrong.”*

Richard Feynman

The Micro Booster Neutrino Experiment (MicroBooNE) is located at the Liquid Argon Test Facility at the Fermi National Accelerator Laboratory site, in Batavia, Illinois. MicroBooNE will act as the middle detector of the Short Baseline Neutrino (SBN) program, along with the Short Baseline Near Detector (SBND), and the Imaging Cosmic And Rare Underground Signals (ICARUS) detector, which will act as the near and far detectors when they begin running. MicroBooNE receives on-axis neutrinos from the Booster Neutrino Beam (BNB), being positioned around 470m downstream of the target. In addition to the on-axis neutrinos received from the BNB, MicroBooNE is able to detect off-axis neutrinos from the Neutrinos at the Main Injector neutrino beam.

The MicroBooNE detector contains 170 tons of liquid argon within a cylindrical cryostat, and employs a Liquid Argon Time Projection Chamber (LArTPC) as its primary detection technology, along with a light collection system in the form of 32 PhotoMultiplier Tubes (PMTs). The cryostat is surrounded by a Cosmic Ray Tagging system (CRT), which is composed of layers of plastic scintillator and provides around 85% coverage.

3.1 The Booster Neutrino Beam

The Booster Neutrino Beam (BNB) is a conventional, single-horn focused neutrino beam, which provides neutrinos peaking in energy at around 800 MeV [62]. Historically, the main target for the beam has been the MiniBooNE detector.

The Fermilab Booster is a 474-meter-circumference synchrotron which operates at 15 Hz. Protons with an energy of 400 MeV are injected into the booster from the Fermilab LINAC [63] and are then accelerated to 8 GeV kinetic energy [64]. From here, proton pulses are delivered into the target hall and onto a beryllium target with a 5 Hz average rate. An aerial view of the Fermilab beamlines can be found in Figure 3.1.

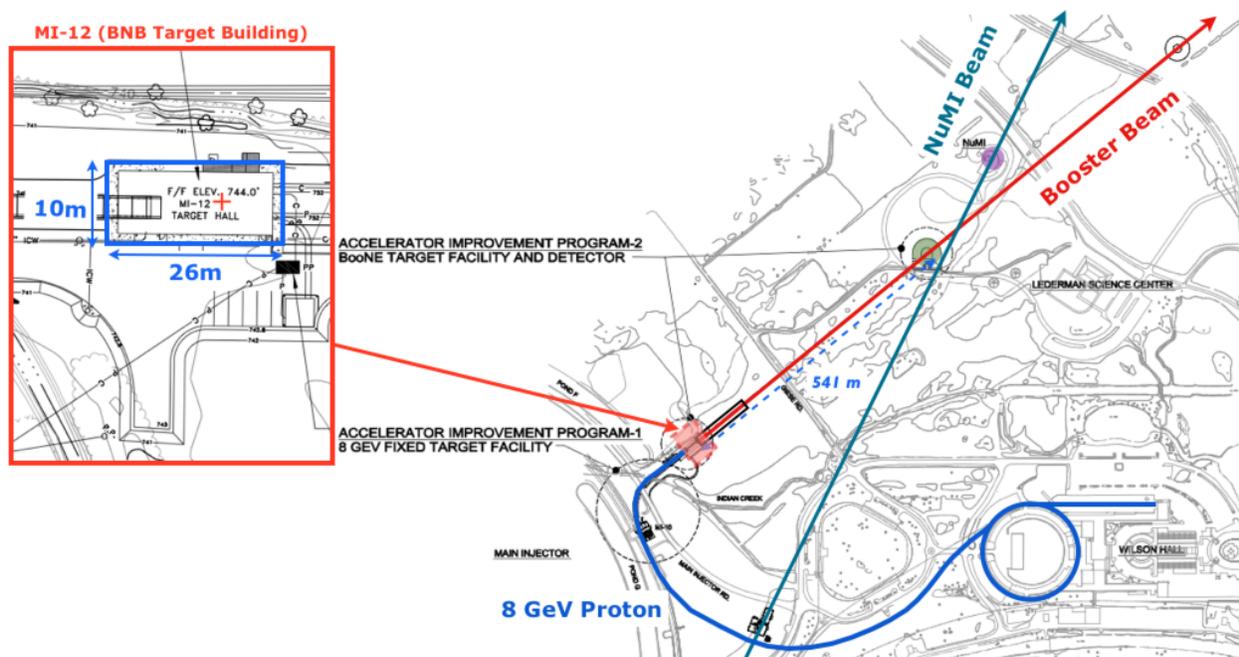


Figure 3.1: Aerial view of the Fermilab beamlines. The Booster neutrino beamline is displayed in red, and the path of the protons is displayed in blue. The NuMI neutrino beam is also displayed in green. The light green circle shows the position of the MiniBooNE detector, which MicroBooNE sits slightly upstream of.

Each $1.6 \mu\text{s}$ beam spill is expected to contain 4.5×10^{12} protons [7]. The protons incident on the target induce interactions, providing a cascade of secondary particles, mainly pions and kaons. Selected-sign particles are focused towards the beamline by use of a single focusing horn pulsed

at 174 kA, while wrong-sign¹ particles are de-focused. The selected-sign charged particles are passed through a collimator and allowed to decay in a 50 m air-filled decay pipe. A steel and concrete beam stop at the end of the decay region removes the majority of secondary particles and any remaining charged particles are absorbed by the dirt to leave a beam of pure neutrinos. This is displayed graphically in Figure 3.2.

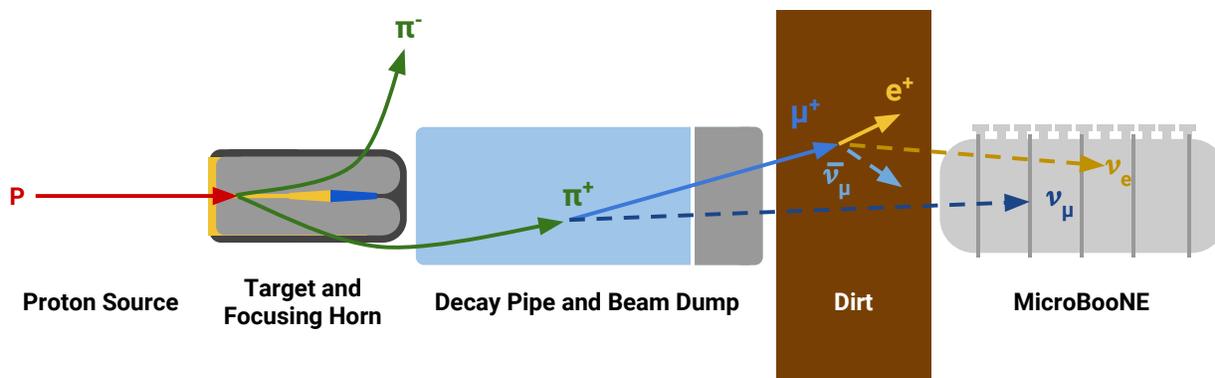


Figure 3.2: Graphical representation of the BNB. Protons with a kinetic energy of 8 GeV are impinged on a target, embedded in a focusing horn which acts to defocus wrong-signed secondary particles and focus right-signed secondary particles. The secondary particles are then allowed to decay in a decay pipe, with residual particles being stopped in the beam dump, or the dirt which follows. Neutrinos from secondary (and tertiary) particle decays pass through both the beam dump and dirt and arrive at the MicroBooNE detector.

3.1.1 Beam Composition

Secondary particles produce mainly ν_μ , with a small contamination of ν_e , $\bar{\nu}_\mu$ and $\bar{\nu}_e$, as shown in Figure 3.3. The composition of the beam is well understood. In neutrino mode, it is dominated by the ν_μ component which is estimated to be $\sim 93.6\%$ of the total composition of the beam [7], with the second largest component being the $\bar{\nu}_\mu$. The ν_e and $\bar{\nu}_e$ components are one and two orders of magnitude smaller, respectively. Table 3.1 is taken from the MiniBooNE flux paper [62] and shows the predicted composition of the BNB in neutrino running mode.

The production of ν_μ s is dominated by charged pion decay below around 2.5 GeV. The primary

¹The focusing horn can be used to focus either negatively or positively charged particles.

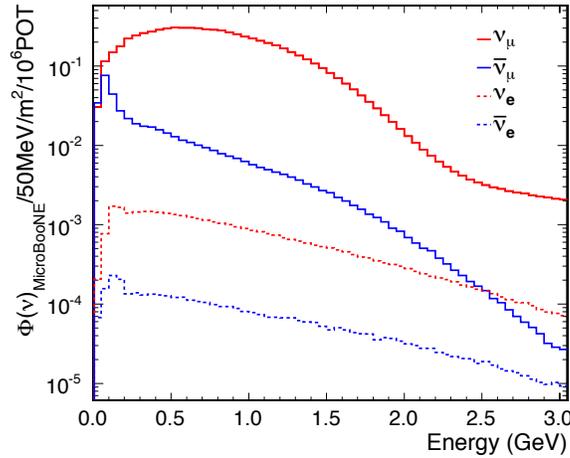


Figure 3.3: The predicted BNB flux at MicroBooNE for neutrino mode running, taken from reference [7]. The dominant component of the beam is ν_μ , although there is significant contribution from $\bar{\nu}_\mu$ at low energies. The expected contamination in the beam from $\nu_e/\bar{\nu}_e$ is at the sub-percent level.

decay mode for charged pions ($\pi^+ \rightarrow \mu^+ + \nu_\mu$) has a branching ratio of 99.9877% due to helicity suppression. Charged pions are produced in a spin-0 state, and decay to a charged lepton and its corresponding neutrino. The two particles must be produced in opposite-handed helicity states in the rest frame of the pion. Because the neutrino has a left-handed chirality and an extremely low mass it is essentially always produced in a left-handed state. The lepton therefore must have a right-handed helicity, and because the muon has a much larger mass than the electron, the $\pi^+ \rightarrow e^+ + \nu_e$ decay is suppressed by a factor of $\sim 10^{-4}$ [65].

The largest non- ν_μ component of the beam is the $\bar{\nu}_\mu$ component which are produced from highly energetic wrong-sign secondary particles (π^- , dominantly), which tend to be very forward going and are not defocused by the focusing horn.

The small ν_e and $\bar{\nu}_e$ components of the beam are primarily from the $\mu^+ \rightarrow e^+ + \nu_e$ decay and the semileptonic decays of K_L^0 s respectively.

	ν_μ	$\bar{\nu}_\mu$	
Flux ($\nu/\text{cm}^2/\text{POT}$)	5.19×10^{-10}	3.26×10^{-11}	
Frac. of Total	93.6%	5.86%	
Composition	π^+ :	96.72%	π^- :
	K^+ :	2.65%	$\pi^+ \rightarrow \mu^+$:
	$K^+ \rightarrow \pi^+$:	0.26%	K^- :
	$K^0 \rightarrow \pi^+$:	0.04%	K^0 :
	K^0 :	0.03%	$K^0 \rightarrow \pi^-$:
	$\pi^- \rightarrow \mu^-$:	0.01%	$K^+ \rightarrow \mu^+$:
	Other:	0.30%	$K^- \rightarrow \mu^-$:
			Other:
			4.43%
	ν_e	$\bar{\nu}_e$	
Flux ($\nu/\text{cm}^2/\text{POT}$)	2.87×10^{-12}	3.00×10^{-13}	
Frac. of Total	0.52%	0.05%	
Composition	$\pi^+ \rightarrow \mu^+$:	51.64%	K_L^0 :
	K^+ :	37.28%	$\pi^- \rightarrow \mu^-$:
	K_L^0 :	7.39%	K^- :
	π^+ :	2.16%	π^- :
	$K^+ \rightarrow \mu^+$:	0.69%	$K^- \rightarrow \mu^-$:
	Other:	0.84%	Other:
			4.62%

Table 3.1: Predicted beam composition of the BNB in neutrino running mode, taken from reference [62]. The total fraction of the flux is displayed, as well as the composition of each of the ν_μ , $\bar{\nu}_\mu$, ν_e , and $\bar{\nu}_e$ components, which are listed in order of size contribution.

3.2 The MicroBooNE Detector

This section is dedicated to a description of the MicroBooNE detector. Section 3.2.1 outlines how signals are generated in LArTPCs. Sections 3.2.2 and 3.2.3 give overviews of the MicroBooNE TPC and light collection system, respectively.

Selected important characteristics of the MicroBooNE detector are outlined in Table 3.2.

Parameter	Value
TPC Dimensions	
Width	2.56 m
Length	10.37 m
Height	2.33 m
Anode Plane Array	
Number of Wire Planes	3
Wire Plane Spacing	3 mm
Wire Pitch	3 mm
Wire Orientation w.r.t Vertical (U, V, Y)	+60°, -60°, 0°
Number of channels (U, V, Y)	2400, 2400, 3256
Bias Voltage (U, V, Y)	-200 V, 0 V, 400 V
Field Cage	
Number of Loops	64
Voltage Step	1.09 kV
Electric Field	
Cathode Voltage	-70 kV
Drift field	273 V/cm
Maximum Drift Time	2.3 ms
PMTs	
Number of PMTs	32
Photocathode coverage	0.9%

Table 3.2: Important variables for the TPC, field cage, and light detection system of the Micro-BooNE detector.

3.2.1 Signals in LArTPCs

Ionisation Signals

The modus operandi of a LArTPC is very simple. A charged particle passing through the liquid argon contained within the detector liberates electrons through the process of ionisation. These

ionisation electrons effectively act to track the path and energy deposition of the initial charged particle. The TPC is held at a constant electric field by holding one side of the TPC (the cathode) at a constant negative voltage. Under the influence of this electric field, the ionisation electrons can be transported over a number of meters with very little distortion to the anode plane of the TPC, where detection instrumentation is situated.

The number of electrons detected by the instrumented anode plane is dependent primarily on two effects: recombination and the purity of the argon inside the detector.

Recombination happens at the point of ionisation, where a fraction of the liberated electrons thermalise and quickly recombine with the argon ions. The scale of this effect is then a function of the applied electric field. The higher the electric field, the more quickly the ions and electrons are separated, and so the smaller the chance of recombination occurring.

Impurities in the liquid argon, primarily oxygen and water, can affect the electron yield at the anode plane by capturing electrons as they traverse the drift. Once again, this is a function of the electric field, as a higher field leads to a shorter drift time, meaning there is less opportunity for electrons to be captured.

As the cloud of ionisation electrons drift through the TPC, the shape of the cloud is modified by diffusion of the electrons in both the direction parallel to the electric field (longitudinal electron diffusion), and perpendicular to the electric field (transverse electron diffusion). This will be discussed in detail in Chapter 5.

Optical Signals

The information in this section is taken mostly from reference [66], an excellent reference for anybody interested in learning about light production in liquid argon.

Liquid argon is a bright scintillator, producing $\mathcal{O}(10,000)$ photons per MeV of deposited energy, with dependence on the dE/dx , the electric field strength, and the argon purity. These photons peak with a wavelength of 128 nm, and come with two distinct time signatures. The fast component occurs within around 6 ns of the interaction, while the slow component occurs

around 1500 ns later.

Both the fast and slow components come from the emission of photons from excited dimers², with the fast component coming from singlet states, while the slow component comes from triplet states, which are delayed by inter-system crossing.

There are two methods in which these excited dimers can be produced. The first method, *self-trapping*, happens when an argon atom becomes excited by an ionising particle, and joins with a neutral argon atom. The second method, *recombination luminescence*, happens when an argon atom is imparted with enough energy to undergo ionisation, and this then combines with another argon ion to produce a charged argon dimer state. The ionisation electrons then thermalise and begin to recombine with the charged argon dimer, leading to an excited argon dimer. This is shown in Figure 3.4.

The recombination method provides an explanation for why scintillation light yield is dependent on the applied electric field; a higher electric field results in a quicker separation of electrons and ions, meaning there are fewer opportunities for recombination. This means that the number of photons from the recombination method is reduced.

In addition to scintillation light, directional Cherenkov light is produced via the usual process: particles traveling faster than the speed of light in the medium results in a directional cone of light.

While oxygen and water are the contaminants of interest for ionisation signals, nitrogen is the primary contaminant of interest for optical signals. An excess of nitrogen makes liquid argon opaque to its own scintillation light, resulting in a reduction in light yield at the optical detectors.

3.2.2 The Liquid Argon Time Projection Chamber

The MicroBooNE TPC measures 2.56 m along the drift (x -direction), with a height of 2.33 m (y -direction), and a length of 10.37 m (z -direction). It is composed of a cathode at high- x , an anode at low- x , and a field cage connecting the two.

A diagram of the MicroBooNE TPC is shown in Figure 3.5.

²A dimer is a molecule composed of two identical molecules or atoms, in this case, argon.

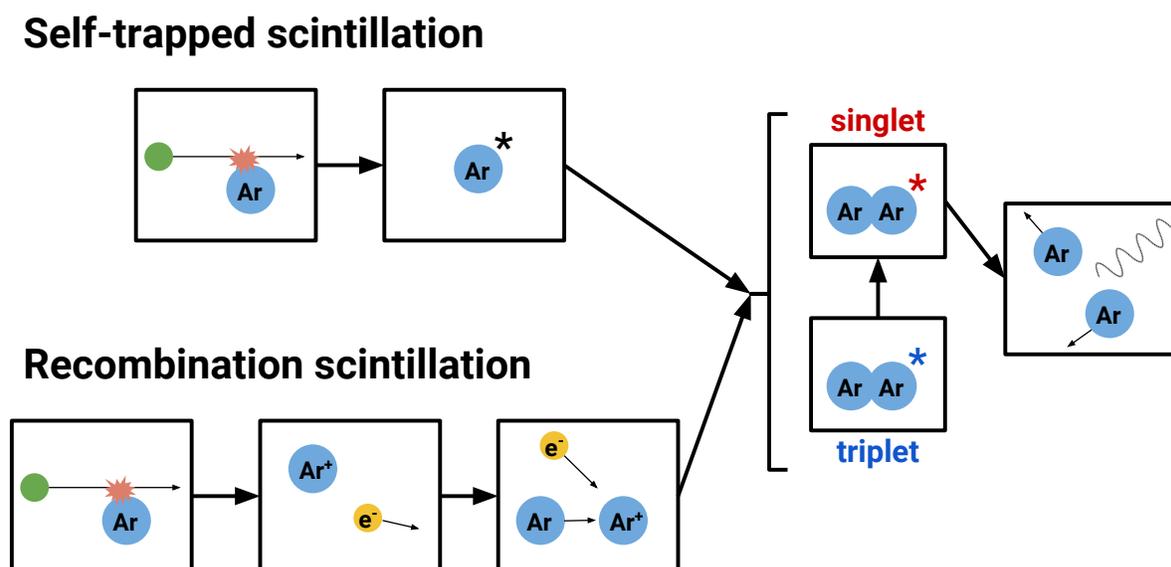


Figure 3.4: Diagram showing the two methods available for scintillation light production in liquid argon. Both the self-trapped and recombination methods can produce either a singlet or triplet excited dimer state, which give rise to the fast and slow components of the scintillation light, respectively. For scintillation produced via the self-trapped method, the ratio of singlet-to-triplet excimers is $\sim 35 : 65$. For the recombination method the ratio is closer to $50 : 50$.

The Cathode

The cathode is composed of 9 individual plates. In order to deliver a uniform electric field throughout the detector, the cathode must be flat and parallel to the Anode Plane Array (APA). A best fit plane from survey data suggests that the two faces of the TPC are within 0.0413° of parallel to each other. The maximum deviations from this best fit are +6.6 mm and -6.5 mm, with 90% of 10,000 survey points within 5 mm deviation of the best fit plane. This indicates that the cathode is slightly bowed with respect to the APA.

The cathode is held at -70 kV, and stepped down to ground over the 64 loops composing the field cage. The resulting electric field has a strength of 273 V/cm in the drift direction.

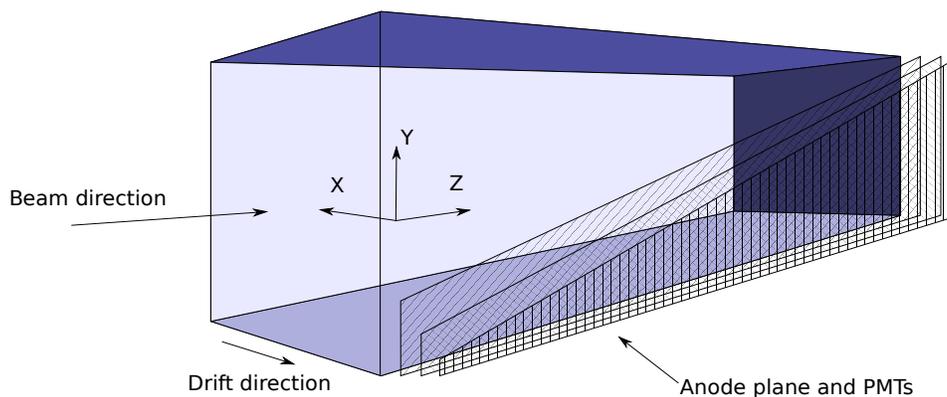


Figure 3.5: A diagram of the MicroBooNE TPC. The orientation is defined such that it makes a right handed co-ordinate system with x increasing along the negative drift direction. The three wire planes are displayed oriented vertically (Y , collection plane), and $\pm 60^\circ$ to this (U and V , induction planes). The light collection system sits just behind the wire planes but is neglected in this diagram.

The Anode Plane Array

The MicroBooNE APA makes use of steel sense wires as its detection technology. The wires are angled vertically, and at ± 60 degrees to the vertical. It should be noted that in contrast to some other current and future LArTPCs, MicroBooNE does not have so-called wrapped wires, meaning there is a one-to-one correspondence between TPC wires and TPC channels, and often these terms are used interchangeably³.

By applying a bias voltage over the three planes of wires it can be ensured that the electrons drift past the first two planes (termed *induction planes*) unimpeded, and are collected in the final plane (the *collection plane*). Each induction plane has 2400 wires, while there are 3456 wires on the collection plane, bringing the total number of wires to 8256. Ionisation electrons produce current on these wire planes; bipolar waveforms are induced on the two induction planes as the electrons pass by, and a monopolar waveform is produced when the electrons are collected on

³There are a number of channels which are not connected to wires, however these have channel numbers greater than 8256, and are often neglected.

the collection plane [6]. Simulated example waveforms are shown in Figure 3.6. For each of the planes, a single minimally ionising particle is simulated at an angle both parallel to the APA and perpendicular to the wire orientation, providing an “ideal” track. The black lines in each diagram represent the contribution of charge drifting in the region within ± 1.5 mm of primary induction or collection wire to the signal, while the red and blue dashed lines represent the signal when contribution from charge drifting in the region bounded by $[-2,+2]$ wires and $[-10,+10]$ wires away are included, respectively. This demonstrates that *induced charge effects* play a large part in modifying the shape of the signals, and must be taken into account. On the V and Y planes, which are each shielded by at least one other plane of wires, these effects are limited to nearby wires, however for the U plane, which receives no shielding, induced charge effects are important to $\mathcal{O}(10)$ wires. This will be an important consideration in later sections of this work.

The wire planes have a pitch of 3 mm, and the planes are spaced 3 mm apart, meaning that millimeter-scale resolution can be attained.

The three wire planes effectively give three different two-dimensional views of each readout window in wire-time space, as shown in Figure 3.7. The three views can be combined in order to recover a full three-dimensional reconstruction of the event. This will be covered in Section 4.6.

The MicroBooNE TPC Electronics

Signals from each channel are amplified and shaped through a metal-oxide semiconductor analog front-end cold Application Specific Integrated Circuit (ASIC). A number of these ASICs are mounted on each Front End Motherboard (FEM), which are located in the liquid argon, close to the sense wires so that the capacitive load, and hence noise, is minimised. From the FEMs, the signals are passed out of the cryostat through a warm flange and into the Data Acquisition (DAQ) machines, whereupon the analogue signals are converted into digital signals through a 12-bit Analogue-to-Digital Converter (ADC), and then processed [68]. This is displayed graphically in Figure 3.8.

The ASICs are able to run at a number of different gains (4.7, 7.8, 14, or 25 mV/fC) and shaping

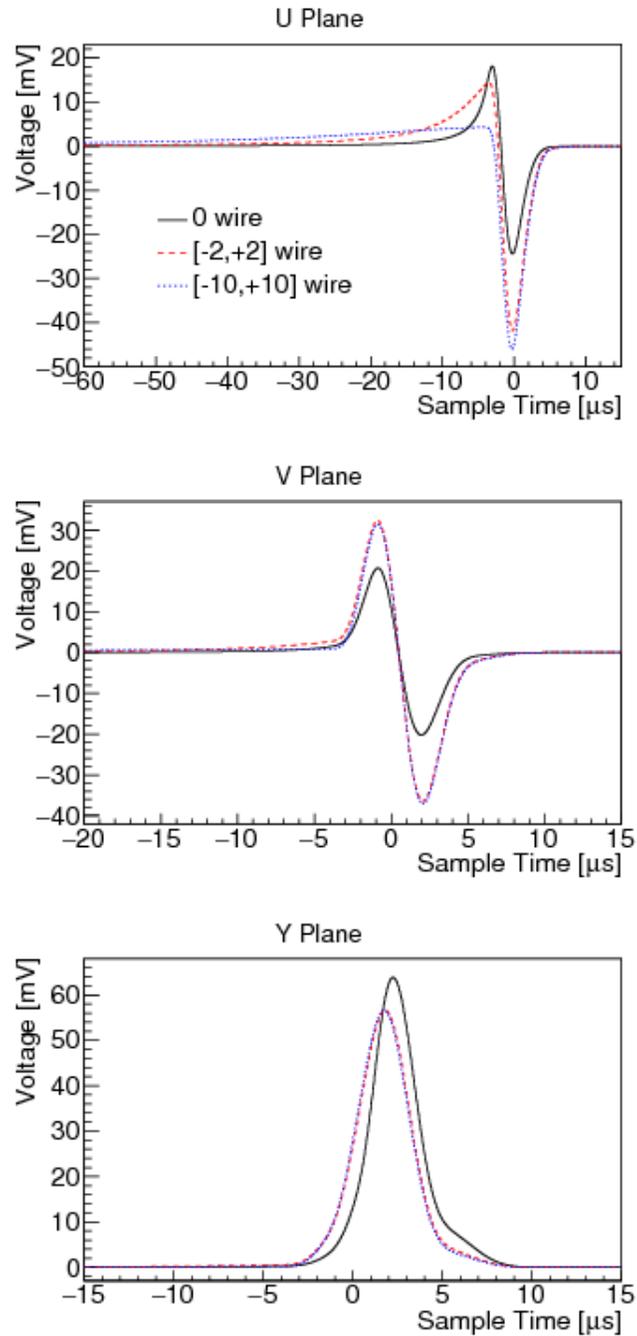


Figure 3.6: Simulated baseline-subtracted TPC signals for ideal tracks traveling parallel to the APA, in the direction perpendicular to the wire orientation for each plane. The “0 wire” depicts the signal for charge arriving at the central wire, while the “[$-N,+N$] wire” plots provide the contribution to the signal on the central wire from ionisation electrons that drift within $\pm N$ wires. Figure taken from reference [67].

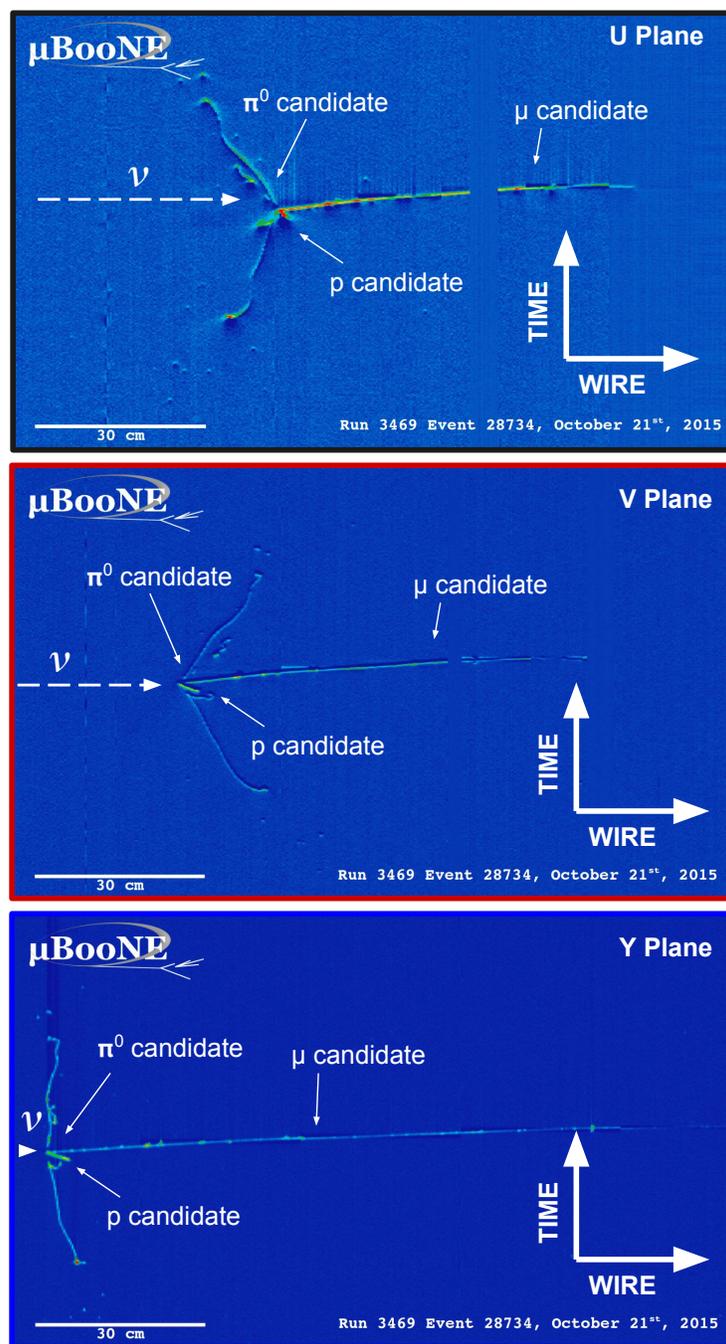


Figure 3.7: Event displays from MicroBooNE Run 1 data showing the U, V, and Y plane views for a candidate $CC\pi^0$ event. Here, the x -axis is equivalent to the wire number, and the y -axis is time.

times (0.5, 1.0, 2.0, or 3.0 μs). Here, the gain defines the peak height of the pulse and the shaping time is defined to be the time between 5% of the peak height and the peak height of each pulse. In

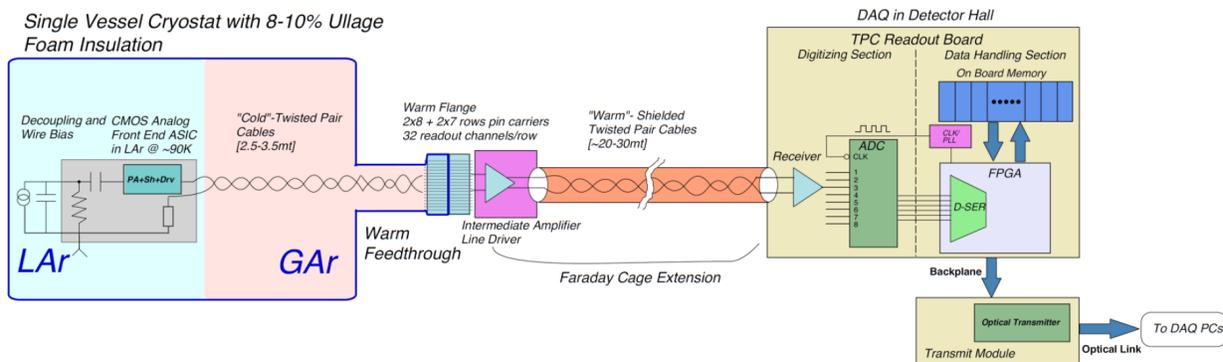


Figure 3.8: A schematic of the MicroBooNE electronics. The left part of this image represents the detector, with the front end motherboards embedded in the liquid argon, while the right part of the image represents the DAQ machines in the detector hall. This is taken from reference [68].

MicroBooNE, each ASIC operates nominally with a 14 mV/fC gain and 2.0 μ s shaping time. The choice of gain is not arbitrary and needs to be given serious consideration: a gain which is too high means that particles which deposit a large amount of energy in a short z -projected distance (such as tracks at high angle to the anode plane) run the risk of saturating the ADC and loss of information. In addition to this, a gain must be chosen to ensure that signals from minimally ionising particles are not swamped by gain-independent noise. Similarly the choice of shaping time must also be given thought: a short shaping time improves peak-resolution, however it increases the level of noise, while the inverse is true for longer shaping times.

Misconfigured Regions

Occasionally during re-configuration of ASICs, a small number of ASICs will report that they have successfully been configured to a gain of 14 mV/fC and shaping time of 2 μ s when in reality they have become stuck on the default gain and shaping time (4.7 mV/fC and 1 μ s, respectively). Table 3.3 shows the misconfigured channels as a function of run number for MicroBooNE Run 1 data.

In Run 1, this affects a relatively small number of channels (\sim 4% of total channels) on the first induction plane, however the fraction of misconfigured channels becomes significantly larger in later runs.

Run Range	Misconfigured Channels	Total Misconfigured Channels
4952-5281	None	0
5282-5810	2016-2111, 2176-2303, 2352-2383	253
5811-6699	2016-2111, 2128-2303, 2320-2383	333
6700-6998	2240-2255	15

Table 3.3: Run 1 misconfigured regions.

U- and Y-shortened Regions

Upon ramping up the wire planes to bias voltage, it was noticed that the field response in some regions of the U and V planes were not what was expected. In the case of the U plane, this is thought to be due to a V plane wire touching the U plane wires and causing them to become shorted. The result of this is that electric field in this region becomes modified and a subset of the electrons are collected on the U plane rather than passing by unimpeded. This means that charge induced or collected in the shadow of the so-called *U-shortened* region is reduced with respect to nominal channels on each plane. In addition to this, there is a suspected short between the V and Y plane, meaning that the V plane takes on a collection-plane-like response in this region.

Dead Regions

Approximately 10% of the MicroBooNE readout channels are functionally “dead”, meaning that no useful information may be extracted from these channels. Only 2 readout planes need to be functional in any region of the x-z plane for reconstruction of high-level objects to be viable, the third plane is used for disambiguation. Taking this into account means that reconstruction is impossible in around 3% of the TPC.

Space Charge Effect

LArTPCs are a relatively slow technology; they have a long read-out with respect to the length of a beam spill. This means that for a detector located on the surface, such as MicroBooNE, neutrino

events are overlaid with $\mathcal{O}(10)$ cosmogenic tracks and showers.

The ionisation of electrons from the argon results in a significant number of argon ions. These are heavier than the ionisation electrons and so they drift much more slowly to the cathode where they are finally dissipated. The result of this is a build up of positive charge in the detector which acts to modify the electric field. This modification can be significant, reaching approximately $\sim 15\%$ for some regions of the detector. This is predicted to have a significant impact on the correct reconstruction of track positions, as well as introducing position-dependence to the recombination and diffusion within the detector. The effects of the space charge effect can be seen in Figure 3.9.

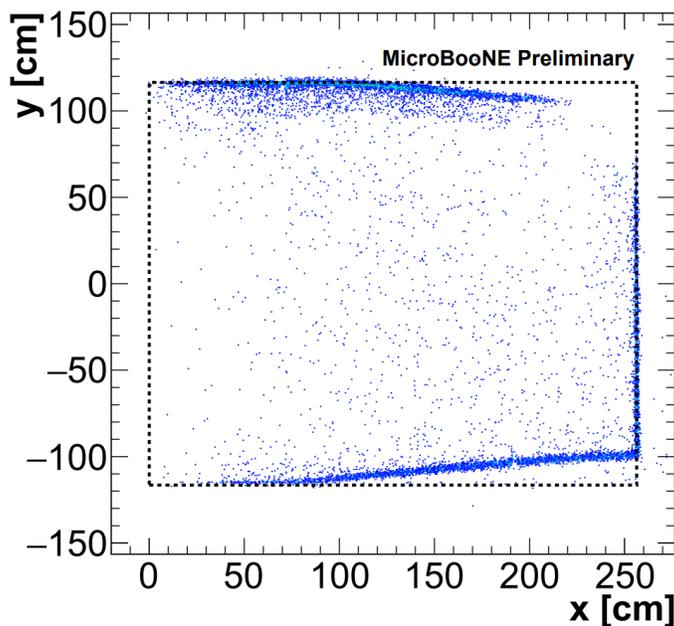


Figure 3.9: Start/end points of tracks tagged by an external muon counter and reconstructed in the MicroBooNE TPC. In the absence of the space charge effect the points are expected to be distributed at the TPC borders, where the dashed line is placed. Figure taken from reference [69].

3.2.3 The Light Collection System

To mitigate challenges related to the high cosmic flux and slow readout, LArTPCs are often paired with some form of light-detection system, which has a much quicker read-out time, and can be

used to perform cosmic rejection.

MicroBooNE PMTs

MicroBooNE implements 32 8" Hamamatsu R5912-02MOD PhotoMultiplier Tubes (PMTs, shown in Figure 3.10a) which are used primarily to provide the interaction time of energy deposited within the TPC.

As shown in Figure 3.10b, the PMTs are most efficient for photons with a wavelength in the 350-450 nm region. Because liquid argon scintillates in the 128 nm region, the PMTs are situated behind tetraphenyl-butadiene (TPB) coated acrylic plates. These plates act to take the 128 nm wavelength scintillation light and shift it to longer wavelengths (peaking at 425 ± 20 nm), where the PMTs are most efficient.

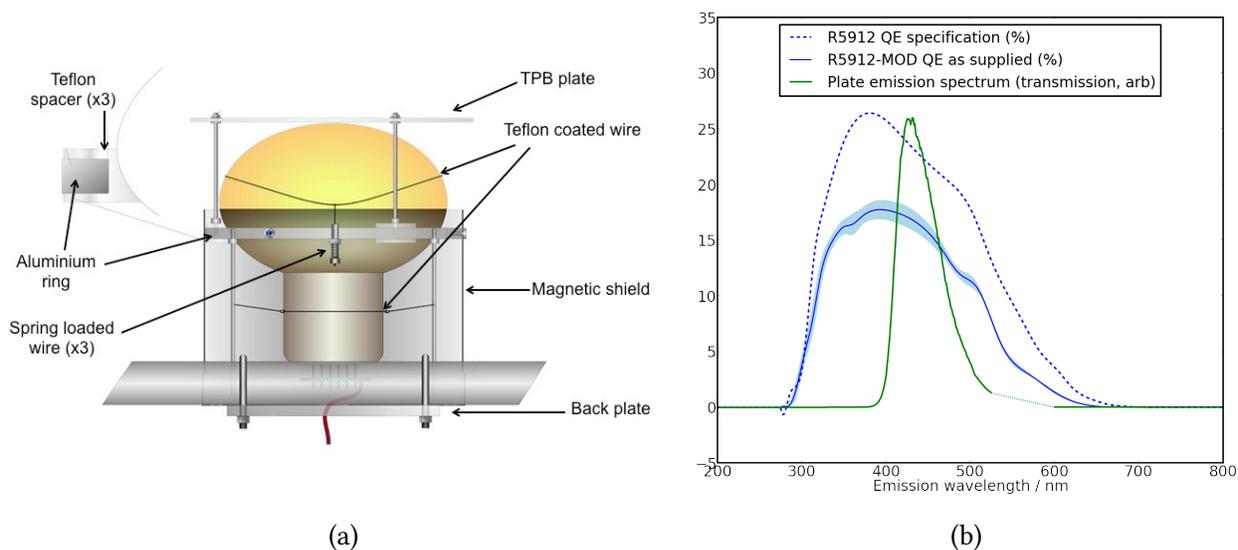


Figure 3.10: Schematic of the MicroBooNE PMTs (3.10a), and the quantum efficiency of the PMTs (specification: dashed blue, supplied: solid blue), along with the tetraphenyl-butadiene emission spectrum (green) superimposed in arbitrary units (3.10b). These are taken from reference [6]. Note that the R5912-MOD is a modification of the standard R5912 which has a platinum layer between the photocathode and the glass bulb, which allows the PMT to maintain conductance at the cryogenic temperatures necessary for MicroBooNE. The result of this is that the efficiency of the PMT is reduced.

The Light Paddles

In addition to the primary light detection system, there are 4 light paddles which were installed in the detector for research and development purposes. They are not currently used in the standard MicroBooNE analysis chain.

The MicroBooNE PMT Electronics

Signals from the light collection system are first passed through a splitter circuit, which results in a high-gain and low-gain component of the signal, which carry 18% and 1.8% of the total signal amplitude respectively⁴. These signals are then shaped with a 60 ns shaping time, and finally they are digitised at 64 MHz with a 16-bit ADC.

3.2.4 The Cosmic Ray Tagger

In order to improve MicroBooNE's ability to reject cosmic ray muons and associated cosmogenic particles, work was undertaken to construct a Cosmic Ray Tagger (CRT) which consists of 73 scintillating modules [70].

The modules line the top, bottom and sides of the MicroBooNE detector, although the upstream and downstream ends are left uncovered due to space constraints in the Liquid Argon Test Facility (LArTF) building. Each scintillator module is made up of 16 scintillator strips and is read out by a FEM when photons are induced inside the strip by cosmogenic activity.

The MicroBooNE CRT attains a maximum solid angle coverage of 85%. This is due to the lack of CRT panels at the upstream and downstream ends of the detector, and the fact that the top panel had to be elevated from the detector in order to accommodate the DAQ machine racks.

The CRT side and bottom panels were installed preceding the beginning of Run 2 data taking, and the final top panels were installed prior to the beginning of Run 3 data taking.

⁴There is an 80% attenuation in the signal. This is purposeful: the electronics expect a certain gain from the the PMTs, which would mean running at an extremely low voltage. To combat this, the PMTs are run at a close-to-nominal voltage, but the signal is attenuated to match the expected gain.

3.3 Data-Taking Triggers

Data-taking *triggers* are the primary method in which MicroBooNE is able to overcome its considerable cosmogenic backgrounds. Here, the two triggers which are employed are described.

3.3.1 Accelerator Trigger

MicroBooNE is an accelerator-based neutrino experiment. This means that beam timing can be leveraged. Preceding delivery of protons onto the target, a trigger signal is issued from the accelerator complex to MicroBooNE, which tells the detector to expect beam to be delivered, and to begin reading out an event at some defined time in the future.

The accelerator trigger allows data to be read from the MicroBooNE detector in chunks rather than with a continuous readout, however the flux of the BNB combined with the low neutrino cross section mean that even after this trigger only approximately 1 in 600 triggered events will contain a neutrino interaction in the TPC.

In addition to the beam trigger is the external trigger, which can be issued at any time and tells the detector to read out one readout window. This is useful for taking cosmic-only (*off-beam*) data.

3.3.2 PMT Trigger

Light from the PMT system is used to trigger readout on detection of a flash of scintillation light containing more than 6.5 photoelectrons within the beam window. Because light is detected at the PMTs within $\mathcal{O}(\text{ns})$ of the neutrino interaction, this trigger is *on-line* in the sense that it is a decision made by the DAQ after the TPC and PMT signals have been recorded.

The software trigger allows a huge improvement on the signal to background ratio of the MicroBooNE data; after application of the hardware trigger approximately 1 in 600 beam spills is expected to induce a neutrino interaction in the TPC due to the low cross-section of neutrinos. By application of the software trigger this can be reduced to around 1 in 6.

Documentation on the specifics of the software trigger algorithm can be found in reference [71].

3.4 Detector Operations

MicroBooNE began taking data in August 2015, and its data is split into several *run periods*.

- Run 1 describes the period from the beginning of operations through to the beginning of the CRT installation, covering runs up to run 7955, which occurred on September 26, 2016.
- Run 2 describes the period from the beginning of the CRT installation through to the time when one of the PMTs became non-responsive.
- Run 3 describes the period from when one of the PMTs became non-responsive.

The data used in this thesis was taken from Run 1 of data taking, shown in Figure 3.11. The POT on tape shown in this figure represents the POT delivered from the BNB weighted by the detector up-time. This includes the short periods between runs where the DAQ is being configured as well as any extended down-time of the detector during running of the BNB.

It should be noted that because only data from MicroBooNE Run 1 has been used in this work, the CRT has not been leveraged for any of these analyses. This dataset is used because MicroBooNE is pursuing a blind analysis strategy for its flagship analysis. Due to this, most analyses are limited to using data taken in Run 1.

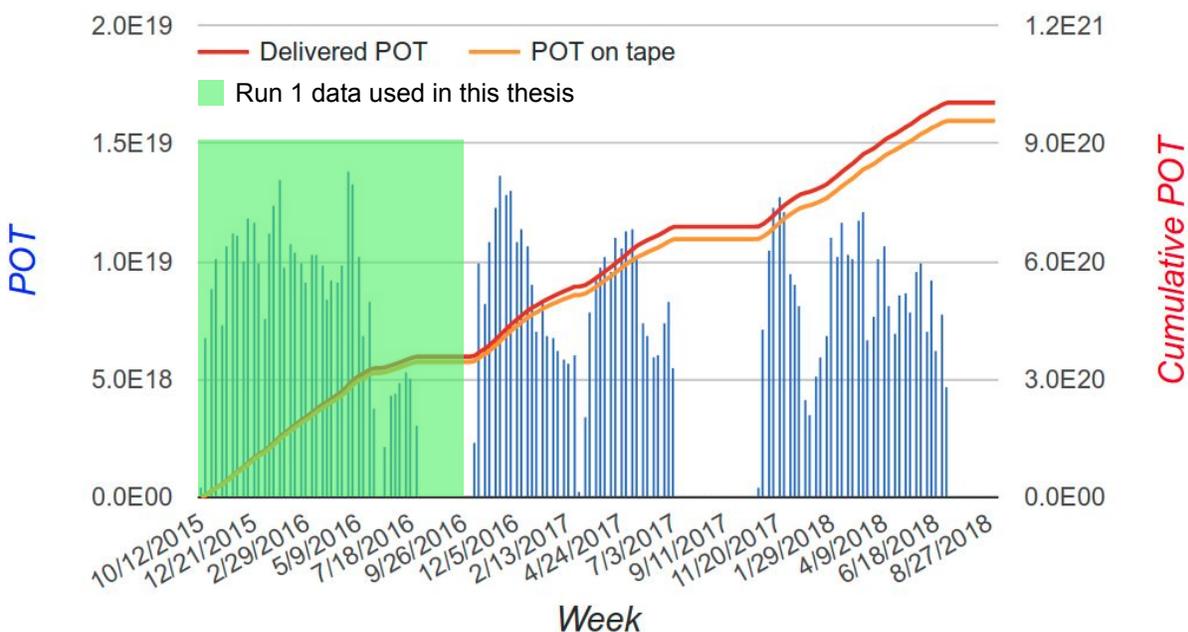


Figure 3.11: BNB performance for MicroBooNE's first three years of data taking. The blue histogram represents the POT delivered on a week-by-week basis, while the red line shows the cumulative POT. The orange line shows the delivered POT weighted by detector up-time.

Chapter 4

MicroBooNE Simulation and Reconstruction

*“A computer once beat me at chess, but it was no
match for me at kick boxing.”*

Emo Philips

Analysis software is an essential part of any experiment. It defines how the events from the data are interpreted through reconstruction, and is an important ingredient in interpreting the results of the experiment through comparisons of the simulated prediction to the data.

This chapter will outline both how neutrinos are generated and propagated through the MicroBooNE detector, and will detail how the reconstruction used in this thesis is performed.

4.1 Overview

The MicroBooNE experiment makes use of the LArSoft (Liquid Argon Software) framework [72], which uses the *art* event-processing framework [73] for simulation and reconstruction of events in the MicroBooNE detector.

The *art* framework provides a number of utilities and templates which can be used to act on input data. It also provides a standard organisation of files, where data is split into runs, sub-runs,

and events. Each event then contains a set of experiment- (or user-) defined classes called *data products*.

LArSoft encompasses a body of code shared between many LAr experiments, and provides many standard C++ classes for use by experiments and users.

4.2 Neutrino Flux Simulation

The purpose of the flux simulation is to predict the number of ν_μ , ν_e , $\bar{\nu}_\mu$, and $\bar{\nu}_e$ coming from the BNB as a function of energy, per proton on target, and per unit area. MicroBooNE makes use of the MiniBooNE flux simulation (rewritten in C++) as documented in reference [62], with some modifications which include an additional dataset in the fit to K^+ production, and use of a spline fit to HARP data in order to estimate systematic uncertainties.

Simulation of primary protons incident on the target are performed by a Geant4-based Monte Carlo [74, 75]. This tracks the interactions of the protons with the beryllium (*Be*) target, and production and subsequent decay of secondary mesons and muons which result in neutrino production.

The secondary mesons and muons are tracked through the decay volume and truth-level information related to each neutrino which impinges upon a specified plane (the *flux window*) is saved so that systematic uncertainties can be estimated through a re-weighting procedure.

4.2.1 Fits to World Data

The hadron production cross sections in this simulation have been tuned to world data [76], primarily from:

- BNL E910 data for π^\pm , K^\pm , and K^0 production in $p + Be$ interactions (at 6.0, 12.3 and 17.5 GeV/c) [77].
- CERN HARP data for π^\pm , K^\pm , and K^0 production in $p + Be$ interactions (at 8.9 GeV/c) [78].

- SciBooNE data for K^+ production [79], which is especially important as this is performed in the BNB. This is the main way in which the BNB MC differs from the original MiniBooNE implementation.
- KEK K^0 production in the 12 GeV/c proton beamline [80].

The double-differential inclusive cross sections for secondary mesons (π^\pm , K^0) can be well described by a Sanford-Wang (SW) parameterisation:

$$\frac{d^2\sigma(p + Be \rightarrow S + X)}{dpd\Omega} = c_1 p^{c_2} \left(1 - \frac{p}{p_B - c_9} \right) \exp \left[-c_3 \frac{p^{c_4}}{p_B^{c_5}} - c_6 \theta (p - c_7 p_B \cos^{\epsilon_8} \theta) \right], \quad (4.1)$$

where p_B is the proton (“beam”) momentum, p is the outgoing meson momentum, and θ is the angle between the proton and meson. The $c_1 \dots c_9$ are parameters of the model which are constrained by world data. Importantly, c_9 is related to the momentum threshold in meson production, which is not fit for pions, but set to 1 GeV/c. For fits to π^+ data, c_3 is held at 1 GeV/c and for π^- , c_3 is held a 5.454 GeV/c.

Originally, MiniBooNE used the SW fit in order to extract uncertainties on the π^+ production, however this was shown to overestimate the systematic uncertainties, and so they developed a new method, based on spline fits to the HARP data, resulting in reduced systematic uncertainties.

Charged kaon production cross sections are not fit using the SW parameterisation, but are modeled using Feynman scaling (FS) instead [81]. Here, the invariant cross section is modeled only on the transverse momentum of the outgoing kaon, p_T , and the Feynman scaling factor, x_F , which is the ratio of the longitudinal to maximum longitudinal momentum in the center of mass frame,

$$x_F = \frac{p_L^{CoM}}{p_L^{CoM,Max}}. \quad (4.2)$$

The invariant cross section can then be written

$$E \frac{d^3\sigma}{dp^3} = c_1(1 - |x_F|) \exp[-c_2 p_T - c_3 |x_F|^{c_4} - c_5 p_T^2 - c_7 |p_T \times x_F|^{c_6}], \quad (4.3)$$

where E is the total energy, and $c_1 \dots c_7$ are once again parameters to be constrained by world data.

The results of the SW fit to π^\pm data from HARP and E910, and the FS fit to K^\pm data can be found in Figures 4.1, 4.2, and 4.3, respectively.

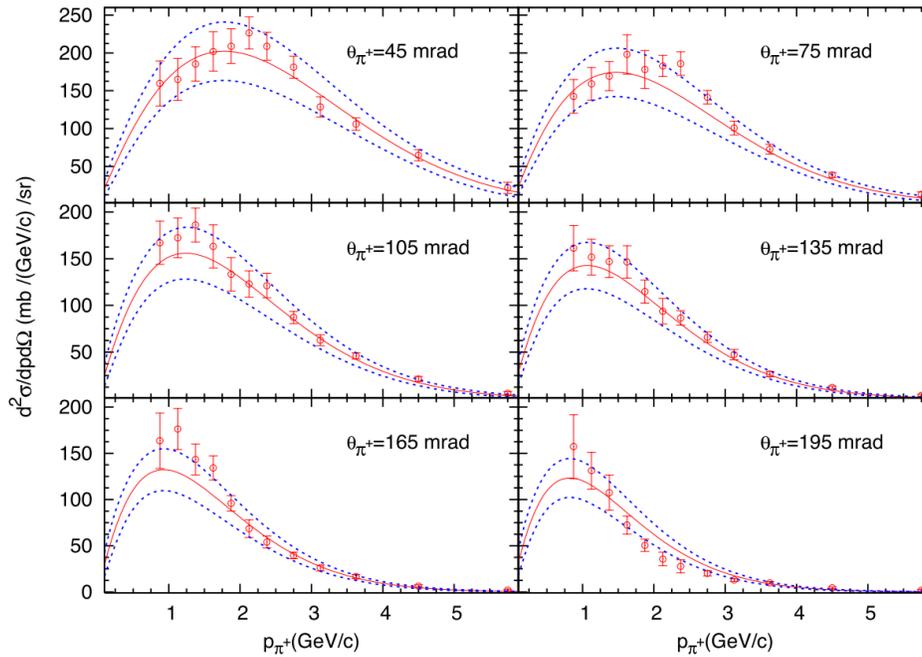
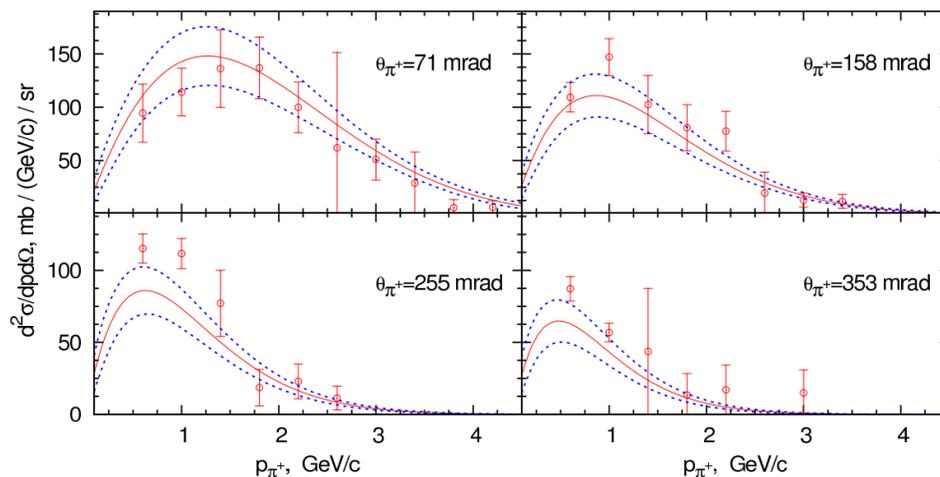


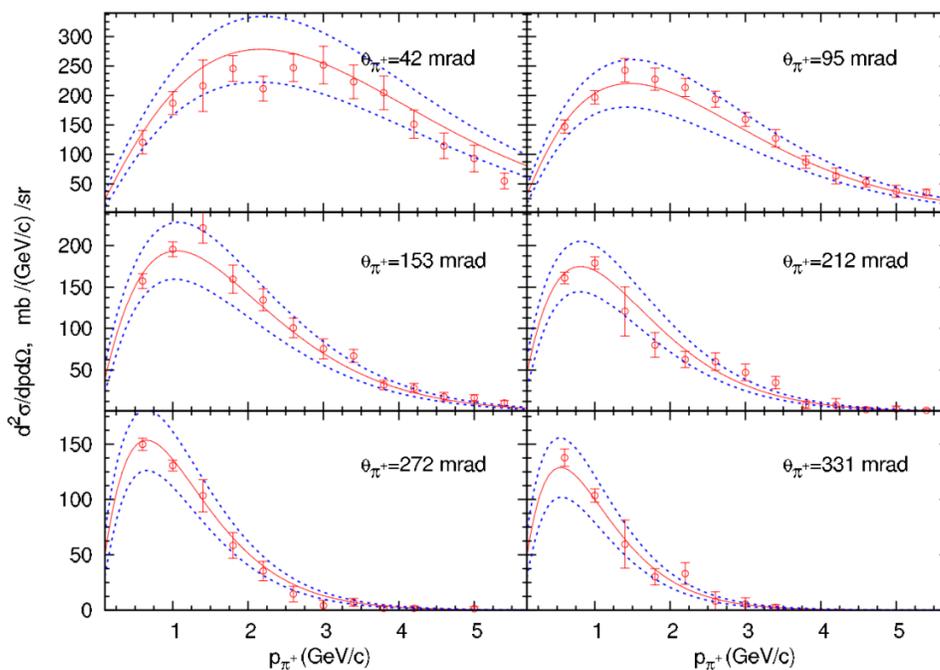
Figure 4.1: HARP π^\pm production data (with 8.9 GeV/c incident protons) compared with the best fit SW prediction (red, solid), and the $\pm 1\sigma$ bounds (blue, dashed). Taken from reference [82].

4.3 Neutrino Interaction Simulation

Neutrinos from the beam Monte Carlo are passed to the the GENIE neutrino generator [83, 84]. For the work presented in this thesis, GENIE version 2.12.0 has been utilised. MicroBooNE uses the default GENIE model set, with a number of chosen “alternate tunes” to assess the model dependence of analyses. Here, the default model set is described.



(a) 6.4 GeV/c incident protons



(b) 12.3 GeV/c incident protons

Figure 4.2: E910 π^+ production data compared with the best fit SW prediction (red, solid), and the $\pm 1\sigma$ bounds (blue, dashed) for 6.4 GeV and 12.3 GeV/c incident protons. Taken from reference [82]

In the default model set, the simulation of the nuclear environment is based on the Bodek-Ritchie relativistic Fermi gas model (RFG) [85]. A standard RFG models the nucleus as a set of non-

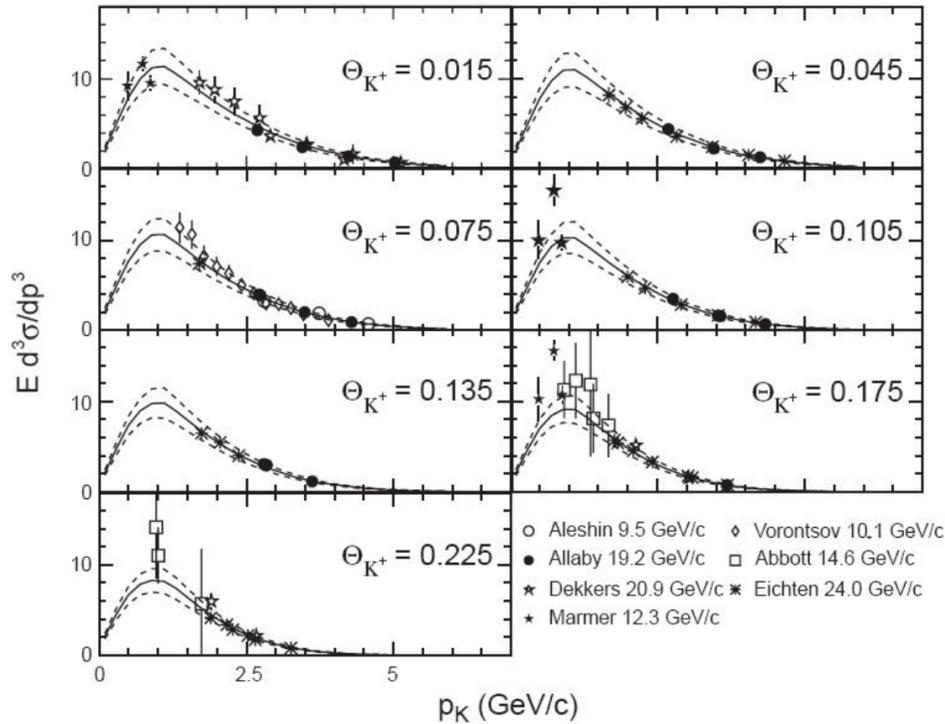


Figure 4.3: Feynman scaling fit to 8 experiments which have been scaled to $p_T = 8.89$ GeV/c. Note that these plots are taken from reference [82], which precedes the results in reference [79], and so they do not include the dataset from SciBooNE. Taken from reference [82]

interacting fermions, and assumes that the only interactions between fermions are due to Pauli blocking. In this way, nucleons are stacked up in energy levels leading to a uniform distribution of particle momenta up to the Fermi momentum, p_F . The Bodek-Ritchie modification takes into account short-range correlations between nucleons by introduction of a phenomenological tail at high-momentum.

The energies of neutrinos coming from the BNB are generally around 700 MeV, and this means that MicroBooNE cares primarily about quasi-elastic (QE) and resonant (RES) interactions (see Section 2.3), the models for which are described below.

GENIE uses the Llewellyn-Smith formalism for QE interactions [86]. This model is built upon the idea of *form factors*, which describe the spatial distributions of electric charge within the nucleon. Several such form factors have been measured from electron scattering experiments,

however the axial vector component,

$$F_A(q^2) = \frac{F_A(0)}{\left(1 - \frac{q^2}{M_A^2}\right)^2}, \quad (4.4)$$

cannot be measured because electron scattering only occurs through exchange of a photon. $F_A(0)$ has been measured from β decay to high precision, and so this means that there is one dominant uncertainty in this formalism of QE neutrino interactions: the value of the axial mass, M_A .

This model deals with interactions on unbound nucleons, and so GENIE deals with the binding energies internally.

This version of GENIE also includes an empirical Meson Exchange Current (MEC) model, which increases the power of the QE cross section in the relevant regions of muon kinematic phase space in accordance with the MiniBooNE ν_μ CCQE data. For these events, an additional nucleon produced at the point of interaction, however this has neither a theoretical or experimental motivation.

Resonant production of pions in neutrino interactions is modeled by a GENIE-specific version of the Rein-Sehgal model [87].

4.4 Cosmic Particle Simulation

MicroBooNE is effectively located on the surface, leaving it open to large cosmogenic backgrounds¹. This, combined with the slow readout of LArTPCs means that simulation of cosmic particles must be given consideration. MicroBooNE has investigated several packages and package configurations in reference [88].

Simulation of cosmogenic particles is done by the Cosmic Ray Simulations for Cascade (CORSIKA) package [89], using the Constant Mass Configuration, which simulates interactions of p ,

¹cosmogenic backgrounds are those particles produced when cosmic rays interact in the upper atmosphere and produce a shower of secondary particles. In general these are observed in the detector as muons which tend to be close to vertical.

