

# Measurement of the muon-neutrino charged-current single charged-pion cross-section on argon with the MicroBooNE detector

**Andrew David Smith**

University of Cambridge  
Girton College



This thesis is submitted for the degree of  
*Doctor of Philosophy*  
January 2022



# Abstract

Several important questions remain open in the field of neutrino oscillation physics, including the possibility of CP-violation in the lepton sector, the ordering of the neutrino mass states and the possible existence of sterile neutrinos. At present, the ability to answer these questions is limited by uncertainties on neutrino-nucleus interaction cross-section models. Consequently, it is key that these uncertainties are constrained by precise cross-section measurements made using experimental data. The MicroBooNE experiment utilises a 90-tonne active mass Liquid Argon Time Projection Chamber to image neutrino interactions at the millimetre scale and is ideally suited to measure complex neutrino-argon interactions.

This thesis presents a measurement of the muon-neutrino charged-current single charged pion ( $\text{CC}1\pi^\pm$ ) cross-section on argon using data from MicroBooNE in the Fermilab Booster Neutrino Beam. The total flux-integrated forward-folded cross-section is found to be  $[22.4 \pm 0.9(\text{stat.}) \pm 5.2(\text{syst.})] \times 10^{-41} \text{ cm}^2$ , with an efficiency of  $[18.8 \pm 1.3] \%$  and is consistent with the prediction of the GENIE generator. Additionally, the world's first measurement of the proton-exclusive  $\text{CC}1\pi^\pm$  cross-section is performed with a  $300 \text{ MeV c}^{-1}$  proton momentum threshold. Finally, the differential cross-section is extracted with respect to the muon and pion momenta and directions. The pion momentum measurement on argon is also the first to be made.

To facilitate these measurements, the Pandora pattern recognition software is employed to identify and reconstruct particle trajectories in MicroBooNE data. A key stage of this process is the identification and removal of cosmic-rays that form the main background to all analyses of neutrino interactions. The approach presented in this thesis is capable of removing 46% of such backgrounds at the cost of only 1.7% of neutrino-induced activity.





## Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text. It does not exceed the word limit of 60,000 words prescribed by the Degree Committee for the Faculty of Physics & Chemistry.

Andrew David Smith  
Thursday 6<sup>th</sup> January, 2022



## Contributions

This thesis is organised into three parts. The research contributions made by the author of this thesis to each part are summarised below and detailed in the text.

- Part I: “Neutrinos at the Micro Booster Neutrino Experiment”, provides context for the work presented in the later parts of this thesis but does not include any research work that was contributed directly by the author.
- Part II: “Reconstruction and identification of neutrino interactions in dense cosmic ray environments”, details a scientific software package of which the author of this thesis is a developer and has made contributions at the technical level, in collaboration with others. In particular, the development of the “consolidated approach” described in Section 4.6 is one of the primary technical contributions of the author. The research contributions made to this package are detailed in Chapter 5 and relate to the identification of cosmic rays which form a major background to all analyses of neutrino interactions with MicroBooNE.
- Part III: “Charged-current single charged pion muon-neutrino interactions”, constitutes the main research work presented in this thesis and was performed solely by the author unless specified in the text, with inputs (e.g. data, simulations, calibrations, etc.) provided by the MicroBooNE collaboration.



## Acknowledgements

Firstly, I wouldn't be where I am today without the endless love, encouragement and support that I am incredibly lucky to have from my family — Mum (Pam), Dad (Robert), Nic and Nat. You've all taught me so much and have always been there for me no matter what. Thank you! I'd also like to thank the many friends and family members (including Smiths, Jones, Higgs, McWilliams and Davies) that have kept me sane over the last five years.

My PhD journey can be split into three distinct phases: starting out in Cambridge, my long term attachment at Fermilab, and the time I spent living in Liverpool after returning to the UK. I'll give thanks for each chronologically, but I first owe a special thank you to Lorena Escudero for her continuous guidance through all phases of this endeavour. Independent of the city, country or time zone you've always given me your time and have been an indispensable fount of knowledge and support. Thank you also for reading this thesis and providing feedback while I was performing the analysis. Once this PhD malarkey is official I'm looking forward to giving my Daruma a second eye and celebrating with a kebab on Madingley road.

I began in Cambridge under the supervision of Mark Thomson and thank him for training me to be a better scientist in my first few years. I also learnt so much from John Marshall who had to deal with my stupid physics questions, poorly written pull requests, and **TeXTable** — so I thank you very much for your patience, tuition and for fostering in me a passion for software. I also was lucky to work alongside fellow PhD students Steve, Bono, Joris, Jack and Stefano (many of whom have now graduated). Thank you for making Cambridge an enjoyable place to work. In particular, thanks to Jack and Stefano for the fun we had together in the US.

As part of my PhD, I had the excellent opportunity to spend a year living in Fermilab, IL, USA at the site of the MicroBooNE experiment. I'm very grateful to have made many friends during this time who are unfortunately too numerous to mention by name, but if you ever: played gin sling, ordered tacos through a hatch in KV shots, wore a costume at Site 56, shared a pitcher of Atom Smasher, took a dodgy trip in Desmond (thanks Dom!), used the freezing cold paddling pool, or survived my (very long) pub

quiz — thank you for making that year one of the most enjoyable of my life so far. I’d specifically like to thank Jack, Iker and Nicola for being great housemates, and especially Tom for accompanying me and Rhiannon on all of our adventures, being our personal trainer and a great friend. I’d also like to thank the friends and family that came all the way from the UK to visit us, and especially Soo and Rowan for the road trip around the great lakes — that was a truly unforgettable experience and I’m glad we were able to do it together.

During my time in the states, the Cambridge neutrino group came under the new leadership of my supervisor Melissa Uchida, and the group has since grown to include Leigh Whitehead, Steve Dennis, and fellow PhD students Alex, Karolina, Jing and Philip. Although I haven’t had the opportunity to spend much in-person time with many of you, I’ve thoroughly enjoyed working alongside you, even if it was via Zoom. In particular, I’d like to thank Leigh and Melissa for reading and giving me comments on this thesis — a job of which I am not envious! Special thanks also go to Philip for agreeing to continue with my analysis, I know the work is in good hands and I hope you find everything sufficiently well documented. After returning to the UK, I’ve primarily spent my time living in Liverpool and would also like to thank their neutrino group — especially Costas, Marco and Steve D. (again!) — for including me in trips to the pub, meals out, and for adopting me at conferences. This thesis would not have been possible without the many MicroBooNE collaborators who have offered guidance and advice in reconstruction and cross-section meetings and whose research this work builds upon. In particular, I’d like to thank: Giuseppe Cerati for his tuition during my time at Fermilab, the attendees of  $CC1\pi^\pm$  meetings (especially Matt and Jarek for their insightful comments), and specifically Kirsty Duffy for all of her significant contributions to this analysis. Additionally, thank you to Andy Furmanski and Steve Dytman for serving on the editorial board.

My final thanks go to my partner (lab and otherwise!), fellow PhD student, pool adversary, travel buddy, best friend, family member, and all of the above — Rhiannon. Right from the start, when working on our MPhys project in Manchester, we’ve been really lucky to share this experience with each other. I could not have done it without you! I’m really proud of both of us and I’m eternally grateful for all you have done for me.

At the imminent risk of becoming too soppy, I will simply finish by asking that you don’t kill me when I insist on singing *Dr Jones, Dr Jones, calling Dr Jones — Dr Jones, Dr Jones, wake up now* from now on every morning I wake up before you.

*To Rhiannon*





# Contents

<b>Acronyms</b>	<b>xix</b>
<b>1 Introduction</b>	<b>1</b>
<b>I Neutrinos at the Micro Booster Neutrino Experiment</b>	<b>5</b>
<b>2 Neutrino physics</b>	<b>7</b>
2.1 History of the neutrino . . . . .	7
2.1.1 Prediction and discovery of the neutrino . . . . .	7
2.1.2 Historical anomalies and neutrino oscillations . . . . .	8
2.1.3 The present status . . . . .	10
2.2 Neutrino oscillations . . . . .	11
2.2.1 Mass and flavour eigenstates . . . . .	11
2.2.2 Oscillation probabilities . . . . .	12
2.2.3 The PMNS matrix . . . . .	15
2.2.4 Summary of existing measurements . . . . .	17
2.2.5 Overview of oscillation experiments . . . . .	19
2.2.6 CP violation and the mass ordering . . . . .	21
2.2.7 The importance of cross-sections for oscillations . . . . .	22
2.3 Neutrino interactions . . . . .	23
2.3.1 The weak interactions of neutrinos . . . . .	23
2.3.2 Neutrino interactions with point-like particles . . . . .	25
2.3.3 Quasi-Elastic neutrino-nucleon scattering . . . . .	26
2.3.4 Inelastic neutrino-nucleon scattering . . . . .	28
2.3.5 Neutrino-nucleus interactions . . . . .	33
2.3.5.1 Fermi-motion and Pauli-blocking . . . . .	34
2.3.5.2 Nucleon-nucleon interactions . . . . .	35
2.3.5.3 Final State Interactions . . . . .	35

2.3.6	The role of generators . . . . .	36
2.4	The CC1 $\pi^\pm$ topology . . . . .	36
2.4.1	Definition . . . . .	36
2.4.2	Kinematic parameters . . . . .	37
2.4.3	Existing measurements . . . . .	39
2.4.4	Motivation for CC1 $\pi^\pm$ at MicroBooNE . . . . .	43
<b>3</b>	<b>The MicroBooNE experiment</b>	<b>45</b>
3.1	The Booster Neutrino Beam . . . . .	46
3.2	The MicroBooNE detector . . . . .	49
3.2.1	Charged particles interactions in liquid argon . . . . .	49
3.2.2	Operational principles of the MicroBooNE LArTPC . . . . .	50
3.2.3	The Light Collection System . . . . .	56
3.2.4	Cryogenic system . . . . .	57
3.2.5	The Cosmic-Ray Tagger . . . . .	58
3.2.6	Readout and trigger systems . . . . .	59
3.3	Experimental challenges . . . . .	62
3.3.1	The rate of cosmic rays . . . . .	62
3.3.2	The Space-Charge Effect . . . . .	63
3.3.3	Unresponsive detector regions . . . . .	64
3.4	MicroBooNE simulations . . . . .	65
<b>II</b>	<b>Reconstruction and identification of neutrino interactions in dense cosmic ray environments</b>	<b>67</b>
<b>4</b>	<b>Pandora pattern-recognition in MicroBooNE</b>	<b>69</b>
4.1	Introduction . . . . .	69
4.2	Input and outputs of the Pandora framework . . . . .	70
4.2.1	Inputs to Pandora . . . . .	70
4.2.2	Output from Pandora . . . . .	73
4.3	Typical particle signatures in MicroBooNE . . . . .	74
4.4	Pandora's algorithms for cosmic-ray signatures . . . . .	76
4.4.1	2D clustering . . . . .	76
4.4.2	Track cluster matching . . . . .	77
4.4.3	Delta ray building . . . . .	79
4.4.4	3D Space-point and vertex reconstruction . . . . .	79

4.5	Pandora’s algorithms for neutrino interactions . . . . .	81
4.5.1	Neutrino interaction vertex reconstruction . . . . .	81
4.5.2	Shower cluster matching . . . . .	81
4.5.3	Interaction hierarchy reconstruction . . . . .	84
4.5.4	Track vs. Shower characterisation . . . . .	84
4.6	Neutrino interactions in dense cosmic-ray environments . . . . .	85
4.6.1	The consolidated approach . . . . .	85
4.7	Performance of Pandora in MicroBooNE . . . . .	87
4.7.1	Performance metrics . . . . .	87
4.7.2	Performance on CC-Res $1\mu + 1\pi^+ + Np$ events . . . . .	89
4.8	High-level reconstruction . . . . .	92
4.8.1	Track fitting . . . . .	92
4.8.2	Calibrations . . . . .	92
<b>5</b>	<b>Identifying cosmic-ray activity in MicroBooNE</b>	<b>93</b>
5.1	Classification of reconstructed particles . . . . .	93
5.1.1	Neutrino purity and significance . . . . .	94
5.1.2	Classification scheme . . . . .	96
5.2	Methods of identifying cosmic rays . . . . .	99
5.2.1	Timing information . . . . .	99
5.2.2	Through-going particles . . . . .	101
5.2.3	Contained particles . . . . .	103
5.2.4	Angular information . . . . .	104
5.2.5	Topological information . . . . .	105
5.2.6	Relationships between particles . . . . .	106
5.3	CR-tagging algorithm . . . . .	106
5.3.1	Algorithm logic . . . . .	106
5.3.2	Algorithm performance . . . . .	107
5.3.3	Comments and improvements . . . . .	112
5.4	Neutrino slice identification . . . . .	113
5.4.1	The topological score . . . . .	113
5.4.2	Technical contributions . . . . .	115
5.4.3	Topology-agnostic neutrino identification . . . . .	116

<b>III Charged-current single charged pion muon-neutrino interactions</b>	<b>119</b>
<b>6 Identification of <math>\text{CC1}\pi^\pm</math> interactions in MicroBooNE</b>	<b>121</b>
6.1 Introduction . . . . .	121
6.1.1 Signal definition . . . . .	122
6.1.2 Aims and analysis strategy . . . . .	122
6.2 $\nu_\mu$ CC-inclusive pre-selection . . . . .	125
6.2.1 Overview of the pre-selection . . . . .	126
6.2.2 Events passing the pre-selection . . . . .	127
6.3 Identification of particle species . . . . .	129
6.3.1 The use of calorimetric information . . . . .	129
6.3.2 Features for particle identification . . . . .	132
6.3.3 BDT training and responses . . . . .	142
6.4 Selection of $\text{CC1}\pi^\pm$ events . . . . .	146
6.4.1 Generic selection . . . . .	146
6.4.2 Golden selection . . . . .	158
6.5 Performance . . . . .	159
6.5.1 Event selection performance . . . . .	159
6.5.2 Particle identification performance . . . . .	161
6.6 Estimation of particle momenta . . . . .	163
6.6.1 Overview of energy estimation methods . . . . .	163
6.6.2 Momentum estimation by range . . . . .	163
6.6.3 Muon momentum estimation by MCS . . . . .	167
<b>7 <math>\text{CC1}\pi^\pm</math> cross-section measurement</b>	<b>169</b>
7.1 Cross-section definition . . . . .	169
7.1.1 Flux-integrated cross-section . . . . .	169
7.1.2 Accounting for inefficiencies and imperfect resolution . . . . .	170
7.1.3 Forward-folding versus unfolding . . . . .	172
7.1.4 Total and proton-exclusive cross-sections . . . . .	174
7.2 Treatment of uncertainties . . . . .	176
7.2.1 Multivariate Gaussian parametrisation . . . . .	176
7.2.2 Quantities presented by this analysis . . . . .	177
7.2.3 Comparing a prediction to data . . . . .	177
7.2.4 Techniques for assessing systematic uncertainties . . . . .	179

7.3	Sources of uncertainty . . . . .	181
7.3.1	Statistical uncertainties on data . . . . .	181
7.3.2	Statistical uncertainties on the simulation . . . . .	181
7.3.3	Flux uncertainties . . . . .	182
7.3.4	POT counting uncertainties . . . . .	183
7.3.5	Neutrino interaction cross-section uncertainties . . . . .	184
7.3.6	Detector uncertainties . . . . .	184
7.4	Cross-section measurement . . . . .	188
7.4.1	Binning choice . . . . .	188
7.4.2	Total cross-section extraction . . . . .	192
7.4.3	Proton-exclusive, and differential cross-sections . . . . .	196
7.5	Summary and discussion . . . . .	203
7.5.1	Total cross-section . . . . .	203
7.5.2	Differential cross-sections . . . . .	204
7.5.3	Future improvements . . . . .	205
<b>8</b>	<b>Conclusions</b>	<b>207</b>
	<b>Bibliography</b>	<b>213</b>
<b>A</b>	<b>CC1<math>\pi^\pm</math> cross-section binning selection</b>	<b>223</b>
A.1	Muon kinematics . . . . .	223
A.2	Pion kinematics . . . . .	228
A.3	Muon-pion opening angle . . . . .	233
A.4	Proton multiplicity . . . . .	233



# Acronyms

The following acronyms are defined in the text and are included here for reference. The pages on which each acronym is used are also specified.

**1p-1h** One-particle one-hole. 35

**2p-2h** Two-particle two-hole. 35

**ADC** Analogue-to-Digital Converter. 54, 60

**APA** Anode Plane Assembly. 50, 51, 53, 60

**ASIC** Application Specific Integrated Circuit. 59, 60, 64

**BDT** Boosted Decision Tree. xvi, 129, 132, 136, 142–145, 148–150, 158, 189, 194, 203, 205, 206

**BNB** Booster Neutrino Beam. iii, xiv, 2, 37, 39, 43, 45–48, 55, 59, 61, 62, 65, 74, 90, 127, 128, 132, 159, 160, 162, 174, 182, 192, 196–202

**CC** Charged-Current. xv, xvi, 23–33, 36, 37, 39, 41, 65, 89–91, 106, 121, 125–128, 131, 142–146, 155, 159–161, 188, 208, 223

**CC1 $\pi^\pm$**  Muon-neutrino charged-current single charged pion. iii, x, xiv, xvi, xvii, 4, 7, 36–43, 121–206, 208, 209, 223–235

**Coh** Coherent. 33, 37, 39

**CP** Charge-Parity. iii, xiii, 14, 16, 17, 21

**CPA** Cathode Plane Assembly. 50, 51

**CR** Cosmic Ray. iii, vii, xiv, xv, 3, 20, 62, 66, 67, 74, 76, 79, 81, 85–87, 89–117, 121, 126, 127, 129, 142, 153, 161, 185–187, 208

**CRT** Cosmic-Ray Tagger. xiv, 49, 56, 58, 59, 113, 115, 116

**CV** Central-Value. 179, 180, 184

**DAQ** Data Acquisition. 60–62

**DIS** Deep Inelastic Scattering. 28, 31–33, 35, 37, 43

**DUNE** Deep Underground Neutrino Experiment. 21, 43

**EDM** Event Data Model. 73, 74, 115

**FEM** Front End Module. 60, 61

**FKR** Feynman-Kislinger-Ravndal. 30

**FoM** Figure of Merit. 78

**FSI** Final State Interactions. xiii, 34–37, 65, 170

**FV** Fiducial Volume. 89, 90, 94, 98, 117, 122, 126, 171

**IA** Impulse Approximation. 34

**IO** Inverted-Ordering. 17, 21

**LAr** Liquid Argon. 56, 57, 92, 167

**LArTPC** Liquid Argon Time Projection Chamber. iii, xiv, 3, 4, 39, 43, 45, 49, 50, 53, 55, 59, 70, 115, 207–209

**LCS** Light Collection System. xiv, 49, 56–60, 113, 115, 116

**LFG** Local Fermi-Gas. 34, 35, 41

**LY** Light-Yield. 184, 186, 187

**MC** Monte-Carlo. 65, 133, 134, 137–140, 143–145, 147, 148, 150, 151, 154–159, 194, 205

**MCS** Multiple Coulomb Scattering. xvi, 163, 167, 168

**MEC** Meson Exchange Currents. 35, 41, 65

**MicroBooNE** Micro Booster Neutrino Experiment. iii, vii, ix, x, xiii–xvi, 1–5, 33, 36, 37, 41, 43, 45–66, 69–117, 121–169, 173, 184–188, 193, 194, 203, 205, 207–209



**MiniBooNE** Mini Booster Neutrino Experiment. 1, 39, 41, 42, 45, 65, 188, 207

**MIP** Minimum Ionising Particle. 131–136, 143–145, 164

**MSW** Mikheyev-Smirnov-Wolfenstein. 19

**NC** Neutral-Current. 23–25, 126

**NO** Normal-Ordering. 17, 21

**NuMI** Neutrinos at the Main Injector. 39, 45, 61

**PDF** Parton Distribution Function. 32

**PDF** Probability Density Function. 176, 178–180

**PE** Photoelectron. 116, 187

**PFO** Particle-Flow Object. 73, 78, 79, 81, 82, 84–91, 93–96, 99, 101–112, 116, 135

**PMNS** Pontecorvo, Maki, Nakagawa, and Sakata. xiii, 15, 16

**PMT** Photomultiplier Tube. 51, 56, 57, 60–62, 66, 112, 115–117, 186, 187

**POT** Protons-On-Target. xvii, 47, 48, 127, 128, 159, 160, 171, 183, 192, 193, 195, 208

**QE** Quasi-Elastic. xiii, 26, 28, 33, 35–37, 41, 65

**Res** Resonant. xv, 28–30, 33, 35, 37, 41, 43, 65, 89–91

**RFG** Relativistic Fermi-Gas. 34, 41

**SBN** Short Baseline Neutrino. 45, 46

**SBND** Short Baseline Near Detector. 45, 46

**SCE** Space-Charge Effect. xiv, 63, 66, 92, 102, 112, 129, 130, 148, 163, 184–186, 203

**SM** Standard Model. 1, 7, 10, 14, 23, 207

**SSM** Standard Solar Model. 8, 9

**SVM** Support Vector Machine. 81, 84, 113, 114, 126, 136, 139, 155

**SW** Sanford-Wang. 65, 182

**TB** Trigger Board. 61

**TPB** Tetraphenyl butadiene. 56

**TPC** Time Projection Chamber. 48, 49, 51, 52, 56–60, 62–66, 90, 96, 99–103, 106, 110, 112–115, 124, 126, 127, 142, 148, 163, 185–187

**VUV** Vacuum Ultraviolet. 50, 56

*“Not everything that can be counted counts,  
and not everything that counts can be counted.”*

— Albert Einstein ... possibly



# Chapter 1

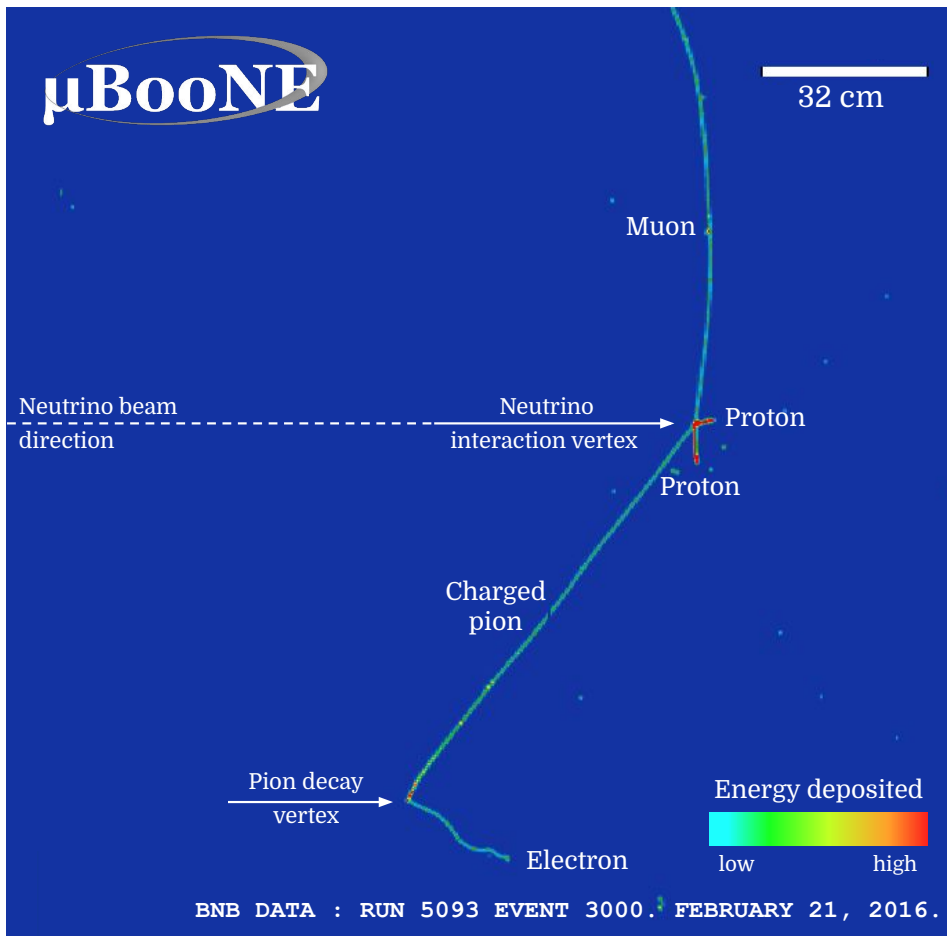
## Introduction

On the microscopic scale, our current best understanding of the universe is described by the Standard Model (SM) of particle physics. This model catalogues the known fundamental particles and describes their interactions, with the exclusion of gravity. Of all the particles of the SM, neutrinos are arguably the most elusive as they rarely interact. This makes them challenging to observe, despite being the second most abundant in our universe after the photon. For example, billions of neutrinos from the Sun reach every square centimetre of the Earth's surface each second, but the vast majority pass freely through to the other side of the globe without any interaction. In contrast, the visible photons from the Sun can be stopped by a piece of paper (observed as a shadow).

Despite their elusive nature, over the last 65 years neutrinos from various natural and human-made sources have been detected by numerous experiments around the globe. Through this data, it is well established that there exist three flavours of neutrino (electron  $\nu_e$ , muon  $\nu_\mu$ , and tau  $\nu_\tau$ ) that participate in the weak interactions of the SM. Additionally, in contrast to the assumptions of the SM, neutrino flavours are observed to change over time. Many of the parameters that govern this phenomenon, known as *neutrino oscillations*, have been measured experimentally and the determination of the remaining parameters is one of the primary aims of the field. The history, properties and interactions of neutrinos are discussed in Chapter 2.

Although the three-neutrino paradigm is broadly successful in describing the results of neutrino oscillation experiments, anomalous results from the LSND experiment [1] and the Mini Booster Neutrino Experiment (MiniBooNE) [2] do not appear to fit into this picture [3]. In particular, MiniBooNE observed an excess of low-energy  $\nu_e$  in a beam of primarily  $\nu_\mu$  that was not consistent with the  $\nu_\mu \rightarrow \nu_e$  oscillations seen by other experiments. The successor Micro Booster Neutrino Experiment (MicroBooNE) [4] is

the focus of this thesis and has the primary goal of confirming or denying this *low-energy excess* signal. MicroBooNE started taking data in October 2015 and will publish its first results regarding this anomaly in the near future. The MicroBooNE detector, described in Chapter 3, is capable of taking high-resolution (millimetre scale) digital images of the particles produced in neutrino interactions. Consequently, MicroBooNE can readily distinguish between the types of particles produced and can measure their properties with fine detail. An example of one such image is shown in Figure 1.1.



**Figure 1.1:** An example image of a candidate neutrino interaction from the Booster Neutrino Beam at MicroBooNE. The particle labels are the author’s interpretation of the image. A beam of neutrinos is aimed at the detector which is filled with liquid argon. Neutrinos enter from the left of the image but can not be detected directly. A neutrino interacts with an argon nucleus to produce other particles which leave trails along their trajectories which are detected as the coloured lines in the figure. Each particle species leaves a distinct pattern which is used to infer the flavour of the neutrino involved, and the nature of its interaction with the argon.

The images taken by early particle detectors, known as bubble chambers, were photographs taken on film. These images would be projected onto a desk, scanned by eye to identify particle trajectories, and analysed by hand using rulers and protractors. The identification of particles, their species and their properties from raw images is known generally as *reconstruction*. Typically, a single person may be able to process  $\sim 200$  images per day in this way [5]<sup>1</sup>. As MicroBooNE has recorded *millions* of digital images, each containing multiple particle trajectories, the automation of this step is crucial to the success of the experiment.

Although pattern recognition comes naturally to the human eye, it is highly non-trivial to design a robust algorithmic process to achieve the same results computationally. One of the primary approaches used by MicroBooNE (and other current and future Liquid Argon Time Projection Chamber (LArTPC) experiments) is the Pandora pattern-recognition software package, described in Chapter 4. Pandora was created by the Cambridge particle physics group, and is now developed primarily by UK institutions. The contributions of the author of this thesis to Pandora involve the development and integration of this software for MicroBooNE.

Another major challenge for any analysis of neutrinos in MicroBooNE are cosmic rays (CRs). These particles continuously bombard the Earth’s surface and are produced when particles from outer space interact in the atmosphere. As cosmic rays are so common (the rate of cosmic-ray muons at MicroBooNE is  $\sim 100 \text{ m}^{-2} \text{ s}^{-1}$  [6]) and neutrino interactions are so rare, MicroBooNE will only save an image if sufficient activity occurs in the detector at the same moment as neutrinos from the beam are expected to pass through (described in Section 3.2.6). Nevertheless, a typical image contains  $\sim 20$  CR-muons, but only one in  $\sim 6$  saved images contain a neutrino interaction [7]. Additionally, even if a neutrino interaction is recorded, it can commonly be obscured by nearby cosmic rays which can make reconstruction challenging. The work presented in Chapter 5 describes an algorithm developed to identify and remove cosmic-ray activity using Pandora. This is a crucial step for all Pandora-based analyses on MicroBooNE and other existing and near-future surface-based LArTPC experiments.

---

<sup>1</sup>Some undergraduate physics courses still have students analyse these images in this way to learn about particle interactions — and to appreciate the advancements made in the last  $\sim 50$  years!

In order to achieve the goals of current and future neutrino oscillation experiments, it is critical to be able to accurately predict the rate at which neutrinos reaching a detector will interact, and to predict the properties of the resulting particles. These predictions are given by the *cross-section* of the relevant neutrino interactions, discussed in Chapter 7. Cross-sections are needed to produce simulations of neutrino interactions in a detector at an appropriate rate to allow for experimental analysis methodologies to be developed and tested. Cross-section predictions are typically derived from theoretical models of neutrino interactions, which are validated and tuned using experimental data. Commonly, it is not possible for a measurement (e.g. one of the neutrino oscillation parameters) to be entirely independent of the cross-section models used for the simulation. Consequently, any uncertainty in the cross-section prediction will result in an uncertainty in the quantity measured. Ultimately, cross-section prediction uncertainties are constrained using experimental data.

The neutrino-argon interactions relevant to MicroBooNE can occur via a wide variety of processes, discussed in Chapter 2. The primary analysis of this thesis (presented in Part III), is a measurement of the rate of neutrino interactions in MicroBooNE that result in a muon ( $\mu$ ), a charged pion ( $\pi^\pm$ ), and any number of protons or neutrons - known as the muon-neutrino charged-current single charged pion (CC1 $\pi^\pm$ ) channel. Figure 1.1 shows a CC1 $\pi^\pm$  candidate image. This result is compared to the cross-section prediction used by MicroBooNE's simulation and can be used to constrain uncertainties for future neutrino oscillation experiments. Previous data of the CC1 $\pi^\pm$  channel on argon is limited and so this measurement is of importance for future LArTPC experiments. Additionally, the measurement of the charged-pion momentum and proton multiplicity presented in this thesis will, when published, constitute a world's first.



## Part I

# Neutrinos at the Micro Booster Neutrino Experiment



# Chapter 2

## Neutrino physics

As described in Chapter 1, it is widely accepted that there are three flavours of neutrinos that participate in the weak interactions of the Standard Model. Additionally, experimental data shows the existence of neutrino flavour oscillations. In this chapter, Section 2.1 describes the historical context that lead to this paradigm, Section 2.2 describes neutrino oscillations on a more mathematical footing and discusses the governing parameters that have been (and will be) measured experimentally. Section 2.3 discusses the weak interactions of neutrinos that are most relevant for neutrino oscillation experiments. Finally, Section 2.4 describes the  $CC1\pi^\pm$  interaction channel (introduced in Chapter 1) and provides an overview of the existing measurements of these interactions.

### 2.1 History of the neutrino

#### 2.1.1 Prediction and discovery of the neutrino

In his famous letter of 1930 beginning “Dear Radioactive Ladies and Gentlemen” [8], Pauli proposed a new particle in light of recent results from  $\beta$ -decay experiments. Prior to his proposal, it was understood that (among other processes) a radioactive isotope can decay by emitting a single particle, called a  $\beta$ -ray — now known to be the electron or positron depending on the particle’s electric charge. In this case, the conservation of energy stipulates that the  $\beta$ -rays must all be emitted with the same energy. However, experimental evidence [9] indicated the spectrum of  $\beta$ -ray energies was continuous. Pauli hypothesised that an additional particle must be produced in the decay that is invisible to the experiment as a “desperate remedy” to save the conservation of energy. In 1933,

Fermi formalised [10] this theory in the language of quantum mechanics and named this new particle *the neutrino*.

The (anti)neutrino was first detected experimentally by Reines and Cowan in 1956 [11]. In order to obtain a sufficiently strong neutrino flux, their detector was situated close to a nuclear reactor in South Carolina, USA (as neutrinos are produced as a by-product of nuclear fission). The experiment used the inverse beta decay process, where an antineutrino interacts with a proton to produce a neutron and a positron, which produced signals that could be observed by the experiment <sup>1</sup>.

In 1962, Lederman, Schwartz and Steinberger discovered a second kind, or *flavour*, of neutrino [12]. Their experiment collided 15 GeV protons with a beryllium target to produce pions which decayed in flight to produce neutrinos. A 13.5 m thick iron shield was used to stop all particles but the neutrinos which reached a spark chamber where they could be detected. When these neutrinos interacted, a muon <sup>2</sup> was observed instead of an electron as would be expected for the neutrinos found so far. These neutrinos are now named muon neutrinos ( $\nu_\mu$ ), and those initially discovered were (anti)electron neutrinos ( $\nu_e$ ). When the tau ( $\tau$ ) lepton was discovered in 1975 [13], a third neutrino flavour,  $\nu_\tau$ , was implied and was observed in 2001 by the DONUT experiment [14].

The properties of the  $Z$  boson (one of the particles that mediate the weak interactions in which neutrinos participate) were measured in 1989 using the LEP collider at CERN [15]. The results constrained the number of light active neutrinos to be three ( $\nu_e$ ,  $\nu_\mu$ ,  $\nu_\tau$ ) and this paradigm is widely accepted today. It is still possible that further types of neutrinos could exist that do not participate in weak interactions, known as *sterile neutrinos*. However, at present, there is no conclusive evidence to confirm their existence.

### 2.1.2 Historical anomalies and neutrino oscillations

In 1968, the Homestake experiment (led by Davis and Bahcall) measured the rate of electron neutrinos on Earth that were produced in the Sun [16]. The result was found to be approximately a third of the expected rate, as predicted by Standard Solar Model (SSM). This provided the first evidence of what became known as the *solar neutrino*

---

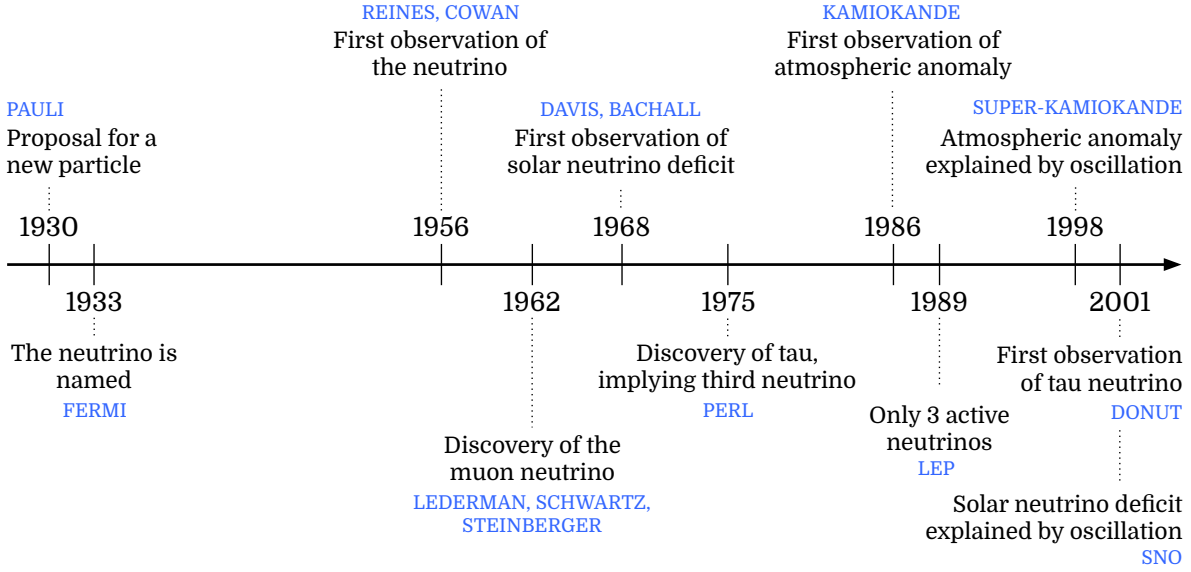
<sup>1</sup>The antineutrino and the positron are the *antiparticles* of the neutrino and electron respectively. Antiparticles behave very similarly to their particle counterparts but have opposite physical charges.

<sup>2</sup>A muon,  $\mu$ , is a fundamental particle that is nearly identical to an electron,  $e$ , but with a larger mass. There also exists the tau particle,  $\tau$ , which is of the same type with a yet higher mass. Collectively, these particles are known as the charged leptons.

*problem.* A similar deficit was later observed by the Kamiokande experiment in 1989 [17] and the GALLEX [18] and SAGE [19] experiments in 1992 and 1994, confirming the problem. In 1998, the Super-Kamiokande experiment measured the neutrino flux as a function of the direction of the neutrinos to confirm that they were indeed of solar origin [20]. The SSM was sufficiently well understood that the observed deficit could not be explained entirely by uncertainties in the predicted rate [16]. This instead implied a missing piece in our understanding of the neutrino.

In 1967, Pontecorvo had proposed that neutrinos might change between flavours  $\nu_e \leftrightarrow \nu_\mu$  [21] — known as *neutrino flavour oscillations*. If this were true, then the solar neutrino problem may be resolved if some portion of the  $\nu_e$  produced in the Sun were found to oscillate to other flavours, which were not observed by the existing experiments. In 2001, the SNO experiment simultaneously measured the rate of solar electron neutrinos and the rate of solar neutrinos of all flavours. The former was consistent with the existing results, and the latter was consistent with the SSM prediction [22]. This provided strong evidence for the existence of solar neutrino oscillations.

The Kamiokande experiment published in 1986 the first observation of a possible second anomaly, this time relating to neutrinos that originate in the atmosphere [23]. In 1998, its successor experiment, Super-Kamiokande, measured the rate of atmospheric  $\nu_\mu$  and  $\nu_e$  as a function of the neutrino energy and the angle of the incoming neutrino [24]. Those neutrinos originating from above the detector travel a significantly shorter distance than those originating from below, which were produced in the atmosphere on the opposite side of the globe. The rate of  $\nu_e$  was consistent with the expectation, whereas a deficit in the rate of  $\nu_\mu$  was found that depended on the distance travelled by the neutrino. This measurement provided strong evidence for the existence of atmospheric neutrino oscillations. Together, the measurements from Super-Kamiokande and SNO resolved the atmospheric and solar neutrino anomalies, and the 2015 Nobel Prize in Physics was awarded to the key contributors for the discovery of neutrino oscillations. A summary of some of the key historical discoveries in neutrino physics is shown in Figure 2.1.



**Figure 2.1:** A timeline of some of the key dates in neutrino physics in the period 1930-2001.

### 2.1.3 The present status

Within the Standard Model, neutrinos are assumed to be massless as this was consistent with historical observations before neutrino oscillations were discovered. As detailed in Section 2.2, the three neutrino flavours ( $\nu_e$ ,  $\nu_\mu$ ,  $\nu_\tau$ ) are most generally expressed as quantum mechanical superpositions of three states of definite mass ( $m_1$ ,  $m_2$ ,  $m_3$ ). These states propagate with frequencies dependent on their mass and can interfere to give rise to the oscillation phenomenon. If the masses were equal then this interference would not occur. Hence, the observation of neutrino oscillations implies that the neutrino masses are different and so at least one must be non-zero, in contradiction to the assumptions of the SM. In particular, it is the difference between the squares of the neutrino masses ( $m_2^2 - m_1^2$ ) and ( $m_3^2 - m_2^2$ ) that govern the rate of the oscillations observed. In addition to these two squared mass differences, neutrino oscillations depend on a further four parameters that must be determined experimentally. The details of these parameters are also discussed in Section 2.2.

In the 20 years since SNO's results, neutrinos from numerous natural and human-made sources have been measured over a wide range of energies and distances. In general, these results are well explained by oscillations between three neutrino flavours ( $\nu_e$ ,  $\nu_\mu$ ,  $\nu_\tau$ ), and the majority of the parameters that govern these oscillations have been measured. However, the value of one of the remaining unknown parameters, known as *delta-CP* ( $\delta_{CP}$ ), has important consequences. This parameter governs the differences between the

oscillations of neutrinos and antineutrinos<sup>3</sup> and may help explain why the universe is made primarily of matter and not antimatter [25]. The T2K experiment has already presented the first hint that such differences may exist [26, 27]. Additionally, the order of the squared masses ( $m_1^2 < m_2^2 < m_3^2$  or  $m_3^2 < m_1^2 < m_2^2$ ) is still to be identified, defined by the parameter  $\text{sign}(m_3^2 - m_1^2)$ . One of the primary aims of the future generation of neutrino oscillation experiments is to measure these parameters [28].<sup>4</sup>

## 2.2 Neutrino oscillations

### 2.2.1 Mass and flavour eigenstates

In quantum mechanics, a particle can be described by a state vector,  $|\psi(t)\rangle$ , whose time-evolution is governed by the Schrödinger equation [29]:

$$i \frac{d}{dt} |\psi(t)\rangle = \hat{H} |\psi(t)\rangle \quad (2.1)$$

Here,  $\hat{H}$  is the Hamiltonian operator of the system. Natural units are used throughout this chapter unless otherwise specified.

If  $|\psi(t)\rangle$  represents a single free particle and is an eigenstate of  $\hat{H}$ , then the corresponding eigenvalue,  $E$ , is the total energy of the particle:

$$\hat{H} |\psi(t)\rangle = E |\psi(t)\rangle. \quad (2.2)$$

Hence, the time evolution of this state over the interval  $t \rightarrow t + \Delta t$  is given by a plane wave solution:

$$|\psi(t + \Delta t)\rangle = \exp(-iE\Delta t) |\psi(t)\rangle. \quad (2.3)$$

In the case of the three active neutrinos, there are three distinct eigenstates of the Hamiltonian operator, known as the mass states:  $|\nu_i(t)\rangle$ , where  $i = 1, 2, 3$ . Each state represents a neutrino with mass,  $m_i$ , has a corresponding energy eigenvalue,  $E_i$ . Neutrino

<sup>3</sup>Each neutrino flavour,  $\nu_e, \nu_\mu, \nu_\tau$  has an associated antineutrino,  $\bar{\nu}_e, \bar{\nu}_\mu, \bar{\nu}_\tau$ .

<sup>4</sup>The other primary goals in the field are to determine the absolute neutrino masses (as currently only the squared mass differences are known) and to determine if neutrinos and antineutrinos are equivalent (known as Majorana particles) or distinct (known as Dirac particles). However, these questions can not be answered by studying neutrino oscillations and so are not described in detail.

oscillations arise when one considers the possibility that these mass states are not the same as the flavour states that participate in the weak interaction:  $|\nu_\alpha(t)\rangle$ , where  $\alpha = e, \mu, \tau$ . Throughout this section, Latin indices  $(i, j, k)$  indicate mass eigenstates, whereas Greek indices  $(\alpha, \beta, \gamma)$  indicate flavour eigenstates. As the Hamiltonian is a Hermitian operator, the mass eigenstates form an orthonormal basis,  $\langle \nu_i | \nu_j \rangle = \delta_{ij}$ , where  $\delta_{ij}$  is the Kronecker delta. Consequently, at the time of the production of the neutrino,  $t = 0$ , the flavour states can be represented as linear combinations of the mass states by:

$$|\nu_\alpha\rangle = \sum_{i=1}^3 U_{\alpha i}^* |\nu_i\rangle, \quad (2.4)$$

where  $U_{\alpha i}$  are the coefficients of a unitary matrix,  $U^\dagger = U^{-1}$ <sup>5</sup>, and for brevity the notation  $|\nu_i\rangle = |\nu_i(0)\rangle$ ,  $|\nu_\alpha\rangle = |\nu_\alpha(0)\rangle$  is used. The matrix  $U$  must be unitary in order for the total probability of the neutrino being in *any* of the flavour states to be unity. This condition also implies the flavour states form an orthonormal basis,  $\langle \nu_\alpha | \nu_\beta \rangle = \delta_{\alpha\beta}$ . Consequently, the mass states can also be expressed in terms of the flavour states as:

$$|\nu_i\rangle = \sum_{\alpha=e,\mu,\tau} U_{\alpha i} |\nu_\alpha\rangle. \quad (2.5)$$

The time evolution of the flavour states can be calculated from Equations 2.3-2.5 as:

$$\begin{aligned} |\nu_\alpha(t)\rangle &= \sum_{i=1}^3 U_{\alpha i}^* \exp(-iE_i t) |\nu_i\rangle \\ &= \sum_{\gamma=e,\mu,\tau} \left( \sum_{i=1}^3 U_{\alpha i}^* \exp(-iE_i t) U_{\gamma i} \right) |\nu_\gamma\rangle \end{aligned} \quad (2.6)$$

### 2.2.2 Oscillation probabilities

A neutrino initially produced in a flavour state  $|\nu_\alpha\rangle$  is measured in a flavour state  $|\nu_\beta\rangle$  after a time  $t$  with probability given by [29]:

$$P_{\alpha \rightarrow \beta}(t) \equiv |\langle \nu_\beta | \nu_\alpha(t) \rangle|^2. \quad (2.7)$$

---

<sup>5</sup> $U^\dagger$  denotes the Hermitian conjugate and  $U^{-1}$  denotes the matrix inverse.



The expression can be evaluated using Equation 2.6 and the orthonormality of the flavour states to give:

$$\begin{aligned}
 P_{\alpha \rightarrow \beta}(t) &= \left| \sum_{i=1}^3 U_{\alpha i}^* \exp(-i E_i t) U_{\beta i} \right|^2 \\
 &= \sum_{i=1}^3 \sum_{j=1}^3 U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^* \exp(-i (E_i - E_j) t).
 \end{aligned} \tag{2.8}$$

The relativistic energy-momentum relation can be used to write the energies of the  $|\nu_i\rangle$  states in terms of the corresponding masses:

$$\begin{aligned}
 E_i &= \sqrt{\underline{p}^2 + m_i^2} \\
 &\approx |\underline{p}| + \frac{m_i^2}{2|\underline{p}|}.
 \end{aligned} \tag{2.9}$$

Here the Taylor approximation holds for ultra-relativistic particles ( $|\underline{p}| \gg m_i$ ) which is appropriate for the neutrinos studied in oscillation experiments. The momentum,  $\underline{p}$ , is assumed to be the same for all mass states and it is common to further define  $E \equiv |\underline{p}|$  as the energy of the neutrino in the massless approximation. With this approximation made, the difference in the energies of two mass states is given by:

$$E_i - E_j = \frac{m_i^2 - m_j^2}{2E} = \frac{\Delta m_{ij}^2}{2E} \tag{2.10}$$

Here,  $\Delta m_{ij}^2 \equiv m_i^2 - m_j^2$  is known as a mass-splitting.

Additionally, under the ultra-relativistic approximation the neutrino travels at near the speed of light such that the distance travelled,  $L$ , by the neutrino in time,  $t$ , is simply given (in natural units) by  $L = t$  and so Equation 2.8 becomes:

$$\begin{aligned}
 P_{\alpha \rightarrow \beta}(L/E) &= \sum_{i=1}^3 \sum_{j=1}^3 A_{ij\alpha\beta} \exp(-2i\phi_{ij}), \\
 A_{ij\alpha\beta} &\equiv U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*, \\
 \phi_{ij}(L/E) &\equiv \frac{1}{4} \Delta m_{ij}^2 \frac{L}{E}.
 \end{aligned} \tag{2.11}$$

Here,  $\phi_{ij}$  controls the phase of the oscillations (with a frequency driven by the mass-splitting  $\Delta m_{ij}^2$ ) and  $A_{ij\alpha\beta}$  controls the amplitude of the oscillations (driven by the elements of  $U$ ). Although Equation 2.11 contains complex terms, the oscillation probability is real. This can be seen more clearly in the form:

$$\begin{aligned}
 P_{\alpha \rightarrow \beta}(L/E) &= \delta_{\alpha\beta} - 2 \sum_{i=1}^2 \sum_{j=i+1}^3 \left[ \underbrace{2 \operatorname{Re}(A_{ij\alpha\beta}) \sin^2(\phi_{ij})}_{\text{CP conserving}} - \underbrace{\operatorname{Im}(A_{ij\alpha\beta}) \sin(2\phi_{ij})}_{\text{CP violating}} \right].
 \end{aligned} \tag{2.12}$$

Antineutrinos are related to neutrinos by a Charge-Parity (CP) transformation<sup>6</sup>, and antineutrino oscillations are also described by Equation 2.12, with the replacement  $U_{\alpha i} \rightarrow U_{\alpha i}^{CP} = U_{\alpha i}^*$  [29]:

$$\begin{aligned}
 \operatorname{Re}(A_{ij\alpha\beta}) &\rightarrow \operatorname{Re}(A_{ij\alpha\beta}^{CP}) = \operatorname{Re}(A_{ij\alpha\beta}), \\
 \operatorname{Im}(A_{ij\alpha\beta}) &\rightarrow \operatorname{Im}(A_{ij\alpha\beta}^{CP}) = -\operatorname{Im}(A_{ij\alpha\beta}).
 \end{aligned} \tag{2.13}$$

Hence, the form of Equation 2.12 can be understood as the sum of three terms:

1.  $\delta_{\alpha\beta}$ , is the probability if no oscillations are present.
2. The term with amplitudes  $\operatorname{Re}(A_{ij\alpha\beta})$  is the component of the oscillation probability that is *symmetric* under a CP transformation (CP conserving), and is the same for neutrinos and antineutrinos.
3. The final term with amplitudes  $\operatorname{Im}(A_{ij\alpha\beta})$  is *antisymmetric* under a CP transformation (CP violating), and is different for neutrinos and antineutrinos.

---

<sup>6</sup>Both charge and parity are required as Standard Model neutrinos are left-chiral whereas antineutrinos are right-chiral.

### 2.2.3 The PMNS matrix

The matrix  $U$  is known as the Pontecorvo, Maki, Nakagawa, and Sakata (PMNS) matrix. In general a  $3 \times 3$  complex matrix such as  $U$  has 18 parameters, however 9 are constrained by the unitarity conditions (one for each combination of  $\alpha$  and  $\beta$ ):

$$(UU^\dagger)_{\alpha\beta} = \delta_{\alpha\beta}. \quad (2.14)$$

As shown in Equation 2.15, there is freedom to modify the elements,  $U_{\alpha i}$ , by an overall phase without changing the physical observable,  $A_{ij\alpha\beta}$ .

$$\begin{aligned} U_{\alpha i} &\rightarrow U'_{\alpha i} = e^{i(\zeta_\alpha - \zeta_i)} U_{\alpha i} \\ A_{ij\alpha\beta} &\rightarrow A'_{ij\alpha\beta} = A_{ij\alpha\beta} \end{aligned} \quad (2.15)$$

This transformation can be interpreted more generally as a modification of the overall phase of the mass and flavour eigenstates in Equation 2.5:

$$\begin{aligned} |\nu_i\rangle &\rightarrow |\nu_i\rangle' = e^{i\zeta_i} |\nu_i\rangle, \\ |\nu_\alpha\rangle &\rightarrow |\nu_\alpha\rangle' = e^{i\zeta_\alpha} |\nu_\alpha\rangle, \end{aligned} \quad (2.16)$$

The 6 phases,  $\zeta_i$  ( $i = 1, 2, 3$ ) and  $\zeta_\alpha$  ( $\alpha = e, \mu, \tau$ ) are not independent. An arbitrary state,  $|\psi\rangle$ , can be expressed in either the mass or the flavour basis and the transformations of the basis states (given by Equation 2.16) must result in the same transformation  $|\psi\rangle \rightarrow |\psi\rangle'$  in either basis:

$$\begin{aligned} |\psi\rangle &= \sum_{i=1}^3 \psi_i |\nu_i\rangle = \sum_{\alpha=e,\mu,\tau} \psi_\alpha |\nu_\alpha\rangle, \\ &\downarrow \\ |\psi\rangle' &= \sum_{i=1}^3 e^{i\zeta_i} \psi_i |\nu_i\rangle = \sum_{\alpha=e,\mu,\tau} e^{i\zeta_\alpha} \psi_\alpha |\nu_\alpha\rangle. \end{aligned} \quad (2.17)$$

Consequently, with the restriction of Equation 2.17, there is a freedom to define 5 of these 6 phases in a manner that is not observable via neutrino oscillations. Hence,  $U$  can be parametrised by  $18 - 9 - 5 = 4$  real values <sup>7</sup>.

---

<sup>7</sup>Two additional parameters are in principle observable via non-oscillatory phenomena, but are not considered here. Namely, if neutrinos are Majorana fermions (a particle that is its own antiparticle) then two phases can not be absorbed into the neutrino field [30].

In order to describe the form of these 4 parameters, consider the special case where all elements of  $U$  are real,  $U_{\alpha i} = U_{\alpha i}^*$ . In this case, the PMNS matrix describes an orthogonal transformation  $U^{-1} = U^T$  which is parametrised by 3 angles. Additionally, if all elements of  $U$  are real then  $\text{Im}(A_{ij\alpha\beta}) = 0$  and hence neutrino oscillations would be unchanged by a CP transformation. In the general case,  $U_{\alpha i} \neq U_{\alpha i}^*$ , the remaining parameter is a complex phase,  $\delta_{CP}$ , that acts to introduce an imaginary component,  $\text{Im}(A_{ij\alpha\beta}) \neq 0$ , and hence sets the extent of CP violation. The PMNS matrix is most commonly parametrised by the three mixing angles  $\theta_{12}$ ,  $\theta_{23}$ , and  $\theta_{13}$ , and the CP-violating phase  $\delta_{CP}$  as [29]:

$$\begin{array}{ccc} \begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} & = & \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}, \\ \text{flavour} & & U \text{ (PMNS)} \quad \text{mass} \end{array} \quad (2.18)$$

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{CP}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (2.19)$$

where  $c_{ij} = \cos(\theta_{ij})$  and  $s_{ij} = \sin(\theta_{ij})$ . With this notation in place, neutrino oscillations are governed by six parameters which must be determined experimentally:  $\Delta m_{21}^2$ ,  $\Delta m_{31}^2$ ,  $\theta_{12}$ ,  $\theta_{23}$ ,  $\theta_{13}$  and  $\delta_{CP}$ . The squared mass differences are not independent, as shown by Equation 2.20, however the choice to omit the parameter  $\Delta m_{32}^2$  in particular from this list is arbitrary, and the ordering of the indices,  $\Delta m_{ij}^2 = -\Delta m_{ji}^2$ , is chosen by convention.

$$\Delta m_{12}^2 + \Delta m_{23}^2 + \Delta m_{31}^2 = 0. \quad (2.20)$$

The parameters to which a given experiment is sensitive depends on the oscillation channel ( $\alpha \rightarrow \beta$ ) probed, and the baseline-to-energy ratio,  $L/E$ . By restoring  $c$  and  $\hbar$ , the phases  $\phi_{ij}$  in Equation 2.11 can be expressed in real-world units as [29]:

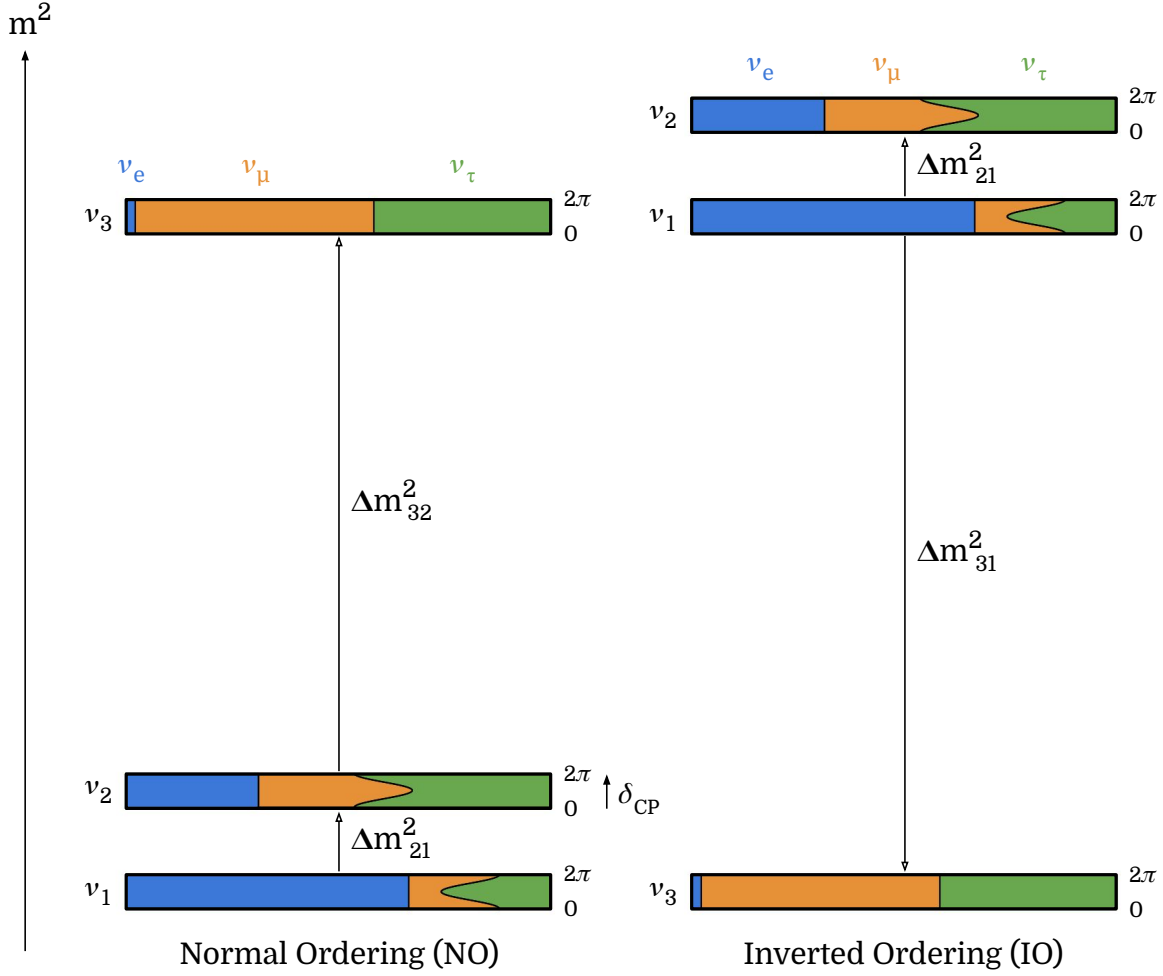
$$\begin{aligned} \phi_{ij} &= \left( \frac{\text{fm GeV}}{4\hbar c} \right) \times \frac{\Delta m_{ij}^2}{\text{eV}^2 c^{-4}} \times \frac{L}{\text{km}} \times \frac{\text{GeV}}{E} \\ &\approx 1.27 \times \frac{\Delta m_{ij}^2}{\text{eV}^2 c^{-4}} \times \frac{L}{\text{km}} \times \frac{\text{GeV}}{E}. \end{aligned} \quad (2.21)$$

### 2.2.4 Summary of existing measurements

The majority of the six parameters that govern neutrino oscillations have been accurately measured, however the CP-violating phase  $\delta_{CP}$ , and the sign of one of the mass-splittings,  $\text{sign}(\Delta m_{31}^2)$  are not yet known. Table 2.1 gives the result of a fit to global neutrino data [31], which is also shown visually in Figure 2.2. The Normal-Ordering (NO) case,  $\text{sign}(\Delta m_{31}^2) = +1$ , is defined such that  $m_1^2 < m_2^2 < m_3^2$ . In the Inverted-Ordering (IO) case,  $\text{sign}(\Delta m_{31}^2) = -1$ , and  $m_3^2 < m_1^2 < m_2^2$ . The fit presented in Table 2.1 prefers NO with a  $2.5\sigma$  statistical significance, and the  $3\sigma$  range of  $\delta_{CP}$  in this case is  $(0.71 \rightarrow 1.99)\pi$  [31]. Here, the phase,  $\delta_{CP}$ , is defined on the interval  $0 \leq \delta_{CP} < 2\pi$ , and the CP violating component of the oscillation probability is zero when  $\delta_{CP} = 0$  or  $\pi$ , and is maximal for  $\delta_{CP} = \pi/2$  or  $3\pi/2$ .

Parameter	Mass hierarchy	Best fit	+1 $\sigma$	-1 $\sigma$
$\Delta m_{21}^2 / 10^{-5} \text{eV}^2 c^{-4}$		7.50	0.22	0.20
$ \Delta m_{31}^2  / 10^{-3} \text{eV}^2 c^{-4}$	NO	2.55	0.02	0.03
	IO	2.45	0.02	0.03
$\sin^2(\theta_{12}) / 10^{-1}$		3.18	0.16	0.16
$\theta_{12} / ^\circ$		34.3	1.0	1.0
$\sin^2(\theta_{23}) / 10^{-1}$	NO	5.74	0.14	0.14
	IO	5.78	0.10	0.17
$\theta_{23} / ^\circ$	NO	49.26	0.79	0.79
	IO	49.46	0.60	0.97
$\sin^2(\theta_{13}) / 10^{-2}$	NO	2.200	0.069	0.062
	IO	2.225	0.064	0.070
$\theta_{13} / ^\circ$	NO	8.53	0.13	0.12
	IO	8.58	0.12	0.14
$\delta_{CP} / \pi$	NO	1.08	0.13	0.12
	IO	1.58	0.15	0.16

**Table 2.1:** The result of a global fit to neutrino oscillation data in the three-neutrino framework [31]. The best fit parameters are given for Normal-Ordering (NO),  $\Delta m_{31}^2 > 0$ , and Inverted-Ordering (IO),  $\Delta m_{31}^2 < 0$ , of the neutrino masses. The parameters  $\Delta m_{21}^2$ , and  $\theta_{12}$  are not dependent on the mass hierarchy.



**Figure 2.2:** A visualisation of the oscillation parameters in Table 2.1. The vertical axis represents the squared neutrino mass, and each multicoloured band is a neutrino mass state ( $\nu_1, \nu_2, \nu_3$ ) assuming normal ordering (left) and inverted ordering (right). The vertical axis is arbitrary and not to scale but the property  $\Delta m_{21}^2 \ll |\Delta m_{31}^2|$  is reflected. For each mass state,  $i$ , the width of the internal coloured bands represents the contribution  $(U_{\alpha i} U_{\alpha i}^*)$  of each flavour state,  $\alpha$  ( $\nu_e$  in blue,  $\nu_\mu$  in orange,  $\nu_\tau$  in green). This contribution is shown for all possible values of the phase parameter with  $\delta_{CP} = 0$  at the bottom and  $\delta_{CP} = 2\pi$  at the top of each band. As this parameter is not yet known, the curved line between the orange and green bands represents the uncertainty in the contribution of  $\nu_\mu$  and  $\nu_\tau$  to  $\nu_1$  and  $\nu_2$  as a result of  $\delta_{CP}$ .

### 2.2.5 Overview of oscillation experiments

The values of the parameters listed in Table 2.1 are the culmination of the work of many experiments from around the globe. Each experiment is sensitive to some subset of the full parameter list, typically driven by the source of the neutrinos under study. This section provides a broad overview of different neutrino sources that are relevant for neutrino oscillation experiments and the parameters to which those experiments are most sensitive.

#### Solar neutrinos

Solar neutrino experiments, measure electron neutrinos produced in thermonuclear reactions in the Sun with energies from 0.1 to 20 MeV [32]. These neutrinos undergo many oscillation periods before reaching Earth <sup>8</sup> so solar experiments can (naively) only measure the average oscillation probability, with  $\langle \sin^2(\phi_{ij}) \rangle = 1/2$  [33]. For a  $\nu_e$  disappearance measurement ( $\nu_e \rightarrow \nu_e$ ), Equation 2.12 becomes:

$$\langle P_{e \rightarrow e} \rangle = 1 - \frac{1}{2}(\cos^4(\theta_{13}) \sin^2(2\theta_{12}) + \sin^2(2\theta_{13})) \approx 1 - \frac{1}{2} \sin^2(2\theta_{12}). \quad (2.22)$$

This approximation would become an equality in the limit  $\theta_{13} \rightarrow 0$ . Consequently,  $\theta_{12}$  is sometimes referred to as the solar mixing angle.

This description is naive as it neglects neutrino interactions that occur within the Sun, which can be modelled by adding potential term to the vacuum Hamiltonian [32]. In general,  $\nu_e$  experience a different potential to  $\nu_\mu$  and  $\nu_\tau$  as solar neutrinos are typically insufficiently energetic to undergo interactions that produce a  $\mu$  or  $\tau$  [32]. These *matter effects* modify the oscillation probability such that the neutrinos leaving the Sun correspond to a different mixture of mass-states than naively assumed in vacuo, described by the Mikheyev-Smirnov-Wolfenstein (MSW) effect [32]. As a result, the disappearance rate of  $\nu_e$  at the Earth corresponds to an averaged oscillation probability that has an additional dependence on  $\Delta m_{21}^2$  [33], which is hence known as the solar mass splitting. Importantly, matter effects allow for the sign of  $\Delta m_{21}^2$  to be determined using a disappearance measurement [33] <sup>9</sup>, which is not possible using vacuum oscillations (as  $\text{Im}(A_{ij\alpha\alpha}) = 0$ , and  $\sin^2(\phi_{ij}) = \sin^2(-\phi_{ij})$  in Equation 2.12).

<sup>8</sup>The Sun-Earth distance is  $\sim 150 \times 10^6$  km [33]. Using the parameters in Table 2.1, the first oscillation maxima for  $E = 10$  MeV neutrinos occur at  $\sim 10$  km (for  $\Delta m_{12}^2$ ) and  $\sim 300$  km (for  $|\Delta m_{31}^2| \approx |\Delta m_{32}^2|$ ).

<sup>9</sup>By convention the mass state with the largest  $\nu_e$  component is chosen to be  $\nu_1$ .

## Reactor neutrinos

Reactor neutrino experiments commonly measure  $\bar{\nu}_e$  produced as a byproduct in commercial nuclear power plants at typical energies of  $\sim 1-8$  MeV [32]. Long-baseline ( $\sim 200$  km) reactor experiments are sensitive to the first oscillation maxima for mass-splittings around  $\Delta m^2 \sim 10^{-5} \text{ eV}^2 c^{-4}$ . The value of  $\Delta m_{21}^2$  in Table 2.1 is primarily driven by the results of the KamLAND experiment [31], which also supplements the measurements of  $\theta_{12}$  from solar experiments.

Short-baseline reactor experiments often measure  $\bar{\nu}_e$  disappearance over typical distances of  $\sim 1-2$  km, and hence are sensitive to the first oscillation maximum for  $\Delta m^2 \sim 10^{-3} \text{ eV}^2 c^{-4}$ . In this case the oscillations driven by the solar mass splitting are small, and so Equation 2.12 can be approximated as:

$$\begin{aligned} P_{\bar{e} \rightarrow \bar{e}}(L/E) &\approx 1 - \sin^2(2\theta_{13}) \left[ \cos^2(\theta_{12}) \sin^2\left(\frac{1}{4}\Delta m_{31}^2 \frac{L}{E}\right) + \sin^2(\theta_{12}) \sin^2\left(\frac{1}{4}\Delta m_{32}^2 \frac{L}{E}\right) \right] \\ &\approx 1 - \sin^2(2\theta_{13}) \sin^2\left(\frac{1}{4}\Delta m_{atm}^2 \frac{L}{E}\right) \end{aligned} \quad (2.23)$$

The final approximation holds in the limit  $|\Delta m_{31}^2| = |\Delta m_{32}^2| \equiv \Delta m_{atm}^2$ , which is known as the atmospheric mass splitting. Consequently, short-baseline reactor neutrino experiments are sensitive to  $\theta_{13}$  and  $|\Delta m_{31}^2|$  (or equivalently  $|\Delta m_{32}^2|$ ).

## Atmospheric neutrinos

Neutrinos are produced when primary cosmic-ray particles interact in the atmosphere. One of the main neutrino production channels is through the decay of charged pions (e.g.  $\pi^+ \rightarrow \mu^+ \nu_\mu$ ), and the subsequent muon decays (e.g.  $\mu^+ \rightarrow e^+ \nu_e \bar{\nu}_\mu$ ). Atmospheric neutrinos are produced with a wide energy spectrum, and the  $\sim 0.1$  GeV to  $\sim 100$  GeV region is typically relevant to oscillation studies [31]. The distance travelled by a neutrino before reaching a detector on (or near) the surface of the Earth depends on its zenith angle. Those produced in the atmosphere above the detector typically travel  $\sim 20$  km before reaching the Earth's surface, whereas those produced on the opposite side of the globe must also traverse the Earth's diameter of  $\sim 12,700$  km [33]. Consequently, atmospheric neutrino experiments are most sensitive to oscillations driven by  $\Delta m_{atm}^2$ , as the name implies. Typically, atmospheric neutrino experiments can measure (and distinguish between) muon and electron neutrino interactions. Unlike the electron (anti)neutrino disappearance measurements, a measurement involving muon neutrinos (e.g.  $\nu_\mu \rightarrow \nu_\mu$ ) is sensitive to  $\theta_{23}$ .



## Accelerator neutrinos

The final source of neutrinos discussed in this section are those produced using a particle accelerator, which are primarily  $\nu_\mu$  or  $\bar{\nu}_\mu$ . This source is of particular interest to the analysis presented in this thesis and more details are provided in Section 3.1. Accelerator neutrino experiments typically study muon neutrino disappearance ( $\nu_\mu \rightarrow \nu_\mu$ ) or electron neutrino appearance ( $\nu_\mu \rightarrow \nu_e$ ) channels in the few GeV neutrino energy range and are broadly categorised by their baseline. Long-baseline ( $L \approx 300 \rightarrow 1500$  km) experiments are designed with a value of  $L/E$  that is sensitive to the maxima of  $\Delta m_{atm}^2$  oscillations, and are primarily sensitive to  $\theta_{23}$ . These experiments can in principle measure  $\delta_{CP}$  and the mass ordering, however those that are currently operational do not currently provide sufficient sensitivity. Short-baseline experiments ( $L \sim 500$  m) typically search for oscillations that do not fit into the 3-neutrino picture.

### 2.2.6 CP violation and the mass ordering

As indicated by Table 2.1, the CP violating phase,  $\delta_{CP}$  and the ordering of the neutrino masses (NO or IO) are not yet known. Measuring these parameters is one of the major goals of the next generation of long-baseline accelerator neutrino experiments. However, the table does reflect the constraints on these parameters from existing data. In particular, recent results from the T2K collaboration rule out CP conserving values of  $\delta_{CP}$  at the 95% confidence level [27]. The value of  $\delta_{CP}$  is of particular importance as CP-violation is one of the necessary conditions for the hypothesised processes that took place in the early universe to produce the asymmetry between matter and antimatter observed today<sup>10</sup>.

The asymmetry between the neutrino ( $\nu_\mu \rightarrow \nu_e$ ) and antineutrino ( $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ ) oscillation probabilities depends on  $\delta_{CP}$ , and can be approximated (to leading order in  $\Delta m_{21}^2$ ) as [28]:

$$\frac{P_{\mu \rightarrow e} - P_{\bar{\mu} \rightarrow \bar{e}}}{P_{\mu \rightarrow e} + P_{\bar{\mu} \rightarrow \bar{e}}} \approx \sin(\delta_{CP}) \left( \frac{\cos(\theta_{23}) \sin(2\theta_{12})}{\sin(\theta_{23}) \sin(\theta_{13})} \right) \left( \frac{1}{4} \Delta m_{21}^2 \frac{L}{E} \right) + \text{matter effects.} \quad (2.24)$$

Deep Underground Neutrino Experiment (DUNE) will measure beam neutrinos that have travelled through the Earth over a distance of  $L = 1300$  km [28]. At this baseline, matter effects (as described for solar neutrinos) introduce an additional asymmetry

<sup>10</sup>Although CP-violation has been observed in the quark-sector, the amount is insufficient to explain the matter-antimatter asymmetry observed [25].

between the  $\nu_\mu \rightarrow \nu_e$  and  $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$  oscillation probabilities that is dependent on mass ordering. This asymmetry will be sufficiently large that it can be disambiguated from asymmetries due to  $\delta_{CP}$ . Hence, both the mass ordering and  $\delta_{CP}$  can be measured [28].

### 2.2.7 The importance of cross-sections for oscillations

For an oscillation experiment with a detector at a baseline  $L$  from a neutrino source, the rate of neutrino interactions that will be observed with neutrino flavour,  $\beta$ , and energy  $E_0 < E < E_1$  is given by:

$$r_\beta(L) = T \sum_{\alpha=e,\mu,\tau} \sum_f^{\text{final states}} \int_{E_0}^{E_1} \int_{\chi_f} \left[ \phi_\alpha(E, L) \cdot P_{\alpha \rightarrow \beta}(L/E) \cdot \frac{d\sigma_\beta}{d\underline{X}_f}(E, \underline{X}_f) \cdot \epsilon_f(\underline{X}_f) \right] d\underline{X}_f dE. \quad (2.25)$$

Here,  $T$  is a normalisation constant depending on the detector material and volume,  $\alpha$  labels the neutrino flavours produced by the source,  $\phi_\alpha$  is the unoscillated flux of neutrinos reaching the detector with flavour  $\alpha$  and  $P_{\alpha \rightarrow \beta}$  is the oscillation probability. Additionally,  $f$  labels the possible sets of final-state particles produced in a neutrino interaction (e.g.  $\mu + p$ ),  $\underline{X}_f$  represents the kinematic properties of these particles (e.g. momenta, directions) and  $\chi_f$  is the phase-space spanned by these kinematic variables.  $d\sigma_\beta/d\underline{X}_f$  is the (differential) *cross-section* which measures the probability of a neutrino interaction with flavour  $\beta$  that result in the final state  $f$ . Finally,  $\epsilon_f$  is the detection efficiency for this final state.

In order to extract the oscillation probability from Equation 2.25, experiments count the number of neutrinos seen in their detector, which is related to  $r_\beta(L)$ <sup>11</sup>. As a result, the flux, cross-section and detection efficiency must be well-understood. Commonly, accelerator neutrino experiments use multiple detectors to probe oscillations: a near detector, situated close to the neutrino source ( $L \approx 0 \Rightarrow P_{\alpha \rightarrow \beta}(L/E) \approx \delta_{\alpha\beta}$ ), measures the unoscillated rate of neutrinos from the beam and a far detector measures the neutrinos after they have propagated. With this approach, near detector measurements can be used to constrain uncertainties on the flux and cross-section. Nevertheless, the extracted oscillation probability will in general have a dependence on the cross-section models used. Consequently, it is important that these models have been well-validated against dedicated cross-section measurements using external data. Additionally, the

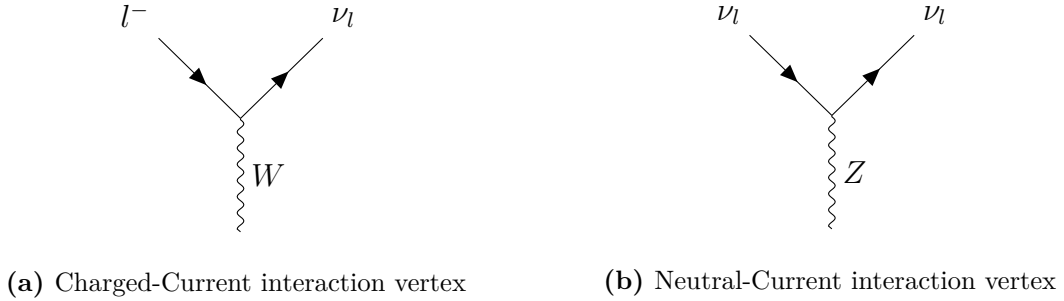
<sup>11</sup>In practice, the exposure of the detector to the source, the resolution with which the neutrino energy can be reconstructed, and any backgrounds need to be accounted for.

neutrino energy,  $E$ , can only ever be estimated using the final-state particles that are visible to the detector. Any energy that is “invisible” (i.e. carried by a particle that is neutral or below the detection threshold) must be estimated using neutrino interaction and detector models.

## 2.3 Neutrino interactions

### 2.3.1 The weak interactions of neutrinos

Standard Model neutrinos only participate in weak interactions involving the exchange of  $W^\pm$  or  $Z^0$ , known as Charged-Current (CC) and Neutral-Current (NC) interactions respectively. The interaction vertices involving neutrinos are shown in Figure 2.3.



**Figure 2.3:** Diagrams for the fundamental neutrino interaction vertices. The label,  $l$ , represents any charged lepton ( $e, \mu, \tau$ ), and  $\nu_l$  is the corresponding neutrino.

In the Standard Model Lagrangian, the terms relevant to weak interactions have the form [34]:

$$\begin{aligned}\mathcal{L}_{CC} &= -\frac{g}{2\sqrt{2}} \left( j_{CC}^\mu W_\mu + j_{CC}^{\mu\dagger} W_\mu^\dagger \right), \\ \mathcal{L}_{NC} &= -\frac{g}{2\cos(\theta_W)} (j_{NC}^\mu Z_\mu).\end{aligned}\tag{2.26}$$

Here,  $W$  and  $Z$  are the fields corresponding to the  $W^\pm$  and  $Z^0$  bosons respectively,  $g$  is the weak coupling constant,  $\theta_W$  is the Weinberg angle, and  $j_{CC}/j_{NC}$  are the charged and neutral currents. Each fermion line in Figure 2.3 contributes a term to the currents

and those in  $j_{CC}$  relevant to leptonic interactions are [34]:

$$j_{CC}^\mu = \sum_{\alpha}^{\text{flavours}} \bar{\nu}_\alpha \gamma^\mu (1 - \gamma^5) l_\alpha + \text{quark terms.} \quad (2.27)$$

Here,  $\nu$  and  $l$  are the neutrino and charged lepton fields respectively,  $\alpha$  labels the lepton flavour, and  $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$  where  $\gamma^\mu$  ( $\mu = 0, 1, 2, 3$ ) are the Dirac gamma matrices. The terms in  $j_{NC}$  relevant to leptonic interactions are [34]:

$$j_{NC}^\mu = \sum_{\alpha}^{\text{flavours}} \left[ \underbrace{\bar{l}_\alpha \gamma^\mu (g_V^l - g_A^l \gamma^5) l_\alpha}_{\text{charged leptons}} + \underbrace{\bar{\nu}_\alpha \gamma^\mu (g_V^\nu - g_A^\nu \gamma^5) \nu_\alpha}_{\text{neutrinos}} \right] + \text{quark terms.} \quad (2.28)$$

The vector couplings,  $g_V^{l/\nu}$ , are coefficients of the components of the current of the form  $\bar{\psi}\gamma^\mu\psi$  (that transform as a Lorentz vector) whereas the axial couplings,  $g_A^{l/\nu}$ , are coefficients of the components of the form  $\bar{\psi}\gamma^\mu\gamma^5\psi$  (that transform as an axial-vector). The neutrino vector and axial couplings are equal,  $g_V^\nu = g_A^\nu = \frac{1}{2}$ , whereas the charged lepton couplings differ,  $g_A^l = -\frac{1}{2} = g_V^l - 2\sin^2(\theta_W)$  [34].

For both CC and NC, the terms involving neutrinos have a vector minus axial-vector ( $V - A$ ) form. This structure is closely related to the concept of chirality. A given fermion field,  $\psi$ , can be decomposed into its chiral components,  $\psi_L$  and  $\psi_R$  (known as the left and right-handed components) via the projection operators  $P_L$  and  $P_R$  [29]:

$$\psi = \underbrace{\frac{1}{2}(1 - \gamma^5)}_{\equiv P_L} \psi + \underbrace{\frac{1}{2}(1 + \gamma^5)}_{\equiv P_R} \psi = \psi_L + \psi_R. \quad (2.29)$$

All terms in the charged and neutral currents involving neutrinos have the  $V - A$  form:

$$\bar{\psi}\gamma^\mu(1 - \gamma^5)\phi = 2\bar{\psi}\gamma^\mu P_L\phi = 2\bar{\psi}_L\gamma^\mu\phi_L. \quad (2.30)$$

Here,  $\psi$  and  $\phi$  represent generic fermion fields. Consequently, the weak interaction only couples to left-handed neutrinos (and right-handed antineutrinos).

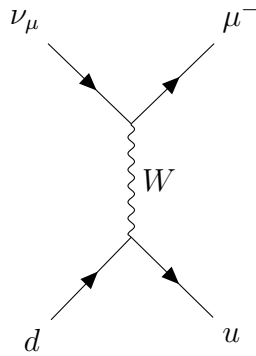
### 2.3.2 Neutrino interactions with point-like particles

Accelerator experiments detect neutrinos by their interactions with a target medium. These interactions result in particles that are visible to the detector and the presence of the neutrino is inferred. Many modern accelerator neutrino experiments typically use nuclear targets (e.g.  $^{12}\text{C}$  or  $^{40}\text{Ar}$ ) with neutrino energies in the  $\sim 100\text{ MeV}$  to  $\sim 10\text{ GeV}$  range, which are the focus of the discussion in this section. The CC and NC interactions, can be discriminated by the presence of a charged lepton in the final state. For example, for muon neutrinos:

$$\begin{aligned}\nu_\mu + A &\rightarrow \mu^- + X \quad (\text{CC}), \\ \nu_\mu + A &\rightarrow \nu_\mu + X \quad (\text{NC}).\end{aligned}\tag{2.31}$$

Here,  $A$  is the target nucleus and  $X$  represents the hadronic final state. This section will focus on CC interactions which are most relevant to the work presented in this thesis<sup>12</sup>.

The neutrino-nucleus interactions relevant to modern accelerator experiments can occur via a number of channels as the target nucleus is a composite object made of constituent nucleons and ultimately quarks. However, it is instructive to first discuss the simplest case of neutrino scattering off a point-like target particle. The tree-level Feynman diagram for the CC interaction  $\nu_\mu + d \rightarrow \mu^- + u$  is shown in Figure 2.4.



**Figure 2.4:** Feynman diagram for the Charged-Current neutrino-quark scattering process.

<sup>12</sup>In particular, emphasis is given to the neutrino interaction models used by the GENIE generator, discussed in Section 2.3.6

The rate at which an interaction can occur depends on the kinematics of the particles involved and the matrix element,  $\mathcal{M}$ , which encodes the underlying particle interactions. The matrix element for the diagram in Figure 2.4 is proportional to the inner product of the leptonic ( $\nu_\mu \rightarrow \mu^-$ ) and quark ( $d \rightarrow u$ ) charged currents, defined as [29]:

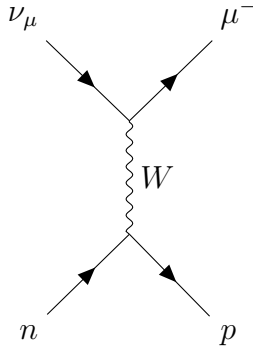
$$\mathcal{M} \propto j_{CC}^l \cdot j_{CC}^q \quad (2.32)$$

$$\begin{aligned} (j_{CC}^l)^\mu &= \bar{u}(p_l) [\gamma^\mu (1 - \gamma^5)] u(p_\nu), \quad (\text{Leptonic current}) \\ (j_{CC}^q)^\mu &= \bar{u}(p_u) [\gamma^\mu (1 - \gamma^5)] u(p_d). \quad (\text{Quark current}) \end{aligned} \quad (2.33)$$

Here,  $u$  is a standard Dirac spinor, and  $p_l$ ,  $p_\nu$ ,  $p_u$  and  $p_d$  are the four-momenta of the charged lepton, neutrino, up-quark and down-quark respectively. The form of Equation 2.33 is ultimately derived from the Charged-Current of Equation 2.27. The cross-section depends on  $\bar{\sum}_i \sum_f |\mathcal{M}|^2$  where  $\bar{\sum}_i$  indicates the average over the spins of the initial particles and  $\sum_f$  is the sum over the final particle spins [29].

### 2.3.3 Quasi-Elastic neutrino-nucleon scattering

For neutrino scattering off point-like particles (such as the single-quark case above), the matrix element can be derived analytically. For composite objects (such as nucleons), a more empirical approach is employed. At neutrino energies between  $\sim 100$  MeV and  $\sim 1$  GeV the dominant CC neutrino-nucleon interaction is Quasi-Elastic (QE) scattering in which the neutrino interacts elastically with the nucleon as a whole but converts it from neutron to a proton. The CC-QE process  $\nu_\mu + n \rightarrow \mu^- + p$  is shown in Figure 2.5.



**Figure 2.5:** Feynman diagram for the Charged-Current Quasi-Elastic neutrino scattering process.

In this case, the quark current in Equation 2.32 must be replaced by an analogous hadronic current,  $j_{CC}^h (n \rightarrow p)$  which reflects the fact the target (neutron) is not point-like. This is achieved by assuming the most general possible form of the hadronic current, given by Equation 2.34 [32], and any free parameters are ultimately constrained by experiment.

$$(j_{CC}^h)^\mu = \bar{u}(p_p) [V^\mu - A^\mu] u(p_n) \quad (2.34)$$

The components of the current involving  $V^\mu$  and  $A^\mu$  are constructed to scale as a vector and an axial-vector respectively from the most general possible combinations of gamma matrices and the four-momenta of the neutron and proton ( $p_n$  and  $p_p$ )<sup>13</sup> [32]. Each term is multiplied by a scalar function  $F(q^2)$  of the four-momentum transfer  $q = p_p - p_n$ , known as a form-factor [32]. For example, a term in  $V^\mu$  may have the form  $F(q^2)\gamma^\mu$ . In the special case of  $F(q^2) = 1$ , this term simplifies to the vector component of the point-like (quark) current. For other values, the form-factor encapsulates the finite structure of the nucleon [29].

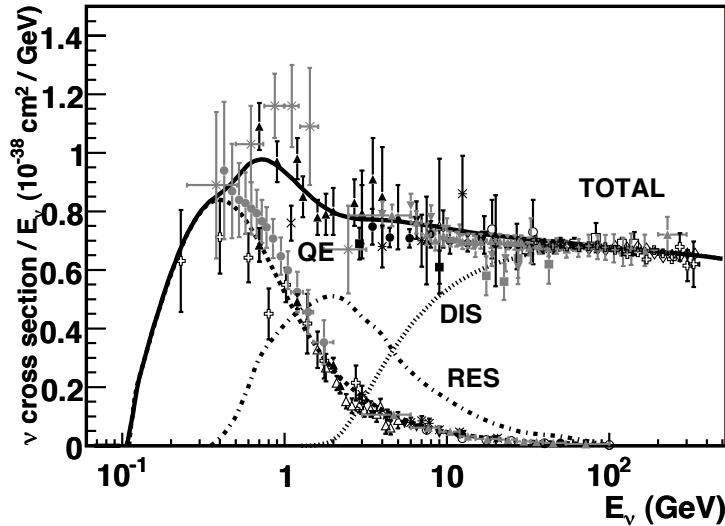
Both  $V^\mu$  and  $A^\mu$  have three possible linearly independent terms with associated form-factors,  $F_i^j(q^2)$  where  $i = V, A$ ,  $j = 1, 2, 3$  [34]. The cross-section for this CC neutrino-nucleon process, written in terms of these most general form-factors, is known as the Llewellyn-Smith model [35].  $F_V^1$  and  $F_V^2$  are known as the Dirac and Pauli form-factors respectively and relate to the distribution of electric charge and the magnetic moment of the nucleon [32]. These factors can be related to analogous quantities in electromagnetic electron-nucleon scattering, and so are constrained by electron scattering experiments.  $F_V^3$  and  $F_A^2$  are known as the scalar and tensor form-factors respectively and are assumed to be zero within the Llewellyn-Smith model [35]. Finally,  $F_A^1$  and  $F_A^3$  are known as the axial and pseudo-scalar form-factors and are related to each other such that only  $F_A^1$  need be determined by neutrino scattering experiments [34]. This factor is commonly parameterised as a dipole  $F_A^1(q^2) \propto (1 - q^2/M_A^2)^{-2}$ , where the parameter  $M_A$  is known as the axial-mass [34].

<sup>13</sup>The following bilinear covariants are the possible combinations of Dirac spinors and gamma matrices of the form,  $\bar{u}\Gamma u$ , that transform as a tensor:  $\Gamma = 1$  (scalar),  $\gamma^5$  (pseudo-scalar),  $\gamma^\mu$  (vector),  $\gamma^\mu\gamma^5$  (axial-vector) and  $\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu]$  (second-rank tensor) [29]. The most general linearly independent terms in  $V^\mu$  have the form  $\gamma^\mu$  (Dirac),  $\sigma^{\mu\nu}q_\nu$  (Pauli) and  $q^\mu$  (scalar), where  $q = p_p - p_n$ . The terms in  $A^\mu$  have the form  $\gamma^\mu\gamma^5$  (axial),  $\sigma^{\mu\nu}q_\nu\gamma^5$  (tensor), and  $q^\mu\gamma^5$  (pseudo-scalar) [32].

### 2.3.4 Inelastic neutrino-nucleon scattering

One of the main aims of early neutrino scattering experiments, which typically used deuterium ( $^2\text{H}$ ) as the target medium, was to measure the axial mass and to test the assumptions made by the QE model described in Section 2.3.3 [34]. This treatment was successful but can not be applied in two important limits: at higher neutrino energies (where inelastic scattering processes occur), and for larger nuclei (where nuclear interactions must be included). The latter nuclear effects are described in Section 2.3.5, and the inelastic processes are described here.

For neutrino energies above  $\sim 300$  MeV, the neutrino is sufficiently energetic to excite the struck nucleon and produce a short-lived baryon resonance. The decay of this resonance produces hadronic final-state particles which are most often a nucleon and a single pion, although multi-pion and strange<sup>14</sup> final states are also possible [34]. For the Charged-Current channel, this process is known as CC-Resonant (Res). At yet higher energies above  $\sim 2$  GeV, the neutrino begins to resolve the quark-structure of the nucleon. This process is known as Deep Inelastic Scattering (DIS) whereby the neutrino scatters off a quark in the nucleon resulting in a hadronic shower [29]. The rate of each process depends on the neutrino energy, as is shown in Figure 2.6 for the CC channel.



**Figure 2.6:** The Charged-Current muon neutrino-nucleus cross-section per nucleon scaled by neutrino energy. This figure shows world data (as of 2013) and predictions of the Quasi-Elastic, Resonant and Deep Inelastic Scattering components provided by the NUANCE generator. Taken from [34]

<sup>14</sup>Strange final states are those involving a strange quark.



### Resonant single-pion production

The neutrino-nucleon CC-Res processes that result in a nucleon and single pion hadronic final state can be written generally as:

$$\begin{aligned}\nu_\mu + N &\rightarrow \mu^- + R \\ R &\rightarrow N' + \pi\end{aligned}\tag{2.35}$$

Here,  $N$  is the target nucleon,  $R$  is a baryon resonance,  $N'$  is the final-state nucleon and  $\pi$  is a charged or neutral pion. The conservation of electric charge restricts the possible channels to [34]:

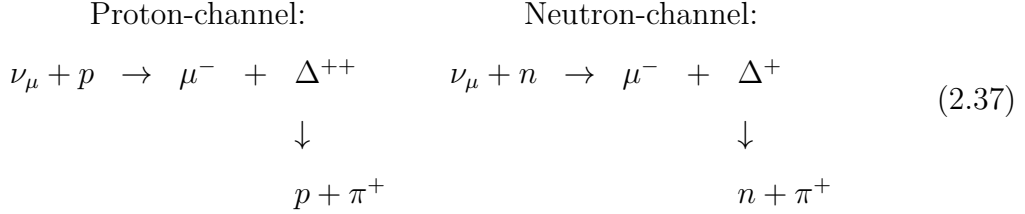
$$\begin{aligned}\nu_\mu + p &\rightarrow \mu^- + p + \pi^+ \quad (Q = 2), \\ \nu_\mu + n &\rightarrow \mu^- + n + \pi^+ \quad (Q = 1), \\ \nu_\mu + n &\rightarrow \mu^- + p + \pi^0 \quad (Q = 1).\end{aligned}\tag{2.36}$$

Here,  $Q$  is the charge of the resonance. Each final state can be reached by numerous resonances and the four with the lowest mass are shown in Table 2.2 [36] [33]. The resonances are differentiated by their mass, isospin,  $I$ , total angular momentum,  $J$ , and parity  $P$ . For baryons consisting of only up and down quarks, those with  $I = \frac{1}{2}$  are known as nucleon resonances and have two possible isospin projections with quark content:  $N^0 = udd$ ,  $N^+ = uud$ , whereas those with  $I = \frac{3}{2}$  are known as delta-resonances have four projections:  $\Delta^- = ddd$ ,  $\Delta^0 = udd$ ,  $\Delta^+ = uud$ ,  $\Delta^{++} = uuu$ . Here the superscript indicates the electric charge,  $Q$ .

Resonance Symbol	$I(J^P)$	$\Gamma$ / MeV	BR( $N' + \pi$ )
$P_{33}(1232)$	$\frac{3}{2}(\frac{3}{2}^+)$	114-120	99.4%
$P_{11}(1440)$	$\frac{1}{2}(\frac{1}{2}^+)$	250-450	55-75%
$D_{13}(1520)$	$\frac{1}{2}(\frac{3}{2}^-)$	100-120	55-65%
$S_{11}(1535)$	$\frac{1}{2}(\frac{1}{2}^-)$	125-175	32-52%

**Table 2.2:** The lowest mass baryon resonances considered by the Rein-Sehgal model [36], with modern values are from PDG [33]. The baryon naming structure is:  $L_{2I} \ 2J(M)$ , where  $L = 0, 1, 2, \dots = S, P, D, \dots$  is the orbital angular momentum,  $I$  is the isospin,  $J$  is the total angular momentum and  $M$  is the mass measured in MeV  $c^{-2}$ . The isospin, total angular momentum and parity,  $P$  is also given explicitly. The full Breit-Wigner decay width,  $\Gamma$ , is supplied and the branching ratio (BR) shown is for decays to a nucleon + single pion final state (as illustrated in Equation 2.35).

The lowest mass  $\Delta$ -resonance  $P_{33}(1232)$  is dominant at energies of  $\sim 1$  GeV and decays primarily to the nucleon + single pion final state. Consequently, a positively charged pion is most commonly produced via the CC-Res interactions shown in Equation 2.37.



The Rein-Sehgal model [36] is commonly used to describe the cross-section for the Res single-pion processes and considers essentially all known resonances with masses less than  $2 \text{ GeV } c^{-2}$ . The total amplitude for a given final state (listed in Equation 2.36) can have contributions from each of the resonances considered, depending on their isospin. For example, only the  $I = \frac{3}{2}$  resonances have a projection with  $Q = 2$  and so can contribute to the  $p + \pi^+$  hadronic final state. Conversely, the  $Q = 1$  final states ( $n + \pi^+$  and  $p + \pi^0$ ) have contributions from both  $I = \frac{1}{2}$  and  $I = \frac{3}{2}$  resonances. The cross-section ultimately depends on the square of the total amplitude with the appropriate sum over the spins of the particles involved. In total the Rein-Sehgal model considers 18 resonances and includes a treatment of their interferences [36]. However, the implementation of this model in the GENIE generator, used in this thesis, does not include interference terms and instead the cross-section is computed as the incoherent sum of the contributions from each resonance [37].

Within the Rein-Sehgal model, the amplitude for a given resonance,  $R$ , in a definite state of charge, isospin and helicity is written in terms of two factors:

$$f(N \rightarrow R) \cdot \eta(R \rightarrow N' + \pi). \tag{2.38}$$

Here,  $f(N \rightarrow R)$  is the amplitude for the production of the resonance and  $\eta(R \rightarrow N' + \pi)$  is the decay amplitude. The production amplitudes are calculated using the Feynman-Kislinger-Ravndal (FKR) model [38] in which baryons are considered as a system of three quarks which are coupled as relativistic harmonic oscillators. The amplitudes  $f(N \rightarrow R)$  depend on a vector and an axial-vector form-factor which are taken to have a dipole form with corresponding parameters,  $M_V^{Res}$  and  $M_A^{Res}$  respectively [36]. These parameters are found experimentally.

The decay amplitude  $\eta(R \rightarrow N' + \pi)$  includes a Breit-Wigner factor such that the contribution of a single resonance to the cross-section has the form [36]:

$$\sigma_R(W) \sim \frac{\Gamma}{(W - M)^2 + \Gamma^2/4} \quad (2.39)$$

Here,  $W$  is the hadronic invariant mass,  $M$  is the mass of the resonance, and  $\Gamma = 1/\tau$  is the decay width where  $\tau$  is the mean lifetime of the resonance.  $\eta(R \rightarrow N' + \pi)$  also includes a factor to account for the branching-ratio for the  $N' + \pi$  final state, as given in Table 2.2. In addition, the Rein-Sehgal model includes an ad-hoc non-resonant background contribution for each final state, which is modelled like a  $P_{11}$  (nucleon type) resonance but with the Breit-Wigner factor replaced by an adjustable constant [36].

In the Rein-Sehgal model, the charged-lepton mass is assumed to be negligible. The work of Kuzmin, Lyubushkin and Naumov [39] extends this model to include lepton mass effects and the Berger-Sehgal model [40], used in this thesis, further includes the pion-pole contribution to the hadronic axial vector current. The predictions of these models are consistent with independent calculations performed by Graczyk and Sobczyk [41, 42].

## Deep Inelastic Scattering

At the highest neutrino energies, neutrino-nucleon interactions are dominated by DIS whereby the neutrino can resolve the quark-structure of the nucleon. The CC-DIS process can be written generally as:

$$\nu_\mu + N \rightarrow \mu^- + X \quad (2.40)$$

Here, the hadronic final state  $X$  can consist of multiple particles resulting from the break-up of the nucleon. The underlying first-generation quark-level processes that contribute to DIS are: <sup>15</sup>

$$\begin{aligned} \nu_\mu + d &\rightarrow \mu^- + u, \\ \nu_\mu + \bar{u} &\rightarrow \mu^- + \bar{d}. \end{aligned} \quad (2.41)$$

Within the quark-parton model [29], the nucleon is considered as a state of valence quarks ( $uud$  or  $udd$  for  $p$  and  $n$  respectively) and gluons,  $g$ , which are not probed by weak interactions but result in a sea of quarks through pair production  $g \rightarrow q\bar{q}$ . The

---

<sup>15</sup>Heavier sea quarks can also contribute to the DIS process, but are not included in this discussion.

$\nu_\mu + d$  process has contributions from valence and sea  $d$ -quarks, however the  $\nu_\mu + \bar{u}$  process has only sea  $\bar{u}$ -quark contributions.

The distributions of the four-momenta of the quarks are specified via Parton Distribution Functions (PDFs),  $q_N(x)$ , which relates to the probability density of a quark,  $q$ , in a nucleon  $N$  having momentum  $p_q = xp_N$ , where  $p_N$  is the nucleon four-momentum in the appropriate frame<sup>16</sup>. The Bjorken- $x$  is a Lorentz-invariant quantity given by [29]:

$$x \equiv \frac{Q^2}{2p_N \cdot q}. \quad (2.42)$$

Here,  $q = p_\nu - p_l$  is the four-momentum transfer and  $Q^2 = -q^2$ . The PDFs must be found experimentally. The neutrino-quark elastic scattering cross-section can be calculated using the matrix element given in Equation 2.32. The differential cross-section for CC-DIS neutrino-proton scattering is naively the sum of the cross-sections for each neutrino-quark (and neutrino-antiquark) processes, weighted by their respective PDFs. At high energies, for which the particle masses can be neglected, this is given by:

$$\frac{d^2\sigma^{\nu p}}{dx dy} = \frac{G_F^2 s}{\pi} \left[ x d_p(x) - x(1-y)^2 \bar{u}_p(x) \right]. \quad (2.43)$$

Here,  $G_F$  is the Fermi coupling constant,  $s = (p_\nu + p_N)^2$  is the centre-of-mass energy, and  $y$  is a Lorentz-invariant quantity known as the inelasticity:

$$y \equiv 1 - \frac{p_N \cdot p_l}{p_N \cdot p_\nu}. \quad (2.44)$$

The analogous cross-section for neutrino-neutron scattering is related by isospin symmetry:  $d_n(x) = u_p(x)$ ,  $\bar{u}_n(x) = \bar{d}_p(x)$ . Alternatively, following similar arguments to Section 2.3.3, the cross-section can be derived from the most general possible hadronic current, giving (in the high-energy limit) [29]:

$$\frac{d^2\sigma^{\nu p}}{dx dy} = \frac{G_F^2 s}{\pi} \left[ xy^2 F_1^{\nu p}(x, Q^2) + (1-y) F_2^{\nu p}(x, Q^2) + xy \left(1 - \frac{y}{2}\right) F_3^{\nu p}(x, Q^2) \right] \quad (2.45)$$

Here,  $F_i^{\nu p}(x, Q^2)$ , are known as structure functions which must be determined experimentally. Equation 2.43 can be restored from Equation 2.45 by identifying  $F_2^{\nu p} = 2xF_1^{\nu p}$  and  $F_1^{\nu p}$  and  $F_3^{\nu p}$  as linear combinations of  $d_p(x)$  and  $\bar{u}_p(x)$ . In this case the structure functions are only dependent on  $x$ . In practice, the dependence on  $Q^2$  is non-zero and

<sup>16</sup>In the quark-parton model the momentum is evaluated in the *infinite momentum frame* where  $p_N = (E_N, \underline{p}_N)$ ,  $\underline{p}_N = (0, 0, E_N)$  and  $E_N$  is the nucleon energy.

arises due to strong interactions within the nucleus. In this thesis, the Bodek-Yang inelastic scattering model [43] is used which is applicable over wide range of  $Q^2$ .

### 2.3.5 Neutrino-nucleus interactions

The models described in Section 2.3.3 and Section 2.3.4 are applicable to neutrino interactions with a single nucleon. However, for experiments that use larger nuclear targets, the dynamics of the nucleons within the nucleus are important. At low momentum transfer ( $Q^2 \leq 0.1 \text{ GeV}^2$ ), the CC-Coherent (Coh) pion production process can occur:

$$\nu_\mu + A \rightarrow \mu^- + A + \pi^+. \quad (2.46)$$

In this case, the neutrino interacts with the entire nucleus,  $A$ , and transfers negligible energy, leaving it in its ground state. However, for many accelerator experiments, this process is rare as the neutrinos involved commonly transfer sufficient energy to resolve individual nucleons (so the QE, Res, and DIS neutrino-nucleon processes are more prevalent) <sup>17</sup>. Consequently, the CC-Coh single pion production channel is not described in detail. Neutrino-nucleus interactions can often be described as the convolution of the following components:

Nuclear model  $\otimes$  Primary interaction  $\otimes$  Hadronisation  $\otimes$  Hadron transport.

- The **nuclear model** describes the initial-state dynamics of the nucleons within the nucleus and any nucleon-nucleon interactions. As a result of these interactions, discussed in Section 2.3.5.2, the neutrino may interact with correlated groups of nucleons known as a multi-nucleon interaction.
- The **primary interaction model** describes the initial neutrino-nucleon (or multi-nucleon) interaction, e.g. QE, Res, DIS.
- If a DIS interaction occurs, the **hadronisation model** describes the formation of the resulting hadronic shower from the break-up of the nucleon <sup>18</sup>.

<sup>17</sup>GENIE predicts that the CC-Coh process contributes only  $\sim 1\%$  of the charged current single pion final states at MicroBooNE.

<sup>18</sup>The GENIE AGKY hadronisation model [44] is used in this thesis.

- The **hadron transport model** describes the secondary interactions of the hadronic final-state particles produced in the primary interaction as they traverse the nucleus known as Final State Interactions (FSI).

### 2.3.5.1 Fermi-motion and Pauli-blocking

A number of the nuclear effects that modify the aforementioned neutrino-nucleon interactions can be explained by a Relativistic Fermi-Gas (RFG) nuclear model [32]. In this model, the nucleons are assumed to feel a mean-field potential due to the average of their interactions with other nucleons, commonly taken as a finite square well. The allowed energies are quantised to energy levels within the well and nucleons are subject to the Pauli exclusion principle. Consequently, each energy level can be occupied by at most two identical particles with opposing spins, and protons and neutrons are considered as independent systems. When the nucleus is in its ground state, all energy levels up to the Fermi-energy,  $E_F$ , are occupied and all levels above this threshold are unoccupied [32].

As a result of the exclusion principle, the nucleons have non-zero momentum known as Fermi-motion, described by a spectral function,  $S(p)$ , that expresses the probability density that a nucleon has four-momentum,  $p$ . In the ground state, the nucleon momenta are bounded by the Fermi-momentum,  $|\underline{p}| < |\underline{p}_F|$ , which relates to the Fermi-energy. Additionally, the momentum transferred,  $q$ , to the nucleon must be sufficient for it to transition into an unoccupied energy level,  $|\underline{p} + \underline{q}| > |\underline{p}_F|$ . This effect is known as Pauli-blocking and leads to a suppression of the cross-section at low momentum transfer [32].

The simplest method by which a neutrino-nucleus cross-section can be evaluated is via the Impulse Approximation (IA) [32]. In this case, the neutrino is assumed to interact with a single nucleon as if it were free and particles produced in the neutrino-nucleon interaction evolve independently of the recoiling nucleus. The neutrino-nucleus cross-section is then the integral of the neutrino-nucleon cross-section over the kinematically-allowed target nucleon four-momentum weighted by the spectral function and the number of nucleons in the nucleus.

In practice, the IA is often insufficient to describe neutrino scattering data on large nuclei due to the presence of nucleon-nucleon interactions and FSI. Nevertheless, the Fermi-gas model is commonly used as starting-point upon which to build a more nuanced nuclear model. In this thesis, a Local Fermi-Gas (LFG) [44] model is used whereby mean-field felt by a nucleon is dependent on the nuclear density,  $\rho(r)$ , as a function of

the radial position,  $r$ , of the nucleon. Consequently, the nucleon momentum distribution is additionally dependent on  $r$ . For the Res and DIS channels, the primary interaction is considered to be between the neutrino and a single nucleon where an LFG model is used to define the initial nucleon kinematics. In addition, the LFG model defines the binding energy of the nucleon which suppresses the free nucleon cross-section at low momentum transfer where the struck nucleon is insufficiently energetic to escape the nucleus [44]. The propagation of the hadronic final state particles through the nucleus is described in Section 2.3.5.3.

### 2.3.5.2 Nucleon-nucleon interactions

Modelling of interactions between nucleons is required to describe modern neutrino scattering data [45]. These nucleon-nucleon correlations can effectively change the strengths of the electroweak couplings from their free nucleon values [46]. Nucleon interactions can also yield additional multi-nucleon neutrino interaction processes [47]. A neutrino interaction involving a single nucleon that transitions to an energy level above the Fermi-energy (leaving behind an unoccupied level) is known as a one-particle one-hole (1p-1h) excitation. In cases where momentum is transferred to multiple correlated nucleons, the result is a two-particle two-hole (2p-2h) or higher multiplicity excitation which can result in the ejection of multiple nucleons from the nucleus [48]. Mutli-nucleon interactions primarily arise due to Meson Exchange Currents (MEC) whereby a pair of nucleons interact via the exchange a virtual meson [32]. In this thesis, the Valencia model [46] [47] is used to describe the 1p-1h (QE) and 2p-2h processes which result in a hadronic final state containing only the ejected nucleons and the recoiling nucleus. This model includes nucleon-nucleon correlation effects.

### 2.3.5.3 Final State Interactions

The hadronic final-state particles of the primary neutrino interaction can undergo secondary interactions as they propagate through the nucleus. In particular, the particles can scatter elastically, undergo charge exchange (e.g.  $\pi^- + p \leftrightarrow \pi^0 + n$ ), pions can be absorbed by the nucleus, or indeed additional pions can be produced <sup>19</sup>. Consequently, the particles that escape the nucleus (and so may be detected experimentally) are not necessarily the same as those produced in the primary neutrino interaction. Therefore, it

<sup>19</sup>This thesis uses the GENIE hA (2018) FSI simulation [44].

is important to make the distinction between a neutrino interaction mode (e.g. CC-QE) and a final-state topology (e.g.  $1\mu + 1p$ ) which defines the multiplicity of each particle species that escape the nucleus after FSI.

### 2.3.6 The role of generators

As illustrated by this section, neutrino-nucleus scattering is a complex phenomenon which, at present, can not be fully described by a single theoretical model. Instead, each model describes a specific process (or set of processes) and may have free parameters that must be estimated experimentally. A generator is a software package that combines the predictions of multiple theoretical models to simulate neutrino interactions in a given detector medium. The role of a generator is to bridge the gap between the theoretical and experimental domains.

The topologies and kinematics of the simulated neutrino interactions are governed by the theoretical models employed by the generator. In general, a given final-state topology will have contributions from multiple primary neutrino interaction modes. The result can be compared to experimental data and used to determine the validity of the models used and tune any free parameters to better represent reality. In this thesis, the GENIE event generator [44] is used as this is the generator adopted by MicroBooNE. In particular, a custom tune of the GENIE v3.0.6 configuration G18\_10a\_02\_11a is used which was produced by the MicroBooNE experiment [44, 49], and the models employed have been listed throughout this section.

## 2.4 The $\text{CC}1\pi^\pm$ topology

### 2.4.1 Definition

The muon-neutrino charged-current single charged pion ( $\text{CC}1\pi^\pm$ ) topology is the focus of the primary analysis presented in Part III of this thesis, and is defined as:

$$\text{CC}1\pi^\pm = 1\mu + 1\pi^\pm + N_p p + N_n n. \quad (2.47)$$

Here, the nucleon multiplicities ( $N_p$  and  $N_n$ ) can take any value, including zero. An example of a  $\text{CC}1\pi^\pm$  candidate with  $N_p = 2$  (in MicroBooNE data) was shown in



Figure 1.1. The  $\text{CC}1\pi^\pm$  topology is commonly the result of the CC-Res interactions shown in Equation 2.37. However, other interaction modes such as CC-DIS and CC-QE contribute. In the CC-QE case, the charged pion is produced during FSI.

The relative contributions from each interaction mode depend on the neutrino energies that are measured (governed by the flux of the beam used). As the neutrino energy is not known on an event-by-event basis and can only be estimated in a model-dependent way, it is common for cross-section measurements to be *flux-integrated* (defined formally in Chapter 7), whereby the result is dependent on the energy distribution of the neutrino beam used. Consequently, measurements from different experiments can not (in general) be directly compared. However, a generator can be used to predict the expected flux-integrated cross-section for a given beam and set of interaction models which can be compared to the data.

At the neutrino energies relevant to MicroBooNE ( $\langle E \rangle \approx 800 \text{ MeV}$ <sup>20</sup>), GENIE predicts that the primary contributions to the  $\text{CC}1\pi^\pm$  topology are from CC-Res interactions ( $\sim 81\%$ ), with subdominant contributions from CC-DIS ( $\sim 13\%$ ) and CC-QE ( $\sim 4\%$ ). The contribution of the CC-Coh mode is predicted to be at the  $\sim 1\%$  level.

## 2.4.2 Kinematic parameters

Modern measurements of the  $\text{CC}1\pi^\pm$  topology are commonly made as a function of the lab-frame three-momenta,  $\underline{p}_P$ , of the final-state particles,  $P = \mu, \pi$ . The momenta are parametrised using a spherical polar coordinate system,  $\underline{p}_P = (p_P, \theta_P, \phi_P)$ , shown in Figure 2.7. The zenith angle,  $\theta_P$ , is defined with respect to the momentum of the neutrino,  $\underline{p}_\nu$ :

$$\cos(\theta_P) = \left( \frac{\underline{p}_P}{p_P} \right) \cdot \left( \frac{\underline{p}_\nu}{E} \right). \quad (2.48)$$

Here,  $E \equiv |\underline{p}_\nu|$ <sup>21</sup> is the neutrino energy. As  $\underline{p}_\nu$  is not known a priori, the unit vector  $(\underline{p}_\nu/E)$  is usually taken as the neutrino beam axis. The interactions are independent of the azimuthal angle,  $\phi_P$ , and so this parameter is not commonly presented. Additionally,

<sup>20</sup>The flux of the Booster Neutrino Beam used by MicroBooNE is shown in Figure 3.3.

<sup>21</sup>In this case the neutrino masses are negligible.

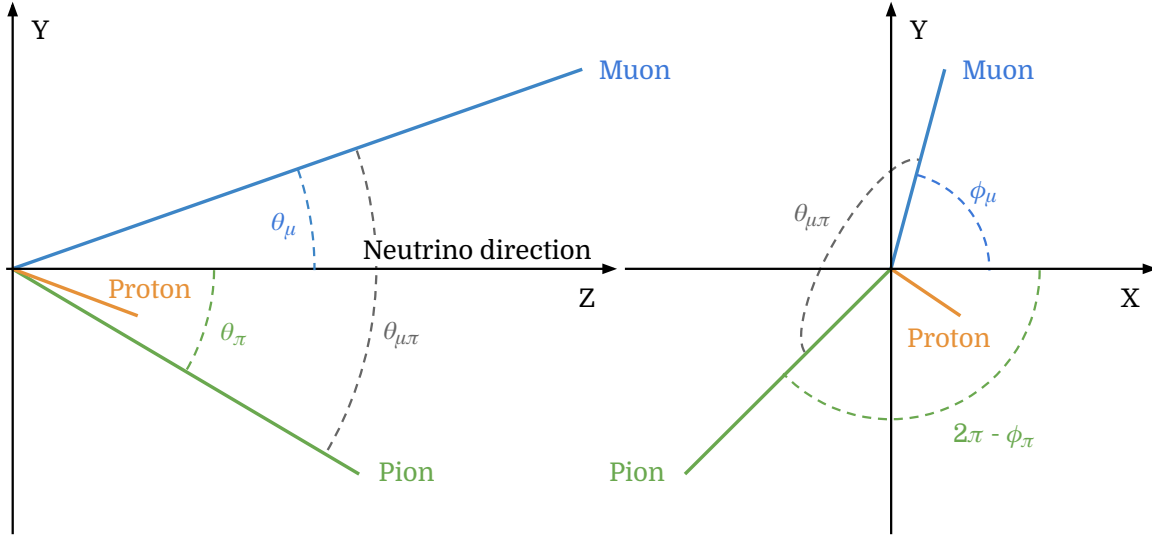
measurements of  $\text{CC}1\pi^\pm$  are often made with respect to the muon-pion opening angle:

$$\cos(\theta_{\mu\pi}) = \left( \frac{\underline{p}_\mu}{p_\mu} \right) \cdot \left( \frac{\underline{p}_\pi}{p_\pi} \right). \quad (2.49)$$

Instead of momentum, some experiments use kinetic energy,  $T_P$ , which is given by the relativistic energy-momentum relation for a particle with mass,  $m_P$ :

$$(T_P + m_P)^2 = m_P^2 + p_P^2. \quad (2.50)$$

Historically, no attempts have been made to measure dependence of the  $\text{CC}1\pi^\pm$  cross-section on the nucleon multiplicities or their kinematics. In this thesis a measurement of the proton multiplicity,  $N_p$ , is presented for the first time.



**Figure 2.7:** The spherical coordinate system commonly used for the  $\text{CC}1\pi^\pm$  topology. The  $Z$  axis is aligned with the axis of the neutrino. The  $X$  and  $Y$  directions are orthogonal to  $Z$  and each other but are otherwise arbitrary. In this thesis,  $Y$ , is chosen to align with the upwards vertical direction and a right-handed coordinate system is used. Here a single proton is shown however any number of nucleons are permissible.

In addition to flux-integrated cross-sections, some experiments quote results with respect to quantities that are not directly measurable, for example the neutrino energy,  $E$ , or four-momentum transfer,  $Q^2 \equiv -(p_\mu - p_\nu)^2$  (where in this case  $p_\mu$  and  $p_\nu$  are the *four-momenta* of the muon and neutrino respectively). These parameters can only be estimated from the visible final states and are not presented in this thesis as they are necessarily model dependent.

### 2.4.3 Existing measurements

Early data of the  $\text{CC}1\pi^\pm$  topology is available from bubble chamber experiments based at the Argonne [50] and Brookhaven [51] national laboratories in the USA. These experiments used hydrogen ( $^1\text{H}$ ) and deuterium ( $^2\text{H}$ ) as a target medium. However, this section will focus on more recent measurements on larger nuclear targets but those focussed on probing the CC-Coh interaction mode are not discussed [52–57]. A summary of the existing measurements of the  $\text{CC}1\pi^\pm$  topology is shown in Table 2.3.

The measurements from ArgoNeuT [58] and MINER $\nu$ A [59] were made using the NuMI beam [60] (on-axis) at Fermilab, USA with neutrino energies in the few-GeV region. ArgoNeuT is a small-scale LArTPC detector and published results [61] with  $\mathcal{O}(100)$   $\text{CC}1\pi^\pm$  neutrino and antineutrino interactions on argon. MINER $\nu$ A is a larger detector situated just upstream of ArgoNeuT and has measured  $\text{CC}1\pi^\pm$  neutrino interactions on plastic scintillator (CH) <sup>22</sup>. The published results [62–64] from this experiment include  $\mathcal{O}(1000)$   $\text{CC}1\pi^\pm$  interactions. The MiniBooNE experiment is also situated at Fermilab but measures neutrinos from the Booster Neutrino Beam with energies of  $\sim 1$  GeV. MiniBooNE measures neutrino interactions on mineral oil ( $\text{CH}_2$ ) and the published results [65] include  $\mathcal{O}(10,000)$   $\text{CC}1\pi^\pm$  interactions, allowing for double-differential cross-sections to be measured.

The long-baseline oscillation experiment K2K used the near-detector, SciBar, to measure  $\text{CC}1\pi^\pm$  interactions on polystyrene ( $\text{C}_8\text{H}_8$ ). SciBar was based at the KEK facility in Japan and measured beam neutrinos with energy  $\sim 1.3$  GeV. The published results [66] include  $\mathcal{O}(1000)$   $\text{CC}1\pi^\pm$  interactions. Finally, the successor T2K experiment has measured  $\text{CC}1\pi^\pm$  interactions in the ND280 near-detector (based at J-PARC, Japan) at neutrino energies of  $\sim 0.8$  GeV. The FGD1 and FGD2 sub-detectors are used to measure neutrino interactions on polystyrene and water ( $\text{H}_2\text{O}$ ) respectively <sup>23</sup>. The published results [67, 68] include  $\mathcal{O}(1000)$   $\text{CC}1\pi^\pm$  interactions.

<sup>22</sup>MINER $\nu$ A is also capable of measuring neutrino interactions on other target material, however the results for the  $\text{CC}1\pi^\pm$  topology use CH.

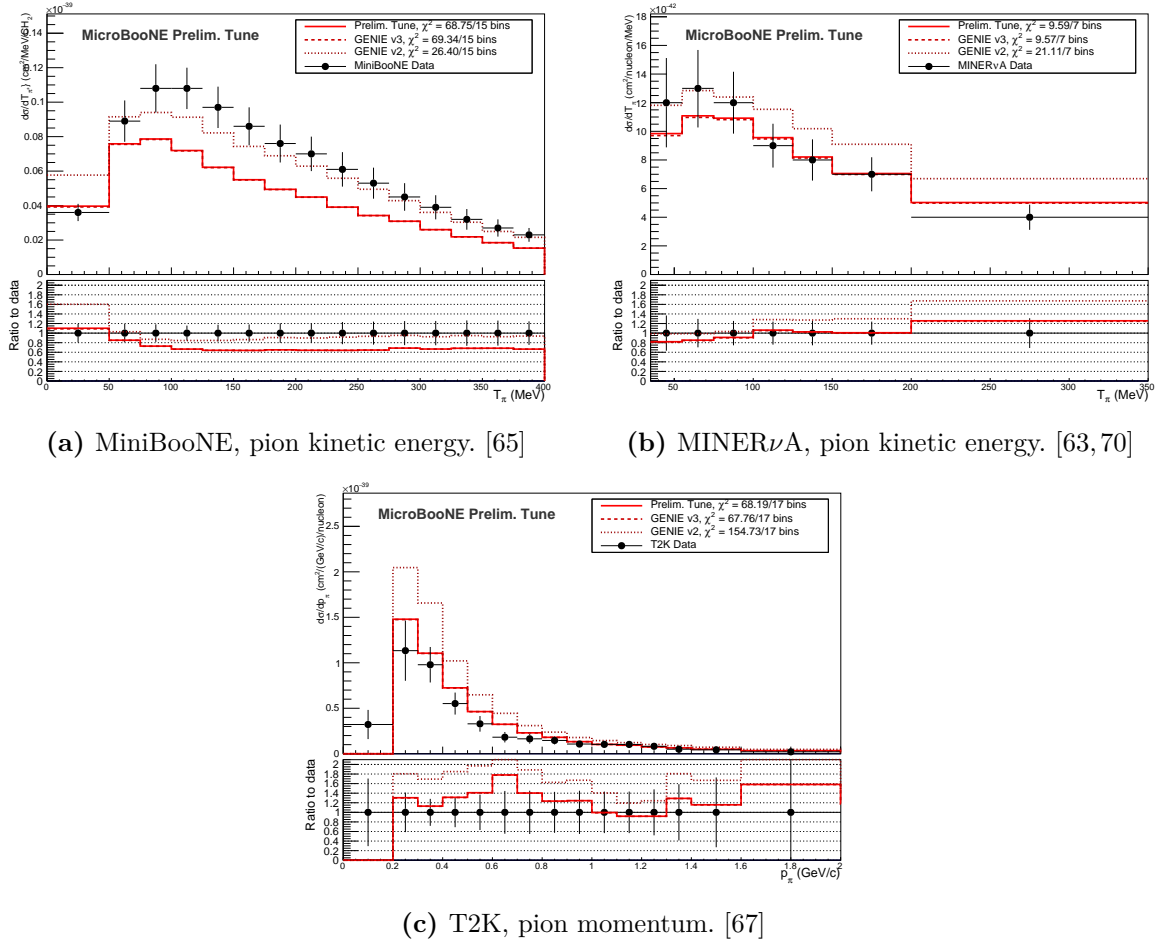
<sup>23</sup>Both FGD1 and FGD2 are composed of scintillator (polystyrene) bars organised in layers. FGD1 is comprised of active scintillator layers only whereas FGD2 incorporates active scintillator and inactive water layers [67]

Experiment	Target	Beam	Measurements made
ArgoNeuT	Ar	$\nu_\mu, \bar{\nu}_\mu$	$\sigma^{CC1\pi^\pm}$ [61] $\frac{d\sigma^{CC1\pi^\pm}}{dX} : X = p_\mu, \theta_\mu, \theta_\pi, \theta_{\mu\pi}$
MINER $\nu$ A	CH	$\nu_\mu$	$\frac{d\sigma^{CC1\pi^\pm}}{dX} : X = \theta_\pi, T_\pi$ [62]
		$\nu_\mu, \bar{\nu}_\mu$	$\sigma^{CC1\pi^\pm}(E_\nu)$ [63] $\frac{d\sigma^{CC1\pi^\pm}}{dX} : X = p_\mu, \theta_\mu, Q^2$
		$\bar{\nu}_\mu$	$\sigma^{CC1\pi^-}(E_\nu)$ [64] $\frac{d\sigma^{CC1\pi^-}}{dX} : X = p_\mu, \theta_\mu, T_\pi, \theta_\pi, Q^2$
MiniBooNE	CH <sub>2</sub>	$\nu_\mu$	$\sigma^{CC1\pi^+}(E_\nu)$ [65] $\frac{d\sigma^{CC1\pi^+}}{dX} : X = T_\mu, T_\pi, Q^2$ $\frac{d\sigma^{CC1\pi^+}}{dX}(E_\nu) : X = T_\mu, T_\pi, Q^2$ $\frac{d^2\sigma^{CC1\pi^+}}{dXdY} : \{X, Y\} = \{T_\mu, \theta_\mu\}, \{T_\pi, \theta_\pi\}$ $\frac{\sigma^{CC1\pi^+}}{\sigma^{CCQE}}(E_\nu)$ [69]
K2K	C <sub>8</sub> H <sub>8</sub>	$\nu_\mu$	$\frac{\sigma^{CC1\pi^+}}{\sigma^{CCQE}}(E_\nu)$ [66]
T2K	H <sub>2</sub> O	$\nu_\mu$	$\frac{d\sigma^{CC1\pi^+}}{dX} : X = p_\mu, \theta_\mu, p_\pi, \theta_\pi, \theta_{\mu\pi}, E_\nu$ [68]
	C <sub>8</sub> H <sub>8</sub>		$\frac{d\sigma^{CC1\pi^+}}{dX} : X = p_\pi, \theta_\pi, \theta_{\mu\pi}, Q^2, \cos(\theta_{\text{Adler}}), \phi_{\text{Adler}}$ [67] $\frac{d^2\sigma^{CC1\pi^+}}{dp_\mu d\cos(\theta_\mu)}$

**Table 2.3:** Modern accelerator-based CC1 $\pi^\pm$  cross-section measurements (extended from [33]). The angles,  $\theta_{\text{Adler}}$  and  $\phi_{\text{Adler}}$  are defined in [67] and refer to the pion direction in the proton-pion rest frame (although T2K calculate these angles using the lepton-pion kinematics as the proton is not observed).

It is not trivial to compare these existing measurements of the  $\text{CC}1\pi^\pm$  topology as each are produced with a different neutrino energy spectrum and apply different conditions to define their signal. However, the MicroBooNE collaboration has compared [49] the results of MiniBooNE, MINER $\nu$ A and T2K experiments to the predictions of the GENIE generator, shown in Figure 2.8. In this figure, the differences in the predictions of MicroBooNE’s tune of GENIE (solid red line) and the default GENIE v3 configuration (dashed red line) are small. This is because the former is a tune of the latter in which only CC-QE and CC-MEC model parameters were modified. However, there are significant discrepancies between the predictions of GENIE v2 (dotted red line) and v3 (dashed red line). This is because the CC-Res and nuclear models were updated in v3. In particular, v2 uses the Rein-Sehgal Res model with an RFG nuclear model, whereas v3 uses the updated Berger-Sehgal Res model with an LFG nuclear model. Additionally, the models employed were tuned by the GENIE collaboration to the available data at the time of release. Consequently, the values of the model parameters shared between the two versions are not necessarily equal.

Figure 2.8 highlights the tensions between existing  $\text{CC}1\pi^\pm$  measurements. In particular, data from MiniBooNE favours GENIE v2 whereas MINER $\nu$ A and T2K prefer v3. Additional data is required to help understand the source of these discrepancies.



**Figure 2.8:** Existing differential cross-section measurements of the CC1 $\pi^\pm$  topology as a function of the pion kinetic energy or momentum, taken from [49]. The black points show the cross-section measured by each experiment and are compared to predictions made by the GENIE generator [44] shown as red lines. The solid red line is the tuned configuration of GENIE used in this thesis, whereas the other lines are the default predictions from GENIE versions three (dashed) and two (dotted).

#### 2.4.4 Motivation for $\text{CC1}\pi^\pm$ at MicroBooNE

One of the primary goals of the MicroBooNE experiment, introduced in Chapter 1, is to measure a suite of precision cross-sections over a diverse range of final-state topologies. As is discussed in Chapter 3, the LArTPC technology employed by MicroBooNE is particularly well suited to this task as final-state particle trajectories can be resolved with high spatial and calorimetric resolution. In particular, the proton-multiplicity measurement developed in this thesis will constitute a world's first.

The LArTPC technology is also crucial for the long-term goals of the neutrino oscillation field as it will be utilised by the future DUNE experiment. DUNE will be subject to a neutrino flux with energies primarily in the Res-DIS cross-over region  $\sim 1 - 5$  GeV [71], and so measurements of the  $\text{CC1}\pi^\pm$  topology on argon, such as those presented in this thesis, are particularly relevant. At present, only the results from ArgoNeuT are available. However, it is important to note MicroBooNE can only probe  $\text{CC1}\pi^\pm$  interactions at energies in the  $\sim 0.5 - 2$  GeV region using neutrinos from the Booster Neutrino Beam.

The results presented in this thesis include  $\mathcal{O}(1000)$   $\text{CC1}\pi^\pm$  interactions in MicroBooNE using data from the first of five runs. Additionally, the methods developed in this thesis allow for the determination of the pion momentum,  $p_\pi$ , which has not yet been measured for the  $\text{CC1}\pi^\pm$  topology on argon by other experiments.





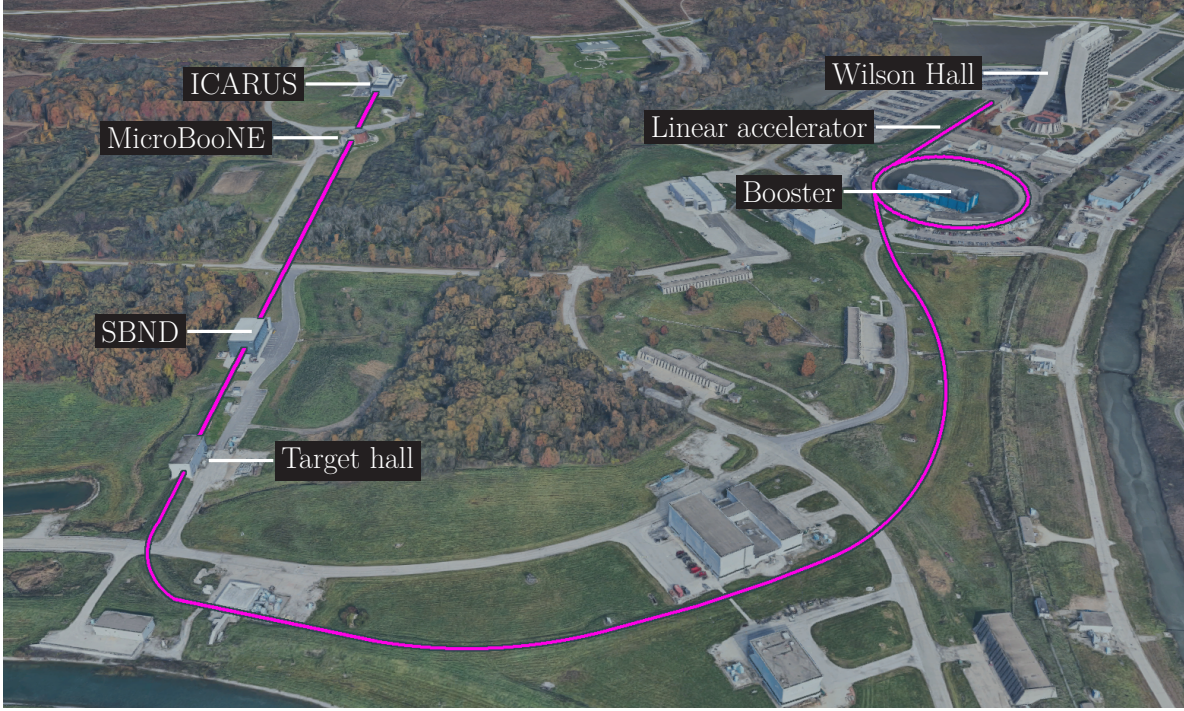
## Chapter 3

# The MicroBooNE experiment

The Micro Booster Neutrino Experiment (MicroBooNE) is a LArTPC detector located in Fermilab, IL, USA [4]. At a baseline of 470 m [4] from the Booster Neutrino Beam (BNB) source, MicroBooNE primarily measures muon neutrinos with an energy spectrum peaking at  $\sim 600$  MeV [72], as described in Section 3.1. MicroBooNE also receives an additional off-axis component of the Neutrinos at the Main Injector (NuMI) beam [73], which is the primary neutrino source for other experiments at Fermilab, however this beam is not discussed in detail as it is not immediately pertinent to the main content of this thesis.

One of the primary physics goals of MicroBooNE is to study Short Baseline Neutrino (SBN) oscillations. In particular, the experiment aims to investigate the low-energy excess of electron neutrino candidates from the BNB observed by the precursor MiniBooNE experiment when compared to the expected oscillated spectrum under the 3-neutrino hypothesis [74]. Additionally, MicroBooNE will form the middle detector of the upcoming SBN programme — along with the Short Baseline Near Detector (SBND) and ICARUS (far detector) [75], shown in Figure 3.1. However, most relevant to this thesis are MicroBooNE’s goals to measure precision cross-sections of various neutrino interactions on argon, and to develop experimental techniques (such as automated event reconstruction) to be used by future LArTPC experiments.

This chapter begins with a discussion of the BNB in Section 3.1, followed by the MicroBooNE detector itself in Section 3.2. Finally, Section 3.3 lists a number of experimental challenges related to LArTPCs, which will need to be addressed in the later chapters of this thesis.