He, *N*, *Mg*, and *Fe* in the atmosphere. The choice was also made to use the FLUKA package for hadronic interactions below 80 GeV instead of the default GEISHA package.

One drawback of the cosmic-ray simulation at MicroBooNE is that there are currently no systematic uncertainties associated with the simulation. This will be addressed in upcoming analyses by overlaying simulated neutrino events on top of cosmic data, however this has not been leveraged for this analysis.

4.5 MicroBooNE Detector Simulation

After the particle production, the particles are propagated through the LAr using Geant4 and are passed to the detector simulation. During this stage, the detector simulation has to combine the effects of the initial electron ionisation with recombination, diffusion, attenuation, and the detector response, as well as simulating scintillation light.

4.5.1 TPC Simulation

Simulation of the TPC begins by taking groups of up to 600 ionisation electrons and instantly transporting them to the readout wires at the anode, taking into account drift-dependent effects such as electron diffusion and electron lifetime, and are grouped into an energy deposition per readout channel.

Once an energy deposition per readout channel has been produced, the detector response is applied to the simulated signal.

The detector response can be thought of as being a convolution of an electronics response and a field response.

The electronics response captures how the electronics respond to a delta-like signal, and has been measured from the data using a pulser, whereby a delta function-like signal can be input into the ASICs and the electronics response can be measured in the absence of the field response.

The field response then encapsulates the drift of the ionisation electrons, which leads to in-

duced currents on the readout wires. It is measured from the MicroBooNE data by taking averaged signal pulses originating from near to the anode plane (so as to neglect drift-dependent effects) and removing the measured electronics response. This is done as a function of Y-Z space in the detector so as to capture regions which do not have a nominal response. This method is known to have a bias which over-estimates the width of the field response due to the averaging of waveforms, however for most analyses this effect is calibrated out at a later stage.

After the signal has been generated with the correct response, noise is applied. This simulation uses a data-driven implementation of the noise model, which takes into account the magnitude of the noise in the time-domain and the shape of the noise in frequency space, as a function of wire length.

In order to model this noise, data with the high voltage turned off is used. This is done in order to minimise the effect of signal pulses. The shape of the distribution is measured by taking an average of a number of collection plane wires, taking the Fourier transform, and performing a fit to the resultant spectrum. A single representative collection plane wire is then taken, ensuring that there is no signal present on the channel and a Poisson distribution is fit in bins of equal width in frequency (84 kHz bins) in order to capture the changing magnitude of the Fourier transform across the range. This procedure is shown in Figure 4.4.

This information can then be used to construct a noise model in frequency space, and the inverse Fourier transform can be taken to give the final noise spectrum in the time domain. An example waveform is shown in Figure 4.5.

Note that this fit only needs to take into account frequencies up to 1 MHz because MicroBooNE samples 9600 ticks in approximately 4.8 ms: a frequency of 2 MHz. Nyquist theorem then states that the highest frequency able to be resolved is

$$v_{Nyquist} = \frac{1}{2} v_{sampling} = \frac{1}{2} \cdot 2MHz = 1MHz \quad (4.5)$$

In general, the inherent noise from the ASIC is a function of the capacitance of the system, which includes the cables and connections to the ASIC, and importantly the readout wire [68].

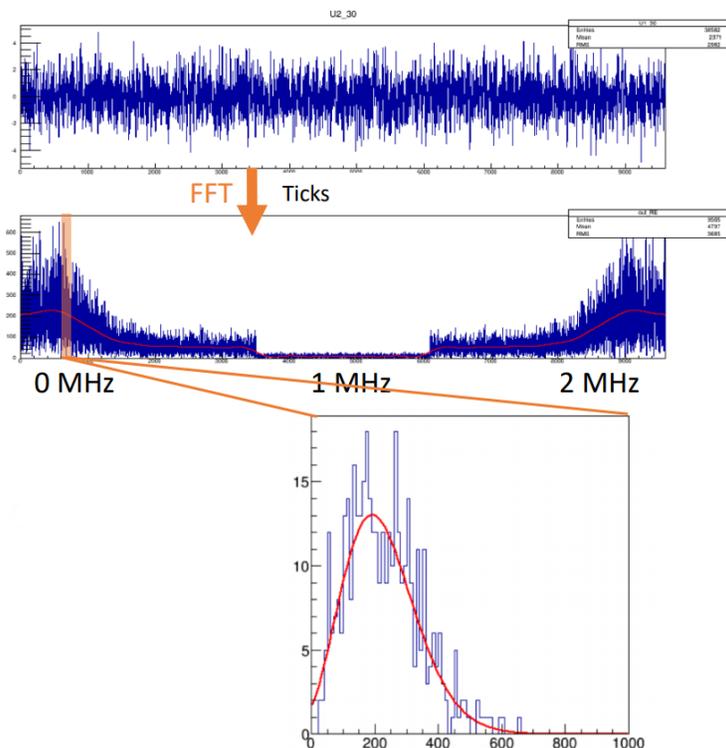


Figure 4.4: Example of the fitting procedure used to extract the data-driven noise model used in the MicroBooNE simulation. The shape of the spectrum in frequency space is a feature of the way the Fast Fourier Transform (FFT) is performed, resulting in the real part of the waveform being represented between 0 and 1 MHz, and the imaginary part of the waveform being represented by the remaining 1 MHz to 2 MHz part. The reduction in magnitude from 730 kHz to 1270 kHz is due to removal of high-frequency noise. In practice, this part of the spectrum is filtered during reconstruction, and so there is no effort made to reproduce this part of the spectrum.

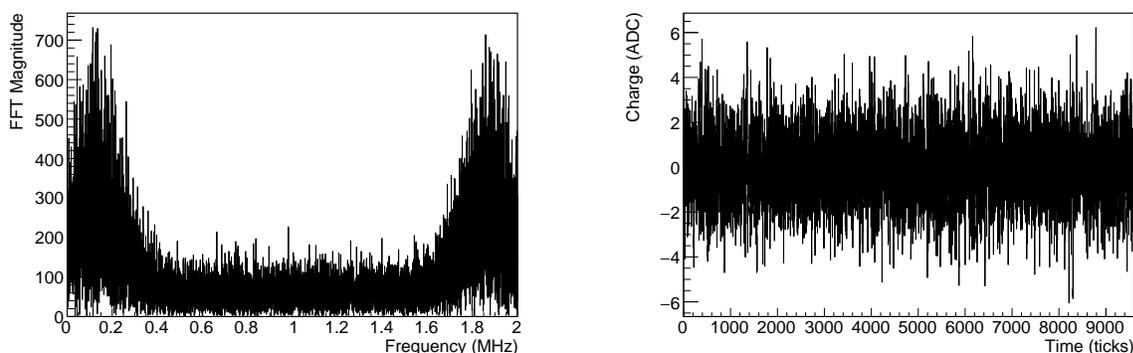


Figure 4.5: Data-driven noise spectrum from MicroBooNE simulation in both the frequency- and time-domains.

This indicates that the level of noise on each readout wire is correlated with its length. This effect is also contained within the model, and the resultant noise level as a function of the channel number can be found in Figure 4.6.

A further effect included in the simulation is the Space Charge Effect (SCE), described in Chapter 3.2.2. This is encapsulated through the use of a position map and an electric field map which are generated from a model of SCE in the MicroBooNE detector. The first of these modifies the position of energy depositions in the detector as a function of the deposition's position in the detector. The second uses the predicted electric field at each position within the detector to ensure the correct amount of recombination takes place.

There is one significant deficit of the MicroBooNE TPC simulation: *induced charge* effects are currently not simulated. The assumption is made that only charge drifting within ± 1.5 mm of the primary readout wire is detected by the wire. In fact, the primary readout wire also receives a significant contribution from wires within ± 2 wires on the second induction plane and the collection plane, and within ± 10 wires on the first induction plane, as shown in Figure 3.6 and demonstrated in Figure 4.7.

Because these effects are not included in the simulation but are present in the data, the reconstruction is known to impact the two data sets differently. A preliminary simulation of induced charge is currently used to estimate a systematic uncertainty resulting from this deficiency in the simulation.

4.5.2 Optical Simulation

The propagation of scintillation light is not performed on a photon-by-photon basis due to computational challenges related to the yield of scintillation light on LAr. Instead, the light production is parameterised through use of a photon library. The TPC is first segmented into 3-dimensional segments known as *voxels* ($75 \times 75 \times 400$) and a photon bomb is simulated in each voxel. A full simulation is performed, and every photon is tracked through the TPC until it leaves or is collected by the PMTs. The photon library then stores information related to each PMT and its acceptance

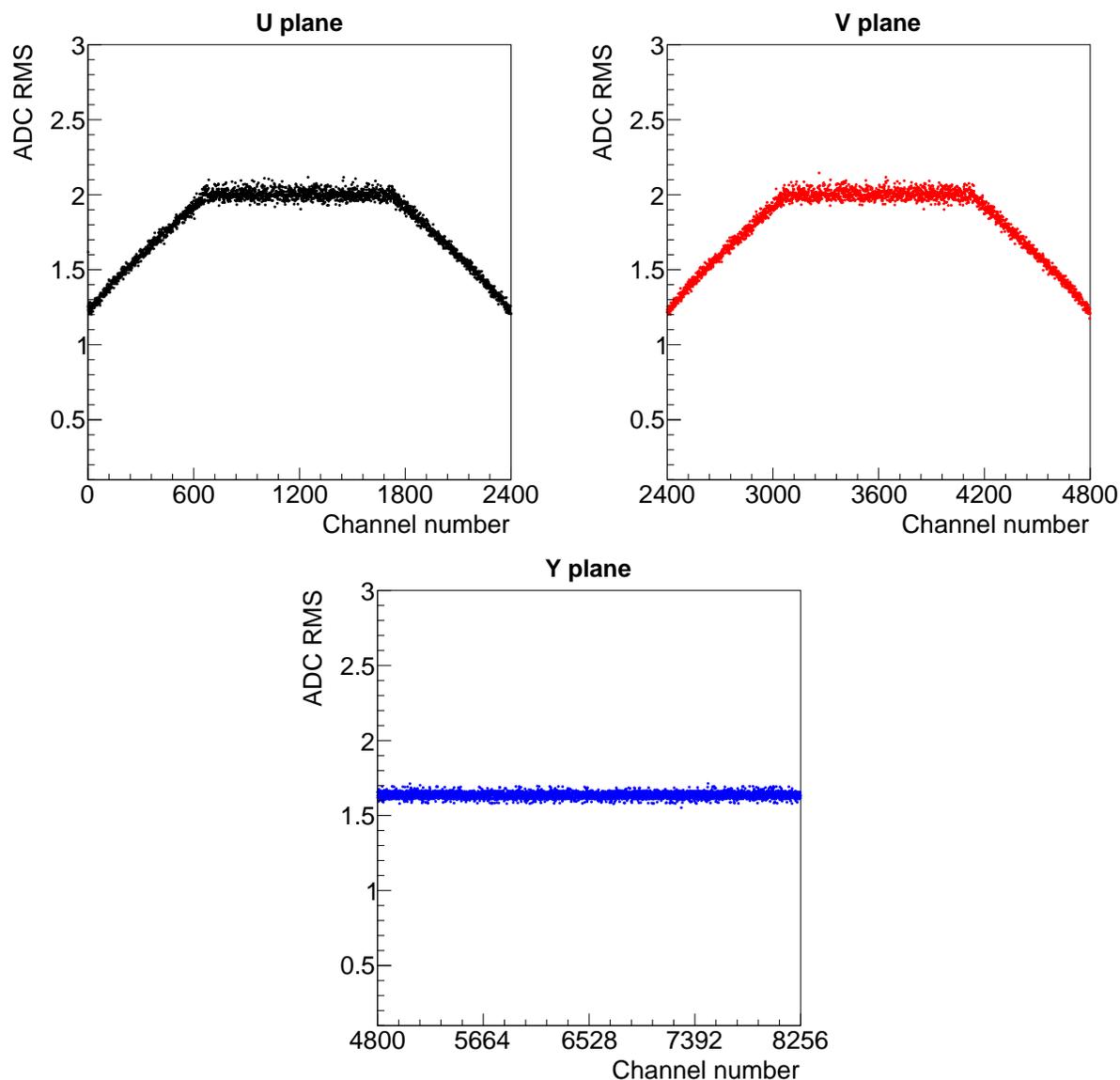


Figure 4.6: The noise level, measured in the RMS of the ADC values, as a function of channel number. The trapezoidal shape of the distributions on the U and V planes are due to the length of the wires becoming shorter at high and low channel number due to the angles of the wires ($\pm 60^\circ$ to the vertical). The collection plane, on the other hand, are vertical, leading to a flat distribution across the channels.

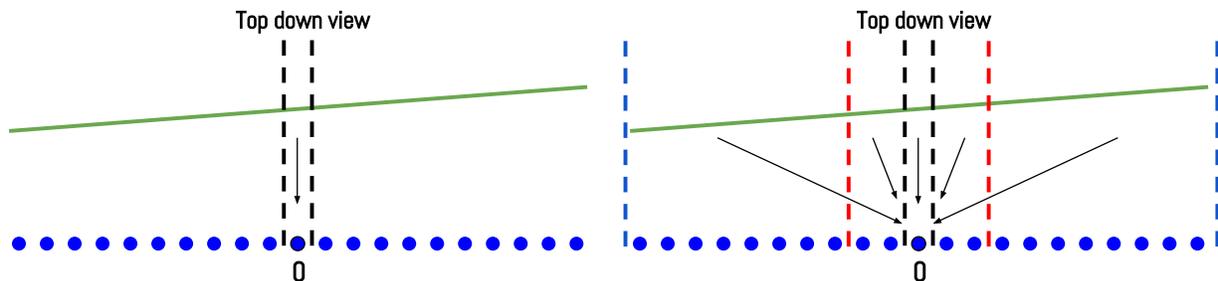


Figure 4.7: Demonstration of induced charge effects. The current MicroBooNE simulation assumes only charge drifting within ± 1.5 mm is detected by the primary readout wire, however induced charge from charge drifting within the ± 10 wires region can also contribute significant signal.

of photons in each voxel.

4.6 Reconstruction

Reconstruction is the act of taking the raw output of the detector and processing it in such a way that physics measurements can be performed.

4.6.1 Signal Processing

MicroBooNE has spent a significant amount of time developing novel signal processing techniques. References [67] and [90] in particular go into great depth about the techniques employed by the experiment, but a brief overview is presented here.

The first stage of the signal processing involves reconstruction of the charge detected on each readout wire. This is done through the process of *deconvolution*. This is a method of extracting the true signal $S(\omega)$ from a measured signal, $M(\omega)$ by removing the response function $R(\omega)$,

$$S(\omega) = \frac{M(\omega)}{R(\omega)} \cdot F(\omega), \quad (4.6)$$

where ω is in units of angular frequency, and $F(\omega)$ is a filter function, which stops the Fourier

transform of $S(\omega)$ becoming overwhelmed by noise². The result of including the filter function is that, provided the response function well approximates the data, the output signal has a shape defined by the filter function. MicroBooNE chooses to use a Gaussian filter, meaning that the shape of the deconvolved waveforms approximates a Gaussian functional form. There are significant benefits to this choice of filter, which are outlined in [67] and not detailed here, however one result of this filter which is worth noting is that the shape of the induction plane signals is modified from a bipolar shape to a monopolar Gaussian shape which aid in downstream reconstruction.

This procedure can be thought of as deconvolving in one dimension: time, but can be expanded to a second dimension: the wires. In this way, induced charge effects can be taken into account in the signal processing. This is discussed in some depth in reference [67], and a comparison of one-dimensional and two-dimensional deconvolution can be found in Figure 4.8. This two-dimensional technique is not utilised in current MicroBooNE analyses, but will be made use of in future iterations.

In order to reduce processing time, the deconvolution process is only applied in specific Regions Of Interest (ROIs) which are identified by an ADC threshold in the time domain on the raw waveforms.

After the deconvolution process, the resulting deconvolved ROIs contain signal which is approximately Gaussian in shape. A procedure of peak finding is then performed in order to find candidate signals, and Gaussian functional forms (*hits*) are fit to the identified signals. These form the basis of the downstream reconstruction.

4.6.2 The Pandora Pattern Recognition Software

This work makes use of the Pandora multi-algorithm pattern recognition framework[91]. This is a particle flow style reconstruction where it takes reconstructed hits as an input and produces Particle Flow Objects (PFOs), which contain Particle Flow Particles (PFPs) as output. Particle flow

²The response function $R(\omega)$ decreases substantially at high frequency, but as shown in Figure 4.5, the noise spectrum persists at high frequencies. The result is that when $M(\omega)/R(\omega)$ is calculated, the contribution from noise at high frequencies is amplified.

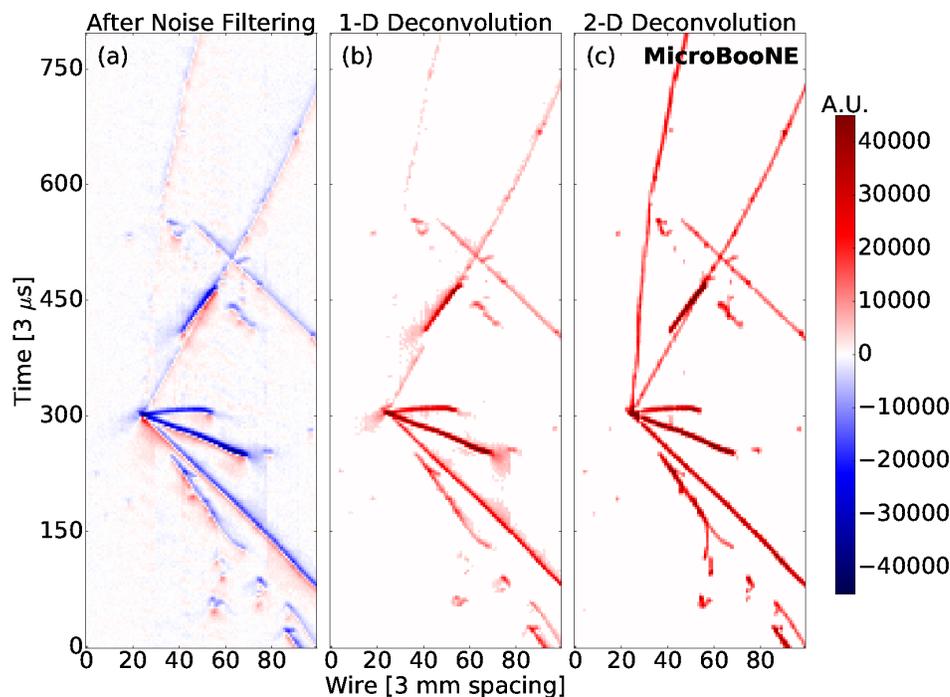


Figure 4.8: Example of a one-dimensional deconvolution compared to a two-dimensional deconvolution for a single data event. When the more sophisticated two-dimensional deconvolution is applied, signal pulses become much more well defined. Figure taken from reference [67].

reconstruction means that each reconstructed object is related to other reconstructed objects in a hierarchy. A PFP is a representation of a reconstructed object which has information about related PFPs; a related PFP may be either a parent (closer to the reconstructed neutrino vertex), or a child (further from the reconstructed neutrino vertex).

The pandora reconstruction chain consists of two passes. The first pass is tuned to target reconstruction of cosmic rays and is more strongly track-oriented, attempting to reconstruct tracks of cosmic origin (with the assumption of tracks being downward going) with daughter delta rays being reconstructed as showers. The second pass is tuned to reconstruct neutrinos coming from the BNB. Candidate neutrino vertices are first identified, and this is used as a handle to reconstruct the particles emerging from the candidate vertex with a preference for reconstructing forward going trajectories. Tracks and showers are more carefully identified in this pass.

Both the cosmic and the neutrino pass undergo functionally very similar processes, and these

are described below.

Two-Dimensional Reconstruction

Firstly, a two-dimensional reconstruction is applied. Hits are clustered together on each plane into collections which are unambiguous straight lines, with a preference for maintaining cluster purity rather than completeness³. That is: every time there is some significant change in direction or other ambiguity, a new cluster is produced. A series of cluster algorithms are then applied to these two-dimensional clusters in order to merge them, increasing the cluster completeness without sacrificing its purity. This is done by identifying clusters which are in close proximity or are pointing at each other. In the case of the neutrino pass the two-dimensional cluster information is also used to identify a candidate neutrino interaction vertex, which is used for downstream reconstruction, and are used to identify each cluster as track-like or shower-like.

Three-Dimensional Reconstruction

Secondly, a three-dimensional track reconstruction is applied. The challenge here is taking the clusters determined by the two-dimensional reconstruction and matching them across planes. Firstly, clusters are matched across planes using the time co-ordinate, which is common between the three planes. These matches are then interrogated by a sophisticated suite of algorithms which aim to identify any ambiguities (for instance, two clusters in the collection plane are matched to single clusters in the induction planes, or a single cluster in the collection plane is matched to two clusters on both of the induction planes), and this information can be used to split or merge clusters in a repetition of the two-dimensional reconstruction. This back and forth between the two- and three-dimensional reconstruction is repeated until there are no ambiguities to be addressed. The result is a series of reconstructed showers and tracks.

³Essentially, this begins as a rather conservative approach, whereby a cluster is preferred to be incomplete (a true particle may be split across multiple clusters), so long as the cluster is pure (there are few hits which are produced by a different particle)

Particle Flow Object Creation

The final step in the neutrino pass is construction of a PFO, which is done by first creating a PFP for the neutrino and identifying this with the reconstructed neutrino vertex. Reconstructed primary particles are then identified and included as daughters of the reconstructed neutrino PFP, and the process is repeated until every particle has a parent and is stored in the PFP hierarchy. An example hierarchy for a simulated interaction is displayed in Figure 4.9.

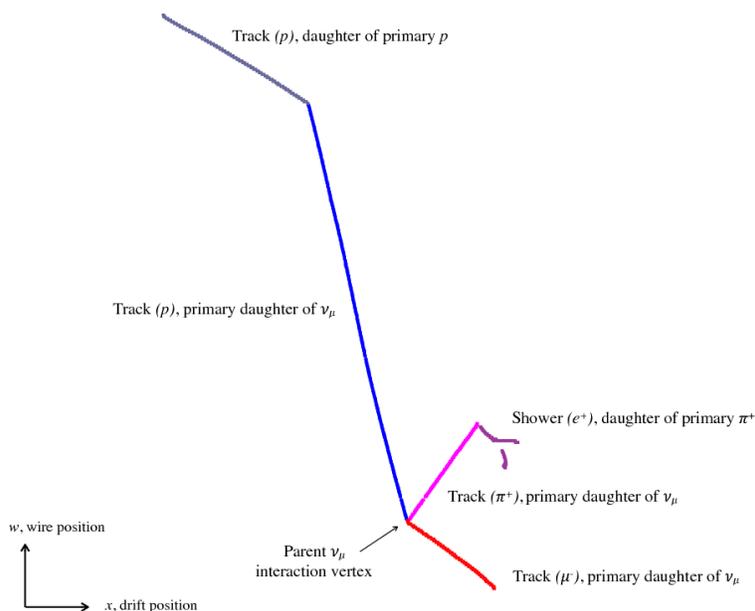


Figure 4.9: Example simulated ParticleFlow Object which contains a PFP hierarchy.

4.6.3 Optical Reconstruction

The objective of the optical reconstruction is to take in raw waveforms from the optical read-out, and produce a high-level reconstructed object, an *optical flash*, which characterises the light information from an interaction in the TPC in a given time period.

Combination of High-Gain and Low-Gain Channels

The first stage in the optical reconstruction is the combination of the high- and low-gain channels into a single gain-corrected waveform. In the case that the high-gain channel is saturated, the

low-gain channel is used with a gain correction factor applied. If the high-gain channel is not saturated then the waveform is taken from this stream only.

Optical Hit Reconstruction

The gain-corrected waveform is taken and optical hits are reconstructed.

The first step of this process is finding the baseline. For the cosmic discriminator, this is simply taken to be the ADC of the first sample. For the beam discriminator, the baseline is calculated taking a rolling mean in a sliding window, and extrapolation is performed to find the baseline throughout the signal region.

Once a baseline has been established, the signal region is interrogated to find the peak, width and area of the pulse, amongst other things, and this information is stored in an optical hit object.

Optical Flash Reconstruction

Once optical hits have been reconstructed, these are grouped into a high-level optical flash. This is done by looking for optical hits which are coincident in time with each other. Once these have been found, a dead time of $8 \mu\text{s}$ is applied such that no other optical flashes may be reconstructed in that region.

4.7 Energy Scale Calibration

The objective of the energy scale calibration is to normalise the calorimetric response across the detector, and to reconstruct the energy deposition per unit length (dE/dx) from the ADC counts which are read out from the detector.

The energy scale calibration actually encompasses two calibrations. The *dQ/dx calibration* acts to make the detector response uniform across the detector volume, while the *dE/dx calibration* acts to set the absolute energy scale of the detector.

4.7.1 Misconfigured Region Correction

Misconfigured regions of the detector are first accounted for by deconvolving the misconfigured gain and shaping time and re-convolving with the desired gain and shaping time. The result of this is that channels connected to misconfigured ASICs have a larger fractional noise contamination than their correctly configured counterparts, however signal is still able to be picked out above the noise baseline.

4.7.2 dQ/dx calibration

Many effects which can introduce non-uniformities in the detector have been discussed: in the Y-Z plane, these are introduced through misconfigured ASICs, shorted regions, and channel-to-channel gain variations, while in the drift direction, these are primarily due to effects like electron attenuation and electron diffusion.

To perform this calibration, MicroBooNE makes use of cosmogenic muons. The charge deposition per unit length (dQ/dx) is then mapped in three dimensional volumes (*voxels*) within the detector, and a three dimensional correction map is generated.

4.7.3 dE/dx calibration

Once the detector has been given a uniform response, the absolute energy scale can be set.

Conversion from dQ/dx to dE/dx is performed using the modified box model of recombination[92],

$$\frac{dE}{dx}^{calib.} = \frac{\exp\left(\frac{\frac{dQ}{dx}^{calib.}}{C} \cdot \frac{\beta_p W_{ion}}{\rho\epsilon}\right) - \alpha}{\frac{\beta_p}{\rho\epsilon}}, \quad (4.7)$$

where

- W_{ion} is the work function of Argon (23.6 MeV/electron).
- ϵ is the electric field (0.273 kV/cm).

- ρ is the density of LAr (1.38 g/cm^3).
- α and β are the recombination constants measured by the ArgoNeuT experiment at 0.481 kV/cm ($-0.93 \pm 0.002 \text{ kV/cm}$, -0.93 ± 0.02).
- C is the calibration constant used to convert ADC to a number of electrons.

It is then clear that to set the absolute energy scale, the calibration constant, C , must be measured. To do this, many reconstructed dE/dx distributions are built as a function of the distance to the end of the track (the *residual range*), and are fit with a Landau-convoluted Gaussian distribution in order to extract the most probable value of the distribution in each residual range bin. A two dimensional residual range- dE/dx distribution can then be constructed, and a simple χ^2 can be minimised between this and the theoretical curve by modifying C .

4.8 Continuous Integration Validation

MicroBooNE uses a Continuous Integration (CI) system to perform validation of its simulation and reconstruction packages. This package was developed in part by the author.

Each time a release is tagged, 10,000 cosmic events and 10,000 BNB events are automatically generated and reconstructed, and a series of variables of interest are plotted against results from the previous week and results from a configurable base release.

The first of these comparisons is used to identify any unexpected changes in a timely manner and so helps with quick implementation of bug fixes. The second of these comparisons is used to avoid a “slow-drift” where incremental changes occur over a number of weeks.

These plots are then monitored by experiment shifters, and CI experts liaise between the shifters and simulation and reconstruction developers.

The two streams (“cosmic” and “BNB”) are leveraged for their different strengths. The cosmic stream allows for easy recognition of changes to the detector simulation and some high-level reconstruction effects such as having a handle on how often Pandora splits single particles into

multiple tracks. It can also be used to perform quick and easy data/simulation comparisons⁴. The BNB stream simulates only BNB interactions with no cosmics, and so it can be used to estimate the neutrino vertex resolution and reconstructed-to-truth match purity and completeness (See Appendix B).

In addition to incremental validation of the simulation and reconstruction release, the CI validation framework is used to do detailed validation of candidate production releases of the simulation and reconstruction.

⁴The BNB+Cosmic stream is not ideal in this case because every simulated event has a neutrino event, where the data does not

Chapter 5

Longitudinal Electron Diffusion in MicroBooNE

*“Chaos is found in greatest abundance wherever
order is being sought. It always defeats order,
because it is better organized.”*

Terry Pratchett

5.1 Diffusion in LArTPCs

Ionisation electrons in MicroBooNE travel under the influence of an applied electric field to an anode plane where signals are read out from three readout wire planes. During transport, the shape of the electron cloud is modified by several physical processes. This in turn means that the signal pulses which are measured at the wire planes change as a function of the distance traveled by the electron cloud. In particular, electron diffusion acts to smear the initial cloud. It is known that under the influence of an electric field, diffusion is non-isotropic [93, 94, 95], and in general this is parameterised in terms of a transverse component (D_T , perpendicular to the drift direction) and a longitudinal component (D_L , parallel to the drift direction). The longitudinal and transverse components can differ significantly, as shown in Figure 5.1 and noted in reference

[95]. It should be noted that the values of D_L and D_T shown in Figure 5.1 are taken from a fit to world data performed in reference [95], however this is not what is nominally implemented in the MicroBooNE simulation, where the diffusion values correspond to MicroBooNE’s original design electric field, 500 V/cm.

Because D_L and D_T are expected to be different, each component must be measured separately in order to minimise the impact of the other. D_T impacts the spatial resolution of a Time Projection Chamber (TPC) in the plane transverse to the readout wires as a function of drift distance, while D_L impacts the spatial resolution in the electron drift direction; see Figure 5.2. Measuring D_L can also be useful in tagging the t_0 (actual arrival time) of tracks in TPCs, as shown by analysis carried out by the DUNE 35 ton prototype experiment [96].

This measurement is an important one in MicroBooNE. The current uncertainties on D_L , estimated from world data, are large $6.2^{+57\%}_{-47\%}$ cm²/s, and this has proven to be a significant effect in shower reconstruction. This constitutes one of the larger detector-related uncertainties in the ν_μ CC π^0 cross section measurement [97]. This is then an important parameter to constrain in the context of a ν_e search, such as MicroBooNE’s low-energy excess investigations.

Currently, there are few measurements of electron diffusion in liquid argon in the literature: measurements of the transverse component of diffusion have been reported for fields above 1500 V/cm, initially by Derenzo, et al. [98], followed by Shibamura, et al. [99], while measurements of the longitudinal component have been reported by Cennini, et al. [94] and Li, et al. [95] for fields between 100 and 2000 V/cm. It should be noted that, for E-fields between 100 and 350 V/cm, Cennini’s measurements show reasonable agreement with theory [93]; Li’s measurements, however, are systematically higher than both. The ICARUS experiment is currently in the process of refining its measurement of the longitudinal diffusion coefficient, and the most recent results can be found in reference [100].

A summary of the current world data can be found in Figures 5.3, 5.4, and 5.5. Reference [95] reports results in terms of electron energy, ϵ_L , which is related to D_L by the equation

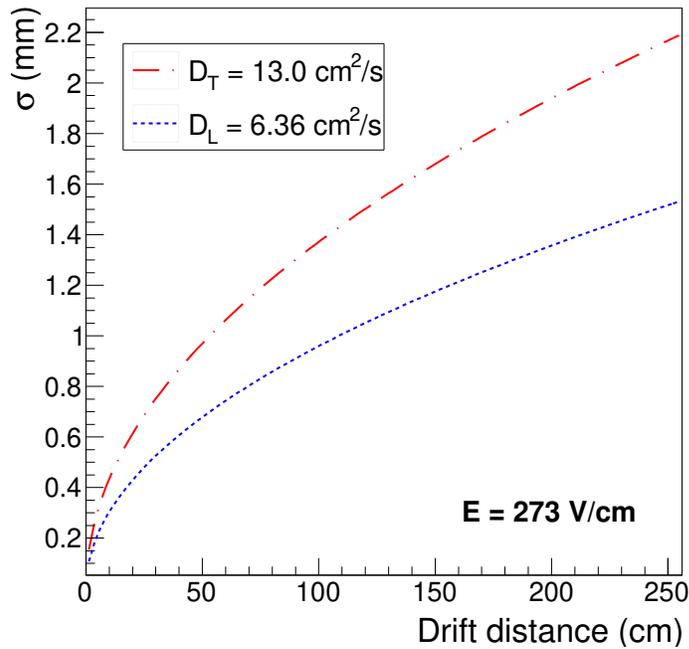


Figure 5.1: Expected contribution to the width of the ionisation electron cloud (σ) from transverse (red dot-dashed), and longitudinal (blue dotted) diffusion as a function of drift distance at MicroBooNE’s nominal electric field, 273 V/cm. The longitudinal diffusion value used here is calculated using the parameterisation outlined in reference [95]. Due to the shortage of transverse diffusion measurements at MicroBooNE-scale electric fields in liquid argon, the transverse diffusion value is estimated from the treatment in Atrazhev and Timoshkin [93]; however this is in disagreement with data reported in Derenzo, et al. [98] and Shibamura, et al. [99].

$$D_L = \frac{\mu \epsilon_L}{e} \quad (5.1)$$

where μ is the electron mobility — which characterises how quickly an electron can pass through a material in an applied electric field, and is measured in $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ — in liquid argon and e is the electron charge. For fields below ~ 0.3 kV/cm, electron mobility changes relatively slowly in liquid argon; for an order-of-magnitude conversion to D_L below these electric fields, one can take ϵ_L and multiply it by a factor of $500 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$. Note, however, that this relationship rapidly changes at higher electric fields, as shown in Figure 5.4, and so this approximate relationship no longer applies.

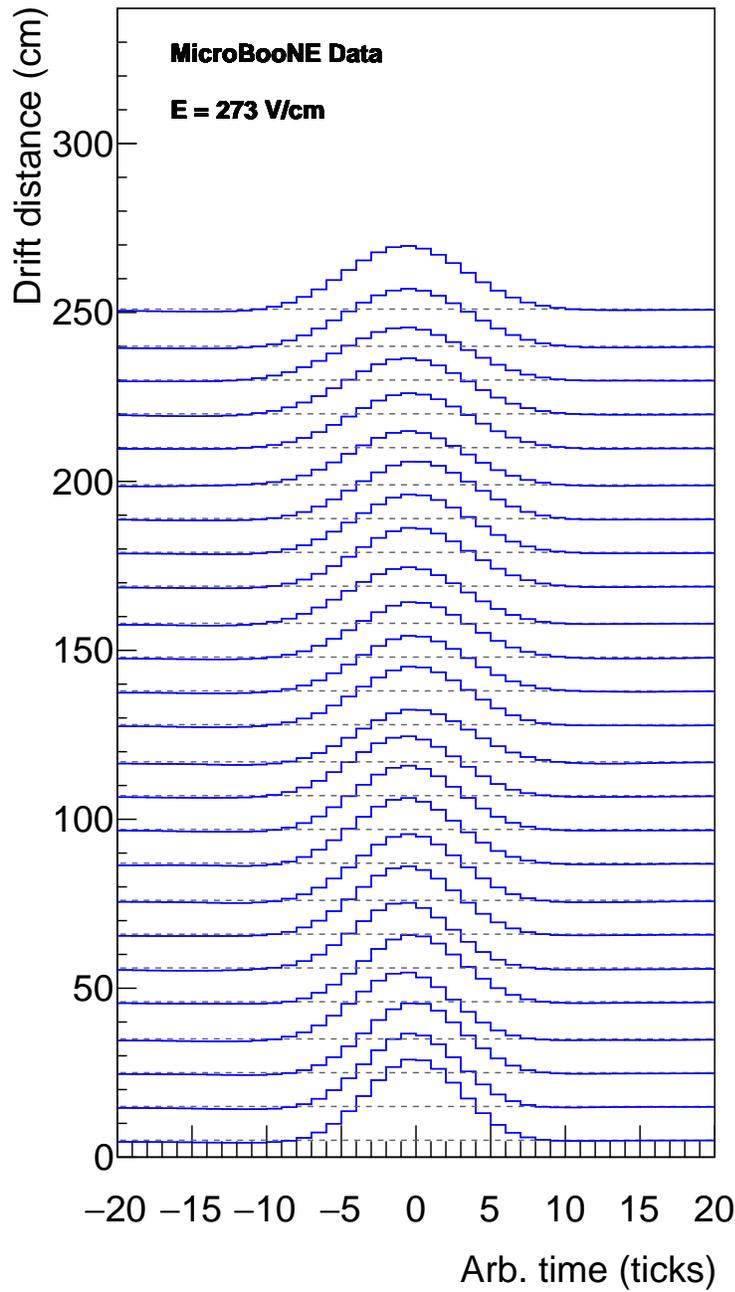


Figure 5.2: Averaged waveforms as a function of drift distance (see Section 5.2 for more details on the waveform averaging technique), taken from MicroBooNE Run 1 beam-off data. The x -axis on this plot is measured in ticks (1 tick = $0.5 \mu\text{s}$), but is arbitrary in the sense that all waveforms have been aligned at zero in order to aid in comparison of waveforms. For the same reason, the pulse heights are also arbitrary. As the drift distance shown on the y -axis increases, the effects of diffusion become more pronounced and the pulses become more smeared.

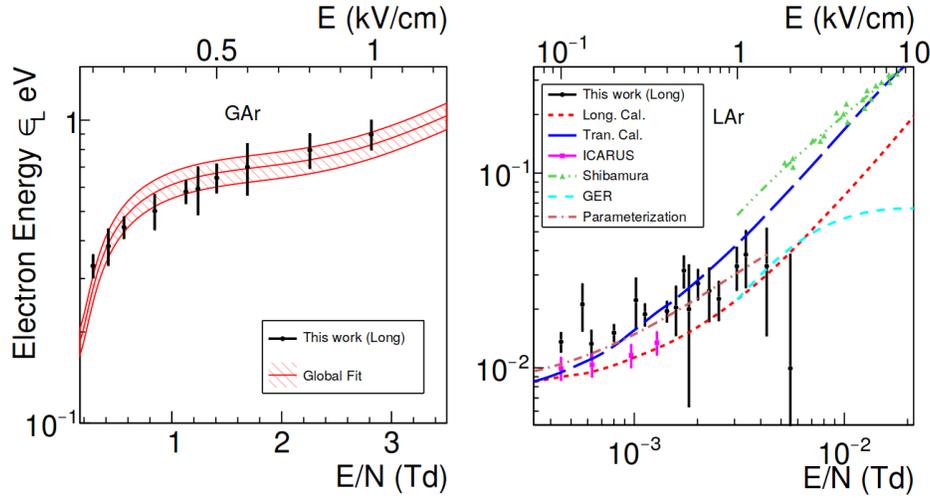


Figure 5.3: The measured electron energy, which is proportional to D_L , as a function of the electric field for gaseous argon (left) and liquid argon (right), taken from reference [95]. This also shows results from ICARUS, and results from reference [99], although the latest ICARUS results can be found in Figure 5.5. It is clear that there is a significant spread in the values claimed from each measurement.

The focus of this analysis is the measurement of the longitudinal diffusion constant, D_L . This can be described to first order by the equation

$$\sigma_t^2(x) \approx \sigma_t^2(0) + \left(\frac{2D_L}{v_d^3} \right) x, \quad (5.2)$$

This, in effect, means that the squared time width of a signal pulse measured in μs^2 , σ^2 , can be related to some inherent pulse width squared, $\sigma_t(0)^2$, plus some width squared which is influenced by D_L , and the drift velocity, v_d , as a function of drift distance, x .

5.2 Method for Effective Longitudinal Diffusion Coefficient Extraction

Equation 5.2 shows that to first order there is a linear relationship between the drift distance of a waveform and the square of its width. To extract the longitudinal electron diffusion constant,

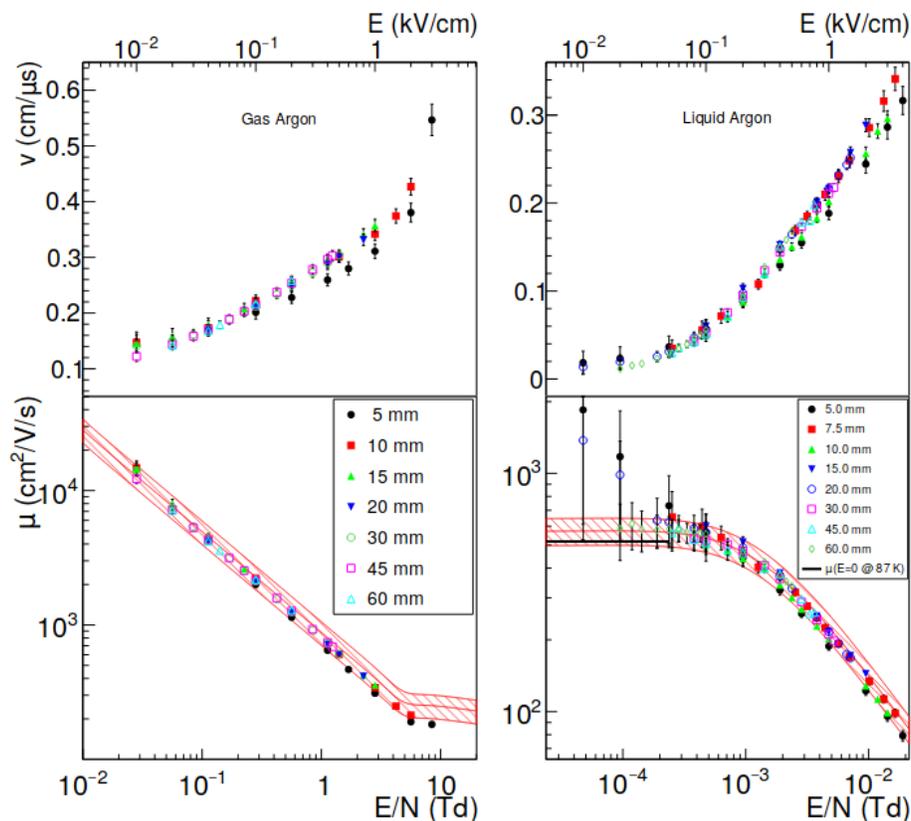


Figure 5.4: World data for electron drift velocity (top), and a fit to world data for electron mobility (bottom), taken from reference [95]. The left plots are for gaseous argon, the right for liquid argon. The fit for liquid argon is limited by the expected μ at 0 electric field with $T=89$ K, 551.6 cm²/V/s.

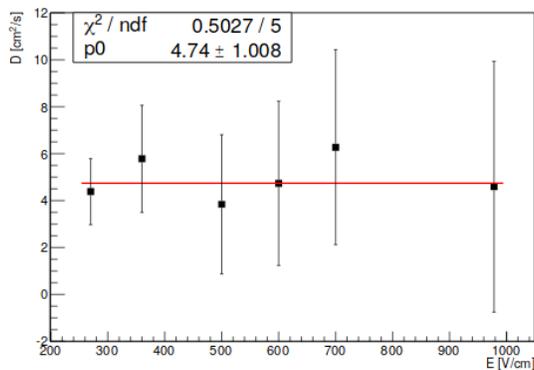


Figure 5.5: The latest ICARUS results, taken from reference [100]. This is taken across several different electric fields between approximately 250 V/cm and 1000 V/cm, and shows relatively little dependence. The uncertainties here are explained to be largely statistical.

therefore, a straight line can be fit to a plot of σ_t^2 vs. x .

This analysis makes use of deconvolved waveforms. These are zero-suppressed waveforms with Regions Of Interest (ROIs) preserved around peaks which exceed a threshold. This is displayed in Figure 5.6. The deconvolution process is described in some detail in Chapter 4.6.1.

For this analysis a Gaussian functional form is fit to the sum of deconvolved collection plane waveforms as a function of drift distance. In practice, this is performed by separating the drift direction into a number of bins (*drift bins*), and waveforms which fall within each of these are summed together into a single waveform. The standard deviation of the Gaussian fit is used as a measure of the width of the waveform.

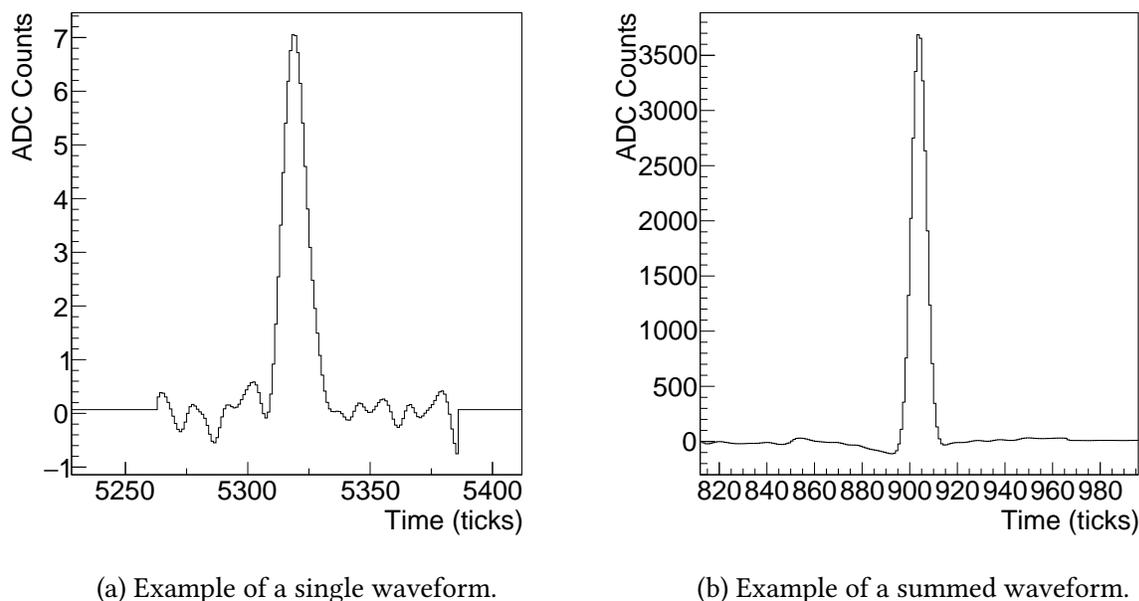


Figure 5.6: Example of (a) a single deconvolved ROI waveform and (b) the sum of many waveforms, extracted from MicroBooNE Run 1 cosmic data. Here, one tick = $0.5 \mu\text{s}$. Where the waveform changes from flat to noisy near the edges of the plot (most easily visible on the left-hand plot) defines the ROI which has been retained. The summed waveform on the right has a significantly improved signal-to-noise ratio. The dip preceding the peak in the right plot is known to be an effect of imperfect deconvolution.

Sections 5.2.1, 5.2.2, and 5.2.3 outline the selection of high quality waveforms from a set of tracks, and how these waveforms are averaged. Method validation has been performed in Section

5.3, and a measurement of the longitudinal electron diffusion coefficient from MicroBooNE Run 1 data is presented in Section 5.4. Finally, systematic studies are performed in Section 5.5

5.2.1 Selection of Tracks

To perform the measurement, off-beam cosmic muon tracks are leveraged. Given MicroBooNE's proximity to the surface, cosmic muons are plentiful, and serve as excellent calibration sources for detector physics measurements. The initial selection for the D_L analysis is simple. Tracks are selected based only on:

- Their length, which must be at least 50 cm.
- Whether they have a reconstructed interaction time (t_0) which has been geometrically tagged, as outlined in reference [101]. The reason this is necessary is that the only information which is associated with a given energy deposition is the amount of charge detected and the time at which it arrived at the readout wires. This means that an energy deposition which truly occurred at some time (t_0, x_0) could also be reconstructed as having happened at (t_1, x_1) or (t_2, x_2). This is displayed graphically in Figure 5.7.

In addition to these, a set of angular cuts will be developed in order to reduce the impact of biases to the measurement.

5.2.2 Selection of high quality waveforms

In MicroBooNE reconstruction, Gaussian functional forms, or “hits” are fit to deconvolved waveforms, and so by performing quality cuts on the hits, it can be ensured that only high-quality waveforms are accessed. The quality demands placed on the hits are as follows:

- **The hit multiplicity** must be exactly 1, ensuring that there is only a single hit in the selected ROI. The intention of this is to try and select out a clean sample of waveforms which are not contaminated by hits from nearby tracks.

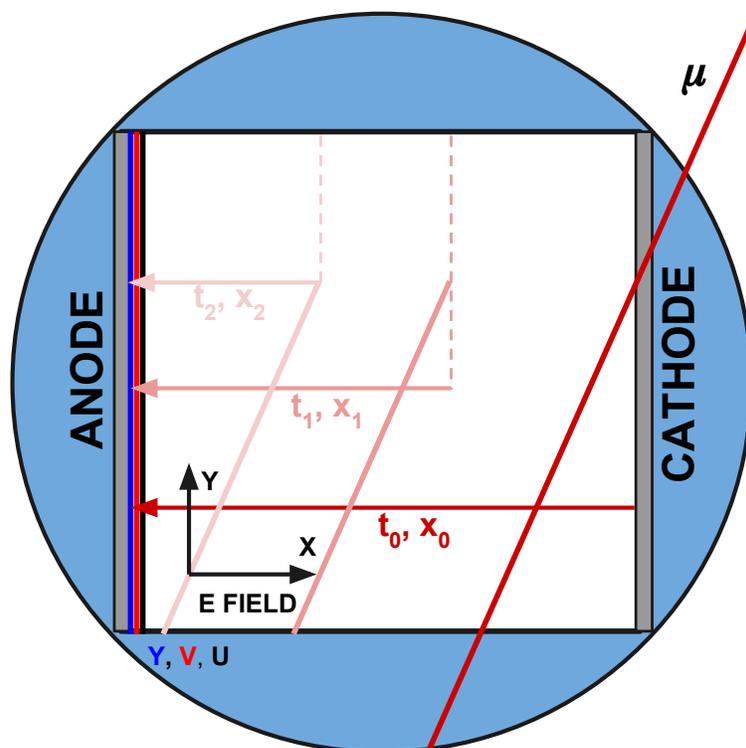


Figure 5.7: Diagram showing the necessity for t_0 -tagging. Without t_0 -tagging it is unknown what the true x -position of the track is.

- **The hit goodness of fit** is demanded to be less than 1.1 to ensure that the deconvolved waveform underlying the hit is reasonably Gaussian. The distribution of this variable for data and simulation is shown in Figure 5.8. This metric defines how well the Gaussian functional form fits the underlying hit. The cut at 1.1 was chosen by-eye to remove hits which do not well describe the underlying waveform.
- **The hit channel** is demanded to be on the collection plane, and an additional demand is made to ensure that the channel does not fall within the shorted region (see Chapter 3.2.2 for details).

The t_0 -corrected peak time of each hit which passes the quality cuts is logged, and a window of size n_i/n_b is opened around this value and a histogram of the waveform is created. Here, n_i is the number of ticks corresponding to one full drift window, which is constant at approximately 4600 ticks, and n_b is the chosen number of drift bins, which is set to be 25 by default. From here,

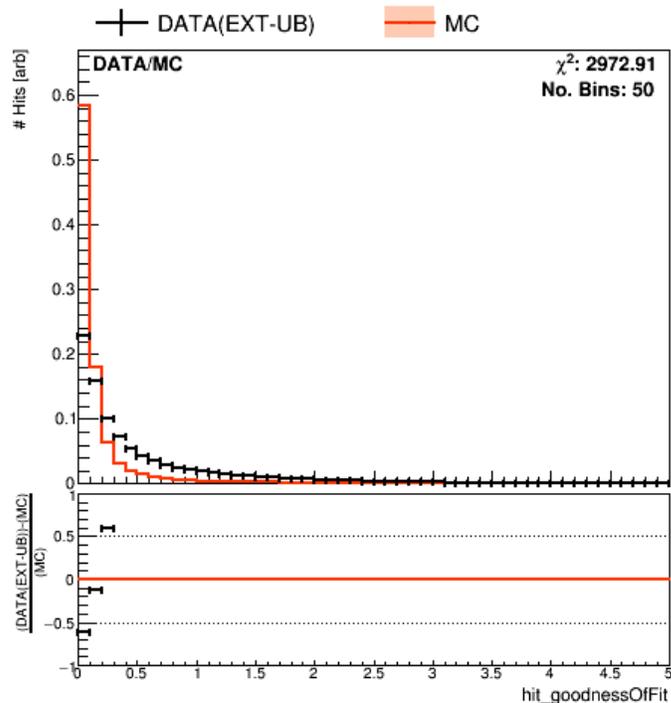


Figure 5.8: Hit goodness of fit plot, taken from a CI validation release. Here the goodness of fit is a metric which defines how well the Gaussian functional form fits the underlying hit. Note that these distributions are area normalised, and although there are differences between the data (black) and the Monte Carlo (red), it is reassuring that they both peak at low goodness of fit values, indicating that the pulses can be well described by a Gaussian curve.

the waveform is manipulated in two ways.

Firstly, the peak is moved to the center of the drift-bin into which it falls. In order to avoid complications with peak finding, a peak is only selected in the case that there is exactly one peak above 3 ADC counts in the chosen drift bin. This value is motivated by Figure 5.9, where it can be observed that all hits in both the data and simulation have a peak ADC count which is larger than 3.

A baseline correction is then applied. The largest peak is located and the signal region is defined to be the part of the waveform which falls within 20 ticks of this peak. The average value of the non-signal ROI is found, and the histogram is baseline-subtracted by this amount.

The final waveform quality cut removes waveforms which have outlying values of σ^2 . This is referred to as the *dynamic sigma cut*, and is shown in Figure 5.10. The region around the median

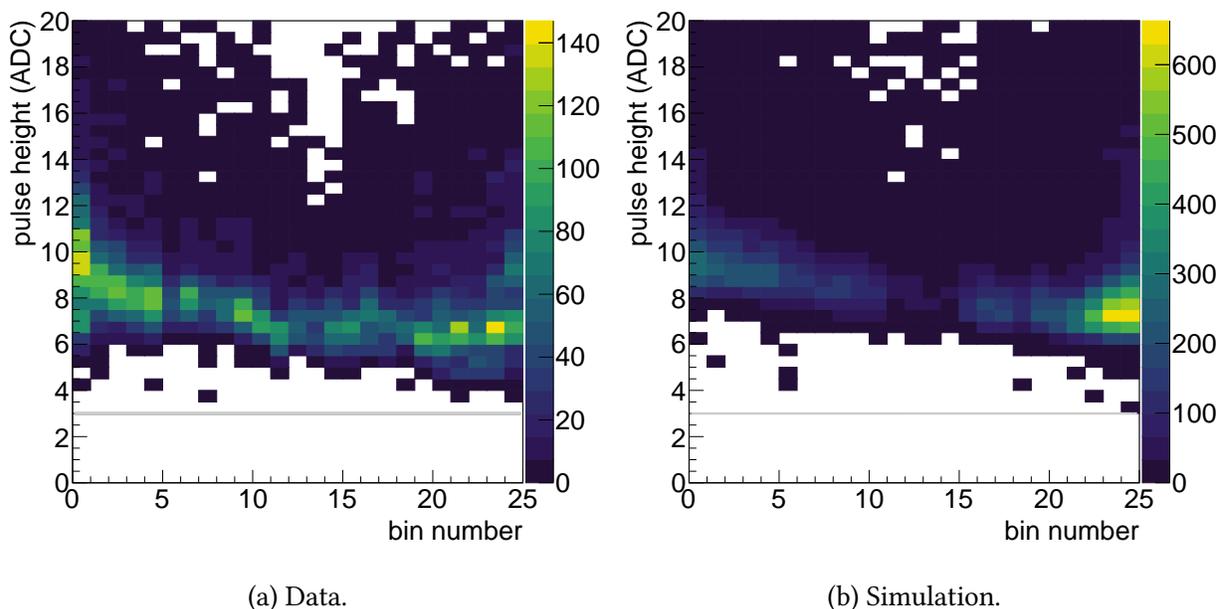


Figure 5.9: Pulse height versus drift bin number for a cosmic data sample (5.9a) and a CORSIKA Monte Carlo sample (5.9b). This contains only collection plane hits. Each drift bin spans approximately 10 cm. The grey line here is indicative of the pulse height threshold used to identify signals. On the collection plane it is expected that a signal-to-noise ratio of around 40:1 can be attained, meaning that the chosen cut value should be significantly above expected pulse heights from noise.

is defined by the standard deviation of the σ^2 values in the chosen drift bin.

5.2.3 Waveform Averaging Technique

The technique developed for this analysis takes the averaged waveform in each drift bin. The goal of this is to minimise the effect of noise and other features in the waveforms. The result of doing this on MicroBooNE data can be found in Figure 5.6. The plot on the left here is an example of an individual deconvolved waveform, complete with noise in the ROI. After applying the waveform averaging technique, the noise is greatly reduced, as is shown in the plot on the right. Here, the slight baseline undershoot on the left-hand side of the averaged waveform is known to be an effect of imperfect deconvolution.

The merging of the histograms which contribute to the average waveform must be done care-

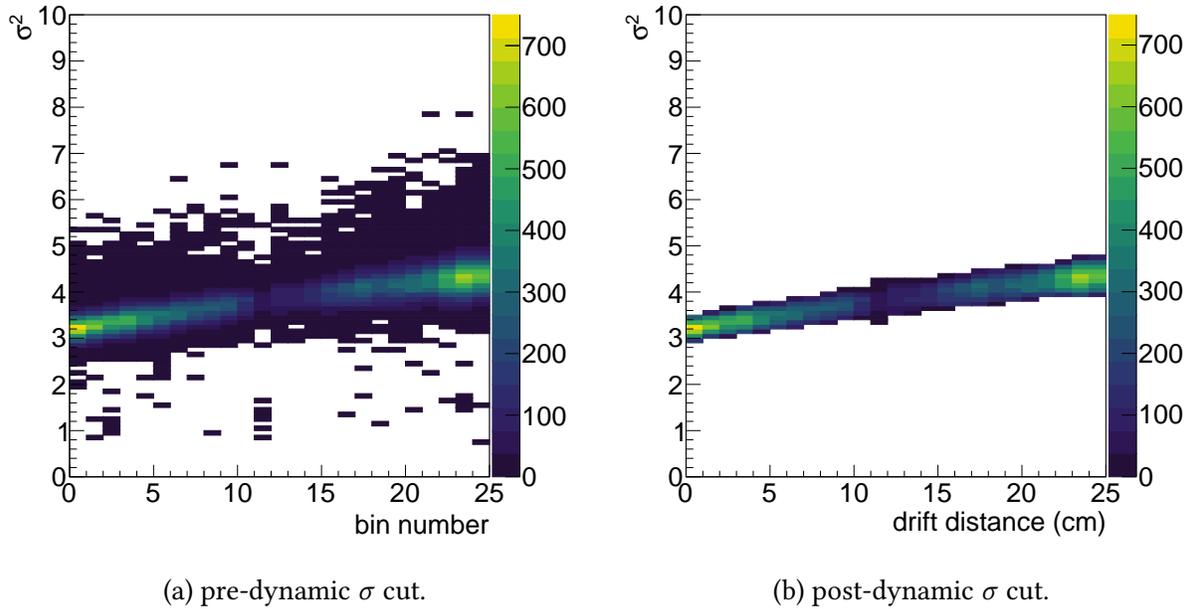


Figure 5.10: σ_t^2 versus bin number pre- (a) and post- (b) dynamic sigma cut for a MicroBooNE cosmic simulation with input $D_T = 0$ and $D_L = 6.36 \text{ cm}^2/\text{s}$.

fully such that no extra widening from misalignment of waveforms can contribute to widening of the pulses. Beginning with the peak value of the averaged waveform and the input single waveform separated by five ticks, the two waveforms are summed and an RMS^2 value is computed above a threshold which is 10% of the pulse height. By iteratively moving the single waveform and the averaged waveform closer together by a single bin and calculating the RMS^2 value for the summed waveform at each step, the combination which produces the lowest RMS^2 can be chosen, as shown graphically in Figure 5.11.

5.2.4 Extracting the Diffusion Value

Performing waveform averaging in each drift bin results in an association between the mean drift distance of the individual waveform peaks and the pulse width of the averaged waveform in that bin. As shown by Equation 5.2, the longitudinal diffusion constant, D_L is, to first order, proportional to the gradient of the plot that has been constructed, and so the diffusion coefficient can be calculated. For simulated data, the true drift velocity, $0.111436 \text{ cm}/\mu\text{s}$, is taken for v_d , while

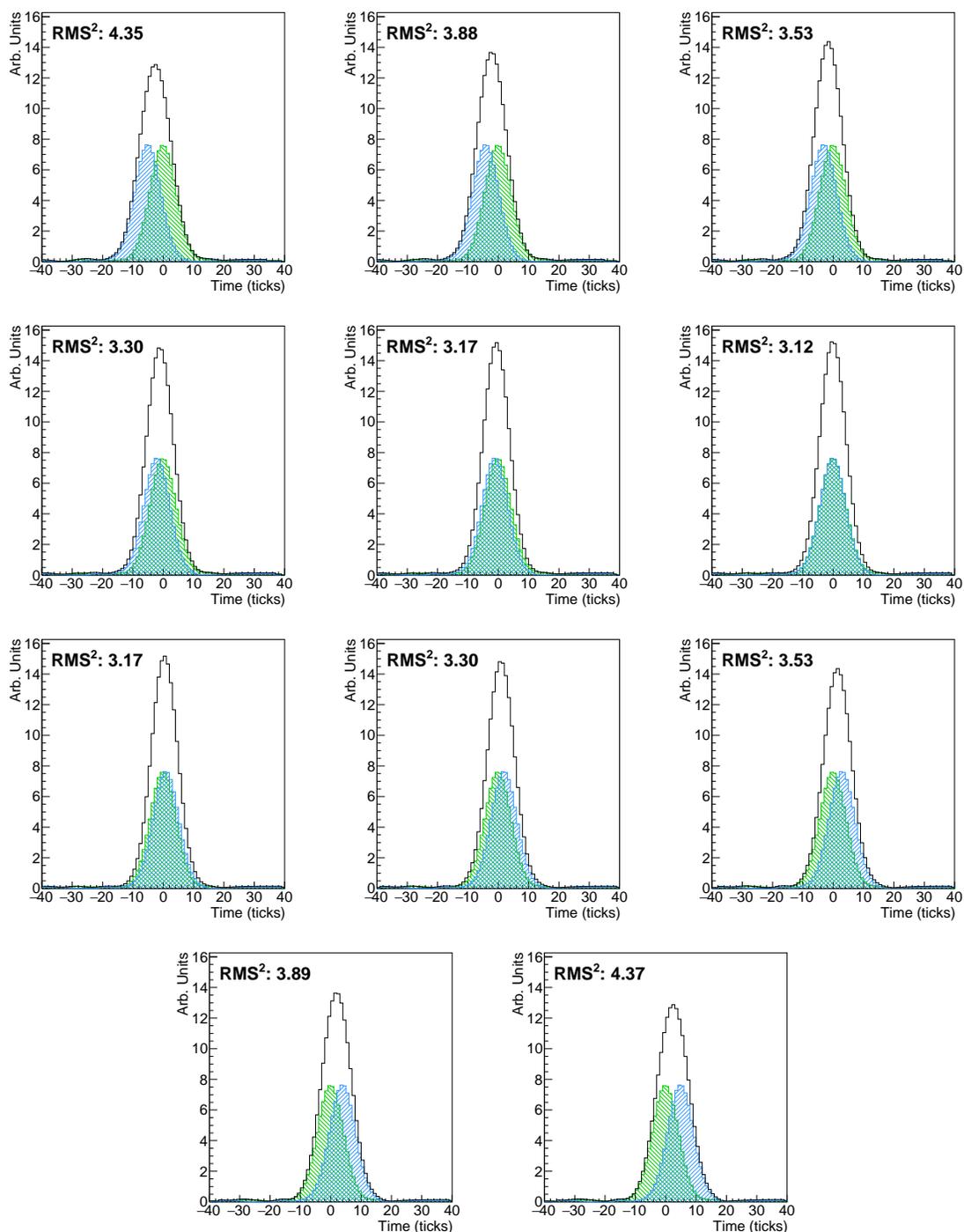


Figure 5.11: Illustration of the method used to merge waveforms. As described in the text, two initial waveforms (shown in blue and green) are first separated with the peak of the first waveform placed five ticks before the peak of the second waveform. The first waveform is iteratively moved forward by one tick until the peak is five ticks after the peak of the second waveform, and at each iteration the sum of the waveforms (shown in black) and RMS^2 are calculated. The combination which results in the minimum RMS^2 is selected (in this case, 3.12, shown on the right in the second row), and merged with the next waveform. This example uses Monte Carlo waveforms from Dataset 1, as defined in Table 5.1.

for data, the drift velocity is taken to be $0.1098 \text{ cm}/\mu\text{s}$, which is derived from measurements of the electric field in the MicroBooNE data.