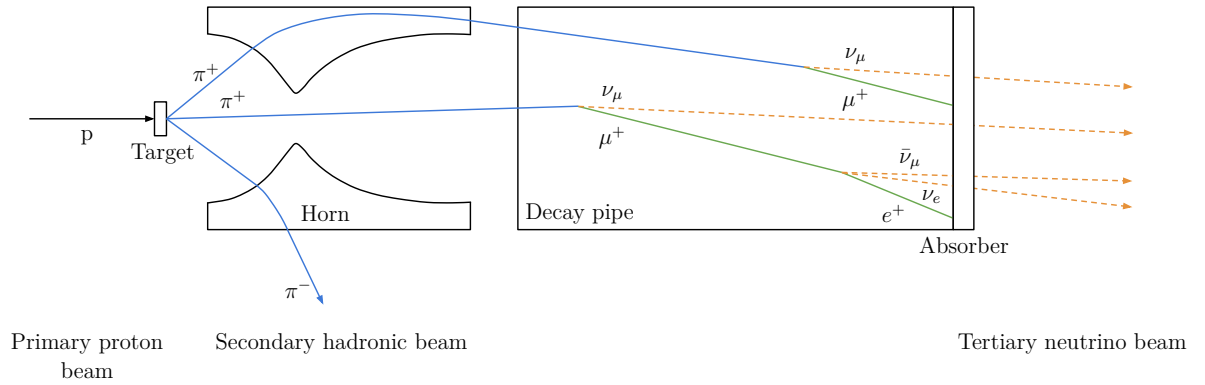


**Figure 3.1:** A map of the Fermilab site showing the path of the BNB (in magenta) and the locations of the three detectors in the SBN programme. Figure adapted from [76].

### 3.1 The Booster Neutrino Beam

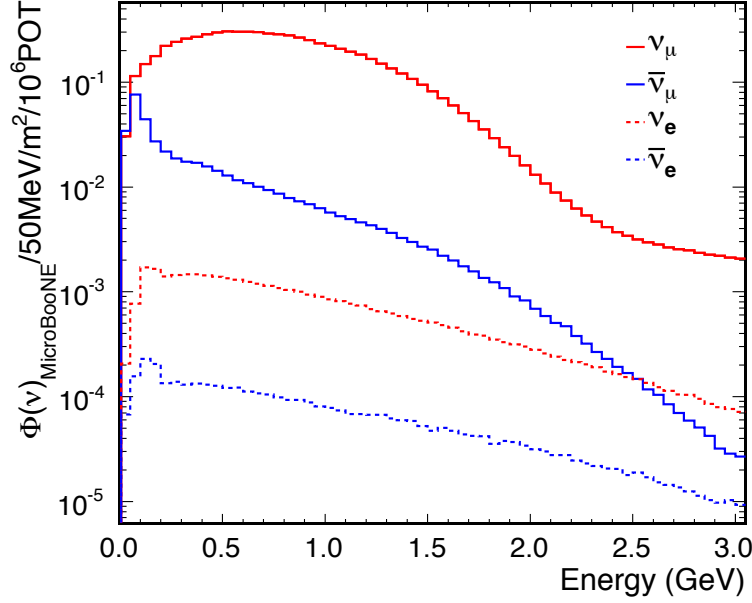
The BNB is designed to produce a pure beam of  $\nu_\mu$  with as few impurities from other neutrino flavours as possible. As shown in Figure 3.1, the Fermilab Booster synchrotron accelerates protons from a linear accelerator to a kinetic energy of 8 GeV [77]. These protons are then extracted and magnetically steered such that they impact upon a beryllium target in the target hall. The result is a secondary hadron beam composed primarily of charged pions [72]. As illustrated in Figure 3.2, this beam then passes through a magnetic focusing horn which — in neutrino mode — acts to collimate the positively charged hadrons, and deflect those charged negatively. In antineutrino mode, negative hadrons can be focused instead by reversing the horn polarity. The resulting collimated beam is comprised primarily of same-charge pions which are allowed to decay via  $\pi^+ \rightarrow \mu^+ + \nu_\mu$  to produce  $\nu_\mu$  in neutrino mode and  $\bar{\nu}_\mu$  via the conjugate process in antineutrino mode. At the end of the 50 m decay pipe, a concrete and steel absorber acts to stop particles from these decays which can further produce neutrinos of other undesired types [72]. The dominant undesired secondary decay is the  $\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu$  channel which contributes 51% of the total intrinsic  $\nu_e$  component of the beam with the remainder primarily from  $K^+$  and  $K^0$  decays [72]. The decay pipe length is chosen so

as to allow the maximal number of pions to decay with minimal contribution from the undesired secondary decays. The result, in neutrino mode, is a tertiary beam comprised primarily of  $\nu_\mu$  ( $\sim 93.6\%$ ), and  $\bar{\nu}_\mu$  ( $\sim 5.9\%$ ) peaking at an energy of 500–700 MeV, with  $\nu_e$  impurities at the  $\sim 0.5\%$  level below energies of 1.5 GeV [72] — this is the BNB. Figure 3.3 shows the predicted flux of  $\nu_\mu$ ,  $\bar{\nu}_\mu$ ,  $\nu_e$ , and  $\bar{\nu}_e$  from the BNB (in neutrino mode) through MicroBooNE as a function of the incoming neutrino energy [72].

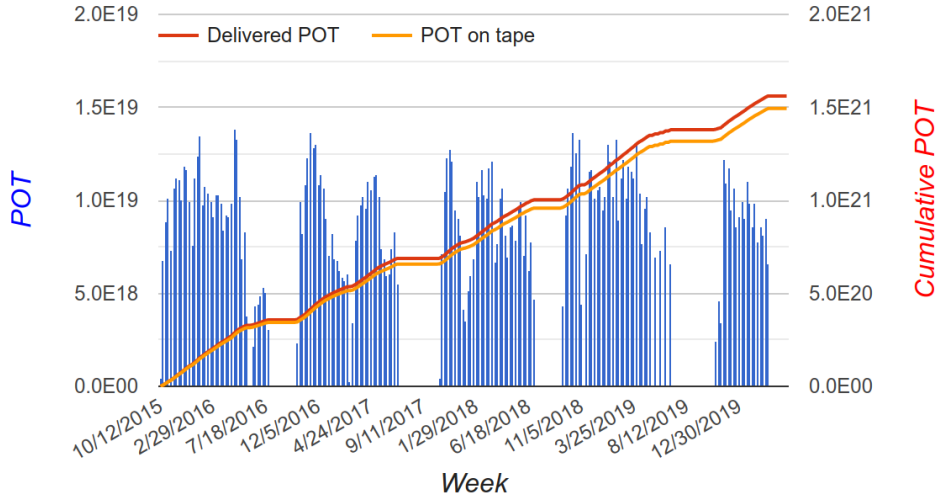


**Figure 3.2:** A schematic (not to scale) of the production of neutrinos in the BNB running in neutrino mode. The particle in black is a proton which impacts upon the target. The secondary hadrons are shown in blue, and the tertiary neutrinos in orange. The particles shown in green are produced with the neutrinos in the hadronic decays and are stopped by the absorber.

Nominally, each  $1.6 \mu\text{s}$  spill from the booster contains  $\sim 5 \times 10^{12}$  protons, which are delivered to the BNB target at a maximum allowable average rate of 5 Hz [72]. The amount of data collected by MicroBooNE is typically measured by the number of Protons-On-Target (POT), where (using the nominal values stated here, and assuming 9 month run-time) one calendar year maximally corresponds to  $\sim 5.8 \times 10^{20}$  POT. In practice, the actual POT rate available to MicroBooNE is somewhat less than this estimate and is shown in Figure 3.4 to be on average  $\sim 3.45 \times 10^{20}$  POT/year. The difference between the orange and red lines in this figure indicate the detector uptime which is  $\sim 96\%$ .



**Figure 3.3:** The BNB neutrino flux prediction through MicroBooNE, averaged through the TPC volume. The peak of  $\bar{\nu}_\mu$  at low energies arises primarily from the pion-to-muon decay chain. Taken from [72].



**Figure 3.4:** The Protons-On-Target delivered to the BNB per week. Each blue bar shows the POT delivered in a given week. The red line shows the cumulative POT delivered, and the orange line refers to the amount of data that was successfully measured and stored on tape by MicroBooNE. The difference between the orange and red lines is a measure of the detector uptime. This plot was obtained from the experiment's internal operational monitoring website, and so the dates should be read in US month/day/year format.

## 3.2 The MicroBooNE detector

Broadly, MicroBooNE consists of three detector subsystems: the Time Projection Chamber (TPC), the Light Collection System (LCS), and the Cosmic-Ray Tagger (CRT). The TPC provides high-resolution positional and calorimetric information, while the LCS and CRT supplement the TPC with additional positional and, crucially, superior timing information. This section begins with a general discussion of argon as a detector medium, followed by the design and operational principles of the three detector subsystems.

### 3.2.1 Charged particles interactions in liquid argon

As a charged particle travels through liquid argon, it can impart energy to the argon atoms leading to excitation or ionisation. In general the number of electrons liberated through ionisation in a small distance  $dx$  travelled by the particle scales with the energy lost  $dE$  by the particle in that distance. The particle will continue along its trajectory, leaving a trail of ionisation electrons in its wake until insufficiently energetic to further ionise the argon. The distribution of the resulting cloud of ionisation electrons defines the trajectory of the particle, and the density defines the rate of energy loss,  $dE/dx$ . Additionally, the particle can undergo elastic or inelastic scatters on the argon atoms which will change its trajectory and, in the latter case, potentially liberate nucleons or produce other hadrons which may themselves leave ionisation trails.

If the argon volume is subject to an external electric field (as is the case for LArTPCs) the resulting ionisation electrons and their argon-ion counterparts will be separated. However, at the point of ionisation some fraction of these electron-ion pairs will attract and recombine. The rate at which the electron-ion pairs are separated, and hence the magnitude of this recombination effect is then dependent on the strength of the external electric field. A stronger field typically leads to lower rates of recombination. [78]

Once separated, the ionisation electrons will drift through the argon under the influence of the electric field. In a uniform field, the electrons initially accelerate but quickly reach a constant drift velocity governed by the field strength and the mobility of the electrons through liquid argon. The mobility is an empirically determined property of a material which defines how quickly an electron will travel through that material. Typically, the thermal velocity of the electrons is significant with respect to their drift velocity. The resulting random walk of the electrons manifests at the distribution level

as a spatial diffusion of the electron cloud. The scale of the diffusion is hence dependent on the amount of time for which the electrons drift. [79]

Although argon is a noble element, a real-world detector will inevitably have non-argon electronegative impurities such as water and oxygen. While drifting, some fraction of the number of initial electrons in the cloud,  $N_0$ , will be captured by these impurities resulting in an exponential decay in the total number of electrons  $N$  as a function of drift time  $t$ , as given by:

$$N(t) = N_0 e^{-t/\tau}. \quad (3.1)$$

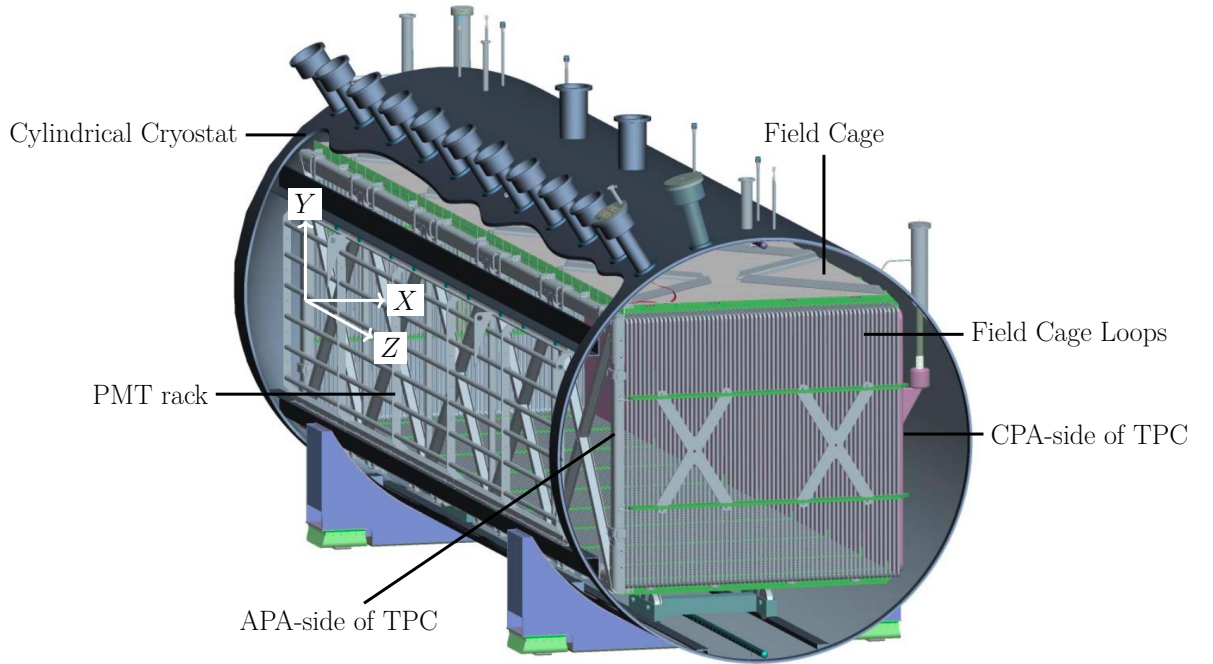
Here,  $\tau$  is the electron lifetime which is governed by the purity of the argon. [4]

In addition to these ionisation signatures, charged particles can also produce optical signals via scintillation and Cherenkov radiation. In liquid argon, scintillation primarily occurs when excited and bonded argon atom pairs (so called excimers) dissociate and reach their respective ground states through photonic emission [80]. The argon excimer has two low-lying excited states which decay to the ground state after characteristic time intervals. Consequently, the scintillation light resulting from a charged particle has a fast ( $\sim 6$  ns), and a slow ( $\sim 1.5$   $\mu$ s) component [80]. In both cases, the emitted scintillation light has a narrow wavelength spectrum peaking at 128 nm in the Vacuum Ultraviolet (VUV) range, and is emitted isotropically [80]. Liquid argon is a bright scintillator producing  $\mathcal{O}(10^4)$  photons per MeV of deposited energy under typical LArTPC electric field strengths, and is also transparent to its own scintillation light [4]. Therefore, scintillation light can readily be measured and is a complementary source of information to the ionisation signatures.

### 3.2.2 Operational principles of the MicroBooNE LArTPC

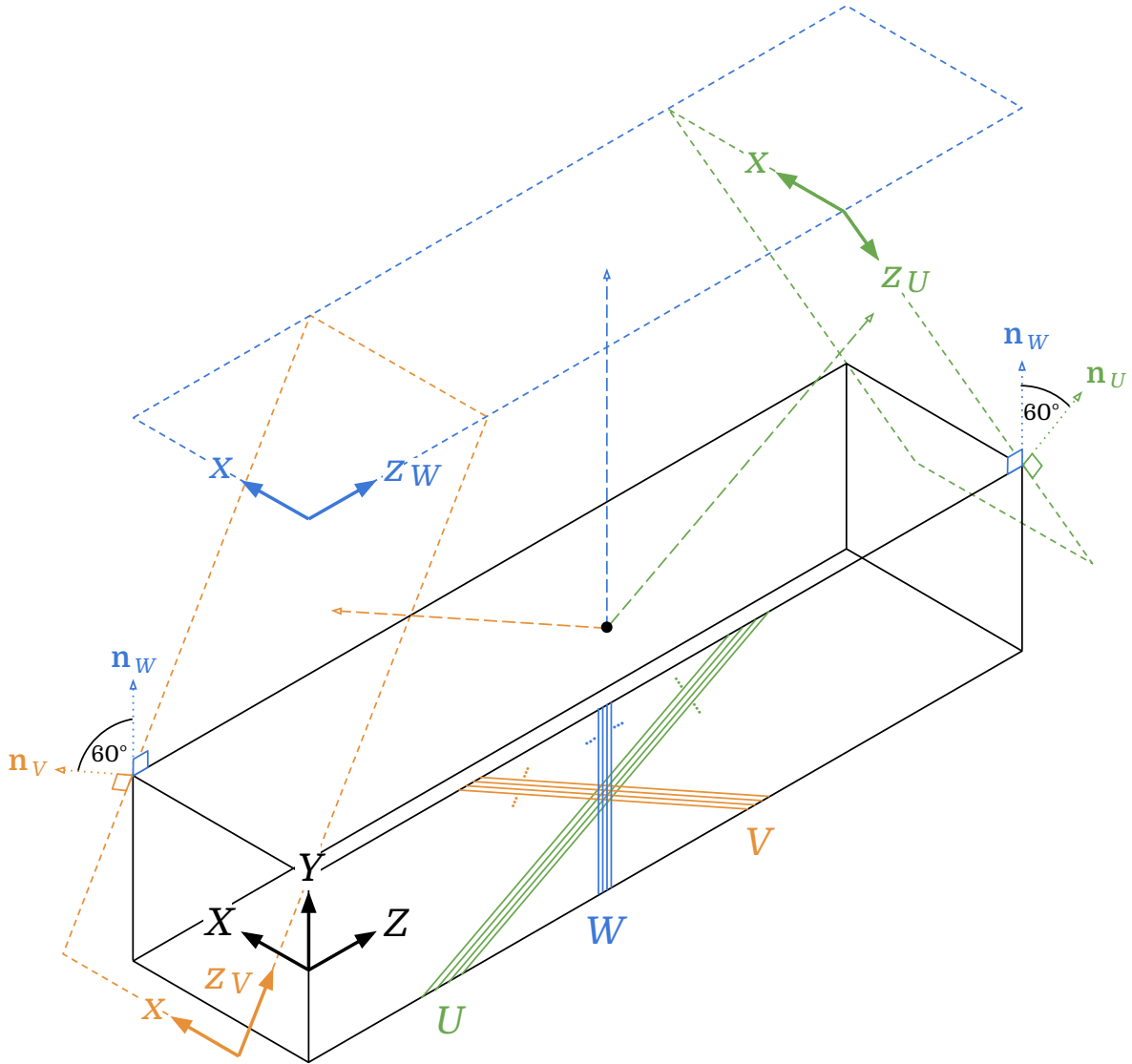
Like all LArTPC detectors, MicroBooNE consists of a volume of liquid argon instrumented to detect the aforementioned ionisation signatures. As depicted in Figure 3.5, a cuboidal field cage surrounds the active detector volume. The cage consists of 64 stainless steel tubes which are shaped into rectangular loops [4]. An electric field is produced via a potential difference between between the Cathode Plane Assembly (CPA) and the Anode Plane Assembly (APA) and a resistor divider chain links the field-cage loops such that the voltage is stepped evenly from the CPA to the APA at each loop to produce a uniform field [4]. Ionisation electrons move toward the APA with a constant drift com-

ponent of their velocity,  $v_d$ , due to the electric field. By convention, the APA is defined to be at  $X = 0$ , and the electrons drift in the negative- $X$  direction. As illustrated in Figure 3.6, the APA consists of 3 wire planes labelled U, V, and W. Each wire plane spans the entire anode side ( $YZ$ ) of the TPC and consists of many finely and equally spaced wires, each pointing at a constant angle in the plane. The  $Y$ -direction is defined along the upward vertical direction, and  $Z$  is along the beam axis such that all three directions,  $X, Y, Z$  are orthogonal. The wire angles are maximally different among the planes, i.e.  $\theta_{YZ} = 0^\circ$ , and  $\pm 60^\circ$  for W, U and V respectively, where  $\theta_{YZ}$  is the angle in the  $YZ$ -plane to the vertical. The planes themselves are finely spaced and ordered in  $X$  where U has the largest  $X$ , followed by V and W in turn. The U and V planes are held at bias voltages so a drifting electron will induce a current on the wires and continue on toward the W plane which is itself biased such that the electron will collect onto a wire producing a current [4]. For this reason, U and V are collectively known as the induction planes, and W is the collection plane. The details of this setup are summarised in Table 3.1.



**Figure 3.5:** The MicroBooNE TPC shown housed in the cryostat. The high- $Z$  end cap of the cryostat is here not shown. Figure modified from [4].





**Figure 3.6:** Illustration of the geometrical connection between the 3D and three 2D coordinate systems used by MicroBooNE. The black cuboid is the TPC active volume and the dashed planes are the projections of the active volume into each of the three views. The solid coloured lines on the low- $X$  side of the TPC illustrate a few of the sense wires for each plane, in reality these wires span the full face. The black dot shows how an example 3D position is projected into each view.



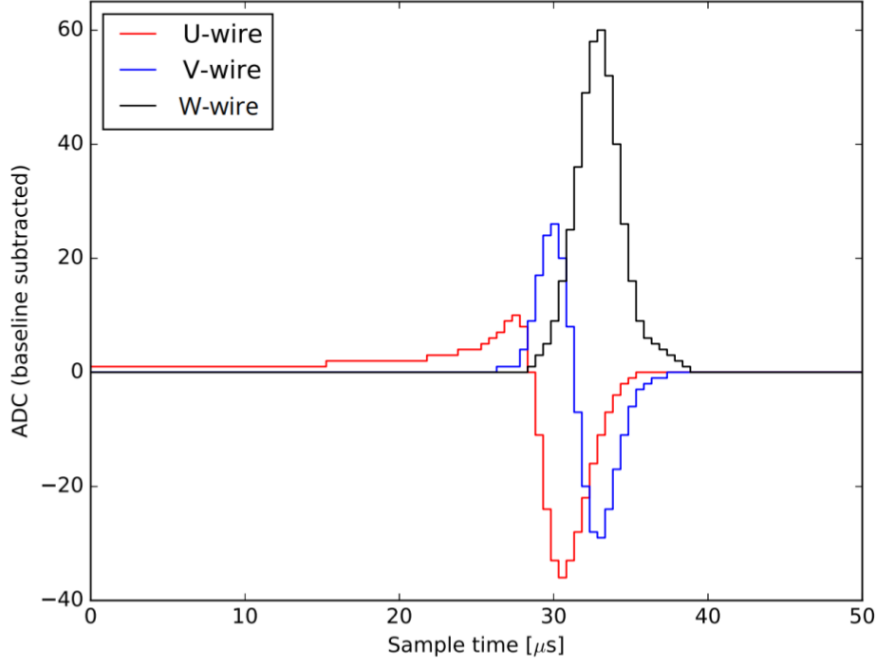
Parameter	Symbol	Value
Width (drift direction)	$L_X$	2.560 m
Height (vertical direction)	$L_Y$	2.325 m
Length (beam direction)	$L_Z$	10.368 m
Wire spacing	$s$	3 mm
Inter-plane spacing		3 mm
Number of U wires		2400
Number of V wires		2400
Number of W wires		3456
U wire angle	$\theta_{YZ}^U$	+60°
V wire angle	$\theta_{YZ}^V$	-60°
W wire angle	$\theta_{YZ}^W$	0°
Field strength		273 V/cm
Drift velocity	$v_d$	$1.1 \times 10^3$ m/s
Proposed field strength		500 V/cm
Drift velocity at proposed field strength		$1.6 \times 10^3$ m/s

**Table 3.1:** Design parameters of the MicroBooNE LArTPC [4]

A localised deposition of ionisation electrons at a point  $(X_0, Y_0, Z_0)$  and time  $t_0$  will on average reach the APA at time,  $t_M$ , given by:

$$t_M = t_0 + \frac{X_0}{v_d}. \quad (3.2)$$

As a result, a signal peak  $I_i^P(t)$ , will be produced on wires labelled by plane,  $P = U, V, W$  and wire number,  $i$ , as a function of time,  $t$ . The shape of these signals will depend on  $P$ , and is bipolar for induction plans and unipolar for the collection plane — as shown in Figure 3.7 [81]. A 3D distribution of charge along multiple particle trajectories will produce signals on multiple wires. By plotting these wire signals as a function of time and the (position ordered) wire number for each plane individually,  $I^P(t, i)$ , we obtain three two-dimensional images of the particle trajectories — one for each plane.

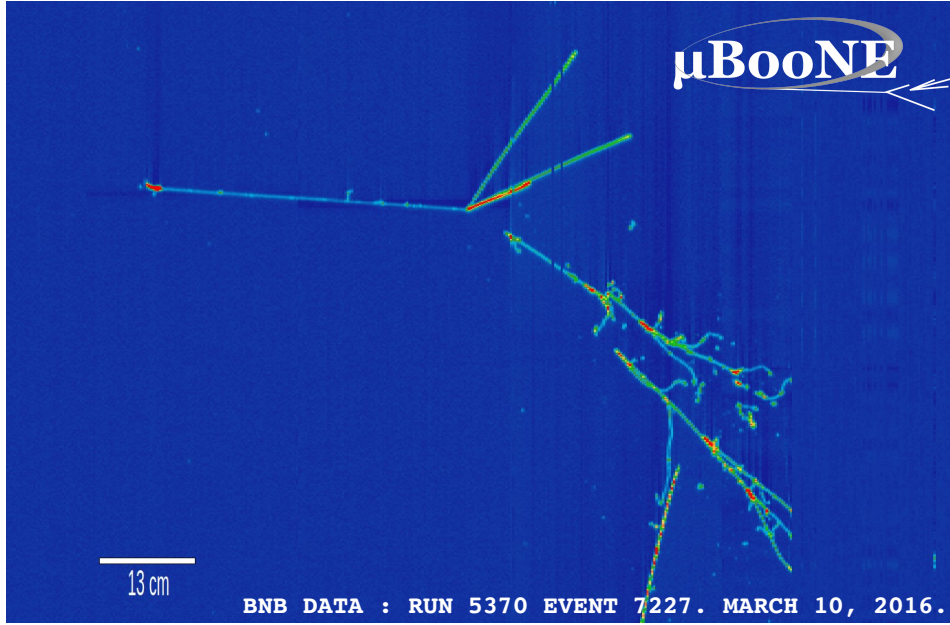


**Figure 3.7:** Simulated responses for a single wire from each of the three readout planes, U, V, W. A bipolar peak is produced on induction planes, and a unipolar peak is produced on the collection plane. The wire responses  $I_i^P(t)$  are here presented in baseline-subtracted ADC which is the output of the Analogue-to-Digital Converter electronics on the raw wire signals. The simulation shown here is for an ideal line source of charge deposited perpendicular to all wires, with no simulated noise. In practice MicroBooNE experiences significant noise which is removed while processing the wire signals. Figure modified from [81], which includes more details of the noise removal process.

To produce such an image,  $t$  and  $i$  are typically scaled by  $v_d$  and the wire spacing,  $s$ , respectively such that they can be compared on the same footing with dimensions of distance, as given by:

$$\begin{aligned} x &= v_d(t - T), \\ z_P &= is. \end{aligned} \tag{3.3}$$

The time  $T$  is conventionally chosen to be the time at which the readout was triggered, such that for charge deposited at  $t_0 = T$ , the position  $x = X_0$ . More details of MicroBooNE’s trigger system are given in Section 3.2.6. An example image from MicroBooNE is shown in Figure 3.8.



**Figure 3.8:** An example BNB data image from the collection plane of charged particle trajectories from a neutrino interaction in MicroBooNE. In this case the wire number coordinate,  $z_W$ , is shown along the horizontal axis and the time coordinate,  $x$ , is along the vertical. The colour scale shows the magnitude of the signal after noise removal and signal processing.

It is important to note that the three readout images (now dropping the redundant labels),  $I(x, z_U)$ ,  $I(x, z_V)$ ,  $I(x, z_W)$ , share the common coordinate  $x$ . Geometrically, this means each of the three 2D coordinate systems  $(x, z_U)$ ,  $(x, z_V)$ ,  $(x, z_W)$  are projections of the full 3D coordinate system  $(X, Y, Z)$  onto the planes defined by normals  $\mathbf{n}_P = (0, \cos(\theta_{YZ}^P), \sin(\theta_{YZ}^P))$ . Said otherwise, the three readout images are orthographic 2D projections of the 3D particle trajectories from different angles of rotation around the  $X$ -axis, as illustrated in Figure 3.6. As such, given any two 2D coordinates  $(x, z_A)$ ,  $(x, z_B)$  one can determine the corresponding 3D position  $(X, Y, Z)$  and hence the projection  $(x, z_C)$ , where  $ABC = UVW$  and cyclic permutations thereof. Consequently, a 3D image of the particle trajectories can be reconstructed from the three 2D images. This is the basic operational principle of a LArTPC.

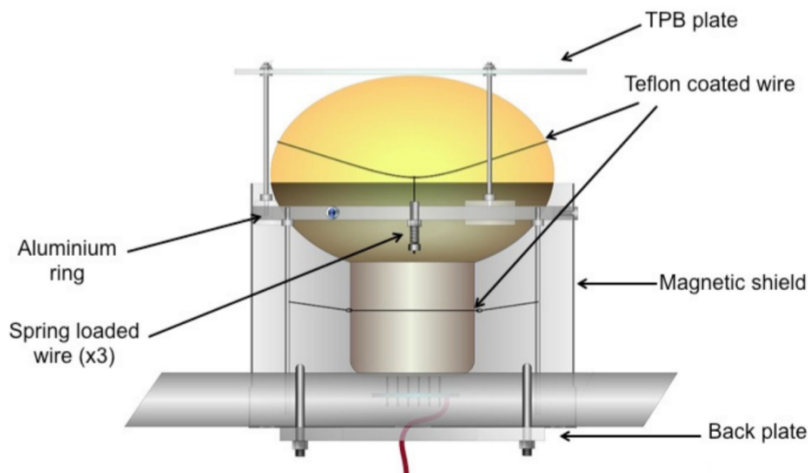
It should also be stressed that the coordinate  $x$  is calculated from the time  $t_M$  at which the signal is measured, not the time  $t_0$  at which the charge was deposited. From Equation 3.2 and Equation 3.3, one can see that for a deposition of charge at  $(X_0, Y_0, Z_0, t_0)$ , the measured position,  $x$  (and equivalently  $X$ ), differs by  $\Delta x$  from  $X_0$ , as given by:

$$\Delta x = x - X_0 = v_d(t_M - T) - X_0 = v_d(t_0 - T). \quad (3.4)$$

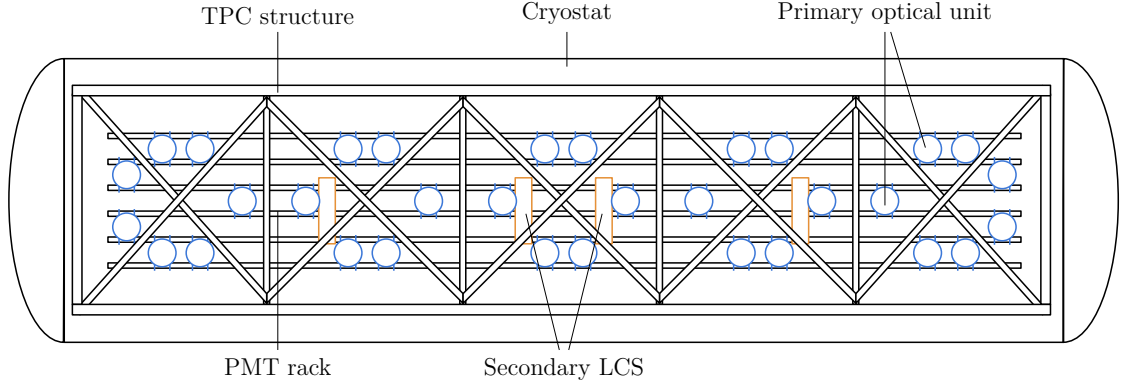
As noted above,  $\Delta x = 0$  for particles produced at the trigger time,  $T$ . However, all other particles will appear shifted by  $\Delta x$ . The  $t_0$  of a given particle typically is not known a-priori, and additional timing information is required from the LCS or CRT to supplement the TPC.

### 3.2.3 The Light Collection System

MicroBooNE’s primary LCS consists of 32 optical units. As shown in Figure 3.9, each optical unit comprises a Photomultiplier Tube (PMT) housed behind a wavelength-shifting plate and surrounded by a “magnetic shield” [4]. As described in Section 3.2.1, Liquid Argon (LAr) scintillation light is emitted in the VUV range. In order to detect this light with the PMTs employed by MicroBooNE, it must be shifted into the visible region; this is the purpose of the acrylic wavelength-shifting plates that are coated with Tetraphenyl butadiene (TPB) — a fluorescent compound that absorbs in VUV and emits in the visible range [4]. The purpose of the mu-metal magnetic shield (a nickel-iron alloy with high magnetic permeability) is to mitigate the effect of the Earth’s magnetic field on PMT response. The optical units are mounted on the PMT rack, shown in Figure 3.5 and Figure 3.10, which is located behind the anode side of the TPC. In addition to this primary system, MicroBooNE also employs a secondary LCS for research and development purposes. This secondary system is not used in the work presented in this thesis and so isn’t described here, see reference [4] for more details.



**Figure 3.9:** A schematic of a MicroBooNE optical unit. Modified from [4].



**Figure 3.10:** A cartoon of the placements of optical units with respect to the structural components of the MicroBooNE TPC. Each blue circle represents one optical unit in the primary LCS. The orange rectangles are components of the secondary R&D LCS. Modified from [4].

### 3.2.4 Cryogenic system

MicroBooNE’s TPC and LCS are housed within the cryostat as shown in Figure 3.5 and Figure 3.10. The cryostat is cylindrical in shape with domed end caps, and is 12.2 m in overall length and 3.81 m in inner diameter [4]. In order to minimise variations in the drift velocity due to temperature gradients, a 41 cm thick spray-on foam insulation covers the outer cryostat surfaces to reduce the impact of ambient temperature differences to acceptable levels. Heat is inevitably added to the system from the active detector components and the external environment. To maintain the requisite stable operating temperatures, MicroBooNE is cooled by a nitrogen refrigeration system. The cooling power is provided by two nitrogen condensers, one used for normal operations and the remainder as backup. [4]

In order to maximise the electron lifetime,  $\tau$  (see Equation 3.1), MicroBooNE implements a LAr purification system. The system circulates, filters and pumps clean LAr into the cryostat such that the entire volume is recirculated every 2.5 days [4]; this ensures the levels of oxygen and water impurities are within the desired threshold. The whole set-up successfully maintains temperature variations below the design threshold of 0.1 K and laminar and turbulent LAr flow rates to below 1 m/s; these limits correspond to fractional errors in the drift velocity below 0.1% [4]. Temperature and purity monitoring systems ensure the cryogenic system performs within the prescribed thresholds.

Indeed, MicroBooNE has achieved a stable purity over all runs that surpasses the design threshold.

### 3.2.5 The Cosmic-Ray Tagger

MicroBooNE is a surface detector and is hence subject to cosmogenic particles. As the time taken for an ionisation electron to drift along the full width of the detector is  $\mathcal{O}(\text{ms})$ , these cosmic-rays are numerous and a significant background to any analysis of neutrino interactions. To help identify these backgrounds, MicroBooNE installed a CRT system after the commissioning of the TPC and LCS systems was complete. Table 3.2 summarises the data available that was collected with the CRT in place. The CRT is not used within the main content of this thesis, but is described briefly here.

The full CRT system is modularised. Each module consists of 16 plastic scintillating strips of width 10.8 cm and thickness 2 cm; the length of the strips varies from 1.3–4.1 m among modules. The strips are placed side-by-side and are encased by a 2 mm thick aluminium casing. The result is a panel of width 1.75 m [82]. Scintillation light produced by a cosmic-ray within a strip is guided via an optical fibre to a silicon photomultiplier where the resulting signal is read out. Each strip contains two fibres in grooves along its long edges, which (by comparing the relative amplitudes from each edge) gives a positional resolution of 1.8 cm in the width direction as measured by laboratory tests [82]. The timing resolution achieved is  $\sim 4 \text{ ns}$ , which is  $\sim 3$  orders of magnitude faster than that of the TPC [82].

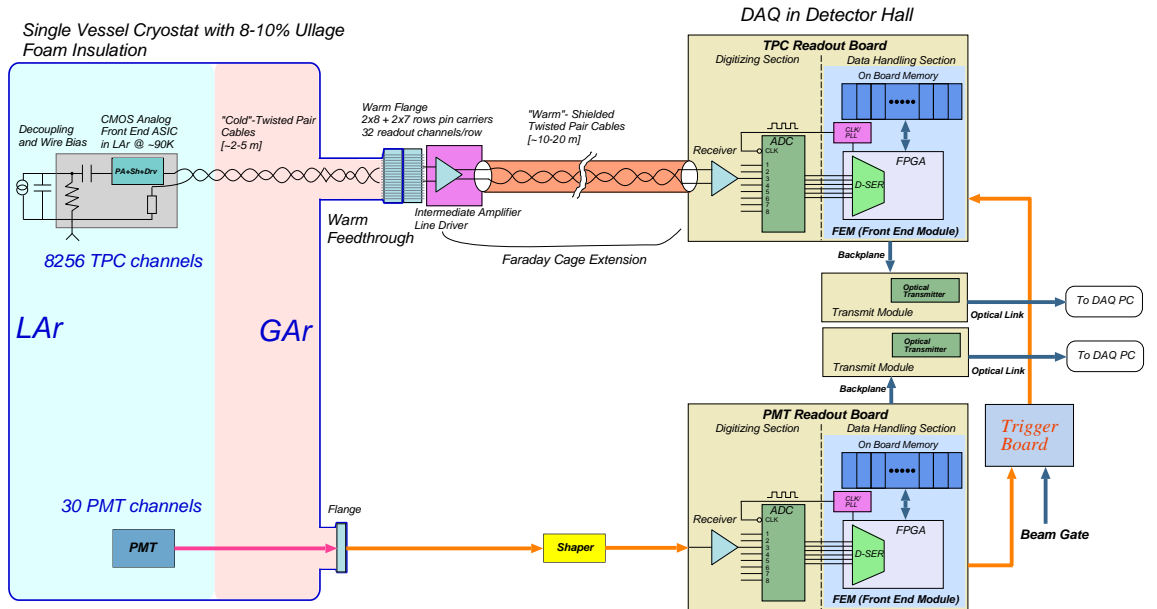
In total, 73 CRT modules cover the anode, cathode, bottom and top sides of MicroBooNE and are external to the cryostat. The modules are arranged so as to maximise coverage without obstructing existing detector components. The total cosmic-ray muon coverage is 85% [82]. On each detector face, the modules are arranged in two layers oriented at right angles; this allows for the simultaneous measurement of orthogonal components of the position of a crossing cosmic-ray.

Run number	Begin	End	CRT available?
1	Oct 2015	July 2016	No
2a	Oct 2016	March 2017	Bottom and side only
2b	March 2017	July 2017	Yes
3	Oct 2017	July 2018	Yes
4	Sep 2018	July 2019	Yes
5	Nov 2019	March 2020	Yes

**Table 3.2:** MicroBooNE’s run periods with and without the CRT installed. Periods between runs correspond to times at which the BNB was shut down, as can be seen in Figure 3.4.

### 3.2.6 Readout and trigger systems

Once a signal is detected by one of the subsystems, it must be read out and saved for future processing. The TPC and LCS are read out together, as is described in this section, whereas the CRT is handled separately and described in reference [82]. Figure 3.11 gives an overview of the readout system described below.



**Figure 3.11:** An overview of the MicroBooNE TPC and LCS readout system [4].

First, the signals from TPC wires are shaped and preamplified by an Application Specific Integrated Circuit (ASIC). Each ASIC is mounted on a cold motherboard that is directly attached to the wire carrier boards within the LArTPC itself, and hence

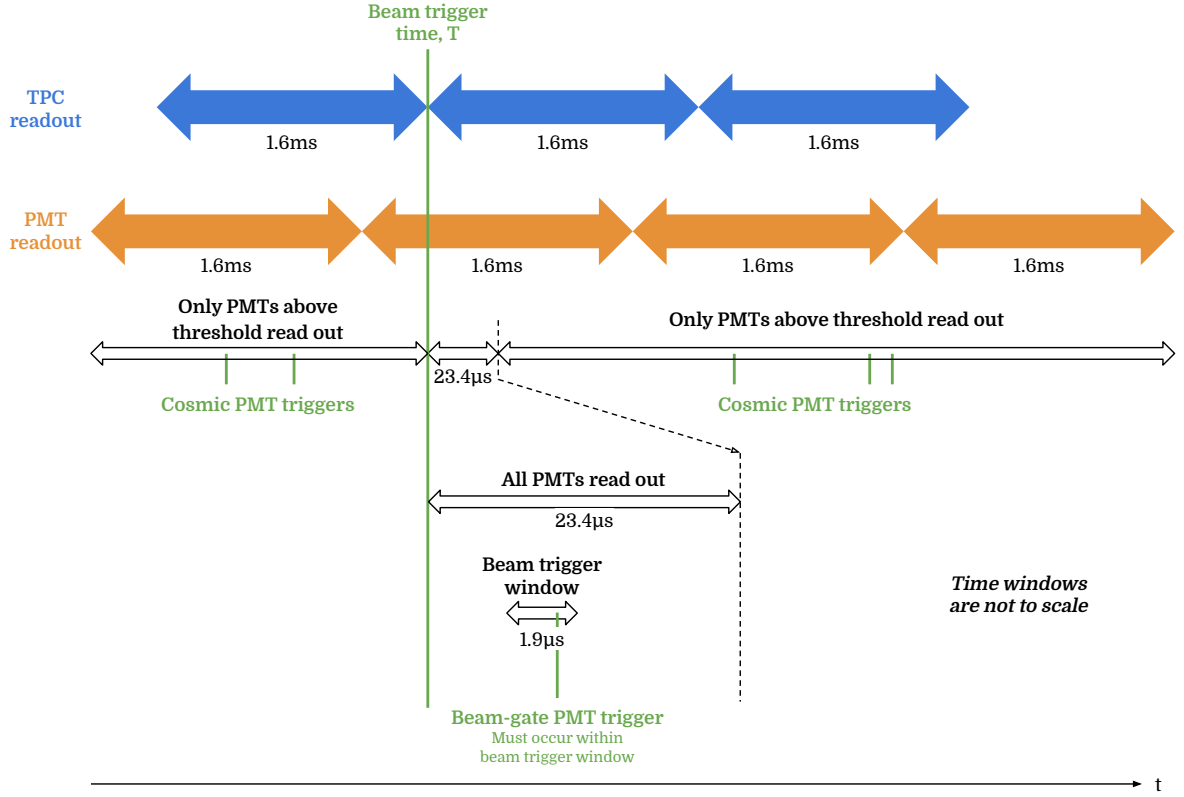
operates at cryogenic temperatures. In this way the wire-to-preamplifier interconnection distance is minimised which results in a high signal-to-noise ratio. The preamplified signals from the motherboards are transmitted  $\sim 2\text{--}4.5\text{ m}$  via so called “cold cables” out of the cryostat. At this point, intermediate amplifiers housed in a Faraday cage provide sufficient gain such that the signals are suitable to be transmitted  $\sim 10\text{--}20\text{ m}$  through warm cables to the readout boards in the detector hall. Upon reaching the readout boards the analogue wire signals are quantised via an Analogue-to-Digital Converter (ADC). This digitised data is handled by the Front End Module (FEM) on the readout board. The FEM compresses and stores the data in memory buffers until triggered to be transmitted via optical cables to dedicated Data Acquisition (DAQ) machines for further processing [4].

Each ASIC can read out 16 wires and each cold motherboard can either take 6 or 12 ASICs (or equivalently 96 or 192 wire channels) depending on its position. Motherboards mounted at the top of the APA have 192 channels (taking 48 U, 48 V, and 96 W wires); those motherboards mounted on the side have 96 channels that take either U or V wires. In total MicroBooNE has 36 top-type and 14 side-type motherboards corresponding to a total of 8256 channels (2400 U, 2400 V, and 3456 W) as given in Table 3.1. Each ADC/FEM readout board can handle signals from 64 wires, and a total of 130 such boards are used. The readout boards are distributed among 9 crates which transmit to 9 DAQ machines — one per crate [4].

The LCS signals are shaped, amplified, digitised, buffered and read out to the DAQ machines in a similar manner to those from the TPC. One main difference is that the signal from each PMT is split into two, and each channel is amplified with different gains. Upon analysis, the high-gain channel is preferentially used for smaller signals, and can be supplemented by the low-gain channel if it begins to saturate for large signals.

MicroBooNE employs two simultaneous data streams to meet different physics goals. The supernova stream reads out continuously with a lossy compression scheme to reduce data rates. The data is stored by the DAQ system for  $\sim 1$  day and only saved permanently for offline analysis in the event of a “supernova early warning system” alert from other larger neutrino detectors around the globe. By contrast, the neutrino stream (pertinent to this thesis) employs a lossless compression scheme, reads out only when triggered and stores all data permanently for offline analysis. The remainder of this section describes the neutrino stream only, and is summarised in Figure 3.12 [4].





**Figure 3.12:** The timing structure of MicroBooNE’s neutrino readout stream in response to a BNB or fake-beam trigger. [83]

The Trigger Board (TB) controls which of the data held in the FEM buffers should be transmitted to the DAQ machines. The TB takes trigger inputs from various sources which are combined with boolean logic to form the final result as described below. Beam trigger inputs fire when a spill occurs for the BNB, NuMI (not discussed here), or a “fake” beam. The fake beam trigger is simply a strobe with configurable frequency used to read out no-beam data in the same manner as if there were a BNB beam spill. The fake beam trigger won’t fire in coincidence with a real beam. Additionally, the TB takes trigger inputs from the PMT FEMs and has other sources for calibration purposes. [4]

For normal data taking, a beam trigger is required to fire for readout to occur. However, on account of the neutrino-argon cross-section, flux and detector volume, the majority of beam spills won’t result in a neutrino interaction in the detector. Consequently, an additional beam-gate PMT trigger is also required to reduce the data rate of neutrinoless events. To fire, the beam-gate PMT trigger demands a sufficiently large optical signal in coincidence with the expected arrival time of a neutrino. The start time of this *beam*

*trigger window* is determined empirically and has a duration of  $1.9\,\mu\text{s}$ , corresponding to the  $1.6\,\mu\text{s}$  BNB spill window with a 300 ns padding.

Once readout is triggered, TPC data is saved by the DAQ system for the interval  $T - 1.6\,\text{ms} \rightarrow T + 2 \times 1.6\,\text{ms}$ , where  $T$  is the time of the beam trigger. The interval 1.6 ms corresponds to the maximum drift time,  $L_X/v_d$ , at the proposed electric field strength. In practice, MicroBooNE operates with a lower field than initially designed (at which  $L_X/v_d = 2.3\,\text{ms}$ ); nevertheless, the readout interval is still sufficient to capture the full drift volume. PMT data is recorded in 1.6 ms frames, and readout occurs for the frame containing  $T$ , the preceding frame and the two following frames. Data from all PMTs is saved in the  $23.4\,\mu\text{s}$  interval following  $T$ , however (to reduce data rates from cosmic-rays) for the remaining time, PMTs must individually fire a cosmic PMT trigger to be read out. This trigger is effectively a threshold on the magnitude of the signal from the PMT. The  $23.4\,\mu\text{s}$  is sufficient to capture light from the primary neutrino interaction and secondary interactions such as particle decays.

### 3.3 Experimental challenges

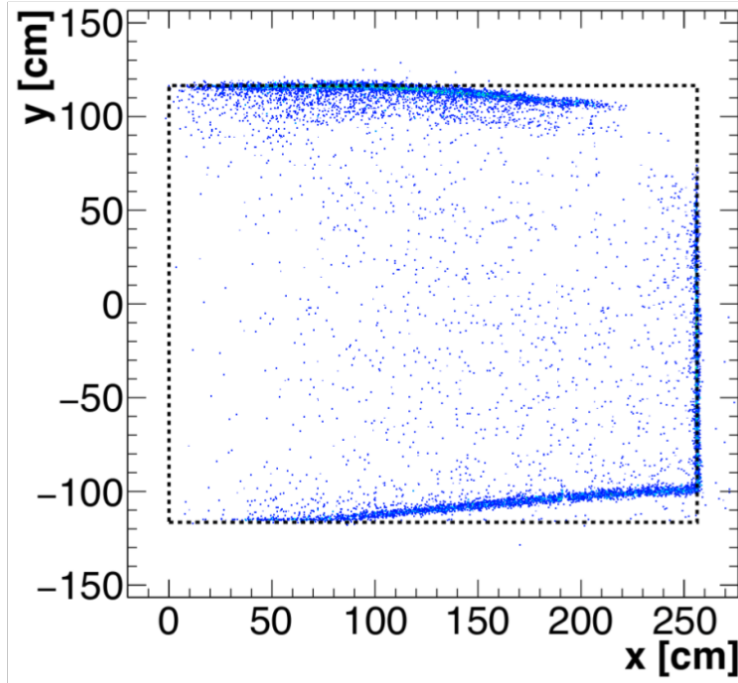
After the data is taken, there exist a number of challenges that must be addressed at the analysis level related to the operation of the detector which are described here.

#### 3.3.1 The rate of cosmic rays

One of the major backgrounds for any analysis of neutrinos with MicroBooNE are cosmic rays. A typical event (corresponding to the readout shown in Figure 3.12) contains  $\mathcal{O}(10)$  cosmic-ray muons that produce ionisation signals in the detector. For a given event it is non-trivial to distinguish the signals due to the neutrino from those due to cosmic rays. The methods by which this is achieved are discussed in Part II of this thesis. Additionally, due to the high rate of cosmic rays, even with the requirement that the beam trigger is coincident with the beam-gate PMT trigger, around 5 in 6 events recorded will not contain a neutrino interaction [7].

### 3.3.2 The Space-Charge Effect

The positive argon ions produced primarily by cosmic-rays drift under the influence of the electric field toward the cathode where they are neutralised. These ions drift significantly more slowly than the ionisation electrons and this, coupled with the large cosmic-ray flux, results in a build up of positive charge throughout the detector. To first order, the density of this positive charge doesn't depend on  $Y$  or  $Z$  and increases linearly in  $X$  from zero at the anode [84]. This positive charge distorts the uniformity of the electric field experienced by the ionisation electrons causing non-linearities in their drift paths and hence deformations of the particle trajectories seen in the output images. Qualitatively, trajectories are deformed toward the centre of the detector, and the strength of the effect increases with distance from the anode. Figure 3.13 shows the impact of this effect on the end points of through-going cosmic-rays in MicroBooNE data. This is known as the Space-Charge Effect (SCE) and has been modelled, simulated and measured in data by MicroBooNE. The result of this work [84] is a lookup table whereby reconstructed 3D positions can be SCE-corrected to account for these deformations.

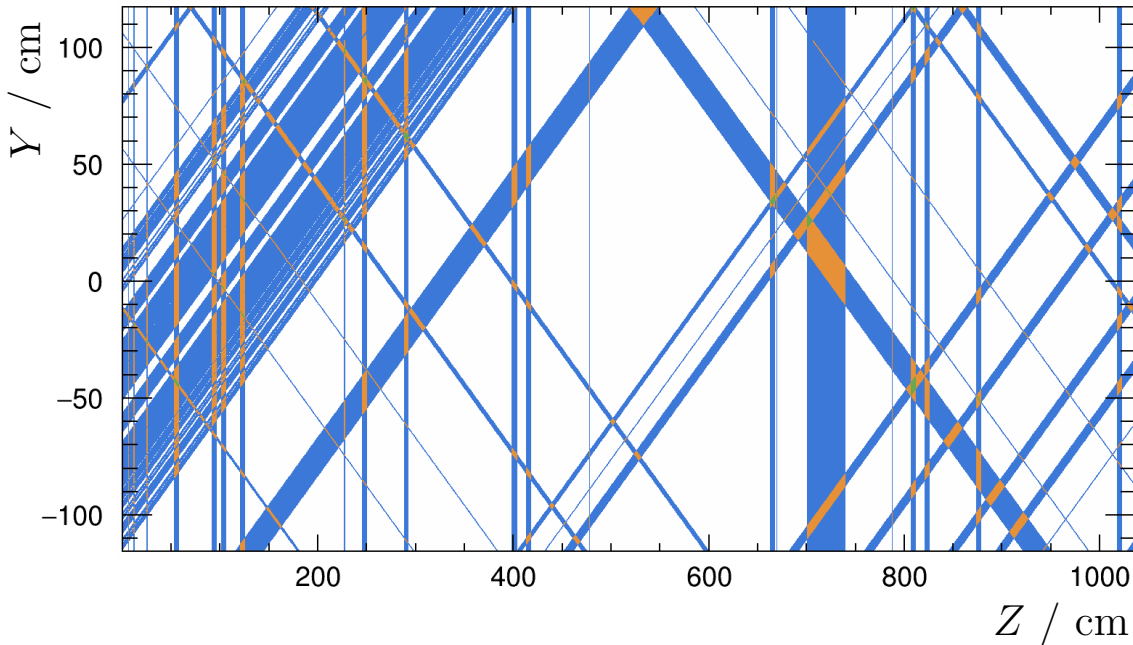


**Figure 3.13:** Start and end points of cosmic-ray muons reconstructed from MicroBooNE off-beam data and  $t_0$  tagged by an external muon counter system. The vast majority of cosmic-ray muons are through-going and hence the end points are expected to pile up at the TPC boundaries (shown by the dotted rectangle). A significant deformation of the particle trajectories is observed due to space charge. Taken from [84].

### 3.3.3 Unresponsive detector regions

A number of the wire channels are functionally dead [81]. The majority of these dead channels are due to touching wires from different planes ( $U \leftrightarrow V$  and  $V \leftrightarrow W$ ) that result in a short circuit. Additionally, some channels are lost due to misconnected or dysfunctional ASICs. The bias voltage on the shorted wires is effected which can lead to charge being collected on  $U$  or  $V$ -plane wires. The result is a shielding effect whereby functional wires on the  $W$ -plane that are behind shorted  $U \leftrightarrow V$  wires see less charge under normal operating conditions. However, this effect is removed by calibration [85].

The dead channels are known and must be accounted for during the reconstruction and analysis stages. As shown in Figure 3.14,  $\sim 17\%$ ,  $\sim 5\%$  and  $\sim 10\%$  of the channels are unresponsive for the  $U$ ,  $V$  and  $W$  readout planes respectively. Consequently,  $\sim 30\%$  of the detector volume is not visible in at least one of the output images. However, as only two planes are strictly required to reconstruct 3D particle trajectories, only  $\sim 2.5\%$  of the detector volume is unusable in this regard. Furthermore, only  $\sim 0.04\%$  of the volume is unresponsive on all planes.



**Figure 3.14:** Unresponsive regions of the MicroBooNE TPC. Regions are colour-coded based on the number of readout planes (0 white, 1 blue, 2 orange, 3 green) that are unresponsive at that point. The X-axis points into the page.

### 3.4 MicroBooNE simulations

The MicroBooNE collaboration uses a Monte-Carlo (MC) simulation of neutrino interactions which can be factorised into five stages, described below.

1. **Flux simulation.** MicroBooNE uses a simulation [86] of the BNB that was initially developed by MiniBooNE [77]. The production rate of secondary hadrons in proton-beryllium collisions is calculated using a double-differential cross-section for each hadron species as a function of the momentum and direction of the hadron produced. For the dominant ( $\pi^\pm$ ) mode, this cross-section is defined using the Sanford-Wang (SW) parametrisation [87] which has been fit to p-Be data at Booster proton energies from the HARP experiment [88] using a replica of the BNB target. The simulation also uses custom cross-sections for the interactions of pions and nucleons on beryllium (target material) and aluminium (horn material) based on world data [77]. The propagation of particles through the target, horn and decay pipe is simulated using the Geant4 toolkit [89] and a custom model is used for the particle decays that result in the production of neutrinos [77]. The rate of simulated neutrinos propagated to MicroBooNE defines the flux shown in Figure 3.3.
2. **Neutrino generation.** The interactions between neutrinos and argon nuclei are simulated using the GENIE event generator [90]. MicroBooNE uses a configuration [49]<sup>1</sup> for which the CC-QE (Nieves) [47] and CC-MEC (Valencia) [91] models are tuned to neutrino interaction  $CC0\pi$  cross-section data from the T2K experiment [92]. Most relevant to the primary analysis of this thesis, the Berger-Sehgal model [37, 39–42] is used to simulate CC-Res interactions. GENIE simulates the dynamics of the argon nuclei and, using the simulated BNB flux, generates neutrino interactions of each mode with a rate and kinematics derived from the aforementioned models. GENIE also accounts for the transport of the resulting particles through the nucleus including the effects of FSI. In addition to neutrino-argon interactions, GENIE is also used to simulate “dirt” background interactions where a neutrino from the beam interacts with material outside of the TPC.
3. **Propagation of final-state particles.** The final-state particles of the neutrino interactions generated by GENIE are propagated through the detector volume using Geant4 [89]. This stage accounts for the energy-loss of particles (due to ionisation)

---

<sup>1</sup>The MicroBooNE tune is based off the GENIE v3 G18\_10a\_02\_11a configuration

and any secondary interactions or decays of the final-state particles, including the formation of electromagnetic showers.

4. **Detector simulation.** Finally, the ionisation electrons simulated by Geant4 are propagated through a simulation of the detector to produce simulated wire signals. This procedure is described in reference [93] and summarised here.

First, the simulated cloud of ionisation electrons is transformed to a distribution of electrons at the wire planes. This transformation accounts for the electron-drift and includes the effects of recombination, diffusion, attenuation due to impurities, and distortions of the drift field due to SCE. The responses of the wires due to induction or collection of the drift electrons is then simulated. This includes the known responses of the front-end electronics (which have been measured by injecting a calibration signal to each channel). The wire responses are digitised and noise due to electronic components is added. The noise model is data-driven and based on the wire responses with the drift field switched off. This simulation has been validated using cosmic rays from beam-off data [94], and performs well. Any residual differences between the data and simulation from this validation procedure are used to assess the uncertainty on the TPC simulation in Part III of this thesis.

Due to the computational cost, scintillation photons are not individually tracked for each simulated event. Instead, the detector is voxelised <sup>2</sup> and the response of each PMT to an isotropic collection of scintillation photons produced in each voxel is simulated and stored in a *photon visibility map*. This map is used to look-up the corresponding PMT responses to the scintillation photon simulated by Geant4. This simulation has also been validated using cosmic rays from beam-off data and any differences are again used to assess the uncertainty on the optical simulation.

5. **Overlaying cosmic-ray data.** The final stage is to add cosmic rays to the simulated neutrino interactions. This is done by adding the simulated wire and PMT signals to those from unbiased beam-off data (produced by firing the so-called “fake” beam trigger). This process is known as overlaying.

---

<sup>2</sup>A voxel is the 3D analogue of a (2D) pixel.

## Part II

# Reconstruction and identification of neutrino interactions in dense cosmic ray environments





# Chapter 4

## Pandora pattern-recognition in MicroBooNE

### 4.1 Introduction

As described in Chapter 3, MicroBooNE produces three 2D images of the interactions that occur in the detector. These images are rich with high-resolution information which allows the experiment to distinguish between (and measure the properties of) a wide range of neutrino interaction topologies. In many cases, the images can be understood by the human brain with relative ease. However, to fully exploit the data, a systematic and computational pattern-recognition procedure is required. This procedure aims to identify individual particle trajectories, and their relationships, to infer their interactions and decays.

The diverse particle signatures seen by MicroBooNE are together challenging for any single pattern-recognition approach. To solve this problem (amongst other approaches) MicroBooNE uses the Pandora pattern-recognition framework — founded and developed by the Cambridge group. Pandora forms the basis of all of the work in this thesis, both directly through algorithm development and indirectly by use of its output products in downstream analyses. Consequently, this chapter is devoted to Pandora and the direct contributions made by the author.

Whilst described here in the context of MicroBooNE, Pandora also provides a general solution to the LArTPC pattern-recognition problem and is indeed used by other such experiments. The Pandora approach is to break the pattern-recognition problem up into many stages, each with defined aims, inputs and outputs. For each stage, multiple algorithms are designed which each target a specific topological signature, and a given algorithm will update the output of the stage only when that signature is clearly present. Each algorithm is hand-written to add a different piece of domain-knowledge to the system (e.g. particles are often represented by continuous lines of hits), which is something the human brain might look for in the images. When required, groups of these algorithms are executed iteratively to facilitate feedback loops. This iteration allows for multiple algorithms to work together, each making small cautious decisions, until the output gradually emerges.

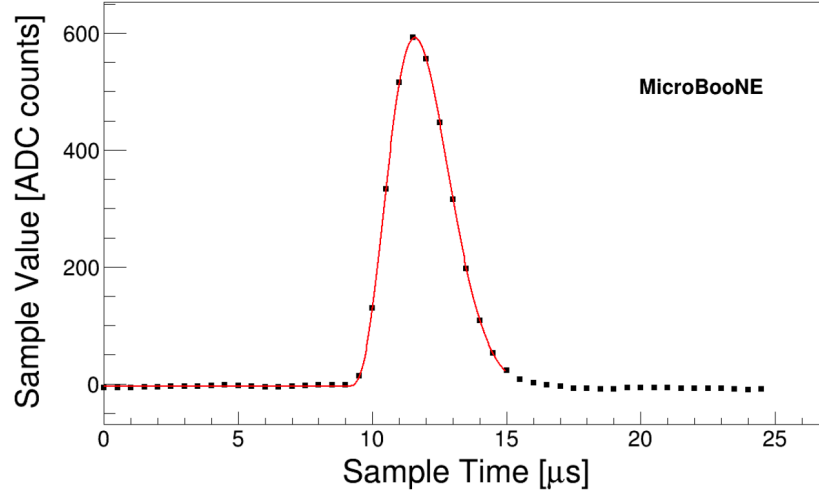
The Pandora content of this chapter describes the state of the project as utilised by the main analysis work of this thesis (presented in Part III) and is largely cited from [95] and [96]. The main contributions of the author of this thesis to Pandora relate to the consolidated approach, described in Section 4.6 and detailed in Chapter 5.

## 4.2 Input and outputs of the Pandora framework

### 4.2.1 Inputs to Pandora

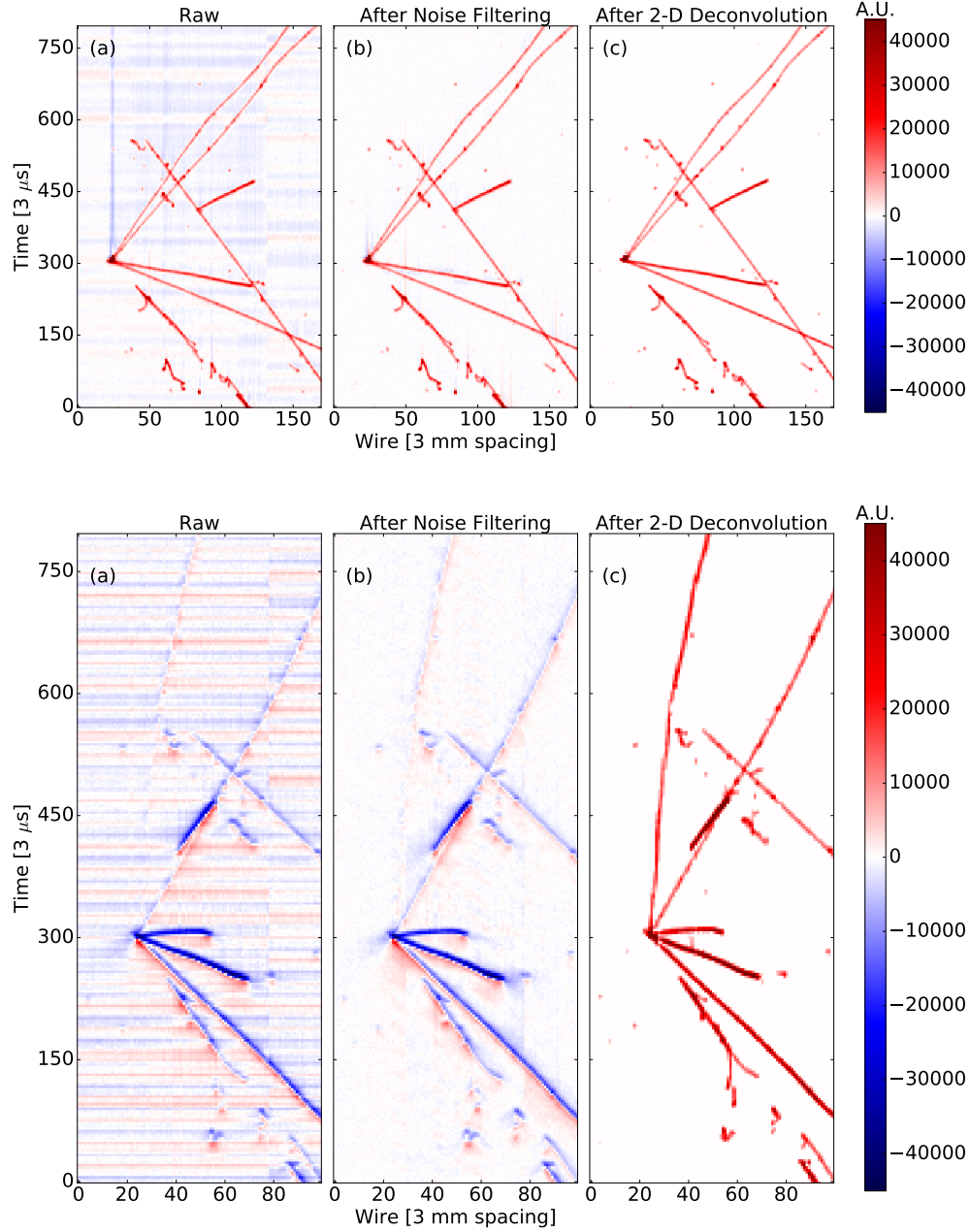
Before pattern recognition can begin, the raw wire signals must be processed to remove noise and detector effects. The measured wire response is the convolution (in the time domain) of the true signal from the ionisation electrons and the known detector response (shown for a collection plane wire in Figure 4.1) with the addition of detector noise at known frequencies due to various components (e.g. power supplies).

The noise is reduced by filtering out the problematic frequencies, and a deconvolution procedure is employed to extract the ionisation signal. In general, the measured response on a given wire has contributions from the signal on the wire itself and also those nearby. Consequently, a two-dimensional deconvolution is used in the time and wire domains. As shown in Figure 4.2, the processed signal is significantly cleaner and unipolar on all planes.



**Figure 4.1:** The digitised waveform due to a calibration signal injected on a single collection plane channel in MicroBooNE. Figure taken from [97].