5.2.5 Estimation of Error

To estimate the statistical error for this measurement, the prescription outlined in reference [95] is followed, whereby the uncertainty in each bin of the summed deconvolved waveform is iteratively increased until the Gaussian fit has a $\chi^2/NDF = 1$. These inflated uncertainties are then propagated to the width parameter which is extracted from the fit.

The drift distance of each summed waveform is taken to be the mean value of the peak positions of the individual waveforms in the chosen drift bin, and the error on this value is taken to be the standard deviation of the peak positions in the selected drift bin.

The linear fit then takes these errors into account. In practice, it is the uncertainty in the x -position in each bin which drives the uncertainty on the fit.

5.3 Method Validation

Having now defined a method for measuring the diffusion value, a staged approach to the analysis is performed. Initially the analysis is performed on a reduced sample, in which many detector and Geant4-level physics effects are turned off. This dataset is then built upon, introducing effects in order to estimate the uncertainty due to each. Once this has been done, the method can be validated on the full MicroBooNE cosmic simulation. A description of the datasets used in this analysis can be found in Table 5.1.

5.3.1 Using a Reduced Simulation Sample

Because a measurement of longitudinal electron diffusion might be expected to be sensitive to a large number of low-level effects—the electronics and field response, electron drift-lifetime, etc.—a staged approach to the analysis is taken. As a first step, a reduced sample is defined which has a

Dataset	Description	Parameter Values
Dataset 1 (multiple samples)	Single Muon	$\theta_{xz} = 0^\circ$ $\theta_{yz} = 0^\circ$ $\Delta\theta_{xz} = 0^\circ$ $\Delta\theta_{yz} = 0^\circ$ Momentum = 1 GeV Start position (x) = 128 ± 128 cm Start position (y) = 0 ± 116.5 cm Start position (z) = 518.5 ± 518.5 cm $D_L = 0.0, 2.0, 4.0, 6.0, 8.0$ cm ² /s $D_T = 0.0$ cm ² /s Electron lifetime = 30 ms No noise simulated No space charge simulated No multiple Coulomb scattering simulated No delta rays simulated No muon decay simulated
Dataset 2	Single Muon	Same as Dataset 1, but with Noise simulated $D_L = 6.36$ cm ² /s
Dataset 3	Single Muon	Same as Dataset 2, but with Multiple Coulomb scattering simulated Delta rays simulated Muon decay simulated
Dataset 4 (multiple samples)	Single Muon	Same as Dataset 2, but with $\theta_{xz} = 8.0^\circ$ $\theta_{yz} = 8.0^\circ$ $\Delta\theta_{xz} = 8.0^\circ$ $\Delta\theta_{yz} = 8.0^\circ$ Start position (z) = 500 ± 0.0 cm $D_T = 11.3, 16.3, 21.6$ cm ² /s
Dataset 5	Cosmic	All detector effects are turned on, with $D_L = 6.36$ cm ² /s $D_T = 0.0$ cm ² /s
Dataset 6	Cosmic	Nominal MicroBooNE simulation
Dataset 7	Cosmic	MicroBooNE Run 1 off-beam data

Table 5.1: Outline of the datasets used in the validation of the longitudinal electron diffusion analysis.

minimal configuration. This is described as Dataset 1 in Table 5.1. An example of a typical event using this configuration is shown in the event display in Figure 5.12.



Figure 5.12: Example collection plane (i.e., vertical wire) event display using deconvolved signal information showing the forward-going muon sample which has been simulated with the configuration outlined in Section 5.3.1. Here, the x -axis corresponds to the direction of the beam, where each pixel represents one wire on the collection plane; the y -axis corresponds to the drift direction, where each pixel represents 1 tick of the TPC clock.

Impact of Tunable Parameters

There are several tunable parameters in the diffusion analysis. Using the sample outlined earlier, the effect of each of these parameters can be understood. Of course, because this sample is so simplistic, each of these checks must be performed once again on a more realistic cosmic sample.

Number of Drift Bins The first, and most obvious, tunable variable is the number of chosen drift bins. Nominally, 25 bins are used, resulting in each drift bin being approximately 10 cm wide. Several other values have been tested, as shown in Figure 5.13. For each of the tested number of bins, the true value of D_L falls within the fit uncertainty, meaning that the analysis is not sensitive to variations in this parameter. As the number of drift bins is increased, the uncertainty on the measured diffusion value decreases. This is because the uncertainty is driven

by the uncertainty on the x -position of the average waveform. Note that it becomes difficult to move to larger numbers of drift bins, as it is necessary to leave enough of a sideband to calculate the waveform baseline, as discussed in section 5.2.2. To strike a balance between the error on the fit and the ability to find the waveform baseline, the nominal 25 drift bins are chosen. This will be reassessed when moving to a cosmic data sample in Section 5.4.

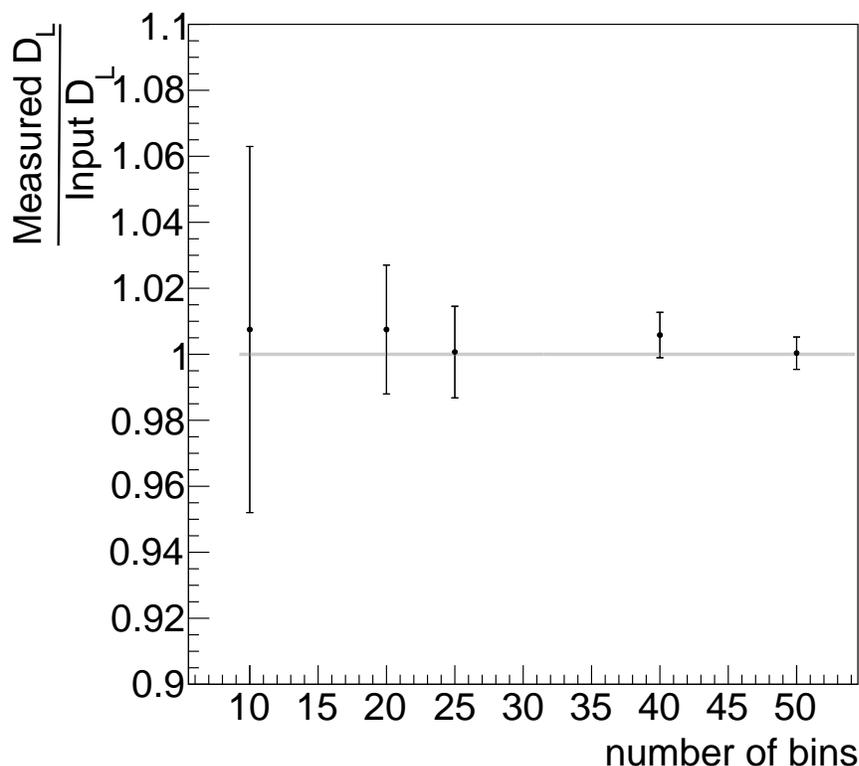


Figure 5.13: Fractional diffusion value as a function of the number of drift bins used. Each sample here used the same 250 tracks, meaning on average the number of tracks used in each drift bin is 25, 12.5, 10, 6.25, and 5, for the respective samples. The reduction in error is because the error is driven by the error on the x -position, which is related to the width of the drift bin.

Dynamic σ Cut The second tunable variable is the dynamic sigma cut value, which only accepts waveforms which have a σ^2 which falls in some region around the median σ^2 value in the relevant drift bin. As explained in an earlier section, the width of this region is defined by the standard deviation of the σ^2 values in the chosen drift bin, multiplied by some factor, which is referred to as the σ cut value. The results of modifying the multiplication factor of the σ cut value are shown in Figure 5.14. The plot shows that the analysis is insensitive to this cut, even for very large values of the multiplication factor. This, however, is likely due to the minimalistic nature of the configuration of the sample used here. For this reason, the multiplication factor is set to

one, however this will be revisited on a cosmic sample.

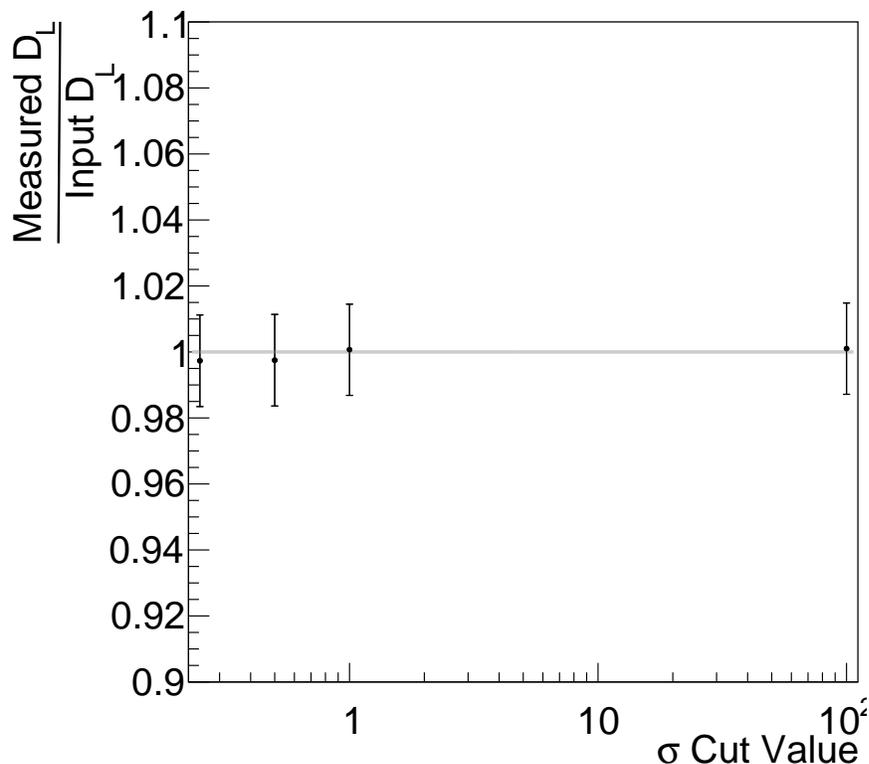


Figure 5.14: Fractional diffusion value as a function of the dynamic σ cut value.

Ensuring Non-Bias Across Many Input D_L Values

As a measure of how well the analysis method works, it is not enough to show that the correct value of D_L can be extracted from a single dataset. To bolster confidence in the method, the analysis is repeated on datasets with input D_L of 0.0, 2.0, 4.0, 6.0, 8.0, and 10.0 cm^2/s . The results of this are shown in Figure 5.15 and Table 5.2. Here the nominal selection and cut values are used, as defined in Table 5.3.

One point of interest, shown in Figure 5.15, is that the error on the fit does not envelope the true input D_L value for the $D_L = 0.0 \text{ cm}^2/\text{s}$ sample. The error is so small here because each σ^2 value is essentially the same for every x , meaning that the fit has very little freedom.

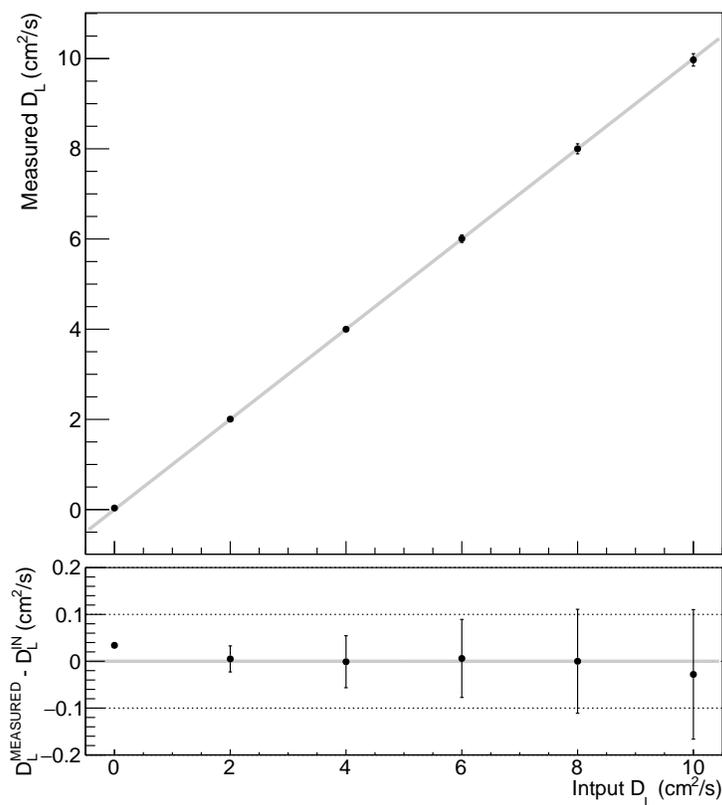


Figure 5.15: Measured D_L values for many different input D_L values, along with associated error.

Input D_L values (cm ² /s)	Measured D_L (cm ² /s)	Error	%difference
0.0	0.034	0.0005	N/A
2.0	2.005	0.0278	0.25
4.0	3.999	0.0555	-0.02
6.0	6.006	0.0833	0.10
8.0	8.000	0.1109	0.00
10.0	9.972	0.1383	-0.27

Table 5.2: Measured D_L values for many different input D_L values. The percent difference column is calculated by $\frac{D_L^{\text{measured}} - D_L^{\text{true}}}{D_L^{\text{true}}} \times 100$, meaning that the $D_L = 0.0$ cm²/s cannot be included.

5.3.2 Including Noise in the Simulation

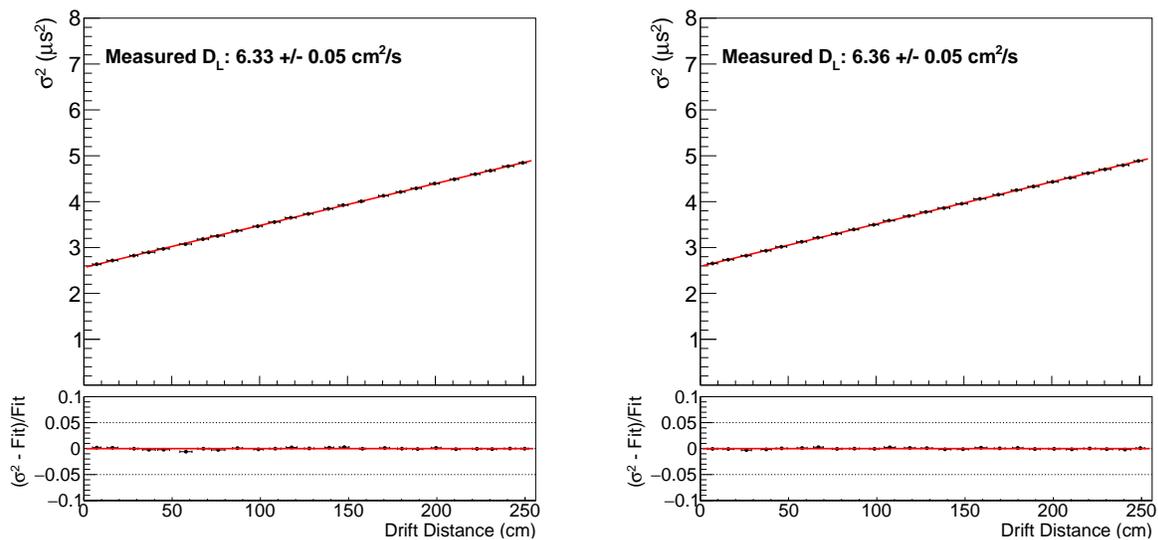
In this section Dataset 2, as defined in Table 5.1, is used. This is identical to Dataset 1, but with the inclusion of simulated electronics noise. The noise model implemented in the MicroBooNE simulation is data-driven in that it reproduces the frequency-domain shape and time-domain RMS of the noise. This is described in Chapter 4.

As can be noted from Figure 5.16a, there is a small difference between the best fit value for

Variable	Value	Bound type
Number drift bins	25	
Drift Velocity (simulation)	0.111436 cm/ μ s	
Drift Velocity (data)	0.1098 cm/ μ s	
X width	256 cm	
Dynamic sigma cut value	1 σ	Upper and Lower
Hit goodness of fit	1.1	Upper
Peak finding threshold	3.0	Lower

Table 5.3: Table summarising nominal cut values and other numbers of interest for this analysis.

the measured D_L coefficient and the true input D_L value; however the true value still falls within the fit error.



(a) Noise on.

(b) Noise and physics effects on.

Figure 5.16: Measured D_L with noise turned on (5.16a) and with noise and Geant4 level physics processes turned on (5.16b). These were produced with Dataset 2 and Dataset 3, respectively.

5.3.3 Including Physics Effects in the Simulation

This makes use of the sample defined as Dataset 3 in Table 5.1. This sample is identical to the one in the previous section, but with physics effects simulated. This includes turning on delta rays, muon decay, and multiple Coulomb scattering. Note that the tracks are still generated precisely

along the beam direction, and transverse diffusion is not simulated at this stage.

Figure 5.16b shows that there is essentially no difference between turning on physics effects and the earlier sample with only noise turned on. From this it can be stated that neither noise nor physics effects will be a significant uncertainty for this analysis.

5.3.4 Moving to a Simulated Cosmic Sample

Now that the method has been validated on a minimally configured reduced sample, investigations can be performed on a cosmic dataset. The intention is to use cosmic datasets to answer three questions:

- Are the number of drift bins and multiplication factor of the σ cut value which were chosen in Section 5.3.1 still appropriate?
- Is it possible to select a sample of t_0 -tagged tracks from a cosmic dataset which can return the simulated D_L value?
- What angular selection is needed to minimise the effects of D_T ?

Number of Drift Bins and σ Cut Value

The plots shown in Figures 5.13 and 5.14 are reproduced using Dataset 5 as defined Table 5.1; see Figures 5.17a and 5.17b. Unlike in the case of Dataset 1, there is a significant variation in the fractional diffusion value depending on the number of bins used. There is also a small dependence on the dynamic σ cut value. Because it has been shown that neither the noise or physics effects are expected to cause significant differences in the measured D_L value, it is assumed that this dependence is due to the angular distribution of the tracks in the cosmic sample.

Because the simulated value of D_L is able to be reasonably accurately measured with the nominal chosen values of 25 bins and a σ cut value of 1, no change is made to these parameters.

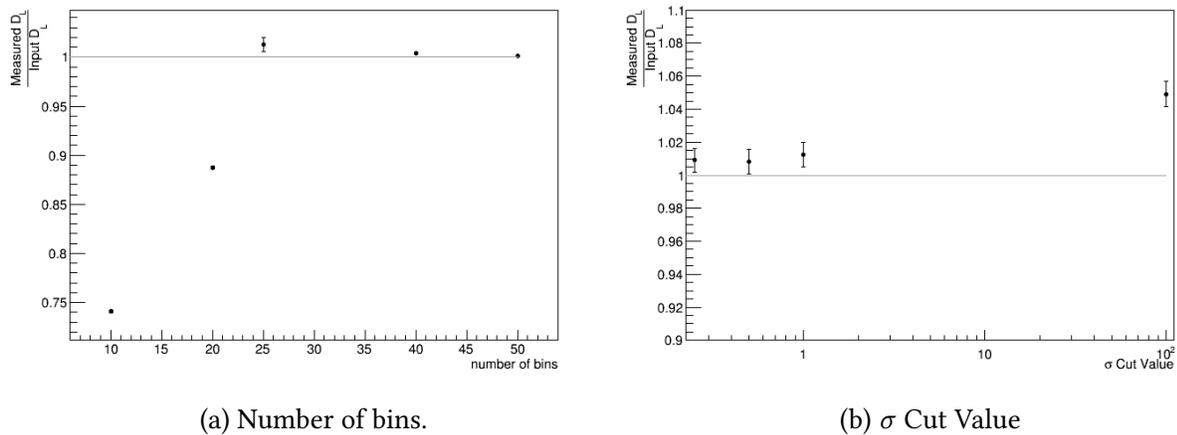


Figure 5.17: Fractional diffusion value as a function of the number of bins used (5.17a, using σ cut value = 1) and the σ cut value (5.17b, using a number of bins of 25) for a simulated cosmic sample. When compared with Figures 5.13 and 5.14 it is clear that the cosmic sample has more dependence on the choice of these variables, and this is assumed to be due to the angular distribution of the tracks.

Selecting t_0 -Tagged Tracks From a Cosmic Dataset

The selection of cosmic tracks is simple. As outlined in section 5.2.1, the only requirements placed upon the tracks are a track length cut and the requirement that a track must have a reconstructed t_0 . As alluded to earlier, an angular selection will now be developed in order to measure D_L using cosmic rays.

For this analysis it is prudent to define the angular cuts in terms of the angle to the anode plane, θ_{xz} , and it's orthogonal angle, θ_{yz} . When using these coordinates, $\theta_{xz} = \theta_{yz} = 90^\circ$ defines a track which is going straight downwards through the detector, $\theta_{xz} = \theta_{yz} = 0^\circ$ is a track going in exactly the beam direction, and $\theta_{xz} = \theta_{yz} = 180^\circ$ is a track going exactly opposite to the beam direction. A diagram showing these angles is provided in Figure 5.18.

The θ_{xz} angle, in particular, is an important angle when selecting angular cuts because of the effect of transverse diffusion, and because the response function can have significantly different effects at different angles during deconvolution.

Figure 5.19 shows the distribution of all tracks above 50 cm in length (left) and those tracks

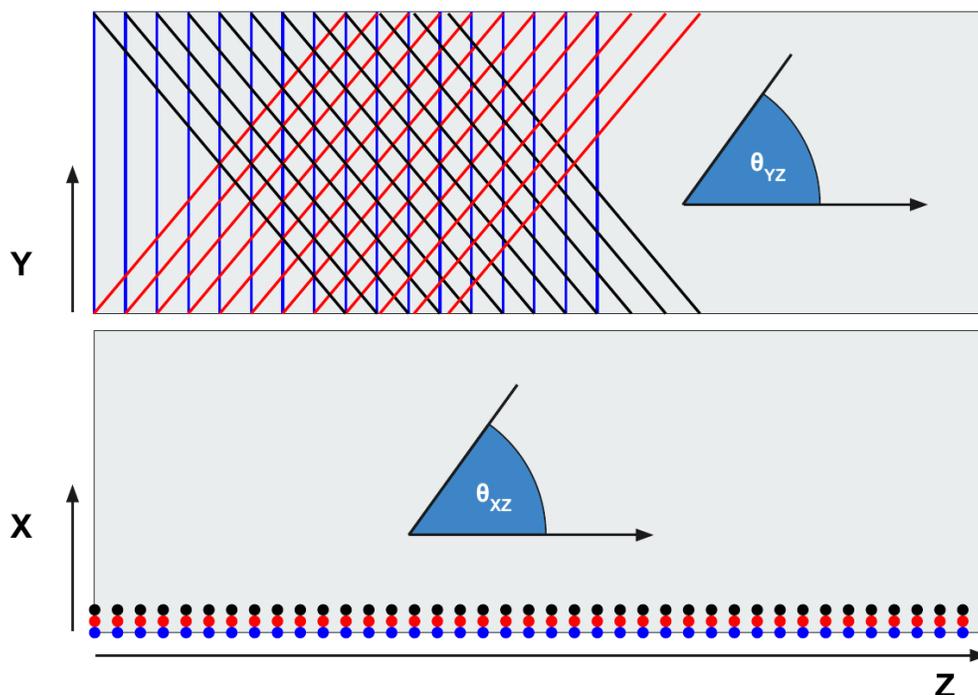


Figure 5.18: Plot showing the definition of θ_{xz} and θ_{yz} . Here, the blue lines and dots represent the collection plane, while the black and red represent the first and second induction planes.

which have been t_0 tagged (right). The demand that tracks are t_0 tagged reduces the statistics of the selected dataset; however, geometric t_0 tagging has been shown to be accurate to the order of several ticks, as shown in Figure 5.20, and in the MicroBooNE geometric t_0 -tagging public note [101]. Because of this high accuracy and the large number of cosmic tracks available in any given MicroBooNE window meaning that this analysis is able to accrue statistics reasonably quickly, this is not currently deemed to not be a significant source of concern. It should be noted that the t_0 -tagging public note states that the purity (the fraction of tagged tracks which are truly anode- or cathode-piercing) is approximately 97%.

Table 5.4 shows the passing rates for tracks as a function of the chosen θ_{xz}/θ_{yz} cut. As a demonstration, this shows a cut which is symmetric in these variables, but this does not necessarily need to be the case, and will be investigated in the next section. The total passing rates appear to be very low; however, this is the passing rate on a track-by-track basis. Knowing that there are approximately 10 cosmic tracks per event increases these passing rates by a factor of

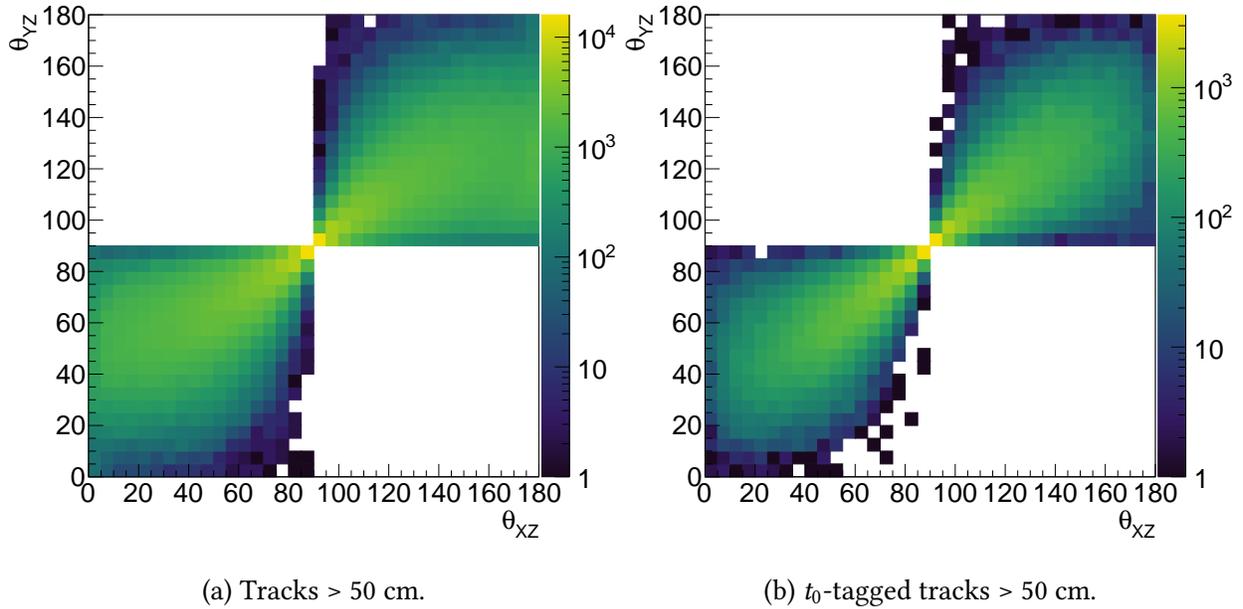


Figure 5.19: Angular distribution of reconstructed tracks with a length greater than 50 cm (5.19a), and the subset of those tracks which have a reconstructed t_0 (5.19b). The t_0 -tagging process reduces the size of the dataset significantly more in the region of interest for this analysis ($\theta_{xz} = \theta_{yz} = 0$ degrees and $\theta_{xz} = \theta_{yz} = 180$ degrees) than in other regions of the phase space. This is an expected effect because the closer a track becomes to forward- or backward- going, the less likely it is to pierce the anode or cathode faces.

10. In this table, the “Selected Tracks” and “% of Total” columns are measured with respect to the total number of tracks above 50 cm, while the “ t_0 -Tagged Tracks” and “ t_0 -Tagging Rate” columns are measured with respect to the selected tracks.

θ_{xz} & θ_{yz} Cut	Selected Tracks	% of Total	t_0 -Tagged Tracks	t_0 -Tagging Rate	Total Pass Rate
5°	219	0.03%	8	3.65%	0.001%
10°	700	0.13%	28	4.00%	0.005%
15°	1568	0.29%	127	8.10%	0.023%
20°	3687	0.67%	456	12.37%	0.083%

Table 5.4: This table shows the total passing rate for tracks in the Monte Carlo as a function of the θ_{xz} and θ_{yz} cut. This assumes that the chosen cut value is symmetric in θ_{xz} and θ_{yz} , but this does not necessarily have to be the case. Note that there is a requirement that every reconstructed track included in this table is at least 50 cm long.

Using a 10 degree θ_{xz}/θ_{yz} upper limit, plus a 50 cm length demand, D_L is measured using the

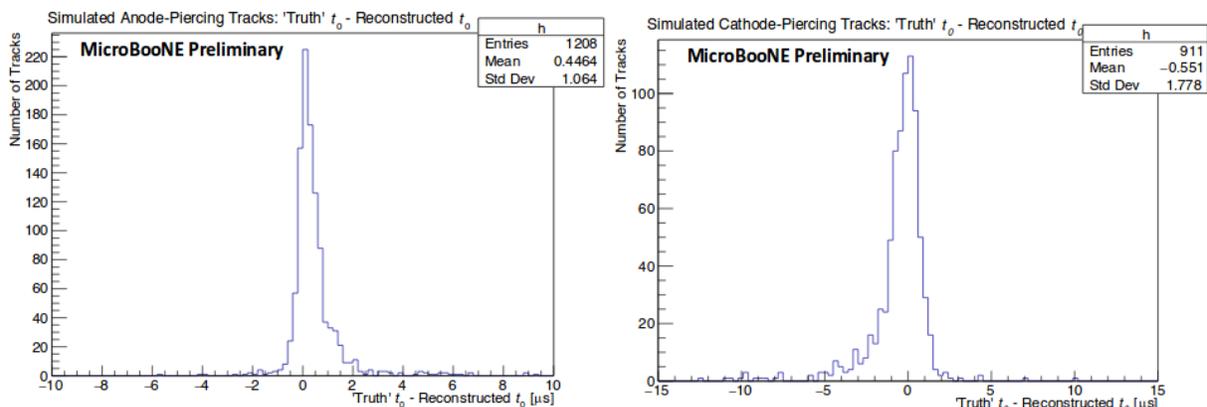


Figure 5.20: Plots taken from the MicroBooNE t_0 -tagging public note (reference [101]) showing the t_0 reconstruction accuracy for anode (left) and cathode (right) piercing tracks. As shown, the accuracy for tracks passing through the anode is higher than for those passing the cathode, but both have a standard deviation of less than 4 ticks. For the diffusion analysis, each drift bin (for 25 bins) corresponds to approximately 184 ticks, and so it is anticipated that this resolution is more than adequate.

dataset of 30k events from Dataset 5. The result is shown in Figure 5.21. There is approximately a 1-2% bias in the central drift bins. This has been investigated and is due to the angular acceptance of each drift bin. Higher angle tracks are significantly more likely to reach the center of the TPC, and so they are more likely to be dominant in those drift bins. Although it has been shown that track angles do not affect the measurement of D_L for a $D_T = 0$ sample up to $\sim 30^\circ$, it does modify the $\sigma_t^2(0)$ values, and so a sample which uses many different track angles, such as this, is liable to bias at the center of the TPC. For this reason, an additional demand is made that each drift bin contain more than 500 waveforms in order to be use in the linear fit. This values has been chosen by-eye. Because few low-angle tracks make it to the center of the TPC, this effectively removes the bias caused by these bins.

Dealing with the Effects of D_T for a Cosmic Dataset

Having now shown that the analysis is able to be performed by placing an angular cut on t_0 -tagged tracks, a study has been performed to understand the bias expected in the measured D_L from the effects of D_T .

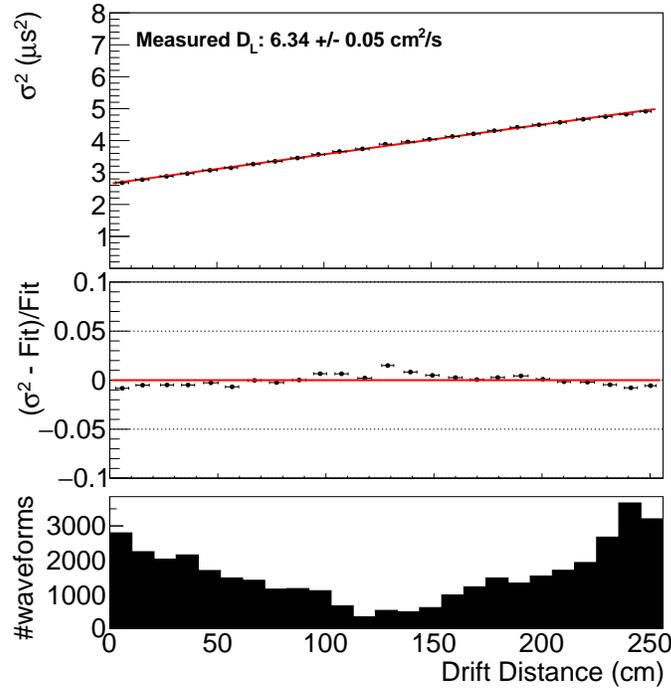


Figure 5.21: Measured D_L for a cosmic sample with input $D_L = 6.36 \text{ cm}^2/\text{s}$, and $D_T = 0.0 \text{ cm}^2/\text{s}$. Here, the number of waveforms in each drift bin is also included.

Transverse diffusion acts to modify the shape of the pulse in the yz -plane. For a track which is at $\theta_{xz} = \theta_{yz} = 0^\circ$, the effect of D_T is minimised because the electron cloud diffusion in the negative z -direction is compensated by the electron cloud diffusion in the positive direction. It is not possible to achieve this using geometrically t_0 -tagged cosmic rays because the sample is inherently angular, and so some effective additional smearing in the longitudinal direction is present. This is shown graphically in Figure 5.22. In this graphic, the green ovals represent single electron clouds. Without D_T , the longitudinal width of the electron cloud is measured to be the width of a single oval in x , however after the application of D_T , neighbouring ovals begin to overlap, and the measured longitudinal width is effectively wider.

In order to proceed with this measurement, a sample of tracks should be chosen which has an angular distribution which minimises the effect of D_T , while maximising statistics. To select the angular distribution, a study has been performed with Dataset 6, as outlined in Table 5.1. These samples use various combinations of D_T , θ_{xz} , and θ_{yz} to determine at which angles the effects of

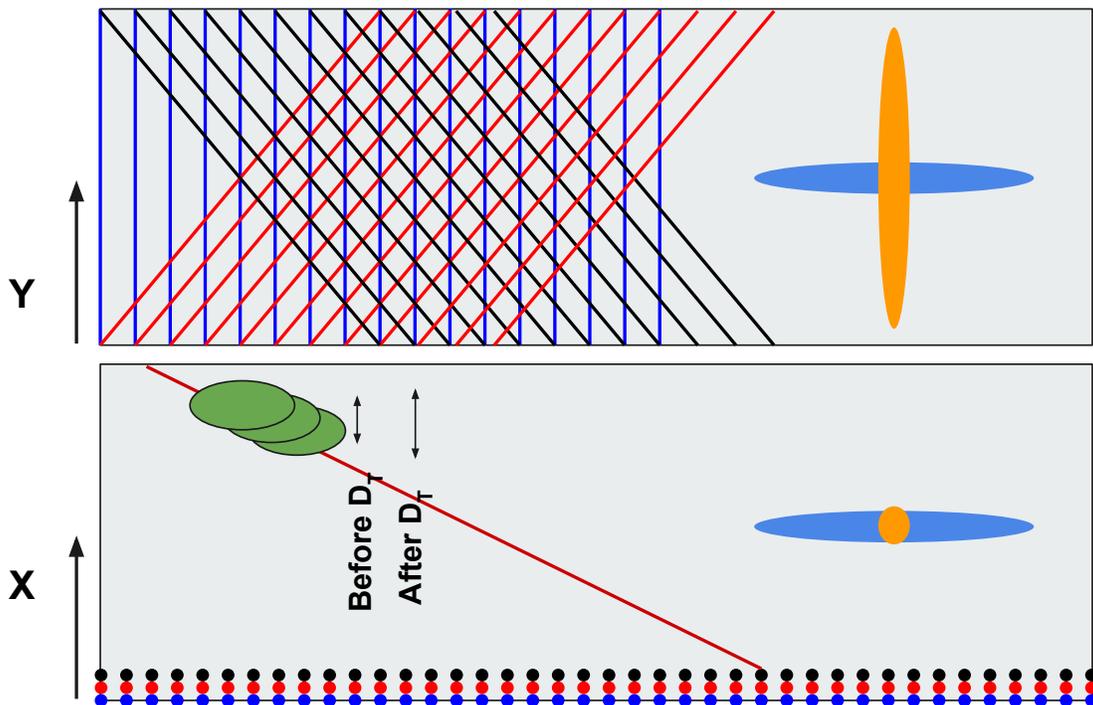


Figure 5.22: Illustration of the effects of transverse diffusion. Transverse diffusion can be thought of as a smearing in the z direction (blue) and the y direction (orange). Because only collection plane (vertical) wires are used, the spreading in Y has little effect because the pulses still arrive at the same time, on the same wires. However, the z component acts to spread the cloud onto neighbouring wires, meaning the x -width of the cloud appears to gain additional width.

D_T become pronounced enough to significantly skew the measurement. The diffusion analysis is performed for datasets using $D_T = 11.3, 16.3, 21.3 \text{ cm}^2/\text{s}$, where the variations were chosen because they are approximately 30% away from the nominal $16.3 \text{ cm}^2/\text{s}$, and this is expected to be a reasonably conservative uncertainty on D_T . For each dataset, different θ_{xz} and θ_{yz} selections have been made between 0 and 16 degrees. The results are presented in Figure 5.23.

For each of these figures, the plot on the left shows four bins in θ_{xz} and four bins in θ_{yz} , each representing the percentage difference between the measured diffusion value and the true D_L value for the different angular selections. The plot on the right of each of these figures represents

the average percentage bias as a function of θ_{xz} (green) and θ_{yz} (brown). It is clear from these plots that the dominant effect on the measured value of D_L is from the θ_{xz} angle, with the θ_{yz} angle having little-to-no effect. For this reason, a very limited θ_{xz} angle of 0 to 4 degrees is chosen, while the θ_{yz} angle is allowed to be between 0 to 16 degrees.

These plots also reveal that, according to the model implemented in Geant4, the bias increases linearly as a function of angle, and the gradient of the D_L bias as a function of θ_{xz} is increased as the value of D_T increases. This raises the interesting prospect of inferring the value of D_T from the measured value of D_L as a function of θ_{xz} . This is beyond the scope of this work, and would require thought before proceeding however; one of the interesting things about diffusion in electric fields is the level of disagreement between data and theory, indicating that there may be effects which are not currently being simulated.

Due to this study, it is found that approximately a +3% bias due to D_T is expected in the nominal simulation for an angular selection of $0^\circ \leq \theta_{xz} \leq 4^\circ$ and $0^\circ \leq \theta_{yz} \leq 16^\circ$.

Table 5.5 summarises the final cut values chosen to perform the analysis with.

Cut	Value
Length	50 cm
θ_{xz}	$0^\circ - 4^\circ$
θ_{yz}	$0^\circ - 16^\circ$

Table 5.5: Cut values used to select out tracks for the diffusion analysis.

5.3.5 Performing the Diffusion Measurement on a Full Monte Carlo Simulation

Taking the nominal values of θ_{xz} and θ_{yz} of 0-4 degrees and 0-16 degrees respectively, as well a demand that each track have a length of at least 50 cm, the analysis can now be performed on a full MicroBooNE simulation including the effects of D_T . In total, 93 tracks were selected from a sample of 200,000 cosmic events, and the result of performing analysis on these tracks is shown in Figure 5.24. The fit-only uncertainty on the measured diffusion value covers the expected value

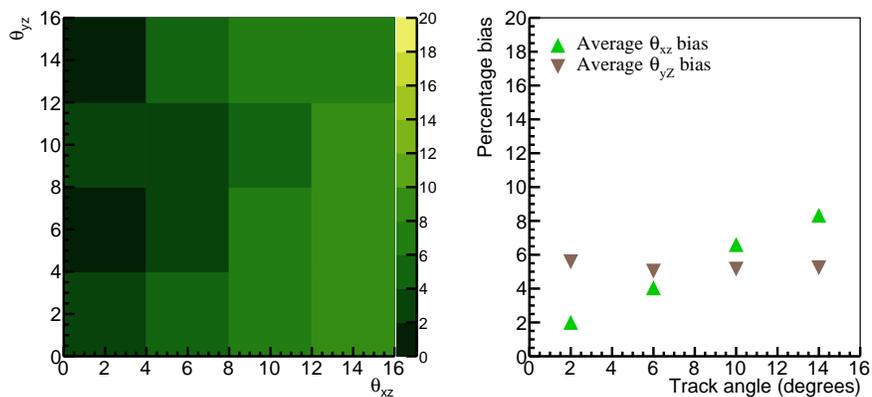
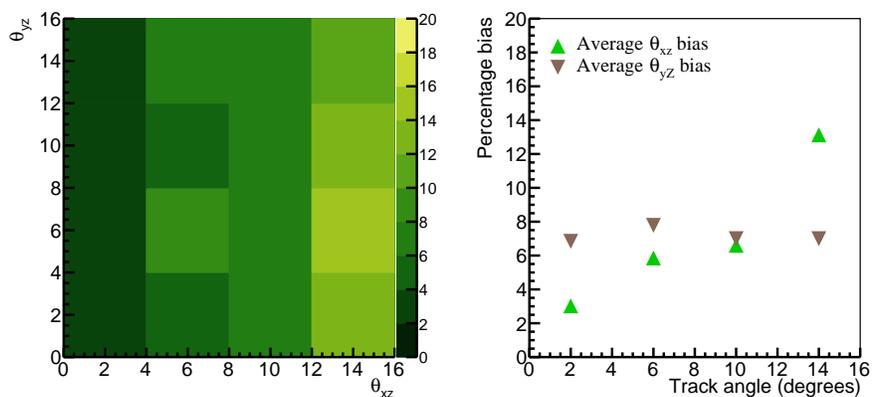
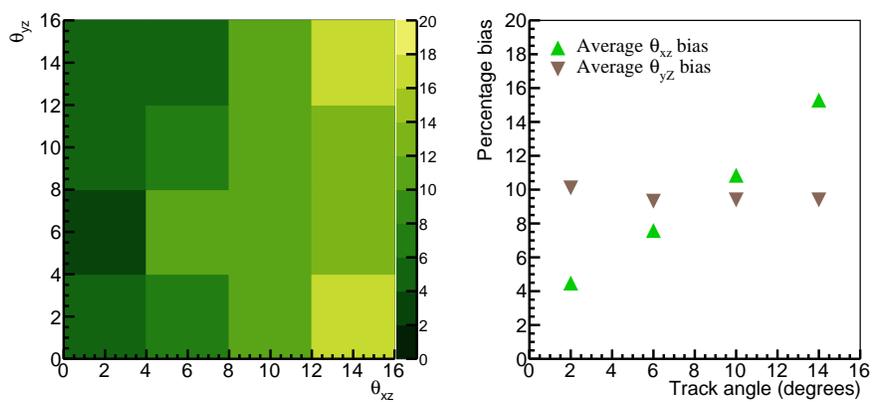
(a) Input $D_T = 11.3 \text{ cm}^2/\text{s}$ (b) Input $D_T = 16.3 \text{ cm}^2/\text{s}$ (c) Input $D_T = 21.3 \text{ cm}^2/\text{s}$

Figure 5.23: Percentage bias in measured D_L values in θ_{xz} - θ_{yz} space, and projected onto each axis. These plots are made using the sample defined in Dataset 6 with $D_T =$ (a) 11.3, (b) 16.3, the nominal value, and (c) 21.3 cm^2/s .

of $6.39 \text{ cm}^2/\text{s}$, where this is defined as the nominal input D_L value of $6.2 \text{ cm}^2/\text{s}$, with an additional expected +3% bias for our angular selection from the effects of D_T .

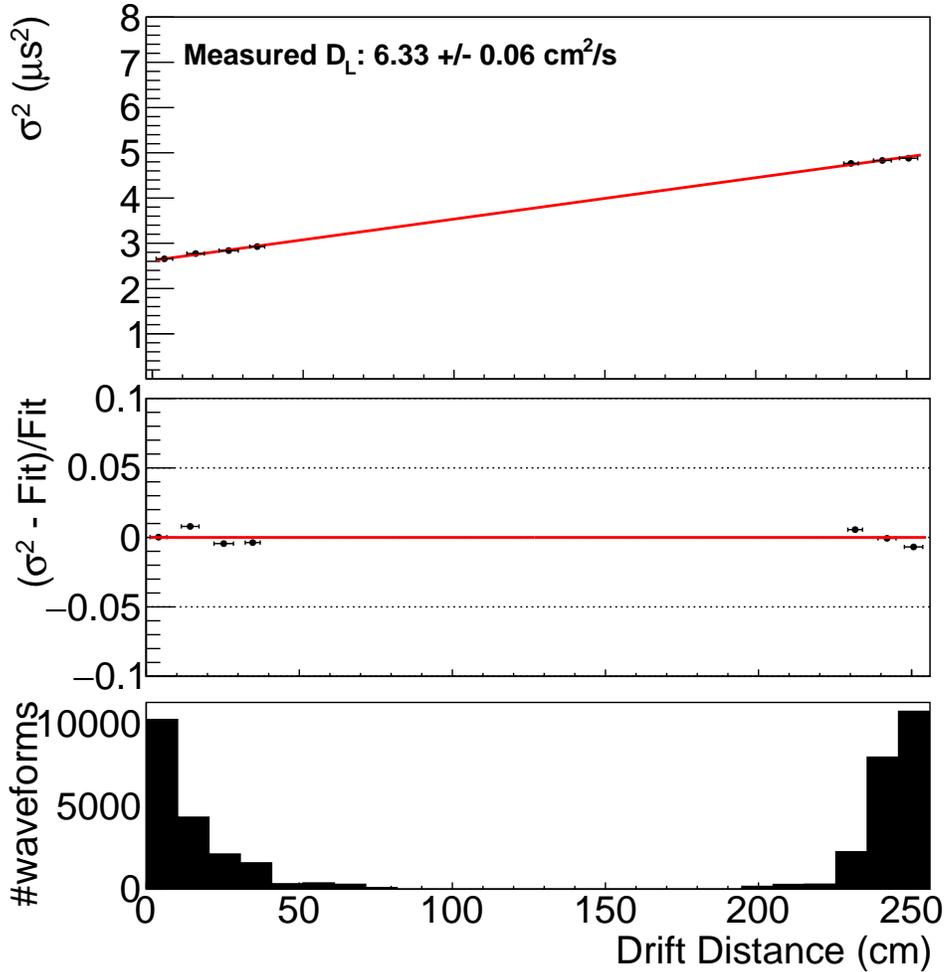


Figure 5.24: Extracted diffusion value for a full cosmic MicroBooNE simulation defined as Dataset 6 in Table 5.1, using the nominal input diffusion values of $D_L = 6.2 \text{ cm}^2/\text{s}$ and $D_T = 16.3 \text{ cm}^2/\text{s}$. The reason for the discrepancy is due to the effects of D_T , as outlined in the text. For this measurement, the angular selection outlined in this section ($\theta_{xz} = [0, 4]$ degrees and $\theta_{yz} = [0, 16]$ degrees) has been used. There is also a requirement that there be at least 500 waveforms in each drift bin, meaning that most of the central bins are excluded. The distribution of waveforms as a function of drift bin is a result of having a sample of tracks which are at very low angles. A track must have a θ_{xz} of approximately 13 degrees, beginning at $z = 0 \text{ cm}$, in order to traverse the whole width of the TPC. Such a track would not pass our angular selection. The maximum distance a track is expected to travel with our angular selection is approximately 90 cm.

5.4 Measuring the Effective Longitudinal Diffusion Coefficient in Data

Applying the same requirements as in Section 5.3.5, a measurement of the longitudinal diffusion coefficient can be measured from MicroBooNE Run 1 data. The central value of the measurement is presented in Figure 5.25.

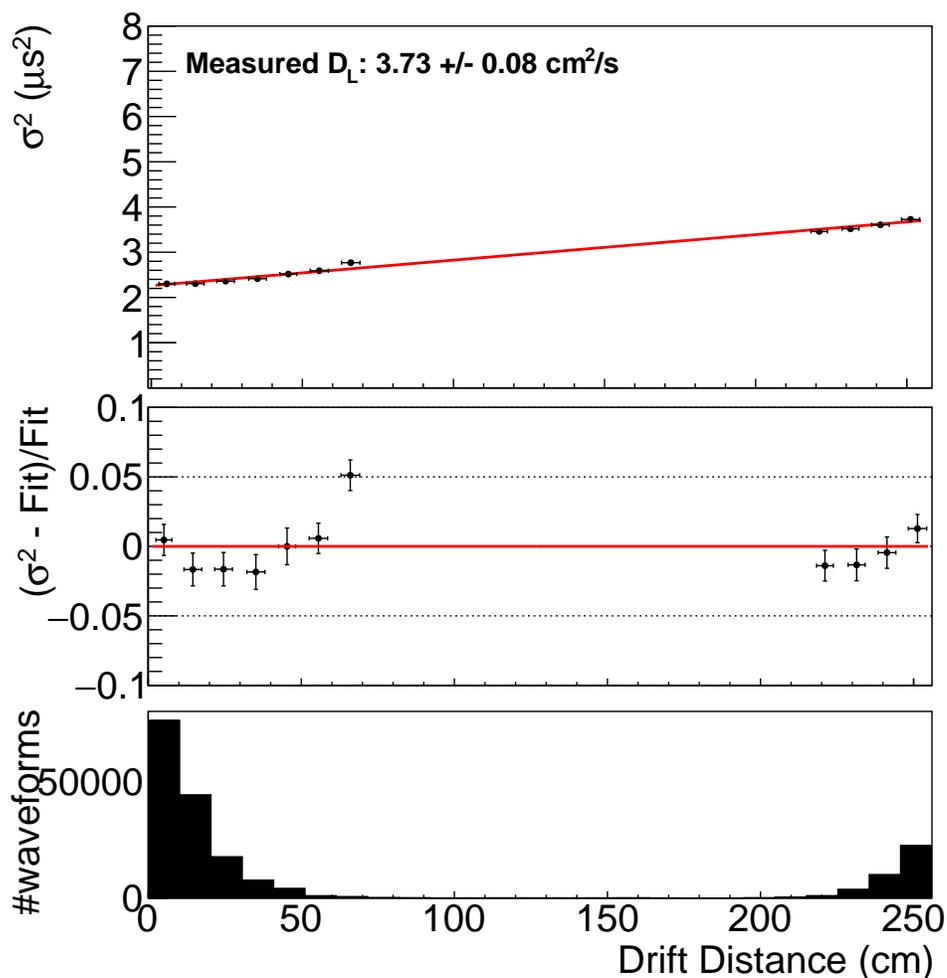


Figure 5.25: Extracted diffusion value from the MicroBooNE Run 1 data. In order for a point to be considered in the linear fit each bin must have a minimum of 500 waveforms.

The measured value is significantly lower than the parameterisation introduced in reference [95], and agrees more with the data presented by the ICARUS collaboration [100]. The uncertainty

presented in Figure 5.25 only contains the error on the fit, and so now it is necessary to understand the systematic uncertainties on this measurement.

5.5 Systematic Studies

The largest systematic uncertainties for this analysis are expected to be those related to uncertainty on the response function, drift velocity, space charge, and D_T . The following sections outline the systematic studies into these effects. A list of smaller systematic uncertainties to be addressed in the future can be found in Section 5.6.

5.5.1 Transverse Diffusion

The effect of transverse diffusion was discussed in Section 5.3.4. Due to the studies outlined there, the systematic uncertainty associated with transverse diffusion is estimated to be 5%.

5.5.2 Response Function

A study of the effect of the uncertainty on the response function has been carried out. From simulation studies, it is known that the data-driven response function has an approximately $1 \mu\text{s}$ widening over the true response function. This broadening has been corrected for, and a sample has been produced using the new, narrower response function in the simulation while using the original broad response function in the deconvolution.

The result is an approximately 15% systematic difference on the measured value of D_L . This represents the dominant systematic uncertainty on this measurement.

5.5.3 Drift Velocity

The drift velocity enters the diffusion equation as v_d^3 (see Equation 5.2), meaning that the D_L parameter has a very strong dependence on it. MicroBooNE does not currently have a full measure-

ment of this parameter with uncertainties, but a central value measurement has been determined from measurements of the electric field.

In simulation, the value of v_d used is that which is input into the simulation, 1.11436 km/s, while in the data, the value measured from the electric field is used, 1.098 km/s.

For the MicroBooNE data, a conservative 5% systematic uncertainty is placed on this parameter. This results in a systematic uncertainty of +9.1%, -8.3%

5.5.4 Space Charge Effect

The Space Charge Effect (SCE) is caused by a build up of argon ions in the TPC, leading to a distortion of the electric field. The effects of this are two fold: the first is that tracks can become bowed, meaning that this may lead to waveforms being placed in the incorrect drift bins, and the second is that the average electric field which an energy deposit travels through is a function of the X-Y-Z position in the detector.

Currently, space charge is simulated by moving the simulated true energy deposit to a new position using a position correction map, at which point the recombination is modified based on an E-field correction map. The energy deposition is then instantly transported to the readout planes without taking into account how changes in the electric field modify the drift velocity of the energy deposit. This presents a problem for estimating the systematic uncertainty, as the drift velocity is always assumed to be nominal.

Current work on the SCE places the variation in the electric field at a maximum of around 15%, but this is position dependent. This uncertainty can be used to calculate a D_L value using the parameterisation presented in reference [95] in order to estimate the scale of the uncertainty which might be expected. Table 5.6 indicates that even for reasonably large changes in the electric field, the effect on D_L should be reasonably limited. For this iteration of the analysis, a 3% systematic is applied due to variations in the electric field.

Electric Field Value (kV/cm)	D_L (cm ² /s)
0.273×0.85	6.24
0.273	6.36
0.273×1.15	6.45

Table 5.6: Expected maximum and minimum electric field values from space charge effect and the predicted D_L value.

5.5.5 Summary of Uncertainties

Table 5.7 outlines the systematic uncertainties which have been investigated. By far the largest systematic is the bias introduced from the response function used in the reconstruction being too wide.

The uncertainties listed for “Noise” and “MCS + delta rays” here are inferred from the studies undertaken with the reduced sample in Section 5.3, where turning on these effects resulted in a %-level change in the measured D_L value.

Systematic	Value
Response Function	15%
Drift Velocity	+9.1%, -8.3%
D_T	5%
Space Charge Effect	3%
Noise	~ 1%
MCS + delta rays	~ 1%
Fit error	~ 2.5%
Total	+18.7%, -18.3%

Table 5.7: An outline of the bounds on systematic uncertainties which have been placed so far. The total uncertainty assumes that the systematic uncertainties are uncorrelated.

There are a number of additional systematic uncertainties which must be investigated for this analysis, and these are outlined in Section 5.6. In general, these are expected to have a small impact on the result.

5.6 Summary and Future Work

Using the systematic uncertainties described in Table 5.7, the D_L measurement is quoted to be $3.73_{-0.68}^{+0.70}$ cm²/s.

There are a number of additional systematic uncertainties which must be addressed and these are outlined in this section. These are thought to be reasonably small when compared to the scale of the uncertainty on the width of the response function, and the drift velocity.

5.6.1 Detector Systematic Uncertainties to Address

The following are the detector-related systematic uncertainties that might be expected to have some impact on the diffusion analysis, although in general the magnitude of the listed effects is expected to be small.

- **Electron lifetime** The electron lifetime in MicroBooNE is extremely good and is simulated as being essentially infinite, and so this is expected to have minimal impact on the analysis.
- **Electron recombination** This takes place at the point of ionisation and so should not be a drift dependent effect.

5.6.2 Geant4-level Physics Uncertainties to Address

The following are the Geant4-level physics systematic uncertainties which might be expected to have some impact on the D_L analysis.

- **Delta ray production rate** Preliminary studies on this show that it is a negligible effect, but this requires further study.

5.6.3 Further Measurements

As discussed in Section 5.3.4, the requirement of t_0 -tagged tracks necessarily leads to a sample of largely high-angle tracks, and the angular selection used in this analysis drastically reduces the

number of selected tracks. The MicroBooNE CRTs – installed after Run 1 – could resolve this problem. The CRTs surround the detector, allowing t_0 tagging tracks which are not necessarily anode- or cathode-piercing. While the same angular selection would likely be used to reduce the effects of D_T , this would allow for much higher statistics, and therefore would mean that the measurement could be performed for lower electric fields for which less data exists. Indeed, existing measurements (as discussed in Section 1) were all performed at a variety of E-fields.

Chapter 6

Particle Identification

“Nobody expects the Spanish Inquisition”

Cardinal Ximénez

Particle Identification (PID) is a critical tool in the development of physics analyses, allowing for selection of exclusive state topologies. LArTPC detectors have a number of options available in order to perform PID.

The output of the MicroBooNE reconstruction is a collection of charged-particle tracks and electromagnetic showers, the distinction between which can act as a coarse PID: if the particle is shower-like then it is likely to be either an electron or a photon, if it is a track then it is likely to be either a muon, a pion, a kaon, or a proton.

Putting each reconstructed particle in the context of the topology of the full event can also shed light on its species: a photon takes time to pair produce and so is identifiable by a gap between the vertex and the start of a shower, whereas an electron shower starts at the vertex. Further to this, identification of a Michel electron can aid in identification of muons, while hard scatters can be used in the identification of hadrons.

In addition to topological information, MicroBooNE is able to make extensive use of the calorimetric information which is read out from the TPC.

6.1 Energy Deposition in Liquid Argon

The mean rate of energy loss for charged particles at energies relevant for MicroBooNE can be described by the *Bethe-Bloch* equation,

$$-\left\langle \frac{dE}{dx} \right\rangle = Zz^2 \frac{K}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 T_{max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right]. \quad (6.1)$$

Here, Z and A are the atomic number and atomic mass of the medium respectively, z is the charge of the incident particle, $m_e c^2 = 0.510$ MeV, T_{max} is the maximum kinetic energy which can be imparted to a free electron in a single collision, I is the mean excitation energy (in eV), and $\delta(\beta\gamma)$ is a correction factor which is important for higher momentum particles. As is standard, $\beta = v/c$ and $\gamma = 1/\sqrt{1-\beta^2}$. It should be noted that the mean dE/dx here is weighted by rare events which deposit large amounts of energy.

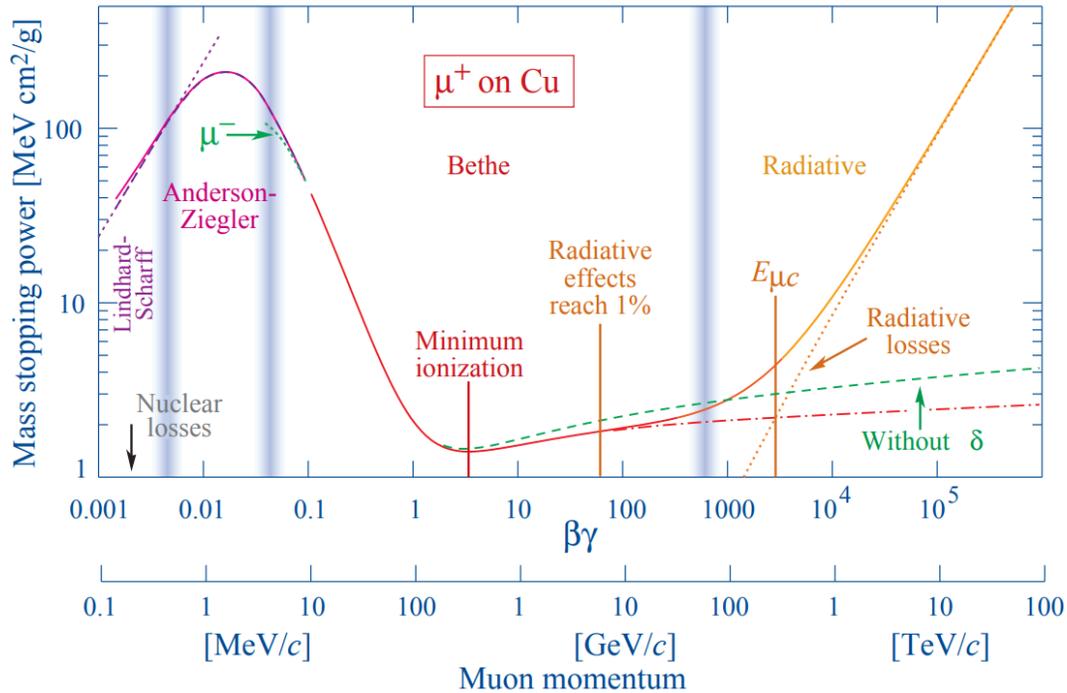


Figure 6.1: Mass stopping power ($=\langle dE/dx \rangle$) for μ^+ on copper, as a function of $\beta\gamma = p/Mc$. Figure taken from reference [102].