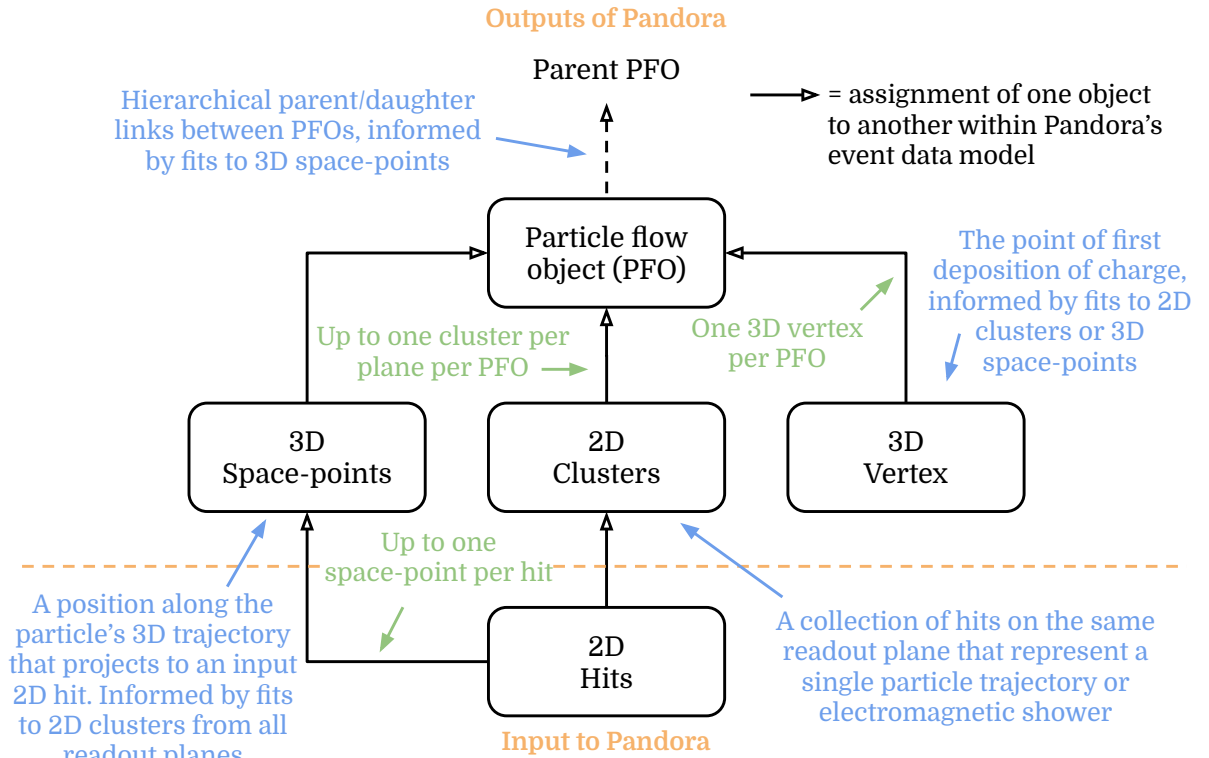
The processed signals on each wire (as a function of time) are evaluated by a peak-finder which identifies “wire-hits”. A hit is produced for each peak and is parametrised by a rectangle in the  $x - z$  plane (as defined in Equation 3.3). The mean and standard deviation of a Gaussian fit to the peak are converted to spatial coordinates via the aforementioned equation, and respectively define the centre and width of the rectangle in  $x$ . The centre and width of the rectangle in  $z$  are respectively defined by the position of the wire and the wire spacing. Finally, each hit is assigned a charge calculated by the integral of the fit. The collection of hits from all wires on all planes form the input to Pandora.



**Figure 4.2:** Event displays from MicroBooNE data of a possible neutrino interaction. From left to right this figure shows waveforms: without processing, after noise filtering and after 2D deconvolution. The interaction is shown on the collection plane (top) and the induction U-plane (bottom). The colour scale indicates the amplitude of the wire responses and is presented in arbitrary units. Figure taken from [97].

### 4.2.2 Output from Pandora

Pandora's primary goal is to identify individual particles from the 2D hit collections and to reconstruct their corresponding 3D trajectories. The Event Data Model (EDM), shown in Figure 4.3, encapsulates all such reconstructed quantities identified by Pandora. The main entity is the Particle-Flow Object (PFO) which represents a reconstructed particle, of which there may be many in a given event. Each PFO is assigned up to three reconstructed clusters (one per readout plane), which are in turn assigned any number of hits. The EDM is designed such that each hit may only be assigned to one cluster, and each cluster to one PFO. Additionally, the reconstructed particles are assigned a 3D vertex which represents the position at which the particle first left an ionisation deposition in the detector. PFOs may also be assigned reconstructed 3D space-points (which are reconstructed 3D positions determined to lie along the particle's trajectory that project to a hit position). Finally, the PFOs are arranged into hierarchies, which indicate parent-daughter relationships in interactions and decays. The EDM forms the output of Pandora and is used as the input to downstream analyses.



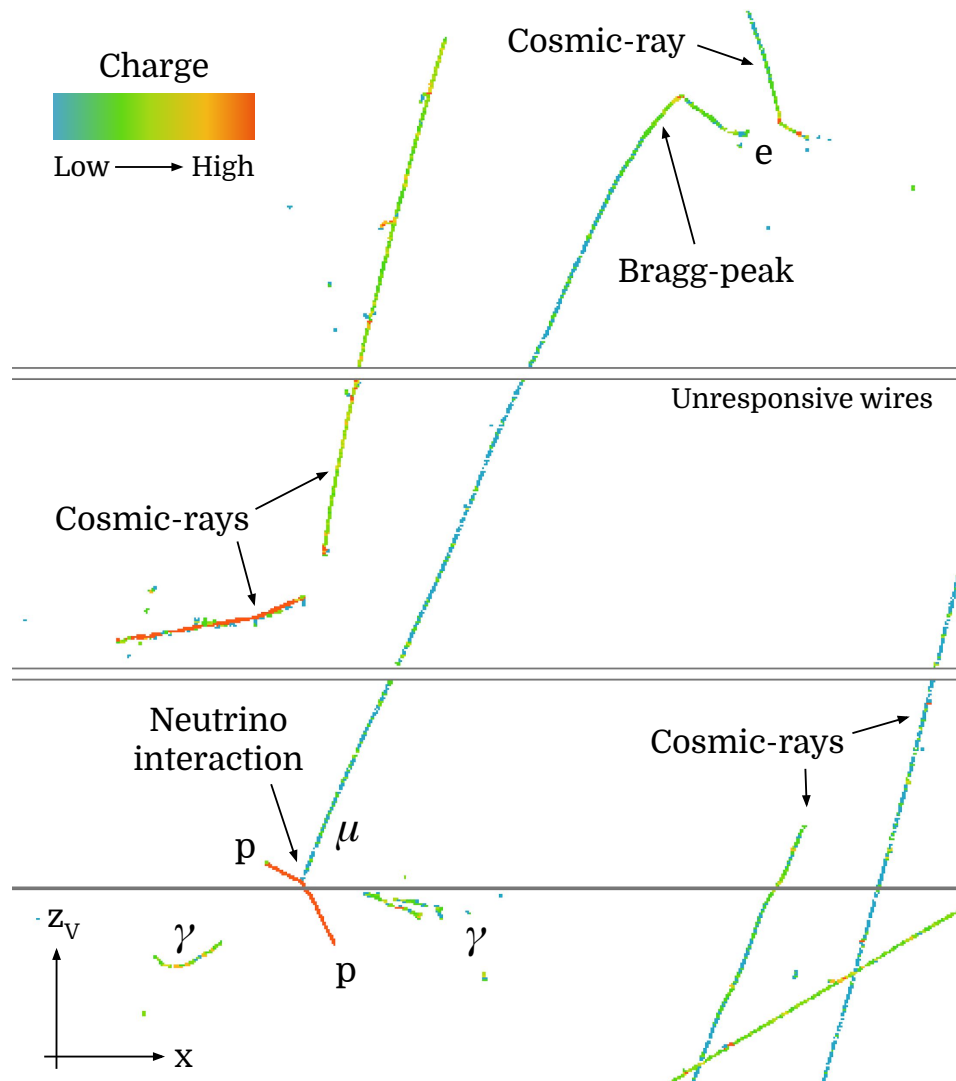
**Figure 4.3:** The Pandora Event Data Model. Black arrows indicate the assignment of one object to another, and the annotations summarise how each object is produced by Pandora's algorithms.

In addition, the EDM represents Pandora’s understanding of the input images during execution. At the beginning of the pattern recognition, the EDM consists only of the input hits which are not yet assigned to any clusters. Pandora’s algorithms act to query and update the EDM, and collectively populate the final output.

### 4.3 Typical particle signatures in MicroBooNE

MicroBooNE is capable of imaging particles with a wide range of complex topological signatures. Track-like particles such as muons, protons and charged pions traverse the detector leaving continuous lines of hits as they gradually lose energy due to ionisation. These particles may scatter (producing deviations in the track trajectory) or undergo secondary interactions (producing further particles). If a track-like particle comes to rest in the detector then a Bragg-peak will be visible in the charges of the hits at the end of the track. The shape of this Bragg-peak allows for discrimination between particle types. Electromagnetic shower-like particles such as electrons and photons initiate a cascade of secondary particles. The electrons and positrons in this cascade leave tracks which together form the shower signature. At BNB energies, a wide range of shower signatures are possible from single electron tracks to a large dense cascade. Discrimination between electron and photon showers is possible through the rate of charge deposition at the start of the shower — a photon shower will begin with depositions from an electron-positron pair, whereas an electron shower begins only with the electron itself. Additionally, a photon can travel from its point of production before initiating a shower whereas an electron will be visible immediately. Figure 4.4 shows a simulated event display with various particle signatures.

Neutrino interactions consist of an interaction vertex position from which the final-state particles emerge, giving rise to a wide range of possible topologies. CRs are typically energetic muons and so are track-like but can also produce small delta ray (energetic ionisation electron) showers. If the origin (CR or neutrino) of a given interaction is known a priori then domain-knowledge of the possible topologies can be used to improve the performance of the pattern recognition. Consequently, Pandora employs two chains of algorithms (summarised in Figure 4.5) which are used by MicroBooNE.

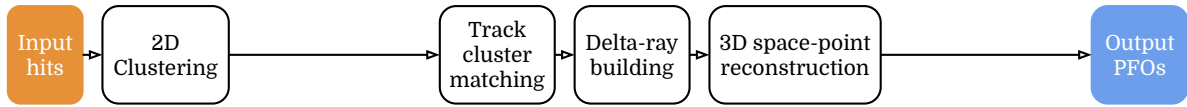


**Figure 4.4:** Reconstructed V-plane hits, coloured by charge, for a simulated neutrino interaction producing the  $\mu + \pi^0 + 2p$  topology overlaid with beam-off cosmic data. The photon showers are from the decay of the neutral pion.

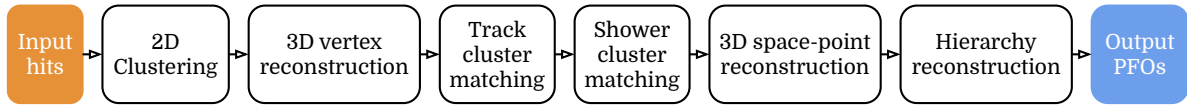
## 4.4 Pandora’s algorithms for cosmic-ray signatures

The cosmic-ray chain is optimised to reconstruct events that contain only cosmogenic activity, and is strongly track-oriented as the majority of the particles of interest are cosmic-ray muons. The neutrino chain is optimised to reconstruct events containing only a neutrino interaction. Under this hypothesis, the interaction vertex is identified early in the algorithm chain to motivate the production of emerging tracks and showers. MicroBooNE events will typically consist of both neutrino and cosmic-ray interactions. Consequently, it is crucial and non-trivial to apply the correct algorithm chain to the correct hits. Section 4.6 and Chapter 5 describe the algorithms developed to handle this problem. The remaining sections in this chapter describe the two algorithm chains in isolation.

Cosmic-ray algorithm chain



Neutrino algorithm chain

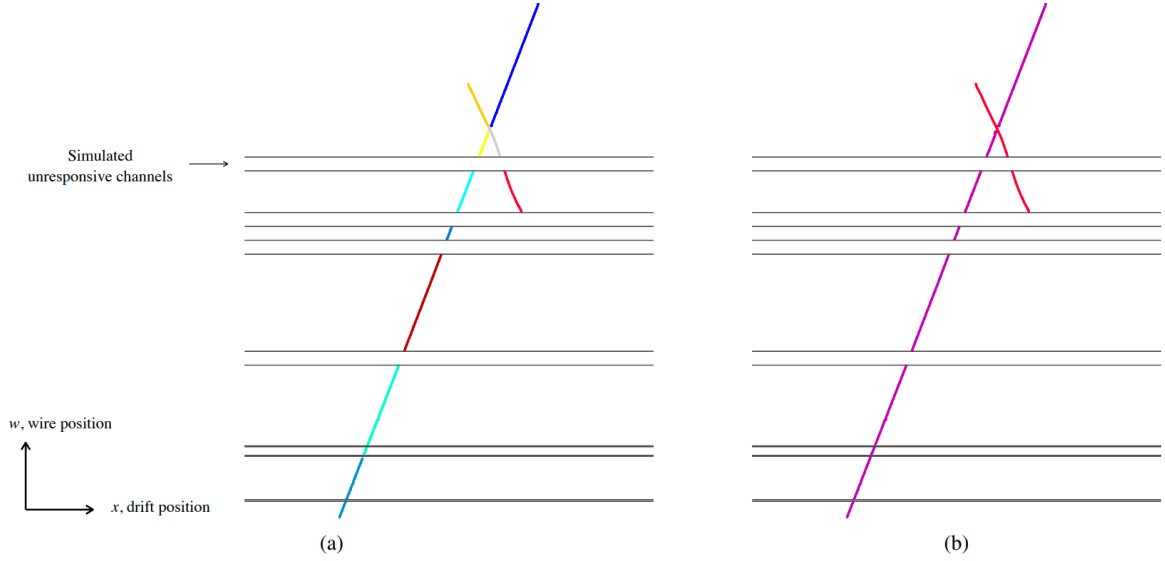


**Figure 4.5:** Flowchart of the steps in the two Pandora algorithm chains used by MicroBooNE. Modified from [96].

### 4.4.1 2D clustering

The cosmic-ray algorithm chain begins with the 2D clustering stage. Working in each view independently, a group of algorithms look to group continuous lines of hits together into clusters. A new cluster is produced at any kink, gap or bifurcation. At this point, the reconstructed clusters are of high purity (each cluster only represents one true particle) but can be of low completeness (the cluster may represent only a fraction of all the true particle’s hits). A collection of cluster refinement algorithms look to merge the existing 2D clusters if specific topologies are identified (such as crossing particles) without compromising purity. Knowledge of unresponsive regions allows clusters to be merged over gaps in the input images.



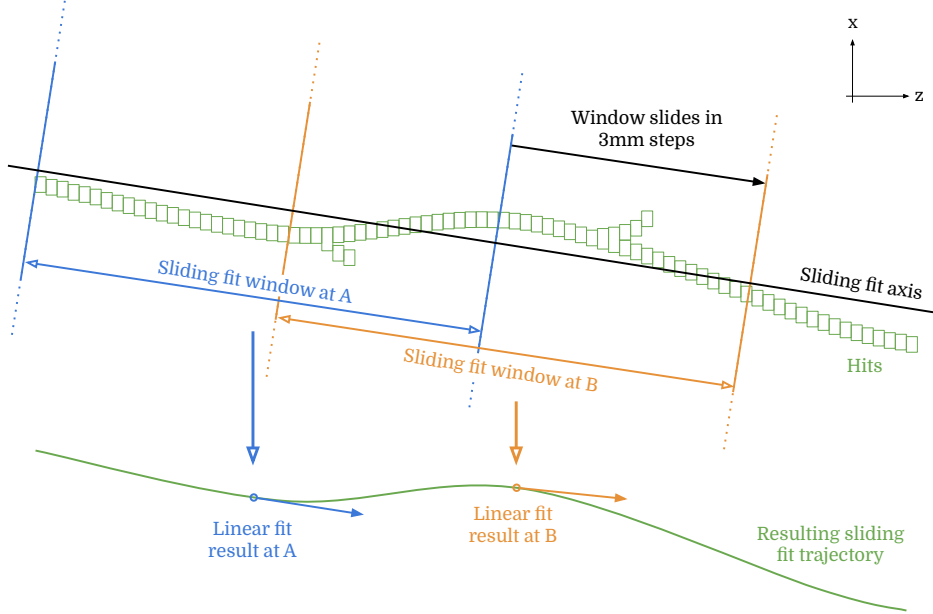


**Figure 4.6:** Event displays showing the output of the 2D clustering stages for two crossing particles. The hits are colour coded by cluster. Figure (a) shows the initial clusters which break at unresponsive regions and the crossing point of the particles. Figure (b) shows the refined clusters which have been extended over the unresponsive regions and merged at the crossing point. Figure taken from [96].

#### 4.4.2 Track cluster matching

After 2D clusters have been formed in the three input images separately, the next step is to identify the clusters in different views that represent projections of the same particle, as is illustrated in Figure 4.8. A sliding linear fit is applied to each cluster so it can be represented as a line  $z_P(x)$  (using the coordinates defined in Equation 3.3). As shown in Figure 4.7, to perform the sliding linear fit, an axis is first defined by a linear fit to the positions of the hits in the cluster. Beginning at one of the two most extremal hits on the axis, a linear fit again is applied to the subset of hits within a rectangular window of width 12 cm aligned along the axis direction. The window is slid across the full length of the axis in steps of 3 mm (set by the wire spacing) until the final hit is reached. The line  $z_P(x)$  is defined by a linear interpolation between the fit results at consecutive window positions.

As shown in Figure 4.8, the extent to which a given pair of clusters match (from different views  $A$  and  $B$ ) can be measured by the overlap of the span of the cluster fits in the  $x$ -direction. One can also determine the expected cluster fit positions in the remaining view,  $C$ , at each common  $x$ -position,  $z'_C(x) = \mathcal{P}(z_A(x), z_B(x))$ , where  $\mathcal{P}$  performs the projection. The consistency between a given triplet of clusters (one



**Figure 4.7:** An illustration of the sliding linear fit procedure. A linear fit to the hits is performed within a window that slides along an axis (black line). The sliding window is displayed at two example positions, A and B, and the result of the linear fit is shown below. The resulting sliding linear fit (green line) is the interpolation between all sliding fit positions.

from each view) can be measured by a Figure of Merit (FoM) calculated as the average squared distance between the projected and actual cluster positions  $(z'_C - z_C)^2$ , over each common value of  $x$  for all combinations:  $UV \rightarrow W$ ,  $WU \rightarrow V$ , and  $VW \rightarrow U$ .

At the beginning of the track cluster matching stage, the  $x$ -overlap and consistency FoM are calculated for all possible triplets of clusters and stored in a rank-3 “tensor” whose indices label the clusters in each view. If every 2D cluster has exactly one consistent match in the other views, then the tensor will be strongly “diagonal”. Off-diagonal entries indicate ambiguities in the matching which typically arise from imperfect 2D clustering. A list of algorithm tools interrogate the tensor and identify cluster matches that are strongly consistent with a track. In these cases, a new PFO is produced and assigned the relevant clusters. If a trajectory projects to an unresponsive detector region, a PFO may also be formed from just two clusters. For ambiguous cases, the algorithms can also split or merge the 2D clusters to see if the overall matching improves. Every such algorithm is designed to identify and resolve a specific topological ambiguity, examples of which are shown in Figure 4.8. The algorithms are ordered and executed consecutively. However, if any algorithm modifies a cluster then the tensor is re-evaluated and execution begins again from the start. The process stops when every algorithm in the list can no longer take an action. Pandora is designed to be as performant as possible

and care has been taken to minimise the time complexity of each algorithm. For example, values that do not change between iterations are cached, and appropriate data structures such as k-d trees are used when searching for the nearest neighbour of a hit. In total, Pandora typically takes  $\sim 3-4$  s to fully reconstruct <sup>1</sup> each MicroBooNE event, which can contain both neutrino and cosmic-ray induced particles.

### 4.4.3 Delta ray building

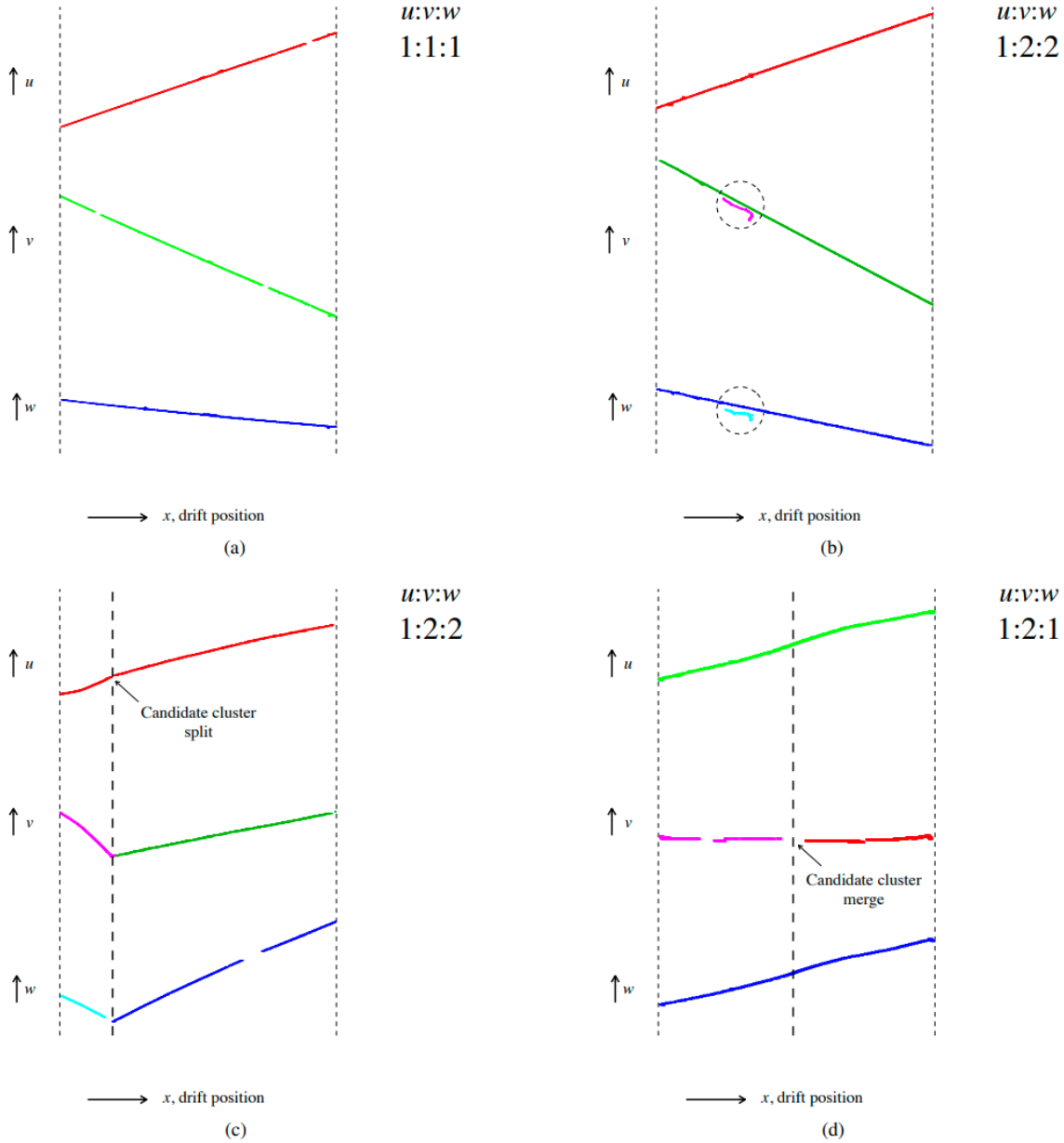
After the track cluster matching is complete, any remaining 2D clusters that weren't matched are reconstructed as delta rays. These clusters are dissolved and their hits are re-clustered to improve their quality. Where possible these clusters are matched between views, and only  $x$ -overlap is considered. Any newly produced delta-ray particles in this step are assigned as a daughter of the closest primary (non delta-ray) particle.

### 4.4.4 3D Space-point and vertex reconstruction

The final step of the cosmic-ray algorithm chain is to reconstruct the 3D space-points which map out the trajectories of the PFOs. Where possible, a space-point is created for each 2D hit. By sampling sliding linear fits to the clusters in a given PFO, a 2D position,  $z_P(x_{hit})$ , can be derived for each cluster at the  $x$ -coordinate of the hit in question. These positions represent the desired projections of the 3D space-point into the 2D views. The 3D position which projects most closely to these desired 2D positions is used to produce the space-point, provided the projection distances are within a quality threshold. Lastly, a sliding linear fit is applied to the space-points of each PFO, and a 3D vertex is assigned at the end of the particle closest to the top of the detector.

---

<sup>1</sup>All of the steps in Pandora's full algorithm flow are shown later in Figure 4.11.



**Figure 4.8:** Event displays showing cluster matching ambiguities. Figure (a) shows a clear match between three clusters. The remaining displays show different topological ambiguities which are resolved during the track cluster matching stage. Figure (b) shows a long track-like particle with a delta-ray cluster (circled). A match is made between the long clusters to resolve the ambiguity. Figure (c) shows a kinked trajectory which isn't visible in the  $U$ -view, but identifiable in  $V$  and  $W$ . The  $U$  cluster is split to resolve the ambiguity. Figure (d) shows a particle with missing hits in the  $V$ -view, but is continuous in  $U$  and  $W$ . The  $V$  clusters are merged to resolve the ambiguity. The triple ratios (e.g.  $1 : 2 : 2$ ) indicate the number of overlapping clusters in each view ( $U : V : W$ ). Figure taken from [96].

## 4.5 Pandora’s algorithms for neutrino interactions

### 4.5.1 Neutrino interaction vertex reconstruction

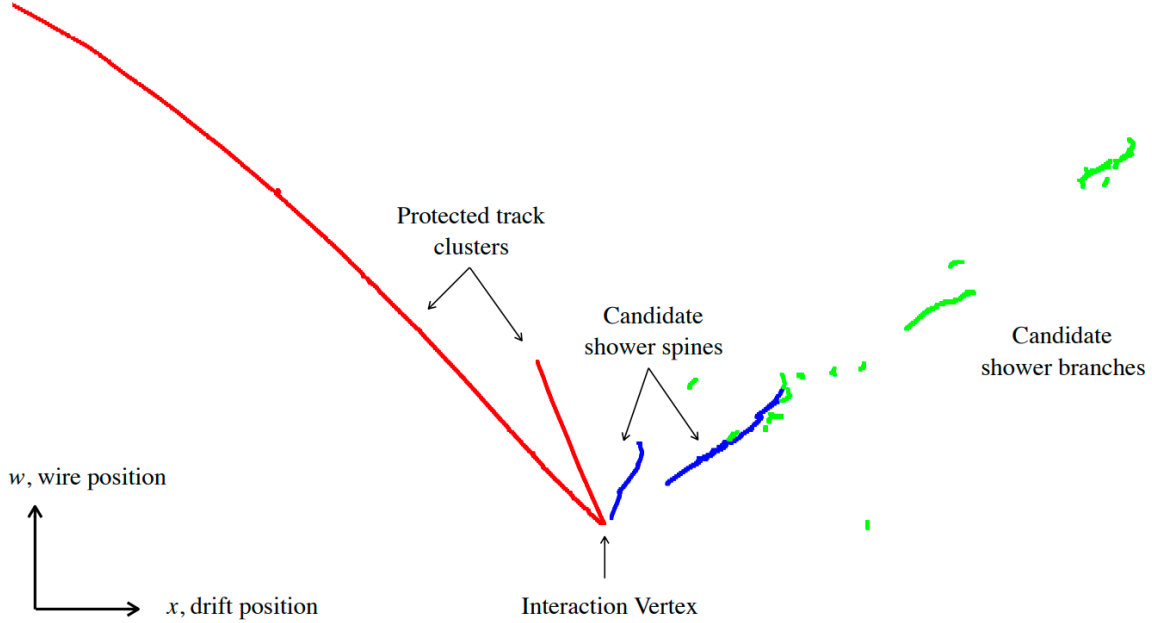
As shown in Figure 4.5, the neutrino and cosmic-ray algorithm chains share a number of common stages. One key difference in the neutrino chain is the early identification of the 3D neutrino interaction vertex, which is used by later algorithms to improve performance. By applying sliding linear fits to the 2D clusters, 3D vertex candidates are produced at positions that project to the cluster end points and crossing points. Each position is scored using a number of metrics that measure the extent to which the surrounding hits point back toward the candidate vertex, and act to suppress candidates that sit along straight tracks or within showers.

The product of these scores is used to rank the candidates, and those with the highest ranks seed groups of nearby vertices with a maximum radius of 10 cm. The individual scores of the seed vertices for each group are fed to a trained Support Vector Machine (SVM) which produces a likelihood that the input candidate vertex represents the true neutrino interaction position. The group with the highest SVM likelihood is selected, and the candidate vertices in that group again interrogated by a second SVM. Finally, the most likely candidate in the group is selected as the final neutrino interaction vertex. At this point, any clusters which straddle the vertex are split and the reconstruction moves on to the track cluster matching stage. The early identification of the neutrino interaction vertex is an important aspect of the neutrino algorithm chain as the vertex position is used to inform the logic of many downstream algorithms.

### 4.5.2 Shower cluster matching

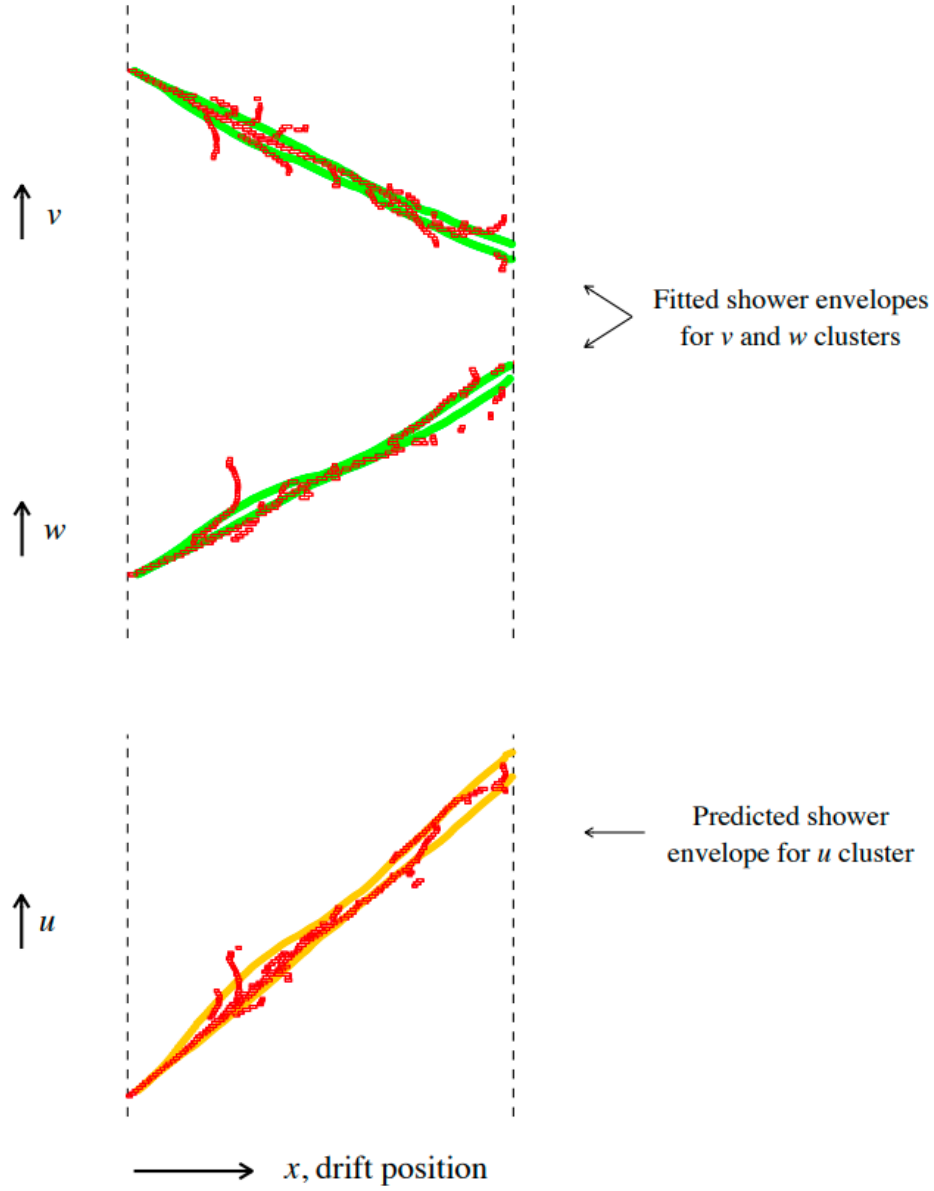
Another important step in the neutrino algorithm chain is the special treatment of showers. After the 2D clustering and track cluster matching is complete, the PFOs are classified as track-like or shower-like by considering the spread of hits in the transverse direction along the particle’s trajectory. Any shower-like particles are dissolved to allow for dedicated 2D shower clustering and the track-like particles emerging from the vertex are protected from being re-clustered. Clusters that represent shower spines are identified, these must be sufficiently long and point back to the vertex. Clusters that represent shower branches are also identified, these are typically shorter and begin along or nearby a shower spine. An example event display is shown in Figure 4.9. An iterative process

grows 2D shower clusters (which are seeded by the spines) by adding nearby branches where appropriate.



**Figure 4.9:** Event display showing the result of the cluster classification. Track-like clusters are shown in red, and shower-like cluster spines and branches are shown in blue and green respectively. Figure taken from [96].

The 2D shower clusters are then matched between views in much the same way as the track-like case. The main difference is that sliding linear fits to the edges of the envelope of the shower clusters (as shown in Figure 4.10) are used to check the consistency between views. Again, 2D shower clusters are iteratively refined when ambiguities are identified and PFOs are produced for clear matches. Finally, after the cluster matching is complete, the shower PFOs are refined to improve their overall completeness. Any remaining clusters which couldn't be matched between views are merged into a nearby shower PFO if sufficiently close.



**Figure 4.10:** Event display showing the envelope (green) fitted to a shower-like cluster (red) in the  $V$  and  $W$  views. The projection of this envelope (orange) is shown in the  $U$  view. Figure taken from [96].

### 4.5.3 Interaction hierarchy reconstruction

After the clustering of the PFOs is finalised, the 3D space-points are produced for each PFO, as described in Section 4.4.4. Next, the particles are arranged into a hierarchy. A neutrino PFO is produced and assigned the neutrino interaction vertex but no clusters or space-points are assigned to it. A 3D sliding linear fit then is applied to the space-points of each track-like and shower-like PFO and those which emerge from the neutrino vertex (based on their proximity and directional information) are identified as the final states of the interaction and added as daughters of the neutrino PFO. By a similar method, secondary particles are also identified and assigned as daughters of the final-state particles. This process continues (adding granddaughters, etc.) until no more hierarchical associations can be made. Any remaining PFOs are assigned as neutrino daughters by default.

### 4.5.4 Track vs. Shower characterisation

The last step in the neutrino algorithm chain is to determine a final track-shower classification of the final-state PFOs. Up to this point in the chain, simple track-shower classifications have been utilised to inform the pattern recognition and to increase performance — for example, in the special treatment of shower-like clusters in Section 4.5.2. However, the result of the characterisation described in this section is a track likelihood score that is used at the analysis-level (downstream of Pandora) and has no impact on the pattern recognition itself.

An SVM is utilised to perform the classification, which is fed with topological features of the 3D space-points belonging to each PFO. The features measure the linearity, the separation of the PFO from the neutrino vertex, the spread of the space-points in the secondary and tertiary directions with respect to the primary axis of the particle, and the angle between the space-points closest to and furthest from the neutrino interaction vertex. Finally, a particle type ( $\nu_\mu$  or  $\nu_e$ ) is assigned to the neutrino PFO based on whether the final-state PFO with most hits is more likely to be track-like or shower-like respectively.



## 4.6 Neutrino interactions in dense cosmic-ray environments

As MicroBooNE is a surface detector with a long ( $\sim$ ms) readout time, a typical event will contain tens of cosmic-ray interactions. These cosmic rays form a major background to any analysis of neutrinos in MicroBooNE. Pandora’s cosmic-ray and neutrino algorithm chains are designed to be performant for cosmic-ray and neutrino interactions respectively, but inherent to their success is the assumption that the input images contain either cosmic rays or a neutrino interaction, not a combination of both. In particular, the neutrino algorithm chain relies on the early identification of the neutrino vertex and assumes that all particles in the input images are either final-state particles or their daughters. The existence of CRs breaks this assumption and negatively affects the pattern-recognition performance of the neutrino algorithm chain.

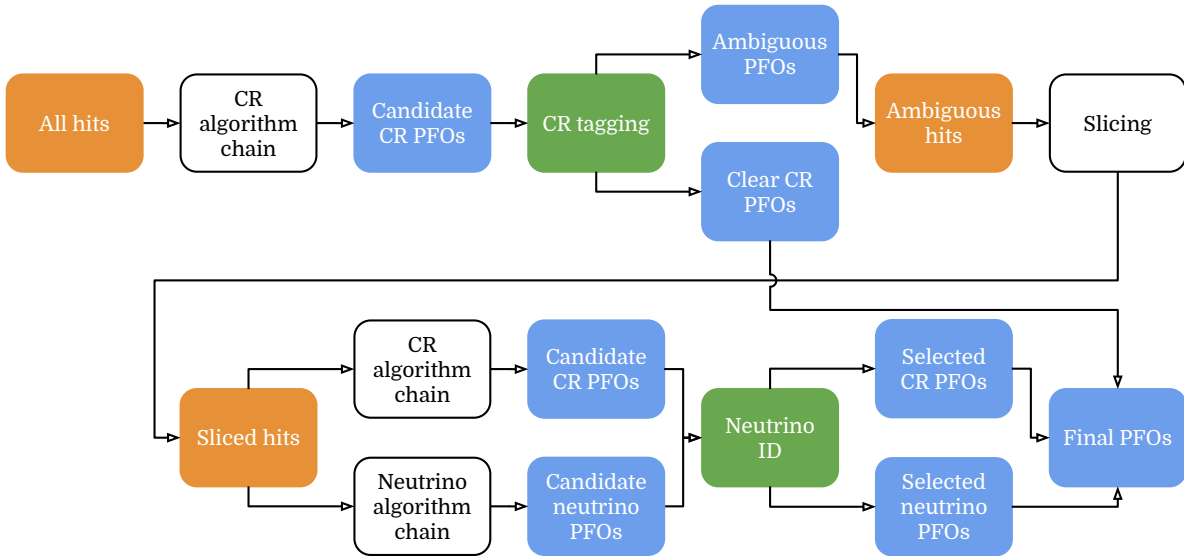
This section describes how the two algorithm chains are consolidated in order to be able to reconstruct both neutrinos and cosmic-ray interactions in the same event. This “consolidated approach” forms a major part of the work in this thesis, both through the development of algorithms to make the requisite key decisions (described in Chapter 5), and through the technical implementation of the approach in the wider analysis framework used by MicroBooNE. The consolidated approach is adopted as the standard used by MicroBooNE and forms the basis of all modern Pandora-based analyses — including the main topic of this thesis.

### 4.6.1 The consolidated approach

The consolidated approach is summarised in Figure 4.11. First, the input hits (which in general can contain both neutrino and cosmic-ray interactions) are fed into the cosmic-ray algorithm chain. The resulting PFOs are then analysed by a cosmic-ray tagging step (described in Chapter 5) that identifies which reconstructed particles are unambiguously cosmogenic. The CR-tagged particles are added to the final output and all remaining PFOs are dissolved so their hits are available for further analysis.

The remaining (CR-removed) hits form the input of the “slicing” stage. During slicing, a fast reconstruction pass is used to build temporary track-like and shower-like PFOs whose space-points are fitted using sliding linear fits and conical fits respectively. Using proximity and directional information from these fits, the PFOs are grouped into topologically distinct regions, called slices, which each represent an individual particle interaction hierarchy (neutrino or CR). Any remaining unclustered hits are added to the closest slice. The PFOs are again dissolved and the hits in each slice are reconstructed separately by both the cosmic-ray and neutrino algorithm chains.

Finally, the neutrino identification step (described in Chapter 5) considers the outcome of the two algorithm chains for each slice, and chooses which slice (if any) contains the neutrino interaction. The PFOs from the appropriate algorithm chain are selected and, together with the unambiguous cosmic-ray PFOs, form the final consolidated output.



**Figure 4.11:** The consolidated algorithm flow. Hit collections are shown in orange, Pandora algorithm chains in black, PFO collections in blue and analysis decisions in green.

## 4.7 Performance of Pandora in MicroBooNE

### 4.7.1 Performance metrics

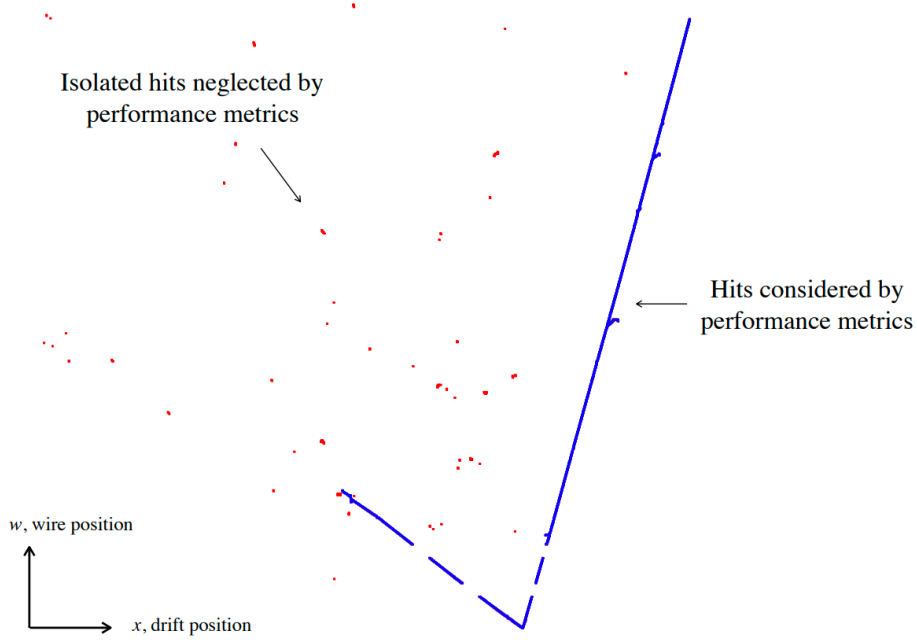
The overall performance of Pandora is assessed using a set of standard metrics which are described in this section. These metrics are calculated using simulated neutrino interactions which may be overlaid with beam-off cosmic data as described in Section 3.4. Firstly, each hit in the event is matched to the simulated neutrino final-state particle which contributes most to the charge of the hit. This charge can either be deposited by the final-state particle directly or by a downstream particle produced in a secondary interaction. Hits with no such matched particles originate from cosmic rays or detector noise. For each true final-state particle, the matched hits are assessed to determine if the particle could plausibly be reconstructed. To do this, the matched hits are filtered and must meet the following criteria:

- The matched neutrino final-state particle (and downstream particles) must in total contribute at least 90% of the charge of the hit. This removes cases in which it is ambiguous as to which neutrino final-state particle the hit belongs.
- The hit must not be matched to a simulated particle that is (or is downstream of) a neutron. This removes sparse and isolated, collections of hits shown in Figure 4.12.
- The hit must not be matched to a simulated track-like final-state particle (muon, charged pion, proton, etc.), and be produced downstream of a far-travelling photon. Again, this restriction removes isolated hit collections.

A simulated final-state particle is deemed reconstructable and a “target” of the pattern recognition if it has at least 5 matched hits on two or more readout planes and a total of at least 15 hits on all planes.

The hits assigned to the reconstructed neutrino final-state PFOs (labelled,  $r$ ) are compared to those belonging to the simulated target particles (labelled,  $t$ ). The purity,  $P_{rt}$ , and completeness,  $C_{rt}$ , of a given pair of reconstructed and simulated particles is:

$$\begin{aligned} P_{rt} &= \frac{Q_{rt}}{N'_r}, \\ C_{rt} &= \frac{Q_{rt}}{N_t}. \end{aligned} \tag{4.1}$$



**Figure 4.12:** Event display showing the hits that are considered (blue), and neglected (red) in the performance metrics. Figure taken from [96].

Here,  $N'_r$  and  $N_t$  are the total number of (filtered) hits assigned to the reconstructed and simulated particles respectively, and  $Q_{rt}$  is the number of hits shared by both.

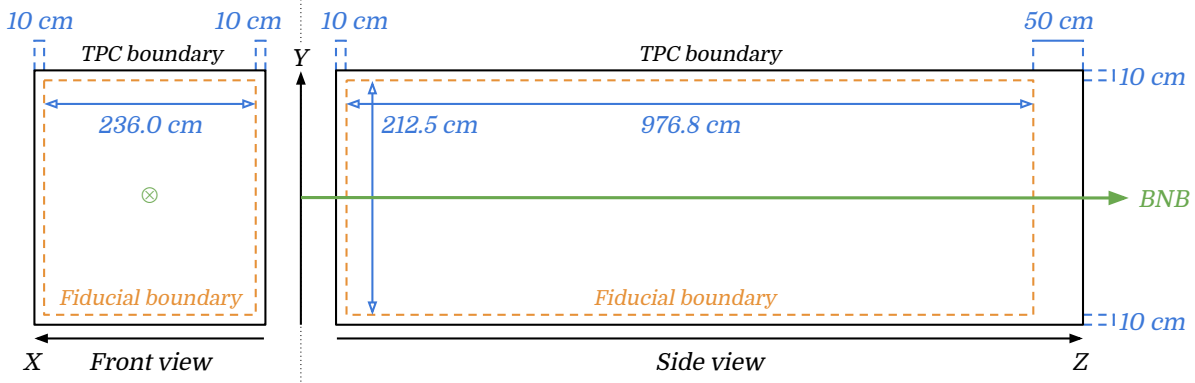
Next, the PFOs are matched to the simulated target particles. A list of all pairs of particles  $(r, t)$  in the event are considered and the pair with the most shared hits is identified as a good match if,  $Q_{rt} > 5$ ,  $P_{rt} > 0.5$  and  $C_{rt} > 0.1$ . If a good match is identified, any pairings involving the matched particles are removed from the list. This process is repeated until all particles are matched or no more good matches can be made between the particles that remain in the list. Finally, any remaining PFOs are matched to the target particle with which most hits are shared provided the match is above the threshold number of shared hits, purity and completeness. After the matching procedure is complete, each target particle is matched to zero, one, or many neutrino final-state PFOs. The number of matched PFOs to each target particle is a useful measure of the performance of the pattern recognition. An event is deemed “correctly reconstructed” if every target particle has exactly one matched PFO.

### 4.7.2 Performance on CC-Res $1\mu + 1\pi^+ + Np$ events

As an example that is most relevant for the work presented in Part III of this thesis, this section shows the performance of Pandora for simulated  $\nu_\mu$  CC-Res interactions with the  $1\mu + 1\pi^+ + Np$  topology. Table 4.1 shows the number of PFOs matched to each target particle for events with a simulated neutrino interaction vertex in the Fiducial Volume (FV) shown in Figure 4.13, and *without* cosmic rays. To produce this table, only the neutrino algorithm chain was executed as the other steps of the consolidated algorithm flow are only required to handle neutrino interactions in the presence of cosmic rays.

Target	Number of matched PFOs				Correct [%]
	0	1	2	3+	
<b>CC-Res <math>1\mu + 1\pi^+</math></b>					$77.3 \pm 0.4$
$\mu$	$5 \pm 0.2$	$92.4 \pm 0.2$	$2.6 \pm 0.1$	$0.08 \pm 0.03$	
$\pi^+$	$9.5 \pm 0.3$	$81 \pm 0.4$	$8.7 \pm 0.3$	$0.75 \pm 0.08$	
<b>CC-Res <math>1\mu + 1\pi^+ + 1p</math></b>					$70.7 \pm 0.4$
$\mu$	$4.5 \pm 0.2$	$93.8 \pm 0.2$	$1.7 \pm 0.1$	$0.01 \pm 0.01$	
$\pi^+$	$9.5 \pm 0.2$	$82.3 \pm 0.3$	$7.7 \pm 0.2$	$0.49 \pm 0.06$	
$p_1$	$10.2 \pm 0.2$	$86.4 \pm 0.3$	$3.2 \pm 0.1$	$0.12 \pm 0.03$	
<b>CC-Res <math>1\mu + 1\pi^+ + 2p</math></b>					$62 \pm 1$
$\mu$	$4.3 \pm 0.5$	$93.8 \pm 0.6$	$1.8 \pm 0.3$	$0.1 \pm 0.1$	
$\pi^+$	$12 \pm 0.8$	$80 \pm 1$	$7.5 \pm 0.7$	$0.5 \pm 0.2$	
$p_1$	$10 \pm 0.8$	$87.8 \pm 0.8$	$2.1 \pm 0.4$	$0.1 \pm 0.1$	
$p_2$	$14.3 \pm 0.9$	$83.6 \pm 1$	$2.1 \pm 0.4$	$0 \pm 0.1$	
<b>CC-Res <math>1\mu + 1\pi^+ + 3p</math></b>					$52 \pm 4$
$\mu$	$6 \pm 2$	$91 \pm 2$	$3 \pm 1$	$0 \pm 0.6$	
$\pi^+$	$15 \pm 3$	$81 \pm 3$	$3 \pm 2$	$1 \pm 1$	
$p_1$	$9 \pm 2$	$89 \pm 3$	$3 \pm 1$	$0 \pm 0.6$	
$p_2$	$22 \pm 3$	$77 \pm 3$	$1 \pm 1$	$0 \pm 0.6$	
$p_3$	$18 \pm 3$	$81 \pm 3$	$1 \pm 1$	$0 \pm 0.6$	

**Table 4.1:** The number of matched PFOs to  $\mu$ ,  $\pi^+$  and protons from simulated  $\nu_\mu$  CC-Res interactions with the  $1\mu + 1\pi^+ + Np$  topology. The events do not contain cosmic rays. The entries in the table give the percentage of target neutrino final-state particles with 0, 1, 2 or 3+ matched PFOs. For the topologies with  $N > 1$ , the target protons ( $p_1, p_2, p_3$ ) are ordered by the number of associated hits. The final column gives the percentage of events with each topology for which every target particle has exactly one matched PFO.



**Figure 4.13:** The Fiducial Volume definition, shown by the dashed orange line. The black rectangle represents the full TPC volume, and the green line shows the direction of the BNB.

For this topology,  $\sim 90\%$  of target muons and  $\sim 80\%$  of target charged pions have exactly one matched PFO. The cases in which zero PFOs are matched primarily occur for low energy target particles and for events with a small muon-pion opening angle. In the latter case the hits from the muon and pion are typically merged into a single PFO. The  $\sim 8\%$  of events for which a target  $\pi^+$  is matched to two PFOs primarily occur when the pion undergoes a secondary interaction or decay and Pandora (correctly) reconstructs the pion and its secondary interaction products as separate PFOs but fails to identify a hierarchical parent-daughter link. For events with  $N > 0$ , the proton with the most hits has exactly one matched PFO in  $\sim 87\%$  of cases. As with muons and pions, the rate at which target protons are missed by the pattern recognition is largest at low energies and when small angles are made to other particles in the event. Consequently, the correct event fraction decreases as the proton multiplicity increases.

Table 4.2 shows the performance of the full consolidated algorithm flow (including CR-tagging, slicing and neutrino identification) for simulated  $\nu_\mu$  CC-Res  $1\mu + 1\pi^+ + Np$  events overlaid with cosmic rays from beam-off data. The first row shows the fraction of events for which no target particles have any matched PFOs that have been identified as a neutrino final state. This failure mode impacts  $\sim 20\%$  of events with this topology and occurs when the event is rejected by the neutrino identification step or a slice containing a cosmic ray is misidentified as a neutrino interaction. It is important to note that the neutrino identification step acts as a pre-selection for analyses downstream of Pandora and correctly rejects  $\sim 95\%$  of slices containing a cosmic ray, which corresponds to reduction in the rate of events without a neutrino interaction by a factor of seven [98].

The second row of Table 4.2 shows the fraction of events not rejected by the slice identification step for which every target particle has exactly one matched PFO. The performance is largely consistent with that shown in Table 4.1, however the correct event fraction is degraded by  $\sim 5\%$ . This degradation primarily occurs due to cosmic rays that intersect the neutrino interaction. In this case a target particle may be merged with the cosmic ray and removed by the CR-tagging step. Alternately, the cosmic ray may be included in the neutrino slice which can cause the reconstructed neutrino interaction vertex to be misplaced.

The final row of Table 4.2 shows the fraction of events for which every target particle has exactly one matched PFO. This represents the bottom-line performance of Pandora’s consolidated algorithm flow for this topology.

$1\mu + 1\pi^+ + \dots$	Number of target protons			
	$0p$	$1p$	$2p$	$3p$
<b>Rejected</b>	$20.4 \pm 0.4$	$16.8 \pm 0.3$	$19 \pm 1$	$19 \pm 3$
<b>Not rejected and correct</b>	$75.2 \pm 0.4$	$66.1 \pm 0.4$	$56 \pm 1$	$44 \pm 4$
<b>Correct</b>	$59.9 \pm 0.5$	$55.0 \pm 0.4$	$46 \pm 1$	$35 \pm 4$

**Table 4.2:** The performance of Pandora’s consolidated approach for simulated  $\nu_\mu$  CC-Res interactions with the  $1\mu + 1\pi^+ + Np$  topology, for  $N = 0, 1, 2, 3$ . If a proton multiplicity has  $T$  events of which  $R$  are rejected (have no PFOs matched to any of the target particles) and  $C$  are correct (have exactly one PFO matched to every target particle), the entries in the table give the percentage of events which are: rejected ( $R/T$ ), not rejected and correct ( $C/(T - R)$ ), and correct ( $C/T$ ).

## 4.8 High-level reconstruction

### 4.8.1 Track fitting

After Pandora’s consolidated output has been produced, the 3D particle trajectories are fitted so they can be readily used in downstream analyses. The track-score, described in Section 4.5.4, is typically used to determine whether a track-fit or shower-fit is appropriate. However this decision is ultimately made on an analysis-by-analysis basis. The main analysis work in this thesis pertains to track-only topologies and so only the track-fit is described here. For each reconstructed particle, a sliding linear fit is applied to the 3D space-points produced by Pandora. This defines the geometrical properties of the track, such as its endpoints and its direction. Each space-point is then projected to the closest point along the fit to produce a smooth collection of points along the track trajectory. These trajectory points are assigned a charge from the hit that was used to produce the corresponding space-point. Additionally, the trajectory points are assigned a range as calculated by the path-integrated distance from the start of the fit. With this information, the rate of energy lost along the track can be calculated allowing for the identification of different particle types.

### 4.8.2 Calibrations

The reconstructed tracks analysed in Part III of this thesis are produced by Pandora and are calibrated using a data-driven approach. Firstly, the distortion of MicroBooNE’s drift field due to the Space-Charge Effect has been measured using anode and cathode piercing cosmic-ray tracks in beam-off data [99], and with a laser calibration system [100]. The trajectory points of the tracks produced by Pandora are displaced to correct for the measured spatial distortions. Additionally, anode-cathode crossing cosmic rays are used to correct for spatial non-uniformities in the charge of each trajectory point due to unresponsive wire regions, diffusion and impurities in the LAr [85]. Finally, the rate of loss of charge  $dQ/dx$  along a track is converted to an energy loss  $dE/dx$  using the “modified-box model” which has been tuned to ArgoNeuT data [101] and has been validated using stopping muons and protons in MicroBooNE data [85].



## Chapter 5

# Identifying cosmic-ray activity in MicroBooNE

This chapter describes the CR-tagging and neutrino identification steps of Pandora’s consolidated algorithm flow, shown in Figure 4.11. The CR-tagging algorithm was developed solely by the author of this thesis and is the primary focus of this chapter. The neutrino identification step was developed in collaboration. The details of this step are given in Section 5.4, and the contributions of the author are made explicit.

### 5.1 Classification of reconstructed particles

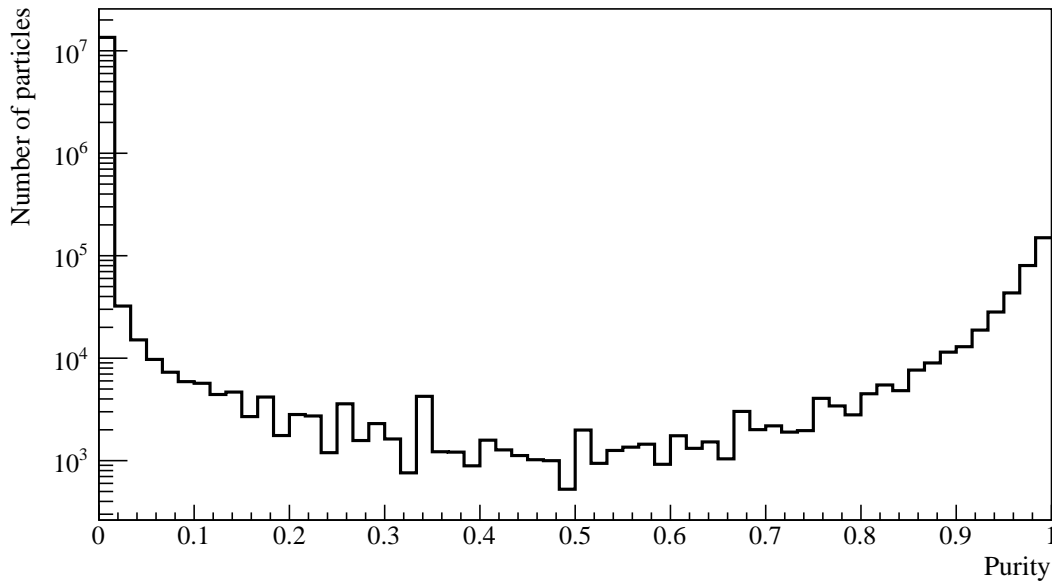
As described in Section 4.6, the inputs to the CR-tagging algorithm are reconstructed particles (PFOs) from Pandora’s cosmic-ray algorithm chain. The task of the CR-tagging algorithm is to determine which of these PFOs are likely produced by a cosmic ray. The hits from these CR-tagged PFOs will not be considered by the later stages in Pandora’s pattern recognition in which neutrino interactions are reconstructed and identified.

The PFOs that are considered by the CR-tagging algorithm represent the *primary particles* identified by the cosmic-ray algorithm chain. However, any hits from descendant PFOs (daughter, granddaughter, etc.) are included in the metrics defined in this section as these PFOs will be tagged by association to their parent. It is also important to note that the cosmic-ray algorithm chain commonly groups multiple neutrino final-state particles together into a single primary PFO or hierarchy.

### 5.1.1 Neutrino purity and significance

In order to measure the performance of the CR-tagging algorithm, it is necessary to determine if a given PFO is CR-induced, neutrino-induced, or some mix of the two. This distinction is made using the neutrino purity,  $P_r$ , of each PFO (labelled  $r$ ). The purity is the fraction of hits in the PFO that are primarily due to a neutrino-induced simulated particle.

The distribution of neutrino purity is shown in Figure 5.1 for events containing simulated neutrino interactions overlaid with cosmic rays from beam-off data, as described in Section 3.4. In this chapter, all studies are restricted to events in which the simulated neutrino interacts within the FV shown in Figure 4.13.



**Figure 5.1:** The neutrino purity of all particles as reconstructed by Pandora’s cosmic algorithm chain. The distribution is bimodal, with particles representing true cosmic rays at a purity of zero, and true neutrino induced particles at a purity of one.

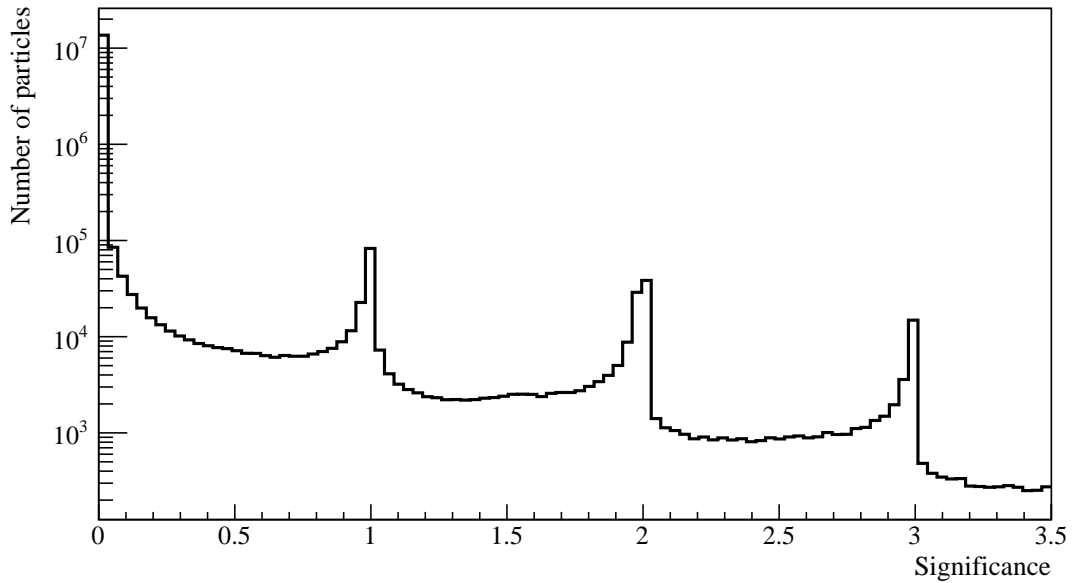
It is also necessary to determine if each PFO represents a significant fraction of a true neutrino final-state particle (which are identified following the procedure described in Section 4.7.1). As given in Equation 4.1, the completeness of a given pairing between a PFO,  $r$ , and a reconstructable true neutrino final-state particle,  $t$ , is defined as the fraction of hits from the true particle that are shared with the PFO. The completeness is defined in the range  $(0, 1)$  and indicates the extent to which the PFO represents the true particle. The neutrino significance,  $S_r$ , of a PFO is further defined as the sum of

the completenesses over all pairings in the event:

$$S_r = \sum_{t=1}^M \frac{Q_{rt}}{N_t}. \quad (5.1)$$

Here,  $M$  is the number of reconstructable final-state particles,  $N_t$  is the number of hits associated with each such particle (labelled  $t$ ), and  $Q_{rt}$  is the number of hits shared between the PFO and the true particle.

This metric, shown in Figure 5.2, is defined in the range  $(0, M)$  and measures the number of true neutrino final-state particles that are represented by the PFO. It is important to restate that the PFOs that form the input to the CR-tagging algorithm are produced using the cosmic-ray algorithm chain. Consequently, the pattern-recognition performance indicated by Figure 5.2 is not comparable to the results presented in Section 4.7 using the full consolidated approach.



**Figure 5.2:** The neutrino significance of all particles as reconstructed by Pandora’s cosmic algorithm chain. The distribution is multimodal, where each peak contains reconstructed particles that represent zero, one, two, etc. true neutrino final-state particles.

### 5.1.2 Classification scheme

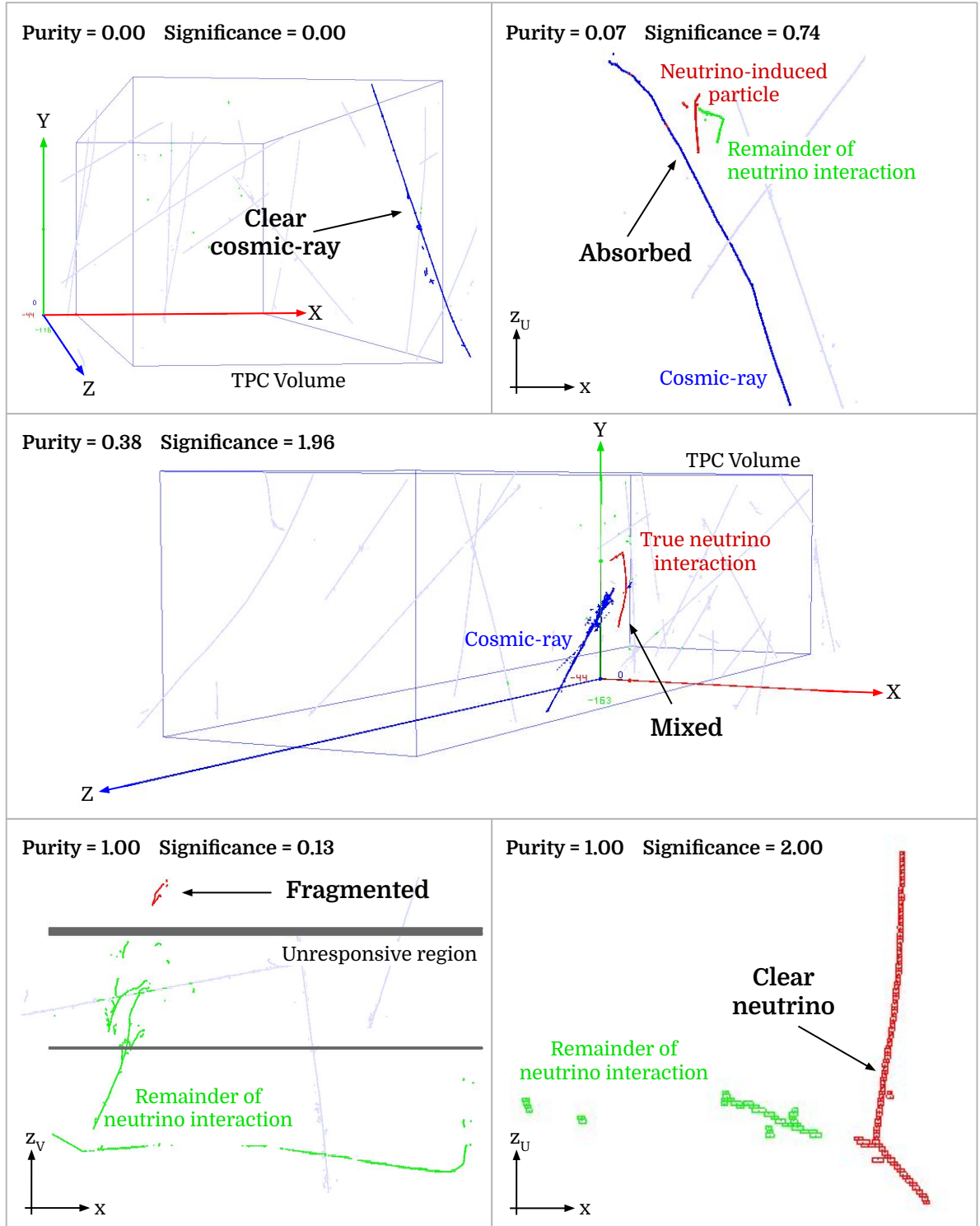
To understand the impact of the CR-tagging algorithm, the PFOs are classified according to Table 5.1. A “clear cosmic-ray ” PFO has no greater than 10% of its hits from a true neutrino interaction and does not represent a significant portion (greater than 50%) of a neutrino final-state particle. Conversely, at least 90% of the hits of a “clear neutrino” PFO must be truly neutrino induced, and the PFO must represent a significant fraction of a neutrino final-state particle.

The “absorbed” category corresponds to PFOs that primarily represent a CR, but have been merged with a significant portion of the neutrino interaction. The “mixed” category includes PFOs that contain hits from both cosmic and neutrino origins. Finally, the “fragmented” category includes PFOs that are neutrino induced, but only represent a fraction of a neutrino induced final-state particle. Examples of each of these categories are shown in Figure 5.3.

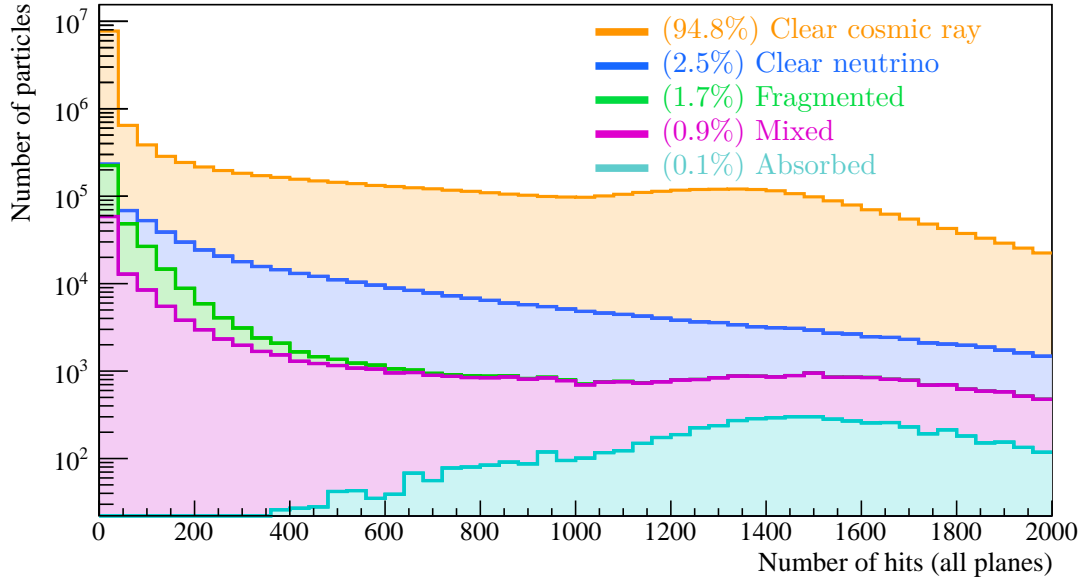
	<b>Insignificant</b> $S_r < 0.5$	<b>Significant</b> $S_r > 0.5$
<b>Impure</b> $P_r < 0.1$	Clear cosmic ray	Absorbed
<b>Mixed</b> $0.1 \leq P_r \leq 0.9$	Mixed	
<b>Pure</b> $P_r > 0.9$	Fragmented	Clear neutrino

**Table 5.1:** The classifications of reconstructed particles by their neutrino purity,  $P_r$ , and significance,  $S_r$ .

The fraction of PFOs with each classification is shown in Figure 5.4 as a function of the number of hits in the PFO over all readout planes. In total,  $\sim 95\%$  of PFOs fall into the clear cosmic ray category. Of these,  $\sim 45\%$  have fewer than 15 hits and typically correspond to small isolated deposits throughout the TPC. The distribution of PFOs among the remaining categories depends strongly on the number of hits. Non-clear CR PFOs with fewer than  $\sim 50$  hits tend to be fragmented (typically corresponding to small shower fragments), or are a mix of CR and neutrino activity. Above  $\sim 100$  hits, clear neutrinos are the dominant (non-cosmic) classification. PFOs in the absorbed category primarily occur above  $\sim 1000$  hits corresponding to cosmic rays that are sufficiently long to cross the TPC.



**Figure 5.3:** Event displays showing examples of the different classifications of reconstructed particles that form the input to the CR-tagging algorithm. In each case, the hits in the event are colour coded to indicate if they are neutrino-induced (red/green) or cosmic-ray induced (blue/grey). Hits that belong to the reconstructed particle in question are red/blue, and other hits in the event are green/grey, and are included for context.



**Figure 5.4:** The number of particles of each classification as a function of the total number of hits (over all readout planes). Only events with a simulated neutrino interaction in the FV are included. The integrated fraction of particles with each classification is given in the legend.

## 5.2 Methods of identifying cosmic rays

This section describes the methods used by the CR-tagging algorithm to differentiate between neutrino-induced and cosmic-ray -induced activity. The specific manner in which each method is exploited to tag cosmic rays is described in Section 5.3.

### 5.2.1 Timing information

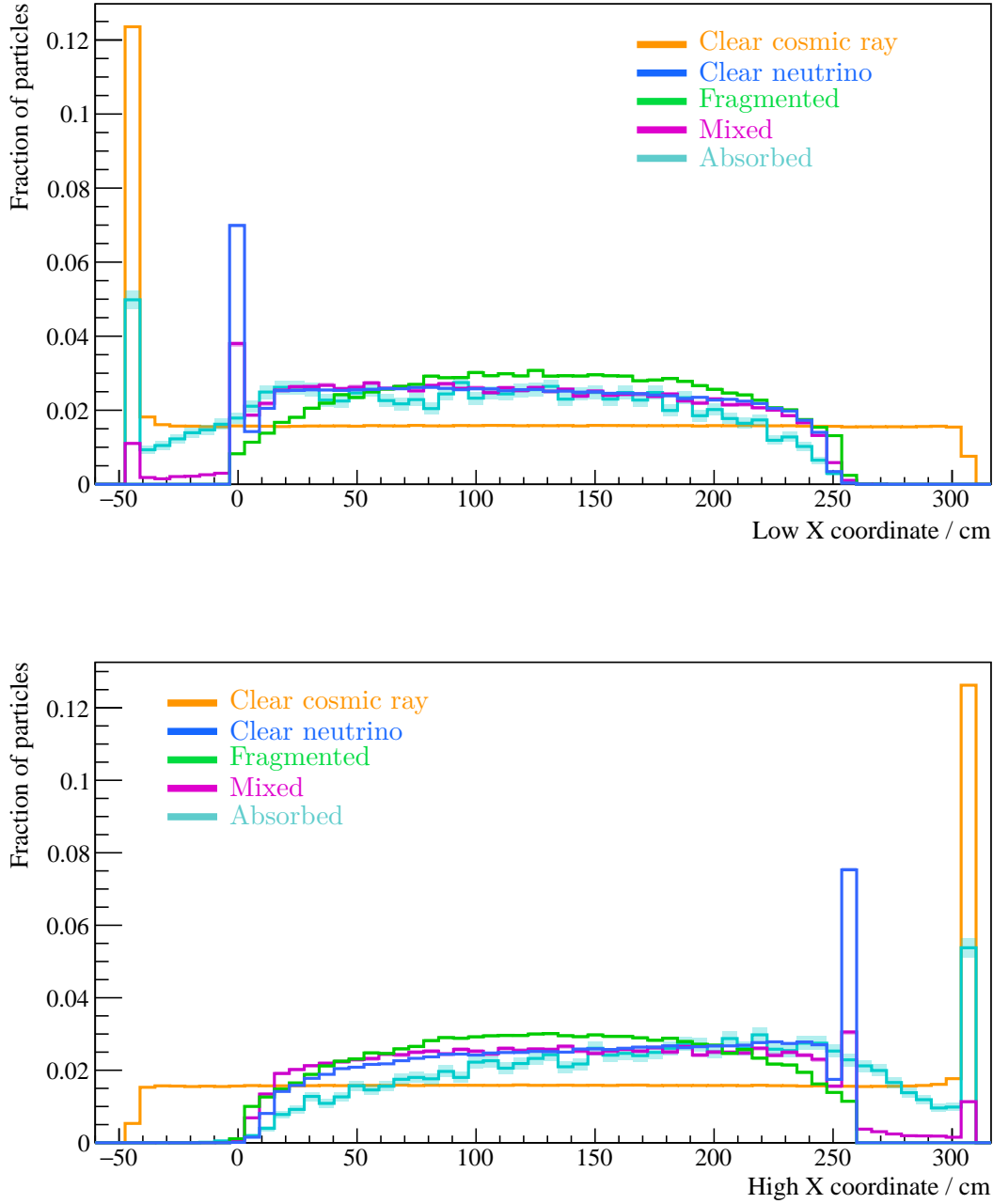
As discussed in Section 3.2.2, care must be taken in order to use the reconstructed drift coordinate of a PFO. In general, the reconstructed coordinate,  $X$ , of a point along a particle's trajectory is displaced from the true point of deposition,  $X_0$ , by:

$$\Delta X = X - X_0 = v_d(t_0 - T). \quad (5.2)$$

Here,  $t_0$  is the time at which the particle deposited energy in the TPC,  $T$  is the trigger time and  $v_d$  is the drift velocity. Neutrino induced particles deposit energy at the trigger time and hence  $\Delta X = 0$ . However, the  $t_0$  of a cosmic ray is not known *a priori*. The result is an ambiguity in the reconstructed  $X$  coordinate of cosmic rays.

This ambiguity can be exploited for CR-tagging. As described in Section 3.2.6, TPC data is read out for the interval  $T - 1.6 \text{ ms} \rightarrow T + 3.2 \text{ ms}$  which (at the design field strength) corresponds to the spatial region  $-L_X < X < 2L_X$ , where  $L_X = 2.56 \text{ m}$  is the length of the TPC in the drift direction. Consequently, and as a result of Equation 5.2, the charge deposited by a cosmic ray can be reconstructed at a position beyond the physical bounds of the TPC ( $0 < X < L_X$ ), whereas the charge deposited by a neutrino interaction can not. In practice, MicroBooNE operates at a lower field strength and TPC data is truncated to the region  $-50 \text{ cm} < X < L_X + 50 \text{ cm}$  to reduce the volume of data stored.

A sliding fit is performed to the space-points associated with each PFO that is considered by the CR-tagging algorithm. The distribution of the  $X$  coordinate of the two fitted end points of each PFO is shown in Figure 5.5. In particular, the end point closest to the anode (low  $X$ ) is shown separately to the end point closest to the cathode (high  $X$ ) however both plots contain the same reconstructed particles from the same events. The PFOs with a high neutrino purity are bound to the dimensions of the TPC, whereas the cosmic-dominant PFOs extend to the edges of the (truncated) readout window, and the mixed classification peaks at both boundaries.



**Figure 5.5:** The  $X$  coordinate of the fitted end-points of the reconstructed particles that form the input to the CR-tagging algorithm. The low- $X$  end is shown in the top plot and the high- $X$  end is shown in the bottom plot. The peaks at the detector boundaries (0 cm &  $L_X = 256$  cm) are due to neutrino-induced activity that escapes the detector at the anode (low- $X$ ) or cathode (high- $X$ ). The peaks at  $-50$  cm and  $L_X + 50$  cm are due to cosmic-induced activity at the edges of the truncated TPC readout region. Statistical-only uncertainties are shown as semitransparent bands.

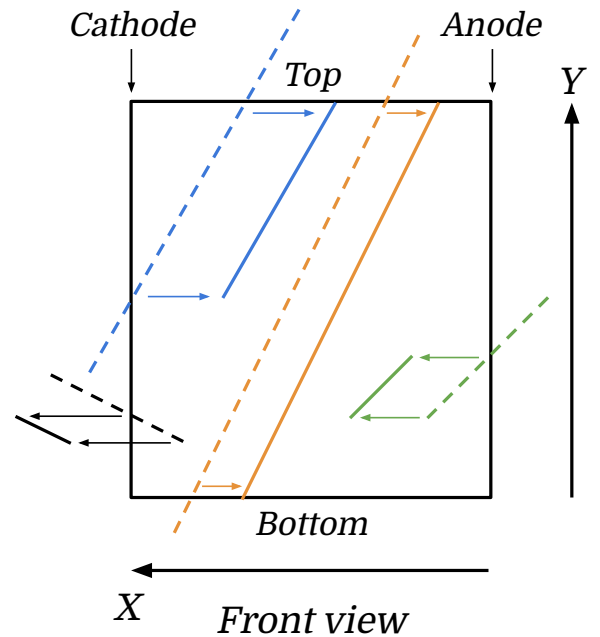


### 5.2.2 Through-going particles

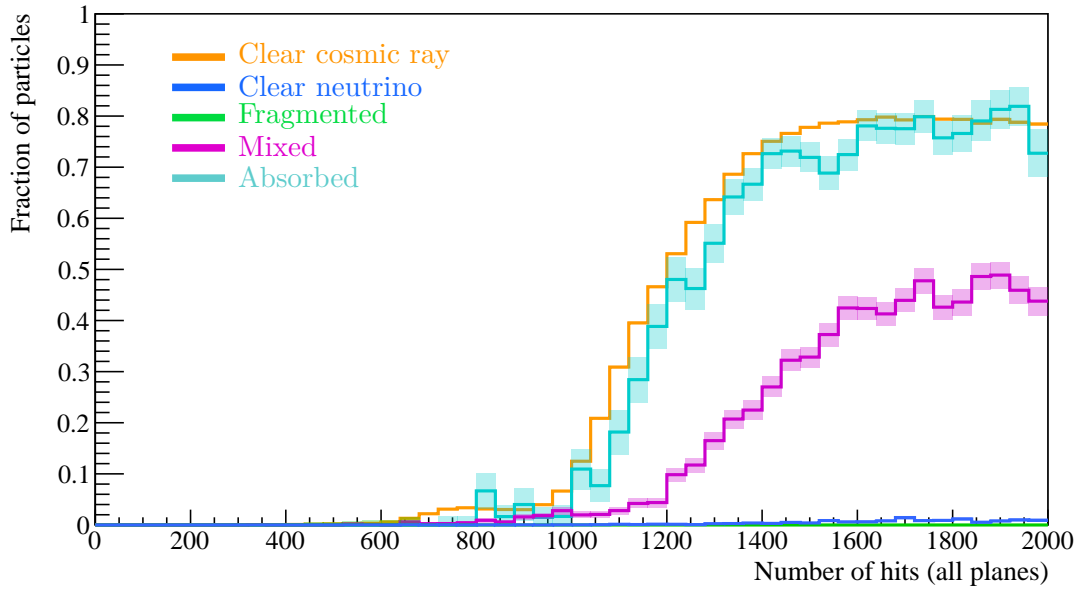
Neutrino induced particles are produced within the TPC and are either contained or escape through one of the detector faces. Clear neutrino PFOs are largely isotropic in the plane perpendicular to the beam direction (and so escape via the top/bottom and anode/cathode faces at an approximately equal rate) but on average point along the beam direction (and so escape via the downstream face more frequently than the upstream face). Conversely, cosmic rays are produced in the atmosphere and enter the TPC through one of the detector faces (primarily the top or side faces), and either stop in the TPC or are through-going.

As illustrated by Figure 5.6, the ambiguity in the  $X$  coordinate means it is non-trivial to exploit these differences. The black/green lines represent stopping cosmic rays that enter through the cathode/anode faces, whereas the blue/orange lines are through-going. As described in Section 5.2.1, the black trajectory is unambiguously cosmogenic as it appears outside of the physical bounds of the TPC. The green and blue trajectories are geometrically equivalent to contained and escaping neutrino-induced particle trajectories respectively. However, the orange trajectory is unambiguously through-going.

**Figure 5.6:** A cartoon showing the differences between true (dashed) and reconstructed (solid) cosmic-ray trajectories as a result of the ambiguity in the drift coordinate. Those trajectories that are shifted left/right were deposited after/before the trigger time.



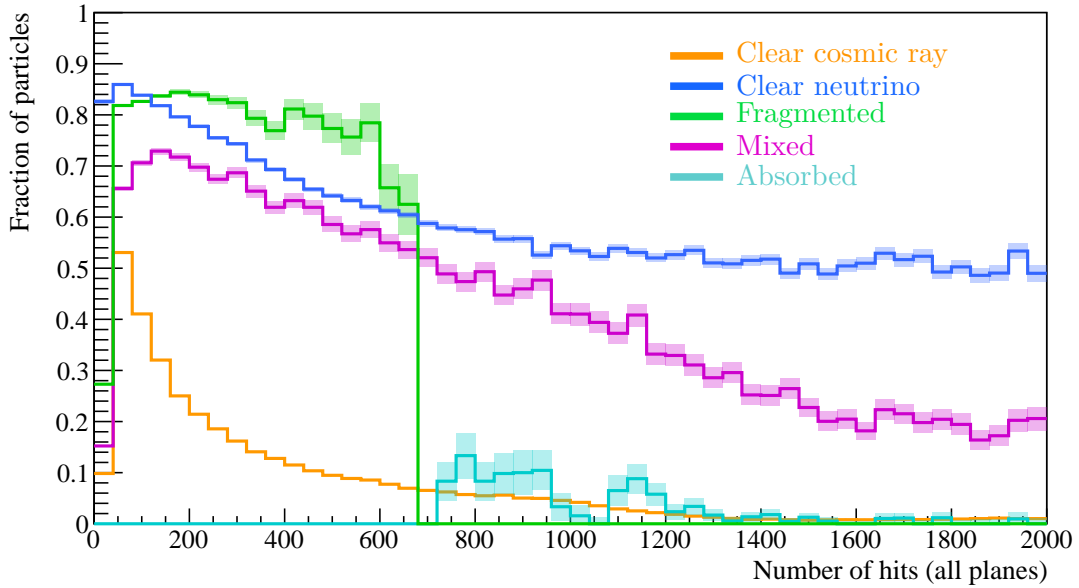
In principle, these ambiguities can be resolved by matching TPC and optical activity such that the  $t_0$  of the cosmic rays can be estimated. Additionally, the direction of stopping particles can be determined by identifying a Bragg peak which can disambiguate between the trajectory illustrated in blue in Figure 5.6 and an escaping neutrino-induced particle. However, neither calorimetric nor optical information was available for use in Pandora at the time of the development of the CR-tagging algorithm. Consequently, only PFOs that cross the TPC between the top, bottom, upstream or downstream faces can be identified as cosmic rays. Of the possible combinations, those that cross from top-to-bottom are most abundant. Figure 5.7 shows the fraction of PFOs with reconstructed endpoints within 20 cm of the top and the bottom face, as a function of the number of hits in the PFO. To determine the trajectory of the PFOs, a sliding linear fit is used following the procedure described in Section 4.8.1 and at least 15 space-points are required to ensure the quality of the fit. The threshold of 20 cm allows for the spatial distortions introduced by SCE, and the resolution of the 3D reconstruction.



**Figure 5.7:** The fraction of reconstructed particles with each classification with a trajectory from the top to the bottom of the TPC as a function of the total number of hits on all planes. In total 0.1% of clear neutrino PFOs travel from top-to-bottom. Statistical-only uncertainties are shown as semitransparent bands.

### 5.2.3 Contained particles

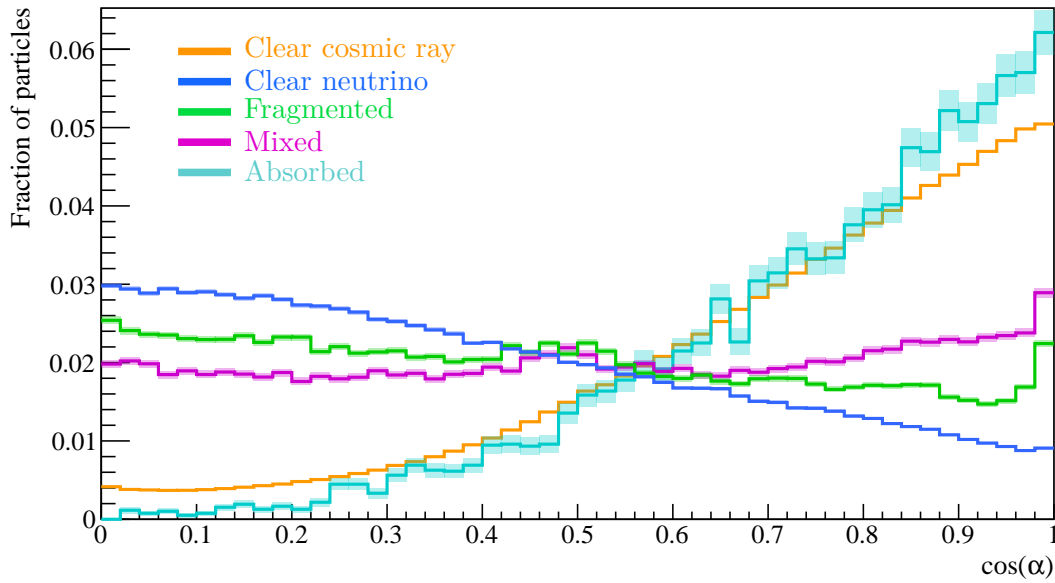
For PFOs that are reconstructed within the physical TPC volume in the drift direction, Figure 5.8 shows the fraction whose end points are also at least 20 cm from the top, bottom, upstream and downstream faces as a function of the number of hits. Many of the clear cosmic-ray PFOs that are contained are those that enter through the anode or cathode faces (as illustrated by the green trajectory in Figure 5.6) or are due to isolated hit deposits (shown in Figure 4.12). The containment fraction of cosmic rays decreases with the number of hits, as the particles become more likely to reach the bottom, upstream or downstream faces. Although the containment of a PFO can not unambiguously differentiate between neutrino and CR-induced particles, it is used in conjunction with other features by the CR-tagging algorithm, as described in Section 5.3.



**Figure 5.8:** The fraction of PFOs reconstructed within the physical TPC volume in the drift direction that are contained within 20 cm of the TPC borders in the  $Y - Z$  plane, as a function of the number of hits. The fragmented (green) and absorbed (cyan) classifications are truncated above and below 700 hits respectively due to limited statistics, as shown in Figure 5.4. Statistical-only uncertainties are shown as semitransparent bands.

### 5.2.4 Angular information

The angular distribution of PFOs is another feature that distinguishes cosmic-ray and neutrino induced particles. Figure 5.9 shows the angle,  $\alpha$ , made between the reconstructed direction of a PFO and the vertical axis. This angle is always defined between  $0 < \alpha < \pi/2$ , as it is not known if a PFO is upwards or downwards going. As expected, clear cosmic-ray PFOs tend to point along the vertical direction, and the reverse is true for clear neutrino PFOs.



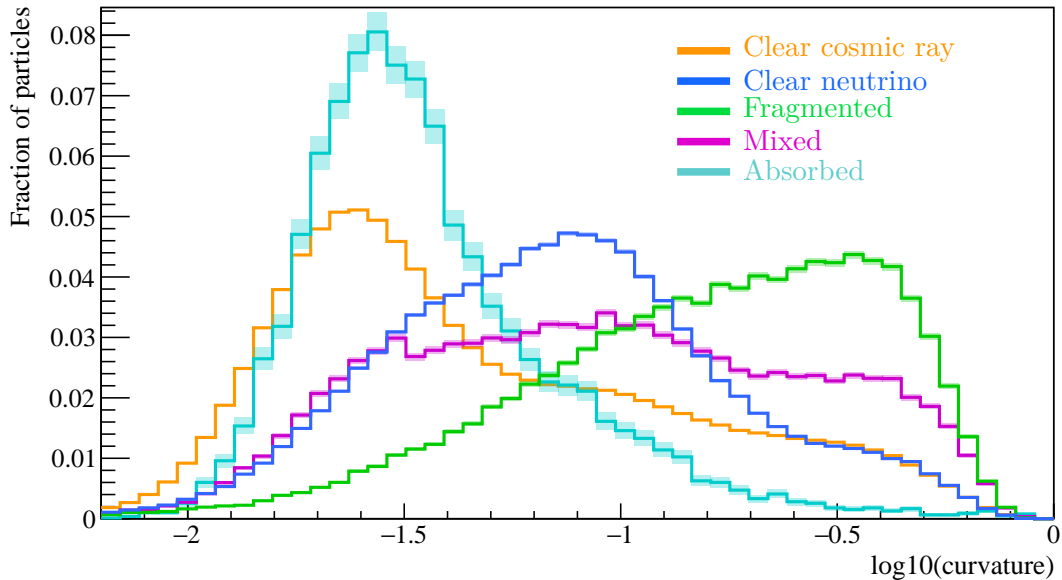
**Figure 5.9:** The angle made by reconstructed particles to the vertical axis, broken down by their classification. Statistical-only uncertainties are shown as semitransparent bands.

### 5.2.5 Topological information

The CR-tagging algorithm also exploits the *curvature*,  $C$ , of a PFO, defined by:

$$C = \frac{1}{N} \sum_{i=1}^N \left| \underline{n}_i - \frac{1}{N} \sum_{j=1}^N \underline{n}_j \right|. \quad (5.3)$$

Here,  $N$  is the number of positions at which the sliding linear fit to the reconstructed space-points in the PFO is calculated, and  $\underline{n}_i$  is a unit vector pointing in the direction of the fit at each such position. This feature measures the average (local) deviation of the direction of the sliding linear fit from the (global) mean direction. In the special case where  $C = 0$ , all of the space-points in a PFO lie along a straight line. Figure 5.10 shows the distribution of this feature for each PFO classification. Fragmented PFOs, which typically represent small electromagnetic shower segments, tend to have a larger curvature. The distribution of clear cosmic-ray PFOs at high curvature are typically due to isolated hit deposits, whereas the PFOs in the peak of the clear cosmic rays are typically straight muon tracks. This feature also has sensitivity to cases in which clear neutrino PFOs represent multiple neutrino final-state particles ( $S_r > 1$ ).



**Figure 5.10:** The curvature of reconstructed particles, broken down by their classification. Statistical-only uncertainties are shown as semitransparent bands.

### 5.2.6 Relationships between particles

The final method discussed in this section is the geometrical relationship between PFOs. In Figure 5.3, the fragmented and clear-neutrino examples show cases in which the PFO in question represents part of a neutrino interaction, and the other PFOs in the event point back to a common interaction vertex. The CR-tagging algorithm identifies that these PFOs are related to others in the event using proximity and directional information. As described in Section 5.3, groups of related PFOs are considered together to avoid tagging parts of a neutrino interaction (for example an escaping muon from a CC interaction), that might appear cosmogenic when considered in isolation.

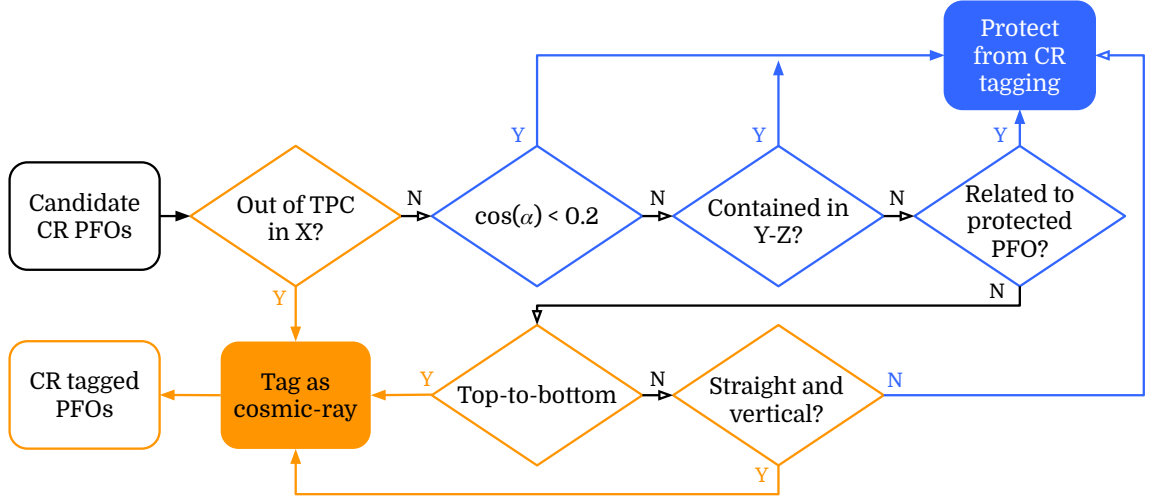
Using this procedure, 4% of clear cosmic-ray PFOs are found to relate to another PFO in the event, of which 96% are also clear cosmic rays, and only  $\sim 3\%$  are clear neutrinos or neutrino fragments. Conversely, 43% of clear neutrino PFOs and 35% of neutrino fragments have a relation, of which 84% and 90% are also clear neutrinos or neutrino fragments respectively.

## 5.3 CR-tagging algorithm

### 5.3.1 Algorithm logic

The logic of the CR-tagging algorithm is summarised in Figure 5.11. First, timing information is used to identify PFOs with an  $X$  coordinate (shown in Figure 5.5) that extends more than 5 cm beyond the physical bounds of the TPC as a cosmic ray. The next stage is to identify PFOs which are likely to be neutrino induced and protect them from CR-tagging. A PFO is protected if it is near orthogonal to the vertical direction (with  $\cos(\alpha) < 0.2$ , shown in Figure 5.9), or is at least 20 cm from the TPC borders in the  $Y - Z$  plane (shown in Figure 5.8). As discussed in Section 5.2.6, relationships between PFOs are identified and the PFOs are organised into connected groups. If a group contains a PFO that has been protected from CR-tagging, then all PFOs in the group are also protected. At this stage, 80% of clear neutrino PFOs and 49% of neutrino fragments are protected from CR-tagging.

The unprotected PFOs are then tagged as a cosmic ray if they are through-going from top-to-bottom. The final cut, shown in Figure 5.12, tags PFOs that are vertical pointing ( $\cos(\alpha) > 0.6$ ) and with low curvature ( $C < 0.04$ ).



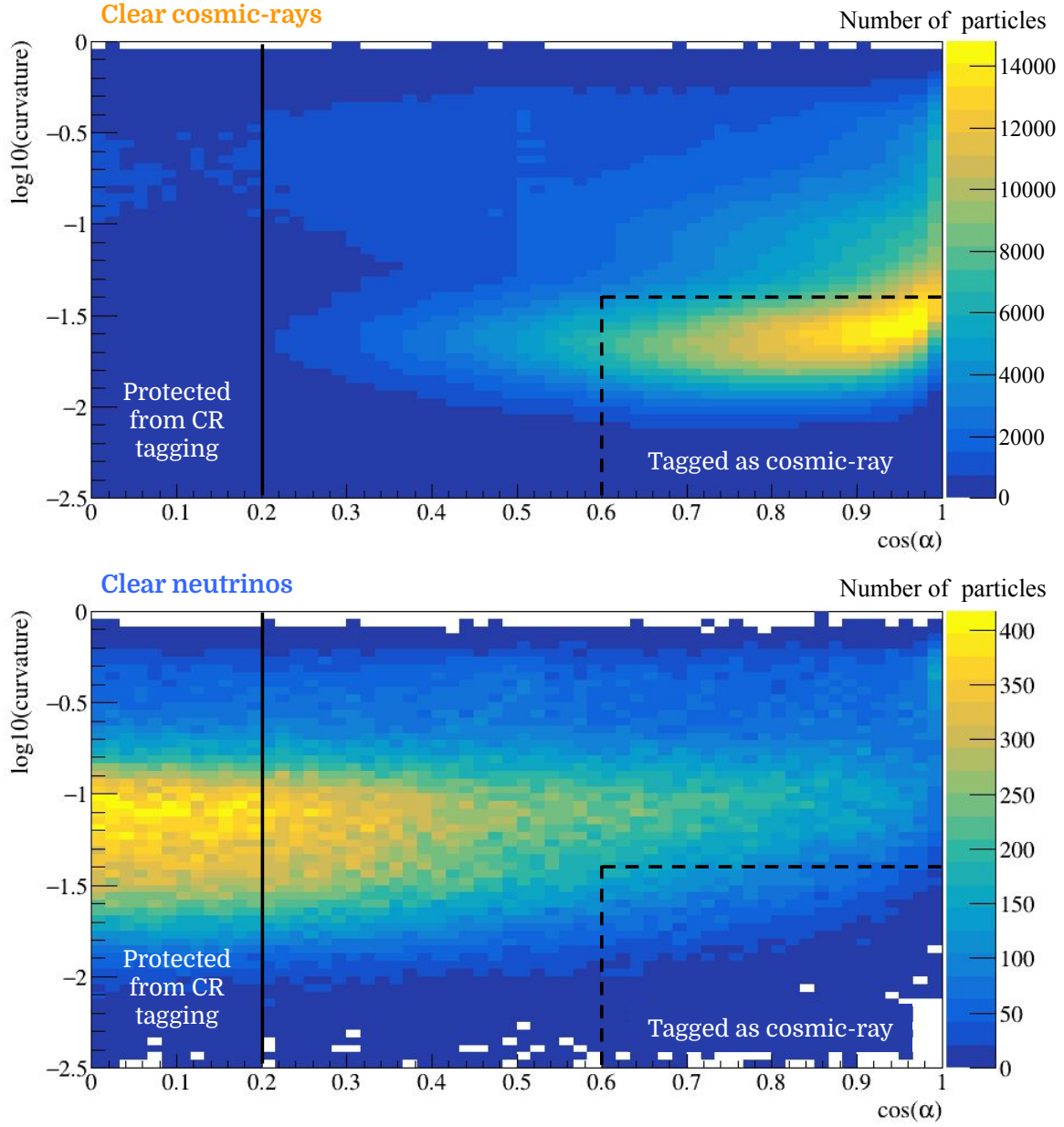
**Figure 5.11:** A flowchart showing the CR-tagging algorithm logic. Each cut either tags a PFO as a cosmic ray (orange), or protects the PFO from being tagged (blue). PFOs following a black line have yet to be tagged or protected.

### 5.3.2 Algorithm performance

The performance of the CR-tagging algorithm is summarised in Table 5.2, and the impact of each cut on clear cosmic ray and clear neutrino PFOs is shown in Figure 5.13.

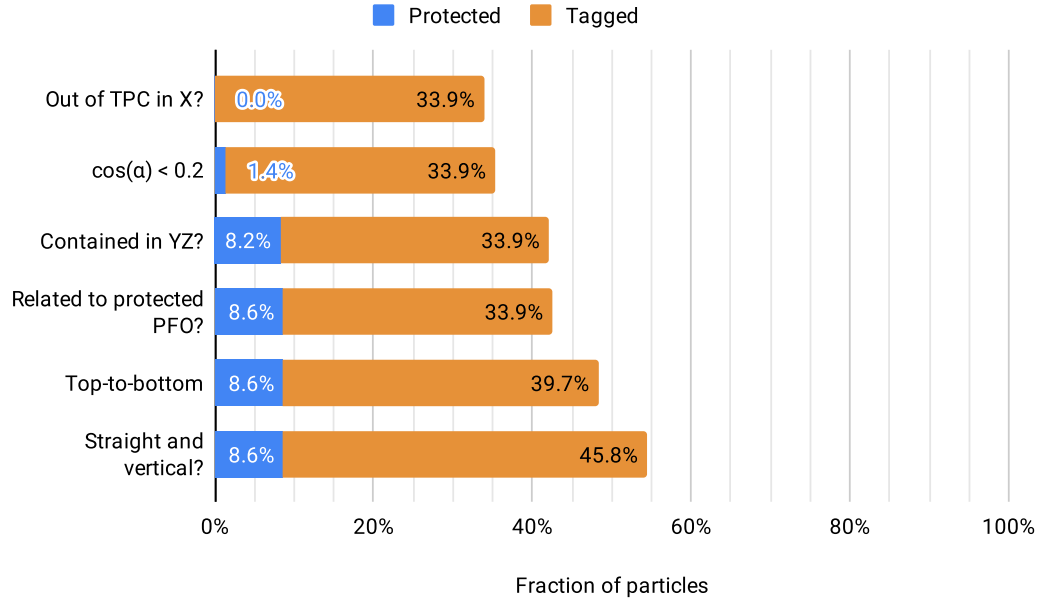
Classification	Fraction CR-tagged
Clear cosmic ray	45.8%
Absorbed	81.8%
Mixed	9.6%
Fragmented	0.1%
Clear neutrino	1.7%

**Table 5.2:** The performance of the CR-tagging algorithm on each reconstructed particle classification.

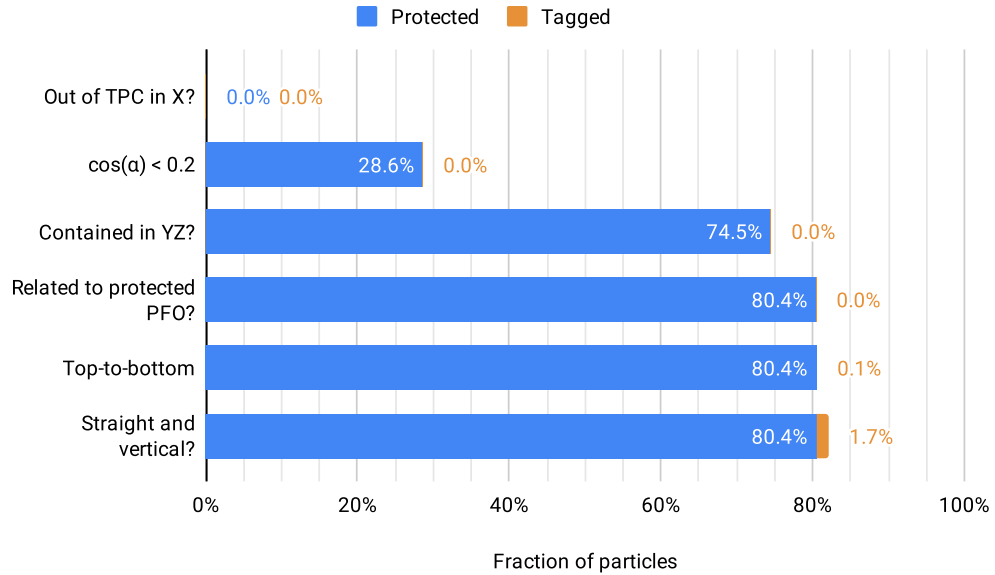


**Figure 5.12:** The distribution of  $\cos(\alpha)$  versus curvature for clear cosmic-ray PFOs (top) and clear neutrino PFOs (bottom). The solid line indicates the region in which PFOs are protected from CR-tagging. The dashed line indicates the region in which otherwise unprotected PFOs are tagged as a cosmic ray.





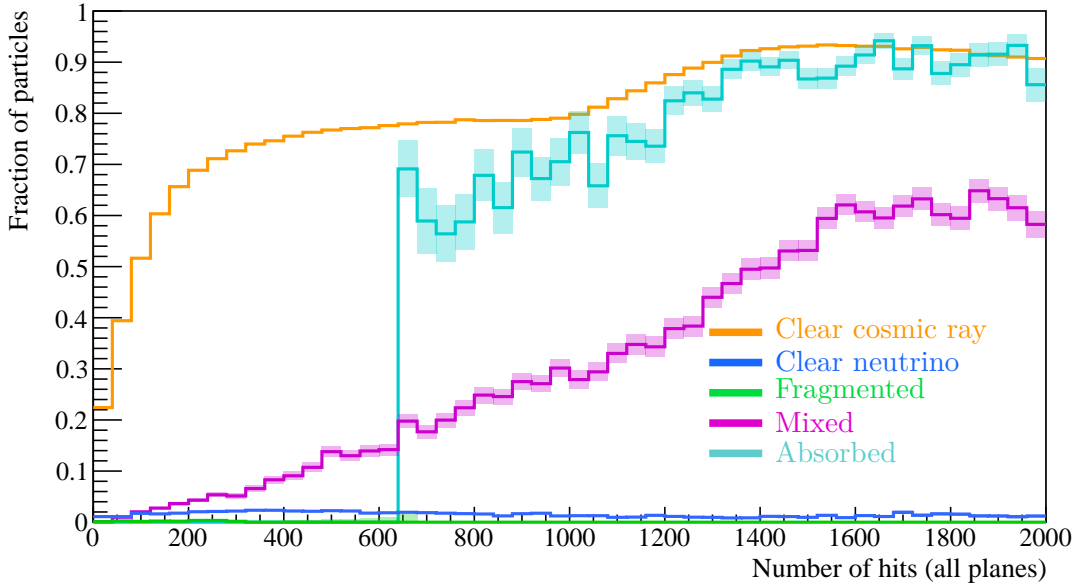
(a) Clear cosmic ray PFOs



(b) Clear neutrino PFOs

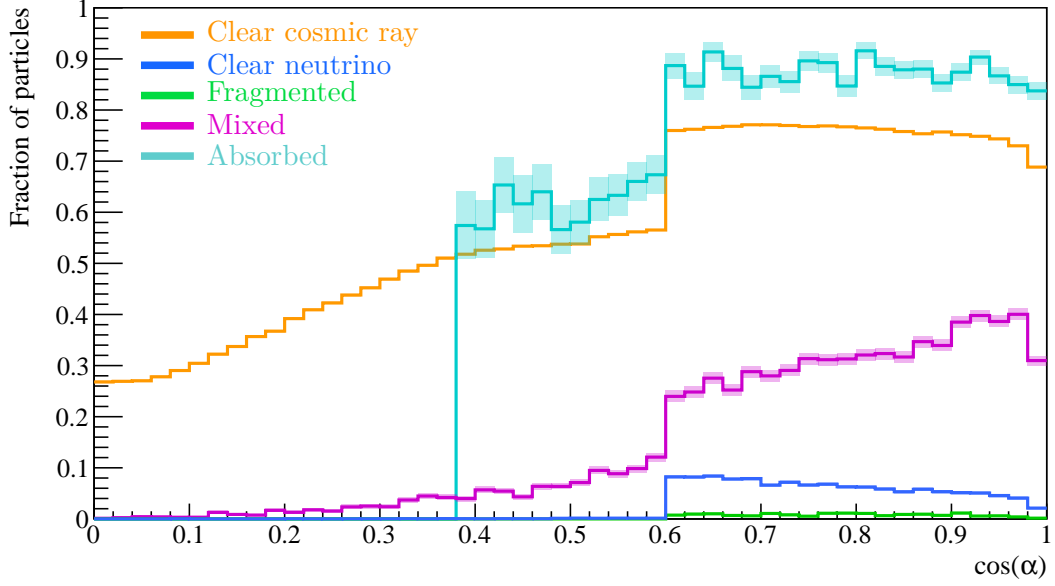
**Figure 5.13:** The fraction of clear cosmic ray (top) and neutrino (bottom) reconstructed particles that are protected from CR-tagging (blue) and are tagged as a cosmic ray (orange) by each of the selection cuts.

Figure 5.14 shows the fraction of PFOs tagged as a function of the number of hits. The tagging rate of clear cosmic rays is lowest below  $\sim 250$  hits over all planes (corresponding to an average track-length of  $\sim 45$  cm) as short PFOs are more likely to be contained in  $X$  and  $Y - Z$ . For longer PFOs, the tagging rate is 75%-90% and the increase above  $\sim 1000$  hits is due to PFOs that are sufficiently long to pierce the TPC from top-to-bottom, as shown in Figure 5.7. For clear neutrino PFOs and neutrino fragments, the tagging rate is largely independent of the number of hits.

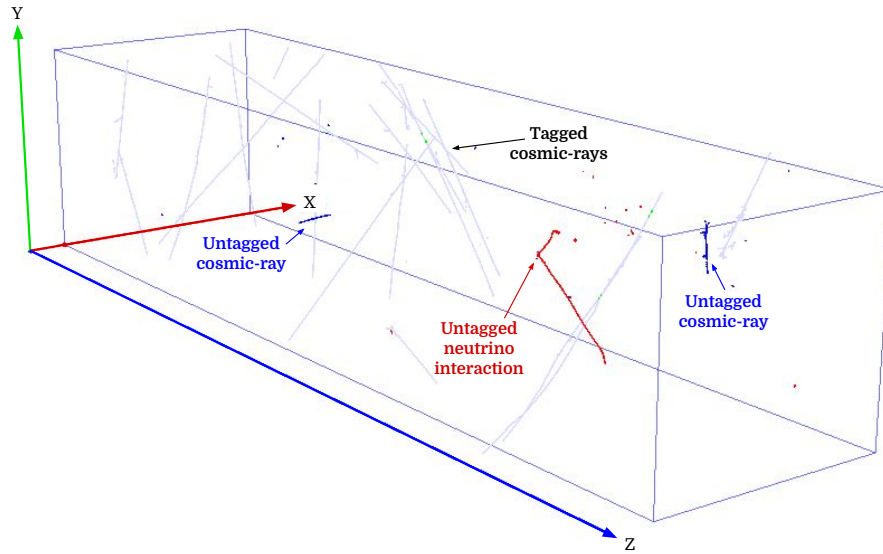


**Figure 5.14:** The CR-tagging rate of each reconstructed particle classification as a function of the number of hits in the PFO over all planes. The absorbed category is truncated at low numbers of hits due to limited statistics in this region. Statistical-only uncertainties are shown as semitransparent bands.

Figure 5.15 shows the CR-tagging rate as a function of the angle made by the PFO to the vertical. In the region  $0 < \cos(\alpha) < 0.6$ , the fraction of clear cosmic rays tagged increases with  $\cos(\alpha)$ . This is driven by the  $Y - Z$  containment cut, as those cosmic rays that appear contained (shown in green in Figure 5.6) enter through the anode or cathode face, which is less likely for more vertically-pointing PFOs. The rate at which clear neutrino PFOs or neutrino fragments are tagged in this region is  $\sim 0.1\%$ . The increase in the tagging rate for  $\cos(\alpha) > 0.6$  is directly due to the cut on the angle and curvature. Finally, the performance of the algorithm is shown visually for an example event in Figure 5.16.



**Figure 5.15:** The CR-tagging rate of each reconstructed particle classification as a function of reconstructed angle to the vertical. In order to reduce misreconstruction, the angle is only found for PFOs with at least 15 space-points. The absorbed category is truncated at low  $\cos(\alpha)$  due to limited statistics in this region. Statistical-only uncertainties are shown as semitransparent bands.



**Figure 5.16:** An example event display showing the result of the CR-tagging algorithm. The grey points belong to reconstructed cosmic-ray particles that have been tagged, whereas blue and red points belong to untagged cosmic rays and neutrino induced activity respectively.

### 5.3.3 Comments and improvements

The CR-tagging algorithm described in this chapter is capable of removing a large fraction of the CR-induced activity in each event while preserving the majority of neutrino-induced activity. The primary failure modes occur when neutrino-induced particles are merged with cosmogenic activity by Pandora’s cosmic-ray algorithm chain. Additionally, the misidentification rate is dependent on the angle to the vertical,  $\alpha$ , as the CR-tagging algorithm is most aggressive for vertical-pointing particles.

Further improvements could be made by using calorimetric and optical information. In particular, calorimetric information could be utilised to identify stopping cosmic rays that enter through the top, upstream or downstream faces by searching for a Bragg-peak. Optical information could be used by matching activity in the TPC (which has superior spatial resolution) with the PMT responses (which have superior timing resolution). Any PFOs that are consistent with optical activity within the beam trigger window could be protected from CR-tagging which would allow the subsequent cuts to be more aggressive.

Additionally, anode or cathode piercing cosmic rays could be identified using optical information. This could be done by considering the displacement,  $\Delta X$ , of the PFO trajectories in the drift direction that would cause the upper end-point of the PFO to touch the anode or cathode. The corresponding value of  $t_0$  for this displacement could then be calculated using in Equation 5.2. If a PFO is truly an anode or cathode piercing cosmic ray, then there should be corresponding optical activity found at this time that is consistent with the PFO’s trajectory in the  $Y - Z$  plane. Finally, the use of SCE-corrected coordinates would allow for the spatial tagging thresholds to be reduced.

## 5.4 Neutrino slice identification

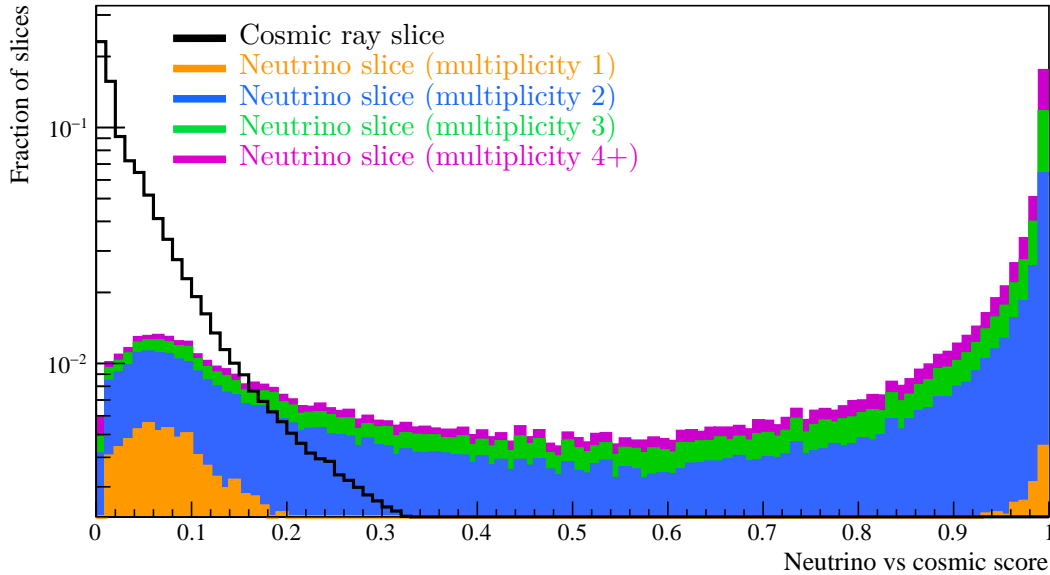
As discussed in Section 4.6, after the CR-tagging stage is complete, the remaining hits are grouped into slices which form the inputs to the neutrino identification stage. A slice is a topologically distinct group of hits within an event that is constructed to represent either a neutrino interaction (including all final-state particles and their secondary interactions) or a cosmic ray. Each slice is reconstructed using both the neutrino and cosmic-ray pattern-recognition algorithm chains and the outputs are available for analysis by the neutrino identification algorithm. The aim of the algorithm is to identify, for a given event, which slice (if any) represents a neutrino interaction. The result of the neutrino algorithm chain will be saved for the neutrino-identified slice and the result of the cosmic-ray chain will be stored for all other slices.

### 5.4.1 The topological score

In general, all of MicroBooNE’s detector subsystems (TPC, LCS, and CRT) are valuable to the neutrino identification step. However, information from MicroBooNE’s LCS and CRT systems is currently inaccessible to a Pandora algorithm. Consequently, a TPC-based neutrino identification algorithm was initially developed by the author of this thesis. This algorithm makes use of a SVM which uses various topological properties of the slices as input. These properties are sensitive to:

- The multiplicity of particles emerging from the vertex. A high-multiplicity vertex is commonly indicative of a neutrino interaction.
- The directions of the particles. Slices which are generally oriented along the beam direction are more likely to contain a neutrino interaction and slices oriented along the vertical are more likely to contain a cosmic ray.
- The proximity of the particles to the top and bottom faces. The details of the use of this information are discussed in Section 5.2.2.

The output of the SVM is a score which classifies each slice on a scale from 0 (CR-like) to 1 (neutrino-like). Figure 5.17 shows this score for slices that truly represent a cosmic ray (black) and those that represent a neutrino (coloured stack). The neutrino slices are broken down by the multiplicity of true final-state particles that are deemed “targets” of the reconstruction, as defined in Section 4.7.1. The separation is largest for slices containing neutrino interactions with a high-multiplicity (3+) of final-state particles that emerge from the interaction vertex, as these slices are topologically distinct from a typical cosmic ray. However, the single-particle neutrino final states (multiplicity 1) are poorly distinguished from cosmic rays using topological information.



**Figure 5.17:** The neutrino-vs-cosmic “topological score” of cosmic-ray and neutrino dominated slices, shown on a logarithmic vertical scale. Only events where the slicing procedure does not split the neutrino interaction into multiple slices and for which the slice containing the neutrino interaction is not contaminated by cosmic-ray activity are shown. For the former constraint, no more than 10% of the hits of a target neutrino final-state particle can be outside of the neutrino slice. For the latter constraint, at most 20% of the hits in the neutrino slice can be of cosmic origin. The neutrino slices are broken down by the multiplicity of target neutrino final-state particles in the event.

As cosmic-ray slices are not clearly isolated from low-multiplicity neutrino interactions, the initial TPC-based algorithm could not reject events in which no neutrino interaction is present. Instead the algorithm would simply identify a neutrino candidate in every event as the slice with the largest topological score.

The neutrino identification algorithm acts as a general-purpose pre-selection for all MicroBooNE analyses using Pandora. Consequently, it is important that any inefficiencies introduced by the algorithm minimally bias the selected neutrino interactions. In general, the use of topological information can introduce such selection biases (as illustrated by Figure 5.17) and must be used with caution. As a result, the TPC-based neutrino identification algorithm described in this section is insufficient for general-purpose use. However, the topological score is used on an analysis-by-analysis basis (including in Part III of this thesis) where care can be taken to ensure no such biases are introduced.

### 5.4.2 Technical contributions

In order to develop a topology-agnostic neutrino identification algorithm, significant technical work was required to access information from the LCS and CRT subsystems, which was performed by the author of this thesis. Although this technical work constituted a significant time commitment, it is only discussed briefly as it is more closely related to software development than to physics research. Nevertheless, this work was necessary to produce MicroBooNE’s Pandora-based physics.

Pandora is designed to be agnostic to many of the details of the experiments it serves. For LArTPC experiments, the input hit objects are abstractions of peaks in a wire waveform. In another detector technology, a hit might represent a “flash” observed by a PMT. Pandora also accepts static experiment-specific information (such as detector geometries or a list of wire-channels that are unresponsive). Otherwise, Pandora is designed to be independent of the wider software environments used by each experiment and can indeed operate as a stand-alone application.

In principle, Pandora could be extended to accept information from MicroBooNE’s LCS and CRT systems. However, this approach was not appropriate as significant LCS and CRT reconstruction developments had already been undertaken by MicroBooNE which were infeasible to be reimplemented within Pandora. Instead, the solution was to allow the wider software framework used by MicroBooNE (LArSoft) access to intermediate states of Pandora’s EDM. In particular, the results of the cosmic-ray and neutrino algorithm chains on each slice were exposed. In this way, the neutrino identification decision can be made outside of Pandora, where access to LCS and CRT information is available.

In addition, the final “event building” step (where CR-tagged PFOs and cosmic-ray or neutrino identified slices are combined to form the final output) was reimplemented within LArSoft. This required a significant refactor of the code that implements the Pandora-LArSoft interface. Consequently, MicroBooNE analyses using Pandora could benefit from an improved topology-agnostic neutrino identification step (using LCS and CRT information) with no code-changes required.

### 5.4.3 Topology-agnostic neutrino identification

With these technical developments made, the implementation of the topology-agnostic neutrino identification algorithm used in the remainder of this thesis (and by other MicroBooNE Pandora-based analyses) was performed in conjunction with other MicroBooNE collaborators. However, the physics logic of this algorithm [98] was *not* developed by the author of this thesis but is summarised below for completeness.

The analysis presented in Part III only uses data from MicroBooNE’s first run for which CRT information is not available (as shown in Table 3.2). Consequently, the neutrino identification algorithm used in this thesis primarily makes use of *optical* information from the LCS.

First, the PMT responses are evaluated in the beam trigger window shown in Figure 3.12. A coincident response of multiple PMTs, known as the “beam flash”, is required with a sufficient number of Photoelectrons (PEs) measured over all PMTs. Initially, the beam flash is assumed to be due to a neutrino interaction. The consistency between the beam flash and each slice is then considered. To do this, the PE-weighted distribution of the positions of the PMTs in the  $Y - Z$  plane is parametrised by a Gaussian, with a mean position  $(Y, Z)$  and standard deviation  $\Delta Y$  and  $\Delta Z$  in each direction. The charge-weighted distribution of the positions of the reconstructed space-points in each slice are also parametrised in the same way. Finally, the ratio of the total charge of each slice to the total PE of the beam flash is evaluated. The consistency of these parameters is used to identify slices that are incompatible with the beam flash. These slices are assumed to be due to a cosmic ray and are no longer considered as candidate neutrino slices.



Additionally, the reconstructed particles from the cosmic-ray algorithm chain are analysed. The improvements to the CR-tagging algorithm described in Section 5.3.3 that require optical information are implemented to identify slices that are likely to contain a cosmic ray and so should no longer be considered as a candidate neutrino slice. In particular, stopping and anode or cathode piercing cosmic rays are identified.

At this stage, any remaining slices are deemed consistent with the beam flash and do not contain a reconstructed particle consistent with CR-induced optical activity. Of events with a simulated neutrino interaction in the FV,  $\sim 16\%$  have no remaining slice,  $\sim 73\%$  have exactly one slice remaining, and  $\sim 11\%$  have multiple slices remaining [98]. The fraction of events with no slices remaining represent the inefficiency of the neutrino identification algorithm <sup>1</sup>. If multiple slices remain, further logic is applied to identify the best neutrino candidate. If the slice with the largest topological score of *all* slices remains ( $\sim 86\%$  of cases [98]), then it is identified as the neutrino. Otherwise, a procedure called flash-matching is employed whereby the expected PMT responses for each reconstructed particle are estimated using the photon visibility map (described in Section 3.4) and the result is compared to the beam flash. A  $\chi^2$  metric is used to measure the consistency between the slice and the beam flash, and the slice with the lowest  $\chi^2$  is selected as the neutrino candidate.

The final stage of the neutrino identification algorithm aims to remove events where no neutrino interaction occurred and instead the beam flash was produced by a cosmic ray. To do this, the flash-match  $\chi^2$  is calculated between the beam flash and all reconstructed particles that are tagged as a cosmic ray by the algorithm described in Section 5.3. The lowest such value,  $\chi_{CR}^2$ , is then found. Additionally, the flash-match  $\chi^2$  is calculated between the beam flash and the selected candidate neutrino slice. If this value is significantly larger than  $\chi_{CR}^2$  then the beam flash is assumed to be due to a cosmic ray and no slices are identified as a neutrino. The impact of this cut on the efficiency is  $< 1\%$  [98].

The slices that are selected by the neutrino identification algorithm form the input to all modern Pandora-based analyses of neutrino interactions with MicroBooNE — including the analysis presented in Part III of this thesis. Together, the cosmic-ray tagging and neutrino identification algorithms (executed as part of Pandora’s consolidated algorithm flow) provide a general purpose method of identifying neutrino interactions in dense cosmic-ray environments.

---

<sup>1</sup>Strictly, this is the convolution of pattern recognition, slicing and neutrino identification losses.



## Part III

### Charged-current single charged pion muon-neutrino interactions



# Chapter 6

## Identification of $\text{CC}1\pi^\pm$ interactions in MicroBooNE

### 6.1 Introduction

This chapter describes the methods by which muon-neutrino charged-current single charged pion interactions are identified in simulated and experimental MicroBooNE data. The selected events are used in Chapter 7 to extract forward-folded differential cross-sections, and the motivation for this measurement was presented in Section 2.4. This selection builds on the work presented in Part II of this thesis and uses particles reconstructed by Pandora that have been identified as neutrino-induced by the consolidated algorithm flow. The work presented in Part III of this thesis is currently under internal review by MicroBooNE and other collaborators are in the process of completing the required final steps (discussed in Section 7.5.3) before the analysis can be published.

This chapter begins by defining the aims and strategy of the analysis. Next, an overview of MicroBooNE's  $\nu_\mu$  CC-inclusive selection is presented in Section 6.2. This is used as a pre-selection. The methods developed to identify the species of Pandora-reconstructed particles are discussed in Section 6.3. Finally, Section 6.4 describes the selection criteria that are used to isolate  $\text{CC}1\pi^\pm$  interactions from cosmic rays and other neutrino-interaction backgrounds.

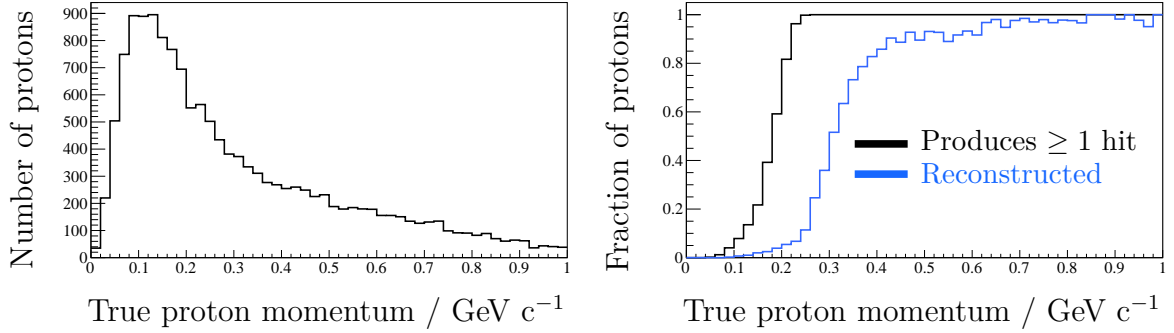
### 6.1.1 Signal definition

In general, the muon-neutrino charged-current single charged pion topology refers to muon neutrino interactions that result exclusively in a muon, a single charged pion, and any number of nucleons in the final state (including zero). The existing measurements of this topology (summarised in Section 2.4) each have subtleties in their signal definition. As the sign of the charge of a particle can not typically be determined in MicroBooNE, this analysis does not make a distinction between the  $\pi^+$  and  $\pi^-$  channels and similarly  $\bar{\nu}_\mu$  interactions (which occur with a  $\mu^+$ ) are included as part of the signal, although they form a small ( $< 1\%$ ) contribution. In addition to the topological definition ( $1\mu^\pm + 1\pi^\pm + N_n n + N_p p$ , where  $N_n, N_p$  can take any value including zero), the label of “signal” is restricted to events in which the true neutrino interaction vertex is within the FV - shown in Figure 4.13. In Chapter 7, additional phase-space restrictions are applied to the signal definition to extract the cross-section but are not applied in this chapter.

### 6.1.2 Aims and analysis strategy

This analysis aims to measure the forward-folded flux-integrated  $\nu_\mu$   $\text{CC1}\pi^\pm$  differential cross-section on argon (defined formally in Chapter 7). The cross-section is extracted with respect to the muon and pion momentum ( $p_\mu, p_\pi$ ), the angle made by these particles to the beam direction ( $\theta_\mu, \theta_\pi$ ), and muon-pion opening angle  $\theta_{\mu\pi}$ . Additionally, the total cross-section of the exclusive channels  $N_p = 0$ ,  $N_p = 1$  and  $N_p \geq 2$  is also found. The measurements of  $p_\pi$  (on argon) and  $N_p$ , when published, will constitute a world’s first. Therefore it is necessary to be able to identify events with a muon, a charged pion and any number of protons, and reconstruct all of these particles accurately.

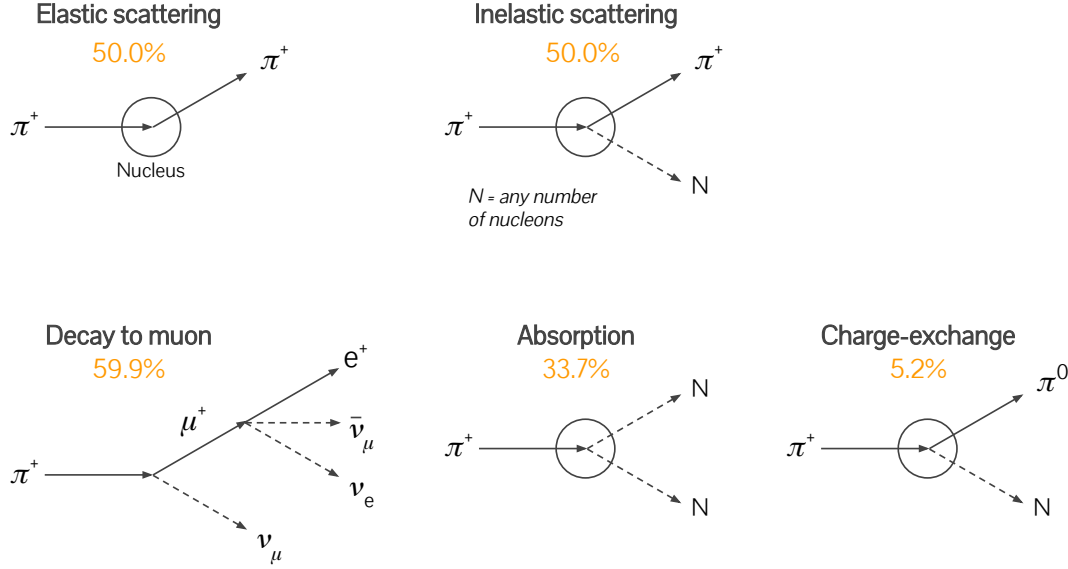
In order to measure the proton multiplicity, low detection thresholds and excellent spatial resolution are required - both of which are possible with MicroBooNE. The efficiency with which Pandora can reconstruct protons in events with the  $\text{CC1}\pi^\pm$  topology was discussed in Section 4.7.2. Figure 6.1 shows the momentum of simulated protons in true  $\text{CC1}\pi^\pm$  events passing the  $\text{CC1}\pi^\pm$  selection used in this analysis (which is described in Section 6.4). The majority of protons with a momentum above  $\sim 200 \text{ MeV } c^{-1}$  produce hits, and are reconstructed above  $\sim 300 \text{ MeV } c^{-1}$ . Consequently, a proton momentum threshold of  $300 \text{ MeV } c^{-1}$  is applied throughout when counting proton multiplicities.



**Figure 6.1:** The true proton momentum distribution in  $\text{CC1}\pi^\pm$  events (left), and the proton reconstruction efficiency (right) in events passing the generic  $\text{CC1}\pi^\pm$  selection. The black line shows the fraction of protons for which a hit is reconstructed. The blue line shows the fraction of protons for which a reconstructed particle is matched.

Without the aid of a magnetic field, the pion momentum measurement is particularly challenging. Charged pions are prone to undergo secondary interactions on the argon. Consequently, the range of a pion’s trajectory is (in general) not a good predictor of its momentum. During an inelastic secondary interaction, the pion transfers energy to the argon which may not result in an ionisation signal, and so a simple calorimetric sum of the deposited charge is also typically not a good predictor of  $p_\pi$ .

Figure 6.2 shows a schematic of the most common charged-pion secondary interactions, split broadly into in-flight (those that result in a charged pion in the final state, i.e. elastic and inelastic scattering), and end-states (those that don’t result in a charged pion). In simulated  $\text{CC1}\pi^\pm$  events (described in Section 3.4), the pion undergoes one or more scatters in 24% of events, with an even split between elastic and inelastic scatters. If a pion undergoes a scatter, it will in general lose energy in the process. A range-based momentum estimation method will not account for the energy lost in such a scatter, and so will tend to underestimate the true momentum. If a pion interacts inelastically with the argon and does not remain in the final state, then again the momentum of the particle (as tracked up to the point of the interaction) will be underestimated by the range-based method. However, the range-based method is applicable if the pion decays, provided that the decay occurs at rest. The most common charged pion decay mode is to a muon, which occurs at rest in 95% of cases in the simulation. Of the pions that come to rest, 99% decay to a muon.