An example energy deposition curve for μ^+ on copper is shown in Figure 6.1. Muons in Mi-

croBooNE generally fall near to the point of minimum ionisation and thus have a mean energy deposition of 2.1 MeV/cm (characteristic of liquid argon), while more highly ionising particles such as protons deposit more energy per unit length. As particles lose kinetic energy in the detector, they begin to deposit more energy, resulting in a Bragg peak, which is a curve characteristic of the particle species. This can be phrased in terms of the distance between each hit and the end of the reconstructed track (the *residual range*). Curves of the $\langle dE/dx \rangle$ as a function of residual range are shown in Figure 6.2. The Bragg peak is often leveraged to perform particle identification of track-like particles.

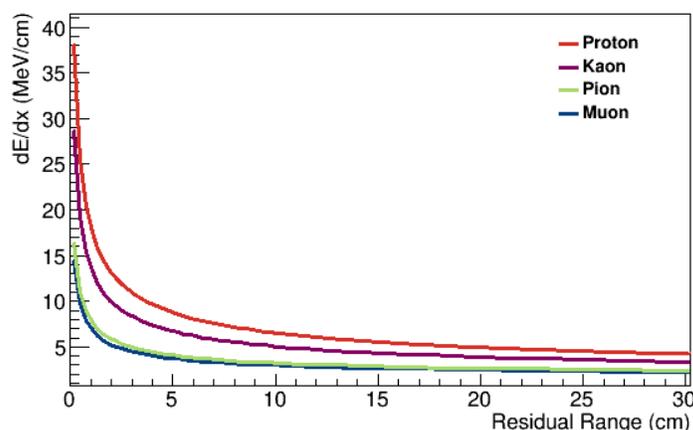


Figure 6.2: Theory curves in residual range- dE/dx space for particles which generally appear as tracks inside LArTPCs: protons, kaons, pions and muons. It should be noted that pions and muons (green and blue respectively) occupy essentially the same space, meaning muon-pion separation is extremely difficult.

Generally, shower-based calorimetric PID algorithms use the dE/dx in the trunk of the shower in order to separate out electrons from photons, as shown in Figure 6.3. Showers which are from electrons have a dE/dx of a minimally ionising particle (MIP), while those which are from photons have twice this energy deposition as they are composed of two electrons which have pair produced from a photon. Note that this is not always the case; photons are able to Compton scatter, producing a shower from a single electron which is not attached to the reconstructed neutrino vertex.

This chapter is dedicated to the description of a tuning which is applied to the calorimetry in

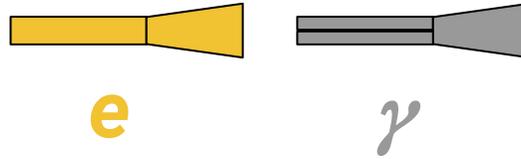


Figure 6.3: Electron versus photon particle identification is performed by making use of the fact that photons pair produce two electrons, meaning the dE/dx is approximately double that of a single electron.

order to make the data and the simulation approximate each other more closely. It also acts as a comprehensive evaluation of the PID algorithms used in the identification of track-like particles in MicroBooNE. Both the tuning outlined in this chapter, and the Bragg Likelihood algorithm for performing PID were developed as part of this work. The primary focus here is on the separation between MIP-like particles and protons such that the ν_μ CC0 π N p signal definition can be selected.

6.2 Simulated Data dE/dx Tuning

This section will describe a tuning which is performed prior to the particle identification outlined in the next sections.

By looking at the dE/dx per hit for hits in extremely MIP-like tracks (as in Figure 6.4 and Figure 6.5), we see that both data and simulation distributions peak around the same value, but that the width of the distribution in data is much larger than in simulation. This behaviour is expected after application of the dQ/dx and dE/dx calibrations as described in Section 4.7. In order to best compare PID performance between the two samples, the simulated data is first tuned by application of a simple smearing in order for it to better match the data. The tuning procedure and a validation of this approach are presented in this section.

A Landau distribution convoluted with a Gaussian distribution is used to represent the dE/dx of charged particles in MicroBooNE. A Landau distribution should be a reasonable approximation of the straggling function, while the Gaussian distribution is used as a catch-all model of detector effects. For both muons and protons, the widths of the Landau and Gaussian distributions are

then fit for. These fits are performed separately as detector effects can impact highly ionising particles in different ways to minimally ionising particles. The fitting procedure is as follows:

1. A distribution of dE/dx is produced for reconstructed hits coming from simulated true muons and protons with a distance from the hit to the end of the track greater than 100 cm and less than 150 cm. This range is used because the dE/dx in this region is slowly varying and can be treated as essentially constant. This distribution is then fit with a Landau-Gaussian functional form.
2. The dE/dx distribution is then produced without discriminating on particle species in order to verify that the distribution is dominated by muons (as protons tend to be very short in length), and is once again fit with a Landau-Gaussian functional form.
3. Step 2 is then repeated on cosmic data. Based on the previous step, the assumption is made that this is muon dominated. When fitting the data, the Landau component width is kept constant with respect to the simulated dataset under the assumption that any differences are due to un-modeled detector effects. This is thought to be a reasonable assumption, as the energy deposition of charged particles passing through a medium is better understood than the response of LArTPCs to such an effect.
4. The Gaussian widths for simulation and data can then be compared, and the residual width is assumed to be due to an uncorrelated Gaussian that represents detector effects that are not currently simulated, or not simulated well.
5. The width of this additional *smearing Gaussian* is estimated empirically using a toy study, as outlined in Appendix A.1.
6. Each dE/dx value along the track is multiplied by a random number drawn from the smearing Gaussian centered at 1.
7. This procedure works for muons, however there are very few protons with an extent of 100 cm or more in data (and there is also no way to reliably identify protons at this stage), and so

a different approach is taken for them. As with muons, the Landau component is estimated from the simulation. To estimate the Gaussian width in the data, the Gaussian width in the simulation is taken, and smeared using the smearing Gaussian which was derived from the muon sample.

This procedure is followed for each of the induction planes and the collection plane individually so that each plane has associated Landau and Gaussian widths for protons and muons. The assumption is made that pions and kaons have the same Landau and Gaussian widths as muons. While the kaon dE/dx profile is known to be quite different, the very low number of kaons expected in MicroBooNE means that this is not expected to be a significant issue. Pions have a very similar dE/dx profile as muons (see Figure 6.2), and so this is expected to be a valid assumption.

The datasets used to determine the widths are described in section 6.3.1: for simulation, a BNB+Cosmic sample is used and for data an on-beam sample is used. Both have been processed through the CC-inclusive selection. This is done so that the selected events have the correct t_0 , and so they are calibrated with the correct x-position.

The results of this procedure for simulation can be found in Figure 6.4 and Table 6.1. Similarly, results for data can be found in Figure 6.5 and Table 6.2. While the χ^2/NDF for many of these fits are unimpressive, it is thought that they describe the distributions well enough that the extracted parameters can be used without significant impact to the results of the algorithms.

A priori, it might be expected that the width of the distributions on the U and V planes would be the same. In the data it is observed that the distribution of dE/dx values is wider on the V plane than on the U plane. This is primarily due to two effects. Firstly, there are expected differences because the field response is different on the two planes due to the U plane not being shielded, meaning that it sees long-range induced signal. Because the front lobe of the response function is so shallow this effectively means that the response is quasi-unipolar and does not vary as much as a function of angle as the V plane. Because the dE/dx distributions are integrating over angle, an additional smearing on the V plane is expected. The second effect is that the V plane contains both U-shortened regions, where only 80% of the nominal charge is collected, and

Y-shorter regions, where the V plane takes on a collection-plane like response. These effects act to broaden the distribution.

As validation of the simulation tuning procedure, the tuned distributions may be fit with a Landau-Gaussian to ensure that the broadening results in an improved agreement between the data and simulation. The results of performing this fit are shown in Figure 6.6 and Table 6.3. These results show that the Gaussian width of the tuned simulation approximates the data to within 5%. In addition to this, a comparison of the data with the pre- and post-tuning simulation can be found in Figure 6.7.

Plane	Particle Species	σ_L	σ_G	χ^2/NDF
U	All Tracks	0.10	0.09	1088/86
U	Muons	0.10	0.09	982.2/86
U	Protons	0.23	0.32	49.2/35
V	All Tracks	0.13	0.19	332.7/86
V	Muons	0.12	0.19	982.2/86
V	Protons	0.19	0.32	40.68/37
Y	All Tracks	0.09	0.09	2628/86
Y	Muons	0.09	0.09	2522/86
Y	Protons	0.13	0.25	48.26/30

Table 6.1: Summary of measured Landau and Gaussian widths for MicroBooNE simulation across the three planes. Both the Landau and Gaussian widths are allowed to float for all fits for the simulation, the assumption being that there are detector effects which more strongly effect highly ionising particles such as protons.

Plane	Particle Species	σ_L	σ_G	χ^2/NDF
U	All Tracks	Fixed: 0.10	0.33	147.4/83
V	All Tracks	Fixed: 0.12	0.53	308.7/87
Y	All Tracks	Fixed: 0.09	0.20	551.7/87

Table 6.2: Summary of measured Landau and Gaussian widths for MicroBooNE data across the three planes. Here, it is assumed that the vast majority of tracks in the 100-150 cm range are muons. Here, ‘‘Fixed’’ alludes to the fact that the σ_L is not allowed to float during the fit.

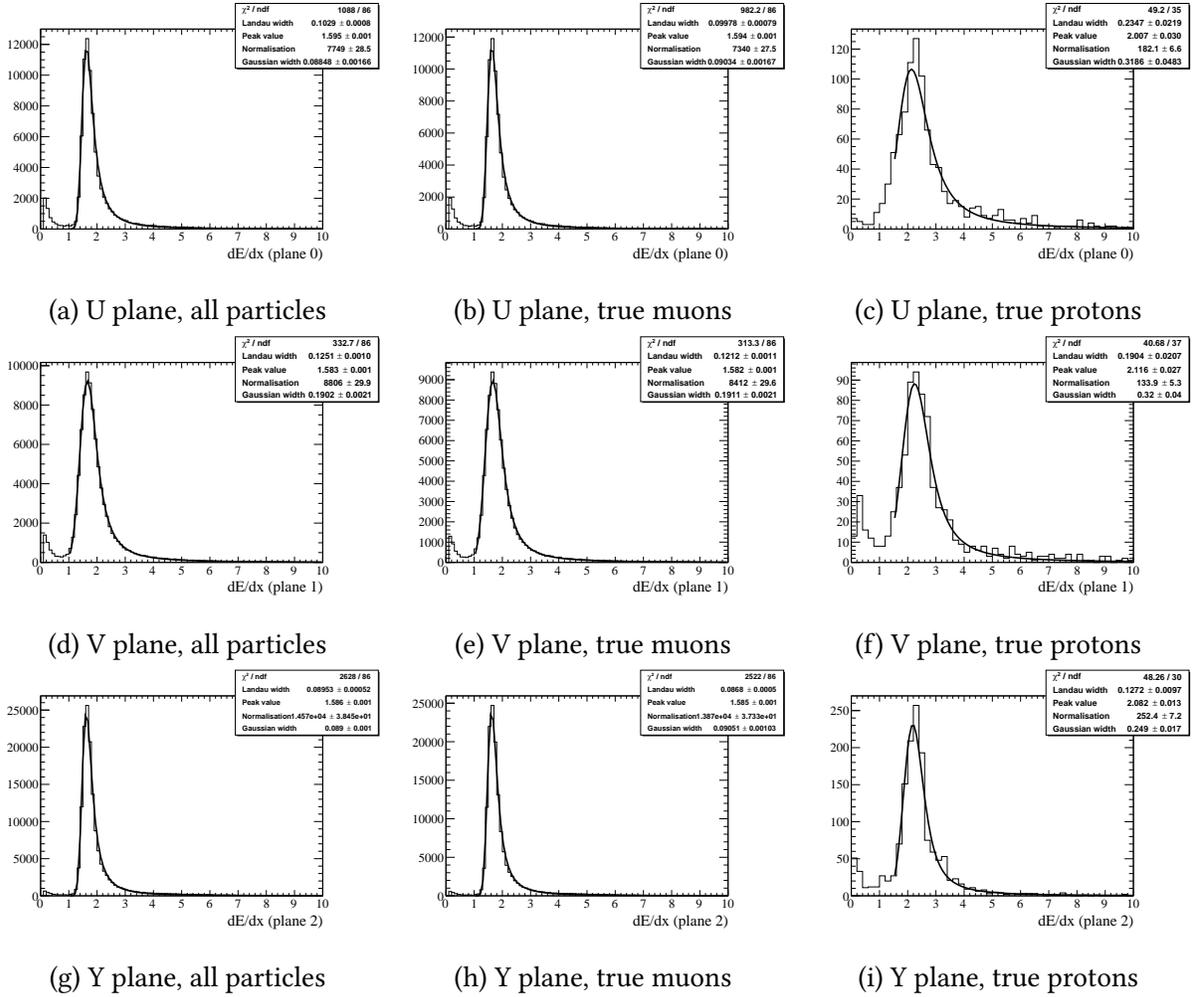


Figure 6.4: Fitted widths for all particles, true muons and true protons for each wire plane in MicroBooNE simulation. The distribution for all particles is very similar to the distribution for true muons due to small numbers of Utons in the sample used. The peaks at 0 dE/dx are due to a known deficiency in the reconstruction as tracks approach being parallel to the wires on a given plane.

Plane	Particle Species	σ_L	σ_G	χ^2/NDF	$(\sigma_D - \sigma_{MC})/\sigma_{MC}$
U	All Tracks	Fixed: 0.10	0.32	228.3/17	0.031
V	All Tracks	Fixed: 0.12	0.56	189.8/17	0.054
Y	All Tracks	Fixed: 0.09	0.19	83.64/17	0.053

Table 6.3: Summary of measured Landau and Gaussian widths for MicroBooNE tuned simulation across the three planes.

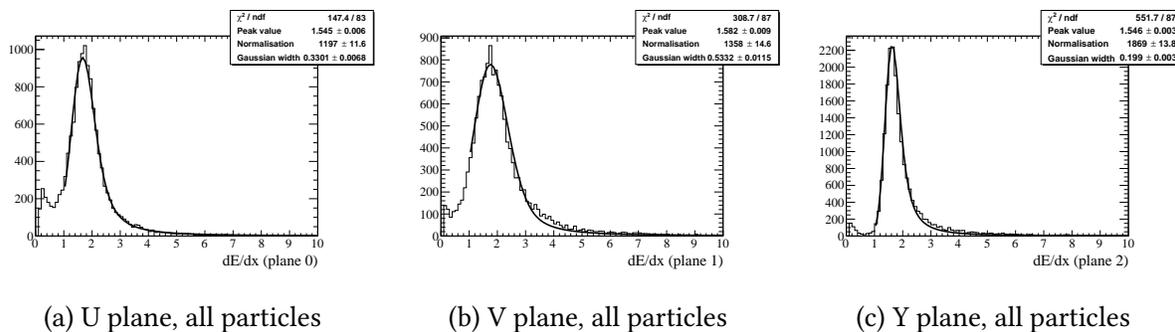


Figure 6.5: Fitted widths for all particles from MicroBooNE data for the two induction planes and the collection plane.

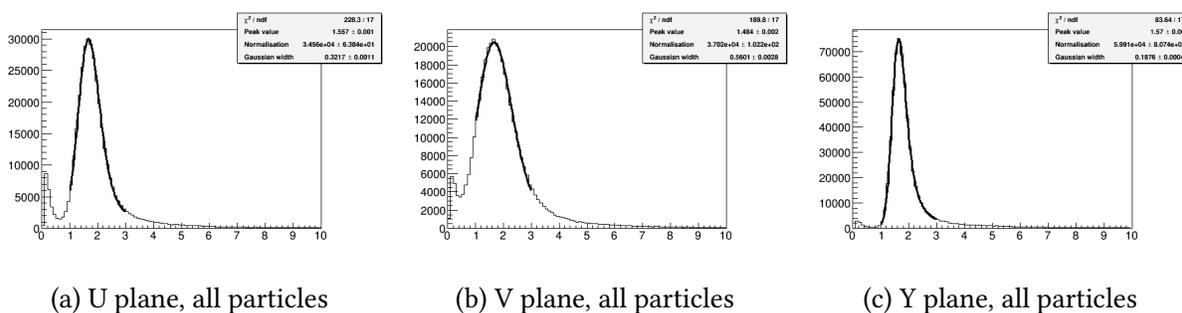


Figure 6.6: Fitted widths for all particles from MicroBooNE tuned simulated data for the two induction planes and the collection plane.

6.3 Algorithm Comparisons

Now that the simulation has been tuned to better match the data, comparisons can be made between the available algorithms in order to evaluate the strengths and weaknesses of each.

This section will compare the output of the algorithms currently being used on MicroBooNE with the Bragg Likelihood algorithm, which has been developed as part of this work. It should be noted here that the tuning which has been applied to the simulation as outlined in Chapter 6.2 is propagated through all of these algorithms such that they can be compared on the same footing.

Because the relative and absolute number of each particle species produced in the simulation is not necessarily correct, the normalisation of each particle species are allowed to float independently in a *template fit*. This allows evaluation of whether the simulation is able to describe the shape of the data for each variable in a way that is decoupled from the complicated physics of

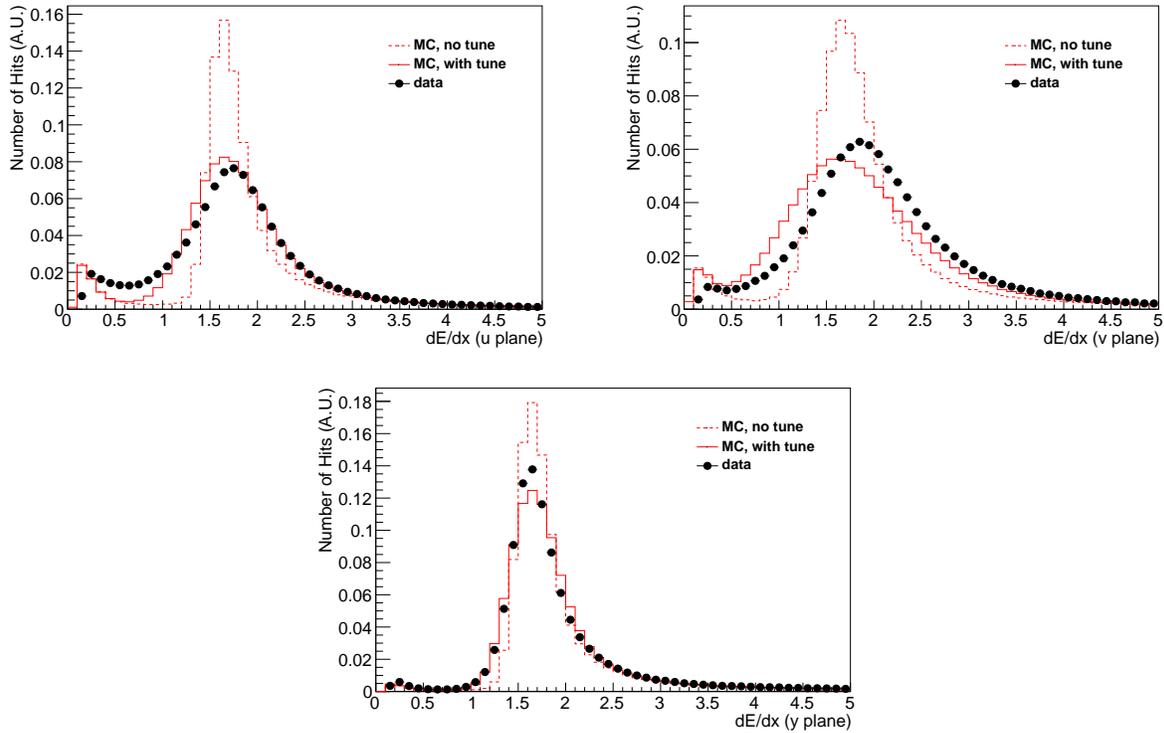


Figure 6.7: Agreement between data and simulation before and after application of the simulation tuning. The post-tuning simulation agrees significantly better with the data, although there is a slight over-smearing on the collection plane.

neutrino interactions. Results of performing this fit for each distribution, for each sample, can be found in this section. The original POT normalised distributions can be found in Appendix A.3.

6.3.1 Sample Information and Scaling Factors

For results that have been normalised by their POT, an overview of the samples and their scaling factors can be found in Table 6.4.

For those distributions which have been template fit, the normalisations of each component of the simulation is allowed to float within $\pm 50\%$ of their nominal POT scaled values. Note that the off-beam and on-beam data are not allowed to float in this fit. Also note that any purities which are calculated are done using the POT normalised distributions so that they are comparable between algorithms.

Within this chapter, the DIC and Birks systematic variation samples have been used as these are expected to have the largest impact on the performance of the PID. A description of these samples can be found in Chapter 7.2.

6.3.2 Deposited Energy Versus Energy By Range Algorithm

The first algorithm which has been investigated is the *depErangeE* algorithm, which compares the energy deposited on the collection plane in the TPC to the energy estimated based on the particle's range under the μ and p assumptions in order to try and attain separation. This is particularly interesting as, unlike the other algorithms investigated, it is not specifically looking at the shape and normalisation of the Bragg peak, and so can act as a somewhat independent handle on PID.

Data-to-simulation comparison distributions are shown for the *depErangeE* under the μ assumption in Figure 6.8. Although the agreement between data and simulation appears to be very good, the separation power of this algorithm is relatively low when compared to the other algorithms.

In addition to the data/simulation comparison plots, Figure 6.8 also contains the efficiency and purity of selected muons, pions and protons as a function of cut value.

Sample	Number of Triggers	POT	Scale Factor
On-beam data	31383149	1.41e+20	-
Off-beam data	33612500	-	0.93
BNB+Cosmic	-	1.96799e+20	0.72
BNB+Cosmic/with no smear	-	1.97421e+20	0.71
BNB+Cosmic/with DIC	-	1.96615e+20	0.72
BNB+Cosmic/with Birks	-	1.96220e+20	0.72

Table 6.4: Information about samples used throughout the development of the dE/dx tuning and algorithm comparisons.

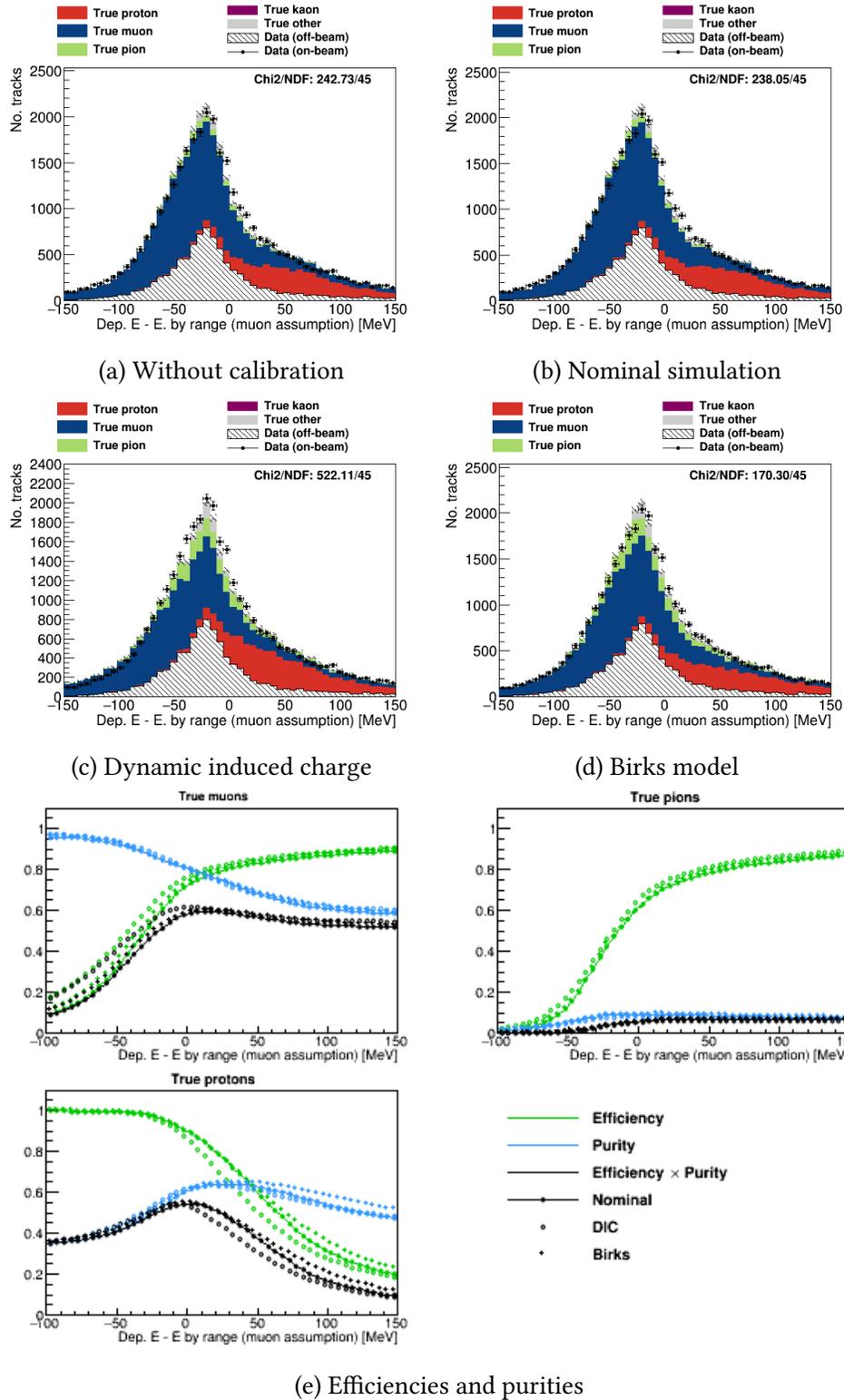


Figure 6.8: Deposited energy on plane 2 versus energy by range under the muon assumption, for four separate samples.

Here, the efficiency is defined to be

$$\epsilon = \frac{\text{Selected true particle species, } s}{\text{Total true particle species, } s}, \quad (6.2)$$

while the purity is defined to be

$$p = \frac{\text{Selected true particle species, } s}{\text{Total selected particles}} \quad (6.3)$$

The efficiency and purity curves are shown for the nominal MicroBooNE simulation and are compared against two detector variations outlined in Chapter 6.3.1 in order to provide information on the robustness of the algorithms.

6.3.3 PIDa Algorithm

The PIDa algorithm has been successfully used by the ArgoNeuT collaboration [92] in the past, however there is a significant level of disagreement between data and simulation when this algorithm is applied at MicroBooNE.

This algorithm parameterises the dE/dx curves as an exponential,

$$dE/dx = aR^b, \quad (6.4)$$

where R is the residual range of the hit and a and b are parameters which are to be fit for. It is found to be a reasonable approximation to use $b = -0.42$, and rely only on the value of a to perform the PID. In practice, a is calculated on a hit-by-hit basis, and the average is taken across the track,

$$PIDa = \frac{\sum \frac{dE}{dx} \cdot R^{0.42}}{N_{hits}}. \quad (6.5)$$

This algorithm drops both the first and last hit in the hit collection as it is known that the first hit can be biased due to being close to the vertex, where there may be other activity, and the last hit can be biased if a particle stops part way between two wires, causing a low dE/dx hit to be reconstructed. Several additional methods have been investigated in the course of this review, for instance taking the median of the PIDa values along the track, or extracting the most probable value using a kernel density estimator, however no significant improvement has been found, and so only the original method using the mean is presented in Figure 6.9.

The efficiency and purity of the PID using PIDa are reasonably good, however there is severe disagreement between the data and simulation in the proton region and it is clear that this variable is extremely sensitive to both the DIC sample and the alternative recombination model sample.

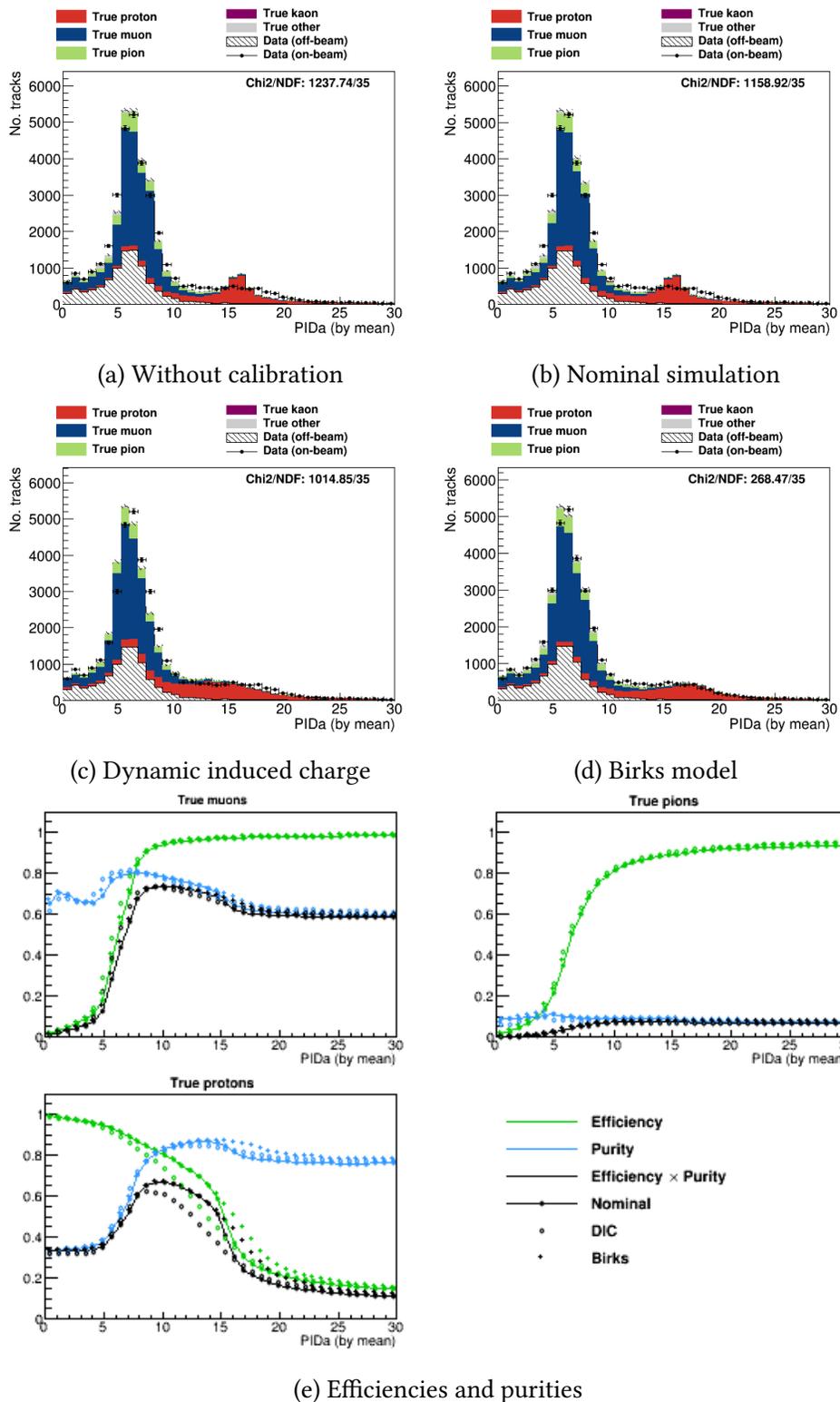


Figure 6.9: Template fit plots of the PIDa-by-mean variable for the collection plane with four different samples.

6.3.4 χ^2 Algorithm

The χ^2 algorithm simply takes the dE/dx and residual range of each reconstructed hit and uses this and the theory curves shown in Figure 6.2 to construct a χ^2 value under the hypothesis of each particle species. One weakness of this algorithm is that it assumes that the underlying dE/dx distribution is Gaussian, and assumes that the track direction and reconstructed end point are correct. It, like the PIDa algorithm, drops the first and last hit from the calculation in order to remove bias.

Many variables can be constructed by combining χ^2 values under different particle species assumptions. In this section a single variable is highlighted, $\chi_\mu^2 - \chi_p^2$, however bare χ^2 distributions can be found in Appendix A.3. This is chosen in the hope that by doing this the scale of the disagreement between data and simulation can be reduced, as any effect which causes a linear increase in the dE/dx of reconstructed tracks should cancel.

Figure 6.10 shows the distributions for each sample and the method efficiencies and purities.

It is unfortunate that even after the template fit, the disagreement between data and simulation is significant. This is particularly true in the proton peak, and in the region between the MIP-like peak and the proton-like peak. This is especially concerning because this is the region where a separator cut might be placed to maximise proton identification efficiency and purity. Note that the extremely high χ^2/NDF for these plots is primarily due to the bins on the right where there are severe differences between the data and simulation. Using the sample which uses the Birks recombination model in the conversion of dQ/dx to dE/dx improves the agreement between the data and simulation, but the χ^2/NDF is still relatively high.

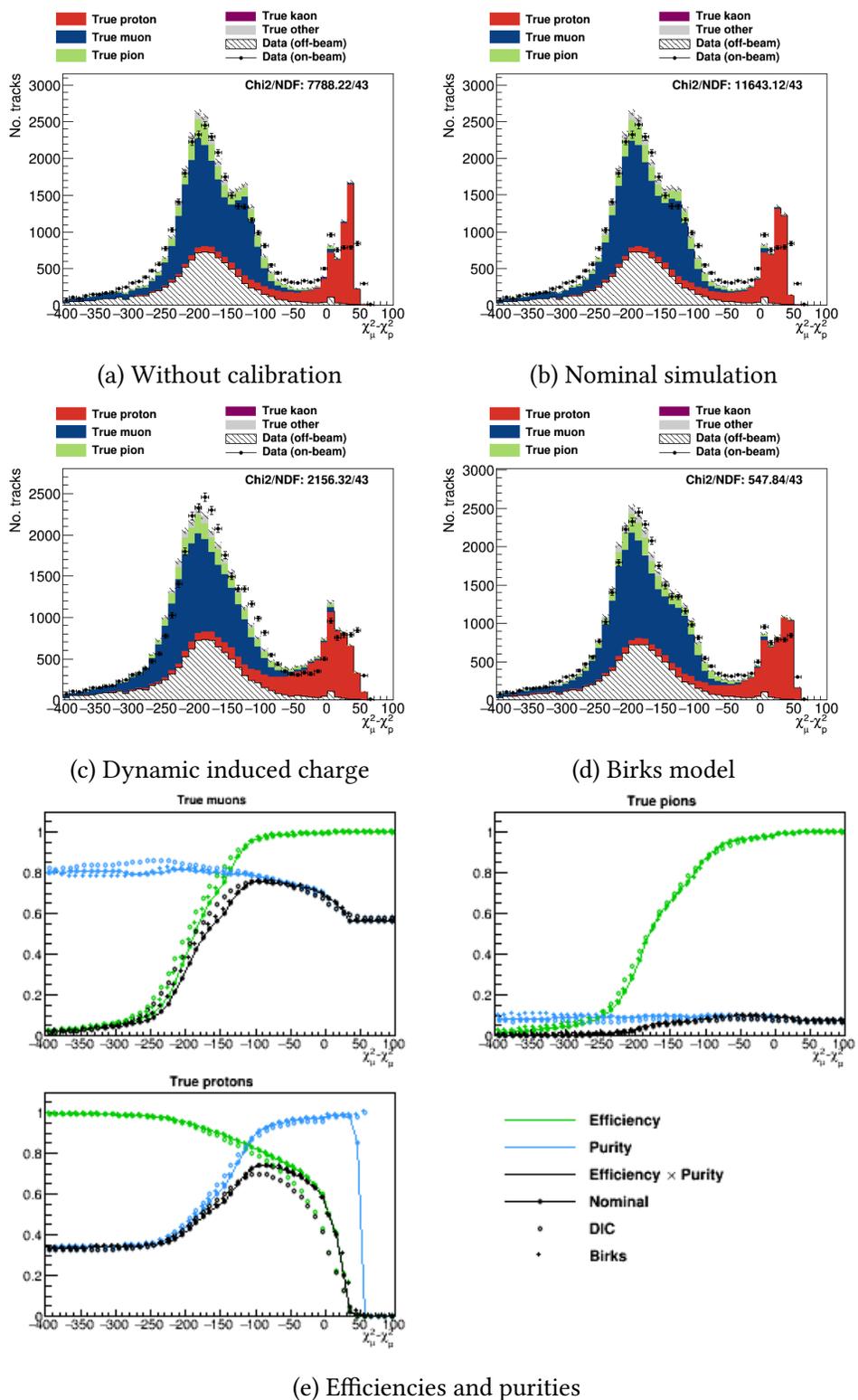


Figure 6.10: Template fit plots of the $\chi_\mu^2 - \chi_p^2$ variable for plane 2 for four different samples.

6.3.5 Bragg Likelihood Algorithm

The final algorithm which is investigated is the Bragg Likelihood algorithm, which has been developed as a part of this work.

Using the Landau and Gaussian widths outlined in Chapter 6.2, a probability map is constructed for each particle species using the theoretical mean dE/dx as a function of residual range close to the Bragg peak (shown for muons, pions, protons, and kaons in figure 6.2), as well as for MIP-like particles which exit the detector and so do not have a Bragg peak. For a given residual range, the theoretical dE/dx prediction is used to determine the mean of the Landau-Gaussian distribution, and the width is taken from the results given in Tables 6.1 and 6.2. The resulting probability maps are shown in Figure 6.11.

These particle maps can be used to construct a likelihood under each particle hypothesis for each track,

$$L_{Total}^s = \frac{\sum_{i=1}^{N_{hits}} L^s(dE/dx_i, R_i)}{N_{hits}}, \quad (6.6)$$

where the sum is over each hit i associated to the track between 0 and 30 cm residual range. The $L^s(dE/dx_i, R_i)$ corresponds to the evaluation of the likelihood map for the particle species, s , at the residual range (R) and dE/dx of the hit. As with previous algorithms, the first and last hits of each track are neglected, as they are known to produce unreliable results.

This calculation is done for each plane separately, but results from multiple planes may be combined to produce a single result by taking the average of the likelihood across multiple planes. Currently, however, this is not performed, as there are known differences in dQ/dx as a function of track angle between data and simulation which are not currently understood. For this reason, only the collection plane is considered for calorimetry.

The algorithm is made to be more robust by the addition of two features:

- For each particle species, the likelihood is first calculated by assuming the track direction is correct, and then the direction is artificially reversed, and the likelihood is recalculated.

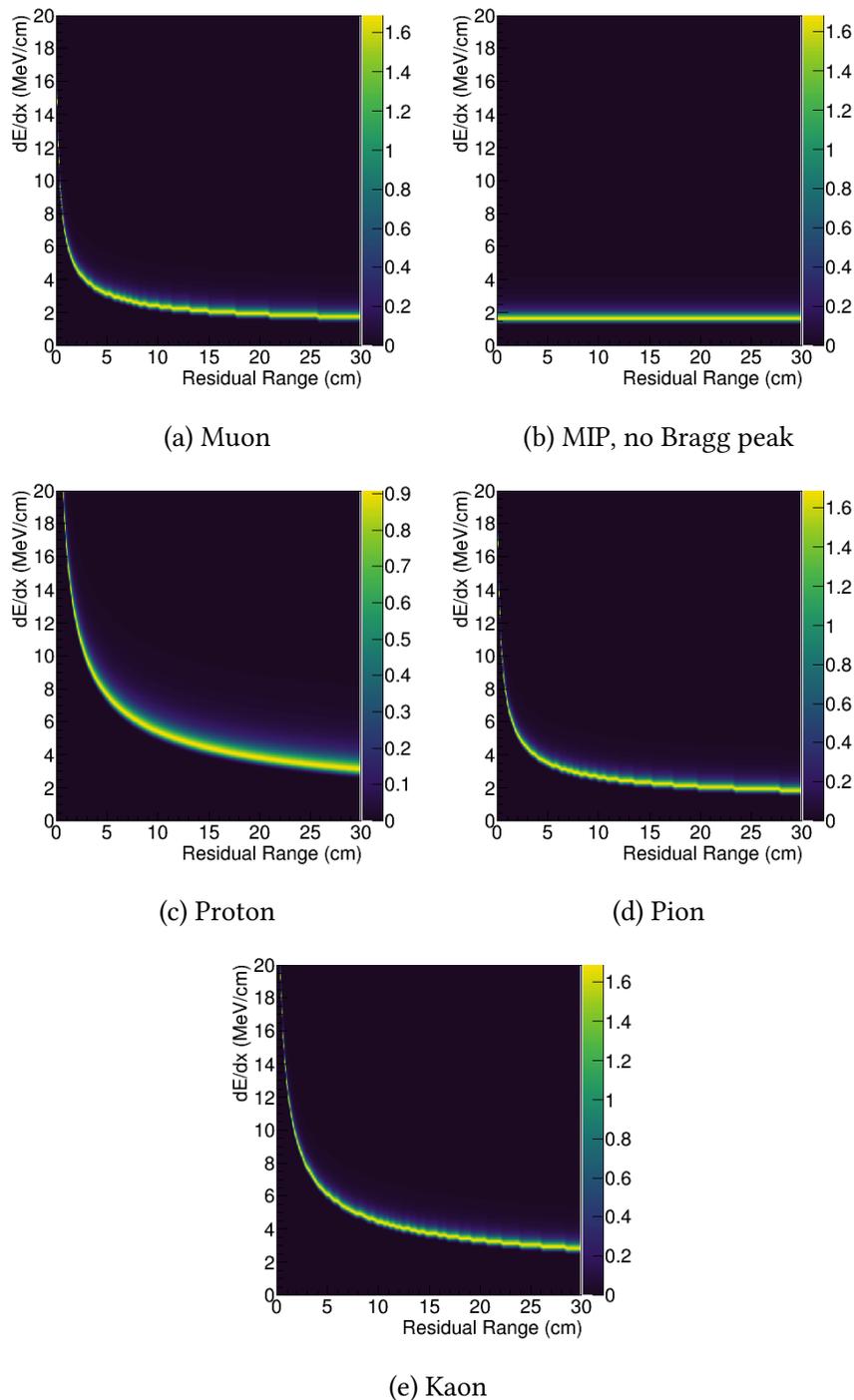


Figure 6.11: Probability maps for each particle species using the Landau and Gaussian widths for the simulated collection plane, as described in the text. Each x bin integrated over y is normalised to an integral of 1 in the range $[0,100]$.

The result which maximises the likelihood is thought to be the correct direction, however both values are stored for future use.

- Additionally, the track end-point resolution is dynamically accounted for. This is done by allowing the hit residual range to float within ± 2 cm, (the estimated 1σ envelope on the end point resolution for tracks from the neutrino pass of the Pandora reconstruction suite, as shown in Figure 6.12), calculating the likelihood for many points with 0.05 cm spacing within these bounds, and taking the maximised likelihood.

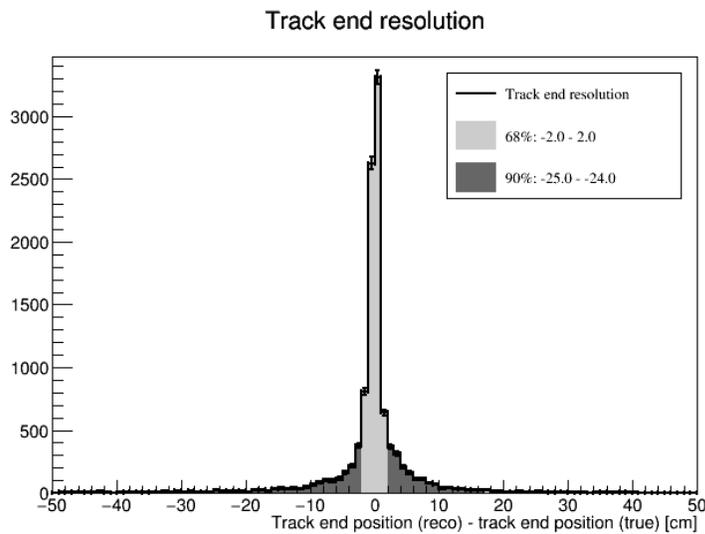


Figure 6.12: End point resolution for BNB-induced tracks reconstructed by neutrino pass of the Pandora reconstruction suite.

Both the data and the simulation suggest that as the the end of the track is approached, the Landau-Gaussian distributions become wider (see, for example, Figures 7.14 and 7.15), however it is believed that this is reconstruction-driven and not physics-driven. The explanation for this is that the end point resolution can have significant impact on the apparent width of the distribution and this is especially dominant at very low residual range, where the dE/dx is rapidly changing. This is shown in Figure 6.13. This effect could be captured by by modifying the width of the distribution as a function of residual range, however due to the dynamic floating of the end-point resolution, as outlined above, this is not necessary.

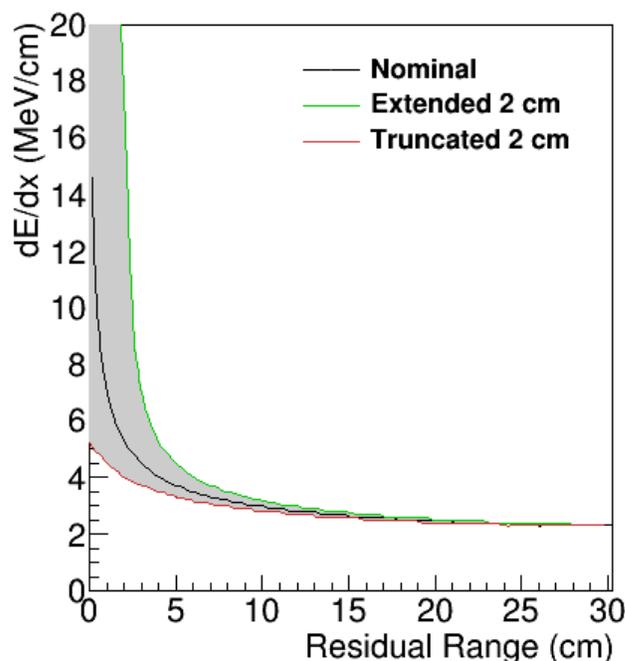


Figure 6.13: Demonstration of the impact of end point resolution on the effective width of the distribution. The black line here is the nominal muon theory as shown in Figure 6.2, the green line is what would be expected if the track was extended by 2 cm, and the red line is what might be expected if the track was truncated by 2 cm. It is clear here that reconstructing an incorrect end point for the track can significantly alter any PID variable which uses the dE/dx and residual range of each hit associated to a track unless this effect is specifically targeted.

Figure 6.14 shows a comparison of data and simulation for the bare likelihoods for muons, pions, protons, kaons, and MIPs. In general, the agreement between data and simulation here is unimpressive. The shape of the proton region is reasonably well approximated in the simulation, however there are significant shape differences in the muon region in each plot. Muons may stop in the detector and exhibit a Bragg peak, or may exit and look like a MIP. Part of the disagreement here may be due to different relative strengths of the different muon populations, although it seems unlikely that this captures the whole problem.

Taking the ratio of some of the variables produced by the BL algorithm can significantly improve the agreement between data and simulation, as shown in Figure 6.15. Here, each likelihood has been normalised by the sum of the likelihoods, therefore enforcing the demand that they

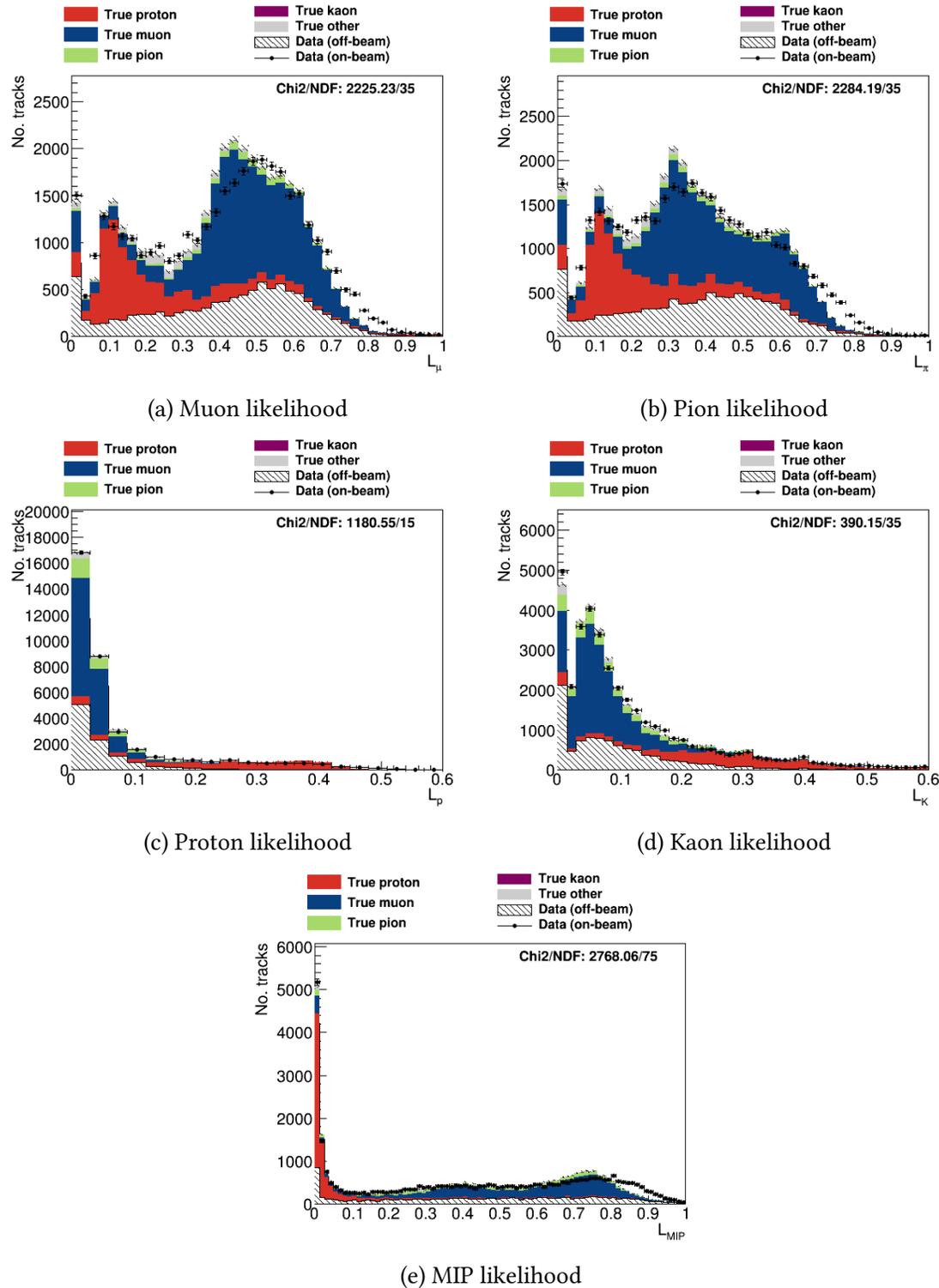


Figure 6.14: Template fit bare likelihoods for tracks under muon, pion, proton, kaon, and MIP assumptions.

range between 0 and 1. The improved agreement seems to indicate that the residual differences between data and simulation for the bare likelihoods in Figure 6.14 are caused by one or more effects which are not contained within the simulation, or where an improved implementation is necessary. Taking the ratio of the likelihood variables helps to mitigate the differences between data and simulation.

For the selection presented in Chapter 7, it is enough to identify tracks as MIP-like (either muon or pion), or proton-like. To maximise separation between these, the variable $\ln(L_{MIP}/L_p)$ is constructed, as shown in Figure 6.16. In this variable we would expect protons to populate the region below 0, and MIP-like particles to populate the region above 0. Both the separation of muons and protons, and the agreement between the data and simulation in this variable are extremely good. The efficiency and purity plots presented in Figure 6.16 also indicate that this algorithm is reasonably insensitive to the detector effects which were investigated.

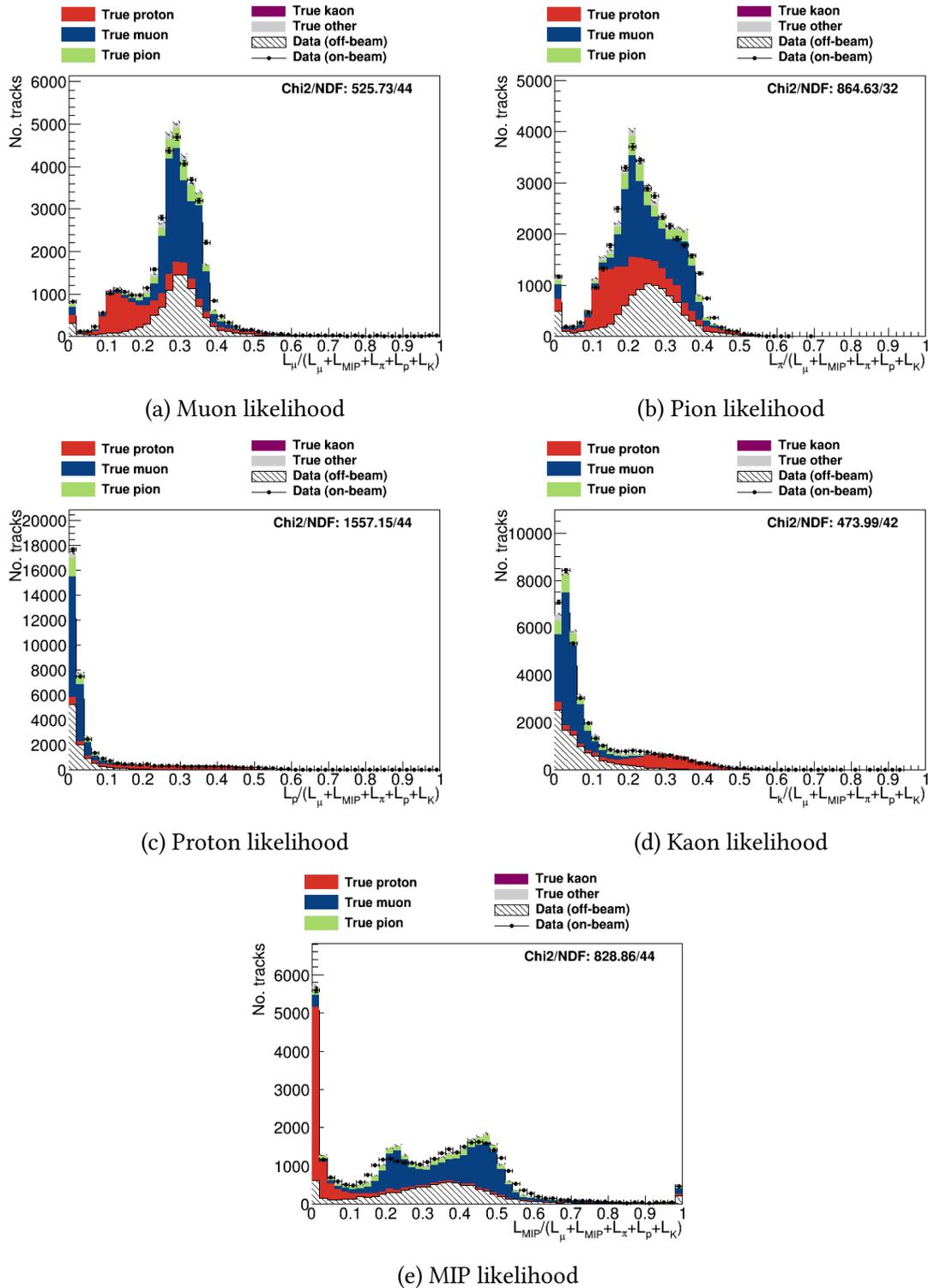
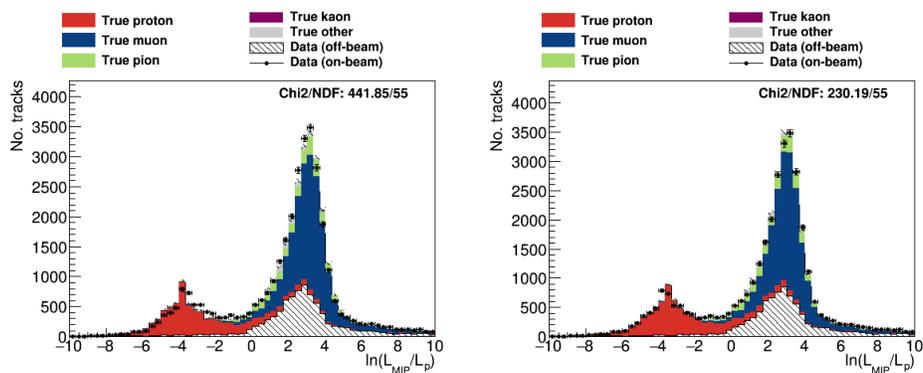
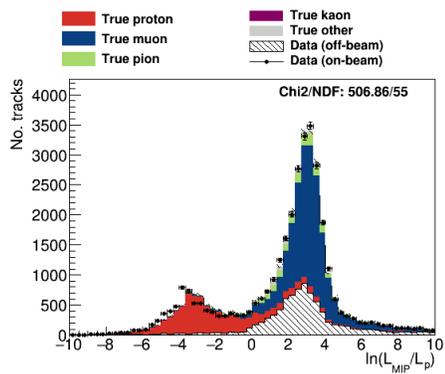


Figure 6.15: Template fit likelihoods for tracks under muon, pion, proton, kaon, and MIP assumptions, normalised from 0 to 1.

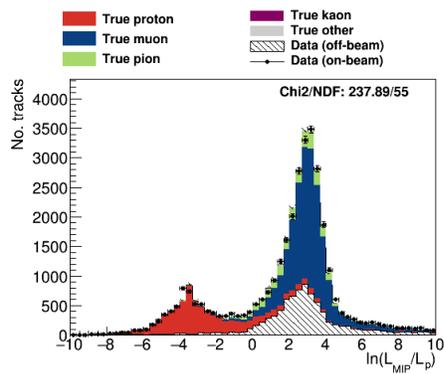


(a) Without Calibration

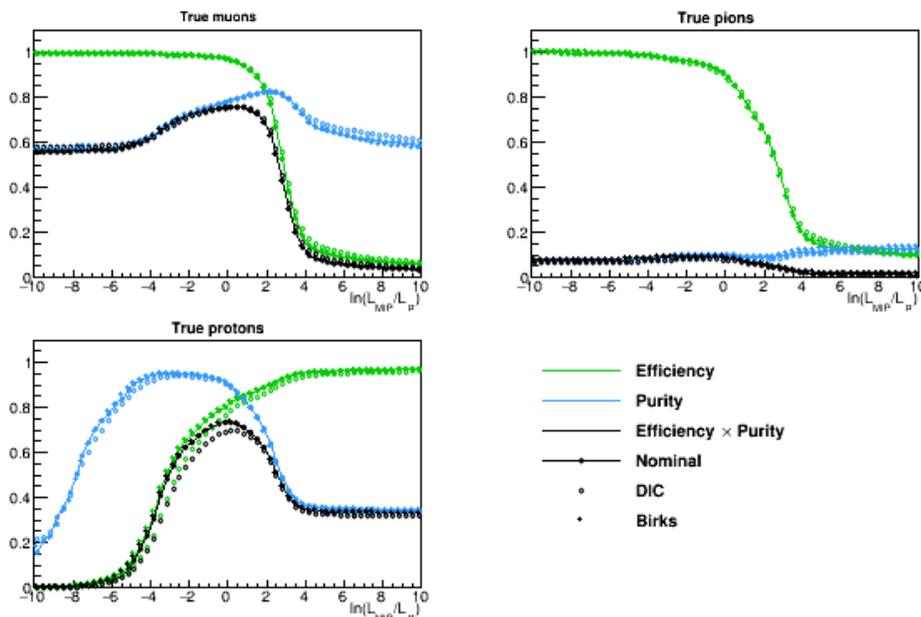
(b) Nominal simulation



(c) Dynamic induced charge



(d) Birks model



(e) Efficiencies and purities

Figure 6.16: Template fit plots of the $\ln(L_{MIP}/L_p)$ ratio for plane 2 for four different samples.

6.4 Discussion

In order to compare the efficiency and purity of the four PID methods outlined in the previous section, Figure 6.17 shows the efficiency versus purity curves for each method, for the nominal simulation, as well as for the two samples used to assess their systematic uncertainties. These are generated by varying the cut value for each algorithm. The curves are separated by particle species. It is clear that the PIDa and *depErangeE* methods underperform compared to the $\ln(L_{MIP}/L_p)$ and $\chi_\mu^2 - \chi_p^2$ algorithms, which are similar in performance to each other.

Using Figure 6.17, a single test cut value can be chosen, and summary tables for muon and proton identification efficiencies and purities across the nominal and two detector variation samples can be found in Table 6.5. The cut value that has been chosen for each variable approximately maximises the proton selection $\epsilon \times p$.

The muon efficiencies and purities for the $\ln(L_{MIP}/L_p)$, $\chi_\mu^2 - \chi_p^2$, and PIDa algorithms are comparable: these algorithms select muons with 95%+ efficiency and around 80% purity, and this can be slightly tuned by choosing cut values for the needs of a specific analysis. However, there is a wide range of possible proton efficiency and purities. The BL algorithm is able to attain excellent proton efficiency without sacrificing too much purity (both approaching around 85%), while the $\chi_\mu^2 - \chi_p^2$ algorithm is able to slightly beat out the BL algorithm on purity, with a sacrifice to the efficiency.