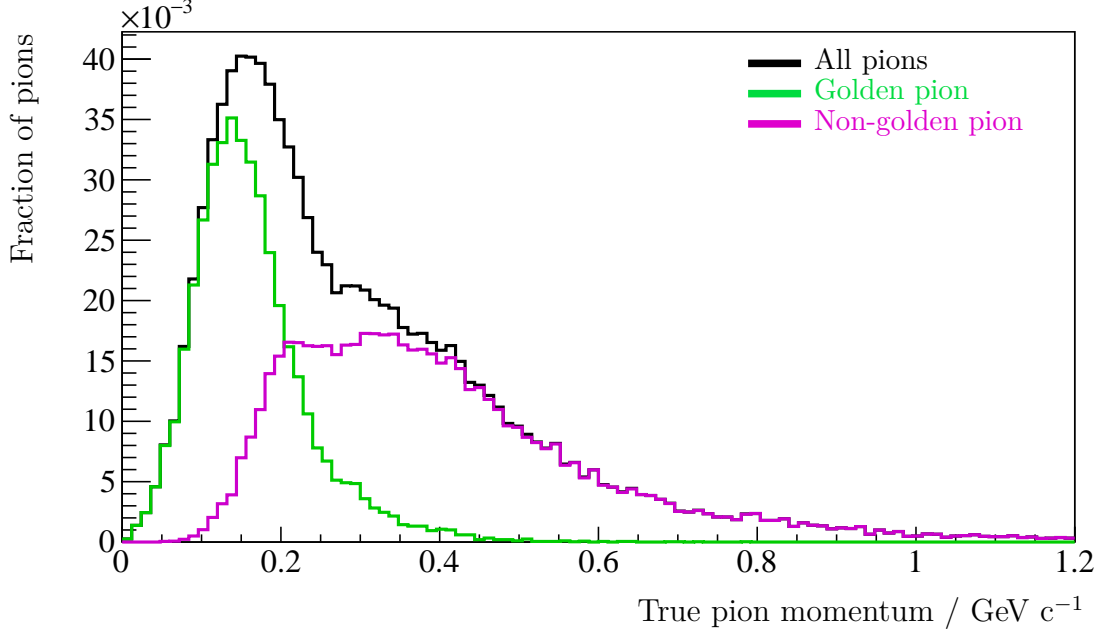


**Figure 6.2:** Overview of the most common charged pion interactions in MicroBooNE. The top row shows in-flight interactions in which the charged pion remains in the final state. The bottom row shows the most common end-state interactions, and the fractions give their rates in  $CC1\pi^\pm$  events.

In this work, the pion momentum measurement is made on a subset of  $CC1\pi^\pm$  events that contain a *golden pion*. A golden pion is defined as one that is contained (with fitted end points at least 5 cm from all TPC faces), does not scatter on the argon, and comes to rest. Hence, the momentum of a golden pion can be found by range, as detailed in Section 6.6. The restriction of the momentum measurement to only golden pions is a trade-off between complexity in the momentum estimation and complexity in the selection. It is important to note that this restriction is not uniform over all pion momenta. As shown in Figure 6.3, golden pions are typically less energetic than their non-golden counterparts. This is primarily because the likelihood of an interaction with the argon increases with the distance travelled by the pion, which in turn depends on its initial momentum.

In this chapter, two complimentary  $CC1\pi^\pm$  selections are presented. The *generic selection* aims to extract a sample of events enriched with  $CC1\pi^\pm$ , but without any regard for the secondary interactions of the pion. This selection is used to extract the cross-section in terms of proton multiplicity ( $N_p$ ), the muon kinematics ( $p_\mu$ ,  $\theta_\mu$ ), the pion angle ( $\theta_\pi$ ), and the muon-pion opening angle ( $\theta_{\mu\pi}$ ). Alternatively, the *golden selection* aims to extract a subset of these events that is enriched with golden pions. This selection is used to extract the cross-section in terms of the pion momentum ( $p_\pi$ ).





**Figure 6.3:** The true momentum of all pions (black) and golden pions (green) and non-golden pion (pink) produced in true  $\text{CC}1\pi^\pm$  events.

## 6.2 $\nu_\mu$ CC-inclusive pre-selection

MicroBooNE’s flagship “low energy excess” search is performed as a blind analysis. Consequently, to access data for a different purpose (such as this work), one must demonstrate that their selection is not sensitive to low energy electron neutrinos. The  $\nu_\mu$  CC-inclusive selection presented in this section has been developed by the MicroBooNE collaboration [98] and meets these unblinding criteria. The selected events are used as a starting point for this analysis. The selection aims to identify muon-neutrino Charged-Current interactions, which have a muon in the final state. All other final-state particles are allowed, including those in the  $\text{CC}1\pi^\pm$  topology.

### 6.2.1 Overview of the pre-selection

The CC-inclusive pre-selection is described in [98] and summarised here. The selection begins with Pandora’s consolidated algorithm flow. Only the events for which a neutrino slice is identified are considered, as described in Section 5.4. To further reduce cosmic-ray backgrounds, the selection applies restrictions on the topological score (presented in Chapter 5), and the extent to which the observed optical activity is consistent with the reconstructed particles in the selected slice.

Contributions from NC and  $\nu_e$ -CC backgrounds are reduced by determining if any of the reconstructed final-state particles are likely to be a muon. Muon candidates are distinguished from electromagnetic showers using the response of Pandora’s track-vs-shower SVM classifier and calorimetric information is used to distinguish muons from protons. Events in which multiple particles are viable muon candidates typically include charged pions that pass the muon selection criteria. In these cases, the longest candidate is simply identified as the muon (which is correct in 83% of simulated  $\text{CC1}\pi^\pm$  passing the pre-selection). Any events without a muon candidate are rejected. Otherwise, the candidate is required to have a track-length of at least 20 cm to meet the unblinding criteria (corresponding to a momentum threshold of  $\sim 150 \text{ MeV}/c$ ). It is important to note, in this analysis, an alternate procedure is used to identify the muon in  $\text{CC1}\pi^\pm$  events to reduce the muon-pion confusion rate (described in Section 6.3).

Electron neutrino backgrounds are further reduced by insisting that the reconstructed final-state particle with the most hits is track-like (using the track-vs-shower SVM score). Finally, the reconstructed neutrino interaction vertex must be within the FV shown in Figure 4.13, and the start position of each final-state particle must be at least 10 cm from all TPC borders.

### 6.2.2 Events passing the pre-selection

Table 6.1 shows a breakdown of the events passing the CC-inclusive pre-selection. In this analysis, only data from run-1 is used corresponding to an exposure of  $1.455 \times 10^{20}$  POT. MicroBooNE’s run periods are shown in Figure 3.4. Events in the “overlay” sample consist of a simulated neutrino interaction with overlaid beam off (cosmic) data, as described in Section 3.4. The overlay sample is constructed so that every event contains a neutrino interaction. Consequently, events from EXT (beam-off) data are added to account for cases where the trigger would have fired but no neutrino interaction occurred. Additionally, simulated “dirt” events are added which account for cases where a neutrino from the BNB interacts with material outside of the TPC. The samples are normalised to the POT of the beam-on BNB data and the combination of overlays, dirt and beam-off data form the predicted event rate. In total 28,417 BNB data events pass the pre-selection, which is under-predicted by the simulation. The data/prediction ratio of 1.07 after the CC-inclusive selection is known to be due to a mismodelling of optical information. At the time of writing, work is ongoing by the collaboration to correct this.

Of all predicted  $\text{CC1}\pi^\pm$  signal events, 54.0% pass the pre-selection and comprise 9.8% of the selected sample. The passing-rate of signal events is largely independent of the kinematic parameters of interest (listed in Section 6.1.2). The main exceptions are in the region  $p_\mu < 150 \text{ MeV}/c$  (due to the restriction on the muon candidate track-length), and for muon angles close to the vertical direction (as this is the region in which the CR-tagging is most aggressive).

Sample	Events	Fraction of events
<u>Signal (simulation)</u>		
Overlay: $1\mu + 1\pi^\pm + Xp$ (non-golden)	1545	5.84%
$1\mu + 1\pi^\pm + Xp$ (golden)	1057	3.99%
<u>Backgrounds (simulation)</u>		
Overlay: $1\mu + Xp$	13120	49.57%
Non-fiducial	1868	7.06%
$1\mu + 1\pi^0 + Xp$	1077	4.07%
$1\pi^\pm + Xp$	336	1.27%
$1\mu + 2\pi^\pm + Xp$	334	1.26%
$1\mu + 1\pi^0 + 1\pi^\pm + Xp$	229	0.86%
$Xp$	152	0.57%
$2\pi^\pm + Xp$	75	0.28%
Other	362	1.37%
Dirt	981	3.70%
EXT data	5335	20.16%
Total signal	2602	9.83%
Total backgrounds	23868	90.17%
Total simulation	26470	
Total BNB data	28417	

**Table 6.1:** The composition of events that pass the CC-inclusive pre-selection. The overlay, dirt and EXT (beam-off) data have been normalised to the POT of the BNB data. Here  $Xp$  refers to any number of protons including zero. Neutrons are not listed in the topology but their multiplicity may be non-zero. The right-most column refers to the fraction of selected events that are of the corresponding topology. In total, only 20% of simulated events in the overlay sample pass the pre-selection, however of those that are rejected, 79% are non-fiducial. The passing rate of  $\text{CC}1\pi^\pm$  signal events is 54.0%.

## 6.3 Identification of particle species

In order to identify  $\text{CC}1\pi^\pm$  events, it is necessary to distinguish between particle types. To achieve this task, this analysis uses three Boosted Decision Trees (BDTs) which are trained using the topological and calorimetric features described in this section. First, the *muon BDT* is used to identify the reconstructed particle in each input event that is most likely to be a muon. Next, the *proton BDT* is used on the remaining particles to identify events in which exactly one is not consistent with a proton - this is assumed to be the pion candidate. Subsequent criteria are employed to remove background events for which this assumption does not hold. Finally, for the golden selection, the *golden pion BDT* is used on the pion candidate to obtain a golden-enhanced sample of events. The full details of this selection are given in Section 6.4.

The features used to distinguish between reconstructed particle types fall into one of two categories: the topological features make use of information regarding the spatial distribution of the hits and space-points assigned to each reconstructed particle; the calorimetric features make use of the charge associated with each particle, and how this charge is distributed along the particle's reconstructed trajectory.

### 6.3.1 The use of calorimetric information

MicroBooNE has developed a procedure to use calorimetric information for track-like particles, described in [85] and summarised here. First, a track-fit is applied to each reconstructed neutrino final-state particle (as described in Section 4.8), and the resulting trajectory is SCE-corrected. The hits assigned to the particle are ordered by their distance,  $x$ , from the start of the track (along the fitted trajectory) and are used to determine the rate of charge deposition  $\frac{dQ}{dx}(x)$  along the trajectory. A calibration procedure (described in Section 4.8.2) using anode-to-cathode crossing CR-muons from beam-off data is used to correct  $\frac{dQ}{dx}(x)$  for any spatial and temporal non-uniformities (for example due to unresponsive wire regions, or changes in the argon purity over time). The rate of energy loss  $\frac{dE}{dx}(x)$  is then calculated using the modified-box recombination model:

$$\frac{dE}{dx} = \frac{1}{\beta} \cdot \rho E_D \left[ \exp \left( \beta \cdot \frac{W_{ion}}{\rho E_D C_{cal}} \cdot \frac{dQ}{dx} \right) - \alpha \right]. \quad (6.1)$$

Here,  $\rho$  is the density of liquid argon,  $E_D$  is the electric drift-field strength,  $W_{ion}$  is the ionisation energy of argon, and  $C_{cal}$  is a calibration constant. The model has two free

parameters,  $\alpha$  and  $\beta$ , whose values were measured by the ArgoNeuT experiment [101], and used by MicroBooNE. In general, the drift-field strength is non-uniform (for example due to SCE), however these variations are assumed to be calibrated-out and so a constant field strength is used for  $E_D$ . The remaining variations of the drift field are covered by the systematic uncertainties described in Section 7.3.6. The value of  $C_{cal}$  is set by fitting the  $\frac{dE}{dx}$  curve as calculated by Equation 6.1, to the expected distribution for a sample of stopping muons from beam-on data [85]. MicroBooNE have also validated the free parameters in this model by tuning to a sample of protons from beam-on data. Any differences are used to motivate the uncertainties described in Chapter 7.

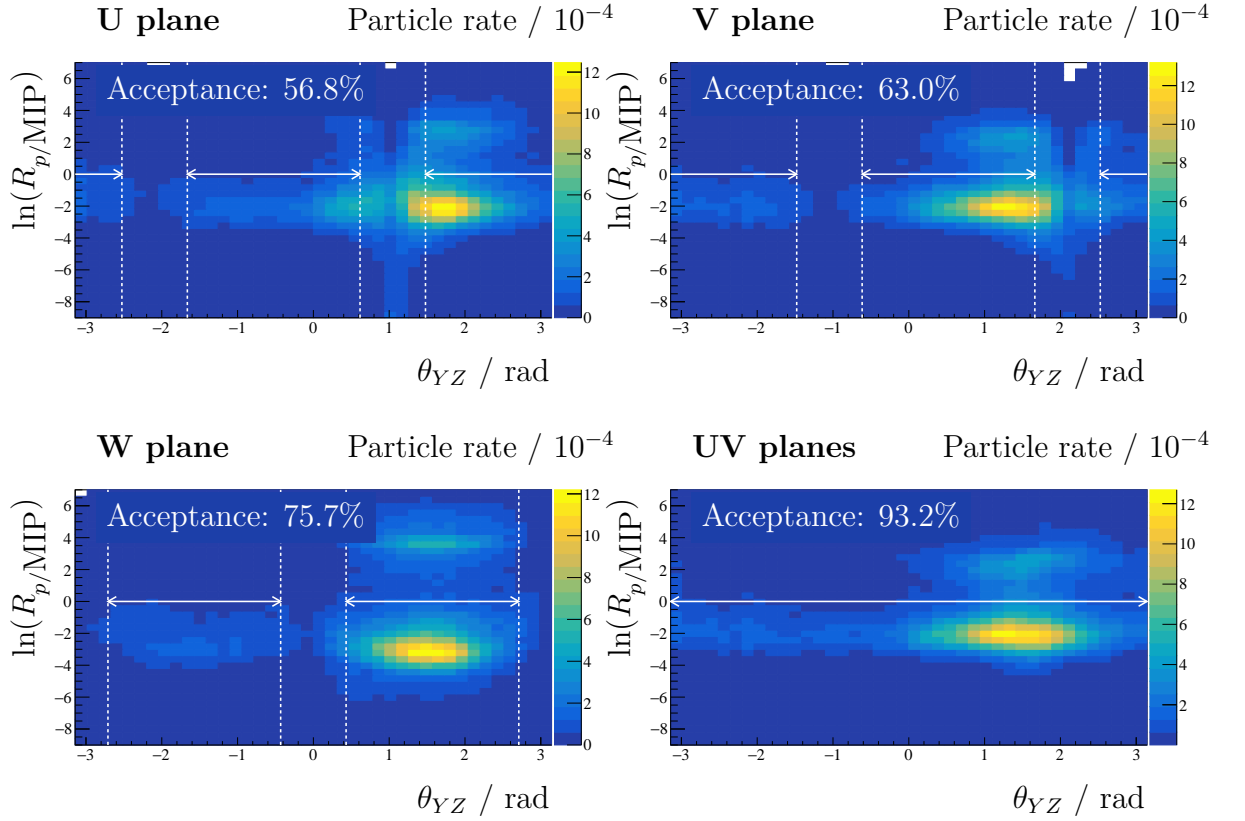
In general, the reliability of calorimetric information on a given readout plane is dependent on the direction of the particle. In particular, if the projection of a particle's direction in the  $Y-Z$  plane,  $\theta_{YZ}$ , is near parallel to one of the wire directions,  $\theta_{YZ}^P$  (where  $P = U, V, W$ ), then all of the charge deposited by the particle will be read-out by a single (or very few) wires on that plane. In this case, the hit-finding procedure (which searches for Gaussian peaks in the wire waveforms, described in Section 4.2) will often poorly estimate the deposited charge as the wire responses are non-Gaussian, and the resulting calorimetric information is typically not reliable. Figure 6.4 shows the distribution of one of the calorimetric features used to distinguish between particle species as calculated on each readout plane as a function of  $\theta_{YZ}$  for reconstructed particles passing the pre-selection. The regions close to  $\theta_{YZ}^P$  are indicated by dashed vertical lines. This feature is shown to illustrate the dependence of the reliability of calorimetric information on the particle direction but is discussed in Section 6.3.2.

In this analysis, a given calorimetric feature,  $f_P$ , (calculated using information from plane  $P$ ) is deemed unreliable if:

$$\sin^2(\theta_{YZ} - \theta_{YZ}^P) \equiv \sin^2(\Delta\theta_{YZ}^P) < 0.175. \quad (6.2)$$

This restriction corresponds to a minimum angle between the reconstructed particle and the wire of  $\Delta\theta_{YZ}^P \sim 25^\circ$ , which is chosen to remove effectively all unreliable information while retaining as much reliable information as possible. The reliable information from all three planes is combined,  $f(f_U, f_V, f_W)$ , to ensure calorimetric information is available over the full range  $-\pi < \theta_{YZ} < +\pi$ . If information from the collection plane is available (i.e. the particle is not in an unresponsive region) and deemed reliable, then it is used to calculate the feature:  $f = f_W$ . If the collection plane can not be used but reliable information is available on both induction planes, a hit-weighted average is used:

$f = w_U f_U + w_V f_V$ , where  $w_U = N_U / (N_U + N_V) = 1 - w_V$ , and  $N_P$  is the number of hits associated with the particle on plane  $P$ . Finally, if reliable information is only available on one of the induction planes, then it is used as:  $f = f_P$ , where  $P = U, V$ . Using this procedure, reliable calorimetric information is available for 98.4% of reconstructed particles passing the pre-selection for which a track can be fitted.



**Figure 6.4:** The proton-to-MIP Bragg likelihood ratio as a function of  $\theta_{YZ}$  for particles in simulated events passing the CC-inclusive selection. The induction planes (U and V) are shown on the top left and right respectively, and the collection plane (W) is on the bottom left. The white vertical dashed lines are drawn either side of the readout wire angles at  $\theta_{YZ}^P \pm \Delta\theta_{YZ}$ . The regions indicated by horizontal white arrows are those in which the calorimetric information is deemed reliable. The regions without arrows are not used in the analysis. The bottom right plot shows the weighted average of the induction planes (described in the text). The upper horizontal band (with a log likelihood ratio of  $\sim +2$ ) is dominated by protons, and the lower band (around  $\sim -2$ ) is dominated by muons. The acceptance written on each figure is the fraction of reconstructed tracks in events passing the pre-selection for which reliable calorimetric information is available on the plane (or planes) indicated. In particular, the UV acceptance is the fraction of tracks with reliable calorimetric information on either (or both) of the U or V planes.

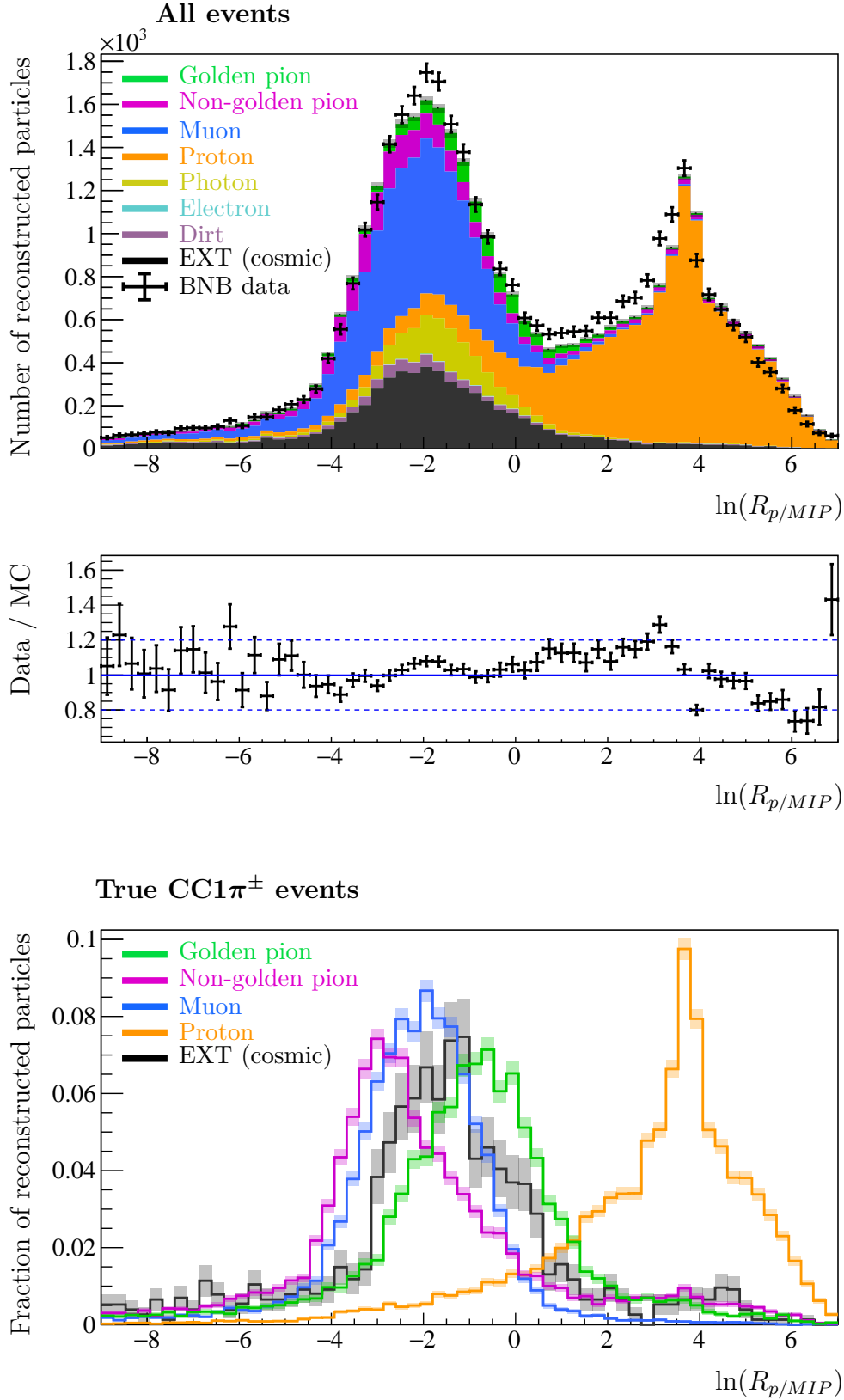
### 6.3.2 Features for particle identification

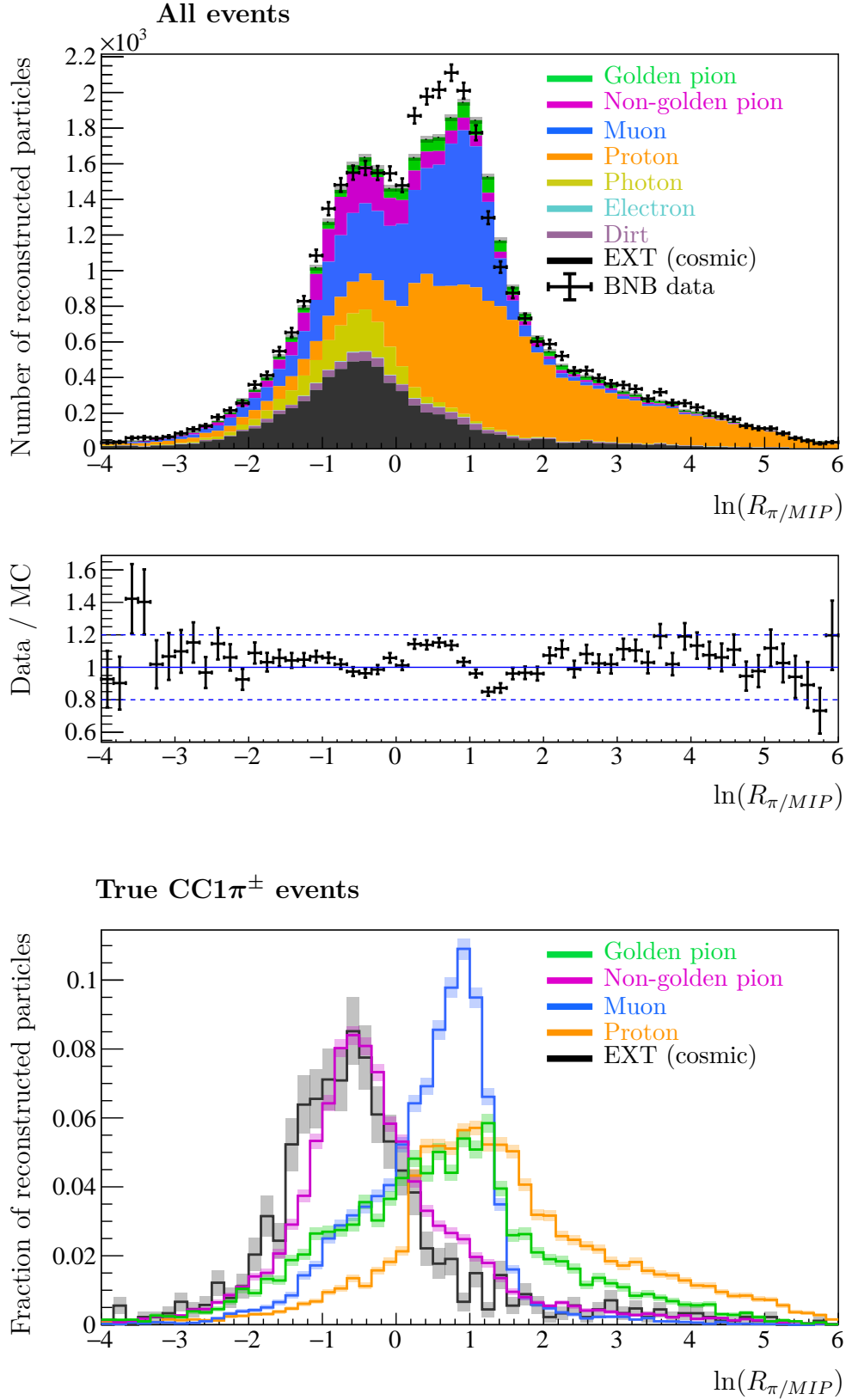
The first two calorimetric features used in this analysis are based on the **Bragg-likelihood** algorithm, developed by MicroBooNE [102]. The algorithm uses the energy loss curve,  $\frac{dE}{dx}(x)$ , over the last 30 cm of a reconstructed track to help distinguish between particle types. The shape of this curve is dependent on the particle species. In particular, the algorithm gives the likelihood,  $L^T = P(\frac{dE}{dx}(x)|T)$ , that a true particle of a given type,  $T$ , would produce the measured energy loss curve. For a given reconstructed particle of unknown type, this analysis considers the ratio of these likelihoods under different hypotheses (e.g.  $R_{p/\mu} = L^p/L^\mu$ ) as a metric for which hypothesis is more probable.

In this analysis, information from all three planes is combined following the procedure described in Section 6.3.1. In particular, the likelihood under the proton, charged-pion and Minimum Ionising Particle (MIP) hypotheses are used. The proton-to-MIP ( $R_{p/MIP}$ ) and pion-to-MIP ( $R_{\pi/MIP}$ ) likelihood ratios were identified as input features to the particle identification BDTs, and are shown in Figure 6.5 and Figure 6.6 respectively.

For all features discussed in this section, the distribution of contained reconstructed neutrino final-state particles is shown for BNB data and simulated events passing the pre-selection. The uncertainties shown are statistical-only. Reconstructed particles from overlay events are broken down by the species of the simulated particle with which most hits are shared (where each hit is associated with the simulated particle which contributes most to its integrated charge). Additionally, the normalised distributions of each particle species are shown for true  $\text{CC1}\pi^\pm$  events passing the pre-selection.

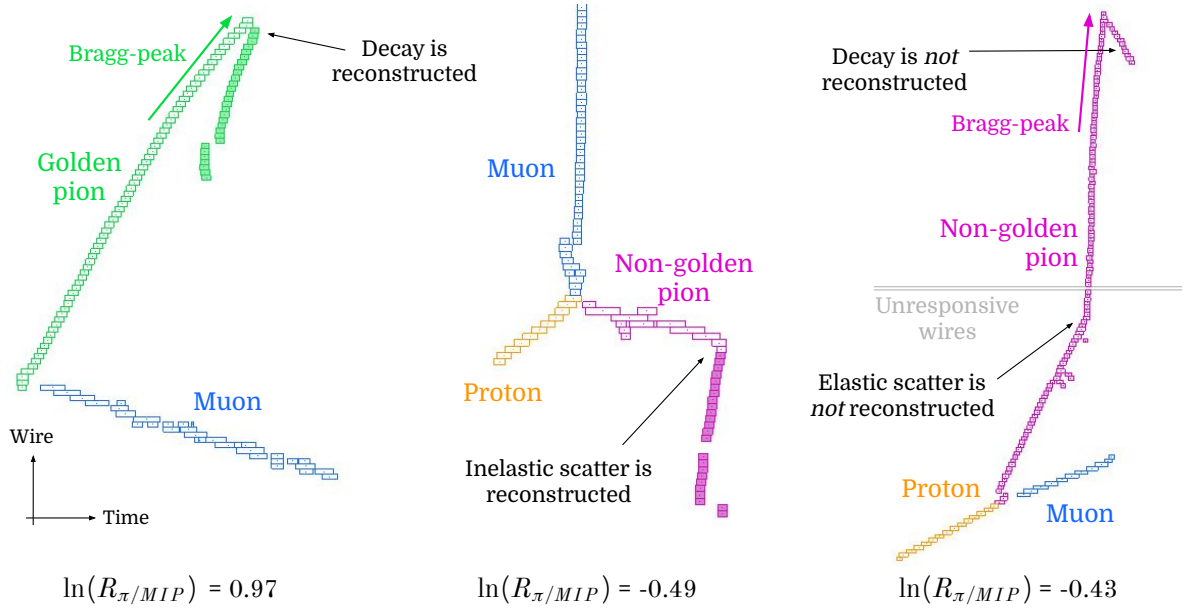


**Figure 6.5:** The 3-plane Bragg likelihood ratio under the proton and MIP hypotheses.



**Figure 6.6:** The 3-plane Bragg likelihood ratio under the pion and MIP hypotheses.

The feature,  $R_{p/MIP}$  can separate protons from other particle species, whereas  $R_{\pi/MIP}$  primarily provides separation power between muons and non-golden pions. Although the Bragg-peak shapes of muons and pions are very similar, the latter separation primarily occurs due to the presence of secondary interactions, as illustrated by Figure 6.7. The left and right displays show a pion that comes to rest before decaying to produce a secondary electron shower. In the centre and right displays, the pion scatters on the argon and so is not “golden”. In the left and centre displays, the secondary interaction is identified by Pandora and the interaction product is reconstructed as a daughter PFO, whereas the decay is missed in the right display. As the track-fit is only applied to reconstructed neutrino final-state particles (and not their daughters), the value of  $R_{\pi/MIP}$  for the pion in the left display is large due to the existence of a Bragg-peak, whereas the pion in the central display appears MIP-like. In the right display, the hits from the decay electron are included in the track-fit and obscure the Bragg-peak, resulting in a low value of  $R_{\pi/MIP}$ .



**Figure 6.7:** Event displays showing three  $\text{CC}1\pi^\pm$  events. The value of  $\ln(R_{\pi/MIP})$  is shown for the pion in each event. The hits belonging to each reconstructed particle are assigned a colour which indicates the best-matched true particle species. Any hits that belong to a reconstructed daughter particle are shaded solid. In the rightmost figure the decay electron is not identified as a daughter of the pion produced in the neutrino interaction and instead both the pion and decay electron are merged into a single reconstructed particle.

Consequently, non-golden pions tend to be distributed at low  $\ln(R_{\pi/MIP}) \lesssim 0$ , whereas muons (which are less likely to undergo visible secondary interactions, and so have a reconstructed Bragg-peak) are primarily distributed at high  $\ln(R_{\pi/MIP}) \gtrsim 0$ . The pion-

MIP likelihood ratio for golden pions is dependent on the existence of a visible decay electron, and whether the electron is reconstructed as a secondary daughter particle.

The next BDT feature is the **truncated mean  $dE/dx$** . This focusses on the average rate of energy loss at the start of the track. As the rate of energy loss is Landau distributed (which has an undefined mean), special care is taken to estimate the average. In each view separately, the  $dE/dx$  of the hits in the first third of the track are collated (excluding the first 3 hits which often include charge from other particles near the neutrino interaction vertex). The median and standard deviation of these values are found, and any hits with a  $dE/dx$  greater than one standard deviation away from the median are ignored (this reduces the impact of hits in the tails of the distribution). The mean of the  $dE/dx$  of the remaining hits is calculated, and the results from all readout planes are combined according to the procedure described in Section 6.3.1. This feature is complementary to the Bragg likelihoods and provides further separation of protons from other particle types, as shown in Figure 6.8.

The remaining features used by the BDTs are topological. The **number of descendants** makes use of the reconstructed particle hierarchy identified by Pandora, by counting the number of descendants (daughters, granddaughters, etc.) of the reconstructed neutrino final-state particle. As shown in Figure 6.9, this feature provides separation of pions from other particles as it is sensitive to their secondary interactions.

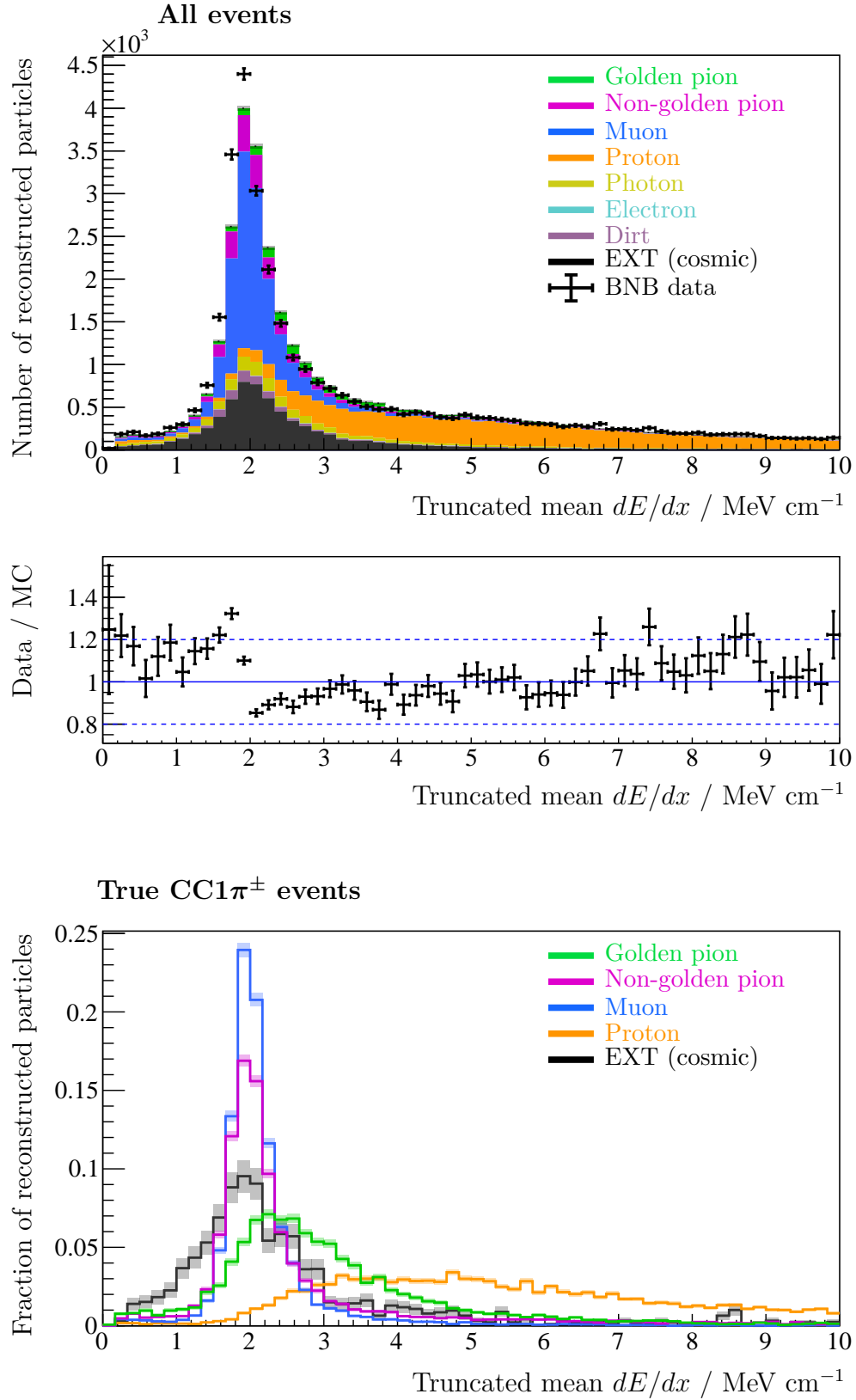
The **track-vs-shower score** is the outcome of the SVM that runs within Pandora and is trained on topological features to distinguish between tracks and showers. As shown in Figure 6.10, this feature provides strong separation of track-like particles from electromagnetic showers and additionally is sensitive to secondary interactions.

Finally, the **wiggleness**,  $W$ , is the standard deviation of the angular differences,  $\Delta_i$ , between the directions  $\underline{n}_i$  at adjacent points along the trajectory of the fitted track:

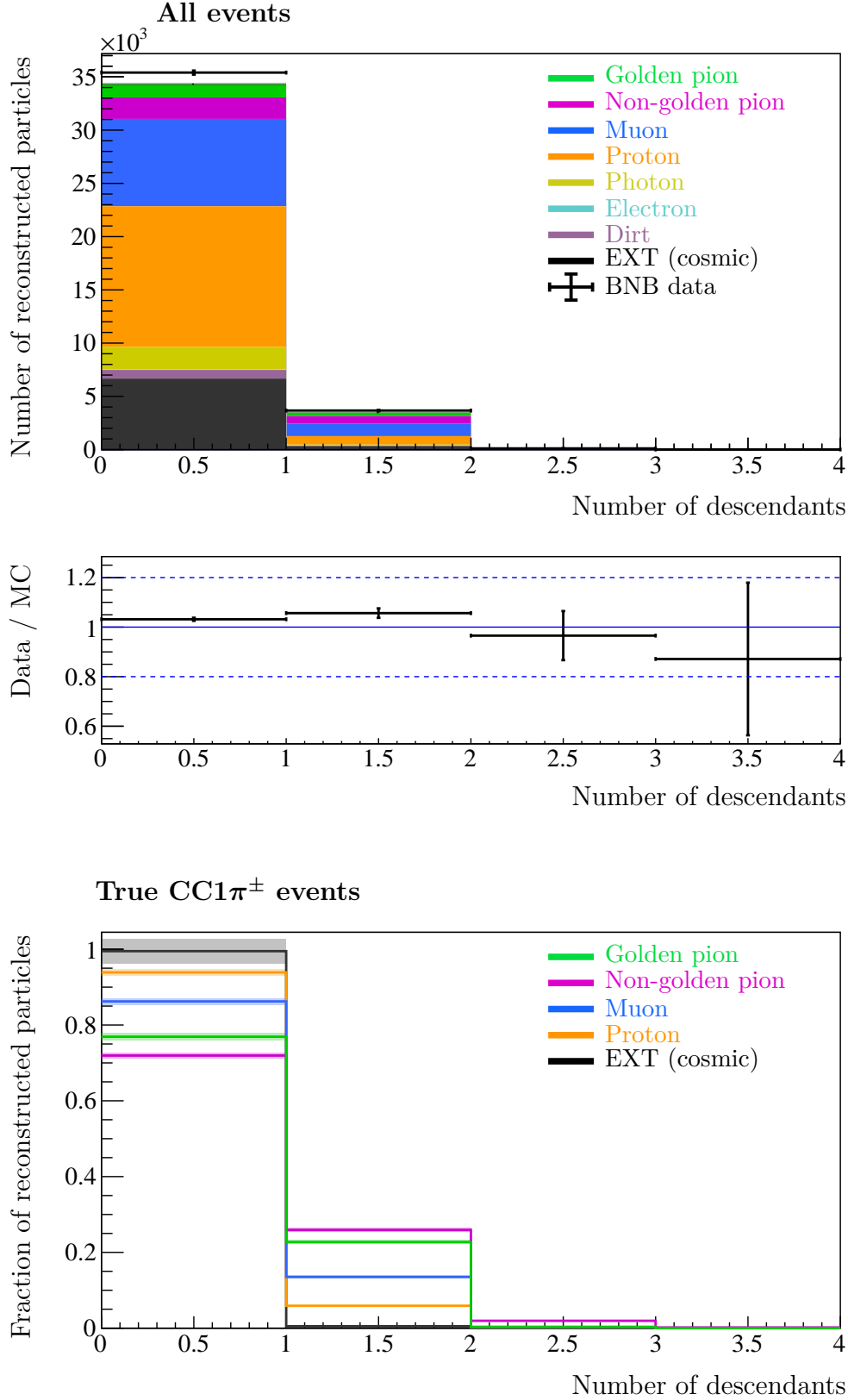
$$W = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \left( \Delta_i - \frac{1}{N} \sum_{j=1}^N \Delta_j \right)^2} \quad (6.3)$$

$$\text{where, } \Delta_i \equiv \cos^{-1}(\underline{n}_i \cdot \underline{n}_{i-1}).$$

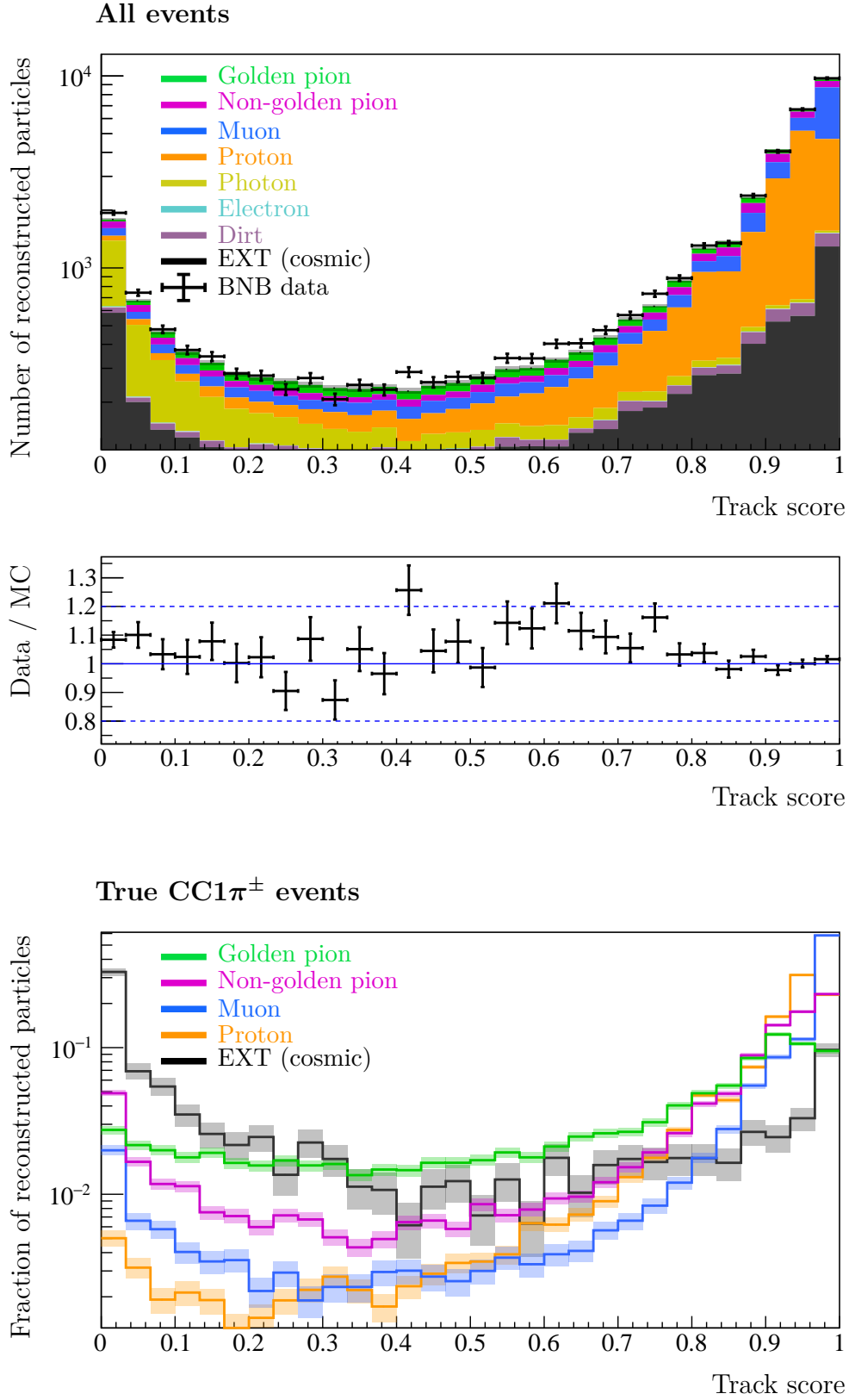
Again, this feature is sensitive to secondary interactions. As shown in Figure 6.11, the distribution of golden pions has two peaks. As illustrated in Figure 6.12, those in the high-wiggleness peak typically have a visible decay electron that is not identified as a secondary particle by the pattern recognition.



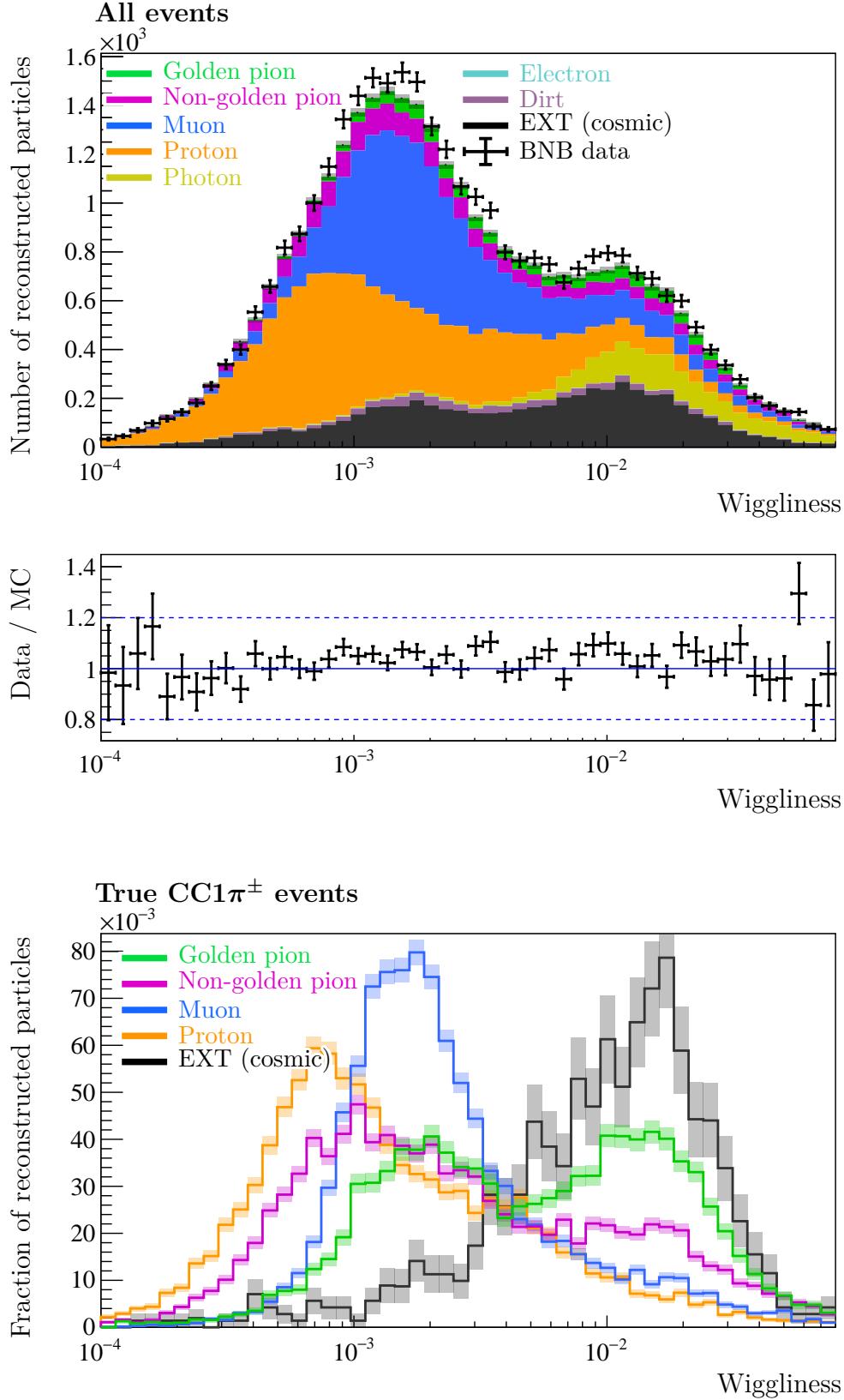
**Figure 6.8:** The 3-plane truncated mean  $dE/dx$  in the first third of the reconstructed particle track.



**Figure 6.9:** The number of descendant reconstructed particles (e.g. daughters, granddaughters) in the particle-flow hierarchy produced by Pandora.

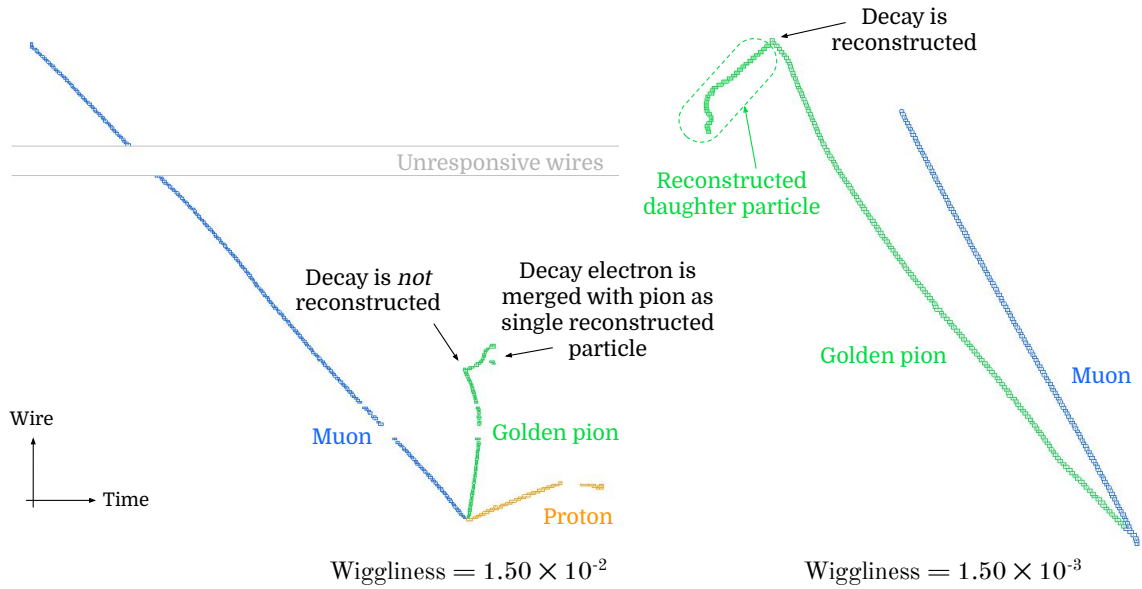


**Figure 6.10:** The response of the track-vs-shower SVM that runs in Pandora.



**Figure 6.11:** The standard deviation of the angular differences between the directions of adjacent points along the track trajectory.





**Figure 6.12:** Event displays showing two simulated  $CC1\pi^\pm$  events with a golden pion from the overlay sample. The value of the wiggleness is shown for the pion in each event. In the left display the pion is in the high-wiggleness golden pion peak shown in Figure 6.11, whereas the pion in the right display is in the low-wiggleness peak. In the left display the decay of the pion is not reconstructed and so the pion and its decay products are merged into a single reconstructed particle, which has high wiggleness. In the right display, the decay is correctly identified and the decay product is reconstructed as a daughter particle of the pion. In this case the reconstructed pion has low wiggleness. In this way, the wiggleness feature is sensitive to the existence of secondary interactions that are missed by the pattern recognition.

### 6.3.3 BDT training and responses

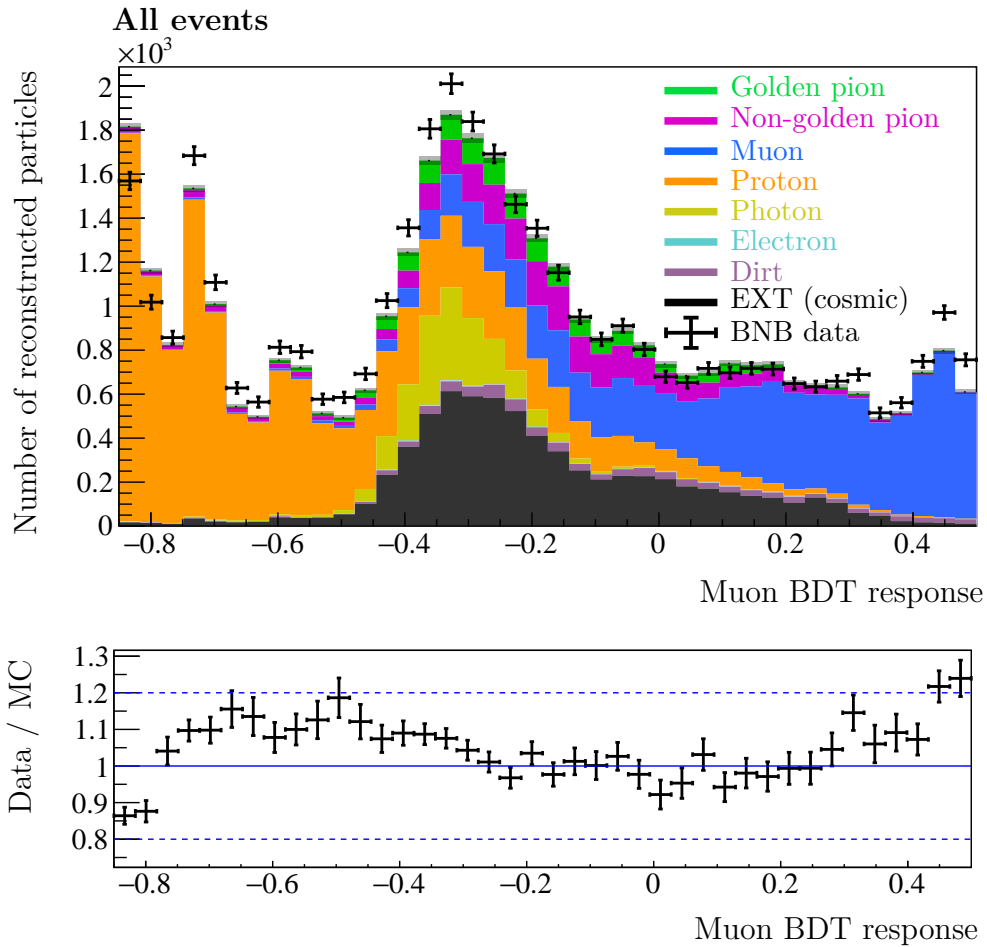
The features described in Section 6.3.2, were selected from a larger set of potential features that had the ability to distinguish between particle species. In order to determine which features to include in each BDT, an  $N - 1$  study was performed <sup>1</sup>. Each BDT was trained with a list of  $N$  features and re-trained with each feature removed in turn (such that  $N - 1$  features remain). The reduction in the performance of the BDT that is caused by removing a feature can be interpreted as a measure of the importance of the feature removed. Importantly, this measure accounts for any possible correlations between features. In the process of this study, if one or more of the features were found to have a sufficiently small impact on the performance of the BDT (within the statistical uncertainty of the sample used), then the feature with the smallest impact was removed. This process was repeated for each BDT until every feature was found to make a significant impact. The remaining features for the muon, proton and golden-pion BDTs are those listed in Section 6.3.2 but the number of descendants is not used for the proton BDT.

In order to train the BDTs, reconstructed neutrino final-state particles from  $\text{CC}1\pi^\pm$  events that pass the CC-inclusive selection are used as training examples. Only contained particles which are at least 5 cm from all TPC faces are used. In total there are five categories of possible particle types: protons, muons, golden pions, non-golden pions, and cosmic rays. Each reconstructed particle is assigned one such label based on the type of true particle with which it shares most hits. The particles are also assigned a training weight based on the completeness of their truth-match. For cosmic rays from external data, a weight of unity is used. In this way, well-reconstructed particles are preferentially treated while training, but the BDT still sees poorer cases of reconstruction to aid robustness. This method was found to perform better than applying a uniform weight for all training examples, or simply not training on particles with low completeness. The full set of simulated  $\text{CC}1\pi^\pm$  events was randomly split into even training and testing datasets and the consistency between the BDT response to these datasets was checked in every case to ensure against overtraining.

---

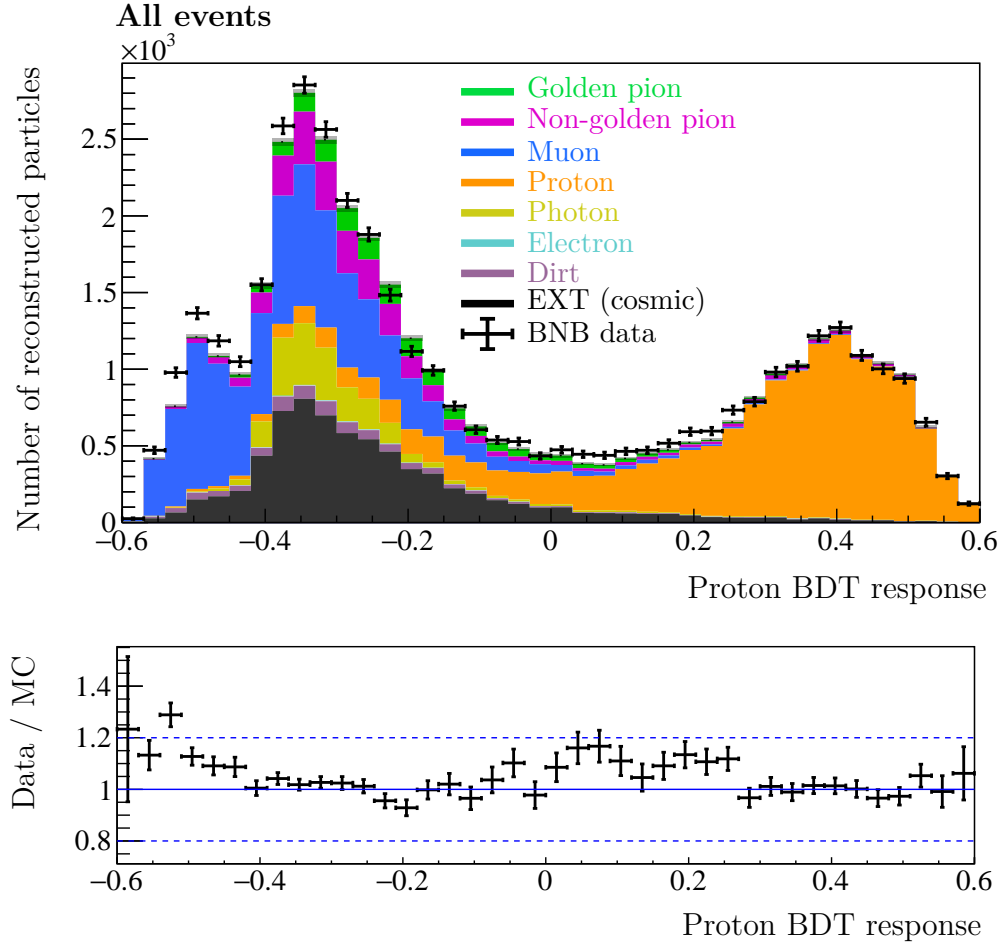
<sup>1</sup>The technical details of this study are not presented in this thesis but are given in [103].

Figure 6.13 shows the response of the muon BDT to all contained reconstructed particles passing the pre-selection. The most important features for separation between muons and protons are the proton-to-MIP likelihood ratio, and the truncated mean  $dE/dx$ . All features contribute toward the separation of muons and pions, but the track score is the most important.



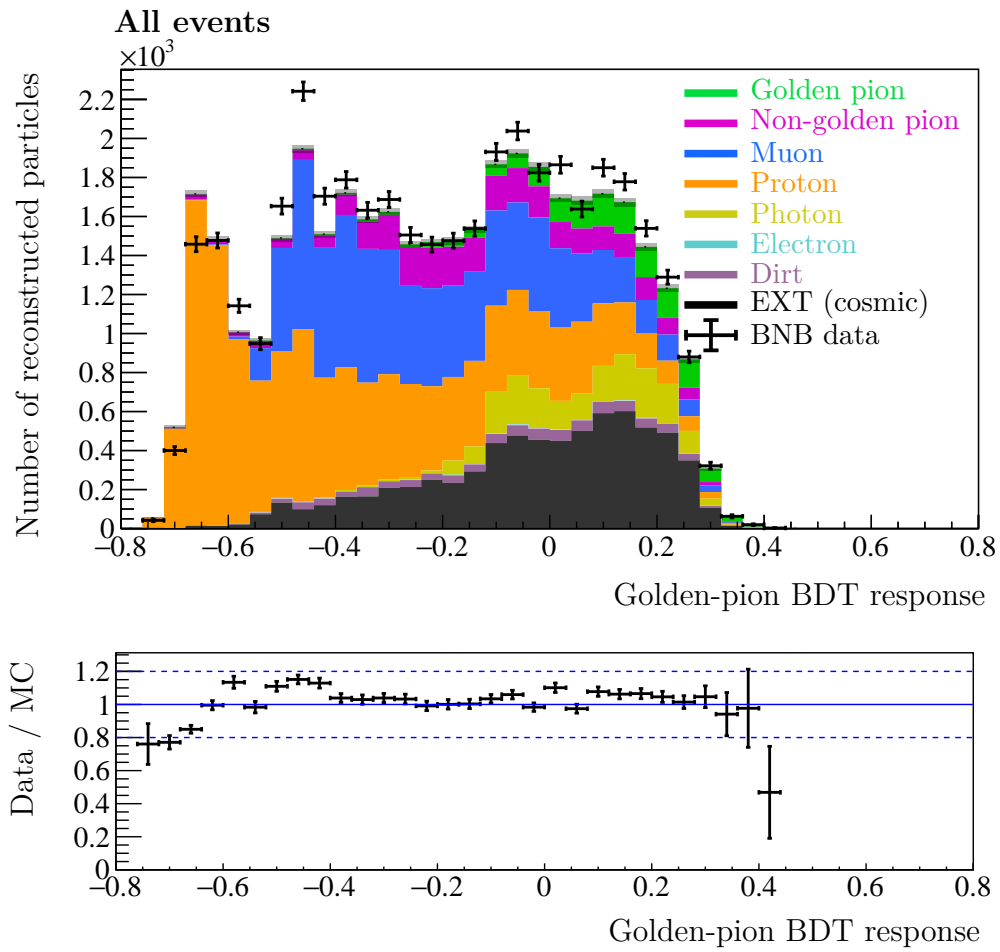
**Figure 6.13:** The response of the muon BDT for contained reconstructed particles in all events passing the CC-inclusive pre-selection. The structure in the distribution at low BDT-response is due to the inclusion of the discrete *number of descendants* feature.

Figure 6.14 shows the response of the proton BDT to all contained reconstructed particles passing the pre-selection. The majority of the separation power between protons and other particles is due to the proton-to-MIP likelihood ratio.



**Figure 6.14:** The response of the proton BDT for contained reconstructed particles in all events passing the CC-inclusive pre-selection.

Figure 6.15 shows the response of the golden pion BDT to all contained reconstructed particles passing the pre-selection. Again, the separation between golden pions and protons is driven by the proton-to-MIP likelihood ratio, and the track-score contributes most to the separation between golden pions and muons. The separation between golden and non-golden pions (for which this BDT is designed) has contributions from many features, of which the pion-to-MIP likelihood ratio is the most important for the reasons described in Section 6.3.2.



**Figure 6.15:** The response of the golden pion BDT for contained reconstructed particles in all events passing the CC-inclusive pre-selection.

## 6.4 Selection of $\text{CC1}\pi^\pm$ events

This section describes the criteria used to identify  $\text{CC1}\pi^\pm$  events, referred to as *selection cuts*. As described in Section 6.1, two event selections are employed. The *generic* selection is optimized to select  $\text{CC1}\pi^\pm$  events without any regard for the golden or non-golden nature of the pion. Events passing the generic selection are used in the extraction of the differential cross-sections in terms of  $p_\mu, \theta_\mu, \theta_\pi, \theta_{\mu\pi}$ , and the proton-exclusive cross-section measurement. The *golden* selection applies a further cut on top of the generic selection to enhance the fraction of selected events with golden pions and is used to extract the differential cross-section in terms of  $p_\pi$ .