Another vital consideration when evaluating the particle ID algorithms is the agreement between data and simulation. If the simulation does not describe the shape of the data well, then this reduces confidence in the ability to select the intended particles with the quoted efficiency and purity. Table 6.6 shows the goodness-of-fit χ^2/NDF between the data and simulation for each of the three detector samples, both for the POT-normalised distribution and after the template fit. These numbers are taken from the figures shown in Section 6.3. In general, a higher level of agreement between data and simulation is observed in the BL-based variables over the other variables. This is particularly true after performing the template fit in which the normalisation of different particle species is allowed to change, but it is also true of the POT normalised distribu-

μ							
Variable	Cut Value	Nom. ϵ	DIC. ϵ	Birks. ϵ	Nom. P	DIC. P	Birks. P
$\ln(L_{MIP}/L_p)$	0.0	0.97	0.97	0.97	0.78	0.78	0.78
$\chi_\mu^2 - \chi_p^2$	-90.0	0.98	0.97	0.97	0.78	0.78	0.78
PiDa (mean)	9.0	0.92	0.92	0.91	0.80	0.79	0.79
depERangeE	10.0	0.76	0.79	0.76	0.78	0.77	0.78
P							
Variable	Cut Value	Nom. ϵ	DIC. ϵ	Birks. ϵ	Nom. P	DIC. P	Birks. P
$\ln(L_{MIP}/L_p)$	0.0	0.81	0.78	0.81	0.90	0.90	0.90
$\chi_\mu^2 - \chi_p^2$	-90.0	0.80	0.77	0.81	0.92	0.90	0.91
PiDa (mean)	9.0	0.83	0.79	0.83	0.80	0.79	0.79
depERangeE	10.0	0.85	0.79	0.90	0.63	0.63	0.63

Table 6.5: Muon (top) and proton (bottom) efficiencies and purities for each algorithm for a cut value chosen to approximately maximise efficiency and purity for selecting protons.

tions presented in Appendix A.2.1. In many of the distributions shown in Section 6.3 it seems that the overall proton normalisation is too high in simulation, and the agreement between data and simulation can be improved by decreasing the total number of protons. This effect will be touched upon in Chapter 7. The *depErangeE* algorithm has a similar level of agreement between the data and simulation as the BL algorithm, but as mentioned above, the μ/p separation is significantly reduced.

Variable	POT-norm. χ^2/NDF			Template Fit χ^2/NDF		
	Nom.	DIC.	Birks.	Nom.	DIC.	Birks.
$\ln(L_{MIP}/L_p)$	28.52	11.71	28.12	4.19	9.22	4.33
$\chi^2_{\mu-p}$	194.32	45.66	41.39	270.77	50.15	12.74
PIDa (mean)	60.07	28.26	39.69	33.11	29.00	7.67
depERangeEMu	29.91	16.18	24.33	5.29	11.60	3.78

Table 6.6: Goodness-of-fit χ^2/NDF showing data/simulation agreement for each algorithm, both for the POT-normalised distribution and after the template fit.

The robustness of the algorithms can be analysed by looking at how the efficiencies, purities, and agreement between data and simulation changes between the three different samples. By looking at Figure 6.17, and Table 6.5 it can be observed that in terms of efficiency and purity, most algorithms are reasonably insensitive to the Birks recombination modification, but they each have a 3-5% sensitivity on the DIC modification. This is somewhat expected as the DIC model results in an additional dE/dx smearing over the nominal sample (that is, a wider distribution in dE/dx) leading to an over-smearing when the tuning is applied to the simulation.

By looking only at the variations of efficiency and purity, each algorithm appears to be reasonably robust, however Table 6.6 tells a different story. Here, the spread in χ^2/NDF values makes it clear that the $\ln(L_{MIP}/L_p)$ and *depErangeE* algorithms, are much more robust against these systematic uncertainties than the remaining algorithms.

One could ask whether the different algorithms are accessing different aspects of the phase space; if they do, then select algorithms could be used in a multi-variate analysis in order to bolster the particle identification capabilities at MicroBooNE. An understanding of the overlap of

the samples can be gained by construction of several Venn diagrams, as shown in Figure 6.18.

Figures 6.18a and 6.18b show the overlap of the BL and χ^2 algorithms with the PIDa and *depErangeE* algorithms. The overlap between the first two of these is very large, which is to be expected as they are utilising the same information. The L_{MIP}/L_p algorithm selects slightly fewer protons than the χ^2 algorithm because it is slightly less efficient. The efficiencies for proton identification for the PIDa and *depErangeE* are reasonably large and so the overlap between these algorithms and the BL/ χ^2 algorithms is reasonably large. For both of these algorithms this efficiency comes with the penalty of worse muon identification rates (indicated by their much lower muon identification efficiency, and lower proton identification purity).

The last comparison which has been made is the overlap of those events selected by PIDa versus those selected by *depErangeE* and this is shown in Figure 6.18c. For protons, the two algorithms are accessing very similar phase space, however for muons there is a significant difference. This indicates that while neither are particularly pure, the muon backgrounds which contaminate the protons are different in each algorithm indicating that they are sensitive to different effects.

Because the overlap between the algorithms presented here is rather large, it seems unlikely that feeding these algorithms into a multi-variate analysis would result in significant improvement to the particle identification.

Based on the work in this chapter, the choice is made to proceed using the $\ln(L_{MIP}/L_p)$ variable as the primary PID method. This choice is made as the algorithm is able to replicate the high efficiency and purity of the $\chi_\mu^2 - \chi_p^2$ variable, but with a significantly improved agreement between data and simulation, and an improved robustness against the systematic uncertainties discussed in this section.

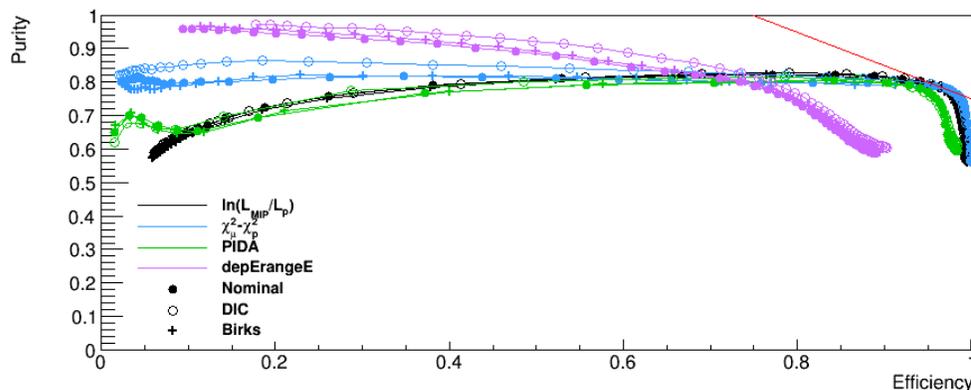
The final test of the particle identification algorithm is to ensure that it is resilient against changes to the underlying interaction model. In absence of, for instance, a sample using a different neutrino interaction generator, this is investigated by looking at the agreement between data and simulation and the efficiency and purity for different track lengths and angles. Particular interest

is paid to low track lengths and tracks near to 90 degrees to the beam direction, as they represent lower energy particles which may be expected to be less well modeled.

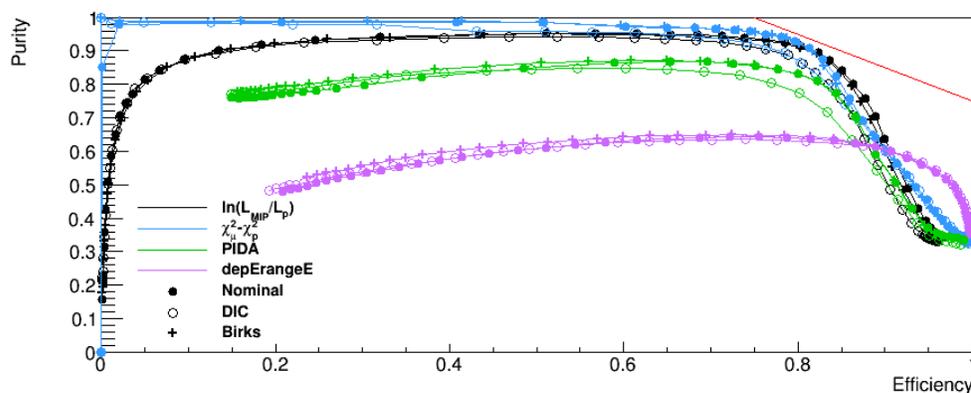
A comparison of data with simulation for the lowest chosen track length bin (between 0 and 2.5 cm) and the lowest chosen $|\theta_x|$ ¹ bin (0 to 30 degrees) can be found in Figures 6.19 and 6.20 respectively. For each of these plots there is a significant disagreement between the data and simulation when the nominal simulation is used, however using the induced charge simulation mitigates those changes to a large extent.

Proton and muon selection efficiencies and purities as a function of track length and $|\theta_x|$ are shown in Figure 6.21. In general these distributions are reasonably flat in both efficiency and purity. The one distribution which does follow this trend is the muon purity as a function of track length, where there is a discontinuity around 20 cm. This is can be simply explained as being due to most muons from neutrino interactions having lengths above 20 cm, while a significant number of cosmic-ray muons also appear below 20 cm.

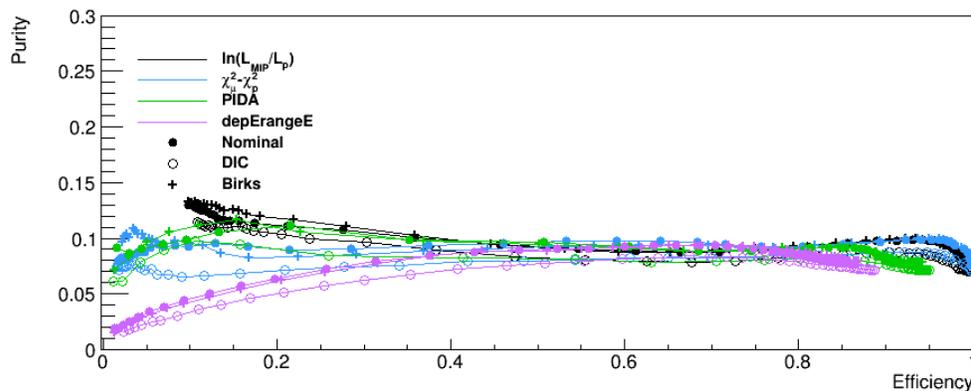
¹ θ_x is defined to be 0 degrees in axis of the drift direction, with ± 90 degrees equivalent to the beam and anti-beam direction. Here, the absolute value of $|\theta_x|$ has been taken.



(a) Muon

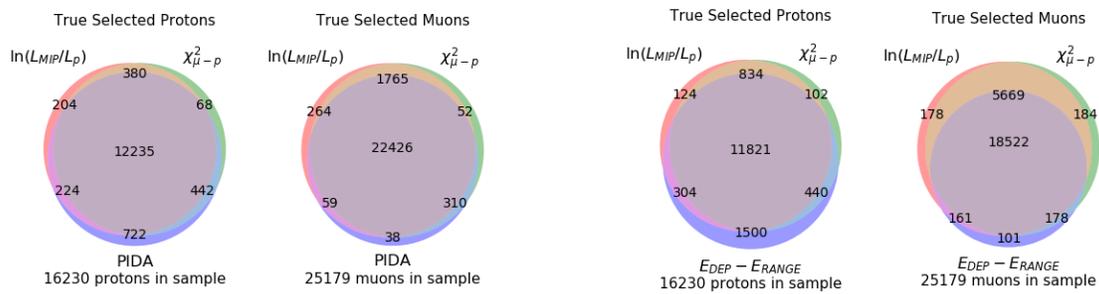


(b) Proton



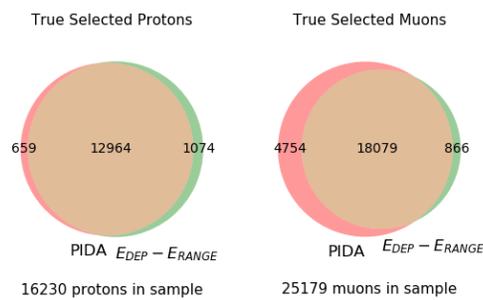
(c) Pion

Figure 6.17: Curves in efficiency/purity space for muons, protons, and pions. For each plot, the performance of the four potential algorithms are presented using the nominal simulation as well as the DIC and Birks recombination samples. The red line on the muon and proton plot represent lines of $\epsilon \times p = 0.75$, and the performance of each algorithm can be qualitatively estimated by how close the curves approach this line.



(a) Overlap of $\ln(L_{MIP}/L_p)$, $\chi_{\mu-p}^2$, and PIDa.

(b) Overlap of $\ln(L_{MIP}/L_p)$, $\chi_{\mu-p}^2$, and $E_{dep} - E_{range}$.



(c) Overlap of the $E_{dep} - E_{range}$ and PIDa algorithms

Figure 6.18: Venn diagrams showing the overlap of each chosen algorithm for true muons selected as muons and true protons selected as protons

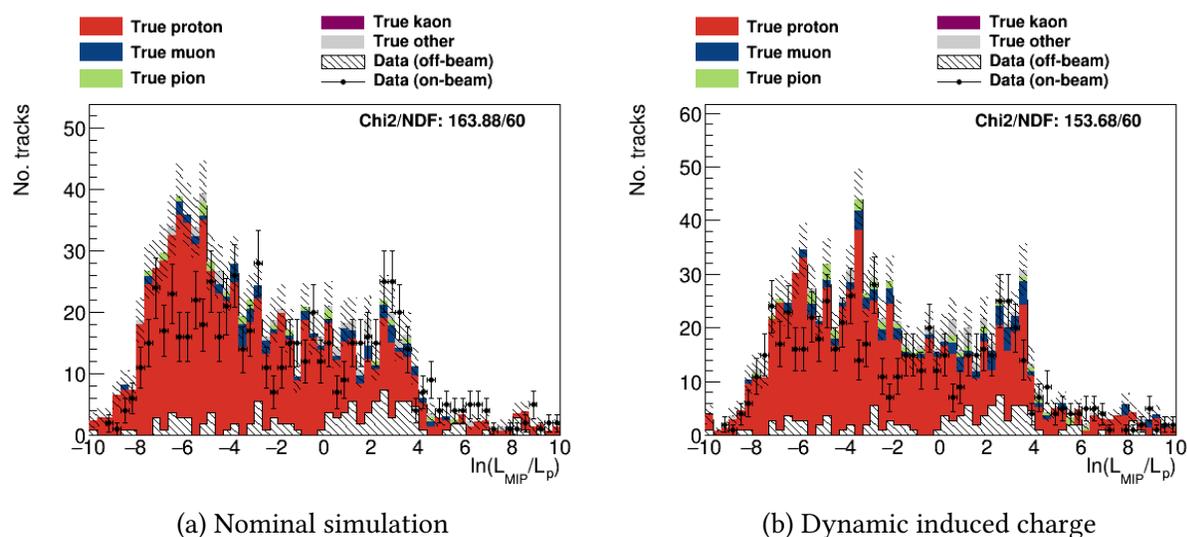


Figure 6.19: A comparison of data and simulation for tracks between 0 and 2.5 cm. This is dominated by protons, which tend to be much shorter than protons. There is a shape disagreement at low $\ln(L_{MIP}/L_p)$ values, however this is almost completely mitigated by the effects of induced charge.

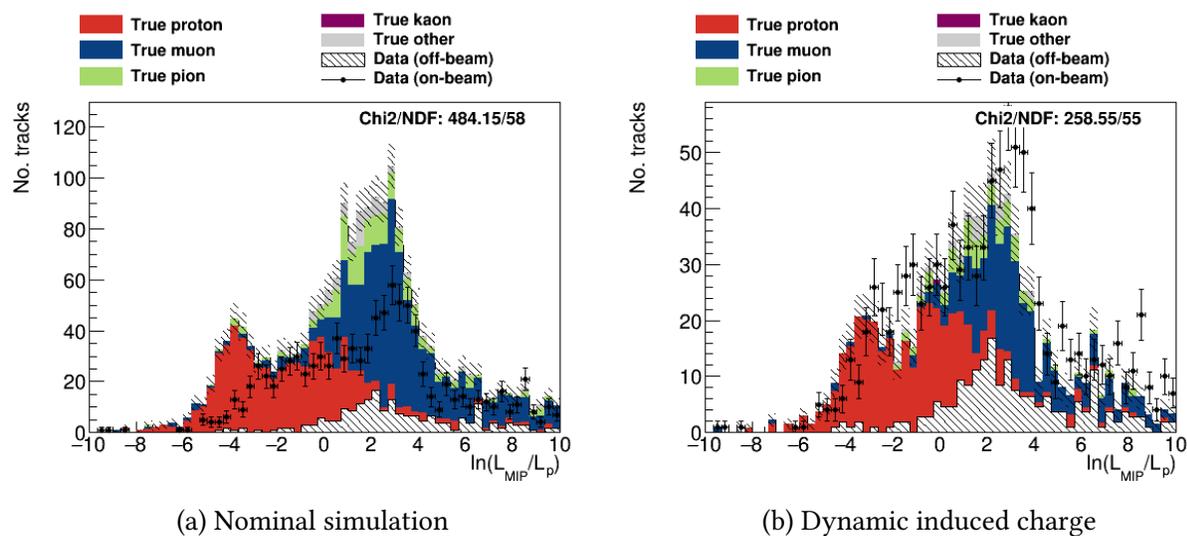


Figure 6.20: A comparison of data and simulation for tracks with $\theta_x < 30$ degrees. There is very clearly a large normalisation and shape difference in the nominal simulation, however including the effect of induced charge largely removes this disagreement, aside from a slight dip in the simulation around $\ln(L_{MIP}/L_p)$, which does not appear to be replicated in the data.

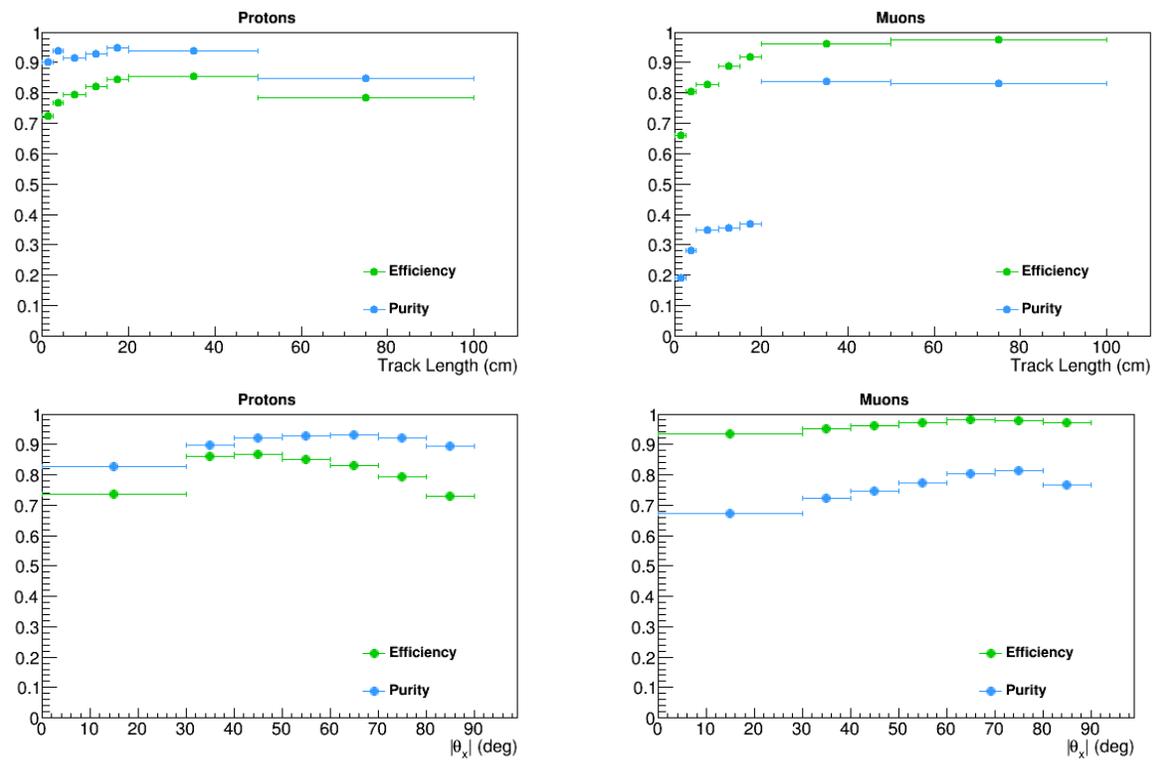


Figure 6.21: Proton and muon selection efficiencies and purities for the $\ln(L_{MIP}/L_p)$ algorithm as a function of track length and θ_x . The efficiencies and purities are generally smoothly varying, and follow shapes which are expected.

Chapter 7

Selection of ν_μ CC $0\pi Np$ Events

“In [physics] almost everything is already discovered, and all that remains is to fill a few unimportant holes”

Phillipp von Jolly, to Max Planck (1878)

The analysis presented in this work has been performed in order to constrain the systematic uncertainties of an electron neutrino dataset. In order to perform the systematic constraint, this chapter will outline the development of a selection of ν_μ CC events with 0 pions and N protons above 40 MeV kinetic energy in the final state (ν_μ CC $0\pi Np$).

7.1 Signal Definition and Backgrounds

This signal definition was chosen in order to approximate the signal definition used by Mini-BooNE (CCQE), while the demand of at least one proton was chosen to minimise the contamination from cosmic rays. The 40 MeV kinetic energy lower limit on the signal definition was chosen to ensure that the proton crosses enough channels such that particle identification can be performed.

ν_μ CC $0\pi Np$ interactions are defined as containing exactly one muon, and some number of

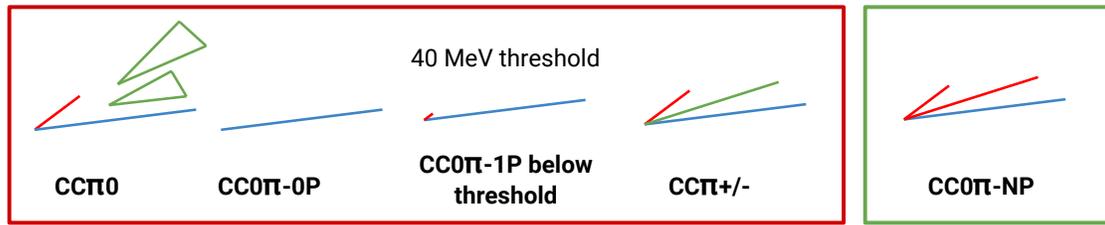


Figure 7.1: Primary ν_μ CC backgrounds to the selection of the signal are shown in the red box, where a blue line represents a muon, a red line represents a proton, and a green line or triangle represents a charged or neutral pion, respectively.

protons above 0 emanating from the vertex. The main backgrounds to this are CC interactions which contain a charged or neutral pion, those without a proton, and NC events. These are shown in Figure 7.1.

Many of the figures in this chapter contain comparisons of MicroBooNE data and simulation, and contain a stacked histogram of the signal and backgrounds. These backgrounds have been separated into the categories described below.

Cosmic: The selected interaction contains only particles of cosmic origin.

Mixed: Due to misreconstruction, the selected interaction contains both neutrino-induced particles and particles of cosmic origin.

OOFV: Events with a true vertex outside of the fiducial volume, where the reconstructed vertex is within the fiducial volume. This primarily happens due to the space charge effect.

Dirt: Events selected from the dirt sample.

NC: Events selected which contain a ν_μ neutrino interaction, but where the interaction is via the neutral current.

$\nu_e, \bar{\nu}_e$: Events selected which contain a neutrino interaction, but where the interacting neutrino is either a ν_e or $\bar{\nu}_e$.

$\bar{\nu}_\mu$: Events selected which contain a neutrino interaction, but where the interacting neutrino is a $\bar{\nu}_\mu$ event.

ν_μ **CC Other**: Selected ν_μ CC interactions which are not signal. These events may contain a pion, or may not contain a proton.

ν_μ **CC0 π NP**: Selected signal events.

Off-beam data Data taken outside of the beam window but with the same constraint on the software trigger.

On-beam data Data taken within the beam window with constraints on the software trigger.

7.2 Data Samples

The data samples which have been used in this analysis are described in this section. A summary of the POT of the datasets used in this analysis is found in Table 7.1.

7.2.1 Run 1 Data

The data used in this analysis represents 5×10^{19} protons on target of collected beam data from MicroBooNE Run 1. This represents less than 5% of the full 13×10^{20} POT which MicroBooNE is expected to collect over its lifetime, and is the data which is currently open for the purposes of developing analyses.

In addition to this, all of the Run 1 off-beam data is used.

7.2.2 Simulated Data

The nominal simulation sample used in this analysis corresponds to around 2,000,000 events, or 1.86×10^{21} POT. These events are generated within the MicroBooNE TPC according to the simulation chain outlined in Chapter 4.

7.2.3 Out-of-TPC Simulation

The nominal simulated dataset contains neutrino interactions which have been simulated within the MicroBooNE TPC. Particles from interactions which take place outside of the TPC may enter and be selected as neutrino candidates. These are referred to as *Out-of-TPC* or *dirt* backgrounds.

This happens primarily because the MicroBooNE TPC is contained within a cylindrical cryostat which is filled with liquid argon. Approximately half of the total mass of liquid argon is not contained within the TPC, and this means that approximately half of the interactions in MicroBooNE occur inside of the cryostat but outside of the TPC. These interactions are able to have a flash in-time with the beam which may be detected by the light collection system, meaning they may be selected as a neutrino candidate. Due to processing constraints, these events are not contained in the nominal simulation, and so a separate simulation of these events is produced.

This dataset corresponds to $1.69\text{e}+21$ POT.

7.2.4 Detector Variation Samples

In order to assess the impact of the detector-related systematic uncertainties in MicroBooNE, several *detector variation* datasets are leveraged. These are simulated samples where parameters thought to have some systematic effect on analyses are varied by their $\pm 1\sigma$ bound, or where specific models are turned on or off in order to gauge the magnitude of their effect. Each detector variation dataset uses the same underlying events from the GENIE generator in order to remove statistical uncertainty as a factor.

One drawback of separating the out-of-TPC events from the simulation is that the detector variation samples described below do not affect the dirt backgrounds, however the out-of-TPC contamination in neutrino selections is generally small due to fiducial volume constraints, and so the effect of this is estimated to be small.

The variation samples presented here are intended to be a conservative treatment, and work is ongoing within the collaboration to constrain many of these through in-situ measurements. Many of the effects listed here are described in greater detail in Chapter 3.

Central Value (CV) The nominal MicroBooNE simulation with no additions or modifications.

Note that this is the same simulation as presented in Section 7.2.2, but it contains the same underlying events as the other detector variation samples such that they are not statistically independent.

Space Charge Effect (SCE) A data-driven correction is applied to the nominal simulation of the SCE. This dataset is then generated using the data-driven SCE model which moves the simulation to better match what is measured in the data. This data-driven correction is applied to both the spatial migration map and the electric field modification map.

Light Yield (LY) In the central value simulation, there were a number of bugs contained in the simulation of scintillation light. The primary bug here is that the simulation assumed that every particle produced light assuming it was a true electron meaning the number of photons/cm is constant for all particle species. This has been corrected in this variation sample.

Longitudinal Diffusion (LD Up and LD Down) There are two variations here: one turns down the longitudinal diffusion coefficient and the other turns the coefficient up. Currently this is estimated from world data to be $6.2 \text{ cm}^2/\text{s}^{+57\%}_{-47\%}$. In future iterations of the systematic uncertainty treatment, the result obtained in Chapter 5 will supersede these values.

Transverse Diffusion (TD Up and TD Down) Both transverse and longitudinal diffusion are simulated assuming MicroBooNE's design voltage, 500 V/cm. In the case of longitudinal diffusion the uncertainties easily cover the difference between this ($6.2 \text{ cm}^2/\text{s}$) and the correct value from a fit to world data ($6.36 \text{ cm}^2/\text{s}$), however in the case of transverse diffusion this is not the case, and so both variations fall below the nominal D_T value and the $\pm 1\sigma$ is $16.3 \text{ cm}^2/\text{s}^{-24.5\%}_{-49.6\%}$. The world data for transverse diffusion does not cover the MicroBooNE electric field, and so the uncertainties here were chosen from theoretical models.

Wire Noise (WN Up and WN Down) The nominal simulation uses a data-driven model of wire noise, as described in Chapter 4. The magnitude of the noise is varied up and down by 10%.

PMT Noise (PMTN Up and PMTN Down) The main noise on the MicroBooNE PMTs is *single PE noise*, which is seen at a rate of 250 Hz. The systematic variations here use a $\pm 1\sigma$ of $\pm 50\text{Hz}$, which is an approximate $\pm 1\sigma$ variation on measurements taken from the data.

Induced Charge (DIC) Induced charge has been widely discussed within this work. There is no induced charge simulation in the nominal MicroBooNE simulation. This detector variation introduces a preliminary simulation of this effect, and is referred to as the *Dynamic Induced Charge* simulation.

Wire Response Function (SQUEEZE RF and STRETCH RF) The wire response function is measured from the data, however studies on simulation have shown that there is some bias on this measurement. This is a roughly 20% bias, and so these detector variation samples modify the width of the response function by \pm this amount.

Removing Channels Prone To Saturation (SAT) A number of channels in the MicroBooNE TPC will occasionally have charge build up on the capacitors in the ASICs, and this results in an amount of dead time on wires connected to these ASICs. This detector variation simulation simulates these channels as being non-responsive in order to understand the effect on the reconstruction.

Removing Misconfigured Channels (MIS) This variation simulates misconfigured channels as being non-responsive in order to place an upper bound on their effect on reconstruction.

Light Outside of the TPC (EXTTPCVIS) The light production from charged particles outside of the TPC is thought to be incorrectly simulated because effects such as the reflectance of the cryostat are not present, and so this variation increases the light yield in this region by 50%.

Electron Lifetime (LT) The electron lifetime in MicroBooNE has been measured to be very high, and is treated to be quasi-infinite in the nominal simulation. This detector variation

estimates a lower bound on the effect of electron lifetime by simulating the lifetime to be 10 ms.

Recombination Model (BIRKS) The nominal MicroBooNE simulation uses the Modified Box model tuned to ArgoNeuT data. This substitutes the Birks recombination model [92], tuned to ICARUS data [103], in the MicroBooNE reconstruction (note that the Modified Box model is still used in simulation).

Sample	Number of Triggers	POT	Scale Factor
On-beam data	10905211	4.89e+19	-
Off-beam data	77329137	-	0.141
BNB+Cosmic	-	1.86e+21	0.026
Out-of-TPC	-	1.69e+21	0.029
CV	-	1.94e+20	0.253
SCE	-	3.92e+20	0.125
LY	-	1.97e+20	0.248
LD Up	-	1.95e+20	0.251
LD Down	-	1.97e+20	0.248
TD Up	-	1.95e+20	0.250
TD Down	-	1.98e+20	0.248
WN Up	-	1.94e+20	0.252
WN Down	-	1.96e+20	0.250
PMTN Up	-	1.96e+20	0.250
PMTN Down	-	1.98e+20	0.247
DIC	-	1.96e+20	0.250
STRETCH RF	-	1.94e+20	0.252
SQUEEZE RF	-	1.94e+20	0.252
SAT	-	1.97e+20	0.248
MIS	-	1.93e+20	0.252
EXTTPCVIS	-	1.97e+20	0.249
LT	-	1.97e+20	0.249
BIRKS	-	1.97e+20	0.248

Table 7.1: Datasets used in the development of the ν_μ CC $0\pi Np$ selection, and understanding of the detector-related systematic uncertainties.

7.3 The ν_μ CC-inclusive selection

The selection of ν_μ CC0 π NP events leverages MicroBooNE's ν_μ CC-inclusive selection as a pre-selection and builds on this to remove both events containing pions and those which do not contain protons. This pre-selection is described in brief in this section. Full details of the selection can be found in reference [104].

The analysis leverages the Pandora multi-algorithm pattern recognition software (see [105, 91], and Chapter 4.6) as part of the reconstruction.

7.3.1 Rejection of Cosmic-Ray Backgrounds

MicroBooNE, being a surface level detector, has a large number of cosmic rays incident upon it, and these make up the most significant background to any neutrino selection. This, combined with its relatively slow readout ($\mathcal{O}(ms)$) means that each event read out from the detector contains $\mathcal{O}(10)$ cosmic ray muons. In order to perform neutrino physics, it is then necessary to process each event through a cosmic removal stage. In practice, this is can be done by using topological, optical, and high-level reconstructed information, and is done in several stages.

Topological Cosmic Rejection

First, cosmic tagging is performed by tagging PFPs which have hits which are not consistent with the neutrino beam spill window. Nominally, the MicroBooNE readout window is around 9600 ticks in time (where 1 tick is $0.5 \mu s$), however neutrinos are known to arrive within a specific 4600 tick window, and so anything which is not fully contained within this window can be identified as being of cosmic origin.

In practice, this cosmic tagging is performed in two ways:

- If a PFP has 4 or more hits in one or more planes reconstructed as being out of time, then the PFP is determined to be of cosmic origin and is not considered in the analysis.

- Remaining PFPs then undergo geometric cosmic tagging to determine whether it both enters and exits the TPC by checking the reconstructed start and end point of each PFP against the fiducial volume (FV) defined in table 7.2. The relatively tight FV cut in the y position is necessary to remove cosmic tracks for which the start and end points have been shifted due to the space charge effect. Note that a smaller fiducial volume is used in event selection.

TPC Dimension	Fiducial Volume
x	10 cm
y	30 cm
z	20 cm

Table 7.2: Summary of the fiducial volume used in the geometric cosmic tagging outlined in the text.

Optical Cosmic Rejection

Following this, optical information is leveraged to remove additional cosmogenic tracks. The geometry of each PFParticle is used to generate a flash hypothesis on MicroBooNE's 32 PMTs and this is compared against the flash coincident with the 1.6 μ s beam window.

The flash in the beam window has a reconstructed Z-position of

$$\langle Z \rangle = \left(\sum_{i=0}^{32} Z_{PMT_i} \times PE_i \right) / \sum_{i=0}^{32} PE_i = Z_{flash} \quad (7.1)$$

with a standard deviation given by

$$\sqrt{\text{Var}(Z)} = \sqrt{\langle Z^2 \rangle - \langle Z \rangle^2} = \Delta Z_{flash}. \quad (7.2)$$

If the number of PE reconstructed on a single PMT is greater than 3σ away from the hypothesised number of PE and the hypothesised Z_{flash} of the flash is not within ΔZ_{flash} of the reconstructed Z_{flash} , then the PFP is identified as being of cosmic origin.

Further cosmic rejection is applied by looking for so-called Anode-or-Cathode Piercing Tracks

(ACPT). For a given track which truly enters the TPC through the top and leaves through the anode, then the reconstructed time of the track is $t_t \sim t_A$, where $t_A = x_A/v_d$ is the time at the anode. The true track should have an associated flash with time t_F , and so we can construct the relationship

$$t_t - t_F \sim t_A \quad (7.3)$$

For each track and flash pair (other than the candidate neutrino flash), $t_t - t_F - t_A$ can be calculated. If this is approximately zero then the track can be identified as being cosmogenic. This can be repeated under the assumption that the track enters the anode and leaves the bottom of the TPC, and under the assumptions that the track pierces the cathode and either the top or bottom planes of the TPC. For each iteration it is ensured that the reconstructed flash position is compatible with the reconstructed track.

Cosmic Rejection Using High-Level Properties

In addition to this, the following two algorithms are aimed at removing cosmic rays which stop in the detector. This type of cosmic ray can be particularly challenging for analysis because they can mimic a neutrino topology in the case where there is a Michel electron. This is complicated by imperfect reconstruction where the electron may be reconstructed as either a track or a shower, may not be reconstructed at all, or may even be merged with the cosmic muon.

- The first algorithm is pitched at removing cosmic muons which have a well reconstructed Michel electron, and uses a series of techniques at the hit-level to try and identify the Bragg Peak of the muon, followed by a drop in the dQ/dx , which represents the presence of the Michel electron.
- The second algorithm makes use of a Multiple Coulomb Scattering (MCS) fit in order to determine the direction of the muon. As a charged particle of momentum p traverses a medium, it undergoes electromagnetic scattering off of atomic nuclei. This can be modeled as a Gaussian functional form centered at 0 with an RMS, σ_0^{HL} , parameterised by the

Highland formula [106],

$$\sigma_0^{HL} = \frac{S_2}{p\beta c} z \sqrt{\frac{\ell}{X_0}} \left[1 + \epsilon \times \ln \left(\frac{\ell}{X_0} \right) \right], \quad (7.4)$$

where ℓ is the distance traveled within the medium, X_0 is the radiation length of the medium (14 cm in liquid argon), z is the magnitude of the charge of the particle (unity, for muons), $\beta = v/c$, and the S_2 and ϵ parameters have been determined to be 13.6 MeV and 0.038, respectively.

σ_0^{HL} is inversely proportional to β , indicating that a track should scatter by higher angles as the momentum decreases. By performing a likelihood fit in each direction of the track, a preferred direction can be determined. If the track is determined to be entering the detector then it is identified as a cosmogenic track.

7.3.2 Flash to TPCObject Matching

Any remaining tracks and showers which are determined to be related are grouped together into a *TPCObject*. This grouping is done by following the PFParticle hierarchy, contained within the ParticleFlow Object, as described in Chapter 4.6, i.e. starting from each reconstructed neutrino and including all of the daughter particles iteratively until all particles in the hierarchy have been included in the TPCObject.

The challenge here is separating out the neutrino candidate TPCObject from any cosmic TPCObjects which passed the cosmic rejection cuts. This is performed by leveraging the light information to do TPCObject-to-optical information matching (or *Flash Matching*, FM).

To perform the FM, each TPCObject is taken, and its x position is modified several times, and at each position the topology of the interaction and geometry of the detector are used to predict the number of PE expected on each PMT. This is compared against the reconstructed number of PE on each PMT, and a likelihood is constructed between the two,

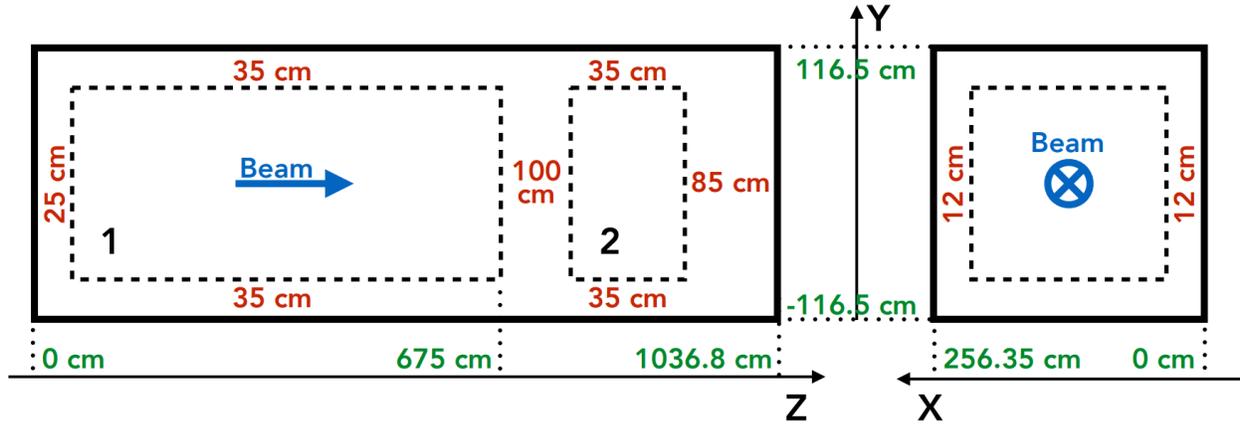


Figure 7.2: Definition of the fiducial volume for the ν_μ CC-inclusive event selection. Figure taken from reference [104]. The left panel here shows a side-on view of the TPC, while the right hand panel shows an end-on view of the TPC.

$$-LL(x) = - \sum_{i=0}^{i=32} \ln(\text{Poisson}(O_i, H_i(x))), \quad (7.5)$$

which is then minimised across the chosen x positions. This is repeated for each TPCObject in the detector and the one which minimises the $-LL(x)$ is chosen as the neutrino candidate.

An additional quality cut is placed at this stage, which demands that the x position chosen by the FM agrees with the assumed x position given by the flash time with some tolerance.

7.3.3 Fiducial Volume

At this stage in the selection, there is a single TPCObject selected inside the cosmic fiducial volume defined in Table 7.2. During this stage of the selection, a tighter fiducial volume is defined, shown in Figure 7.2. This takes into account the known large dead region on the collection plane at around $z = 700$ cm. The large 35 cm fiducial cut in Y is defined such that cosmogenic tracks where the start and end positions have been mis-reconstructed due to the space charge effect are still removed.

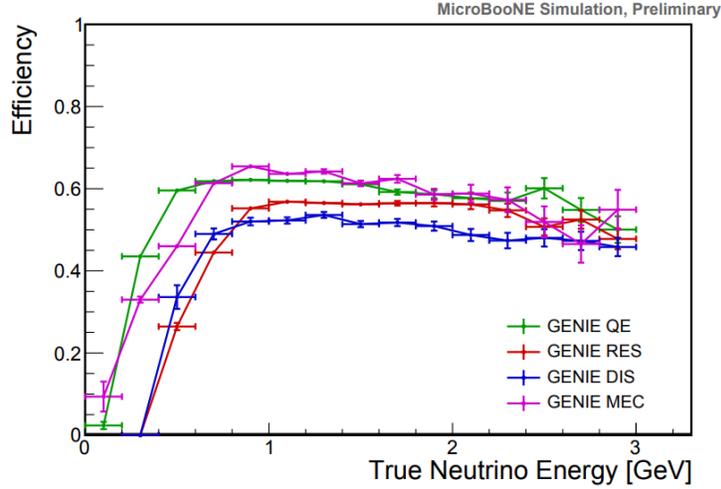


Figure 7.3: The efficiency of selecting ν_μ CC-inclusive events as a function of neutrino energy, separated by interaction type.

7.3.4 Performance of the ν_μ CC-inclusive Selection

At this point it is instructive to quantify the performance of the inclusive part of the selection.

The selection efficiency is defined to be

$$\epsilon = \frac{\text{Selected } \nu_\mu \text{ CC interactions with a vertex in the FV}}{\text{Generated } \nu_\mu \text{ CC interactions with a vertex in the FV}}, \quad (7.6)$$

and the purity is defined to be

$$p = \frac{\text{Selected } \nu_\mu \text{ CC interactions with a vertex in the FV}}{\text{All selected events}}. \quad (7.7)$$

The overall efficiency and purity of the selection are found to be 57.2% and 49.1%, respectively. The efficiency is shown as a function of true neutrino energy, separated by interaction type, in Figure 7.3.

After the CC-inclusive selection has been performed, 8351 on-beam events have been selected from MicroBooNE Run 1 data, with 6523.6 events being predicted from the simulation, and 2708.2 offbeam events, bringing the total predicted events to 9231.8. Of the selected events in simulation, 4531.6 are ν_μ CC-inclusive signal events (3302.9 of which are ν_μ CC $0\pi NP$), and the remaining

selected events are primarily OOFV (670.1), Cosmic (594.5), and Dirt (385.9).

7.4 Selection of ν_μ CC $0\pi Np$ Events

After selecting ν_μ CC interactions, removal of events with pions and identification of muons and protons must take place. In principle, the selection of ν_μ CC $0\pi Np$ interactions requires the identification of the following event characteristics:

- TPCObjects must meet the topological requirement of having at least two reconstructed tracks and zero reconstructed showers.
- A particle identification requirement of having only one muon candidate in order to remove events containing a pion

In practice, each of these steps have complications. Chapter 7.4.1 describes some of the challenges in placing topological cuts on the data, and Chapter 7.4.2 outlines the particle identification leveraged in this analysis.

7.4.1 Topology Cuts

Initial Studies of Reconstructed Tracks and Showers

Initially, investigations into the topological information of the TPC objects were performed in order to develop a selection based on reconstructed tracks and showers, however differences between the MicroBooNE simulation and the MicroBooNE data mean that applying the same reconstruction algorithm to both yields different results. This is displayed succinctly in Figure 7.4. Specifically, it should be noted that the ratio of data to simulated events for the number of reconstructed PFParticles is reasonably flat, whereas the ratios for number of reconstructed tracks and showers are biased low and high respectively, indicating that more showers and fewer tracks are reconstructed in the data than in the simulation.

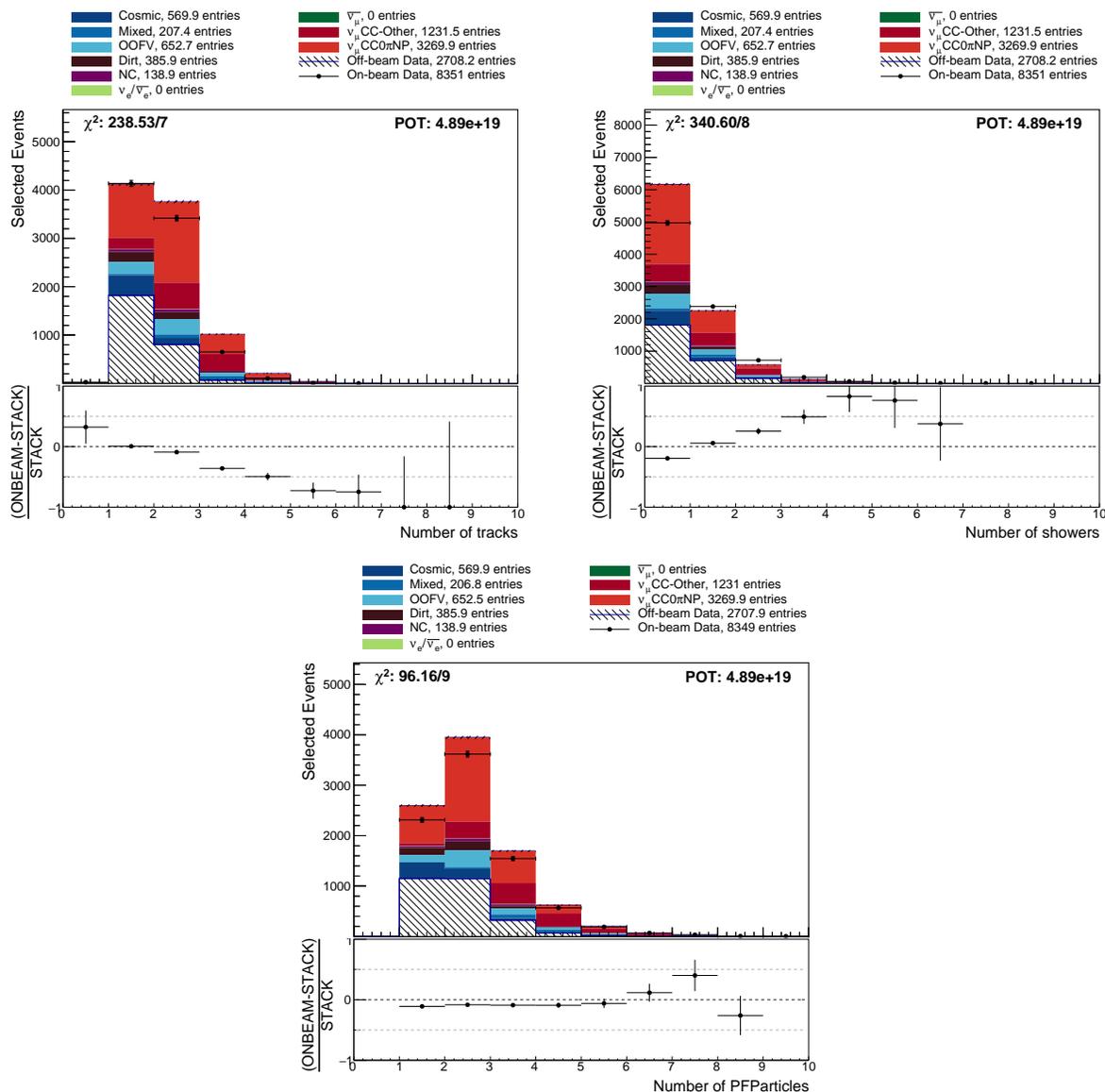


Figure 7.4: Comparison of data and simulation for the number of reconstructed PFParticles, tracks, and showers in a TPCObjects which pass the CC-inclusive selection reconstructed by the Pandora reconstruction algorithm. These plots use the CV dataset, defined above.

While it is feasible that the reason for this difference is due to a mis-modeling in the GENIE generator or an issue with particle propagation in Geant4, it seems more likely that this is due to a missing effect in MicroBooNE’s detector simulation. This is especially true given results from the CC π^0 cross section analysis, which suggests reasonable agreement with GENIE and NuWro [97]. It seems likely that in large part these differences are driven by a lack of induced charge

effects in the nominal simulation. Induced charge effects have been described in Chapters 3.2.2 and 4.5.1.

Including induced charges modifies the shape of the clusters of hits through two primary mechanisms:

- Bipolar induced signals from neighbouring energy depositions can destructively interfere with the primary signal on a selected wire, meaning the signal pulse falls below hit-finding threshold and no hit is created.
- The induced signals on neighbouring wires can act to induce additional hits on a selected wire.

A preliminary implementation of such effects is included in the DIC systematic variation sample described in Section 7.2.4. This can be used as a proxy for the MicroBooNE data, and can be compared against the nominal CV simulation to begin to understand the differences between data and simulation.

To do this, a dataset containing CC-inclusive selected events is used.

The first piece of information which can be gleaned by looking at the DIC sample is that it leads to around 8% fewer events being selected. This is shown in Figure 7.5.

Further to this, the DIC sample can be used to probe how often particles of different species are reconstructed as a track versus reconstructed as a shower in the MicroBooNE data compared with the nominal simulation. This information is presented for muons, protons, pions, and electrons in Figures 7.6, 7.7, 7.8, and 7.9, respectively. These are presented as a function of true kinetic energy and true θ_{xz} , where this defines the angle to the plane¹. Here “true” means the energy and angle of the true particle which has been associated with the reconstructed object (see Appendix B). The blue histogram shows the particles reconstructed in the CV dataset, while the green shows the particles reconstructed when induced charge is included. The grey histogram shows

¹ $\theta_{xz} = 0/\pi$ is forward going, $\theta_{xz} = \pm\pi/2$ is going into or out of the anode plane array.

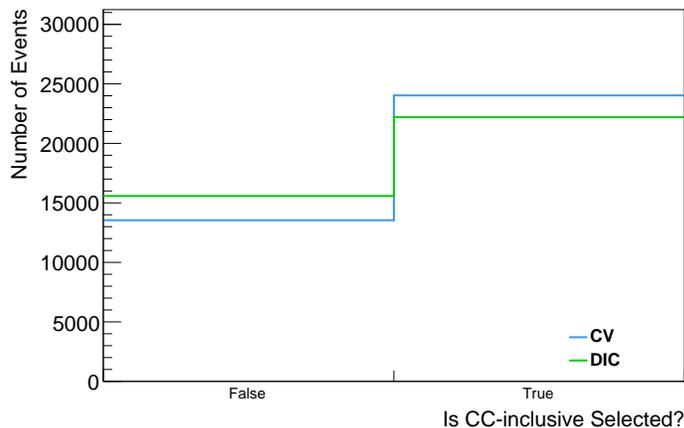


Figure 7.5: Number of events selected by the CC-inclusive selection in the CV dataset and in the DIC dataset

the distribution of the true particles which have been matched to the reconstructed objects, and the bottom plot in each figure shows the ratio of the DIC over the CV dataset.

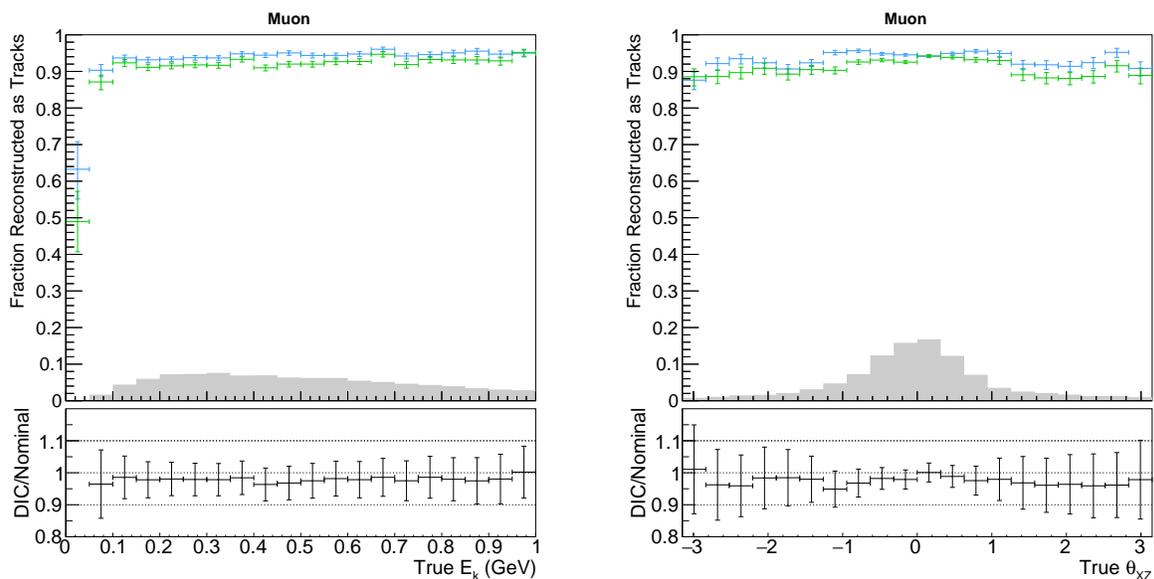


Figure 7.6: Fraction of muons which have been reconstructed as a track according to Pandora using the nominal (blue) and DIC (green) datasets as a function of true kinetic energy (left) and true θ_{xz} (right). Overlaid is the spectrum of energies for reconstructed particles.

While there are differences between the two samples for true electrons and muons, these are generally differences on the order of a few percent. By a significant margin the largest differences

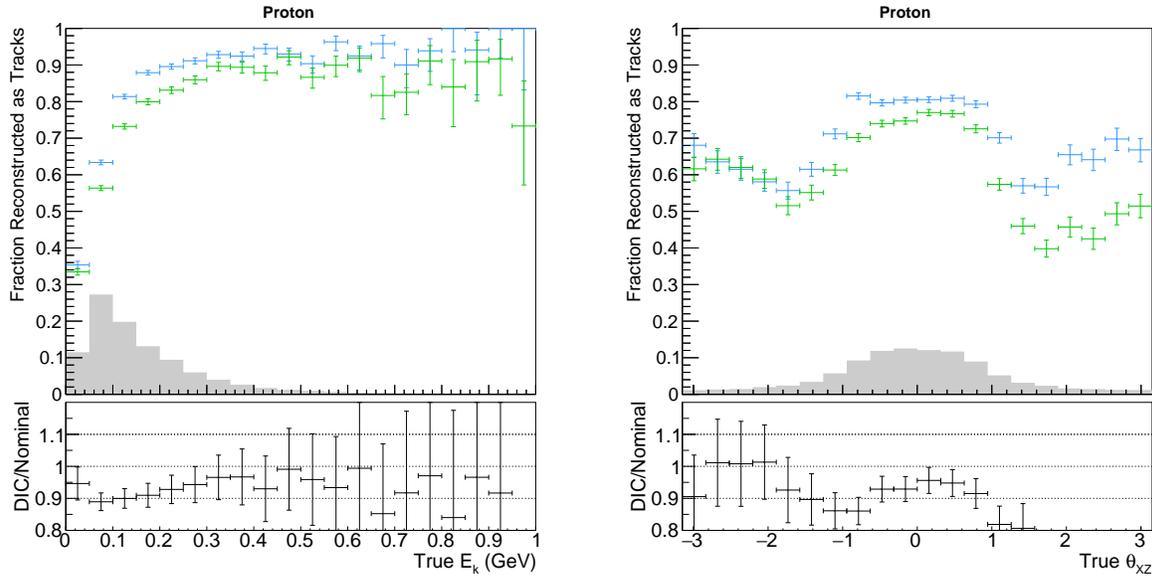


Figure 7.7: Fraction of protons which have been reconstructed as a track according to Pandora using the nominal (blue) and DIC (green) datasets as a function of true kinetic energy (left) and true θ_{xz} (right). Overlaid is the spectrum of energies for reconstructed particles.

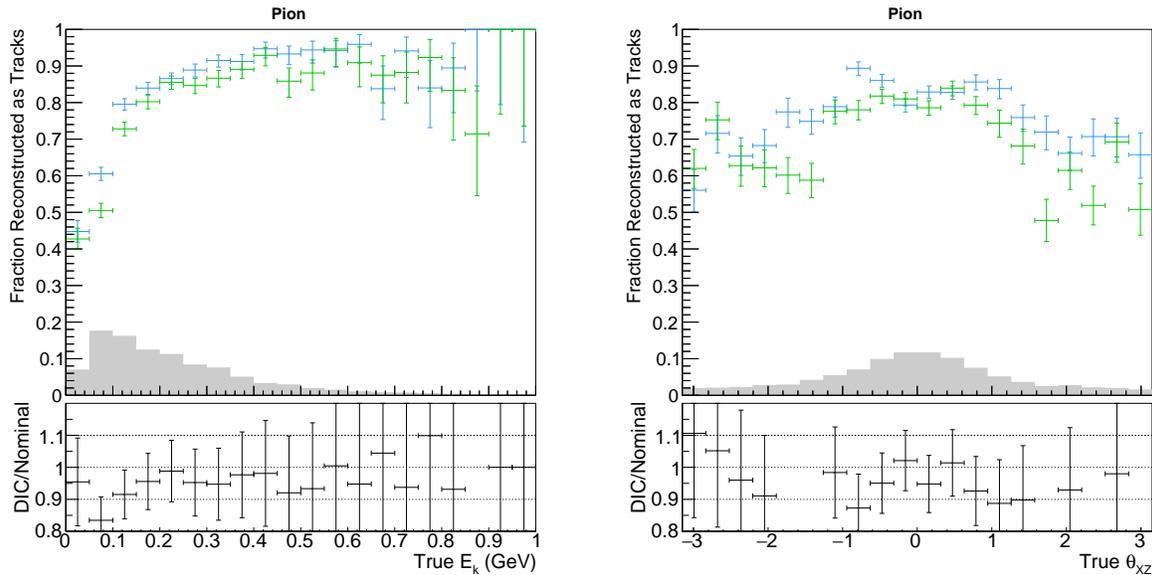


Figure 7.8: Fraction of pions which have been reconstructed as a track according to Pandora using the nominal (blue) and DIC (green) datasets as a function of true kinetic energy (left) and true θ_{xz} (right). Overlaid is the spectrum of energies for reconstructed particles.

are for the true protons and pions, where both show an approximately 10% difference between nominal and DIC in the bin where most of the reconstructed particles lay in E_k . It is also obvious

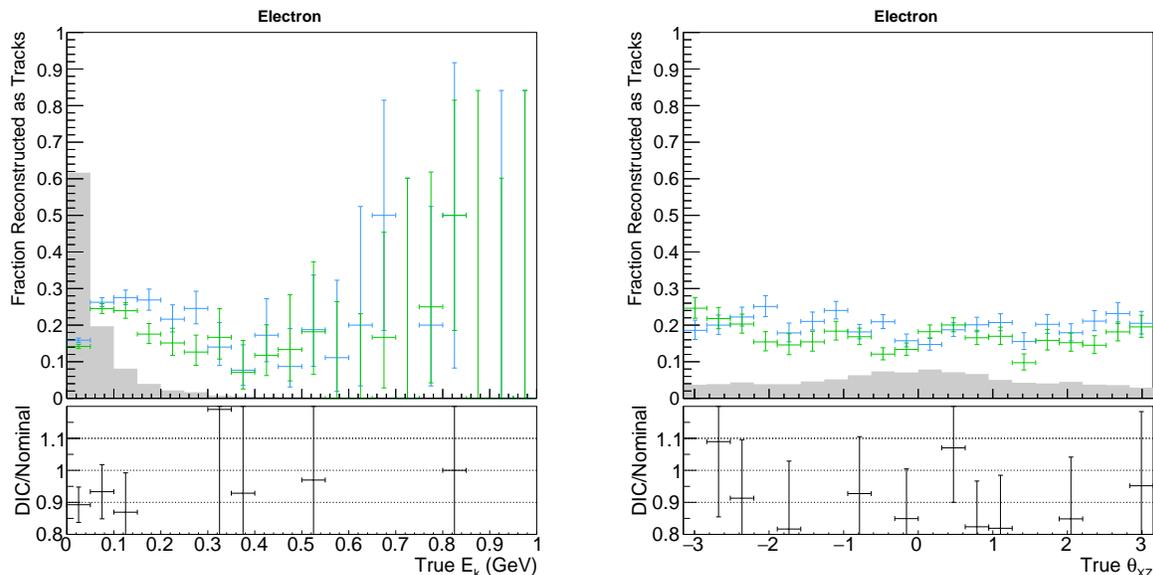


Figure 7.9: Fraction of electrons which have been reconstructed as a track according to Pandora using the nominal (blue) and DIC (green) datasets as a function of true kinetic energy (left) and true θ_{xz} (right). Overlaid is the spectrum of energies for reconstructed particles.

that this is not only a function of energy, where lower energy (and therefore shorter) protons and pions (Figures 7.7 and 7.8) are expected to be reconstructed as a shower more often in the DIC sample, but also of angle. As might be expected from the effects of induced charge, as tracks approach high angles with respect to the plane, the reconstruction becomes worse, and particles are more likely to be reconstructed as showers.

These studies indicate that while it is not currently feasible to develop a selection based on reconstructed tracks and showers, developing a selection based on PFParticles and applying particle identification to each PFParticle in order to separate muons from protons is a viable path forward.

Mitigation Through Fitting Showers As Tracks

In order to leverage the particle identification method outlined in Chapter 6, each PFParticle is required to have a valid set of dE/dx and residual ranges². This poses a problem for PFParticles which have been reconstructed as showers, as a shower residual range is ill-defined.

²Distance to the end of the reconstructed object

In order to proceed, Pandora is asked to build every shower as both a track and a shower object, meaning every PFParticle now has an associated track, which has a well-defined residual range.

PID on Showers Built as Tracks

After building all PFParticles as tracks, the showers-built-as-tracks may be processed through the PID, and the $\ln(L_{MIP}/L_p)$ variable can be constructed in order to select out protons which were built as showers against electrons. The results of doing this can be found in Figure 7.10.

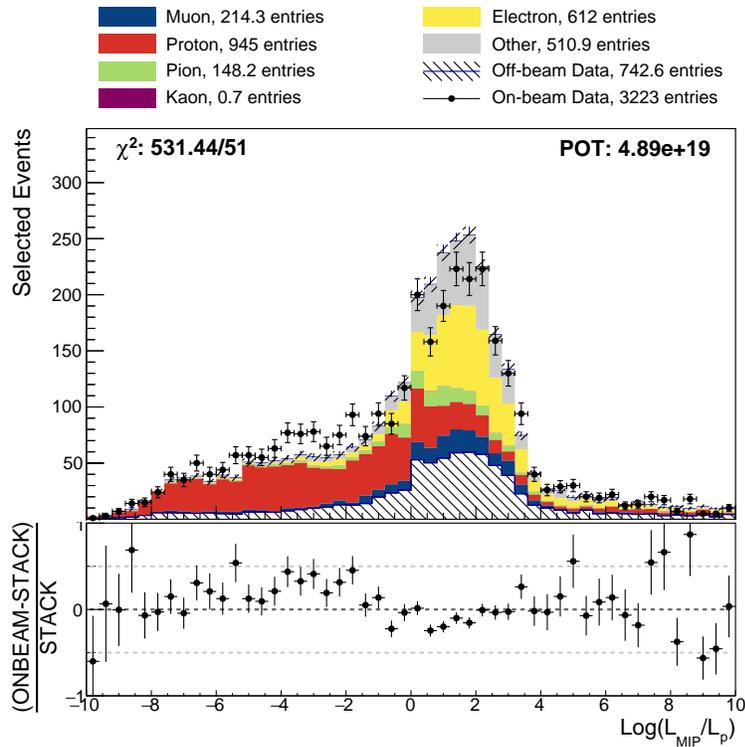


Figure 7.10: PID variable for showers which have been rebuilt as tracks.

The electrons mostly pile up at values greater than zero, meaning that they are more MIP-like than proton-like, while in general protons are located at values less than zero, indicating that they are more proton-like than MIP-like. There is a significant number of protons which also fall within the MIP peak. This is because these protons are at a high angle to the collection plane where the charge reconstruction becomes difficult.

The excess which is present in the data at values in the proton region is consistent with the earlier assertion that more protons are reconstructed as showers in the data than in the nominal simulation.

Including both the tracks and the showers-built-as-tracks, we can reproduce the full PID distribution, and this is shown in Figure 7.11. When including these additional particles which were reconstructed as showers there is a slight deterioration in the χ^2/NDF , however this is minor. Some amount of disagreement remains in the proton peak, however this is largely a normalisation difference, and is thought to be related to induced charge effects meaning fewer protons are reconstructed in the data than in the simulation.

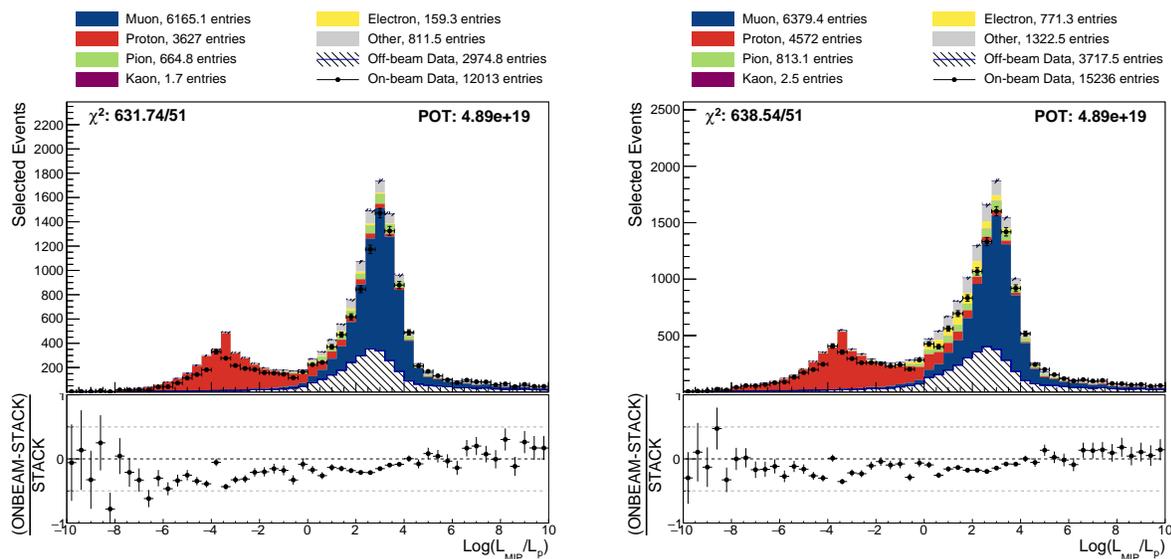


Figure 7.11: Distributions of the MIP/proton separator using only tracks identified as such by Pandora (left) and including in addition the showers built as tracks (right).

Application of Topological Cuts

This section has outlined a method to minimise the effects of differences between data and simulation due to reconstruction effects.

After application of the topological demand that each interaction have at least two PFParticles, the makeup of events is as follows: 6039 on-beam data events are selected compared to a

predicted 6633.8 events, of which 2571.4 events are signal, 1558.1 are off-beam backgrounds, and the remaining backgrounds are predicted from simulation. At this stage, the selection efficiency is $\sim 46.5\%$ with a selection purity of $\sim 38.8\%$. At this stage, the largest backgrounds are from the off-beam data (1558.1 events), CC-other interactions (1162.1 events), and OOFV interactions (517.0 events).

It might be expected that both off-beam data and OOFV interactions are selected, because if they decay, the muon and Michel electron meet the 2-PFParticle criteria. Both these backgrounds and CC-other backgrounds may be targeted by applying particle identification.

7.4.2 Particle Identification

This analysis leverages the Bragg Likelihood algorithm in order to separate out MIP-like particles from protons. This is discussed extensively in Chapter 6, where the performance and robustness of the available algorithms is investigated.

Specifically, this analysis makes use of the $\ln(L_{MIP}/L_p)$ variable, which was shown to have a high selection efficiency and purity, while being relatively robust against systematic variations and having a good level of agreement between the data and simulation.

In order to select out the chosen signal, the demand is made that exactly one track be a muon candidate and the remaining selected tracks be consistent with being a proton. This is done in practice by cutting at -1 in the separator variable displayed in Figure 7.11, and defining any particle which falls to the left of this cut as a proton candidate, and any particle which falls to the right of this particle as a muon candidate.

After application of the particle identification cut, the makeup of events is as follows: The number of on-beam data events selected is 1595, compared to a prediction of 1775.2. Of the predicted events, 1196.3 are signal, meaning the purity of the selection at this stage is 67%, while the efficiency is around 20.5%. The largest backgrounds at this stage of the selection are from off-beam data (292.4 events), followed by OOFV (87.7 events) and ν_μ CC-other (85 events).

It was noted that after building showers as tracks, some events are misreconstructed, and so

now it is important to place quality cuts on the selected interactions.

7.4.3 Quality Cuts

The final cuts, performed after the particle identification, are related to the quality of the reconstructed tracks.

This comes down to three demands:

- All proton candidates must be fully contained within the FV defined in Figure 7.2. This is done because protons in general are very short and travel in straight lines, meaning that it is difficult to measure their energy from multiple Coulomb scattering. This means that the only energy estimators for these particles are their length and the energy they deposit in the TPC, both of which require containment.
- The energy of each *contained* reconstructed particle from its range is similar to that deposited in the TPC. This is applied to remove tracks which are misreconstructed by, for instance, the inclusion of hits which are reconstructed from noise.
- Contained muon candidates must have an energy reconstructed from MCS within 20% of the energy reconstructed from the range. This is a relatively loose cut which ensures that the three energy estimators available for contained muons are consistent with each other.

For the second and third cuts here, the energy from range is calculated under the assumption that the particle is a muon for muon candidates, and under the proton assumption for proton candidates. Plots of the distributions of this variable can be found in Figure 7.12.

There are two features in these distributions: firstly, there is a slight offset from zero in both distributions indicating that the energy from range is slightly higher than the energy deposited in the TPC. This is not fully understood, but has been observed in several MicroBooNE analyses. Secondly, it is clear from these plots that there is a some disagreement in the peak regions in both shape and normalisation. For this reason, reasonably loose cuts are placed on these distributions of $-0.15 \text{ GeV} < E_{dep}^Y - E_{range} < 0.1 \text{ GeV}$.

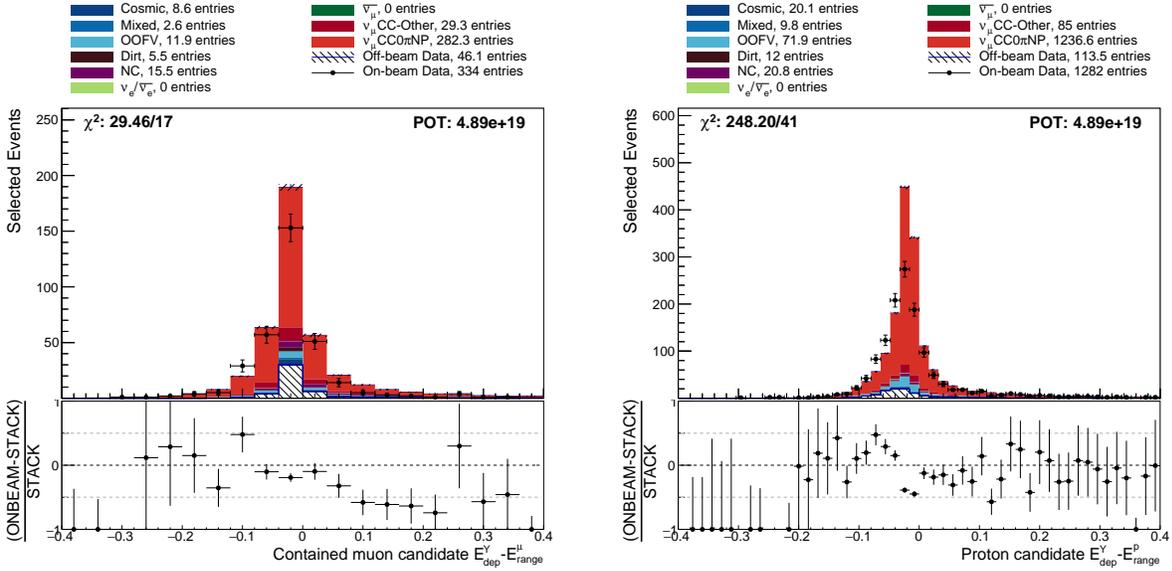


Figure 7.12: $E_{dep}^Y - E_{range}$ for selected contained muon candidates and proton candidates. Note that only contained muons are present in the plot on the left, while all selected protons are present in the plot on the right.

7.4.4 Selection Performance

After application of the full selection chain, the PID variable, $\log(L_{MIP}/L_p)$ can be plotted for selected events, as shown in Figure 7.13. Here, any entries to the left of -1 have been identified as a proton candidate, and any entries to the right of -1 have been identified as muon candidates. There is clearly very little contamination from muons in the proton peak, and low contamination from protons in the muon peak. The shape agreement between the data and simulation is reasonable overall, although there is a normalisation difference in the nominal simulation which can be observed to come from the lack of induced charge effects.

In order to provide further confidence that the selection is acting as intended, the dE/dx versus residual range at the end of the track for contained and uncontained muon candidates, and leading and non-leading proton candidates can be plotted. This is shown in Figure 7.14 for the simulation, and Figure 7.15 and 7.16 for on-beam and off-beam data, respectively.

It is clear from these plots, that the selected proton candidates are more highly ionising than the muon candidates and so there can be some confidence that the PID algorithm is acting as it

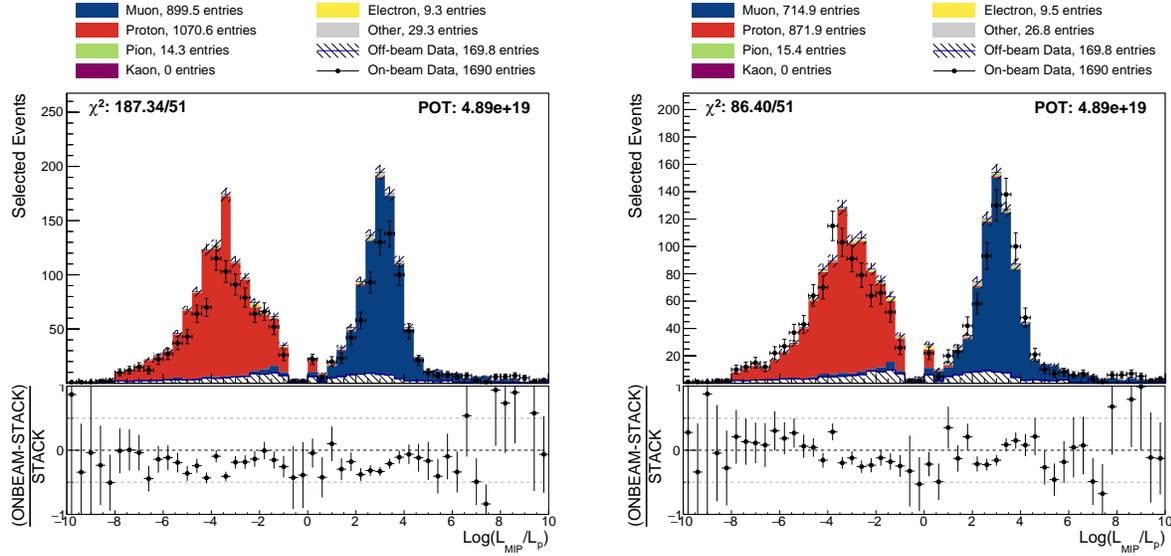


Figure 7.13: Distribution of the PID variable $\log(L_{MIP}/L_p)$ after the full selection chain on the CV and DIC simulated samples. Those entries to the left of -1 are identified as proton candidates, while those to the right have been identified as muon candidates.

should.

Specifically of interest are the off-beam distributions presented in Figure 7.16. It might be natural to ask why there is still a relatively significant contribution of cosmic particles passing the selection after the demand of one muon candidate plus at least one proton candidate, and this plot indicates that the proton candidates which are identified in this sample are indeed consistent with being protons, and this is not some feature of the PID algorithm.

As for the case of the CC-inclusive selection outlined above, the efficiency and purity of this selection can be defined as

$$\epsilon = \frac{\text{Selected } \nu_\mu \text{ CC } 0\pi Np \text{ interactions with a vertex in the FV}}{\text{Generated } \nu_\mu \text{ CC } 0\pi Np \text{ interactions with a vertex in the FV}}, \quad (7.8)$$

and

$$p = \frac{\text{Selected } \nu_\mu \text{ CC } 0\pi Np \text{ interactions with a vertex in the FV}}{\text{All selected events}}, \quad (7.9)$$

respectively

The efficiency of the selection is 14.2%, and the the purity is 79%. The efficiency as a function of

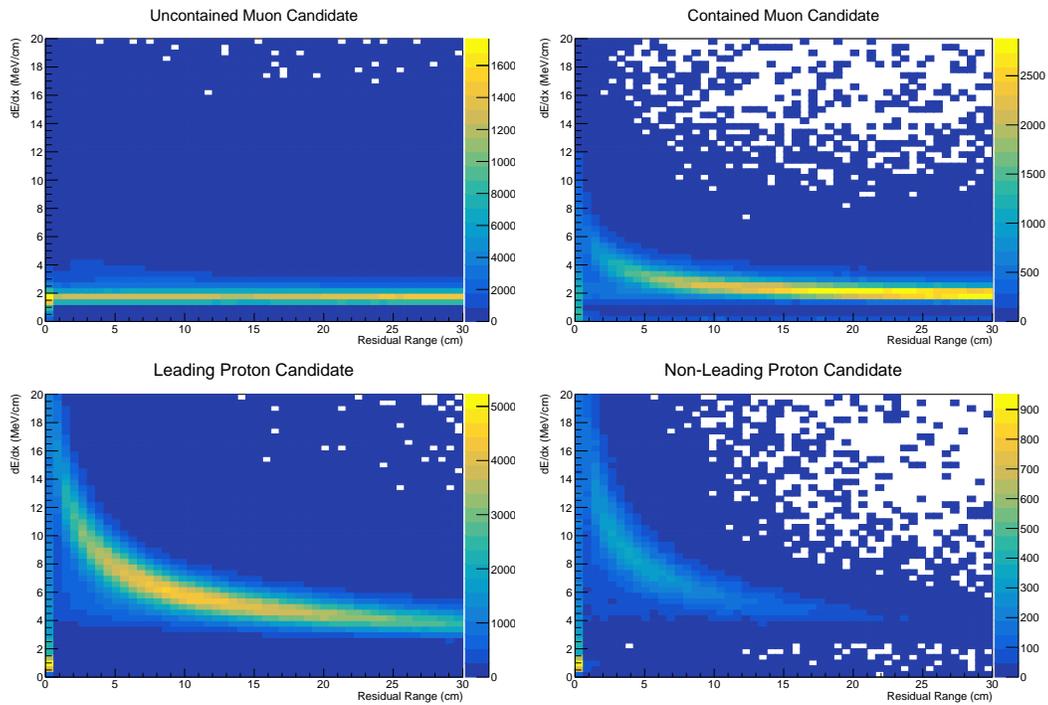


Figure 7.14: dE/dx versus residual range for uncontained and contained muons, and for leading and non-leading protons from simulation.

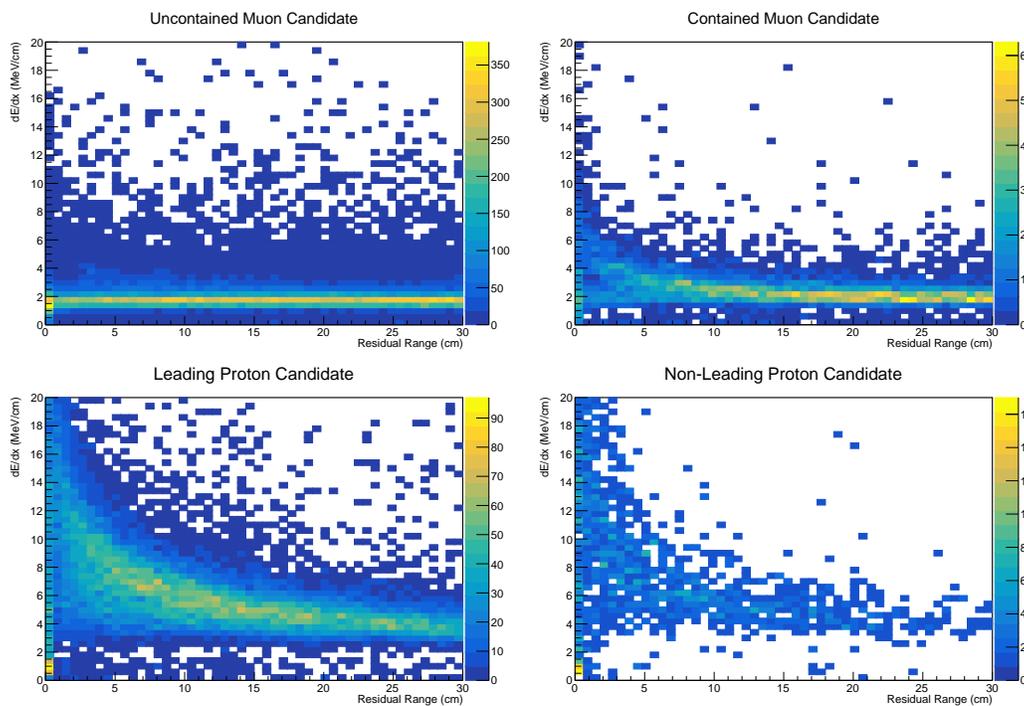


Figure 7.15: dE/dx versus residual range for uncontained and contained muons, and for leading and non-leading protons from on-beam data.

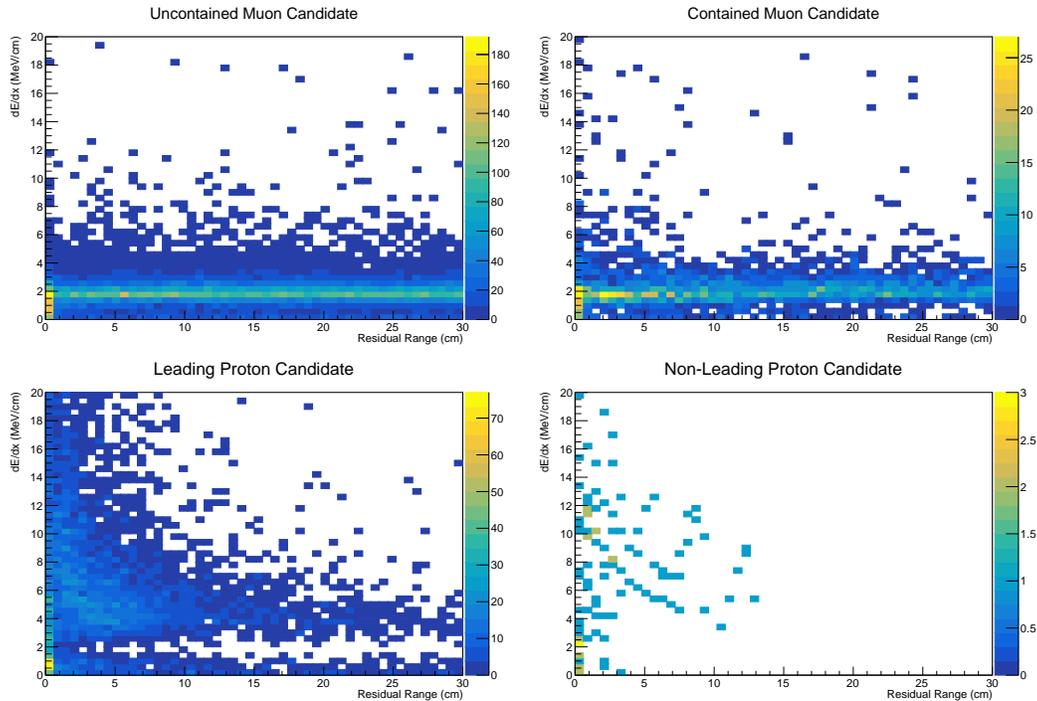


Figure 7.16: dE/dx versus residual range for uncontained and contained muons, and for leading and non-leading protons from off-beam data.

the true neutrino energy and of the true muon momentum are shown in Figure 7.17. Of particular interest for this analysis is the efficiency at low neutrino energy.

After the selection has been applied, a total of 804 on-beam data events are selected, where 89.8 off-beam events are expected, and 941.9 simulated events are predicted, bringing the total expectation to 1031.7 events. The largest backgrounds are from offbeam (89.8 events) and OOFV (52.1 events) events. A cut flow table is presented in Table 7.3.

Quantifying Effects of Building Showers as Tracks

This section has so far demonstrated that application of topological cuts using reconstructed tracks and showers is not reliable, and that this introduces differences between the data and simulation.

A method has been devised in order to minimise this effect using the loose topological requirement of demanding only two PFParticles, but demanding each PFParticle pass a PID cut. It

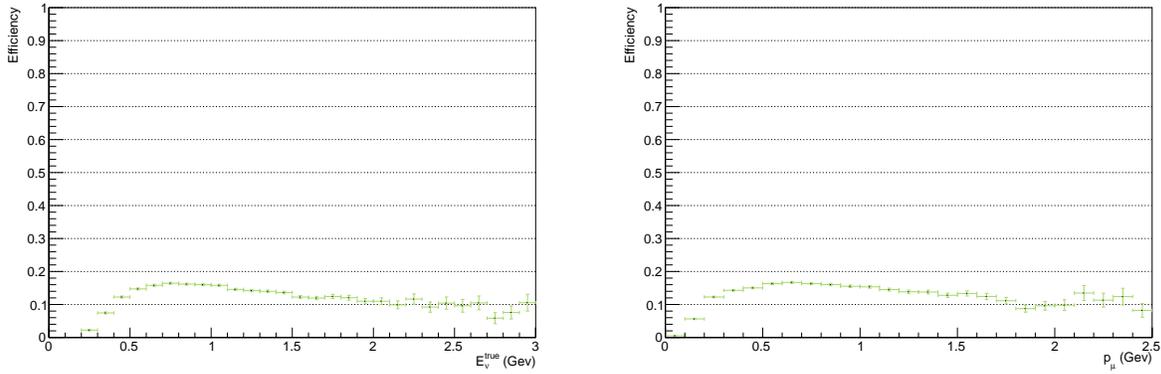


Figure 7.17: Efficiency of the selection as a function of E_ν^{true} and p_μ^{true} .

Cut	Selected On-Beam Events	Predicted Signal	Predicted Background	Efficiency	Purity
CC-inclusive	8351	3302.9	5928.9	60%	35.8%
Topological	6039	2571.4	4062.4	46.5%	38.8%
PID	1595	1196.3	578.9	20.5%	67%
Quality Cuts	804	813.4	218.3	14.2%	79%

Table 7.3: Cut flow table, showing the number of selected on-beam events compared with the predicted number of events, as well as the selection efficiency and purity at each stage.

is now interesting to quantify what magnitude the uncertainty due to DIC can be reduced by if the proposed method of building showers as tracks is followed.

To benchmark the performance of the process of building showers as tracks, two sets of cuts can be defined:

1. **Tight Topology Cuts** Exactly 0 reconstructed showers (as identified by Pandora) and ≥ 2 reconstructed tracks are demanded. Only one candidate track must be consistent with being MIP-like ($\log(L_{MIP}/L_p) \geq -1$), and all other tracks must be consistent with being a proton $\log(L_{MIP}/L_p) < -1$.

2. **Loose Topology Cuts** Demand only ≥ 2 reconstructed PFParticles, but reconstruct every particle as a track, and subject them to a PID cut, making the same demand as above.

This study is performed in only a single variable, the candidate muon ϕ , because it is especially sensitive to induced charge effects. Further distributions can be found in Appendix B.

Results using the tight topology cuts can be found in Figure 7.18. It is clear here that the DIC simulation drastically improves the agreement between data and simulation, but this is at a price of a large uncertainty on the selection efficiency. The ϵ/p for the nominal simulation is $\sim 12.4\% / \sim 84.0\%$ while for the DIC simulation it is $\sim 8.9\% / \sim 80.3\%$, which represents a $\sim 28.2\%$ uncertainty on the efficiency.

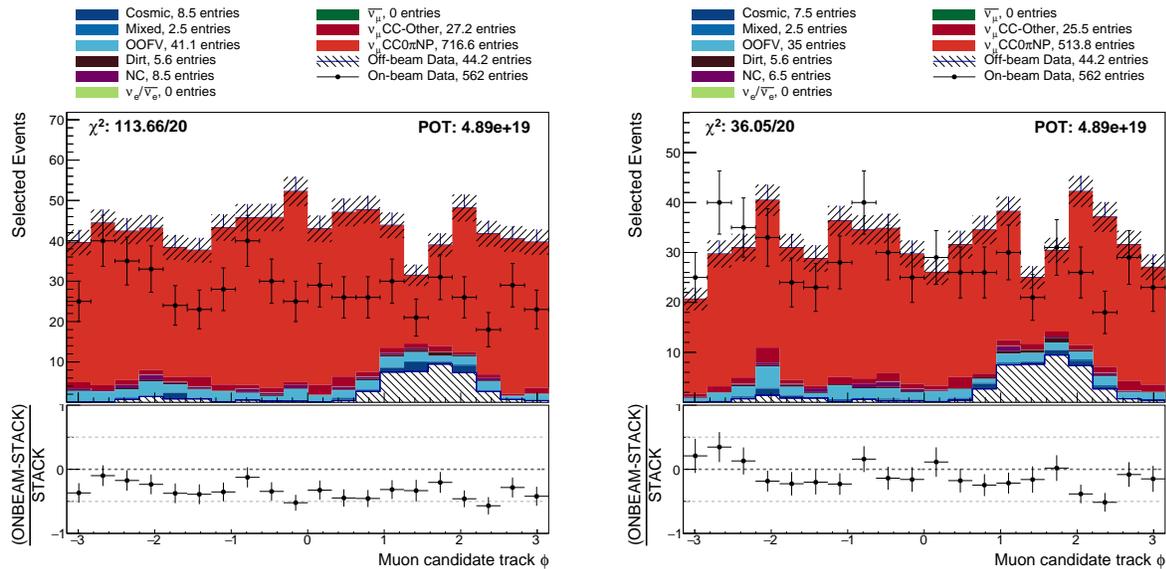


Figure 7.18: Candidate muon track ϕ distributions for the nominal simulation (left) and the DIC simulation (right) using the tight topology cuts.

Results using the loose topology cuts can be found in Figure 7.19. It is immediately obvious here that the agreement between data and simulation in for the nominal sample is significantly improved over the tight topology cuts. The ϵ/p for the nominal sample is $\sim 14.2\% / \sim 79.0\%$ and for the DIC simulation is $\sim 11.1\% / \sim 75.1\%$, which represents a 21.8% uncertainty on the efficiency.

To summarise: by implementing loose topological cuts, the data-to-simulation agreement in the nominal sample can be improved, and the magnitude of the systematic uncertainty due to

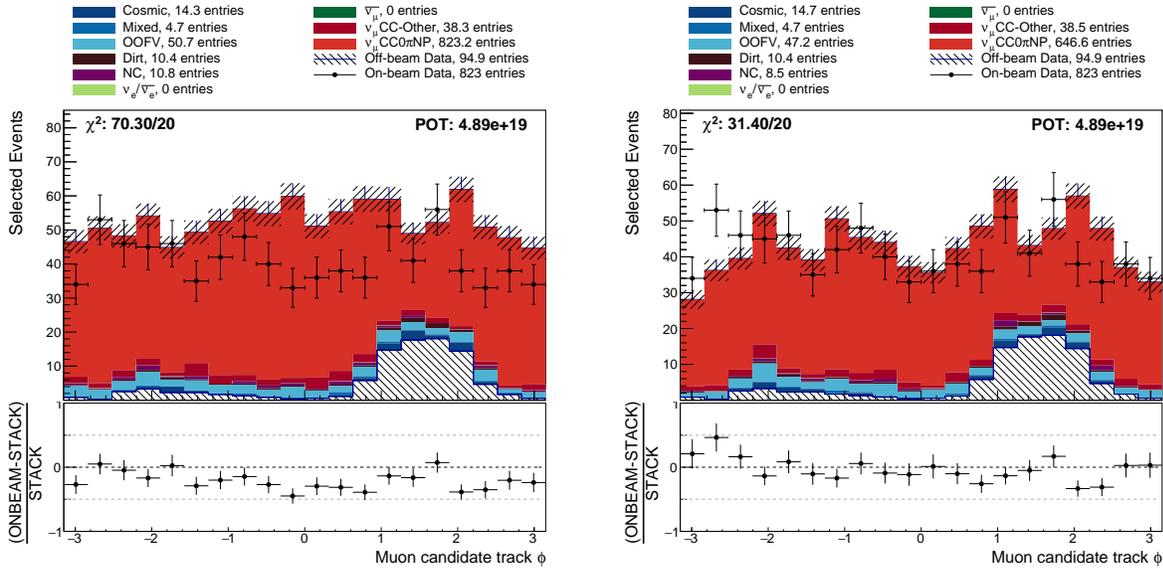


Figure 7.19: Candidate muon track ϕ distributions for the nominal simulation (left) and the DIC simulation (right) using the loose topology cuts.

induced charge effects can be reduced by 22.7%.

One trade-off to the application of the loose topological cuts is that the overall purity is reduced by around $\sim 6\%$, however this appears to mostly be due to an increase in the off-beam data, which can be measured and removed, and so this is thought to be a reasonable trade-off.

Despite this improvement, the overall shape and normalisation agreement between the nominal simulation and the data is still imperfect due to a lack of induced charge effects in the simulation.

7.5 Distributions of Selected Events

Figure 7.20 shows the distribution of reconstructed neutrino vertex positions in the x , y , and z dimensions within the MicroBooNE TPC. From these distributions it is clear that there remains a normalisation offset between the data and simulation, although there appears to be very little in the way of shape difference.

There is a large gap around 700 cm in the vertex z position plot is due to a known dead region

on the collection plane, leading the fiducial volume to neglect this region, as shown in Figure 7.2.

The shape of the OOFV backgrounds here is somewhat nonintuitive with events generally laying at large x positions and at the edges of the distribution in y . This is mostly due to the space charge effect. Particles which emanate from interactions which occur at larger x positions are subject to a higher level of spatial distortion due to an increased amount of time traversing the non-linear electric field caused by space charge. In this way, interactions which truly take place outside of the fiducial volume can be brought inside the volume more often.

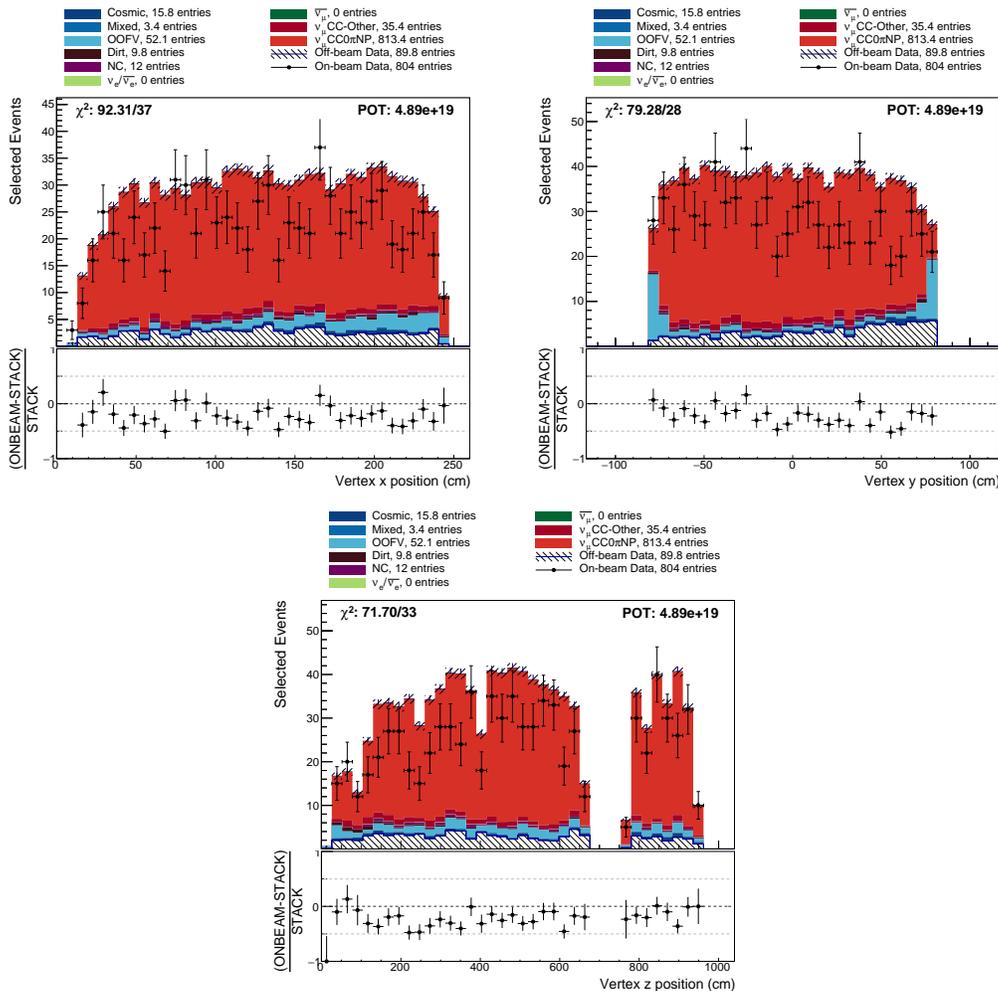


Figure 7.20: Reconstructed neutrino vertex positions in the x , y and z dimensions within the MicroBooNE TPC.

In addition to these plots, distributions of the muon and proton candidate $\cos(\theta)$, ϕ , and energy estimators are shown in Figures 7.21 and 7.22.

The downturn in the data for very forward going tracks in the muon candidate $\cos(\theta)$ plot is similar to that observed by the CC-inclusive analysis [104], however it appears more pronounced. This is thought to be for two reasons. The first is that there is a coherent noise filter run on the MicroBooNE data which is not applied to the simulation. This filter aims to remove noise which is correlated across several channels, and so parts of very forward going particles might be removed, impacting the reconstruction. This would only impact the bin at $\cos(\theta) = 1$. The second reason is due to effects which are not included in MicroBooNE's default tune of GENIE, such as Random Phase Approximation corrections.

There is also some level of disagreement in the muon candidate ϕ plot which is due to the lack of induced charge effects in the simulation. Aside from these effects, and the overall normalisation difference, the agreement between data and simulation is reasonably good.

The same holds true of the proton candidate distributions shown in Figure 7.22, where there is generally an overall normalisation disagreement and a disagreement in ϕ , but otherwise the ratio of the data to simulation is relatively flat.

7.6 Neutrino Energy Reconstruction

After application of the selection outlined in this chapter, a neutrino energy estimator is constructed. The energy estimator which is chosen for this analysis is the sum of the energies of the particles reconstructed as coming from the interaction.

The energy of contained muon and proton candidates are reconstructed by using the reconstructed track range, while the energy of those muons which are uncontained are reconstructed using the momentum obtained from multiple Coulomb scattering.

For each true muon and proton passing the selection, the variable $(E_k^{true} - E_k^{reco})/E_k^{true}$ can be constructed. These values can be binned depending on their E_k^{true} in order to estimate the resolution for each species as a function of their true kinetic energy. This is done separately for uncontained and contained muons, and for protons, and a functional form similar to that of the

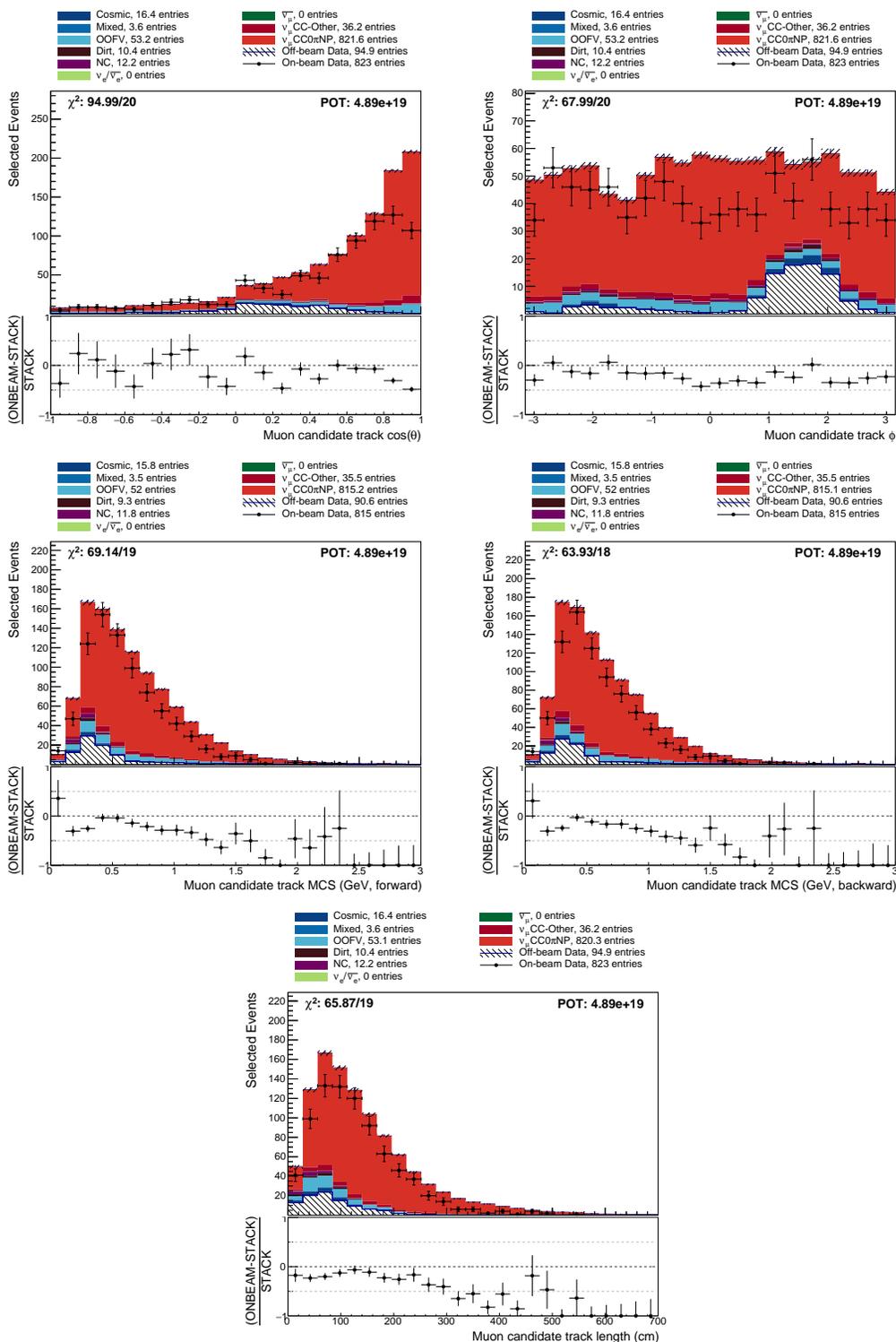


Figure 7.21: Kinematic distributions for the selected muon candidate.

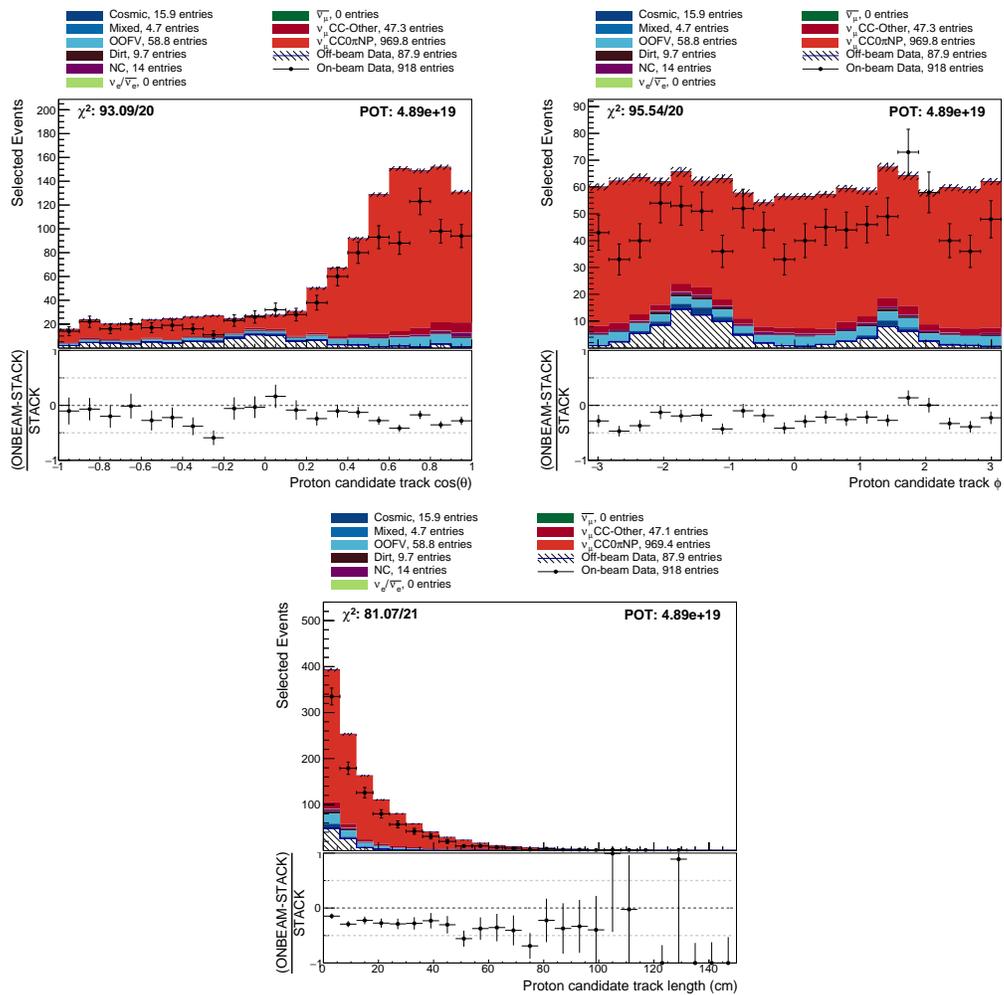


Figure 7.22: Kinematic distributions for the selected proton candidates.

Crystal Ball function (developed in reference [107]) is fit in each of these bins,

$$f(x; \bar{x}, \sigma, k) = e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2}, \text{ for } \frac{x-\bar{x}}{\sigma} \geq -k$$

$$= e^{-\frac{k^2}{2}} + k\left(\frac{x-\bar{x}}{\sigma}\right), \text{ for } \frac{x-\bar{x}}{\sigma} < -k. \quad (7.10)$$

Here, the Gaussian width is taken to be the resolution in each bin. The results of these fits can be found in Figures 7.23, 7.25 and 7.27.

The resolution as a function of E_k for uncontained and contained muons, and for protons, can be found in Figures 7.24a, 7.26a, and 7.28a respectively. These plots show that for uncontained muons, where the energy is estimated from the magnitude of the multiple Coulomb scattering [106], the resolution hovers around 15%, but for contained muons and protons, where the energy can be estimated from the range the energy resolution is generally sub-5%.

In addition to this, the bias of each particle species has been found and is presented in Figures 7.24b, 7.26b and 7.28b. Here, E_k^{true} is plotted against E_k^{reco} , and for each bin in E_k^{true} the modal value of E_k^{reco} is found. A least-squares fit is then performed in order to find the bias.

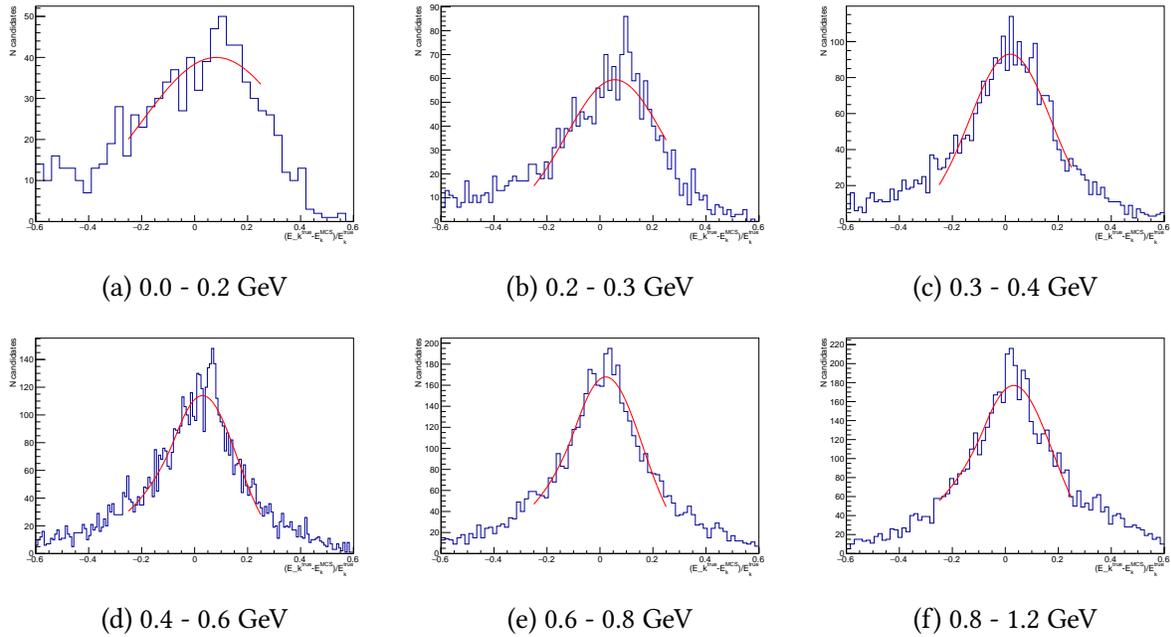


Figure 7.23: Fits for energy resolution of uncontained muons as a function of energy.

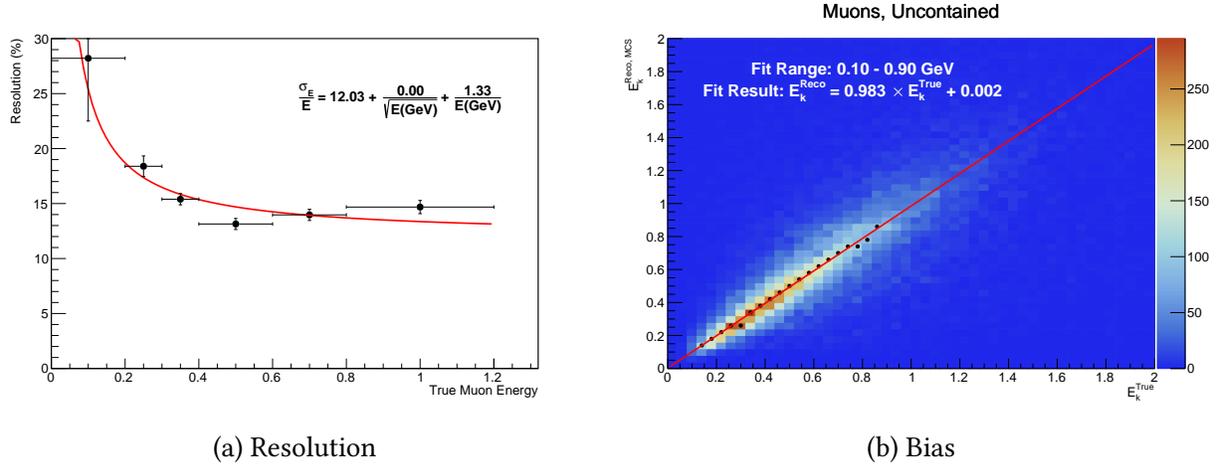


Figure 7.24: Resolution (left) and bias (right) for uncontained muons

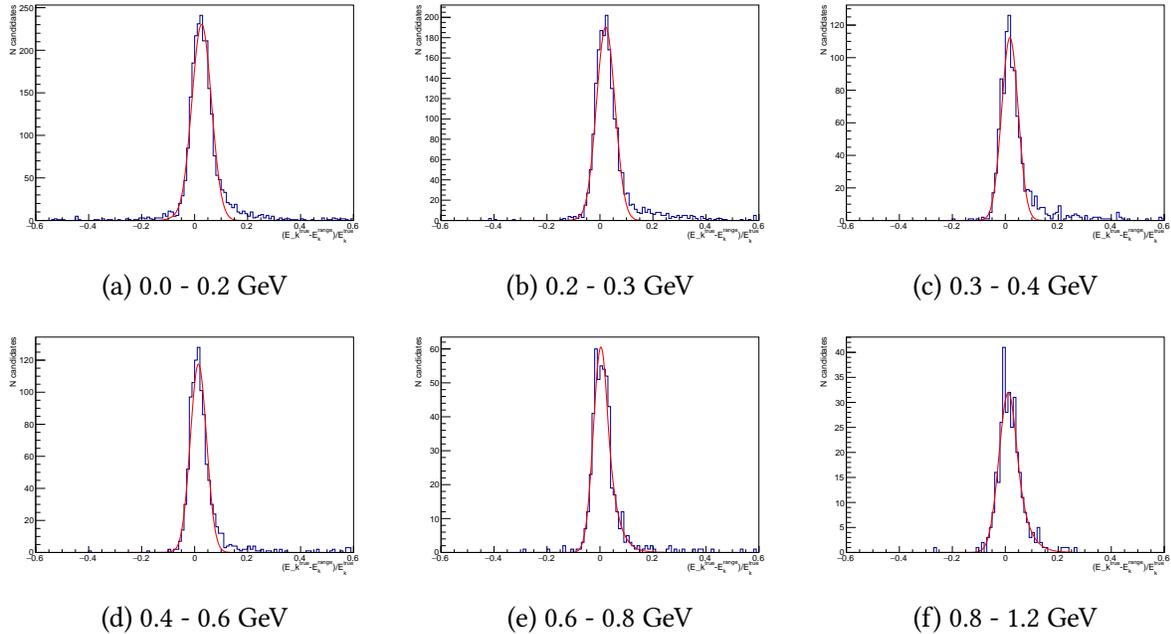


Figure 7.25: Fits for energy resolution of contained muons as a function of energy.

From the fits to the bias plots a method can be devised for each particle species to translate between the true and reconstructed energies. These are shown in Equations 7.11, 7.12m and 7.13.

$$E_k^{\text{true}} = \frac{E_k^{\text{reco}, \mu, \text{uncontained}} - 0.002}{0.983}, \quad (7.11)$$

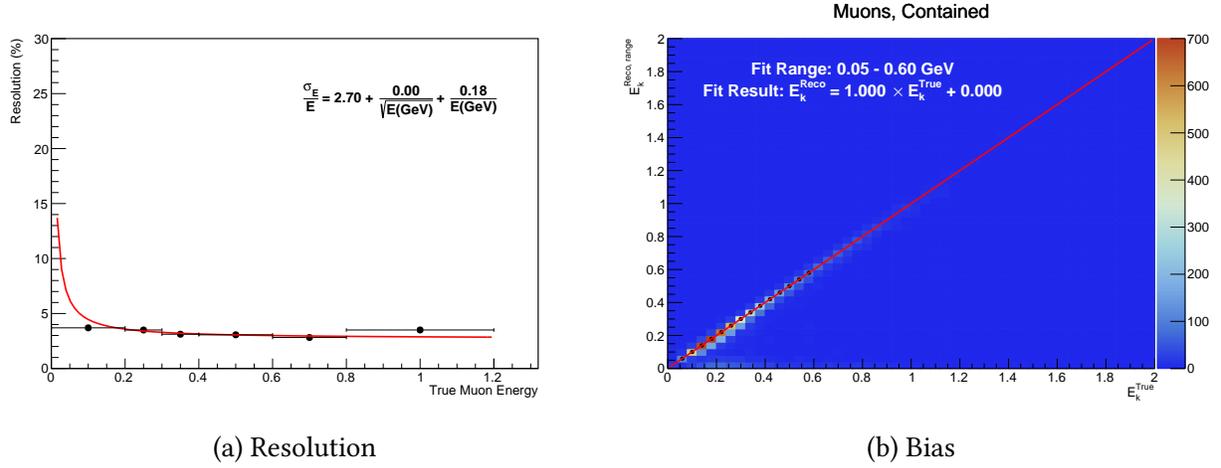


Figure 7.26: Resolution (left) and bias (right) for contained muons

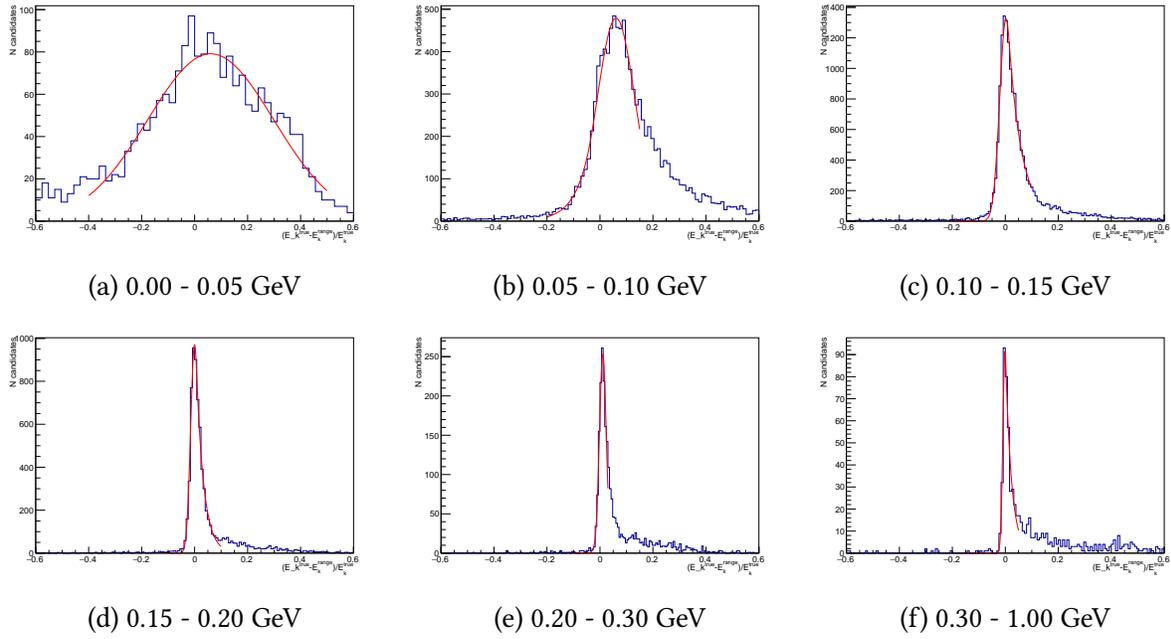


Figure 7.27: Fits for energy resolution of protons as a function of energy.

$$E_k^{\text{true}} = E_k^{\text{reco}, \mu, \text{ contained}}, \quad (7.12)$$

$$E_k^{\text{true}} = \frac{E_k^{\text{reco}, p} - 0.007}{1.030}. \quad (7.13)$$

Using these equations and the reconstructed kinetic energies, the neutrino energy estima-

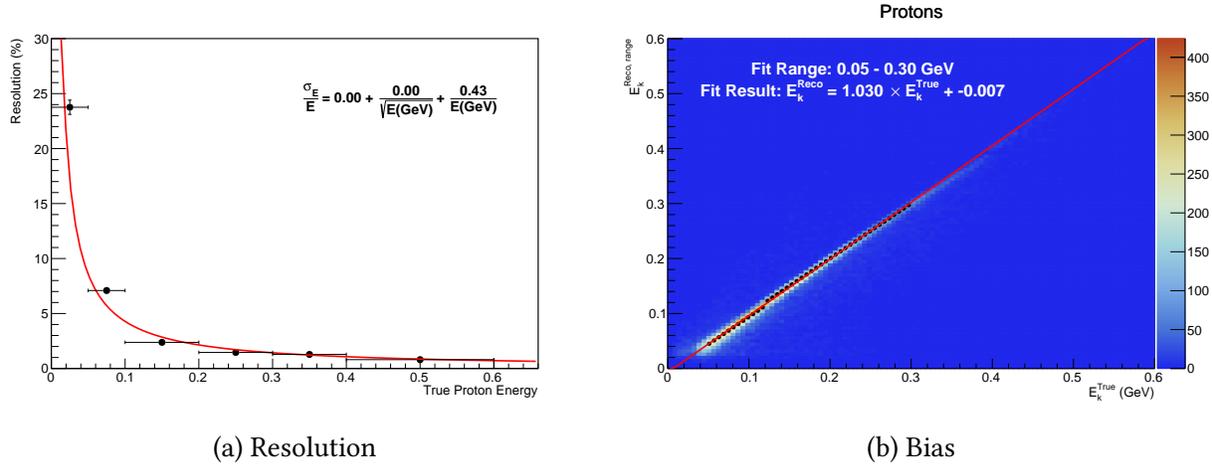


Figure 7.28: Resolution (left) and bias (right) for protons

tor can finally be constructed. Plots of this variable can be found in Figure 7.29 for both the uncalibrated and the calibrated reconstructed energies.

A comparison of the true neutrino energy compared against this neutrino energy estimator can be found in Figure 7.30. It is clear that there is a systematic bias to lower reconstructed energies. This is expected as only charged particles which are reconstructed are able to contribute energy to the reconstructed neutrino energy. Neutral particles and those which are too short to be reconstructed therefore do not contribute to the reconstructed neutrino energy.

7.7 Evaluation of Systematic Uncertainties

Systematic uncertainties for MicroBooNE analyses are generally dealt with in two ways: through *unisims*, or *multisims*.

Unisims refer to a process by which underlying model parameters are varied by $\pm 1\sigma$ of the nominal value (or the effect is turned off in order to get a lower bound on its effect), and the events are re-simulated. This is performed for detector-related systematic uncertainties, as described in Chapter 7.2.

Multisims are used for flux and interaction uncertainties. In these cases, the underlying models have parameters with associated uncertainties which are assumed to be distributed in a Gaus-

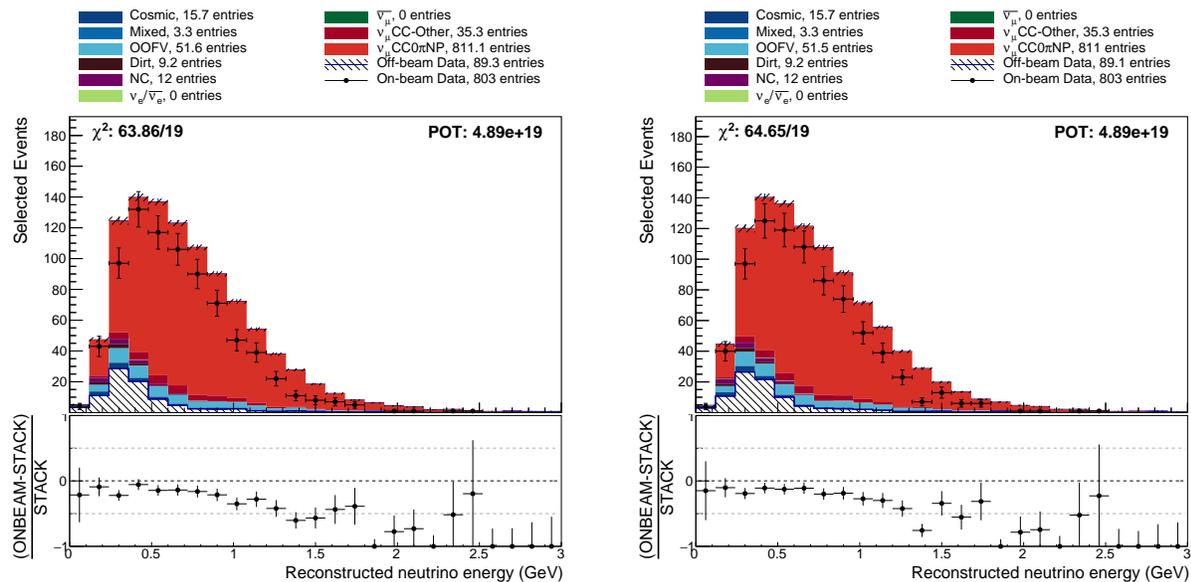


Figure 7.29: Total deposited energy in the TPC before (left) and after (right) calibrating.

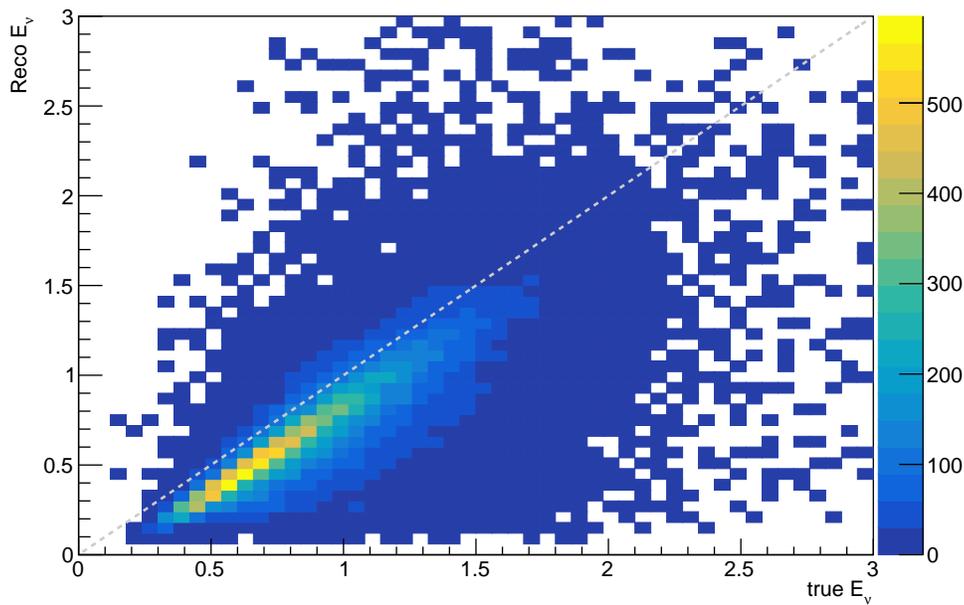


Figure 7.30: True neutrino energy compared with the energy estimator used in this analysis.

sian way. Given these uncertainties, a new value for each parameter can be chosen by randomly sampling the Gaussian function. This random throw results in a new *weight* for each neutrino interaction which describes how much more or less likely it is with the randomly thrown parameter.