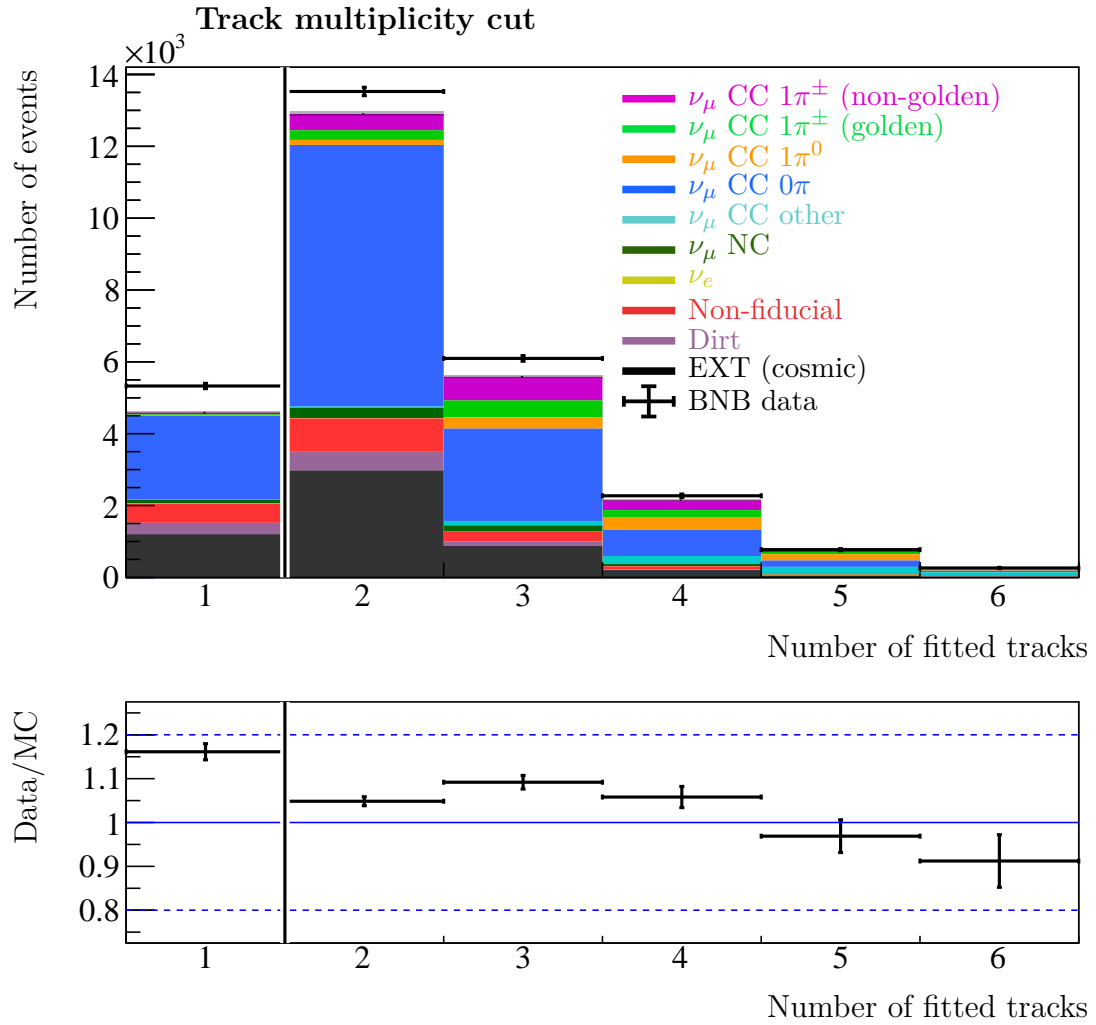
The selection cuts were chosen to maximise the product of the efficiency,  $\epsilon$ , and purity,  $\rho$ , of the selection which are defined as:

$$\begin{aligned} \text{Efficiency} = \epsilon &= \frac{\# \text{ selected signal events}}{\# \text{ generated signal events}}, \\ \text{Purity} = \rho &= \frac{\# \text{ selected signal events}}{\# \text{ selected events}}. \end{aligned} \tag{6.4}$$

For all figures in this section, only the statistical uncertainty on the data is shown.

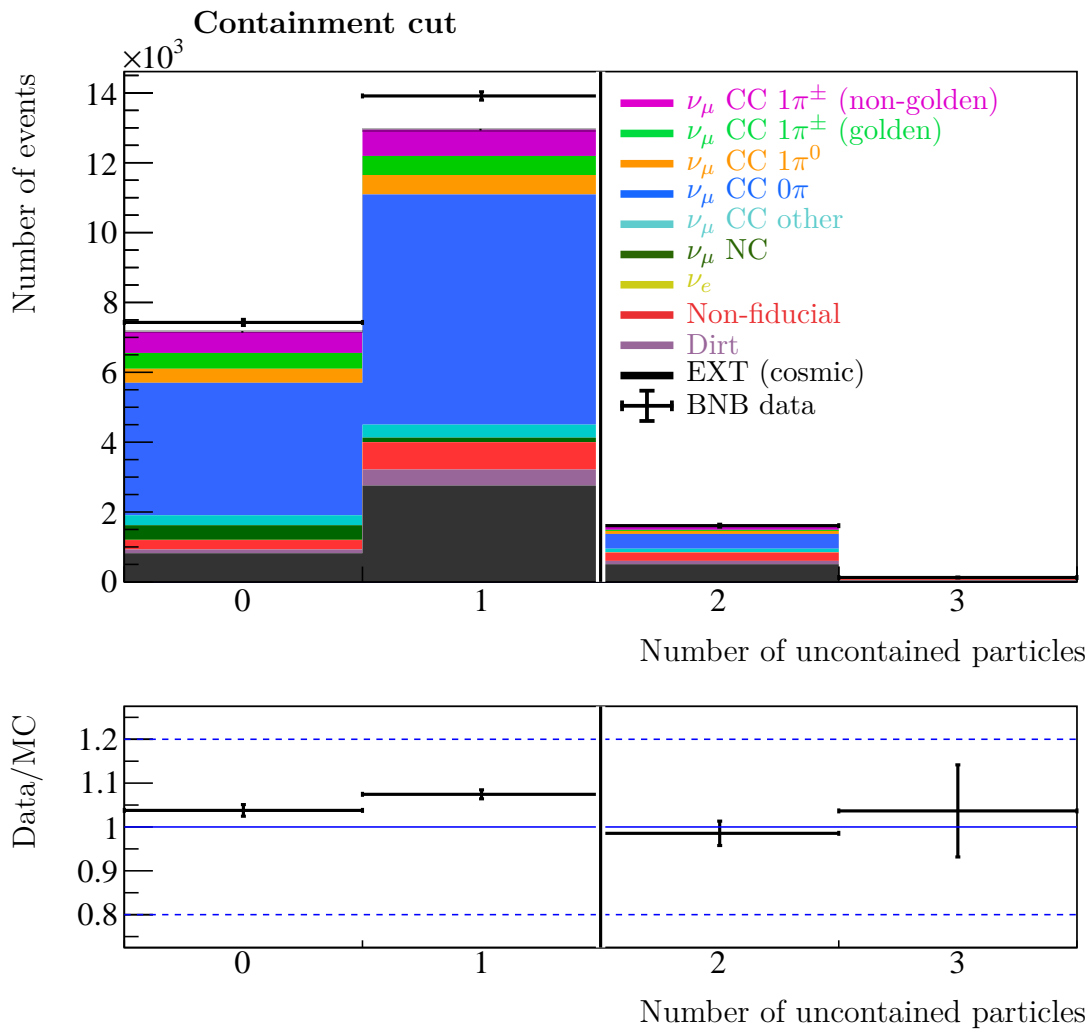
### 6.4.1 Generic selection

After the CC-inclusive pre-selection, the first cut applied is on the **track multiplicity**. A track fit is applied to all reconstructed neutrino final-state particles (including those that represent true electromagnetic showers) but can fail if the particle has too few associated space-points. A minimum of two tracks is required (for the muon and the pion). Figure 6.16 shows the number of tracks per event and is broken down by the true origin of the events.  $\text{CC1}\pi^\pm$  events with and without a golden pion are shown in green and pink respectively. The dominant backgrounds after the CC-inclusive selection are CC  $0\pi$  ( $1\mu + 0\pi + Xp$ ,  $X \geq 0$ ) neutrino interactions, and external cosmic-ray backgrounds — shown in blue and dark grey respectively.



**Figure 6.16:** The number of tracks that were successfully fitted to the reconstructed particles in each event. Events to the left of the black line are rejected by the track multiplicity cut.

Next, a **containment** cut rejects events in which more than one particle leaves the TPC. A particle is considered contained if its reconstructed (SCE-corrected) endpoints are at least 5 cm from all TPC faces. Figure 6.17 shows the number of uncontained particles per event. Following this cut, if a particle escapes, it is identified as the muon candidate (correct in 87.5% of  $\text{CC}1\pi^\pm$  events). However, if all particles are contained then the one with the largest muon BDT response is selected as the muon candidate (correct in 84.5% of  $\text{CC}1\pi^\pm$  events).



**Figure 6.17:** The number of reconstructed particles with a fitted track end-point closer than 5 cm to a TPC face. Events to the right of the black line are rejected by the containment cut.



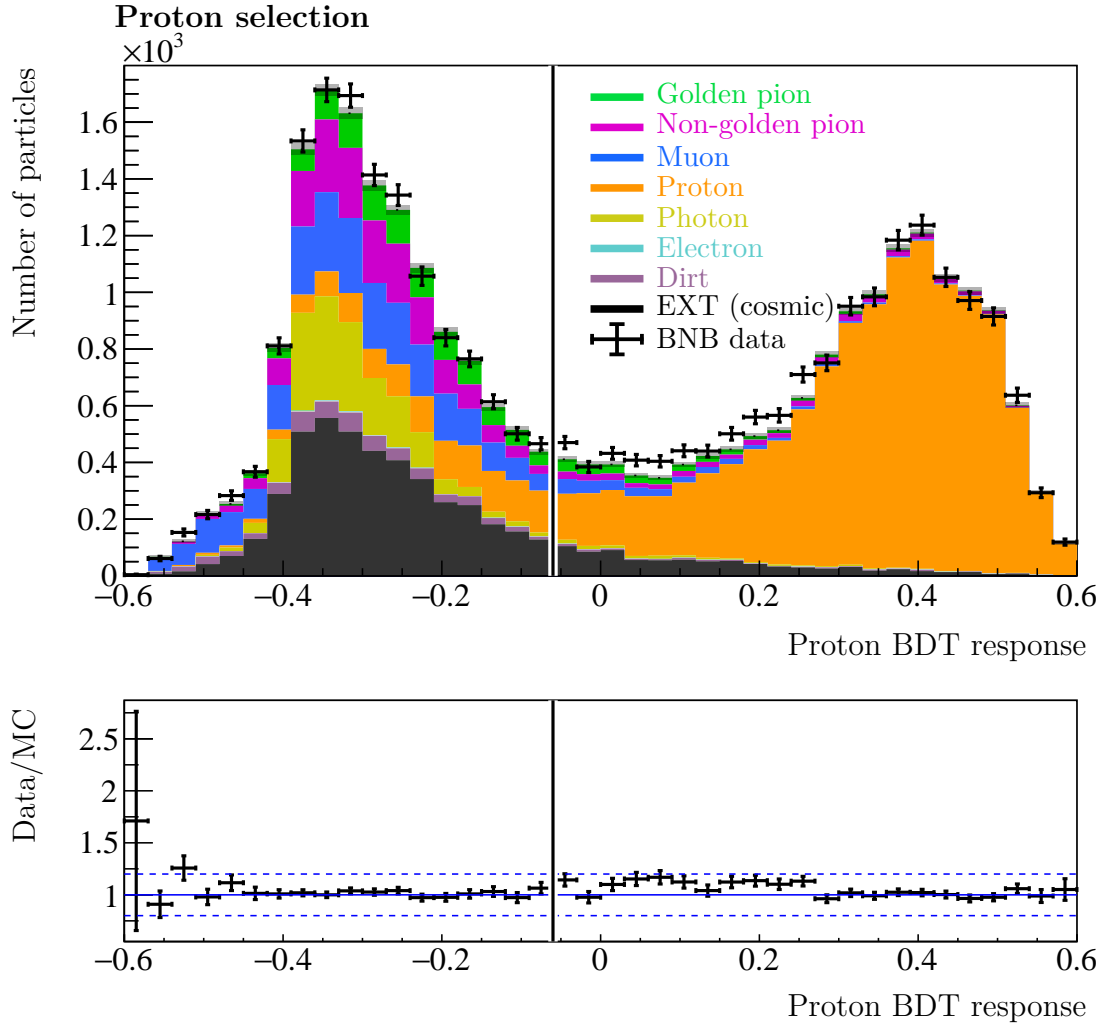
With the muon found, the next step is to identify protons using the proton BDT responses of the remaining reconstructed particles, shown in Figure 6.18. Any particles with a response greater than  $-0.06$  (or for which the BDT response is unavailable) are identified as protons<sup>2</sup>. Figure 6.19 shows the distribution of the number of particles per event that are not identified as a proton. As expected,  $\text{CC}1\pi^\pm$  events primarily have **two non-protons**: the muon (already identified), and the pion. Events with zero or greater than two non-protons are rejected. In the remaining events, with exactly two non-protons, the particle whose species has yet to be identified is taken as the pion candidate. The proton identification BDT cut was chosen to maximize the product of the  $\text{CC}1\pi^\pm$  selection efficiency and purity after the cut.

At this point in the selection, all reconstructed particles have been assigned a type: muon, proton or charged-pion, the accuracy of which is given in Section 6.5.2. The directions of the muon and pion are found using the track fit at the endpoint closest to the reconstructed vertex. The next two cuts are designed to reject events for which this directional information is not reliable.

As described in Chapter 4, the space-points associated with a reconstructed particle are produced by applying a 2D sliding linear fit to the positions of the particle's hits in each view separately. For a hit in a given view, the corresponding positions in the remaining views are found by querying the 2D sliding linear fit results at the  $x$  (time) coordinate of the hit, which is common to all views. These 2D positions should represent the projections of a 3D position that lies along the trajectory of the particle. If the 2D sliding linear fits could be applied in all three views, then the problem is over-constrained as four distinct parameters  $(x, z_U), (x, z_V), (x, z_W)$  are available to identify the three coordinates  $(X, Y, Z)$  of the space-point. An analytical minimisation procedure is used to find the 3D position that projects most closely to the desired 2D positions. However, if the particle lies in an unresponsive detector region in one view, the problem is only critically constrained. In this case, the positions of the space-points (and hence the direction of the muon and pion candidates) can become unreliable. The effect is compounded for particle trajectories near-parallel to the  $Y$ - $Z$  plane ( $\phi \approx \pm\pi/2$ ). In this case, all hits occur at or near the same  $x$  coordinate. This degeneracy means that multiple  $z_P$  positions (wire-coordinate,  $P \in U, V, W$ ) correspond to a single  $x$  position, and so it is problematic to match between views.

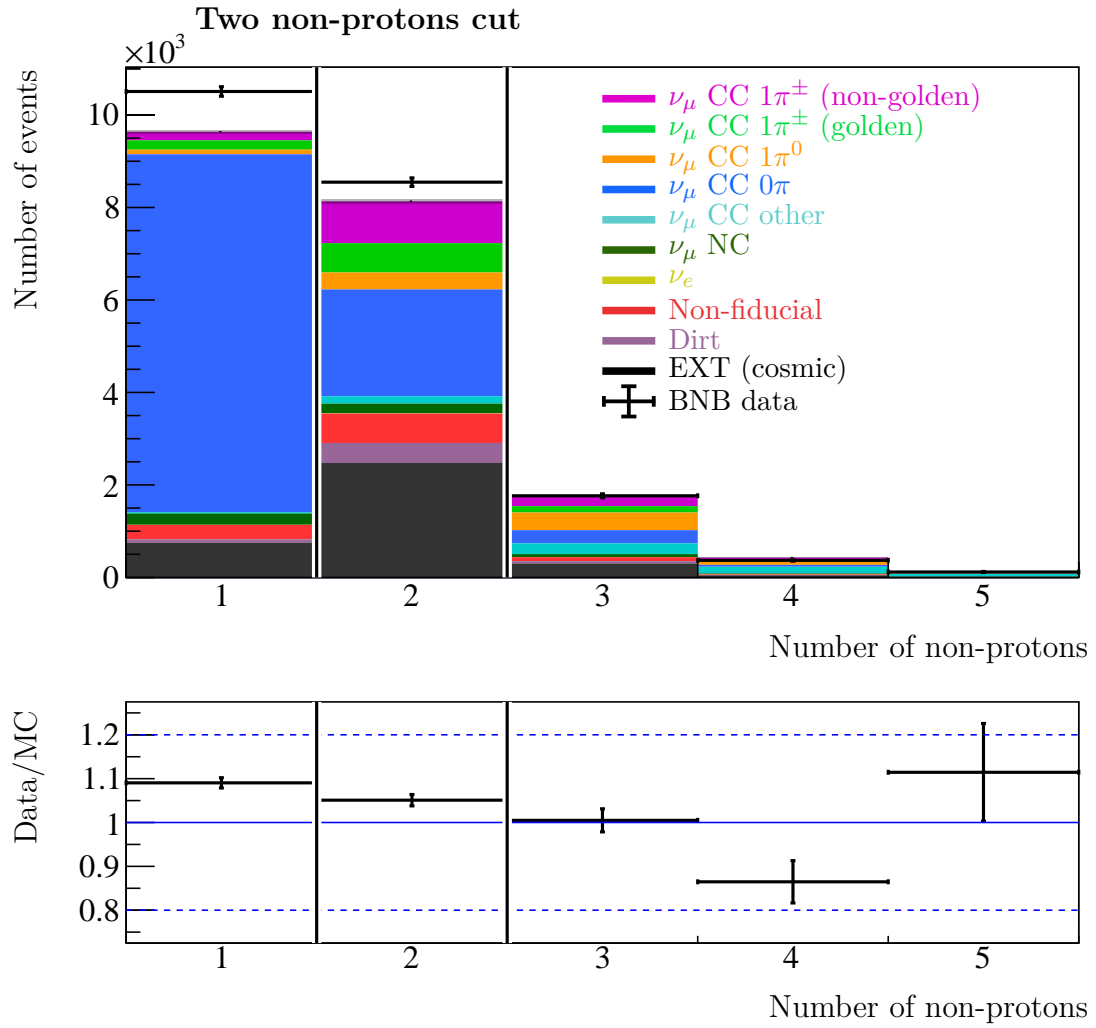
---

<sup>2</sup>The BDT response is unavailable when a reconstructed particle has too few 3D space-points for a track fit to be possible or for the calorimetric features to be calculated - this primarily occurs for low-energy protons. Such particles can not be considered by the BDTs but are included when counting proton multiplicity.

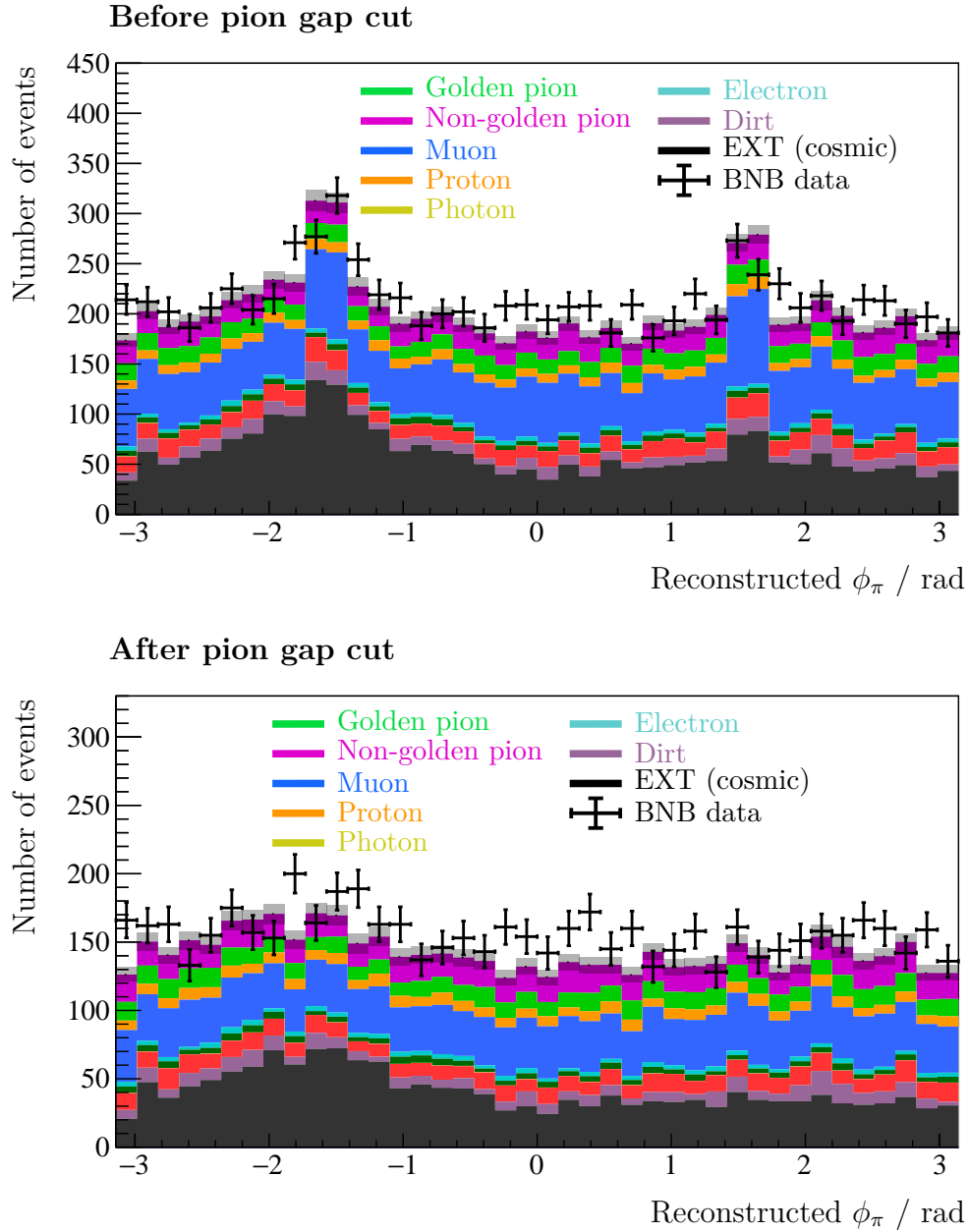


**Figure 6.18:** The proton BDT response for all contained reconstructed particles that have not already been identified as the muon candidate. Particles to the right of the black line are identified as proton candidates.

To ensure good direction reconstruction, events are only selected if at least one hit is associated with the muon and pion in all three readout planes. Figure 6.20 shows the reconstructed pion azimuth,  $\phi_\pi$  (defined in Section 2.4.2), before and after the application of the **pion gap** cut. The efficiencies of these cuts are flat over the true muon and pion directions.



**Figure 6.19:** The number of reconstructed particles that have not been identified as a proton. Events outside of the black lines are rejected by the “two non-protons” cut.



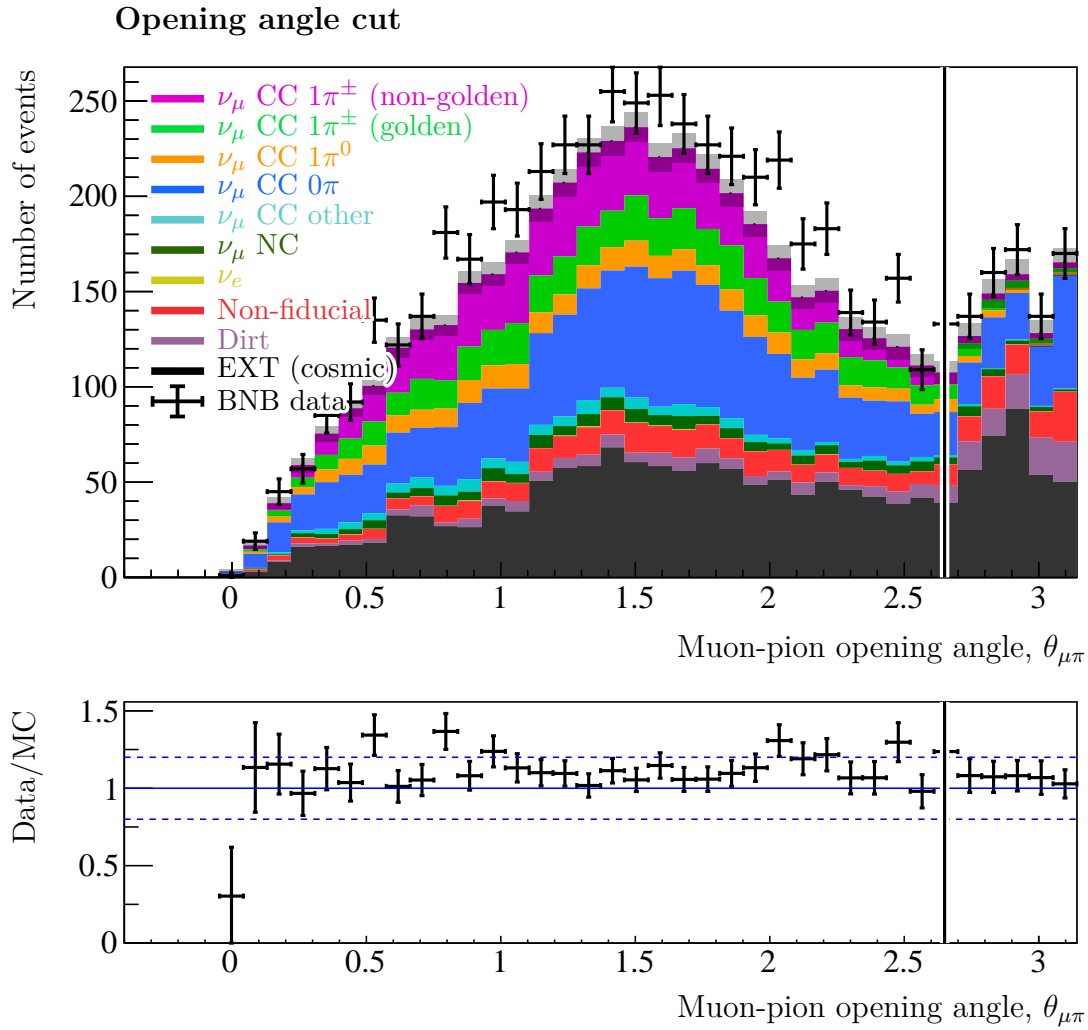
**Figure 6.20:** The azimuthal angle of the pion-candidate reconstructed particle before (top) and after (bottom) the pion gap cut. The excess of events around  $\phi_\pi = \pm\pi/2$  (before the cut) are due to reconstruction failures when the pion is in an unresponsive detector region and are removed by the cut.

After the muon and pion candidates are selected, the major neutrino-induced backgrounds are events with  $1\mu + 0\pi + Xp$  (where  $X \geq 0$ ). In the case where  $X = 0$ , (or  $X > 0$  but the protons have low energy and so produce few hits), the single-muon topology can be mistaken by the reconstruction as two back-to-back particles emerging from a common vertex, especially if the muon undergoes a scatter. This failure mode is also present for cosmic-ray muons. When this occurs, the true muon is reconstructed as two separate particles which are typically identified as the muon and pion candidates by this selection. For events with particle multiplicities greater than one (such as  $\text{CC1}\pi^\pm$ ), this failure mode is significantly less prevalent as the position of the neutrino interaction vertex is more readily visible. Consequently, events in which the reconstructed muon-pion opening angle is large are background-dominated, as is shown in Figure 6.21.

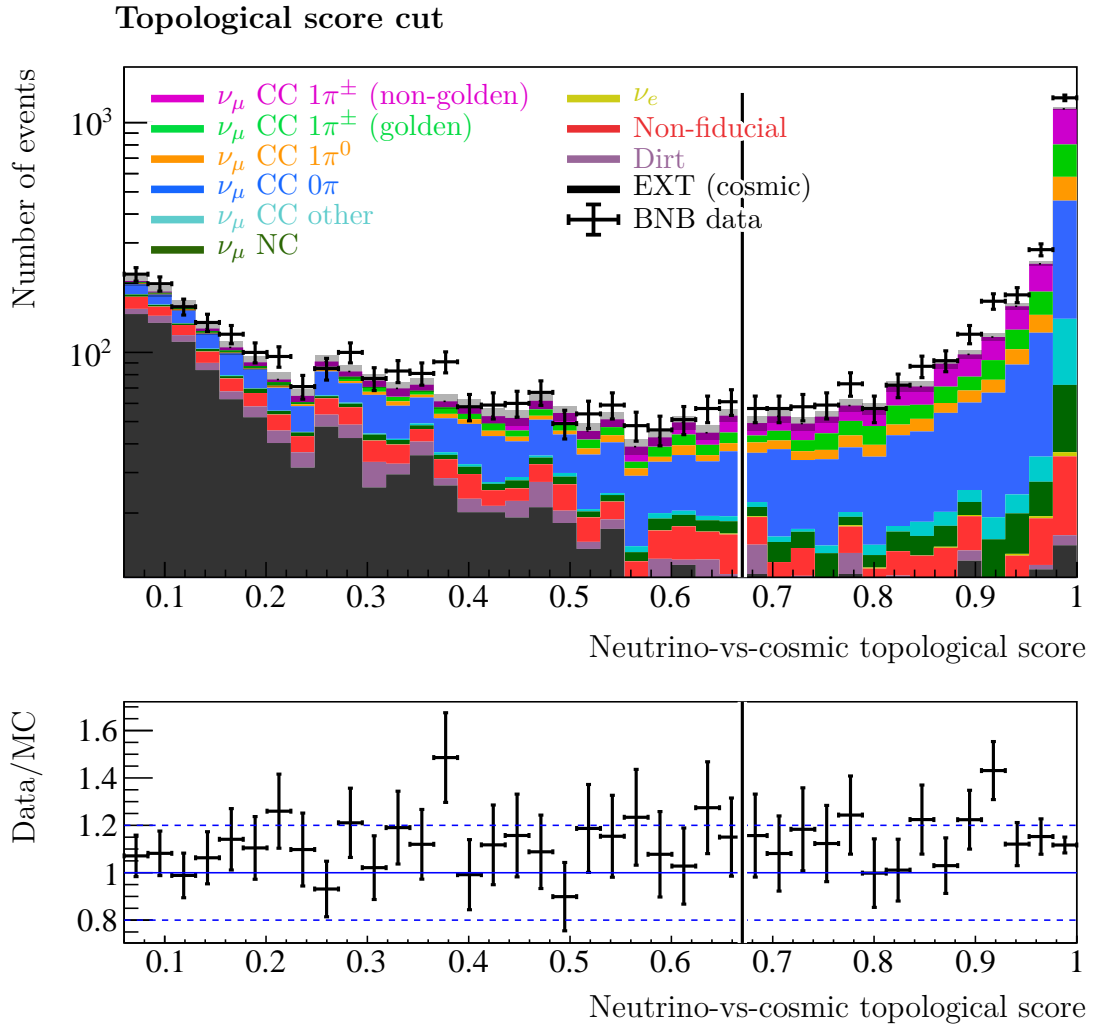
In order to increase the purity of the selection, an **opening angle** cut is applied and events with  $\theta_{\mu\pi} > 2.65$  rad are rejected (chosen to maximise the product of efficiency and purity). In Chapter 7 the phase-space of  $\text{CC1}\pi^\pm$  events that are considered as “signal” is restricted in accordance with this cut. However, this restriction is not made in this chapter.

The next cut uses the **topological score** for neutrino identification, discussed in Chapter 5 and shown in Figure 6.22. The cut value of 0.67 is chosen to maximise the product of efficiency and purity. This cut removes many of the remaining cosmic-ray backgrounds.

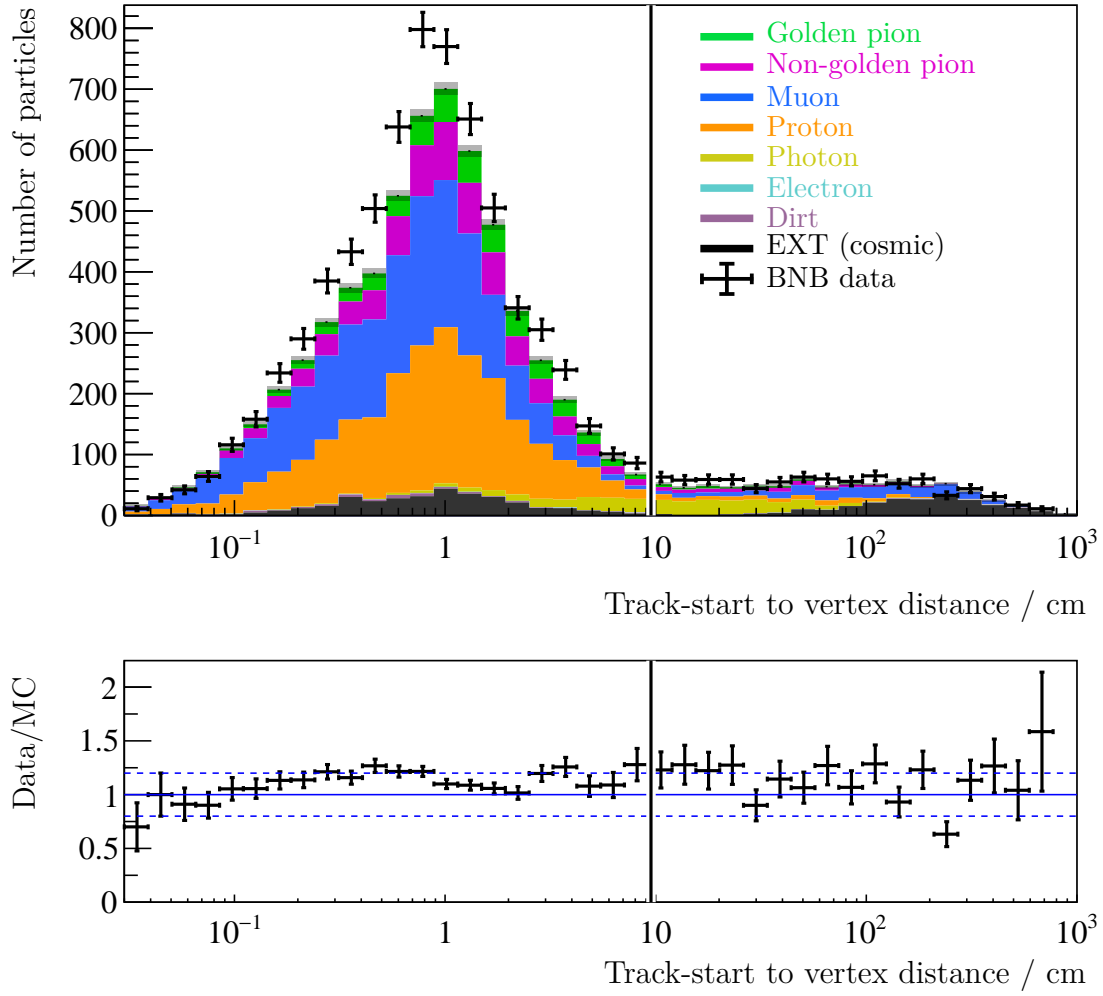
Finally, all reconstructed particles are required to start within 9.5 cm of the reconstructed neutrino interaction vertex. Figure 6.23 shows the “track-start to vertex distance” for all reconstructed particles in events passing the cuts up to this point in the selection. As expected, photons and external backgrounds are most commonly separated by a significant distance from the vertex. Figure 6.24 shows the furthest track-start to vertex distance of all reconstructed particles in each event at this point in the selection, broken down by the true origin of the event. The **start near vertex** cut primarily removes  $1\mu + 0\pi + Xp$  and  $1\mu + 1\pi^0 + Xp$  backgrounds, and was chosen to maximise the product of efficiency and purity.



**Figure 6.21:** The opening angle between the muon and pion candidate reconstructed particles. Events to the right of the black line are rejected by the opening angle cut.

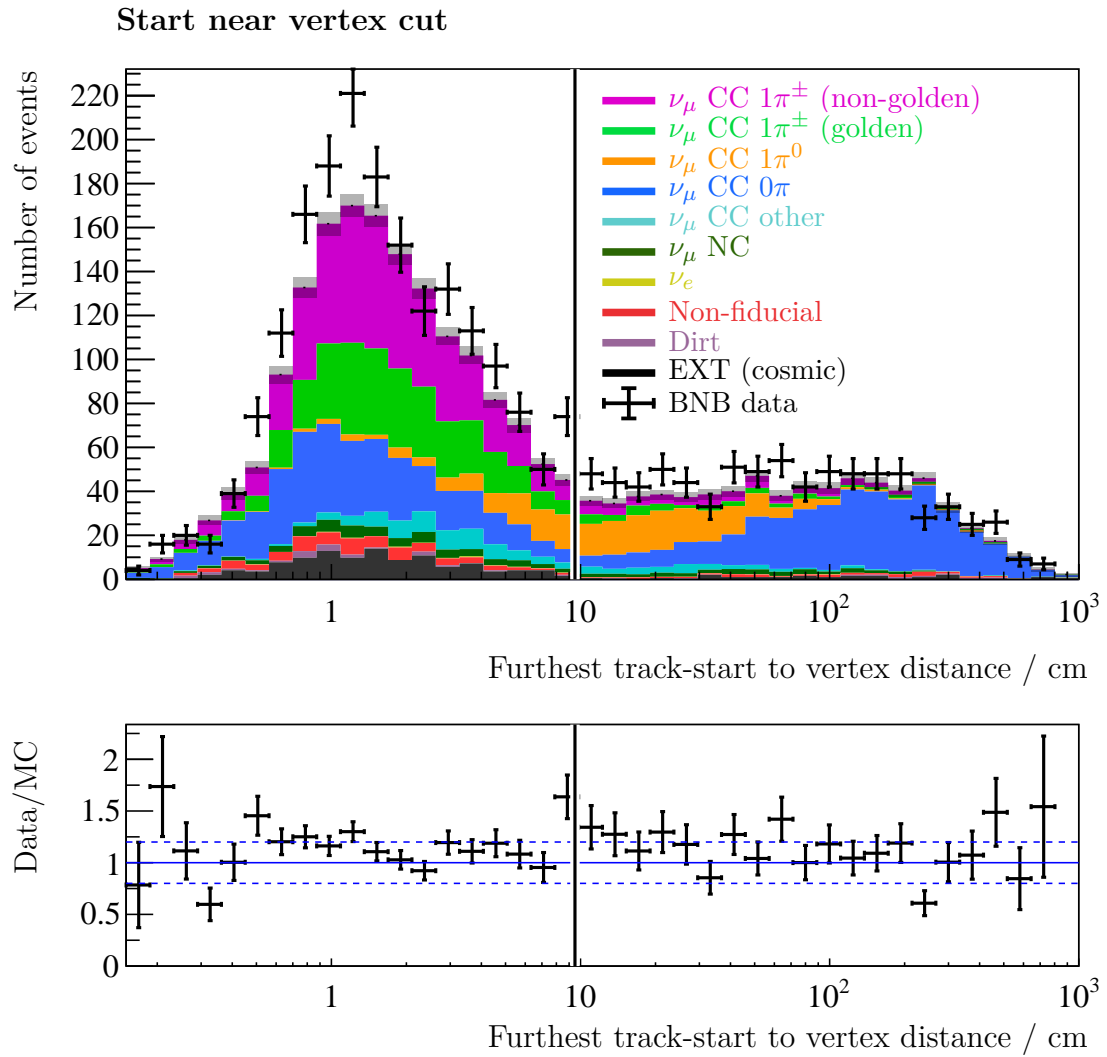


**Figure 6.22:** The response of the topology-based neutrino identification SVM that runs within Pandora. Events to the left of the black line are rejected by the topological score cut. Here, no events have a topological score below 0.06 as these have already been removed by the CC-inclusive selection.



**Figure 6.23:** The distance from the reconstructed vertex to the start-point of each reconstructed particle track. Events with one or more particles to the right of the black line are rejected by the “start near vertex” cut.

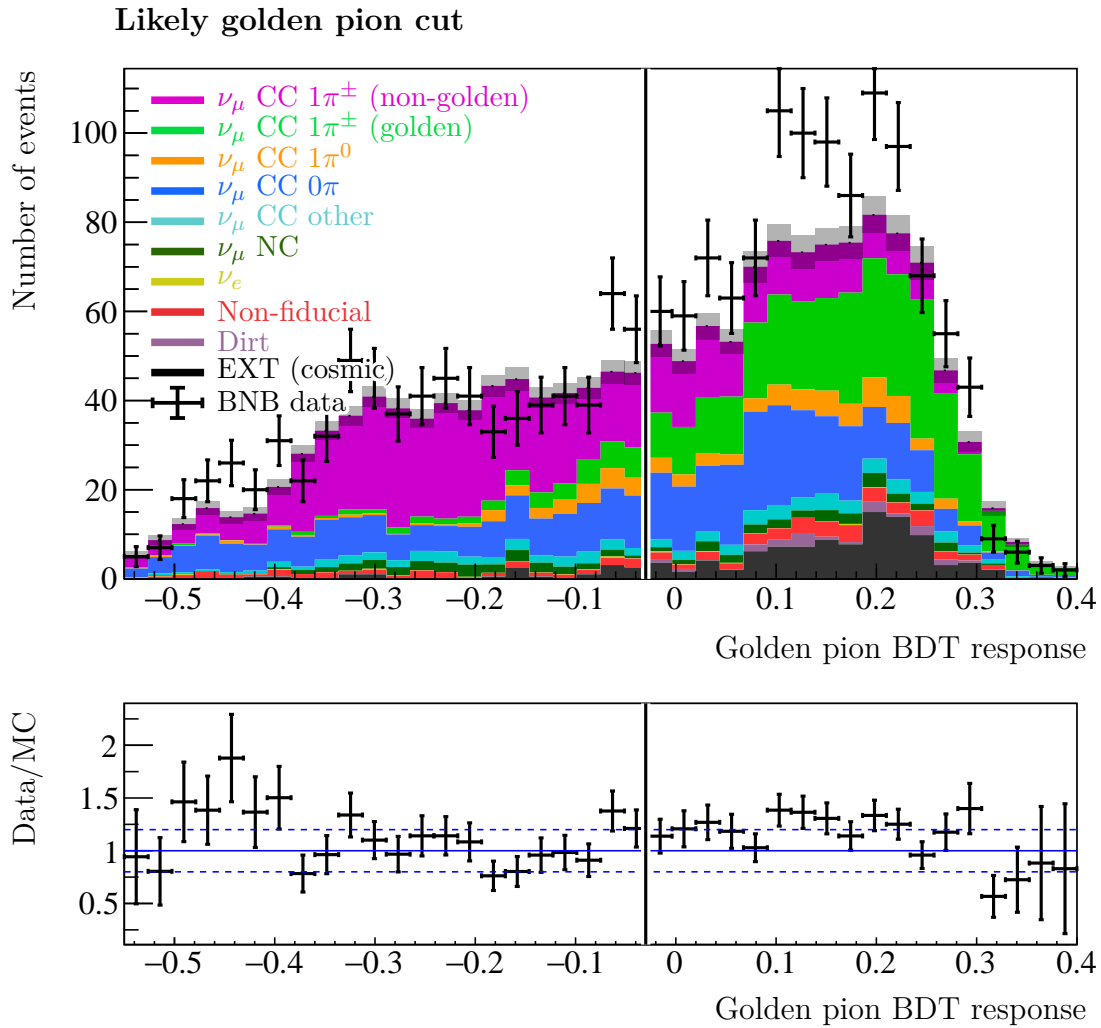




**Figure 6.24:** The distance from the reconstructed vertex to the start-point of the furthest reconstructed particle track. Events to the right of the black line are rejected by the “start near vertex” cut.

### 6.4.2 Golden selection

The golden selection begins with the generic selection and includes an extra cut on the golden-pion BDT response of the pion candidate at  $-0.03$ , shown in Figure 6.25. This cut enhances the fraction of selected events that truly contain a golden pion.



**Figure 6.25:** The response of the golden pion BDT for the pion-candidate reconstructed particle. Events to the left of the black line are rejected by the likely golden pion cut.

## 6.5 Performance

### 6.5.1 Event selection performance

Table 6.2 shows the impact of each of the selection cuts on the efficiency and purity and Table 6.3 shows the breakdown of events passing the generic selection. In the former table the golden-fraction,  $f_G$ , is defined as:

$$f_G = \frac{\# \text{ selected signal events with golden pion}}{\# \text{ selected signal events}}. \quad (6.5)$$

Cut	$\epsilon$	$\rho$	$\epsilon \times \rho$	Signal	$f_G$	Data	Data/MC
CC-inclusive	54.0%	9.8%	0.053	2601.8	40.6%	28417	1.07
<b>Generic selection</b>							
Track multiplicity	52.7%	11.6%	0.061	2539.5	40.6%	23089	1.06
Containment	48.8%	11.7%	0.057	2354.2	42.4%	21341	1.06
Two non-protons	31.9%	18.9%	0.060	1539.4	41.3%	8548	1.05
Pion not in gap	25.6%	21.6%	0.055	1232.0	40.0%	6236	1.10
Muon not in gap	24.3%	22.6%	0.055	1171.9	40.3%	5776	1.12
Opening angle	23.4%	25.7%	0.060	1127.5	39.8%	4923	1.12
Topological score	19.5%	40.6%	0.079	941.0	39.5%	2643	1.14
Start near vertex	17.5%	52.4%	0.092	844.7	39.7%	1857	1.15
<b>Golden selection</b>							
Likely golden pion	9.3%	48.6%	0.045	449.9	65.5%	1109	1.20

**Table 6.2:** Breakdown of the efficiency,  $\epsilon$  and purity,  $\rho$ , of the  $\text{CC}1\pi^\pm$  event selection cuts. The “signal” column shows the number of selected signal events in the overlay sample when scaled to the POT of the BNB data used. The “data” column gives the total number of BNB data events selected. The fraction of selected signal events that contain a golden pion,  $f_G$ , is also shown. Finally, the “data/MC” column gives the ratio of the total number of selected BNB data events to the total number of selected simulated events including both signal and backgrounds ( $\text{MC} = \text{Signal}/\rho$ ).

Sample	Events	Fraction	Fraction removed
<u>Signal (simulation)</u>			
Overlay: $1\mu + 1\pi^\pm + Xp$ (non-golden)	509.6	31.6%	67.0%
$1\mu + 1\pi^\pm + Xp$ (golden)	335.2	20.8%	68.3%
<u>Backgrounds (simulation)</u>			
Overlay: $1\mu + Xp$	370.2	23.0%	97.2%
$1\mu + 1\pi^0 + Xp$	82.0	5.1%	92.4%
Non-fiducial	58.7	3.6%	96.9%
$1\mu + 2\pi^\pm + Xp$	50.9	3.2%	84.9%
$1\pi^\pm + Xp$	25.1	1.6%	92.5%
$2\pi^\pm + Xp$	19.4	1.2%	73.8%
$1\mu + 1\pi^0 + 1\pi^\pm + Xp$	11.9	0.7%	94.8%
$Xp$	3.7	0.2%	97.8%
Other	19.5	1.2%	94.6%
Dirt	15.6	1.0%	98.4%
EXT data	109.6	6.8%	98.0%
Total signal	844.8	52.4%	67.5%
Total backgrounds	766.6	47.6%	96.8%
Total simulation	1611.4		93.9%
Total BNB data	1857		93.5%

**Table 6.3:** The composition of events that pass the  $\text{CC}1\pi^\pm$  generic selection. The overlay, dirt and EXT data have been normalised to the POT of the BNB data. The fraction column refers to the fraction of the selected events that are of the corresponding topology. The fraction removed column gives the fraction of events that pass the CC-inclusive selection that are removed by the  $\text{CC}1\pi^\pm$  selection.

Of the events passing the CC-inclusive selection, 32.5% of signal events pass the generic selection whereas only 3.2% of background events remain. Consequently, the signal-to-background ratio is increased by a factor of  $\sim 10$ . Importantly, the rate at which signal events pass the generic selection is largely independent of the kinematic parameters of interest. The main exceptions are listed below and the selection efficiency as a function of each parameter is shown in Appendix A.

- The selection efficiency reduces at muon and pion momentum below  $\sim 150 \text{ MeV } c^{-1}$ . This occurs as particles with lower momenta produce fewer hits and so are less likely to be reconstructable.
- The selection efficiency drops in the region  $\theta_{\mu\pi} > 2.65$  due to the opening angle cut. Additionally, the efficiency is reduced for small  $\theta_{\mu\pi}$  as the muon and pion begin to overlap and are likely to be merged in the reconstruction.
- The selection preferentially rejects muon angles that are close to the vertical. This effect is primarily introduced by the CR-tagging algorithm.

### 6.5.2 Particle identification performance

Table 6.4 and Table 6.5 show the performance of the particle identification for true  $\text{CC}1\pi^\pm$  events passing the generic and golden selections respectively. Reconstructed particles are labelled by the species of the simulated particle with which most hits are shared. Each reconstructed particle is matched to exactly one simulated particle (or is deemed as a cosmic ray from beam-off overlaid data). However, each simulated particle may have zero, one or many matched reconstructed particles. Unlike the matching procedure described in Section 4.7.1, no thresholds are applied to the purity or completeness of these matches. The true muons and pions are misidentified as each other in  $\sim 11\%$  and  $\sim 8\%$  of events passing the generic and golden selections respectively.

Best true match	Muon candidate	Pion candidate	Proton candidate
Muon	88.7%	10.9%	5.0%
Charged pion	10.6%	87.7%	11.3%
Proton	0.7%	1.4%	83.7%
EXT (cosmic)	0.0%	0.0%	0.1%

**Table 6.4:** Particle identification performance for true  $\text{CC1}\pi^\pm$  events passing the generic selection.

Best true match	Muon candidate	Pion candidate	Proton candidate
Muon	92.1%	7.8%	4.3%
Charged pion	7.6%	90.2%	13.0%
Proton	0.4%	2.0%	82.6%
EXT (cosmic)	0.0%	0.0%	0.2%

**Table 6.5:** Particle identification performance for true  $\text{CC1}\pi^\pm$  events passing the golden selection.

Table 6.6 shows the performance of the particle identification for all events (signal + background) passing the generic selection. The leading backgrounds are BNB neutrino interactions with the  $1\mu + Xp$  topology which primarily pass the selection when a proton is mistakenly identified as a charged pion.

Best true match	Muon candidate	Pion candidate	Proton candidate
Muon	74.1%	16.5%	5.4%
Charged pion	9.2%	52.8%	11.3%
Proton	8.2%	16.7%	77.6%
EXT (cosmic)	8.1%	8.5%	4.0%
Other	0.5%	5.5%	1.8%

**Table 6.6:** Particle identification performance for all events passing the generic selection. The other category includes all other particle types, but is primarily photons from the decay of neutral pions.

## 6.6 Estimation of particle momenta

### 6.6.1 Overview of energy estimation methods

In order to extract the desired differential cross-sections, this analysis needs to be able to measure the momentum of muons and pions. The simplest approach is to estimate the momentum using the range of the particle. Consider a particle with some initial kinetic energy  $T(x=0) = T_0$ , that loses energy according to  $dT/dx(x)$ . Given that the particle stops (reaches  $T(R) = 0$ ) after a distance  $R$ , one can determine  $T_0$  by integration:

$$T_0 = \int_0^R \frac{dT}{dx}(x) dx. \quad (6.6)$$

The initial kinetic energy can then be converted to momentum via  $p = \sqrt{T_0^2 + 2mT_0}$ . In this analysis, this integral is not computed analytically. Instead, an empirical fit is applied directly to the true momenta of simulated muons and pions, as a function of the range of their corresponding reconstructed particle. The details of this process are given in Section 6.6.2.

It is important to note that this range-based method is only applicable for stopping particles and so is not appropriate for particles that escape the detector. Another approach (detailed in [104]) is to use the energy dependence of Multiple Coulomb Scattering (MCS) to estimate the momentum of the particle. This MCS-based method applies to both contained and escaping particles but has only been so far implemented in MicroBooNE for muons. This method is discussed in Section 6.6.3.

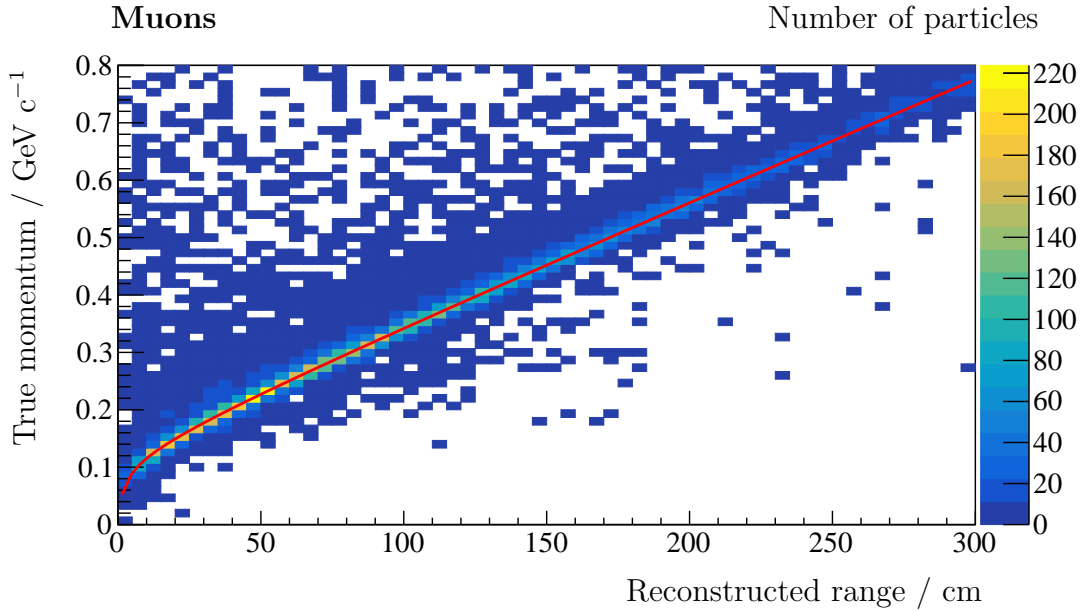
### 6.6.2 Momentum estimation by range

Figure 6.26 shows the aforementioned momentum-range curve for simulated muons from true  $\text{CC}1\pi^\pm$  events. The range is calculated from the track fitted to the reconstructed particle that matches to the muon. Only reconstructed particles that contain at least half of the hits of the simulated particle are used (i.e. truth-match completeness  $>50\%$ ). Additionally, the reconstructed particle must be contained (at least 5 cm from all TPC faces). The range is calculated as the sum of the SCE corrected distances between all

consecutive trajectory points from the track. A function with form,

$$p(R) = a + bR - cR^{-d} \quad (6.7)$$

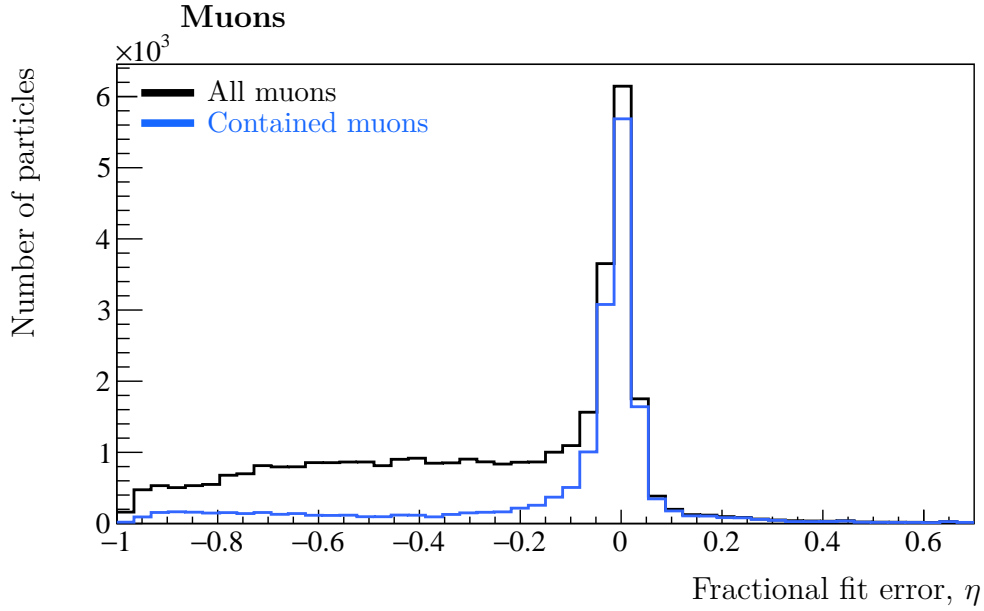
is fitted by allowing the positive-definite parameters  $a, b, c$ , and  $d$  to vary. This functional form is empirical and was chosen for its simplicity and its ability to capture the shape of the distribution (but was not derived from the underlying energy-loss process). Qualitatively, it does possess the desirable property that as  $R \rightarrow \infty$ ,  $p(R) \rightarrow a + bR$ . This linear dependence at high  $R$  is driven by the rate of energy loss of MIPs, for which  $\langle dE/dx(x) \rangle = \text{const.}$  The function was fitted to the most probable momentum in each range bin and the result is the red line in Figure 6.26.



**Figure 6.26:** The true momentum of simulated muons against the range of their best matched reconstructed particle. The fitted line is shown in red.

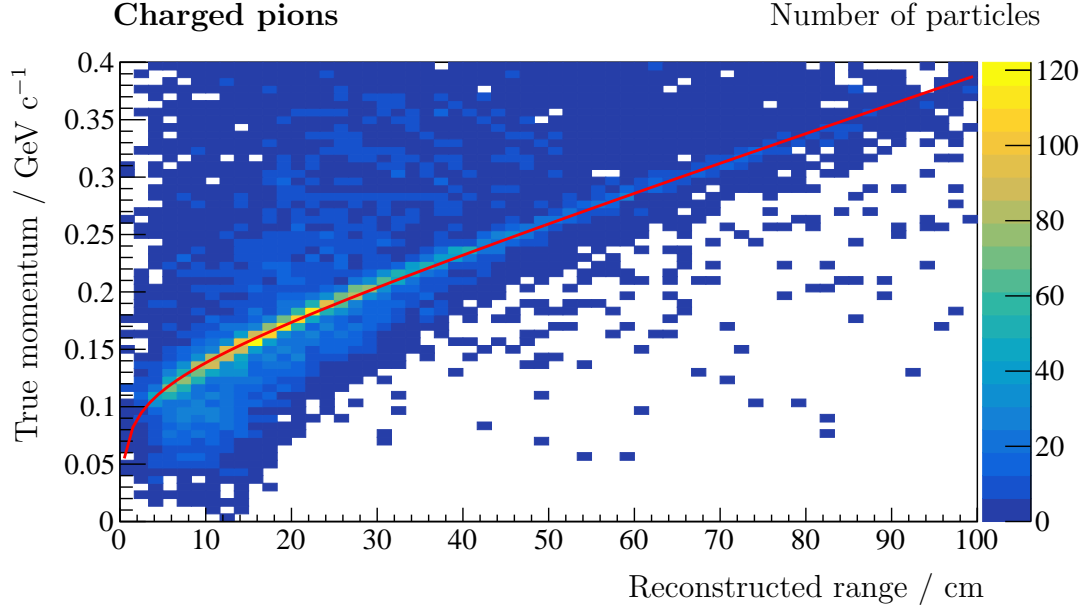
Figure 6.27 shows the fractional fit error,  $\eta \equiv (p_{\text{reco}} - p_{\text{true}})/p_{\text{true}}$  for contained and escaping reconstructed muons with no restriction applied to the completeness of the truth-match. This illustrates explicitly that the range-based method is only applicable for contained muons. The contained reconstructed particles in the tail of the distribution are primarily low-completeness matches.



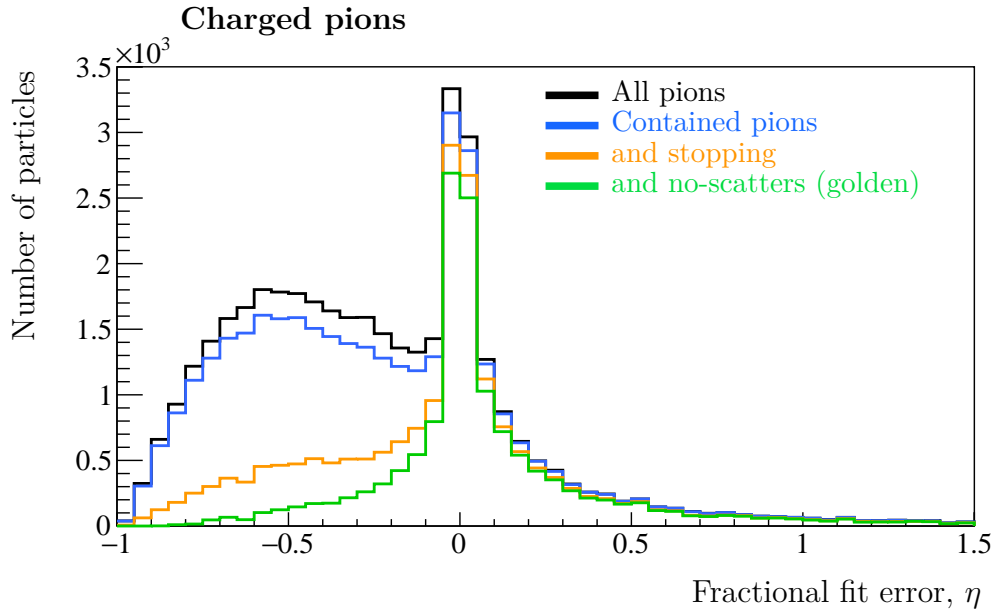


**Figure 6.27:** The fractional error,  $(\text{reconstructed} - \text{truth})/\text{truth}$ , of the range-based momentum of reconstructed particles which match to true muons in  $CC1\pi^\pm$  events. Shown in black are all such particles, and in blue are those that are contained within the detector boundaries.

This procedure was repeated for pions. Figure 6.28 shows the true momentum versus range distribution for reconstructed particles that match to charged pions, along with the fitted curve. As for the muons, only contained reconstructed particles with a truth-match completeness of  $>50\%$  were used for the fit. Figure 6.29 shows  $\eta$  for all pions (with no restriction on the completeness of the truth-match) and illustrates the successive impact of each of the restrictions made to define a golden pion.



**Figure 6.28:** The true momentum of simulated pions against the range of their best matched reconstructed particle. The fitted line is shown in red.



**Figure 6.29:** The fractional error,  $(\text{reconstructed} - \text{truth})/\text{truth}$ , of the range-based momentum of reconstructed particles which match to true pions in  $CC1\pi^\pm$  events. The line colours show the impact of each of the conditions which define a golden pion. Shown in black are all pions, in blue are those that are contained within the detector boundaries, in orange are those that also come to rest, and in green are those that also undergo no scatters. The conditions for the green line are the definition of a golden pion.

### 6.6.3 Muon momentum estimation by MCS

As illustrated by Figure 6.27, the range-based momentum estimation method is only applicable for contained muons. Consequently, for escaping muons, the MCS method is instead used. MCS is the process by which a charged particle traversing a medium (in this case argon) repeatedly scatters off atomic nuclei resulting in a trajectory described by a stochastic random walk. After travelling a distance,  $l$ , through the medium, the direction of the particle will in general differ from its initial direction, as described by a scattering angle. This angle can be modelled as a random variable, distributed according to a Gaussian function with a mean of zero, and standard deviation,  $\sigma_{MCS}$ , given by the Highland formula [104]:

$$\sigma_{MCS} = \frac{S_2 z}{p \beta c} \sqrt{\frac{l}{X_0}} \left[ 1 + \epsilon \ln \left( \frac{l}{X_0} \right) \right]. \quad (6.8)$$

Here,  $p$ ,  $\beta c$  and  $z$  are the momentum, velocity and absolute charge of the particle respectively and  $X_0$  is the radiation length of the target material (14 cm for LAr [104]). Finally,  $S_2$  and  $\epsilon$  are empirical parameters of the model. As  $\sigma_{MCS}$  is dependent on  $p$ , the momentum of a charged particle can be estimated by the distribution of the angles through which it scatters. In particular, MicroBooNE considers the angular deviations between successive segments of a reconstructed track, using a segment length of  $l = X_0$ . In this special case,  $\sigma_{MCS}$  simplifies to:

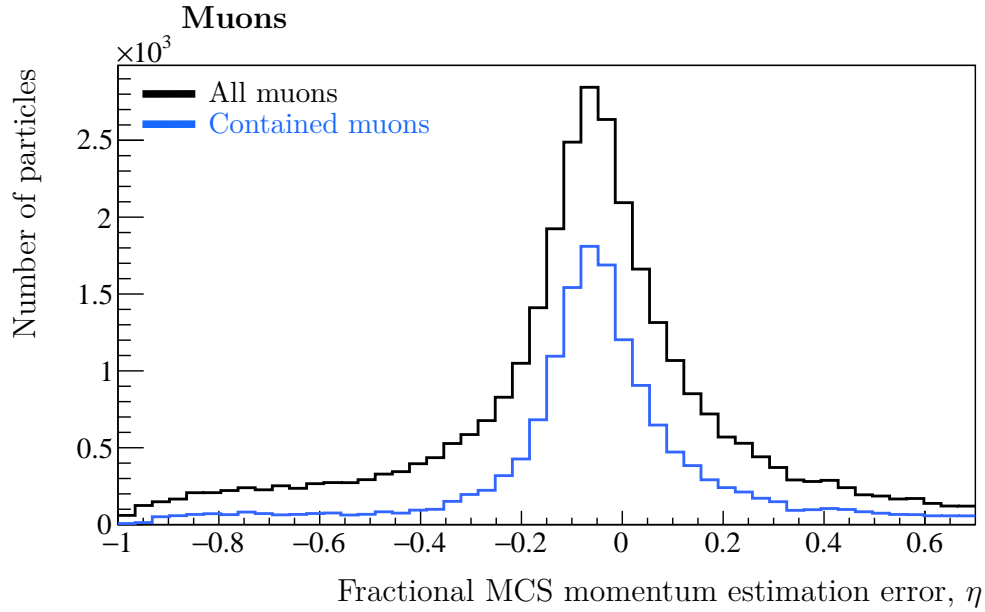
$$\sigma_{MCS}|_{l=X_0} = \frac{S_2 z}{p \beta c}. \quad (6.9)$$

Here, the only relevant model parameter is  $S_2$ . As the Highland formula was initially developed for relativistic particles ( $\beta \approx 1$ ), MicroBooNE has studied its applicability in the region  $\beta < 1$  (relevant for the experiment) using a sample of muons simulated with Geant4 [89]. To describe the simulated data, the constant  $S_2$  must be replaced by a factor,  $\kappa(p)$ , that is dependent on the particle's momentum in the  $p < 1 \text{ GeV c}^{-1}$  region and asymptotically approaches a constant value at high momentum (equivalent to  $S_2$  in the relativistic case) [104].

For an assumed initial particle momentum,  $p_0$ , MicroBooNE determines the likelihood that the measured angular deviations between successive reconstructed track segments would occur using the aforementioned modified Highland formula. In addition, this likelihood accounts for the expected energy loss due to ionisation between each track

segment by assuming the particle is a muon. In order to estimate the momentum of a reconstructed particle via MCS, the value of  $p_0$  that maximises this likelihood is found.

The performance of this method is shown in Figure 6.30, and typically results in a poorer resolution than the range-based method, but crucially is applicable to uncon-  
tained muons.



**Figure 6.30:** The fractional error,  $(\text{reconstructed} - \text{truth})/\text{truth}$ , of the MCS-based momentum estimation of reconstructed particles which match to true muons in  $CC1\pi^\pm$  events. Shown in black are all such particles, and in blue are those that are contained within the detector boundaries.

# Chapter 7

## CC1 $\pi^\pm$ cross-section measurement

### 7.1 Cross-section definition

This chapter begins by defining the flux-integrated forward-folded differential cross-section, as measured in this analysis, and details the motivation for these definitions. The methods by which statistical and systematic uncertainties are assessed are discussed in Section 7.2 and the sources of uncertainty are given in Section 7.3. Finally, the results of the cross-section extraction are presented in Section 7.4.