A new *universe* can be defined by randomly sampling from each underlying model parameter at once (in a correlated way, where necessary). Some number of universes can then be generated, and each universe is assigned a global weight for the neutrino interaction which is just the multiplication of all of the individual weights from the parameters. This process is called *reweighting*.

Using these alternate universes and the central value simulation, a covariance matrix may be constructed,

$$C_{ij} = \frac{1}{U} \sum_{\text{All universes}, u}^U [N_{Nom.}^i - N_{Univ.u}^i] \times [N_{Nom.}^j - N_{Univ.u}^j], \quad (7.14)$$

where $N_{Nom.}^i$ and $N_{Univ.u}^i$ are the number of entries in the i th bin of the nominal simulation and of the alternative universe, respectively, and U is the total number of alternative universes.

A covariance matrix encodes how each bin within the distribution covaries with every other bin. From this, the fractional covariance matrix can be defined,

$$F_{ij} = \frac{C_{ij}}{N_i N_j}, \quad (7.15)$$

in addition to a correlation matrix,

$$X_{ij} = \frac{C_{ij}}{\sqrt{C_{i,i}} \sqrt{C_{j,j}}}. \quad (7.16)$$

The fractional covariance matrix and correlation matrix contain the same information as the covariance matrix. The utility of these as opposed to the regular covariance matrix is that the diagonal of the fractional covariance matrix gives the fractional error in each bin, while the correlation matrix is unitless and has a lower and upper bound of -1 and 1.

It is often instructive to first construct covariance matrices for each source of systematic error, and then combine these covariance matrices. In this case, there are three covariance matrices

to construct: one for the flux uncertainties, one for the neutrino interaction uncertainties, and one for detector-related systematic uncertainties. This means the total covariance matrix can be written

$$C_{syst.} = C_{flux} + C_{xsec} + C_{detector} \quad (7.17)$$

and this can then be combined with the statistical covariance matrix, the entries for which lay on the diagonal

$$C = C_{syst.} + C_{stat.} \quad (7.18)$$

7.7.1 Flux Uncertainties

Flux uncertainties come from uncertainties on the parameters contained within the simulation of the MicroBooNE beam simulation as discussed in Chapter 4.2. These mainly come down to uncertainties on the hadron production, as well as uncertainties on effects such as the current in the focusing horn.

The flux uncertainties on the reconstructed neutrino energy is shown in Figure 7.31. Here, the green histogram is the central value from the nominal simulation and the colour map shows the distribution of systematic universes.

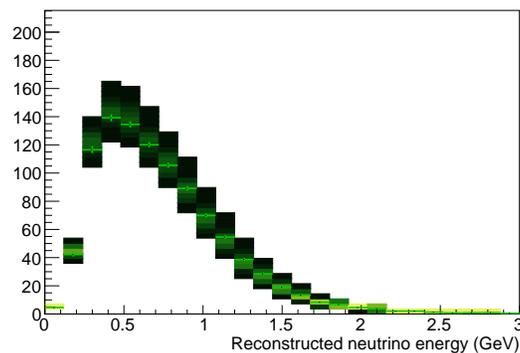


Figure 7.31: Distribution of systematic universes from the MicroBooNE flux simulation. The green histogram is the central value histogram, while the colour map shows the distribution of systematic universes in each bin.

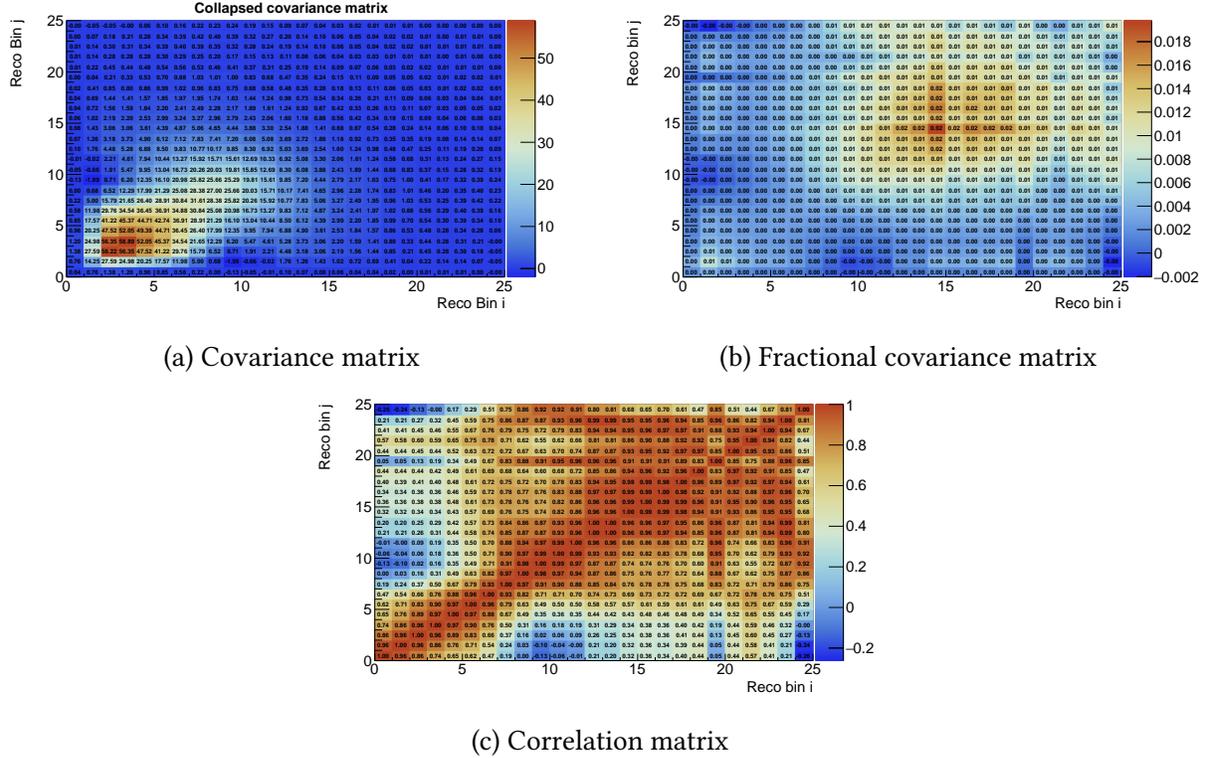


Figure 7.32: Covariance matrices for flux-related systematic uncertainties.

7.7.2 Interaction Uncertainties

Interaction uncertainties come from the GENIE event generator outlined in Section 4.3, where further details may be found in references [83] and [84].

The default tune of GENIE used by MicroBooNE uses the Llewellyn-Smith formalism for quasi-elastic interactions [108], the Rein-Seghal (RS) model for both resonant and coherent pion production interactions [87], the Bodek-Ritchie Fermi gas model of the nuclear environment [85], and the hA model [84] for final state interactions. In addition to this, a model is added in order to capture the effects of meson exchange currents (MEC), however this model is purely empirical.

For ease, the interaction-related systematic uncertainties used in this analysis along with their default values (where quoted in the GENIE manual [84]) and $\pm 1\sigma$ ranges are displayed in Table 7.4. In this table, the first section describes the axial and vector masses for different scattering processes, as well as other parameters related to quasi-elastic scattering and the RS model of resonant interactions. The second section relates to interactions in the deep-inelastic regime, which

is less important at MicroBooNE due to the typical energy of beam neutrinos. The third section relates to hadronization and resonance decay uncertainties, while the fourth accounts for non-resonant backgrounds to interactions with π in the final states. Finally, the last section contains parameters related to intranuclear hadron transport. This work is aimed at selecting interactions which contain only a muon and some number of protons in the final state, meaning that the quasi-elastic interaction uncertainties are much more prominent than those for the resonant and inelastic interactions.

The effect of interaction uncertainties on the reconstructed neutrino energy spectrum can be found in Figure 7.33. The associated covariance matrices can be found in Figure 7.34.

It is known that these interaction systematic uncertainties are incomplete. In particular, there is no uncertainty applied to meson exchange current interactions.

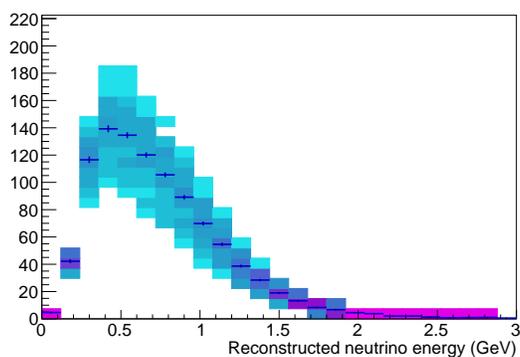


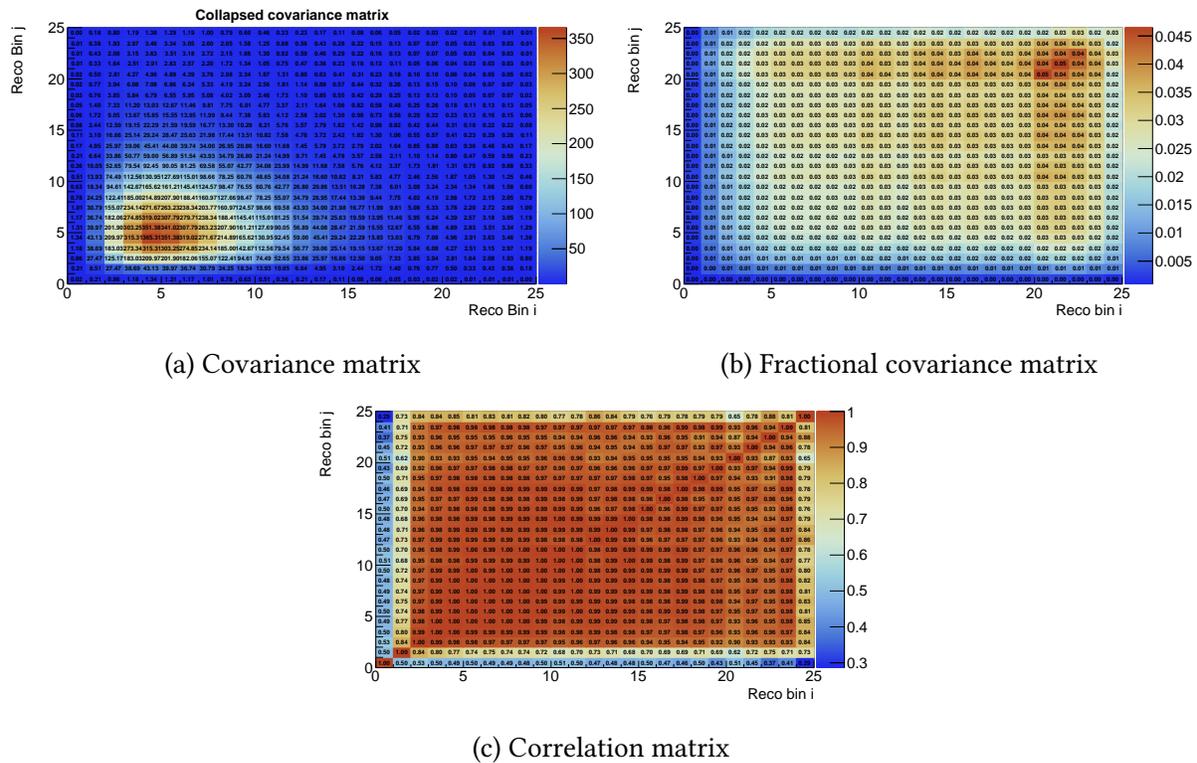
Figure 7.33: Distribution of systmatic universes from the GENIE generator. The blue histogram is the central value histogram, while the colour map shows the distribution of systematic universes in each bin.

7.7.3 Detector Uncertainties

The detector-related systematic uncertainties which are probed here are outlined in Section 7.2.4. The effect of these systematic uncertainties on the shape of the reconstructed neutrino energy can be found in Figure 7.35. A summary of the percentage change in number of selected events are presented in Table 7.5. The covariance matrices associated with detector-related uncertainties

Parameter P	Description of P	Value	$\delta P/P$
M_A^{NCEL}	Axial mass for NC elastic	0.990 GeV	$\pm 25\%$
η^{NCEL}	Strange axial form factor for NC elastic	0.120 GeV	$\pm 30\%$
M_A^{CCQE} A	Axial mass for CC quasi-elastic	0.990 GeV	$^{+15\%}_{-25\%}$
M_A^{CCRES}	Axial mass for CC resonance neutrino production	1.120 GeV	$\pm 20\%$
M_V^{CCRES}	Vector mass for CC resonance neutrino production	0.840 GeV	$\pm 10\%$
M_A^{NCRES}	Axial mass for NC resonance neutrino production	1.120 GeV	$\pm 20\%$
M_V^{NCRES}	Vector mass for NC resonance neutrino production	0.840 GeV	$\pm 50\%$
$M_A^{COH\pi}$	Axial mass for CC and NC coherent pion production	1.000 GeV	$\pm 50\%$
$R_0^{COH\pi}$	Nuclear size para. controlling π absorption in RS model	1.000 fm	$\pm 10\%$
CCQE-PauliSup (p)	CCQE Pauli suppression (via changes in Fermi level k_F)	0.242 GeV	$\pm 35\%$
CCQE-PauliSup (n)	CCQE Pauli suppression (via changes in Fermi level k_F)	0.259 GeV	$\pm 35\%$
A_{HT}^{BY}	A_{HT} higher-twist param. in BY model scaling variable ζ_ω	0.538	$\pm 25\%$
B_{HT}^{BY}	B_{HT} higher-twist param. in BY model scaling variable ζ_ω	0.305	$\pm 25\%$
C_{V1u}^{BY}	C_{V1u} u valence GRV98 PDF correction maram in BY model	0.291	$\pm 30\%$
C_{V2u}^{BY}	C_{V2u} u valence GRV98 PDF correction maram in BY model	0.189	$\pm 40\%$
FZ (pion)	Hadron formation zone	0.342 fm	$\pm 50\%$
FZ (nucleon)	Hadron formation zone	2.300 fm	$\pm 50\%$
BR(γ)	Branching ratio for radiative resonance decays	-	± 50
BR(η)	Branching ratio for single- η resonance decays	-	$\pm 50\%$
$R_{vp}^{CC1\pi}$	Non-resonance bkg in νp CC1 π reactions	-	$\pm 50\%$
$R_{vp}^{CC2\pi}$	Non-resonance bkg in νp CC2 π reactions	-	$\pm 50\%$
$R_{vn}^{CC1\pi}$	Non-resonance bkg in νn CC1 π reactions	-	$\pm 50\%$
$R_{vn}^{CC2\pi}$	Non-resonance bkg in νn CC2 π reactions	-	$\pm 50\%$
$R_{vp}^{NC1\pi}$	Non-resonance bkg in νp NC1 π reactions	-	$\pm 50\%$
$R_{vp}^{NC2\pi}$	Non-resonance bkg in νp NC2 π reactions	-	$\pm 50\%$
$R_{vn}^{NC1\pi}$	Non-resonance bkg in νn NC1 π reactions	-	$\pm 50\%$
$R_{vn}^{NC2\pi}$	Non-resonance bkg in νn NC2 π reactions	-	$\pm 50\%$
x_{abs}^N	Nucleon mean free path (total rescattering probability)	-	$\pm 20\%$
x_{cex}^N	Nucleon charge exchange probability	-	$\pm 50\%$
x_{el}^N	Nucleon elastic reaction probability	-	$\pm 30\%$
x_{inel}^N	Nucleon inelastic reaction probability	-	$\pm 40\%$
x_{mfp}^N	Nucleon absorption probability	-	$\pm 20\%$
x_π^N	Nucleon π -production probability	-	$\pm 20\%$
x_{abs}^π	π mean free path (total rescattering probability)	-	$\pm 20\%$
x_{cex}^π	π charge exchange probability	-	$\pm 50\%$
x_{el}^π	π elastic reaction probability	-	$\pm 10\%$
x_{inel}^π	π inelastic reaction probability	-	$\pm 40\%$
x_{mfp}^π	π absorption probability	-	$\pm 20\%$
x_π^π	π π -production probability	-	$\pm 20\%$

Table 7.4: Neutrino interaction model parameters and associated uncertainties taken from the GENIE generator. These are taken initially from the GENIE Users Manual [84] but were grouped together by the authors of reference [109].



(a) Covariance matrix

(b) Fractional covariance matrix

(c) Correlation matrix

Figure 7.34: Covariance matrices for interaction-related systematic uncertainties.

can be found in Figure 7.36.

Understanding Detector-Related Systematic Uncertainties

In order to understand whether the detector-related systematic uncertainties correctly capture knowledge of the detector, it is possible to make data-driven plots similar to those in Figures 7.6, 7.7, 7.8, and 7.9. Because it is known that the largest detector-related systematic, DIC, primarily affects the reconstruction of protons, only the impact on protons is analysed. To this end, a cut is placed at -1 in the PID distribution in Figure 7.11, and everything to the left is defined to be a proton candidate, resulting in a pure selection of protons. Given this, we can ask how many PFParticles which fall in this area were identified by Pandora as a track versus being identified as a shower. The results are presented in Figure 7.37.

These plots show that in the majority of the phase space the systematic uncertainties cover the disagreement between data and simulation. There are notable exceptions for higher energy

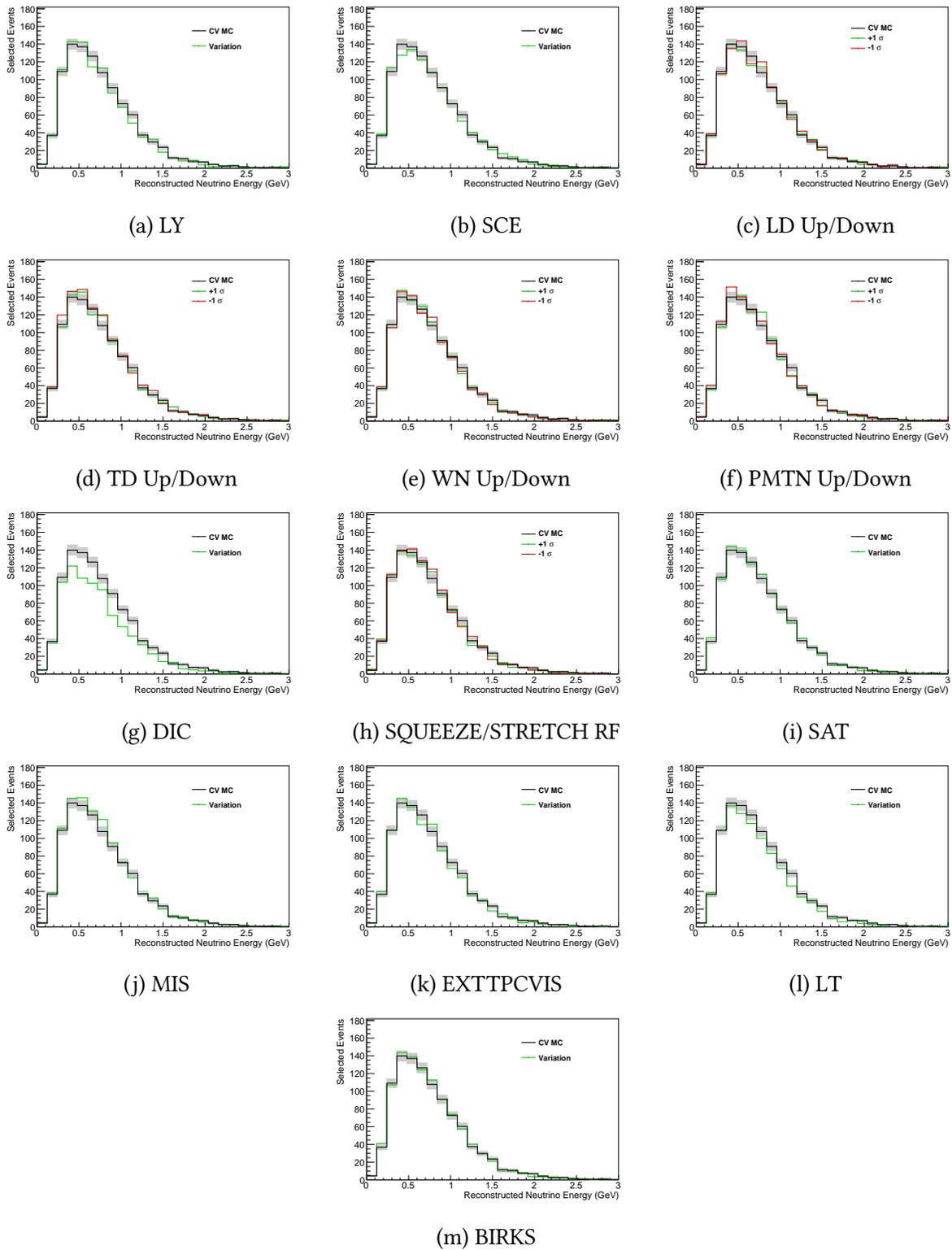


Figure 7.35: Effect of detector-related systematic uncertainties on the reconstructed neutrino energy spectrum. Each detector variation used is described in Section 7.2.4.

Sample	N. Selected Events	$(N_{det.var} - N_{CV})/N_{CV} \times 100$ (%)
CV	1030.82	0
SCE	1016.15	-1.42
LY	1003.14	-2.68
LDUp	1018.88	-1.16
LDDown	1038.79	+0.77
TDU _p	1037.52	+0.64
TDDown	1073.82	+4.17
WNU _p	1031.84	+0.10
WND _{Down}	1026.90	-0.38
PMTNU _p	1023.74	-0.69
PMTND _{Down}	1037.81	+0.68
DIC	843.29	-18.20
SQUEEZERF	1016.69	-1.37
STRETCHRF	1037.57	+0.65
SAT	1035.59	+0.46
MIS	1062.62	+3.08
EXTTPCVIS	1004.92	-2.51
LT	946.425	-8.19
BIRKS	1035.59	+0.46
Total systematic uncertainty		21.1%

Table 7.5: Effect of detector-related systematic uncertainties on the number of selected signal candidates. For the total systematic uncertainty, the larger of $(+1\sigma, -1\sigma)$ is taken.

protons, which are likely to be forwards going. One possible reason for this is that tracks which are forward going are more likely to be impacted by the coherent noise removal, which may make it appear more sparse and therefore more likely to be reconstructed as a shower. It is also likely that this is in part driven by physics which is not captured by the models in the event generator.

7.7.4 Total Uncertainties

Combining the flux, cross section, and detector systematic uncertainties with the statistical uncertainties results in the covariance matrices located in Figure 7.38

From this total covariance matrix it is possible to extract the total error in each bin of reconstructed energy. This is shown in Figure 7.39. In general, the data fall within the systematic error bands, although there is a systematic excess in the simulation. Much of this is due to the lack of

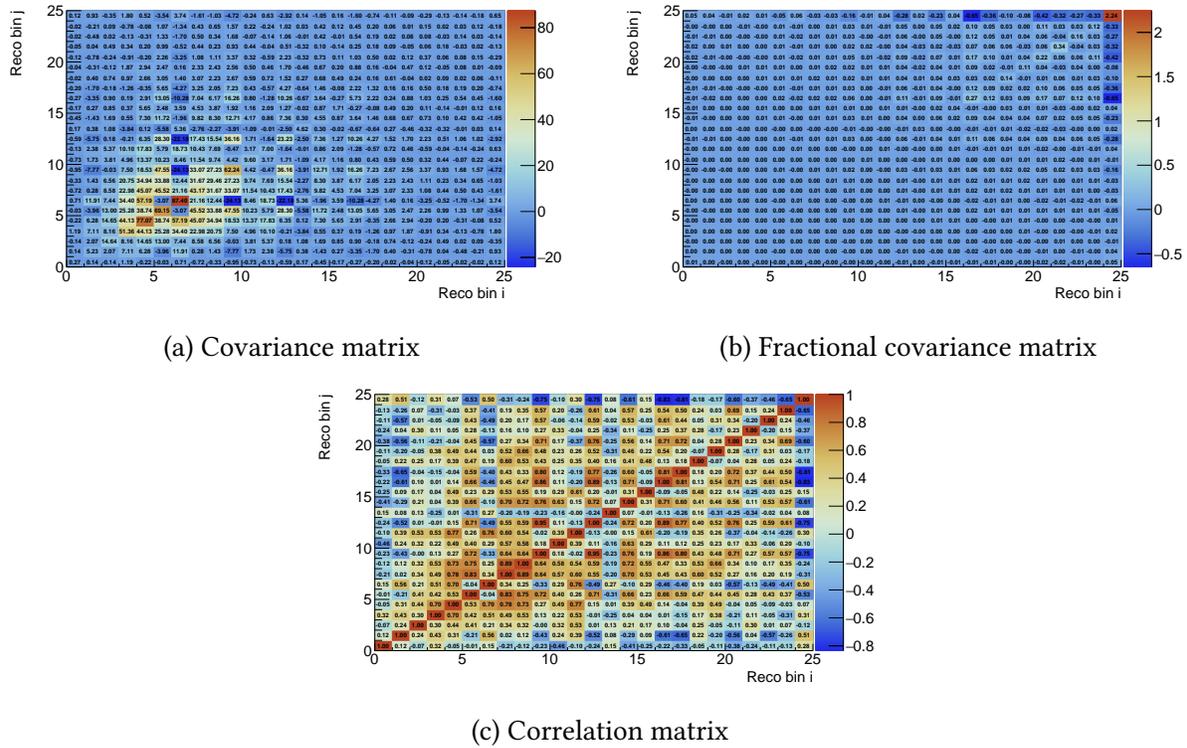


Figure 7.36: Covariance matrices for detector-related systematic uncertainties.

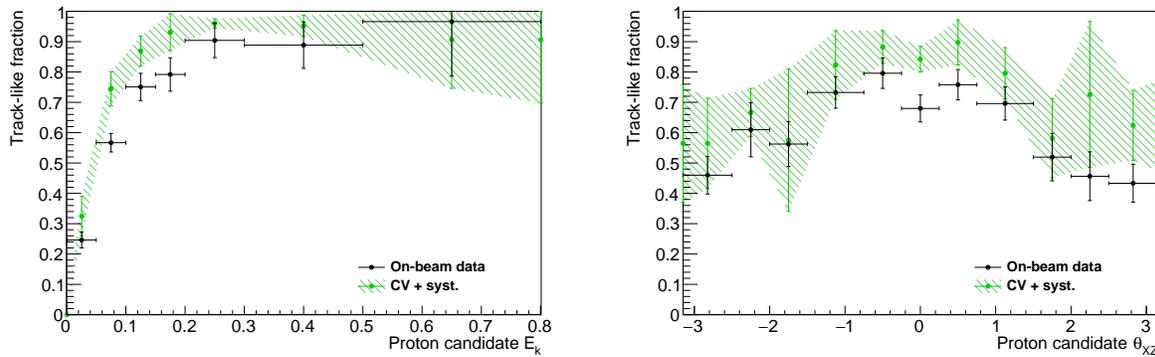


Figure 7.37: Data-driven track fraction plots for on-beam data and for the CV dataset with detector-related systematic uncertainties included (green, hashed).

induced charge in the simulation, as shown in Figure 7.35.

In total, 804 events are selected from the on-beam data, with a statistical uncertainty of ± 28.6 events compared to a predicted 1031.7 ± 32.0 (stat.) ± 157.8 (syst.). The ratio between data and simulation is $R = 0.78 \pm 0.04$ (stat.) ± 0.12 (syst.).

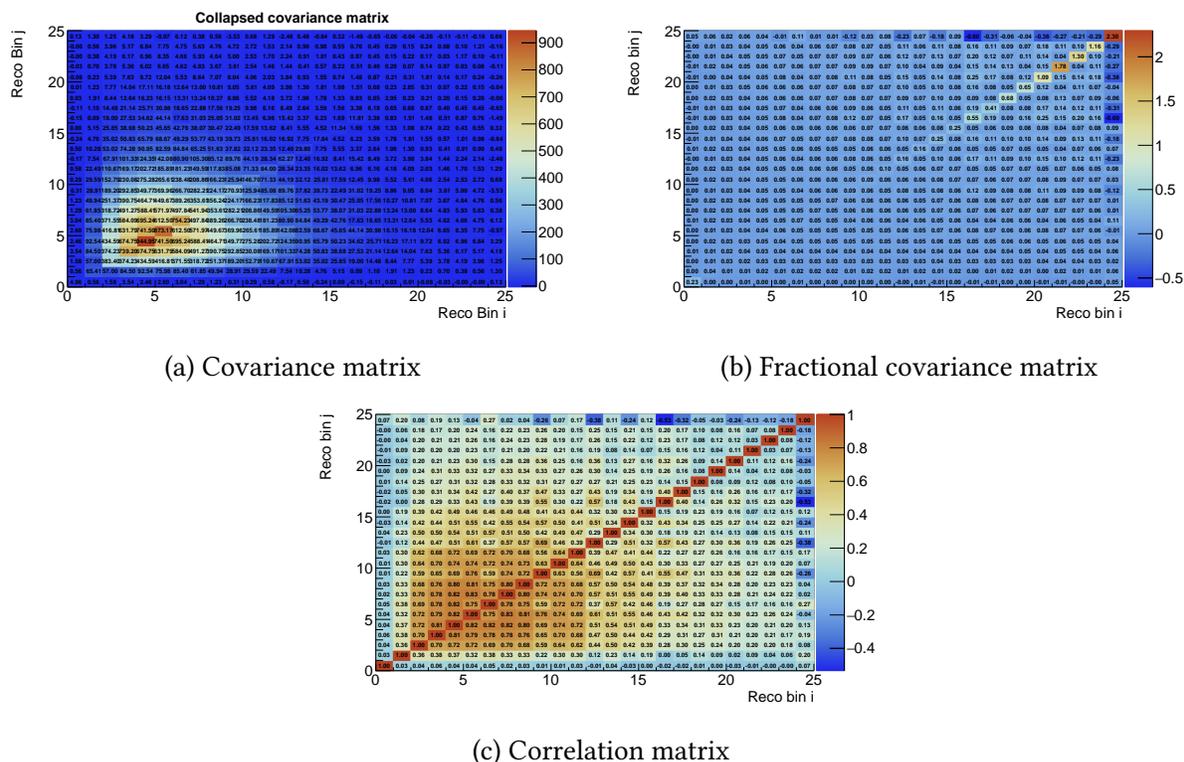


Figure 7.38: Combined covariance matrices for flux, interaction, and detector systematic uncertainties.

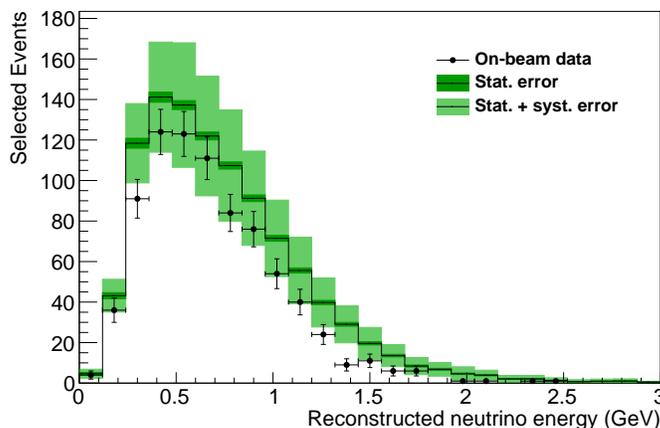


Figure 7.39: Distribution of reconstructed neutrino energy with statistical uncertainty shown in dark green and stat. + syst. uncertainty shown in light green. The black histogram includes the off-beam data as well as signal and backgrounds from the simulation. The black data points are the selected on-beam data events.

Chapter 8

The ν_μ Constraint

“Scientists have calculated that the chances of something so patently absurd actually existing are millions to one. But magicians have calculated that million-to-one chances crop up nine times out of ten.”

Terry Pratchett

The excess of low-energy electron neutrino candidates observed by the MiniBooNE collaboration [11] can be interpreted in two distinct ways. The first is that the excess is electron-like in nature, while the second treats the excess as photon-like.

The MicroBooNE collaboration is actively developing analyses under both of these assumptions. The expectation is that for analyses developed under the electron neutrino hypothesis, an improvement in the sensitivity can be obtained by application of a muon neutrino dataset in order to constrain systematic uncertainties. This chapter outlines how this is performed using the muon neutrino selection outlined in Chapter 7 and a dataset containing electron neutrinos.

The power of the ν_μ constraint comes primarily from the fact that ν_μ and ν_e interactions are expected to be effectively the same up to kinematic differences due to the mass of the lepton, and from the fact that the ν_e and ν_μ in the beam are kinematically related because they originate from the same particles, as shown in Figure 3.2.

Some detector effects are also expected to be constrained, however many such effects will impact the reconstruction of the two topologies differently. An example of this is longitudinal electron diffusion, where reconstruction of showers is expected to be disproportionately affected compared to tracks due to signal pulses falling below threshold. Some evidence for this effect can be derived by comparing the ν_μ CC-inclusive cross section, where longitudinal diffusion contributes a 1.4% uncertainty to the total cross section [104], to the ν_μ CC π^0 cross section, where it contributes a 9% uncertainty to the cross section for single photon events.

In cases such as this one MicroBooNE is pursuing in-situ measurements of the detector effects in order to constrain the systematic uncertainties.

8.1 Understanding the ν_μ Constraint

To estimate MicroBooNEs sensitivity to the MiniBooNE low-energy excess, a χ^2 test statistic is constructed,

$$\chi^2 = \sum_{i,j=1}^{n_b^{\nu_e} + n_b^{\nu_\mu}} \delta_i C_{i,j}^{-1} \delta_j, \quad (8.1)$$

where $n_b^{\nu_e}$ and $n_b^{\nu_\mu}$ are defined to be the number of bins in the ν_e and ν_μ distributions, respectively, and

$$\delta_i = N_i^{data} - N_i^{sim}. \quad (8.2)$$

Here, $N_i^{data/sim}$ is the number of events in i th bin for the concatenation of the ν_e and ν_μ distributions for data or simulation, and $C_{i,j}^{-1}$ is the inverse covariance matrix.

In order to estimate the magnitude of the constraining power obtained by inclusion of the ν_μ in the χ^2 calculation, the prescription outlined in Chapter 4.2.4 of reference [110] is followed.

Here, $n_b^{\nu_\mu} + n_b^{\nu_e}$ fit parameters, N_i^{fit} , are introduced, one for each bin. Δ_i is then defined to be

$$\Delta_i = N_i^{fit} - N_i^{sim}, \quad (8.3)$$

and the χ^2 can be written

$$\chi^2 = \sum_{i,j=1}^{n_b^{v_e} + n_b^{v_\mu}} \Delta_i C_{i,j}^{-1} \Delta_j. \quad (8.4)$$

Minimising this results in $N_i^{fit} = N_i^{sim} \pm \sqrt{C_{i,i}}$, $\chi^2 = 0$.

It should be noted here that taking the derivative of the χ^2 with respect to the fit parameters results in just the inverse covariance matrix with an additional factor of 2,

$$\frac{\partial^2 \chi^2}{\partial N_i^{fit} \partial N_j^{fit}} = 2 \times C_{i,j}^{-1}. \quad (8.5)$$

The constraint can be applied by including a pull term for the ν_μ bins in the χ^2 calculation which applies a penalty if the fit parameters fall outside of the statistical uncertainty of the data,

$$\chi^2 = \sum_{i,j=1}^{n_b^{v_e} + n_b^{v_\mu}} \Delta_i C_{i,j}^{-1} \Delta_j + \sum_{i=n_b^{v_e}+1}^{n_b^{v_e} + n_b^{v_\mu}} \frac{(N_i^{fit} - N_i^{data})^2}{(\sigma_i^{data})^2} \quad (8.6)$$

$$= \sum_{i,j=1}^{n_b^{v_e} + n_b^{v_\mu}} \Delta_i C_{i,j}^{-1} \Delta_j + \sum_{i=n_b^{v_e}+1}^{n_b^{v_e} + n_b^{v_\mu}} \frac{(N_i^{fit} - N_i^{data})^2}{N_i^{data}}. \quad (8.7)$$

This can then be differentiated with respect to the fit parameters to retrieve an effective inverse covariance matrix,

$$\frac{1}{2} \frac{\partial^2 \chi^2}{\partial N_i^{fit} \partial N_j^{fit}} = (C_{i,j}^{-1})^{eff} = C_{ij}^{-1} + \frac{\delta(i=j, i > n_b^{v_e})}{N_i^{data}}, \quad (8.8)$$

where the Kronecker delta in the second term, $\delta(i=j, i > n_b^{v_e})$, implies that this term is only applied to the ν_μ part of the covariance matrix, and is only non-vanishing for the case in which $i=j$.

This matrix can now be inverted, and the constrained uncertainties in bin i of the ν_e spectrum can be taken to be $\sqrt{(C_{i,i})^{eff}}$.

8.2 Demonstration of the ν_μ Constraint

In order to demonstrate the constraint, a dataset containing ν_e has been provided, following the selection outlined in reference [111]. The distribution of selected events are displayed in Figure 8.1 for a sample with and without the LEE signal. The binning chosen here is a modification to the MiniBooNE binning, combining the 4 lowest bins into a single bin due to the low electron neutrino selection efficiency, and therefore low number of events in this region.

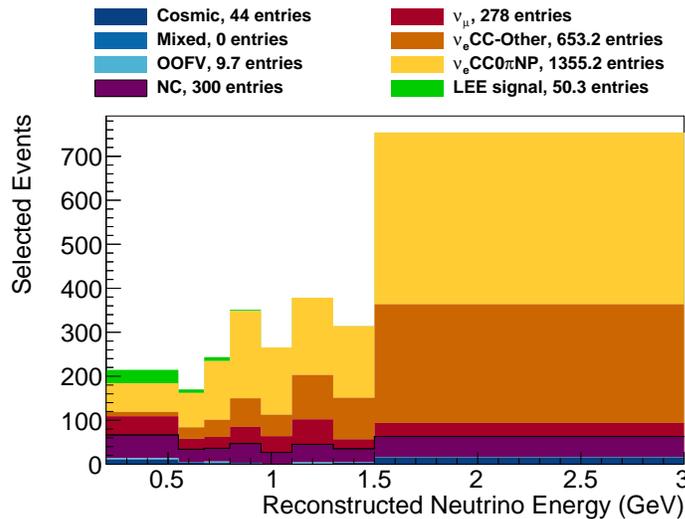


Figure 8.1: Selected events from the electron neutrino selection outlined in reference [111]. The number of selected signal events (LEE signal) is estimated from the number of selected ν_e interactions.

There are some known limitations to this dataset:

- The booster neutrino beam contains only 0.5% intrinsic ν_e . To improve the number of selected events, the selection was performed on a dataset which includes only beam-intrinsic ν_e and simulated cosmic rays overlaid. This means that this dataset does not include a number of backgrounds, such as ν_μ CC and ν_μ NC backgrounds.
- The ν_e intrinsic sample provided corresponds to $4e20$ POT. The expected total POT MicroBooNE will collect is $13e20$ POT, meaning that each event from this sample is weighted up

by a factor of 3.25.

- To estimate the backgrounds, the electron neutrino selection is also performed on the full beam MicroBooNE simulation. The provided dataset corresponds only to $5e19$ POT, meaning very few backgrounds were selected. In order to overcome this, the shape of the ν_μ CC and ν_μ NC events have been estimated from the selected events, and this shape has been used to randomly generate backgrounds of the correct scale. The result of this is that **these backgrounds do not have associated systematic uncertainties**. This means that the systematic uncertainties are under-estimated in this demonstration.
- No detector-related systematic uncertainties were provided for the electron neutrino selection, meaning that effect of the constraint on these uncertainties cannot be studied at this time.

These limitations mean that there is expected to be considerable statistical noise in this dataset due to the low number of simulated events, while the systematic uncertainties are under-estimated. This means that care should be taken when interpreting the results in this chapter, and this should be used as a demonstration of the constraining power of the muon neutrino dataset only.

In order to demonstrate the method outlined in Section 8.1, a covariance matrix describing the covariance between each bin in the electron neutrino and muon neutrino datasets is first constructed, using the no-LEE electron neutrino distribution from Figure 8.1. This is shown in Figure 8.2.

From this covariance matrix, the total error on the electron neutrino bins can be taken to be $\sqrt{C_{i,i}}$ in the first 8 bins. Following the procedure outlined in Section 8.1, the effective correlation matrix can be found, and the constrained errors can be read in the same way.

The result of performing this procedure can be found in Figure 8.3 and Table 8.1. As expected, the systematic uncertainties in each of the electron neutrino bins have been reduced through inclusion of the muon neutrino dataset.

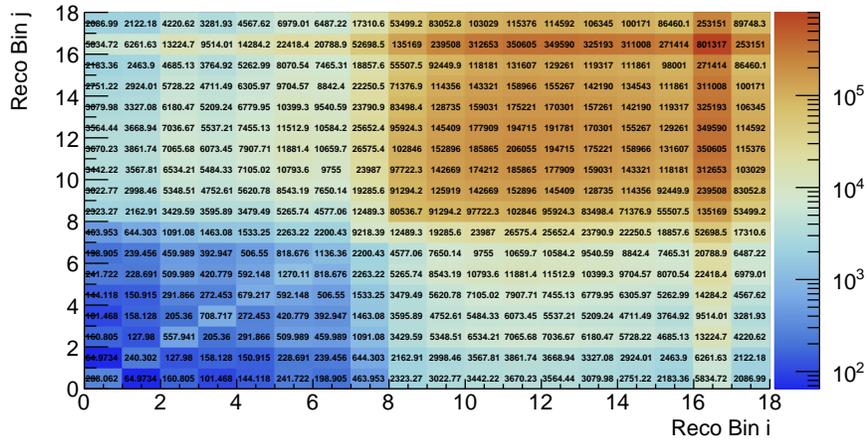
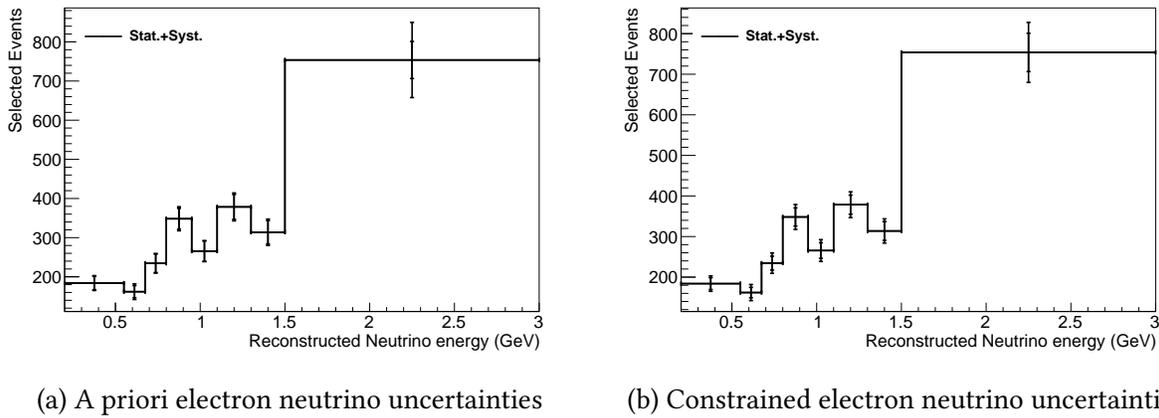


Figure 8.2: Full statistical plus systematic covariance matrix showing the full flux and cross section covariance between the muon neutrino sample and the electron neutrino sample. The first 8 bins correspond to the electron neutrino sample and the remaining 10 bins correspond to the muon neutrino sample.



(a) A priori electron neutrino uncertainties

(b) Constrained electron neutrino uncertainties

Figure 8.3: Total stat. + syst. uncertainties on the electron neutrino selection before and after applying the muon neutrino constraint.

8.3 Frequentist Studies

Given the electron and muon neutrino datasets, the SBNFit framework [112] may be used to pose a statistical question: Given some data with the associated statistical and systematic uncertainties, how well is MicroBooNE able to reject the scenario in which there is no excess of electron-like events at low energy?

In order to give a quantitative answer to this question, two hypotheses are defined:

ν_e bin number	bin range (GeV)	Stat. frac. err.	Syst. frac. err. w/o constraint	Syst. frac. err. w/ constraint	Syst. percentage difference	Tot. frac. err. w/o constraint	Tot. frac. err. w/ constraint	Tot. percentage difference
1	0.200 - 0.550	0.073	0.055	0.029	47%	0.092	0.079	14%
2	0.550 - 0.675	0.079	0.055	0.009	83%	0.096	0.079	18%
3	0.675 - 0.800	0.065	0.077	0.033	57%	0.101	0.073	28%
4	0.800 - 0.950	0.051	0.055	0.036	34%	0.076	0.065	14%
5	0.950 - 1.100	0.061	0.077	0.039	49%	0.098	0.072	27%
6	1.100 - 1.300	0.051	0.079	0.037	53%	0.094	0.062	34%
7	1.300 - 1.500	0.056	0.091	0.047	48%	0.107	0.073	32%
8	1.500 - 3.000	0.036	0.122	0.091	25%	0.127	0.098	23%

Table 8.1: Table showing the fractional error in each electron neutrino bin, split into statistical error, systematic error, and total error before and after application of the muon neutrino constraint.

- H_0 : the null hypothesis. No excess of data exists.
- H_1 : the alternative hypothesis. This is taken to be the excess of data events observed by MiniBooNE. This has been unfolded from the MiniBooNE reconstructed energy to a true energy following the procedure outlined in reference [113].

Given these hypotheses it is possible to produce two probability distribution functions. To do this, two fake datasets, D_0 and D_1 are produced by pulling from the covariance matrix of H_0 and H_1 respectively. Each bin in these datasets is then Poisson fluctuated. By doing this over many pseudo-experiments and calculating a test statistic for each dataset every time,

$$\chi^2 = \sum_{i,j=1}^{n_b^{\nu_e} + n_b^{\nu_\mu}} (D_i^0 - H_i^0) M_{i,j}^{-1} (D_j^0 - H_j^0), \quad (8.9)$$

Dataset	bin 1	bin 2	bin 3	bin 4	bin 5
ν_e	0.2-0.4 GeV	0.4-0.6 GeV	0.6-0.8 GeV	0.8-1.0 GeV	-
ν_μ (0 bins)	-	-	-	-	-
ν_μ (1 bins)	0.1-3.0 GeV	-	-	-	-
ν_μ (2 bins)	0.1-0.8 GeV	0.8-3.0 GeV	-	-	-
ν_μ (5 bins)	0.1-0.3 GeV	0.3-0.5 GeV	0.5-0.8 GeV	0.8-1.0 GeV	1.0-3.0 GeV

Table 8.2: Binning used in the frequentist studies for the ν_μ and ν_e samples.

$$\chi^2 = \sum_{i,j=1}^{n_b^{\nu_e} + n_b^{\nu_\mu}} (D_i^1 - H_i^0) M_{i,j}^{-1} (D_j^1 - H_j^0), \quad (8.10)$$

the distributions can be built up. It should be noted here that every time this test statistic is built, the null hypothesis is also fluctuated within systematic uncertainties.

Results from the frequentist studies can be found in Figure 8.4. Here, a slightly different binning is chosen in order to make the effects of the constraint more apparent. The chosen binning is presented in Table 8.2. The results are presented with the ν_μ dataset binned in four different ways in order to show the effect of additional degrees of freedom on the final sensitivities.

For each plot in Figure 8.4, a chosen χ^2 value can be used to calculate a sensitivity, and this has been done for several χ^2 values in each plot:

- The median sensitivity: the median of the distribution of χ^2 values for H_1 is chosen. This is an important quantity because 50% of the time MicroBooNE would measure a more significant excess than this and 50% of the time it would measure a less significant excess.
- The $\pm 1\sigma$, $\pm 2\sigma$ sensitivities: 68.2% of the time the measured excess will give a sensitivity in the 1σ range, while 95.4% of the time the measured excess will give a sensitivity in the 2σ range.

The median sensitivities for these four sets of frequentist studies show that using the muon neutrino constraint does improve the ability of MicroBooNE to reject the null hypothesis.

It is observed that the largest improvement to the sensitivity is obtained by including only a single muon bin such that the muon neutrino dataset is only acting as a normalisation constraint. Including shape information through increasing the number of bins seems to decrease the sen-

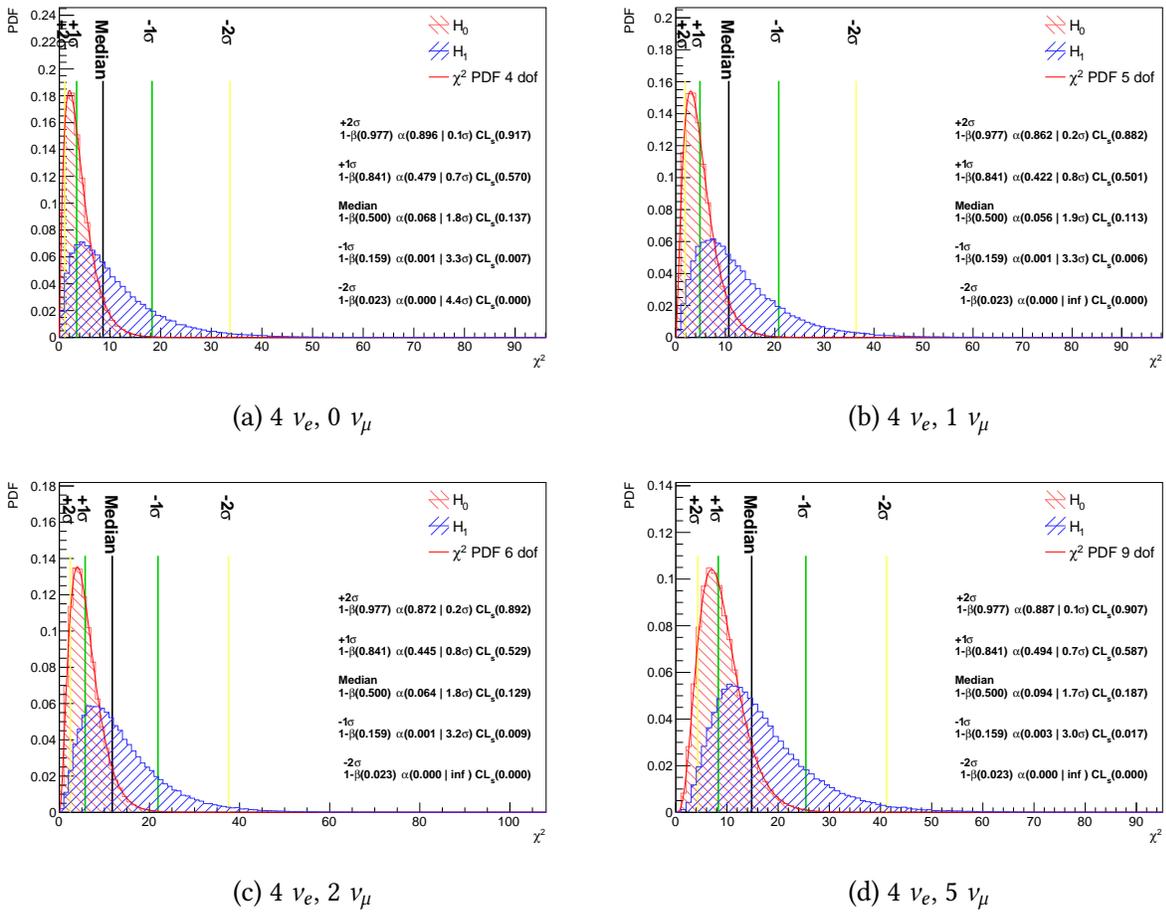


Figure 8.4: Results of performing frequentist studies for the case in which 0, 1, 2 and 5 bins of ν_μ bins are included in the fit.

sitivity. Including 5 muon neutrino bins is shown to make the sensitivity worse than using only the electron neutrino bins.

This is thought to be because the uncertainty in the signal region of the electron neutrinos is dominated by the statistical error, with the systematic uncertainty not becoming dominant until higher energies. This means that with the current electron neutrino selection the constraint can have only a small impact. Including additional bins in the muon neutrino distribution increases the number of degrees of freedom beyond the capability of the dataset to constrain the uncertainties. The constraint is expected to provide a larger effect when the electron neutrino selection attains higher selection efficiency in the signal region, below 800 MeV. This can be investigated by

applying a flat scale factor to the efficiency of the electron neutrino selection, and this is presented in Figure 8.5. As the electron neutrino selection efficiency is scaled up, the expected significance with and without the muon neutrino constraint begin to diverge.

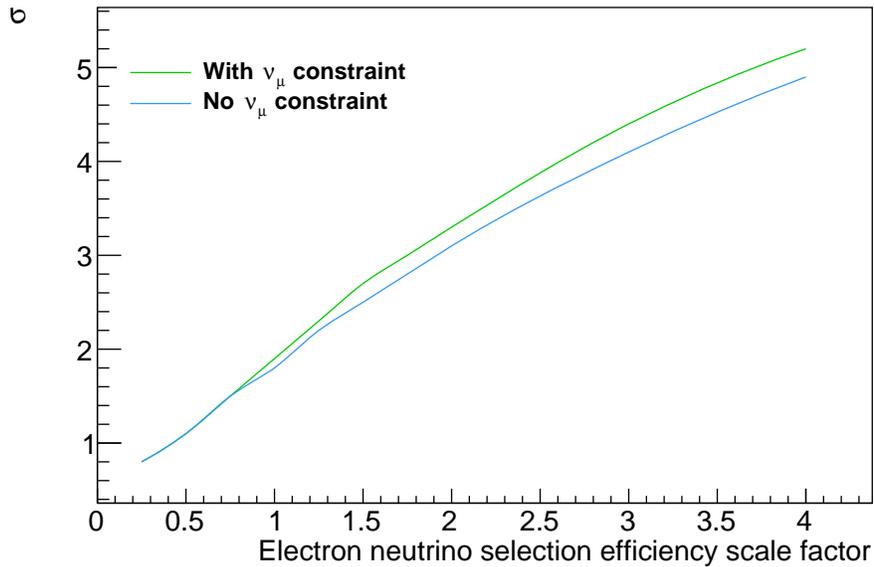


Figure 8.5: Significance with which H_0 can be rejected in favour of H_1 , as a function of electron neutrino efficiency scale factor. As the electron neutrino selection efficiency

If the selection efficiency is able to be increased by a factor of three, then the constraint, as currently implemented, is able to increase the median significance with which the null hypothesis can be rejected from 4.1σ to 4.4σ .

Chapter 9

Conclusions

“I think I’ll stop here.”

Andrew Wiles

The field of neutrino physics has many open questions. It is hoped that some of these questions can be addressed by the current and next generation of neutrino oscillation experiments. Regarding the existence of sterile neutrinos, these hopes are in part pinned on the SBN program at Fermilab, while for CP violation, we must look to the next generation experiments: Hyper-Kamiokande and DUNE.

This thesis has presented a number of analyses intended provide insight into MicroBooNE’s investigation of the excess of low-energy electromagnetic events observed by the MiniBooNE collaboration, and to develop tools and techniques for LArTPCs.

Chapter 5 presents MicroBooNE’s first measurement of longitudinal electron diffusion. This is an important effect to constrain due to the large discrepancies between the theory and measurements, and its effect on electron neutrino searches. This measurement is important not only for MicroBooNE and SBN, but also for DUNE, where the dual-phase far detector module is expected to have a 12 m drift.

In chapter 6, a novel method of performing particle identification in LArTPCs is developed. It is shown that the Bragg Likelihood algorithm marks an improvement in both data-to-simulation

agreement and robustness against systematic uncertainties over the methods previously used by MicroBooNE and other LArTPCs.

Chapter 7 presents a selection of a pionless ν_μ charged-current interactions with at least one proton above a 40 KeV threshold on kinetic energy. This includes a discussion of the importance of the effects of induced charge and its impact on analyses, and presents a method by which the sensitivity to this effect can be reduced. This approach is being pursued by other analyses which are expected to be published on this software release.

In chapter 8, this thesis has demonstrated a path towards the first step of the SBN program: investigation of the excess of low-energy electron neutrino candidates observed by the MiniBooNE collaboration. Specifically, this chapter focuses on the constraint of systematic uncertainties through use of a dataset containing muon neutrinos.

The analyses here are from the first-generation of MicroBooNE analyses. There are a number of improvements which can be made in future iterations.

The first has been widely discussed within this work: improvement to the simulation of drifting ionisation electrons is clearly a priority. This has been worked on extensively over the last year, and is in the process of being incorporated into the MicroBooNE simulation. Additionally, the inclusion of a two-dimensional deconvolution will result in an improved signal processing [67, 90], and it is hoped that this will lead to an improved reconstruction. It is also thought that this simulation improvement will allow for calorimetry to be performed on the induction planes, meaning an improvement in the particle identification capabilities of MicroBooNE.

Further improvement to the muon neutrino constraint can be attained by bringing the electron neutrino selection into a greater degree of alignment with the muon neutrino selection. Ensuring that the same particle identification and fiducial volume requirements are met will give greater power to the constraint due to the similarity of the backgrounds. The sensitivity to a MiniBooNE-style low-energy excess is primarily driven by the efficiency and purity of the electron neutrino selection, and so this is another area which would benefit from an improved reconstruction.

Bibliography

- [1] A. D. Sakharov, “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe,” *Pisma Zh. Eksp. Teor. Fiz.*, vol. 5, pp. 32–35, 1967, [Usp. Fiz. Nauk161,no.5,61(1991)].
- [2] F. Reines and C. L. Cowan, “Detection of the free neutrino,” *Phys. Rev.*, vol. 92, pp. 830–831, 1953.
- [3] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, “Detection of the free neutrino: A Confirmation,” *Science*, vol. 124, pp. 103–104, 1956.
- [4] C. Anderson *et al.*, “First Measurements of Inclusive Muon Neutrino Charged Current Differential Cross Sections on Argon,” *Phys. Rev. Lett.*, vol. 108, p. 161802, 2012.
- [5] P. Cennini *et al.*, “ICARUS II - A Second Generation Proton Decay Experiment and Neutrino Observatory at the GRAN SASSO Laboratory: Proposal by the ICARUS Collaboration,” 1993.
- [6] R. Acciarri *et al.*, “Design and Construction of the MicroBooNE Detector,” *JINST*, vol. 12, no. 02, p. P02017, 2017.
- [7] M. Antonello *et al.*, “A Proposal for a Three Detector Short-Baseline Neutrino Oscillation Program in the Fermilab Booster Neutrino Beam,” 2015.
- [8] R. Acciarri *et al.*, “Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE),” 2015.

- [9] B. Abi *et al.*, “The Single-Phase ProtoDUNE Technical Design Report,” 2017.
- [10] P. Adamson *et al.*, “LArIAT: Liquid Argon TPC in a Test Beam,” 2013.
- [11] A. A. Aguilar-Arevalo *et al.*, “Significant Excess of Electron Like Events in the MiniBooNE Short-Baseline Neutrino Experiment,” *Phys. Rev. Lett.*, vol. 121, no. 22, p. 221801, 2018.
- [12] J. Chadwick, “Intensitätsverteilung im magnetischen Spectrum der β -Strahlen von radium B + C,” *Verhandl. Dtsc. Phys. Ges.*, vol. 16, p. 383, 1914. [Online]. Available: <http://cds.cern.ch/record/262756>
- [13] W. Pauli, “Dear radioactive ladies and gentlemen,” *Phys. Today*, vol. 31N9, p. 27, 1978.
- [14] E. Fermi, “Tentativo di una teoria dell’emissione dei raggi beta,” *Ric. Sci.*, vol. 4, pp. 491–495, 1933.
- [15] H. Bethe and R. Peierls, “The ‘neutrino’,” *Nature*, vol. 133, p. 532, 1934.
- [16] G. Danby, J. M. Gaillard, K. A. Goulianos, L. M. Lederman, N. B. Mistry, M. Schwartz, and J. Steinberger, “Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos,” *Phys. Rev. Lett.*, vol. 9, pp. 36–44, 1962.
- [17] K. Kodama *et al.*, “Observation of tau neutrino interactions,” *Phys. Lett.*, vol. B504, pp. 218–224, 2001.
- [18] G. Aad *et al.*, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett.*, vol. B716, pp. 1–29, 2012.
- [19] C. Giunti and C. W. Kim, *Fundamentals of Neutrino Physics and Astrophysics*. Oxford University Press, 2007.
- [20] J. A. Formaggio and G. P. Zeller, “From eV to EeV: Neutrino Cross Sections Across Energy Scales,” *Rev. Mod. Phys.*, vol. 84, pp. 1307–1341, 2012.

-
- [21] J. N. Bahcall and R. Davis, “Solar Neutrinos - a Scientific Puzzle,” *Science*, vol. 191, pp. 264–267, 1976, [,141(1976)].
- [22] W. C. Haxton, “The Solar Neutrino Problem,” , vol. 33, pp. 459–504, 1995.
- [23] M. C. Gonzalez-Garcia, “Solutions to the atmospheric neutrino problem,” *Nucl. Phys. Proc. Suppl.*, vol. 81, pp. 113–123, 2000.
- [24] M. Gell-Mann and A. Pais, “Behavior of neutral particles under charge conjugation,” *Phys. Rev.*, vol. 97, pp. 1387–1389, 1955.
- [25] Z. Maki, M. Nakagawa, and S. Sakata, “Remarks on the unified model of elementary particles,” *Prog. Theor. Phys.*, vol. 28, pp. 870–880, 1962, [,34(1962)].
- [26] H. Fritzsch and P. Minkowski, “Vector-Like Weak Currents, Massive Neutrinos, and Neutrino Beam Oscillations,” *Phys. Lett.*, vol. 62B, pp. 72–76, 1976.
- [27] S. Eliezer and A. R. Swift, “Experimental Consequences of electron Neutrino-Muon-neutrino Mixing in Neutrino Beams,” *Nucl. Phys.*, vol. B105, pp. 45–51, 1976.
- [28] S. M. Bilenky and B. Pontecorvo, “Quark-Lepton Analogy and Neutrino Oscillations,” *Phys. Lett.*, vol. 61B, p. 248, 1976, [,248(1975)].
- [29] A. Lister, “Neutrino Oscillations in Long Baseline Experiments,” Master’s thesis, Durham U., 2014.
- [30] Y. Fukuda *et al.*, “Evidence for oscillation of atmospheric neutrinos,” *Phys. Rev. Lett.*, vol. 81, pp. 1562–1567, 1998.
- [31] Q. R. Ahmad *et al.*, “Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory,” *Phys. Rev. Lett.*, vol. 89, p. 011301, 2002.
- [32] F. P. An *et al.*, “Observation of electron-antineutrino disappearance at Daya Bay,” *Phys. Rev. Lett.*, vol. 108, p. 171803, 2012.

-
- [33] K. Abe *et al.*, “Search for CP Violation in Neutrino and Antineutrino Oscillations by the T2K Experiment with 2.2×10^{21} Protons on Target,” *Phys. Rev. Lett.*, vol. 121, no. 17, p. 171802, 2018.
- [34] M. A. Acero *et al.*, “New constraints on oscillation parameters from ν_e appearance and ν_μ disappearance in the NOvA experiment,” *Phys. Rev.*, vol. D98, p. 032012, 2018.
- [35] M. G. Aartsen *et al.*, “Searches for Sterile Neutrinos with the IceCube Detector,” *Phys. Rev. Lett.*, vol. 117, no. 7, p. 071801, 2016.
- [36] P. Adamson *et al.*, “Combined analysis of ν_μ disappearance and $\nu_\mu \rightarrow \nu_e$ appearance in MINOS using accelerator and atmospheric neutrinos,” *Phys. Rev. Lett.*, vol. 112, p. 191801, 2014.
- [37] “Nufit wepage,” <http://www.nu-fit.org/>.
- [38] N. Agafonova *et al.*, “Observation of a first ν_τ candidate in the OPERA experiment in the CNGS beam,” *Phys. Lett.*, vol. B691, pp. 138–145, 2010.
- [39] M. G. Aartsen *et al.*, “Measurement of Atmospheric Tau Neutrino Appearance with IceCube DeepCore,” *Phys. Rev.*, vol. D99, no. 3, p. 032007, 2019.
- [40] S. Schael *et al.*, “Precision electroweak measurements on the Z resonance,” *Phys. Rept.*, vol. 427, pp. 257–454, 2006.
- [41] C. Athanassopoulos *et al.*, “Evidence for anti-muon-neutrino \rightarrow anti-electron-neutrino oscillations from the LSND experiment at LAMPF,” *Phys. Rev. Lett.*, vol. 77, pp. 3082–3085, 1996.
- [42] 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.*, vol. D64, p. 112007, 2001.

- [43] J. M. Conrad and M. H. Shaevitz, “Sterile Neutrinos: An Introduction to Experiments,” *Adv. Ser. Direct. High Energy Phys.*, vol. 28, pp. 391–442, 2018.
- [44] A. A. Aguilar-Arevalo *et al.*, “Measurement of $\nu(\mu)$ and anti- $\nu(\mu)$ induced neutral current single π^0 production cross sections on mineral oil at $E(\nu) \sim O(1\text{-GeV})$,” *Phys. Rev.*, vol. D81, p. 013005, 2010.
- [45] K. N. Abazajian *et al.*, “Light Sterile Neutrinos: A White Paper,” 2012.
- [46] M. Antonello *et al.*, “Experimental search for the “LSND anomaly” with the ICARUS detector in the CNGS neutrino beam,” *Eur. Phys. J.*, vol. C73, no. 3, p. 2345, 2013.
- [47] P. Astier *et al.*, “Search for $\nu(\mu) \rightarrow \nu(e)$ oscillations in the NOMAD experiment,” *Phys. Lett.*, vol. B570, pp. 19–31, 2003.
- [48] K. B. M. Mahn *et al.*, “Dual baseline search for muon neutrino disappearance at $0.5\text{eV}^2 < \Delta m^2 < 40\text{eV}^2$,” *Phys. Rev.*, vol. D85, p. 032007, 2012.
- [49] G. Cheng *et al.*, “Dual baseline search for muon antineutrino disappearance at $0.1\text{eV}^2 < \Delta m^2 < 100\text{eV}^2$,” *Phys. Rev.*, vol. D86, p. 052009, 2012.
- [50] P. Adamson *et al.*, “Search for Sterile Neutrinos Mixing with Muon Neutrinos in MINOS,” *Phys. Rev. Lett.*, vol. 117, no. 15, p. 151803, 2016.
- [51] P. Adamson *et al.*, “Limits on Active to Sterile Neutrino Oscillations from Disappearance Searches in the MINOS, Daya Bay, and Bugey-3 Experiments,” *Phys. Rev. Lett.*, vol. 117, no. 15, p. 151801, 2016, [Addendum: *Phys. Rev. Lett.*117,no.20,209901(2016)].
- [52] I. E. Stockdale *et al.*, “Limits on Muon Neutrino Oscillations in the Mass Range $55\text{-eV}^{**2} < \Delta m^{**2} < 800\text{-eV}^{**2}$,” *Phys. Rev. Lett.*, vol. 52, p. 1384, 1984.
- [53] J. N. Abdurashitov *et al.*, “Measurement of the response of a Ga solar neutrino experiment to neutrinos from an Ar-37 source,” *Phys. Rev.*, vol. C73, p. 045805, 2006.

- [54] M. Apollonio *et al.*, “Search for neutrino oscillations on a long baseline at the CHOOZ nuclear power station,” *Eur. Phys. J.*, vol. C27, pp. 331–374, 2003.
- [55] K. Eguchi *et al.*, “First results from KamLAND: Evidence for reactor anti-neutrino disappearance,” *Phys. Rev. Lett.*, vol. 90, p. 021802, 2003.
- [56] F. P. An *et al.*, “Improved Measurement of Electron Antineutrino Disappearance at Daya Bay,” *Chin. Phys.*, vol. C37, p. 011001, 2013.
- [57] G. Mention, M. Fechner, T. Lasserre, T. A. Mueller, D. Lhuillier, M. Cribier, and A. Loutourneau, “The Reactor Antineutrino Anomaly,” *Phys. Rev.*, vol. D83, p. 073006, 2011.
- [58] Y. Ko *et al.*, “Sterile Neutrino Search at the NEOS Experiment,” *Phys. Rev. Lett.*, vol. 118, no. 12, p. 121802, 2017.
- [59] M. Danilov, “Recent results of the DANSS experiment,” in *32nd Rencontres de Physique de La Vallée d’Aoste (La Thuile 2018) La Thuile, Aosta, Italy, February 25-March 3, 2018*, 2018.
- [60] M. Dentler, A. Hernández-Cabezudo, J. Kopp, P. A. N. Machado, M. Maltoni, I. Martinez-Soler, and T. Schwetz, “Updated Global Analysis of Neutrino Oscillations in the Presence of eV-Scale Sterile Neutrinos,” *JHEP*, vol. 08, p. 010, 2018.
- [61] H. L. H. Wong, “Search for a Light Sterile Neutrino at Daya Bay,” *J. Phys. Conf. Ser.*, vol. 888, no. 1, p. 012130, 2017.
- [62] A. A. Aguilar-Arevalo *et al.*, “The Neutrino Flux prediction at MiniBooNE,” *Phys. Rev.*, vol. D79, p. 072002, 2009.
- [63] D. Barak, B. Harrison, and A. Watts, “Concepts Rookie Book,” 2013.
- [64] B. Worthel, J. Crawford, B. Polico *et al.*, “Booster Rookie Book,” 2009.
- [65] M. Thomson, *Modern Particle Physics*. Cambridge, United Kingdom: Cambridge University Press, 2013.

- [66] B. J. P. Jones, “Sterile Neutrinos in Cold Climates,” Ph.D. dissertation, MIT, 2015. [Online]. Available: <http://lss.fnal.gov/archive/thesis/2000/fermilab-thesis-2015-17.pdf>
- [67] C. Adams *et al.*, “Ionization electron signal processing in single phase LArTPCs. Part I. Algorithm Description and quantitative evaluation with MicroBooNE simulation,” *JINST*, vol. 13, no. 07, p. P07006, 2018.
- [68] R. Acciarri *et al.*, “Noise Characterization and Filtering in the MicroBooNE Liquid Argon TPC,” *JINST*, vol. 12, no. 08, p. P08003, 2017.
- [69] The MicroBooNE Collaboration, “Study of Space Charge Effects in MicroBooNE,” 2016.
- [70] C. Adams *et al.*, “Design and construction of the MicroBooNE Cosmic Ray Tagger system,” 2019.
- [71] D. Caratelli, “Study of Electromagnetic Interactions in the MicroBooNE Liquid Argon Time Projection Chamber,” Ph.D. dissertation, Columbia U., 2018. [Online]. Available: <http://lss.fnal.gov/archive/thesis/2000/fermilab-thesis-2018-02.pdf>
- [72] E. L. Snider and G. Petrillo, “LArSoft: Toolkit for Simulation, Reconstruction and Analysis of Liquid Argon TPC Neutrino Detectors,” *J. Phys. Conf. Ser.*, vol. 898, no. 4, p. 042057, 2017.
- [73] <http://art.fnal.gov>.
- [74] S. Agostinelli *et al.*, “Geant4 – a simulation toolkit,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250 – 303, 2003. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0168900203013688>
- [75] J. Allison *et al.*, “Recent developments in geant4,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 835, pp. 186 – 225, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0168900216306957>

- [76] M. Sorel, “The GEANT4-based beam Monte Carlo for MiniBooNE,” *MiniBooNE Technical Note 103 Version 2*, 2005.
- [77] I. Chemakin *et al.*, “Pion production by protons on a thin beryllium target at 6.4-GeV, 12.3-GeV/c, and 17.5-GeV/c incident proton momenta,” *Phys. Rev.*, vol. C77, p. 015209, 2008, [Erratum: *Phys. Rev.*C77,049903(2008)].
- [78] A. Bolshakova *et al.*, “Cross-Sections of Large-Angle Hadron Production in Proton- and Pion-Nucleus Interactions I: Beryllium Nuclei and Beam Momenta of +8.9 GeV/c and -8.0 GeV/c,” *Eur. Phys. J.*, vol. C62, pp. 293–317, 2009.
- [79] G. Cheng *et al.*, “Measurement of K^+ production cross section by 8 GeV protons using high energy neutrino interactions in the SciBooNE detector,” *Phys. Rev.*, vol. D84, p. 012009, 2011.
- [80] F. Abe *et al.*, “Production of Neutral Strange Particles K_S^0 and Λ^0 by 12-GeV Protons on Nuclear Targets,” *Phys. Rev.*, vol. D36, pp. 1302–1319, 1987.
- [81] D. A. A. Wickremasinghe, “HARP targets pion production cross section and yield measurements: Implications for MiniBooNE neutrino flux,” Ph.D. dissertation, Cincinnati U., 2015. [Online]. Available: <http://lss.fnal.gov/archive/thesis/2000/fermilab-thesis-2015-23.pdf>
- [82] D. W. Schmitz, “A Measurement of Hadron Production Cross Sections for the Simulation of Accelerator Neutrino Beams and a Search for ν_μ to ν_e Scillations in the Δm^2 about equals $1 - eV^2$ Region,” Ph.D. dissertation, Columbia U., 2008. [Online]. Available: http://lss.fnal.gov/cgi-bin/find_paper.pl?thesis-2008-26
- [83] C. Andreopoulos *et al.*, “The GENIE Neutrino Monte Carlo Generator,” *Nucl. Instrum. Meth.*, vol. A614, pp. 87–104, 2010.

- [84] C. Andreopoulos, C. Barry, S. Dytman, H. Gallagher, T. Golan, R. Hatcher, G. Perdue, and J. Yarba, “The GENIE Neutrino Monte Carlo Generator: Physics and User Manual,” 2015.
- [85] A. Bodek and J. L. Ritchie, “Further Studies of Fermi Motion Effects in Lepton Scattering from Nuclear Targets,” *Phys. Rev.*, vol. D24, p. 1400, 1981.
- [86] C. H. Llewellyn Smith, “Neutrino Reactions at Accelerator Energies,” *Phys. Rept.*, vol. 3, pp. 261–379, 1972.
- [87] D. Rein and L. M. Sehgal, “Neutrino Excitation of Baryon Resonances and Single Pion Production,” *Annals Phys.*, vol. 133, pp. 79–153, 1981.
- [88] The MicroBooNE Collaboration, “Cosmic shielding studies at microboone,” *MICROBOONE-NOTE-1005-PUB*, 2016.
- [89] D. Heck, G. Schatz, T. Thouw, J. Knapp, and J. N. Capdevielle, “CORSIKA: A Monte Carlo code to simulate extensive air showers,” 1998.
- [90] C. Adams *et al.*, “Ionization electron signal processing in single phase LArTPCs. Part II. Data/simulation comparison and performance in MicroBooNE,” *JINST*, vol. 13, no. 07, p. P07007, 2018.
- [91] R. Acciarri *et al.*, “The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector,” *Eur. Phys. J.*, vol. C78, no. 1, p. 82, 2018.
- [92] R. Acciarri *et al.*, “A study of electron recombination using highly ionizing particles in the ArgoNeuT Liquid Argon TPC,” *JINST*, vol. 8, p. P08005, 2013, <http://iopscience.iop.org/article/10.1088/1748-0221/8/08/P08005>.
- [93] V. M. Atrazhev and I. V. Timoshkin, “Transport of electrons in atomic liquids in high electric fields,” *Trans. Dielectrics and Electrical Insulation*, vol. 5, no. 3, pp. 450–457.

- [94] P. Cennini *et al.*, “Performance of a 3-ton liquid argon time projection chamber,” *Nucl. Instrum. Meth.*, vol. A345, pp. 230–243, 1994.
- [95] Y. Li *et al.*, “Measurement of Longitudinal Electron Diffusion in Liquid Argon,” *Nucl. Instrum. Meth.*, vol. A816, pp. 160–170, 2016.
- [96] T. K. Warburton, “Simulations and Data analysis for the 35 ton Liquid Argon detector as a prototype for the DUNE experiment.”
- [97] C. Adams *et al.*, “First Measurement of ν_μ Charged-Current π^0 Production on Argon with a LArTPC,” 2018.
- [98] S. E. Derenzo, A. R. Kirschbaum, P. H. Eberhard, R. R. Ross, and F. T. Solmitz, “Test of a Liquid Argon Chamber with 20-Micrometer RMS Resolution,” *Nucl. Instrum. Meth.*, vol. 122, p. 319, 1974.
- [99] E. Shibamura, T. Takahashi, S. Kubota, and T. Doke, “Ratio of diffusion coefficient to mobility for electrons in liquid argon,” *Phys. Rev.*, vol. A20, no. 6, p. 2547, 1979.
- [100] M. Torti, “Electron diffusion measurements in the ICARUS T600 detector,” *J. Phys. Conf. Ser.*, vol. 888, no. 1, p. 012060, 2017.
- [101] The MicroBooNE Collaboration, “Establishing a pure sample of side-piercing through-going cosmic-ray muons for lartpc calibration in microboone,” 2017.
- [102] M. Tanabashi *et al.*, “Review of particle physics,” *Phys. Rev. D*, vol. 98, p. 030001, Aug 2018. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>
- [103] S. Amoruso *et al.*, “Study of electron recombination in liquid argon with the ICARUS TPC,” *Nucl. Instrum. Meth.*, vol. A523, pp. 275–286, 2004.
- [104] The MicroBooNE Collaboration, “First muon-neutrino charged-current inclusive differential cross section measurement for microboone run 1 data,” 2018.

-
- [105] J. S. Marshall and M. A. Thomson, “The pandora software development kit for pattern recognition,” *The European Physical Journal C*, vol. 75, no. 9, p. 439, 2015.
- [106] P. Abratenko *et al.*, “Determination of muon momentum in the MicroBooNE LArTPC using an improved model of multiple Coulomb scattering,” *JINST*, vol. 12, no. 10, p. P10010, 2017.
- [107] S. Das, “A simple alternative to the Crystal Ball function,” 2016.
- [108] C. H. Llewellyn Smith, “Neutrino Reactions at Accelerator Energies,” *Phys. Rept.*, vol. 3, pp. 261–379, 1972.
- [109] M. Del Tutto and A. Schukraft, “First muon-neutrino charged-current inclusive single and double-differential cross section measurement for microboone run 1 data,” *MICROBOONE-NOTE-INT*, 2018.
- [110] A. A. Aguilar-Arevalo, “An Improved Neutrino Oscillations Analysis of the MiniBooNE Data,” Ph.D. dissertation, Columbia U., 2008.
- [111] The MicroBooNE Collaboration, “Electron-neutrino selection and reconstruction in the MicroBooNE LArTPC using the Pandora multi-algorithm pattern recognition,” *MICROBOONE-NOTE-1038-PUB*, 2018.
- [112] D. Cianci, A. Furmanski, G. Karagiorgi, and M. Ross-Lonergan, “Prospects of Light Sterile Neutrino Oscillation and CP Violation Searches at the Fermilab Short Baseline Neutrino Facility,” *Phys. Rev.*, vol. D96, no. 5, p. 055001, 2017.
- [113] The MicroBooNE Collaboration, “MicroBooNE low-energy excess signal prediction from unfolding MiniBooNE Monte-Carlo and data,” *MICROBOONE-NOTE-1043-PUB*, 2018.

Appendix A

Additional Information for Particle Identification in MicroBooNE

A.1 Toy Width Study

The aim of the toy width study is to estimate the width of the Gaussian needed to smear the simulation in order to make the dE/dx distributions in physics data and simulated data more similar. This is done under the assumption that the differences are caused by detector effects that are not currently simulated, which can be modeled by an additional uncorrelated Gaussian. The improved data/simulation agreement seen when a smearing of this type is applied supports the assumption that there are some effects that can be modeled in this way.

When fitting dE/dx distributions as in Section 6.2, the Landau distribution width for both data and simulation is constrained to be the same, and so it is sufficient to perform the toy study with only Gaussian functional forms. A Gaussian toy sample is first generated by throwing randomly from a Gaussian distribution with a given mean and width 100,000 times. Each point is then smeared by multiplying its value by another random number thrown from a different Gaussian. By changing the width of this second Gaussian while keeping the first constant, the width of the additional smearing Gaussian can be empirically estimated.

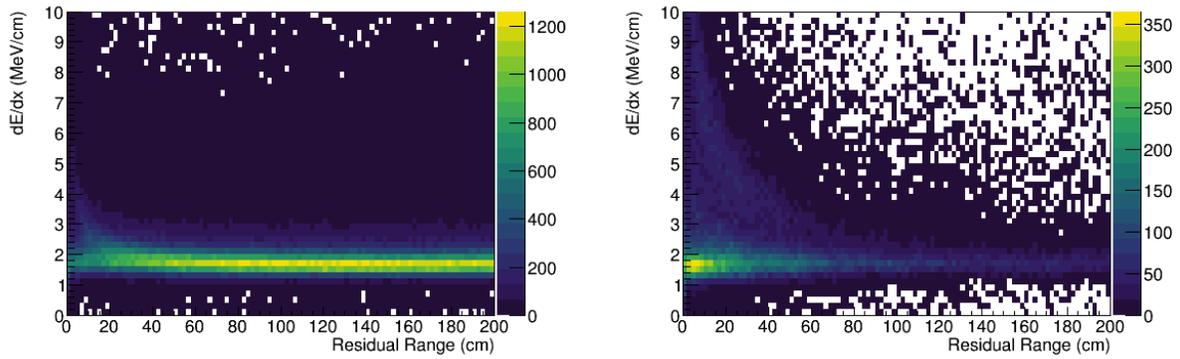
The additional widths needed to smear simulation in order to match the width of the data distribution are estimated for the three planes to be $\sigma_0 = 0.19$, $\sigma_1 = 0.30$, and $\sigma_2 = 0.11$.

In order to understand whether these widths are a reasonable approximation for protons as well as muons, the widths of protons in data have been measured in the collection plane. This is done by first selecting out two-track events from data, and selecting the shorter track, which should be proton-enhanced.

There is obviously significant contamination here. By removing co-linear tracks, which are expected to be broken muons, this contamination can be reduced, but not completely removed without seriously impacting the proton candidate statistics. Instead, we take the region between 65 and 70 cm residual range and fit the muon-like peak with a Landau-Gaussian functional form of fixed Landau and Gaussian width ($\sigma_L = 0.09$ and $\sigma_G = 0.20$, as in section 6.2), and fit the proton peak with a fixed Landau width of 0.13, allowing the Gaussian width to float.

The result of this fit is a proton Gaussian width of 0.34, but with a reasonably large uncertainty of 0.09. In section 6.2, we measured the Gaussian width for simulated protons to be 0.25. We can therefore use the same toy study process as above (smearing by a second Gaussian with different values) to empirically find the value needed to match the simulation width of 0.25 to the data width of 0.34. By doing so, we estimate that the additional smearing needed for protons on the collection plane is $\sigma_2 = 0.14^{+0.07}_{-0.13}$, which is consistent with the value of $\sigma_2 = 0.11$ derived from muon tracks earlier in this section.

This loose piece of evidence combined with the improved agreement between data and simulation in the proton-like region across all tested algorithms gives some confidence that this smearing works reasonably well.



(a) Longer track dE/dx versus residual range

(b) Shorter track dE/dx versus residual range

Figure A.1: dE/dx versus residual range distributions for the longer (left) and shorter (right) track in two-track selected events.

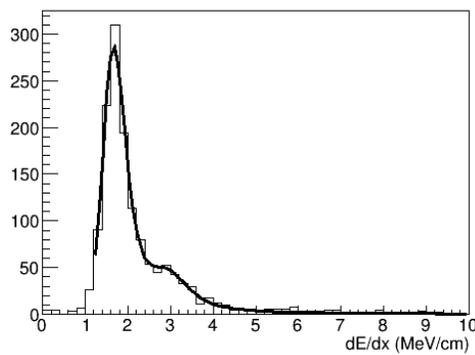


Figure A.2: dE/dx distribution for the shorter track of two-track selected events between 65 and 70 cm residual range, fit with a double Landau-Gaussian curve.

A.2 Bare χ^2 Distributions

This section contains both POT normalised and template fit distributions of the bare χ^2 variables under the muon, pion, proton and kaon assumptions. These are provided in order to complement the $\chi_\mu^2 - \chi_p^2$ variable which is presented in Chapter 6.

A.2.1 POT normalised

Figure A.3 shows POT normalised plots of the χ^2 variables.

A.2.2 Template Fit

Figure A.4 shows template fit distributions of the χ^2 variables.

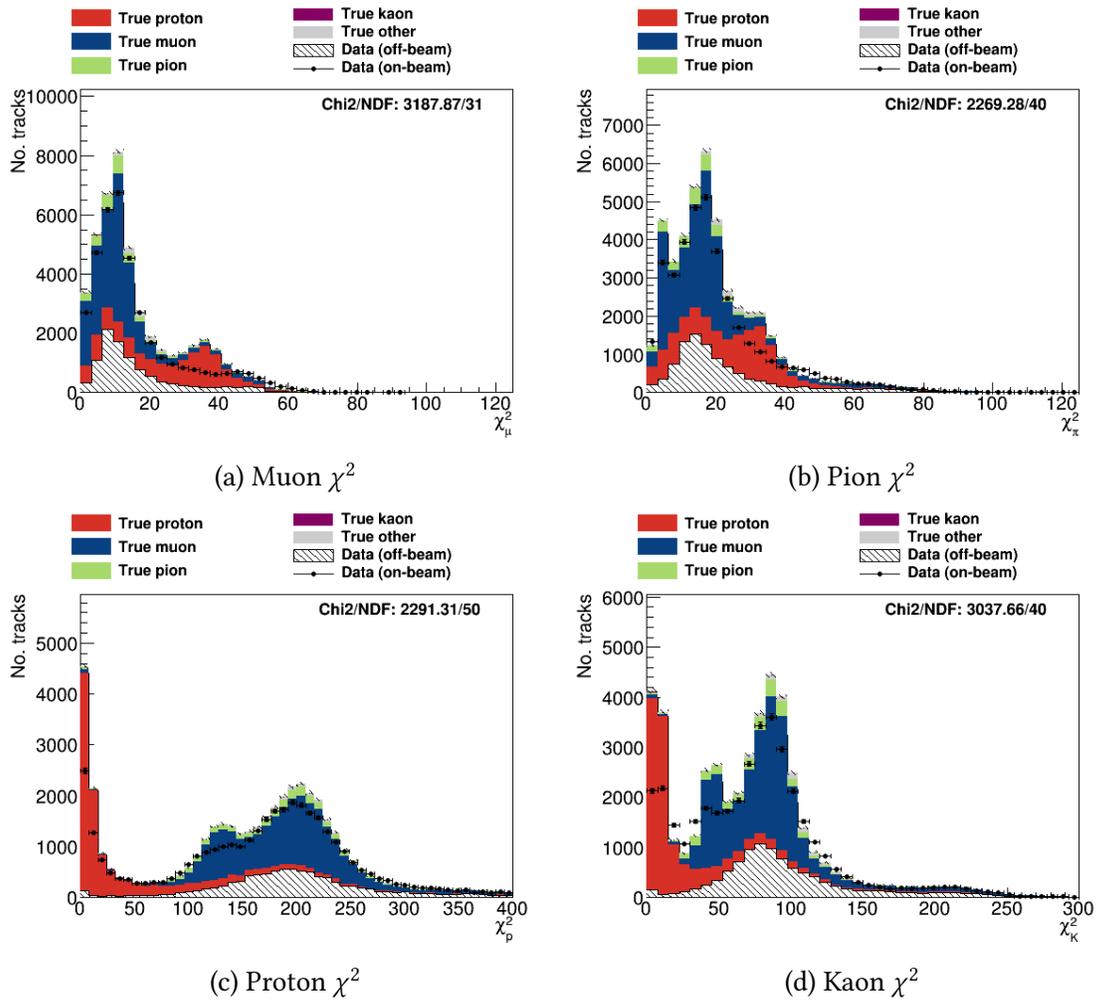


Figure A.3: POT normalised bare χ^2 for tracks under muon, pion, proton, kaon, and MIP assumptions.

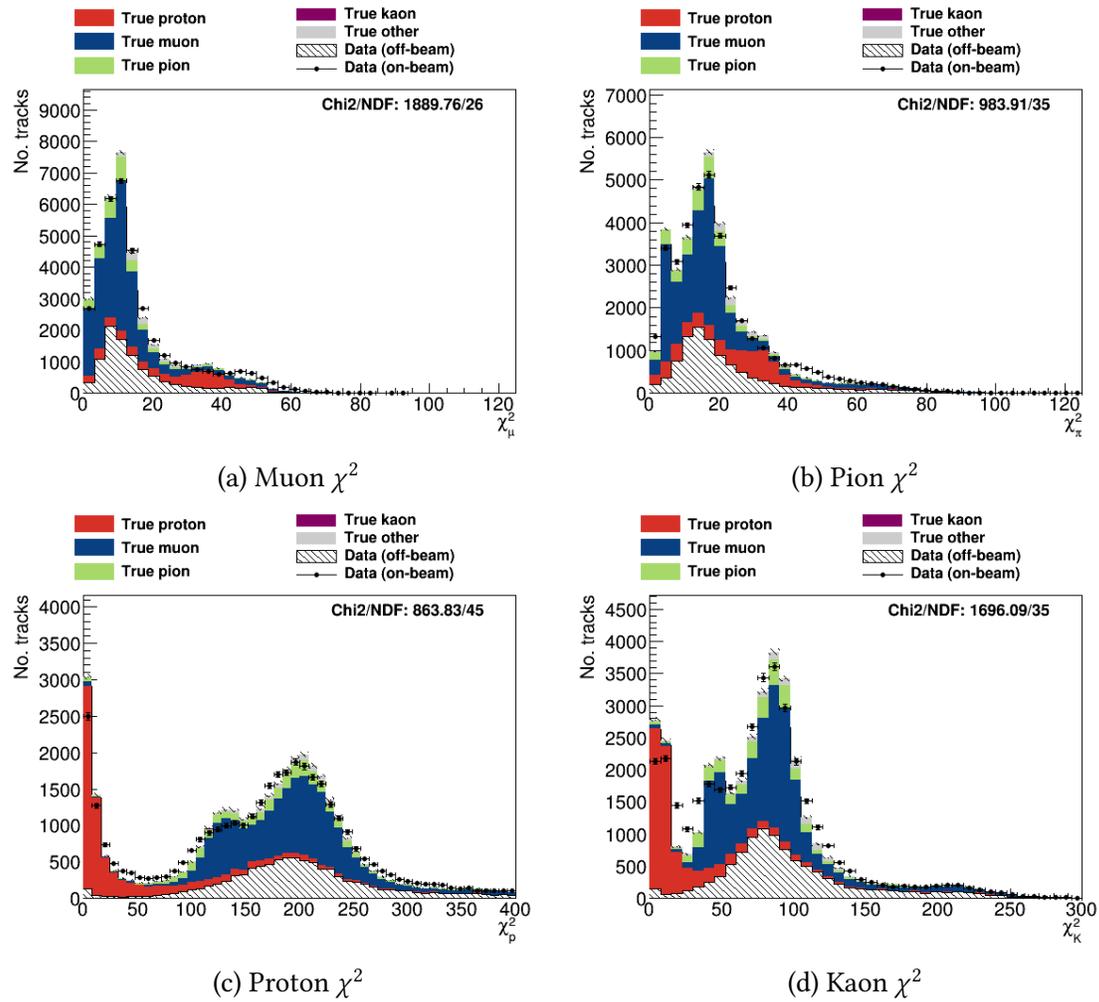


Figure A.4: Template fit bare χ^2 for tracks under muon, pion, proton, kaon, and MIP assumptions.

A.3 POT normalised Distributions

This section shows the POT normalised versions of the plots presented in Section 6.3.

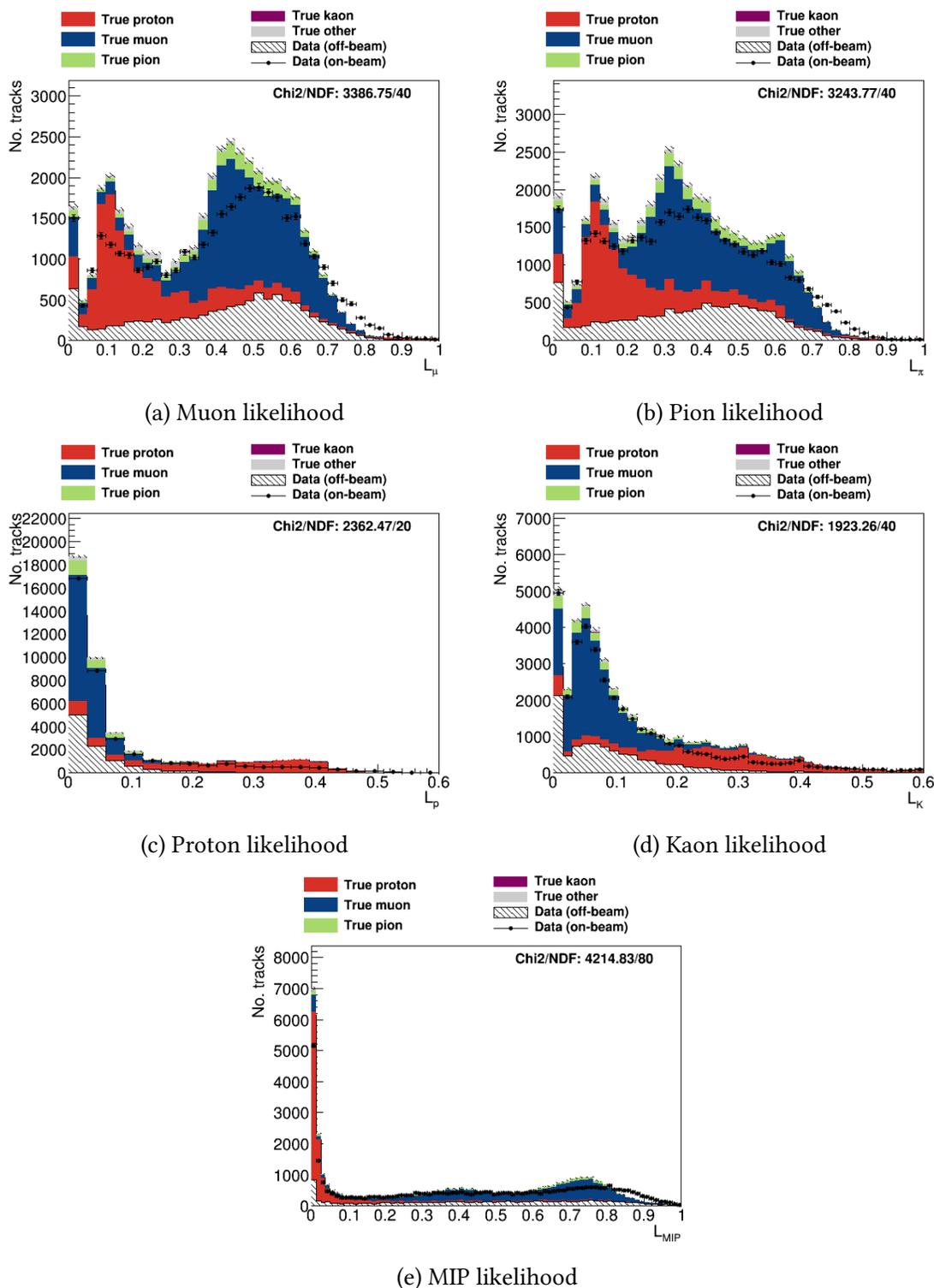


Figure A.5: POT normalised bare likelihoods for tracks under muon, pion, proton, kaon, and MIP assumptions.

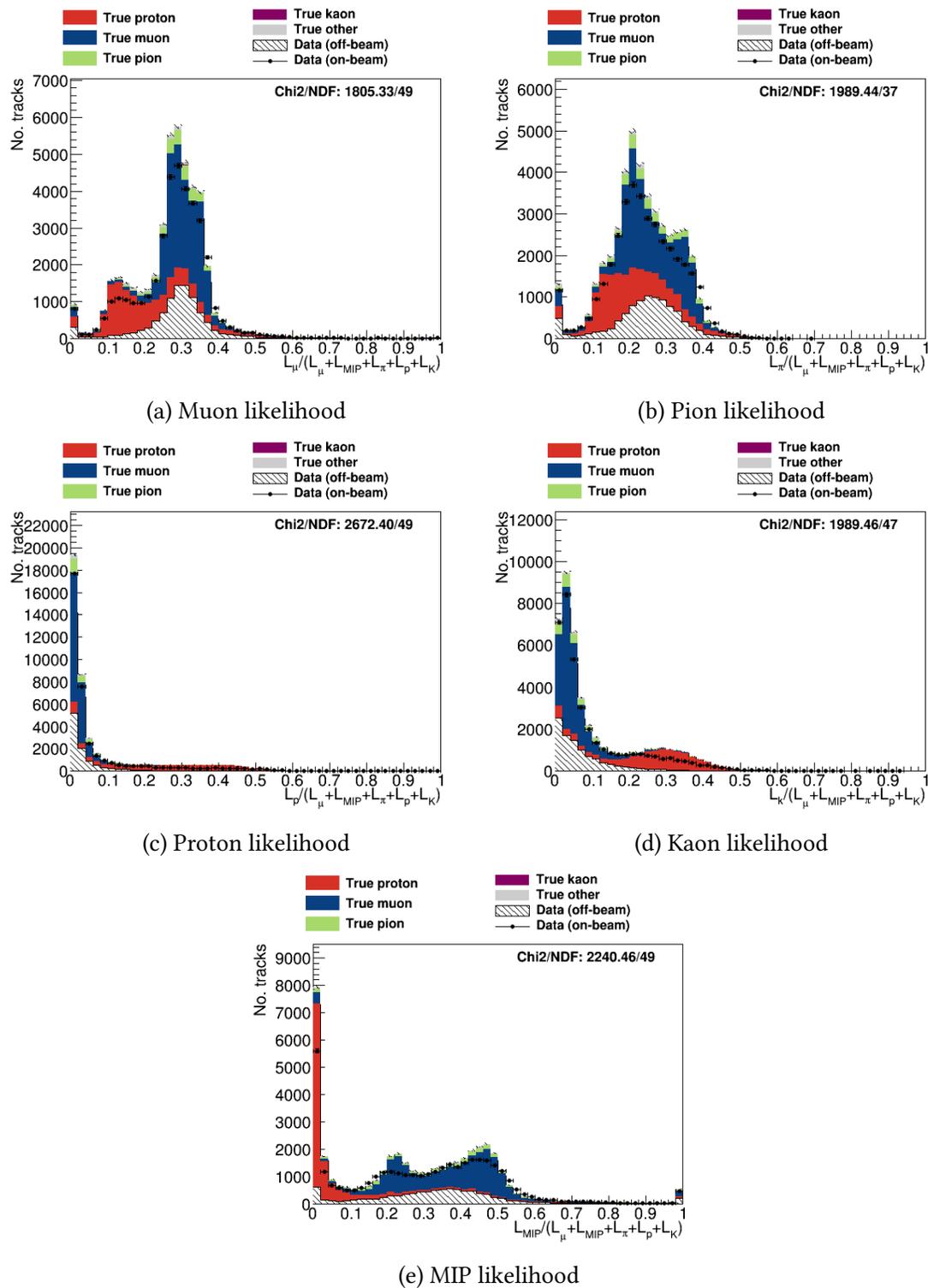


Figure A.6: POT normalised likelihoods for tracks under muon, pion, proton, kaon, and MIP assumptions, normalised from 0 to 1.

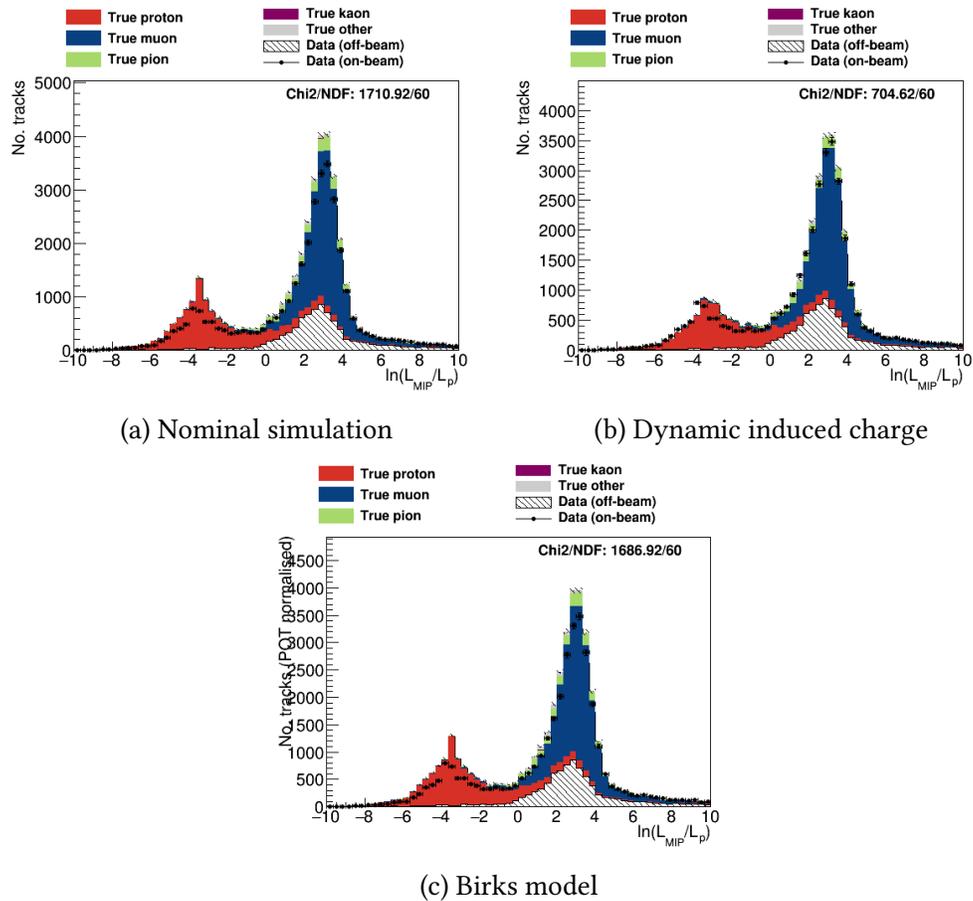


Figure A.7: POT normalised plots of the $\ln(L_{MIP}/L_p)$ ratio for plane 2.

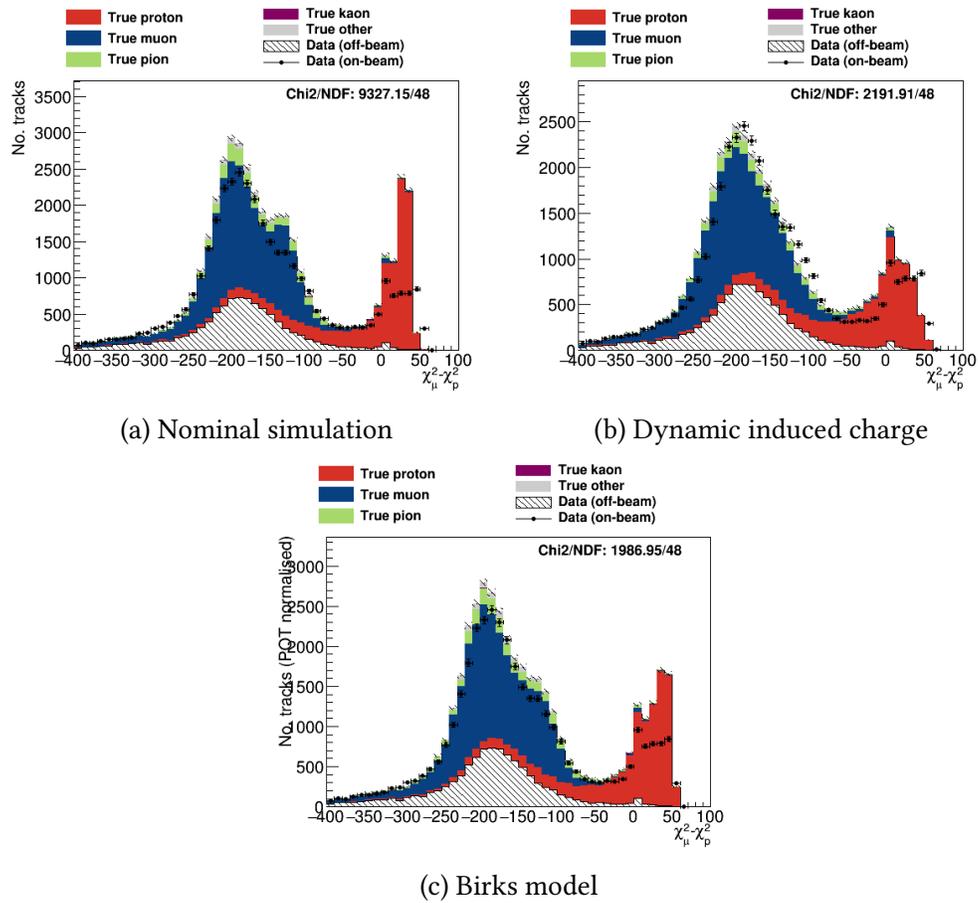


Figure A.8: POT normalised plots of the $\Delta\chi^2_{\mu-p}$ variable for plane 2.

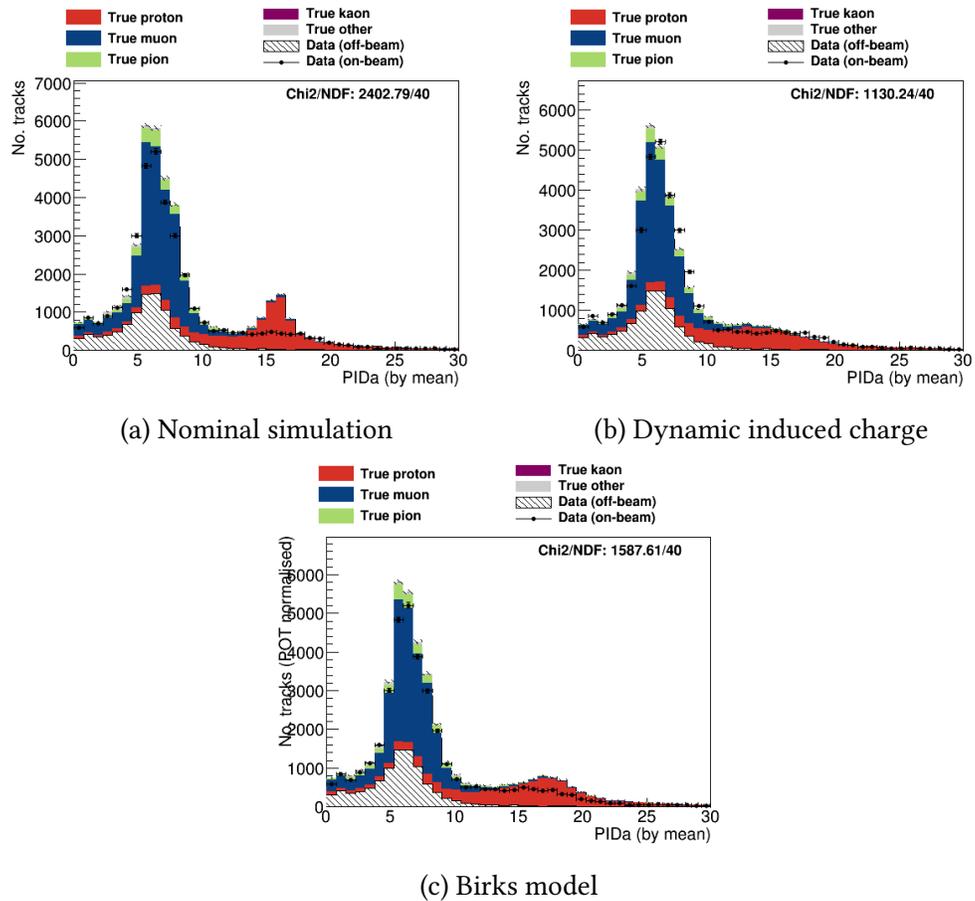


Figure A.9: POT normalised plots of the PIDA-by-mean variable for plane 2.

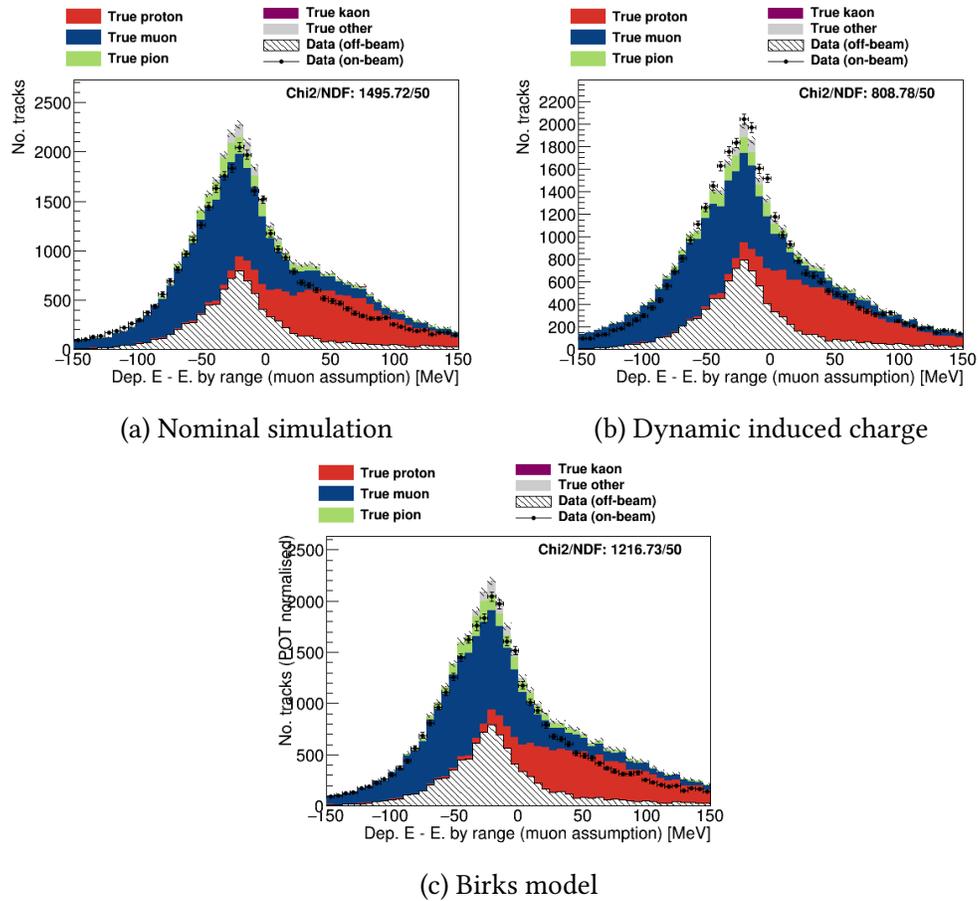


Figure A.10: POT normalised deposited energy on plane 2 versus energy by range under the muon assumption.

Appendix B

Additional Information for the ν_μ CC0 π N p Selection

B.1 A Note on Reconstructed-To-True Object Matching

One tool which MicroBooNE analyses make frequent use of is so called reconstructed-to-true object matching. This is the act of taking a reconstructed object such as a track or shower and finding which true simulated particle this corresponds to.

This is an important concept as it allows for studies of particle resolutions and is necessary for defining a signal.

While simple in concept, this is a somewhat complicated procedure. The reason for this is that due to imperfect reconstruction there are often multiple true particles which are associated to a single reconstructed particle or multiple reconstructed particles associate to a single true particle. This is demonstrated in Figure B.1.

The reconstructed-to-true object matching is performed by taking high level reconstructed objects such as tracks and showers, finding the reconstructed energy depositions which make up these objects, and then finding the true energy depositions which make up each of these. The true particle which deposits the most energy in the reconstructed object is then identified as the

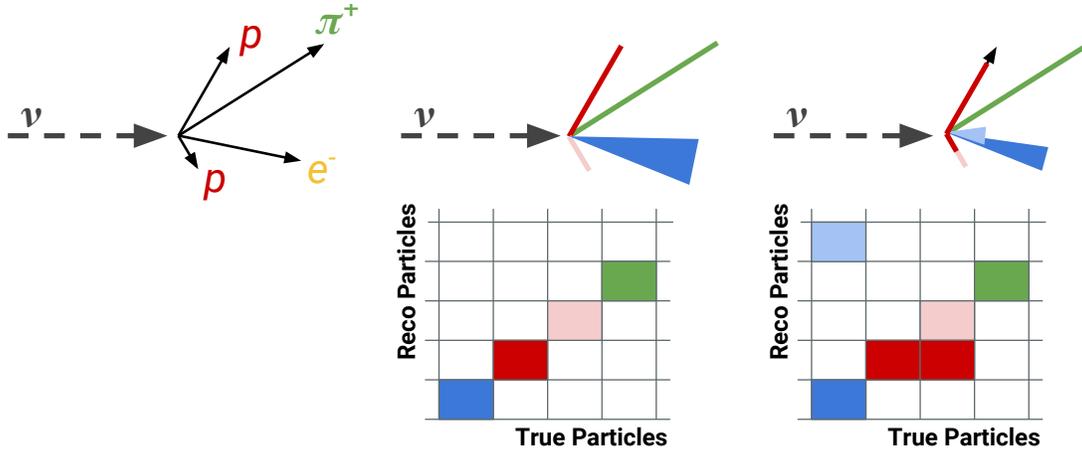


Figure B.1: Cartoon showing the difficulties of matching reconstructed particles to true particles. A set of true particles which leave the interaction would ideally be reconstructed as a series of tracks and showers, but due to imperfect reconstruction multiple reconstructed particles may be associated to a single true particle, and multiple true particles may be associated to a single reconstructed particle.

matched particle, and two quantities are defined to characterise the quality of this match. The purity is defined to be

$$P = \frac{MCParticle_{E_{reco}}^{match}}{Track_E^{match}} \quad (\text{B.1})$$

and the completeness is defined to be

$$C = \frac{MCParticle_{E_{reco}}^{match}}{MCParticle_{E_{total}}^{match}} \quad (\text{B.2})$$

B.2 Kinematic Variables with the DIC Sample

This appendix contains plots produced using the DIC sample. Here, the “physics-less” quantities of the vertex positions of the selected ν_μ CC0 π NP, along with kinematic variables for muons and protons are displayed. These plots are analogous to those found in Chapter 7.5.

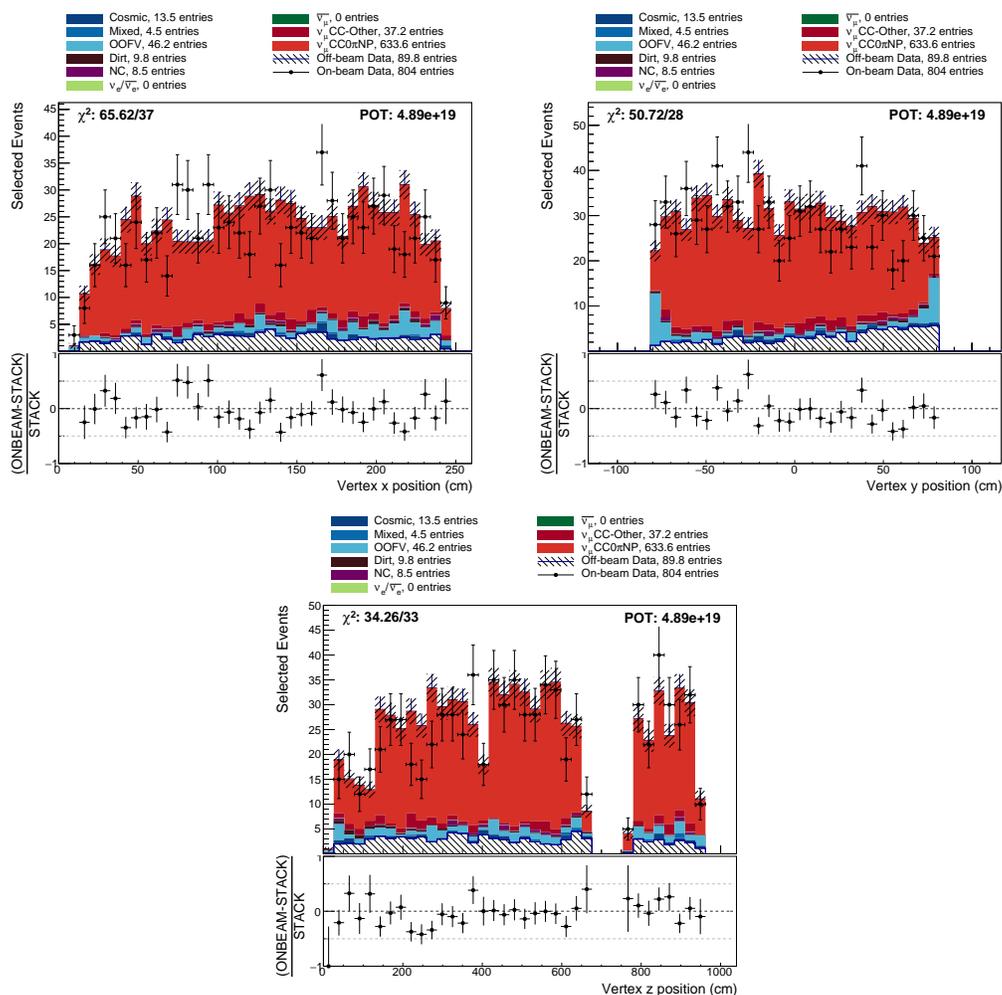


Figure B.2: Reconstructed neutrino vertex positions in the x, y and z dimensions within the MicroBooNE TPC.

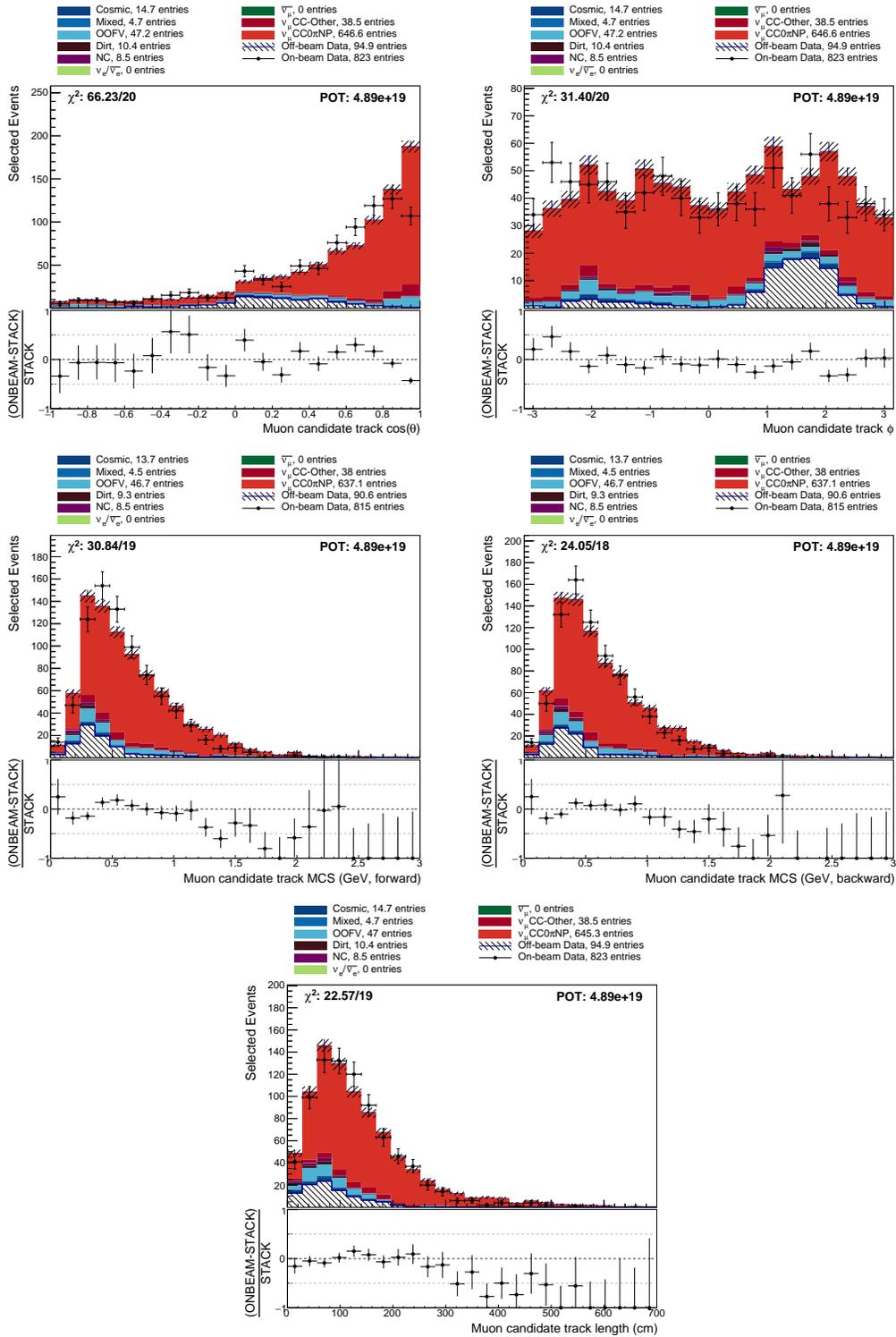


Figure B.3: Kinematic distributions for the selected muon candidate.

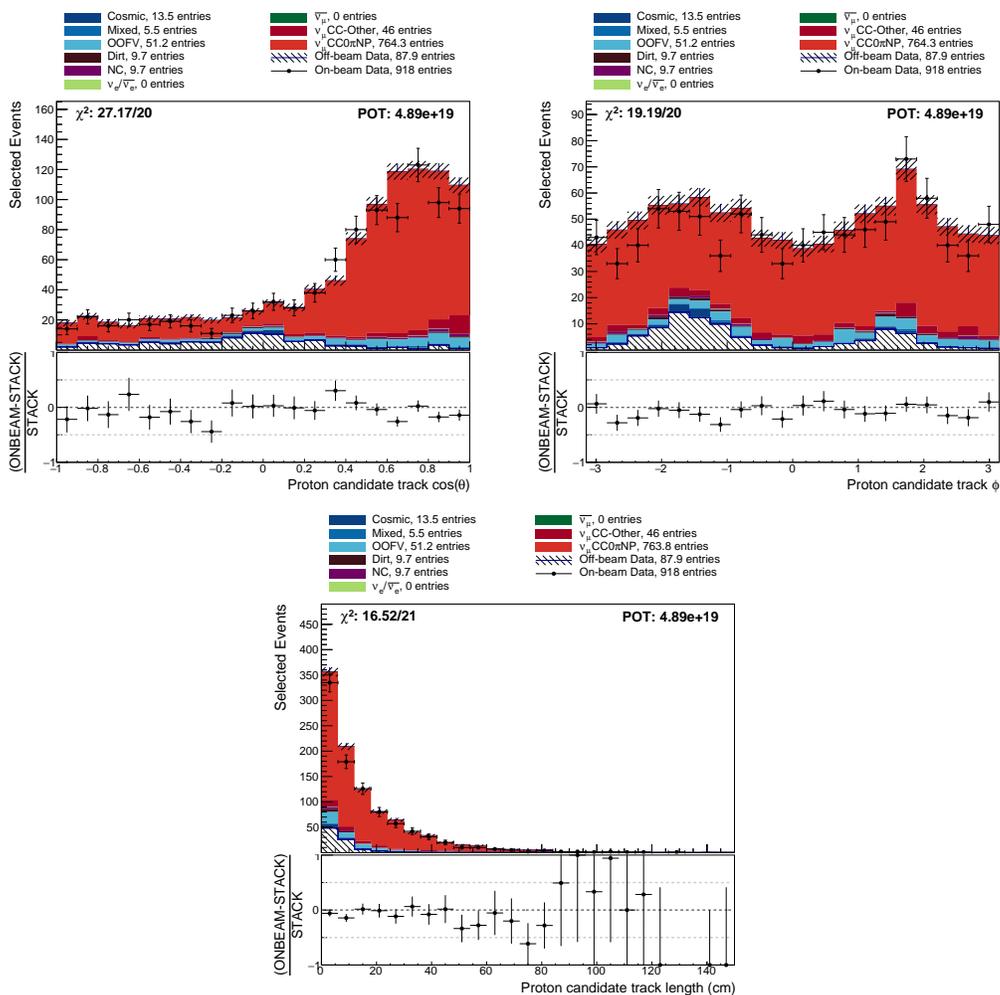


Figure B.4: Kinematic distributions for the selected proton candidates.