#### 7.1.1 Flux-integrated cross-section

The cross-section governs the probability that a neutrino will interact with a target particle. In the case of MicroBooNE, the targets are argon nuclei (as O<sub>2</sub> and N<sub>2</sub> impurities are below the 100 ppt and 2 ppm levels respectively [4]). The rate of neutrino interactions,  $r(E)$ , is given by:

$$r(E) = \phi(E) \times \sigma(E) \times T, \quad (7.1)$$

and depends on:

- The neutrino energy,  $E$ ,
- The flux (rate of neutrinos per unit area reaching the detector),  $\phi(E)$ ,
- The number of target particles,  $T$ , with which the neutrinos might interact,
- The cross-section,  $\sigma(E)$ , for the neutrino-target interactions.

In general, any estimation of  $E$  can only be derived from the final-state particles that exit the nucleus. Consequently, a naive estimation will typically underestimate the true value, as energy can be transferred to the nucleus during FSI and some subset of the exiting particles may not be visible to the detector. In principle, the extent to which this estimation is biased can be estimated and removed. However, such a correction would introduce a dependence on the models used to estimate the bias. In general, a measurement of the cross-section should be as independent of the models used in the simulation as is feasibly possible. In this analysis, to avoid such difficulties the flux-integrated cross-section,  $\sigma^{F.I.}$ , is instead calculated:

$$\begin{aligned}\sigma^{F.I.} &\equiv \frac{1}{\Phi} \int_0^\infty [\phi(E) \times \sigma(E)] dE, \\ \Phi &\equiv \int_0^\infty \phi(E) dE,\end{aligned}\tag{7.2}$$

where  $\Phi$  is the integrated flux. A measurement of  $\sigma^{F.I.}$  is typically simpler and less model-dependent than a measurement of  $\sigma(E)$ . However, these advantages come at the cost of a measurement that is dependent on the flux shape ( $\phi(E)/\Phi$ ). Consequently, two flux-integrated measurements (made with different neutrino beam fluxes) can not be compared directly. Instead, a prediction for the cross section (typically derived from a generator as a function of the neutrino energy) can be averaged over the fluxes of each experiment (using Equation 7.2) and compared to each measurement to see if both are mutually consistent (or inconsistent) with the prediction. For the sake of notational brevity, the label  $F.I.$  is dropped for the remainder of this chapter.

### 7.1.2 Accounting for inefficiencies and imperfect resolution

This analysis aims to measure the dependence of the CC1 $\pi^\pm$  interaction rate on the kinematic quantities of interest defined in Section 6.1.2. This rate is governed by the differential cross-section,  $d\sigma/dX$ , which is dependent on a kinematic quantity,  $X \in \{p_\mu, \cos(\theta_\mu), \dots\}$ . To measure the dependence of  $d\sigma/dX$  on  $X$ , the range of possible values of  $X$  is divided into discrete regions,  $X_t \rightarrow X_{t+1}$ , called *bins* labelled by the index  $t$ . The choice of bin edges,  $\{X_t\}$ , is discussed in Section 7.4.1. The *discrete* flux-integrated differential cross-section,  $(d\sigma/dX)_t$ , is defined as the average of the continuous differential cross-section,  $d\sigma/dX$ , over the bin  $t$ :

$$\left(\frac{d\sigma}{dX}\right)_t = \frac{1}{\Delta X_t} \int_{X_t}^{X_{t+1}} \left[\frac{d\sigma}{dX}(X)\right] dX.\tag{7.3}$$

Here,  $\Delta X_t = X_{t+1} - X_t$  is the width of bin  $t$ . Let  $S_t$  be the number of CC1 $\pi^\pm$  interactions that occur in the FV with  $X$  in bin  $t$  over an exposure  $P$  (measured in POT) of the detector to the beam. In this case,  $S_t$  is related to  $(d\sigma/dX)_t$  via:

$$S_t = \Delta X_t \times P \times T \times \Phi \times \left( \frac{d\sigma}{dX} \right)_t. \quad (7.4)$$

In practice,  $S_t$  can not be counted directly from the data for three reasons:

1. **Efficiency.** The number of interactions that can be measured will be fewer than the number that occurred in reality, due to inefficiencies in the detection, reconstruction and selection of CC1 $\pi^\pm$  events. The number of CC1 $\pi^\pm$  interactions,  $s_t$ , with  $X$  in bin  $t$  that pass the selection, is related to  $S_t$  via Equation 7.5, where  $\epsilon_t$  is the efficiency of the selection in bin  $t$ .

$$s_t = S_t \times \epsilon_t \quad (7.5)$$

2. **Resolution.** The true kinematic quantity  $X$  can only ever be estimated using the reconstruction. The estimated value  $X'$  will have an imperfect resolution and hence a given event may have  $X'$  in bin  $r$ , but  $X$  in bin  $t$ , where  $r \neq t$ . To account for this, let  $M_{rt}$  be the elements of the smearing matrix,  $M$ . Each element,  $M_{rt}$ , gives the conditional probability that a selected CC1 $\pi^\pm$  event with  $X$  in bin  $t$  will be reconstructed with  $X'$  in bin  $r$ . The number of selected CC1 $\pi^\pm$  events,  $s'_r$ , with  $X'$  in bin  $r$  is related to  $s_t$  via:.

$$s'_r = \sum_t M_{rt} s_t, \quad (7.6)$$

where,  $M_{rt} = P(X' \text{ in } r | X \text{ in } t)$ .

3. **Backgrounds.** The selected events will not all, in truth, correspond to the CC1 $\pi^\pm$  topology due to inefficiencies in the reconstruction and particle identification. The number of events,  $n'_r$ , that pass the selection (with  $X'$  in bin  $r$ ) is related to  $s'_r$  via Equation 7.7, where  $b'_r$  is the number of selected background events with  $X'$  in bin  $r$ .

$$n'_r = s'_r + b'_r \quad (7.7)$$

Hence,  $S_t$  (the total number of CC1 $\pi^\pm$  interactions with  $X$  in bin  $t$ ), is related to  $n'_r$  (the number of selected events with  $X'$  in bin  $r$ ) via:

$$n'_r - b'_r = s'_r = \sum_t M_{rt} \epsilon_t S_t. \quad (7.8)$$

The quantity  $n'_r$  is counted from data, whereas the values of  $b'_r$ ,  $M_{rt}$  and  $\epsilon_t$  must be estimated from the simulation.

### 7.1.3 Forward-folding versus unfolding

Broadly, two options are available to account for the complications listed in Section 7.1.2 such that the cross-section can be compared to a theoretical prediction.

1. Present the cross-section in “reconstructed-space” (also known as reco-space, denoted by primes), according to:

$$\left( \frac{d\sigma}{dX} \right)'_r \equiv \frac{n'_r - b'_r}{\Delta X_r \times P \times T \times \Phi}. \quad (7.9)$$

This quantity includes a correction term,  $b'_r$ , for the selected background prediction, but does not include any corrections for the efficiency or smearing effects. A cross-section prediction,  $(d\sigma/dX)_t^{pred}$ , (made in truth-space) must be “forward-folded”, via Equation 7.10, to be compared to a reco-space measurement. To achieve this, the forward-folding matrix,  $\xi$ , must be calculated and presented along with the cross-section. A result presented in reco-space is known as a forward-folded measurement.

$$\left( \frac{d\sigma}{dX} \right)'_r = \frac{1}{\Delta X_r} \sum_t \xi_{rt} \Delta X_t \left( \frac{d\sigma}{dX} \right)_t^{pred}, \quad (7.10)$$

where,  $\xi_{rt} = M_{rt} \epsilon_t$



2. Present the cross-section in “truth-space”, according to Equation 7.11, where  $M^{-1}$  is the inverse of the smearing matrix,  $M$ .

$$\left(\frac{d\sigma}{dX}\right)_t = \frac{\sum_r M_{tr}^{-1} (n'_r - b'_r)}{\Delta X_t \times P \times T \times \Phi \times \epsilon_t} \quad (7.11)$$

This quantity has been corrected for backgrounds, efficiency and smearing. A result presented in truth-space is known as an unfolded measurement.

Unfolded measurements are advantageous in that they can be compared directly to a (flux integrated) theoretical prediction, and are typically easier for a human to comprehend as the effects listed in Section 7.1.2 have been removed. However, the process of unfolding is a challenging statistical inverse problem that can be (and often is) ill-posed [105]. In particular, there can be multiple distributions in truth-space that, when forward-folded, are equally consistent with the measured data in reco-space based on its statistical uncertainties. Typically, some form of regularisation is employed to add information to the system such that the unfolded solution is unique. In doing this, care must be taken that any statistical tests performed using the unfolded measurement give the same results as in reco-space.

To avoid these complications, the results produced in this analysis are presented in reco-space (according to Equation 7.9). In this case, the problem is well-posed. It is important to note that multiple forward-folding techniques are possible. For example, previous results [106] from MicroBooNE have utilised a forward-folding approach similar to Equation 7.9 but have included an effective reco-space efficiency correction term. Other forward-folding approaches [107] opt to present the data ( $n'_r$ ) as-is and instead supply a background prediction along with the forward-folding matrix. As noted by Koch in [107]:

*It might seem like a shortcut for lazy experimental physicists to simply publish the raw data and response matrix to leave the rest to the model builders. This is not the case though, since the construction of the response matrix requires exactly the same understanding of the detector and care to cover all systematics as a classical, unfolding analysis.*

### 7.1.4 Total and proton-exclusive cross-sections

In addition to the differential cross-section,  $(d\sigma/dX)'_r$ , this analysis includes a measurement of the total cross-section for the proton-inclusive and exclusive channels with proton multiplicities  $N_p = 0, 1$  and  $N_p \geq 2$ . In the inclusive case, the forward-folded flux-integrated cross-section is calculated in reco-space as:

$$\sigma'^{data} \equiv \frac{n' - b'}{P \times T \times \Phi}. \quad (7.12)$$

Here,  $n'$  is the number of BNB data events that pass the selection, and  $b'$  is the predicted number of selected background events. In this case, the forward-folding matrix,  $\xi$ , is simply the efficiency,  $\epsilon$ . Hence, by analogy to Equation 7.10, a truth-space flux-integrated cross-section prediction,  $\sigma^{pred}$ , can be “forward-folded” and compared to the data by applying an efficiency correction:

$$\sigma'^{pred} = \epsilon \times \sigma^{pred}. \quad (7.13)$$

In this case, the unfolding problem is trivial (and well posed), hence an efficiency correction of  $1/\epsilon$  could equally be applied to  $\sigma'^{data}$ . For consistency with the differential measurements, this analysis applies the efficiency correction to the prediction following Equation 7.13. However, the unfolded total cross-section is also presented for completeness.

In principle, the proton-exclusive total cross-section measurements could be calculated via Equation 7.12. For example, in the  $N_p = 0$  case,  $b'$  would include any selected CC1 $\pi^\pm$  events with a true proton multiplicity,  $N_p \neq 0$ . However this method would introduce model-dependence as the wrong-multiplicity backgrounds must be estimated from the simulation. Instead, this analysis follows a forward-folding procedure whereby  $N_p$  is treated like a kinematic parameter,  $X$ . The proton-exclusive total cross-section is calculated in reco-space as:

$$\begin{aligned} (\sigma_{N_p}^{data})'_r &= \frac{n'_r - b'_r}{P \times T \times \Phi}, \\ (\sigma_{N_p}^{pred})'_r &= \sum_t \xi_{rt} (\sigma_{N_p}^{pred})_t, \end{aligned} \quad (7.14)$$

where,  $\xi_{rt} = M_{rt}\epsilon_t$ .

Here,  $b'_r$  does not include wrong-multiplicity events. Instead, this effect is encoded in the smearing matrix,  $M_{rt} = P(N'_p \text{ in } r | N_p \text{ in } t)$ .

## 7.2 Treatment of uncertainties

### 7.2.1 Multivariate Gaussian parametrisation

In this analysis, all quantities are assigned uncertainties that are parametrised according to a multivariate Gaussian Probability Density Function (PDF), given generally by Equation 7.15.

$$\begin{aligned}
 x_i &= x_i^{nom} + \delta_i && \text{Random variable} \\
 \mu_i &= \langle x_i \rangle && \text{Mean} \\
 b_i &= \mu_i - x_i^{nom} && \text{Bias} \\
 V_{ij} &= \langle (x_i - \mu_i)(x_j - \mu_j) \rangle && \text{Covariance matrix} \\
 f(\underline{\delta}; \underline{b}, V) &= \frac{1}{\sqrt{(2\pi)^N |V|}} \exp \left[ -\frac{1}{2} (\underline{\delta} - \underline{b})^T V^{-1} (\underline{\delta} - \underline{b}) \right] && \text{PDF}
 \end{aligned} \tag{7.15}$$

Here  $\underline{x}^{nom}$  is the nominal (quoted) value of the quantity in question, such as a total or differential cross-section or forward-folding matrix.  $\underline{x}$  is a random variable that is distributed according to  $f(\underline{\delta}; \underline{b}, V)$ , where  $\underline{\delta}$  is the deviation of the random variable from the nominal value. The bias,  $\underline{b}$ , is the deviation of the mean,  $\underline{\mu}$ , from the nominal value, and the covariance matrix  $V$  sets the scale of the deviations around the mean. The index,  $i$ , labels the bins and  $N$  is the number of bins.

Most sources of uncertainties in this analysis are unbiased ( $\underline{b} = \underline{0}$ ). In this case, the uncertainty is parametrised by the covariance matrix alone. In the cases where the bias is non-zero, an approximation is made - given by Equation 7.16.

$$\begin{aligned}
 E_{ij} &= \langle (x_i - x_i^{nom})(x_j - x_j^{nom}) \rangle = V_{ij} + b_i b_j && \text{Error matrix} \\
 g(\underline{\delta}; E) &= f(\underline{\delta}; \underline{0}, E) \approx f(\underline{\delta}; \underline{b}, V) && \text{Unbiased PDF}
 \end{aligned} \tag{7.16}$$

Here the error matrix describes the scale of the deviations around the nominal value. The elements,  $E_{ij}$ , are composed of a true covariance term  $V_{ij}$ , and a bias term ( $b_i b_j$ ). The biased PDF,  $f(\underline{\delta}; \underline{b}, V)$ , is then approximated by an unbiased function,  $g(\underline{\delta}; E)$ . In this way, all sources of uncertainty are parametrised by an error matrix,  $E$ , and are interpreted as unbiased Gaussian deviations around the nominal value. This approximation is useful as the total PDF for a set of uncorrelated and unbiased Gaussian variations, parametrised by error matrices,  $E^n$ , is itself an unbiased multivariate Gaussian distribu-

tion, parametrised by the error matrix  $E^{tot}$ , where:

$$E_{ij}^{tot} = \sum_n^{\text{sources}} E_{ij}^n = E_{ij}^{stat} + E_{ij}^{flux} + \dots \quad (7.17)$$

### 7.2.2 Quantities presented by this analysis

As described in Section 7.1.3, this analysis presents forward-folded cross-section results. As such, the following quantities are required:

- $\left(\frac{d\sigma}{dX}\right)_r^{data}$  - the CC1 $\pi^\pm$  forward-folded flux-integrated differential cross-section with respect to the kinematic quantity,  $X$ . This quantity has  $N$  bins.
- $E_{rs}^{data}$  - the  $N \times N$  error matrix for  $\left(\frac{d\sigma}{dX}\right)_r^{data}$ , combining all sources of uncertainty.
- $\xi_{rt}$  - the  $N \times N$  forward-folding matrix that transforms a predicted cross-section from truth-space to reco-space.
- $E_{ij}^\xi$  - the  $N^2 \times N^2$  error matrix for the flattened  $N^2 \times 1$  smearing matrix,  $\xi_i^{flat}$ .

Note that  $\xi^{flat}$  is the  $N^2 \times 1$  column-vector form of the  $N \times N$  forward-folding matrix,  $\xi$ , and is constructed via:

$$\begin{aligned} \xi_i^{flat} &= \xi_{(i/N)(i\%N)}. \\ / &\Rightarrow \text{integer division} \\ \% &\Rightarrow \text{modulo operator} \end{aligned} \quad (7.18)$$

For the remainder of this thesis  $\left(\frac{d\sigma}{dX}\right)_r^{data}$  and  $\xi^{flat}$  will be known collectively as “quantities of interest” to this analysis.

### 7.2.3 Comparing a prediction to data

Consider a CC1 $\pi^\pm$  flux-integrated differential cross-section prediction in truth-space,  $(d\sigma/dX)_t^{pred}$ , with an associated error matrix,  $E_{tu}^{pred}$ . In order to be compared to the data,  $(d\sigma/dX)_r^{data}$ , this prediction must be forward-folded using Equation 7.10. The result is a prediction in reco-space,  $(d\sigma/dX)_r^{pred}$ , with an error matrix  $E_{rs}^{pred}$  that contains the uncertainties on both the truth-space prediction,  $E_{tu}^{pred}$ , and the forward-folding matrix  $E_{ij}^\xi$ . Here,  $s$  and  $u$  are bin labels in reco-space and truth-space respectively.

For Gaussian uncertainties, the quantity  $Z$  (given in Equation 7.19) can be used as a test statistic to measure the extent to which the prediction matches the data.  $Z$  is distributed according to a  $\chi_N^2$  distribution with  $N$  degrees of freedom [108], where  $N$  is also the number of bins of the measured cross-section.

$$Z = \sum_{r,s}^N \left[ \left( \frac{d\sigma}{dX} \right)_r^{data} - \left( \frac{d\sigma}{dX} \right)_r^{pred} \right] (E'^{tot})_{rs}^{-1} \left[ \left( \frac{d\sigma}{dX} \right)_s^{data} - \left( \frac{d\sigma}{dX} \right)_s^{pred} \right] \quad (7.19)$$

The total error matrix,  $E'^{tot} = E'^{data} + E'^{pred}$  combines the error matrices for the data and prediction in reco-space.

A multisim technique is used to determine  $E'^{pred}$ , this process is described more formally in Section 7.2.4 and summarised here. A number of “universes” are generated in which  $(d\sigma/dX)_t^{pred}$  and  $\xi_{rt}$  are simultaneously varied to a random value with a probability given by their PDF (given by Equation 7.16, and parametrised by the error matrices  $E_{tu}^{pred}$  and  $E_{ij}^\xi$  respectively). The resulting variation of  $(d\sigma/dX)_r^{pred}$  over these universes is used to define  $E'_{rs}^{pred}$ , via Equation 7.20 with  $M = 10,000$  universes.

$$E'_{rs}^{pred} = \frac{1}{M} \sum_u^M \left[ \left( \frac{d\sigma}{dX} \right)_r^{pred,u} - \left( \frac{d\sigma}{dX} \right)_r^{pred,nom} \right] \left[ \left( \frac{d\sigma}{dX} \right)_s^{pred,u} - \left( \frac{d\sigma}{dX} \right)_s^{pred,nom} \right] \quad (7.20)$$

Here,  $u$  labels the universes and  $u = nom$  is the nominal universe in which  $(d\sigma/dX)_t^{pred}$  and  $\xi_{rt}$  are not varied. The inverse matrix,  $(E'^{tot})^{-1}$  is calculated using an eigendecomposition. Since  $E'^{tot}$  is a real symmetric matrix (as can be inferred from Equation 7.16), it can be expressed as:

$$E'^{tot} = Q D Q^T, \quad (7.21)$$

where  $Q$  is an orthogonal matrix whose columns are the (orthonormal) eigenvectors of  $E'^{tot}$ , and  $D$  is a diagonal matrix whose entries are the eigenvalues of  $E'^{tot}$ . Hence the inverse matrix is calculable via:

$$(E'^{tot})^{-1} = Q D^{-1} Q^T, \quad (7.22)$$

as for an orthogonal matrix,  $Q^{-1} = Q^T$ . The inverse matrix,  $D^{-1}$ , is also diagonal with entries given by the reciprocal of the eigenvalues of  $E'^{tot}$ . The matrices  $Q$  and  $D$  are calculated using the Jacobi eigenvalue algorithm [109], to floating point precision.

### 7.2.4 Techniques for assessing systematic uncertainties

As described in Section 7.1.3, in order for the measured event rate,  $n'_r$ , to be compared to a prediction, an estimate of the flux,  $\phi(E)$ , selected background rate,  $b'_r$ , and forward-folding matrix  $\xi_{rt}$  are required. These estimates are made using a simulation of the experiment, and so any uncertainties on the inputs to the simulation must be included. Broadly, the sources of uncertainty in the simulation fall into the following categories: statistical (due to the finite number of simulated events used), flux, cross-section, detector, other backgrounds (e.g. dirt), and scaling factors (exposure  $P$ , and number of targets  $T$ ). The impact of these uncertainties on the quantities of interest to the analysis (listed in Section 7.2.2) is either estimated using the unisim or multisim techniques (described in [110], and summarised below)

#### The unisim technique

Consider a quantity of interest,  $\underline{x}$ , with nominal bin values,  $x_i^{nom}$ . The unisim technique is used to estimate the uncertainty due to a set of input parameters by considering the change in  $\underline{x}$  when each input parameter is varied in turn by one standard deviation from its nominal value according to its PDF. This is achieved by running the simulation multiple times to produce a set of unisim variation samples. Each sample corresponds to a single parameter which is varied and all other parameters are set to their nominal values. A special Central-Value (CV) sample is also produced in which no parameters are varied. If the set of unisim samples are identical in all but the parameters varied (i.e. no random numbers are reseeded), then any differences in  $\underline{x}$  are assumed to be entirely as a result of the parameter variation and not statistical in nature. If the variations are independent then the fractional error matrix,  $F$ , is given by: [110].

$$F_{ij} = \sum_n^{\text{params}} \left( \frac{x_i^n - x_i^{CV}}{x_i^{CV}} \right) \left( \frac{x_j^n - x_j^{CV}}{x_j^{CV}} \right). \quad (7.23)$$

Here,  $x_i^n$  is the value of  $i$ th bin of the quantity of interest, calculated using the unisim sample in which the  $n$ th input parameter was varied. Similarly,  $x_i^{CV}$  is the value in the  $i$ th bin calculated using the CV sample. This fractional error matrix, is then applied to the nominal bin values to obtain the total error matrix for the set of unisim parameters:

$$E_{ij} = x_i^{nom} F_{ij} x_j^{nom}. \quad (7.24)$$

In general, the nominal and CV samples are not necessarily the same. For the detector uncertainties considered in this analysis, the nominal sample could not be used as the CV sample due to the computational cost of re-running the simulation with each parameter varied. Instead, the CV sample used contains fewer events.

### The multisim technique

Alternatively, the multisim technique is used to estimate the uncertainty on  $\underline{x}$  due to a set of input parameters by considering the distribution of  $\underline{x}$  over multiple variations of the nominal simulation in which all input parameters are simultaneously varied to a random value with a probability given by their PDF. Each variation is known as a universe. Unlike the unisim technique (in which the full simulation is typically re-run for each parameter variation), the value of  $\underline{x}$  in each universe is often obtained by re-weighting each event of the nominal simulation. These weights can be pre-computed by repeating the part of the simulation that is relevant for the parameters in question (e.g. just the flux simulation, or just the generator for cross-section parameters), which can be significantly less computationally intensive than re-running the full simulation. However, this re-weighting procedure is only applicable if input parameters modify the likelihood that a given event will occur, and not the contents of that event. Conceptually, and indeed as is used for the flux and cross-section uncertainties in this analysis, the weight that should be applied to a given event in a given universe is defined by:

$$w(\underline{p}^u, \underline{X}) = \frac{r(\underline{p}^u, \underline{X})}{r(\underline{p}^{CV}, \underline{X})}. \quad (7.25)$$

Here,  $\underline{p}$  is a vector whose elements are the input parameters. The label  $u$  indicates the values of the parameters in the universe in question, and  $CV$  indicates the central values.  $\underline{X}$ , is a vector whose elements are a set of physics parameters (e.g. the neutrino energy) that vary from event-to-event and depend on  $\underline{p}$ . Finally,  $r(\underline{p}, \underline{X})$  gives the simulated rate of events at the supplied values of  $\underline{p}$  and  $\underline{X}$ . The rate,  $r$ , can be computed for a wide range of values of  $\underline{p}$  and  $\underline{X}$ , by re-running the relevant part of the simulation. If the dependence of the rate on  $\underline{X}$  can be parametrised for each value of  $\underline{p}$ , (or simply binned and interpolated at intermediate values), then the relevant weight for a novel event in a given universe can simply be looked-up. Note that any correlations between the parameters,  $p_m$ , are accounted for as all parameters are varied simultaneously.



The bias vector, covariance matrix, and error matrix for  $\underline{x}$  for a given set of multisim parameters with  $M$  universes are given by Equation 7.26.

$$\begin{aligned}
 \mu_i &= \frac{1}{M} \sum_u^M x_i^u \\
 b_i &= \mu_i - x_i^{nom} \\
 V_{ij} &= \frac{1}{M} \sum_u^M (x_i^u - \mu_i) (x_j^u - \mu_j) \\
 E_{ij} &= \frac{1}{M} \sum_u^M (x_i^u - x_i^{nom}) (x_j^u - x_j^{nom}) = V_{ij} + b_i b_j
 \end{aligned} \tag{7.26}$$

Here,  $x_i^u$  is the value of the  $i$ th bin, calculated using re-weighted events in universe  $u$ . Similarly,  $x_i^{nom}$  is the value as calculated using unweighted events.

## 7.3 Sources of uncertainty

### 7.3.1 Statistical uncertainties on data

The cross-section is assigned a statistical uncertainty due to the finite number of events,  $n'_r$ , that pass the selection. The counts  $n'_r$  are considered as random variables drawn from an underlying Poisson distribution (where  $n'_r$  is taken as the best estimate of the mean of this distribution). In the limit where  $n'_r$  is large, this is well-estimated by a multivariate Gaussian distribution with zero bias, and a covariance matrix with elements,  $V_{rs} = \delta_{rs} n'_r$ . The uncertainty on  $n'_r$  is propagated to the measured cross-section, and the corresponding error matrix is given by:

$$E_{rs}^{stat} = \delta_{rs} \left[ \frac{\sqrt{n'_r}}{\Delta X_r \times P \times T \times \Phi} \right]^2. \tag{7.27}$$

### 7.3.2 Statistical uncertainties on the simulation

The bootstrap method [111] is used to account for the statistical uncertainty that results from the finite number of simulated events. In this case, it is non-trivial to derive an analytic form for the elements of the error matrix as the uncertainty enters through somewhat complex objects (such as the forward-folding matrix,  $\xi$ ). Instead, the statistical uncertainty on the simulation is treated as a multisim systematic variation. In

each universe, each event is assigned a random weight drawn from a unit-mean Poisson distribution. The weight takes integer values such that each event is counted 0, 1, 2, etc. times, and the total weight (summed over of all events) is approximately equal to the number of events simulated (provided the number of events is large). In this way, a given universe is approximately equivalent to a set of events that have been randomly sampled from the events in the nominal simulation with replacement. The variation of a quantity of interest,  $\underline{x}$ , over a large number of universes gives a measure of the statistical uncertainty on  $\underline{x}$  [111]. In this analysis,  $M = 1000$ , universes are used.

### 7.3.3 Flux uncertainties

The uncertainties due to the flux simulation (discussed in Section 3.4) are detailed in [77, 86] and summarised here.

The first set of uncertainties relate to the hadron production rates ( $\pi^\pm$ ,  $K^\pm$ ,  $K_L^0$ ) in collisions between Booster protons and the BNB's beryllium target. As discussed in Section 3.4, the flux simulation uses the Sanford-Wang [87] parametrisation for the  $\pi^\pm$  production cross-sections which has been fit to data from the HARP experiment [88]. The uncertainties on the HARP data are propagated to the flux using a multisim approach. In each universe, the HARP data points are varied to a random value within their uncertainties and are fit using a cubic spline. At each pion momentum and angle, the ratio of the flux as calculated using the spline fit to that as using the SW (central-value) fit is used to define a weight. In total, 1000 universes were produced. These weights are used to propagate the flux uncertainty to the quantities of interest to this analysis using Equation 7.26.

The second set of flux uncertainties arise from the interaction cross-sections of nucleons and pions on beryllium (in the target) and aluminium (in the horn). In each case, the total cross-section,  $\sigma_{tot}$ , is broken down as:

$$\begin{aligned}\sigma_{tot} &= \sigma_{el} + \sigma_{in} \\ \sigma_{in} &= \sigma_{QE} + \sigma_{had} + \sigma_{other},\end{aligned}\tag{7.28}$$

where  $\sigma_{el}$  and  $\sigma_{in}$  are the cross-sections for the elastic and inelastic interaction modes,  $\sigma_{QE}$  is the quasi-elastic cross-section,  $\sigma_{had}$  is the hadron production cross-section (described above) and  $\sigma_{other}$  includes all other inelastic interaction modes. Here, elastic scattering is the process by which the incident hadron (nucleon, or pion) scatters co-

herently off the nucleus (beryllium or aluminium) as a whole. All other interactions fall under the inelastic category. In particular, quasi-elastic scattering is the process by which the incident hadron interacts with nucleons in the nucleus as if they were free. For protons, the breakdown of the total cross-section into each mode defines the fraction that scatter out of the target before they can interact to produce a hadron. For pions (produced in such proton-beryllium collisions), the cross-section breakdown defines the rate of absorption within the target or horn [77]. In the simulation,  $\sigma_{had}$ ,  $\sigma_{tot}$ ,  $\sigma_{in}$  and  $\sigma_{QE}$  are modelled explicitly. The remaining cross-sections (for elastic scattering and other interactions) are set implicitly via Equation 7.28. The uncertainties on the total, inelastic, and quasi-elastic cross-sections (for nucleon and pion interactions) are propagated to the flux using the unisim technique. The beam simulation is repeated with each cross-section varied by one standard deviation, and the corresponding variation of the flux is used to define an error matrix. In order to propagate these flux uncertainties to the quantities of interest to this analysis, a set of random weights are generated for each event using the unisim error matrices. These weights are treated by this analysis as the universes of a single-parameter multisim and are propagated using Equation 7.26. Again, 1000 universes are used.

The final set of flux uncertainties considered relate to the horn magnetic field. The uncertainty on the current through the horn (as quoted by the manufacturer of the current transformer) is propagated to the flux using the unisim method described above. In order to propagate the flux uncertainty to the quantities of interest to this analysis, the resulting error matrix on the flux is used to generate event weights in 1000 universes. The final source of uncertainty considered is in the modelling of the skin effect which governs the distribution of the current within the conductor [77]. A unisim approach is taken in which the flux simulation is repeated with skin effect enabled (nominal) and disabled (corresponding to a 100% variation), and the corresponding event weights are produced using 1000 universes.

### 7.3.4 POT counting uncertainties

The POT (which sets the exposure,  $P$ ) is measured by the intensity of the Booster proton beam. Two toroids monitor the rate of protons upstream of the target, and the reading from the toroid closest to the target is used to count the POT. The readings from the toroids are found to agree within 2% [77], and this value is taken to be the uncertainty on the exposure. This is propagated to the cross-section measured in this

analysis as an overall normalisation uncertainty, via Equation 7.24, where  $\sqrt{F_{ij}} = 0.02$  for all  $i, j$ .

### 7.3.5 Neutrino interaction cross-section uncertainties

This analysis makes use of the GENIE neutrino generator [90] to simulate neutrino interactions. Each model parameter used by GENIE has an associated uncertainty, which must be propagated to the quantities of interest for this analysis. For the majority of parameters, a multisim approach is taken with 100 universes and the uncertainties applied to each parameter are listed in [49]. For those parameters modified by Micro-BooNE’s tune (described in Section 3.4), the uncertainties are estimated during the fitting procedure. For all other parameters, the uncertainties recommended by the GENIE collaboration are used. A number of additional parameters, given in [49], are not reweightable in GENIE by default and so aren’t included in the multisim. Instead, a unisim approach is taken in which these parameters are considered uncorrelated.

### 7.3.6 Detector uncertainties

All uncertainties relating to the detector simulation are handled as unisims and are listed in Table 7.1. In each case, a CV sample is varied and the corresponding change in the quantities of interest is used to define a fractional error matrix, via Equation 7.23.

Variation	Description
Wire modifications	Data-driven modification to wire response simulation
Space-Charge Effect	Alternate data-driven SCE displacement map
Recombination	Variation of recombination model parameter
Light-Yield reduction	25% reduction in Light-Yield
Rayleigh scattering	Double Rayleigh-scattering length

**Table 7.1:** A summary of systematic detector variations considered.

#### Wire modifications

The response of the readout wires to charge deposited in the detector depends on a number of effects that are modelled by the simulation (such as recombination, diffusion, electron lifetime, non-responsive wires, etc.). In order to assess the uncertainty on the simulated wire responses, this analysis uses a data-driven approach described in [94]

and summarised here. This approach is advantageous as it accounts for the modelling uncertainties in a manner that is agnostic to the models employed.

A sample of anode/cathode piercing cosmic rays is selected in beam-off data, and an analogous sample is generated using CORSIKA [112] and propagated through the detector simulation. The differences between the data and simulation are assessed in terms of the widths,  $\Delta x$ , and charges,  $Q$ , of the reconstructed hits from these samples. In particular, the dependence of the most probable value of these quantities on the position  $X$ ,  $YZ$ , and direction  $\theta_{XZ}$ ,  $\theta_{YZ}$  of the hits (as calculated from the particle's reconstructed trajectory) is evaluated. Here,  $X$  is the drift coordinate,  $YZ$  is the 2D-coordinate in the plane perpendicular to the drift direction, and  $\theta_{XZ}$  and  $\theta_{YZ}$  are the angles between the reconstructed particle direction and the drift and wire directions respectively. Data/simulation ratio functions are produced that describe the factor by which the simulated  $\Delta x$  and  $Q$  should be scaled to match the data as a function of  $X, YZ, \theta_{XZ}, \theta_{YZ}$  for all readout planes.

In order to produce a detector variation sample, the nominal detector simulation is first executed. A process similar to hit reconstruction looks to identify Gaussian peaks in the wire waveforms, and produces one hit per peak. The width,  $\Delta x$ , is the standard deviation of the Gaussian that best fits the peak, and the charge,  $Q$ , is the integral. Next, the widths and charges of Gaussian peaks that originate from a simulated neutrino interaction are weighted by the appropriate factor from the aforementioned ratio functions. Finally, a time-dependent weight is applied to the simulated wire waveforms such that they are consistent with the modified Gaussian peaks. These modified waveforms form the input to the standard reconstruction chain (including hit finding) which proceeds in the same manner as the nominal simulation.

### Space-Charge Effect

The Space-Charge Effect is modelled by distortion and correction displacement maps, discussed in Section 3.3. These maps allow for the conversion between the true point at which charge is deposited and the reconstructed position. MicroBooNE has measured these displacement maps using through-going cosmic rays [99], and the laser calibration system [100]. The results of these approaches have been combined such that the full TPC volume is covered, but in general the cosmic-ray measurement takes precedence. The distortion map is used in the simulation, and the correction map is used on data (and simulated data) to remove the distortions during reconstruction.

Laser calibration data is used to estimate the residual uncertainty due to the SCE. This process is described in [99] and summarised here. The laser calibration system is used to produce known (straight-line) ionisation deposits. At each point along a reconstructed laser trajectory, the SCE-correction vector is found by querying the combined (cosmic-ray + laser) displacement map. If the map is accurate, this vector should point from the reconstructed (distorted) trajectory point to the corresponding (undistorted) point along the true laser trajectory. As the laser trajectory is known, the accuracy of this SCE-correction can be calculated. By considering many laser trajectory points, an estimation of the bias of the SCE-correction is produced. This is done for all regions of the TPC volume covered by the laser data and is extrapolated elsewhere. This bias arises from differences in the SCE correction derived using cosmic rays and the laser calibration data, and is taken as an estimate of the uncertainty. A detector variation sample is produced in which the combined SCE displacement maps are modified such that the estimates of the laser trajectories are unbiased.

## Recombination

As described in Section 6.3.1, MicroBooNE uses the modified box recombination model to estimate the rate of energy loss  $dE/dx$  along a particle's trajectory from the rate of charge deposition,  $dQ/dx$ . This model is given by Equation 6.1, and reproduced below:

$$\frac{dE}{dx} = \frac{1}{\beta} \cdot \rho E_D \left[ \exp \left( \beta \cdot \frac{W_{ion}}{\rho E_D C_{cal}} \cdot \frac{dQ}{dx} \right) - \alpha \right]. \quad (7.29)$$

The nominal simulation uses the model parameters ( $\alpha$  and  $\beta$ ) derived by ArgoNeuT [101]. However, MicroBooNE have also fit the model to a sample of protons from beam-on data [85]. To assess the uncertainties due to recombination, a detector variation sample is produced in which the parameters are modified to cover the values found from the fit.

## Light-Yield & Rayleigh scattering length

The final two detector variations relate to the optical simulation. The Light-Yield (LY) defines the number of scintillation photons produced per unit of energy absorbed by the argon due to excitation by a charged particle. This quantity is modelled by the optical simulation, and sets the overall scale of the simulated PMT responses to activity in the detector. In general the PMT responses also depend on the simulation of the transport of the scintillation photons, which includes effects such as reflections,

scattering, attenuation etc. This is modelled by the photon visibility map, described in Section 3.4.

In a similar manner to other detector calibration techniques employed by Micro-BooNE, anode and cathode piercing cosmic rays are used to estimate the uncertainty on the photon visibility map used by the simulation. These particles are typically minimally-ionising and so their expected  $dE/dx$  distribution is constant along the particle's trajectory. Hence, the total energy deposited is assumed to be proportional to the reconstructed length of the particle track. For each particle, the corresponding total PE count over all PMT channels is divided by the track length and used as a metric for the average visibility along the particle's trajectory. The distribution of this metric was found separately for anode-piercing and cathode-piercing tracks as a function of the midpoint of the track in the drift direction in both data and simulation. The simulation was found to uniformly overestimate this metric by  $\sim 25\%$  when compared to data. As a result, a detector variation sample was produced in which the LY was reduced by this amount. In this way the varied simulation more accurately represents the data in the PE per track-length metric.

In addition to the overall scale factor, an additional detector variation sample was produced in which the simulated Rayleigh scattering length was modified. This length defines the mean free path through which a photon will travel before scattering and is not well constrained by existing data. Consequently, a 100% uncertainty was assigned and a detector variation sample was produced in which the scattering length is doubled from 60 cm (nominal) to 120 cm (variation) and the photon visibility map was recalculated.

### Dirt uncertainties

The rate of dirt interactions is difficult to simulate as the geometry and composition of the TPC surroundings (e.g. the detector hall and pit) are not well modelled. A 100% uncertainty is assigned to the dirt background, which is handled as a single-parameter 2-universe multisim. The quantities of interest are measured three times; once using the nominal dirt simulation, and in two universes where the dirt events are weighted by a factor of 0 (no dirt) and 2 (double dirt). The error matrix is calculated via Equation 7.26, and the variations were found to be symmetric around the nominal. In the nominal simulation, dirt events make up 1% of the events passing the generic selection and in the majority of cases the reconstructed particles are truly cosmic rays.

## 7.4 Cross-section measurement

### 7.4.1 Binning choice

Appendix A details the choice of bins,  $X_t$ , in which the differential cross-sections are extracted, and the overall strategy is summarised here. The number of possible bins is limited by the available statistics and the resolution with which the kinematic parameter  $X$  can be estimated. In general, the bins were chosen such that each contains  $\gtrsim 100$  selected events (as predicted by the simulation), and such that the smearing matrix is mostly diagonal.

Additionally, regions of the phase-space in which the selection has very low efficiency are excluded and treated as background. In particular, the signal region of the phase-space is restricted to:

- Muon momentum,  $p_\mu > 150 \text{ MeV c}^{-1}$ .
- Muon-pion opening angle,  $\theta_{\mu\pi} < 2.65 \text{ rad}$ .

Given the important challenge for MicroBooNE to address the excess of low energy electrons observed by MiniBooNE, and in order to avoid biases in analyses, data that might contain such particles is blinded until an agreement within the whole collaboration has been reached. Therefore, analyses not pursuing the understanding of such low energy electrons need to incorporate a cut to ensure they are not contained in their selections when using data. The restriction on  $p_\mu$  in this analysis is due to the requirement that the reconstructed muon candidate of the CC-inclusive pre-selection has a track length of at least 20 cm to meet the criteria to un-blind the data. The restriction on  $\theta_{\mu\pi}$  is due to the opening-angle cut which removes reconstruction failures. For the measurement of the proton multiplicity (as illustrated by Figure 6.1), a threshold proton momentum of  $300 \text{ MeV c}^{-1}$  is applied.

The choice of binning for pion momentum (shown in Appendix A) is particularly noteworthy and so is reproduced in Figure 7.1 and described below. The solid black lines in the figure indicate the edges of the bins in which the cross-section is measured.

Very few pion candidates are reconstructed with a momentum below  $\sim 100 \text{ MeV c}^{-1}$ . This threshold corresponds to a reconstructed range of  $\sim 3 \text{ cm}$  or an average of 6 hits per view. Reconstructed particles below these values are not selected as the pion candidate as a result of the conditions required to calculate the features used by the particle

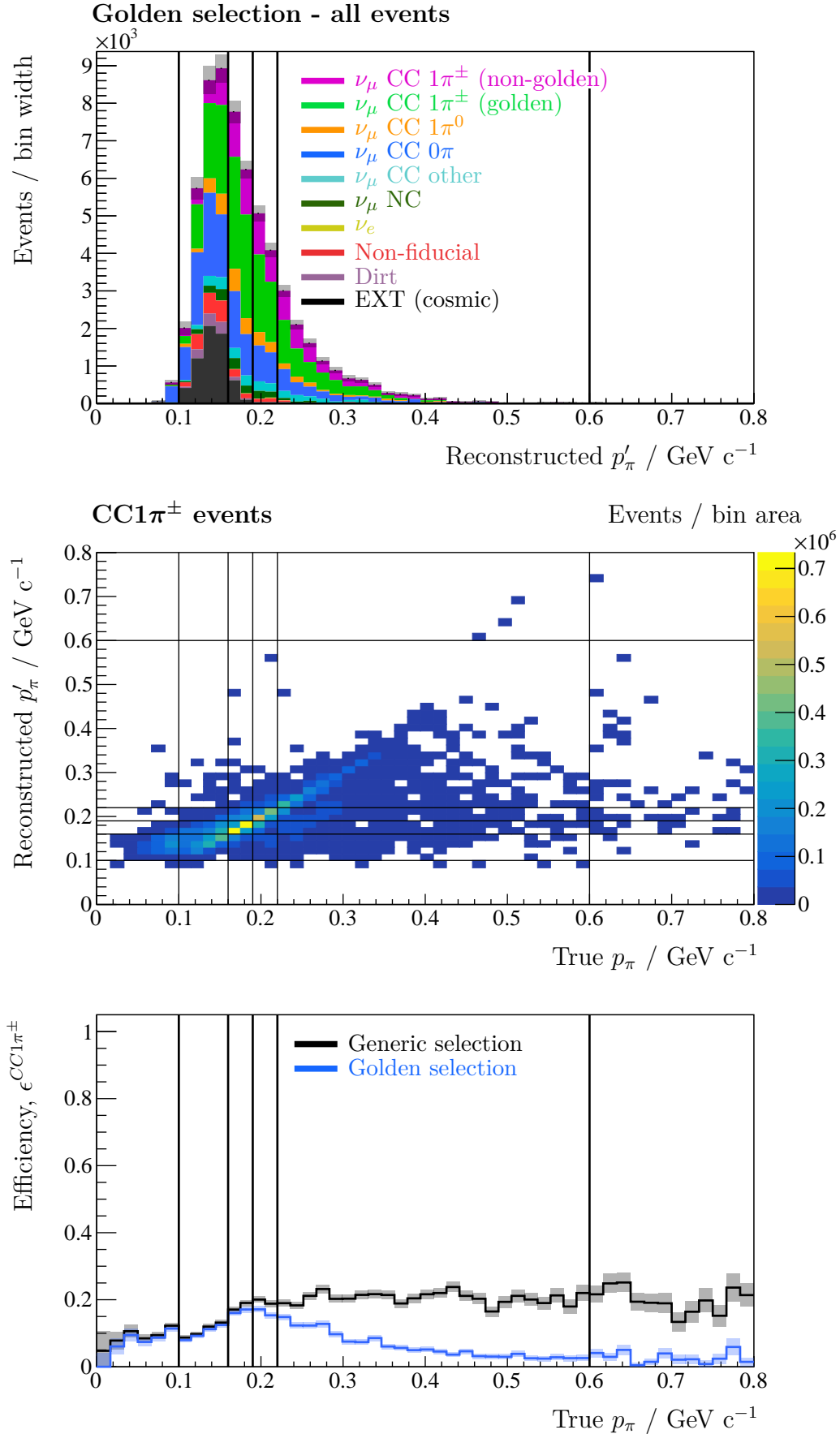


identification BDTs. In particular, the Bragg likelihood method excludes the first and last hits in the fitted track trajectory. Similarly, the truncated mean  $dE/dx$  method excludes the first three hits, and any additional hits with  $dE/dx$  greater than one standard deviation from the mean  $dE/dx$ . Additionally, both methods only consider hits for which a reconstructed space-point was produced.

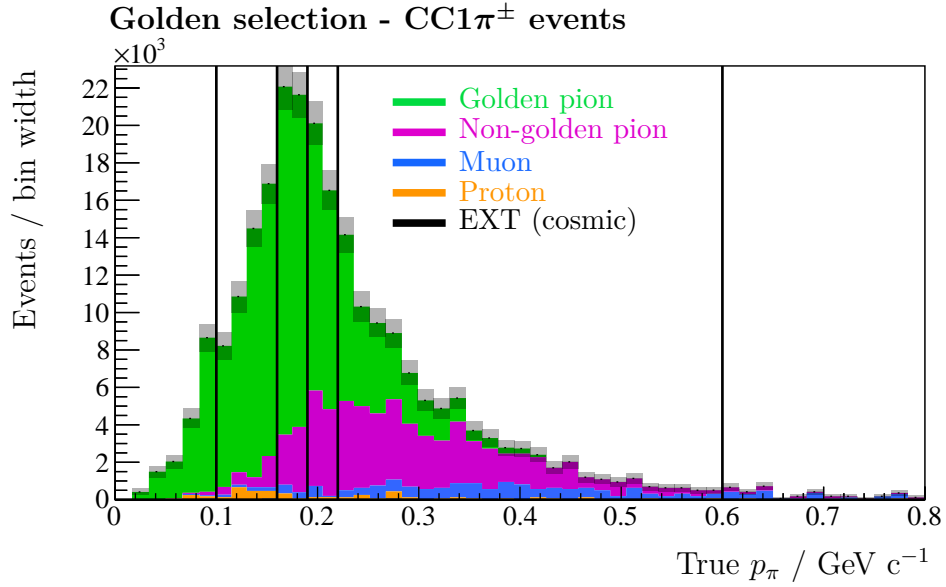
The events for which the reconstructed pion momentum is larger than the true pion momentum ( $p'_\pi > p_\pi$ ) primarily occur at low  $p_\pi$ . In this region, the pion candidate particle most often matches to a true golden pion, as shown in Figure 7.2. The dominant failure mode that causes the pion momentum to be incorrectly estimated is illustrated by Figure 7.3. Low momentum pions typically travel a short distance and produce few hits before stopping and decaying. The resulting decay electron (produced by a short-lived intermediate muon) can itself ionise the argon and produce additional hits. If the pattern recognition does not correctly identify the pion and decay electron as separate particles, then the range of the combined particle is used to calculate the reconstructed pion momentum which hence overestimates the true value. This effect becomes most prevalent at low true pion momentum as the initial pion is typically short and more difficult to distinguish from the decay products.

As a result of these effects, both the generic and golden selections accept events with true pion momenta below the reconstruction threshold of  $\sim 100 \text{ MeV c}^{-1}$ , but the estimation of the pion momentum in these events is inaccurate. Consequently, the cross-section is not presented in this region. These low true momentum events are not taken as background (as in the muon case), as the need to subtract them from higher-momentum bins in reconstructed space would introduce model dependence. Instead, the events are included as an underflow bin in the smearing matrix.

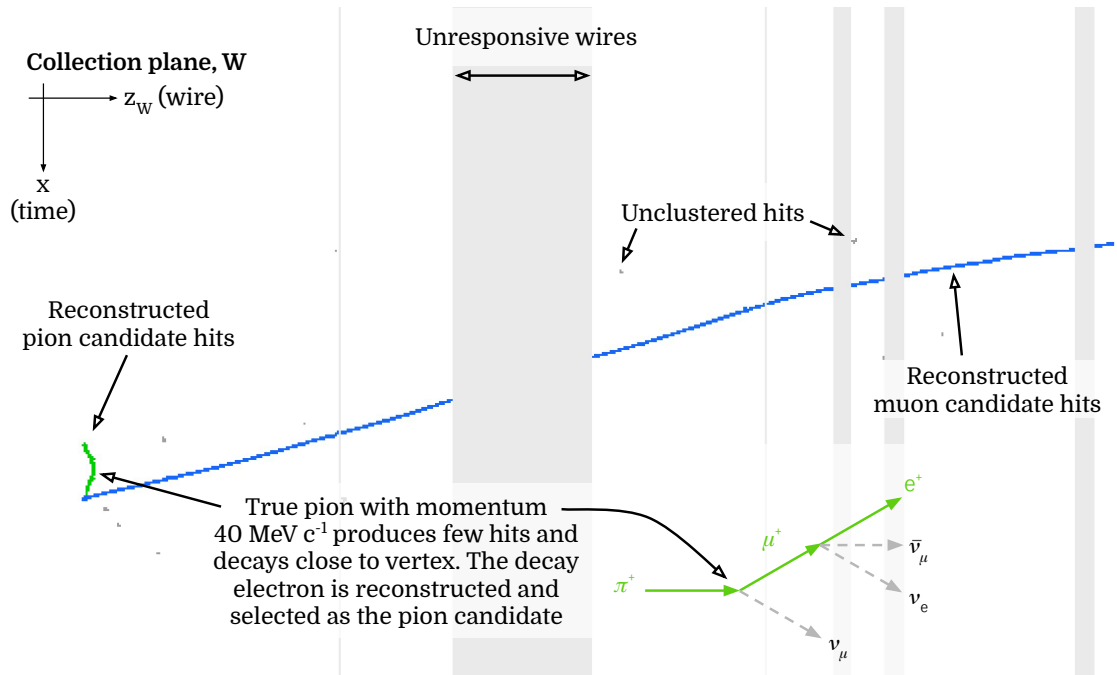
The events for which  $p'_\pi < p_\pi$  primarily contain non-golden pions for which the range-based momentum estimation is not applicable (as illustrated in Figure 6.29). As intended, these entries are largely removed by the golden selection.



**Figure 7.1:** The distribution of events passing the golden selection (top), the reconstructed versus true pion momentum for golden-selected signal events (middle), and the efficiency of the generic and golden selections (bottom). The black lines indicate the edges of the bins in which the cross-section is extracted.



**Figure 7.2:** The true pion momentum in CC1 $\pi^\pm$  events passing the golden selection. The colours indicate the true origin of the reconstructed pion candidate. The black lines indicate the edges of the bins in which the cross-section is extracted.



**Figure 7.3:** Event display showing a common failure mode of the pion momentum estimation.

### 7.4.2 Total cross-section extraction

The number of target nucleons,  $T$ , is given by:

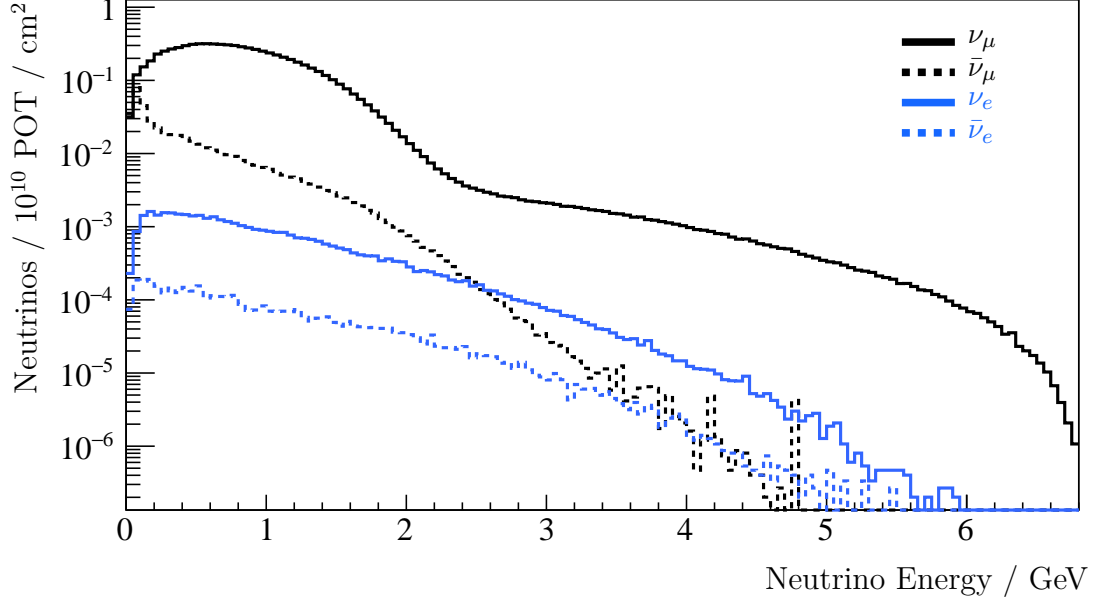
$$\begin{aligned}
 T &= 4.15128 \times 10^{31} \quad [\text{nucleons}] \\
 &= N_{\text{nuc}} \quad 40 \quad [\text{nucleons/atom}] \\
 &\times N_A \quad 6.0221 \times 10^{23} \quad [\text{atoms/mol}] \\
 &\times 1/m_{\text{mol}} \quad 1/39.948 \quad [\text{mol/g}] \\
 &\times \rho_{\text{LAr}} \quad 1.40 \quad [\text{g/cm}^3] \\
 &\times V \quad 4.91746 \times 10^7. \quad [\text{cm}^3]
 \end{aligned} \tag{7.30}$$

Here,  $N_{\text{nuc}}$  is the number of nucleons per argon atom,  $N_A$  is Avogadro's constant,  $m_{\text{mol}}$  is the molar mass of argon,  $\rho_{\text{LAr}}$  is the density of liquid argon, and  $V$  is the fiducial volume. It is important to note that in the literature some cross-section measurements are presented per target atom, and others are presented per target nucleon - the latter is used in this analysis. The neutrino flux prediction is shown in Figure 7.4, and the integrated  $\nu_\mu + \bar{\nu}_\mu$  flux is  $\Phi = 7.83 \times 10^{-10} \text{ cm}^{-2}/\text{POT}$  [86]. The  $\bar{\nu}_\mu$  flux ( $4.56 \times 10^{-11} \text{ cm}^{-2}/\text{POT}$ ) is included as the signal definition does not distinguish between the  $\mu^+$  and  $\mu^-$  channels. Finally, the total exposure is  $P = 1.455 \times 10^{20} \text{ POT}$ .

Table 7.2 summarises the performance of the generic and golden selections after the application of the phase-space restrictions described in Section 7.4.1.

Selection	Selected Events		Selected Backgrounds	Efficiency
	Data	Simulation		
Generic	1828	1581.4	769.8	$(18.8 \pm 1.3)\%$
Golden	1106	920.4	490.7	$(10.0 \pm 0.7)\%$

**Table 7.2:** Summary of the generic and golden CC1 $\pi^\pm$  selections in the restricted phase-space described in Section 7.4.1. The simulated events have been scaled to the POT of the BNB data. The uncertainty quoted on the efficiency includes all systematic uncertainties discussed in Section 7.2 which are detailed in Table 7.5 (generic selection) and Table 7.6 (golden selection).



**Figure 7.4:** The neutrino flux prediction at MicroBooNE averaged over the full TPC volume [86]. This analysis includes the  $\nu_\mu$  and  $\bar{\nu}_\mu$  (black) fluxes in the cross-section calculation. The  $\nu_e$  and  $\bar{\nu}_e$  (blue) fluxes are included in the simulation and contribute to the background prediction.

Table 7.3 gives the total CC1 $\pi^\pm$  flux-integrated forward-folded cross-section as found using each selection. It is important to note that, as shown in Equation 7.12,  $\sigma^{tdata}$  does not include an efficiency correction and so the result of the golden and generic selections can not be compared directly. However,  $\sigma^{tpred}$  is forward-folded and so can be compared to  $\sigma^{tdata}$  from the relevant selection. The  $\chi^2$  test statistic has been calculated to compare the MicroBooNE data and GENIE prediction for both selections. In addition, the p-value has been calculated. Under the assumption that the GENIE prediction is true, the p-value gives the probability that a measurement at least as unlikely as the observed data would be observed due to random fluctuations within the assigned uncertainties. Commonly in particle physics, the data needs to be at least 3 standard deviations from the predicted value to provide evidence that the prediction is inconsistent with the data: this corresponds to a p-value of 0.003 or less. Although it is beyond the scope of this analysis to assess the consistency of various models with the data, by this metric MicroBooNE data does not rule out the GENIE prediction.

Selection	$\sigma'^{data} / 10^{-41} \text{ cm}^2$	$\sigma'^{pred} / 10^{-41} \text{ cm}^2$	$\chi^2$	p-value
Generic	$22.4 \pm 5.3$	$17.2 \pm 1.3$	0.935	0.33
Golden	$13.0 \pm 3.2$	$9.1 \pm 0.4$	1.557	0.21

**Table 7.3:** The forward-folded flux-integrated CC1 $\pi^\pm$  total cross-section as measured in MicroBooNE data and predicted by GENIE. The uncertainties on the data include all statistical and systematic sources, and are detailed in Table 7.5 (generic selection) and Table 7.6 (golden selection). The uncertainties on the prediction include the prior (MC statistical) uncertainty and the uncertainties propagated from the selection efficiency. The  $\chi^2$  test statistic is calculated according to Equation 7.19 and the associated p-value is calculated using this statistic with 1 degree of freedom.

For completeness, Table 7.4 gives the total CC1 $\pi^\pm$  flux-integrated *unfolded* cross-section,  $\sigma^{data} = \sigma'^{data}/\epsilon$ , where the efficiency correction (shown in Equation 7.13) is instead applied to  $\sigma'^{data}$  such that it can be compared directly to the truth-space prediction  $\sigma^{pred}$ .

Selection	$\sigma^{data} / 10^{-41} \text{ cm}^2$	$\sigma^{pred} / 10^{-41} \text{ cm}^2$	$\chi^2$	p-value
Generic	$119 \pm 28$	$91.1 \pm 0.5$	0.961	0.33
Golden	$130 \pm 31$	$91.1 \pm 0.5$	1.585	0.21

**Table 7.4:** The unfolded flux-integrated CC1 $\pi^\pm$  total cross-section as measured in MicroBooNE data and predicted by GENIE. The uncertainties on the data include all statistical and systematic sources. The uncertainties on the prediction include only the MC statistical uncertainty.

Table 7.5 and Table 7.6 show the breakdown of the uncertainties for the generic and golden selections respectively. In both cases, the leading source of uncertainty ( $\sim 17\%$ ) on  $\sigma'^{data}$  relates to the modelling of neutrino-nucleus interactions by GENIE which enter primarily through the background prediction. The flux uncertainty ( $\sim 14\%$ ) is the next most dominant and enters through flux itself and the background prediction. The largest contribution to the detector modelling uncertainty ( $\sim 7\%$ ) relates to uncertainties on the recombination model. This enters the analysis through the calorimetric features that are used by the particle-identification BDTs. The leading source of uncertainty on the efficiency ( $\sim 6\%$ ) is due to the detector modelling. As expected, the efficiency is largely insensitive to the flux uncertainties as the variations generally impact the overall event rate (hence the numerator and denominator of the efficiency typically vary by the same amount), and the selection efficiencies are mostly flat as a function of the neutrino energy. The impact of GENIE uncertainties on the total selection efficiency is a consequence of any selection biases - which are shown in Appendix A by non-uniformities in the selection

efficiency as a function of the kinematic quantities of interest. The golden selection has a stronger dependence (4.7% uncertainty) on GENIE than the generic selection (2.7% uncertainty). This is driven by the inherent bias of golden pions toward lower momenta, shown in Figure 6.3, which manifests as a dependence of the efficiency on the simulated pion momentum distribution.

Source	Fractional uncertainty	
	Cross-section	Efficiency
Data statistics	4.0%	-
MC statistics	1.2%	1.1%
Flux	14.0%	0.4%
POT	2.0%	-
GENIE	16.8%	2.7%
Dirt	1.5%	-
Detector	7.2%	6.0%
Total	23.6%	6.7%

**Table 7.5:** A breakdown of the uncertainties on the measured cross-section,  $\sigma^{data}$ , and the efficiency prediction for the *generic* selection.

Source	Fractional uncertainty	
	Cross-section	Efficiency
Data statistics	5.4%	-
MC statistics	1.8%	1.6%
Flux	12.9%	0.9%
POT	2.0%	-
GENIE	16.8%	4.7%
Dirt	2.5%	-
Detector	6.4%	5.0%
Total	23.0%	7.1%

**Table 7.6:** A breakdown of the uncertainties on the measured cross-section,  $\sigma^{data}$ , and the efficiency prediction for the *golden* selection.

### 7.4.3 Proton-exclusive, and differential cross-sections

Figure 7.5 shows the total CC1 $\pi^\pm$  flux-integrated forward-folded cross-sections in each proton multiplicity bin. As discussed in Section 7.4.1 a momentum threshold of 300 MeV  $c^{-1}$  is applied when counting protons in truth-space. The remaining figures in this section show the differential cross-sections as a function of each kinematic variable. In every case, the total uncertainty on the data is presented as a fractional error matrix:

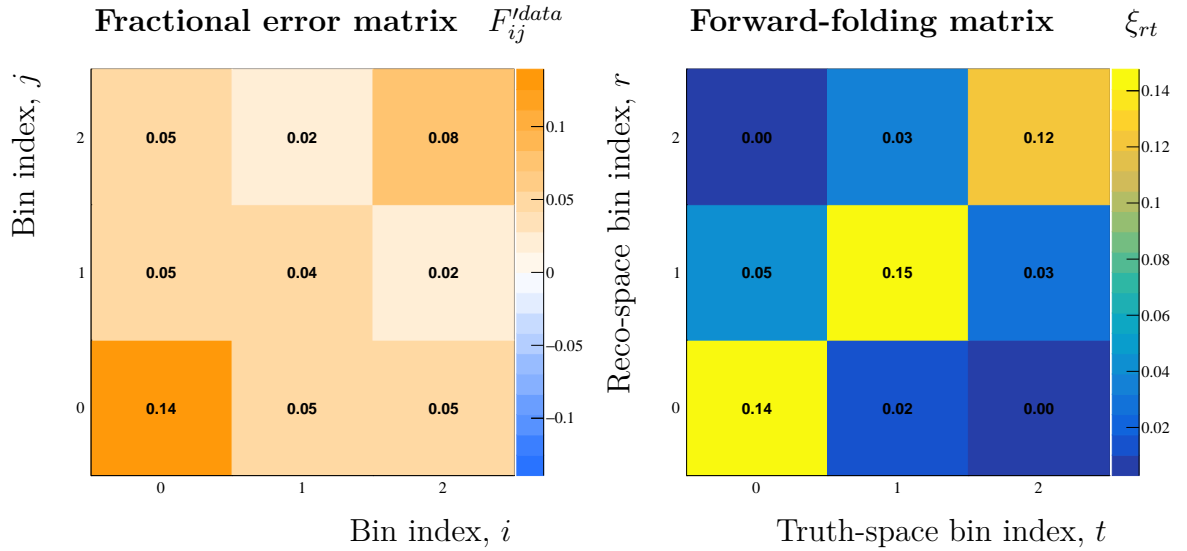
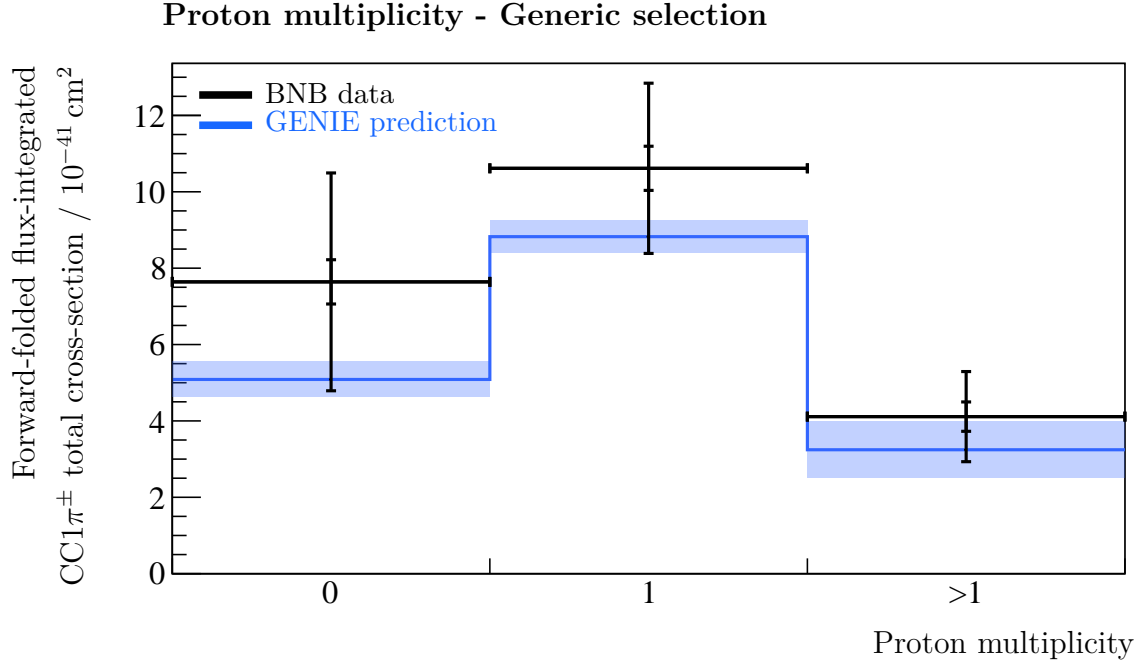
$$F_{ij}^{data} = \frac{E_{ij}^{data}}{\sigma_i^{data} \sigma_j^{data}}, \quad (7.31)$$

where  $\sigma_i^{data}$  is the cross-section (total or differential) in bin  $i$ . The diagonals of the error matrix are shown as the error bars of the BNB data. The inner error bars show the statistical-only uncertainty and the outer error bars show the total uncertainty. Additionally, the forward-folding matrix is shown for each cross-section. The GENIE prediction has been forward-folded using Equation 7.10 and is shown for comparison with the data. The uncertainties on the forward-folding matrix have been propagated to the GENIE prediction and the diagonals of the resulting error matrix are shown as an error band. The test-statistic, given by Equation 7.19, has been calculated for each cross-section measurement using the full error matrices and is given in Table 7.7.

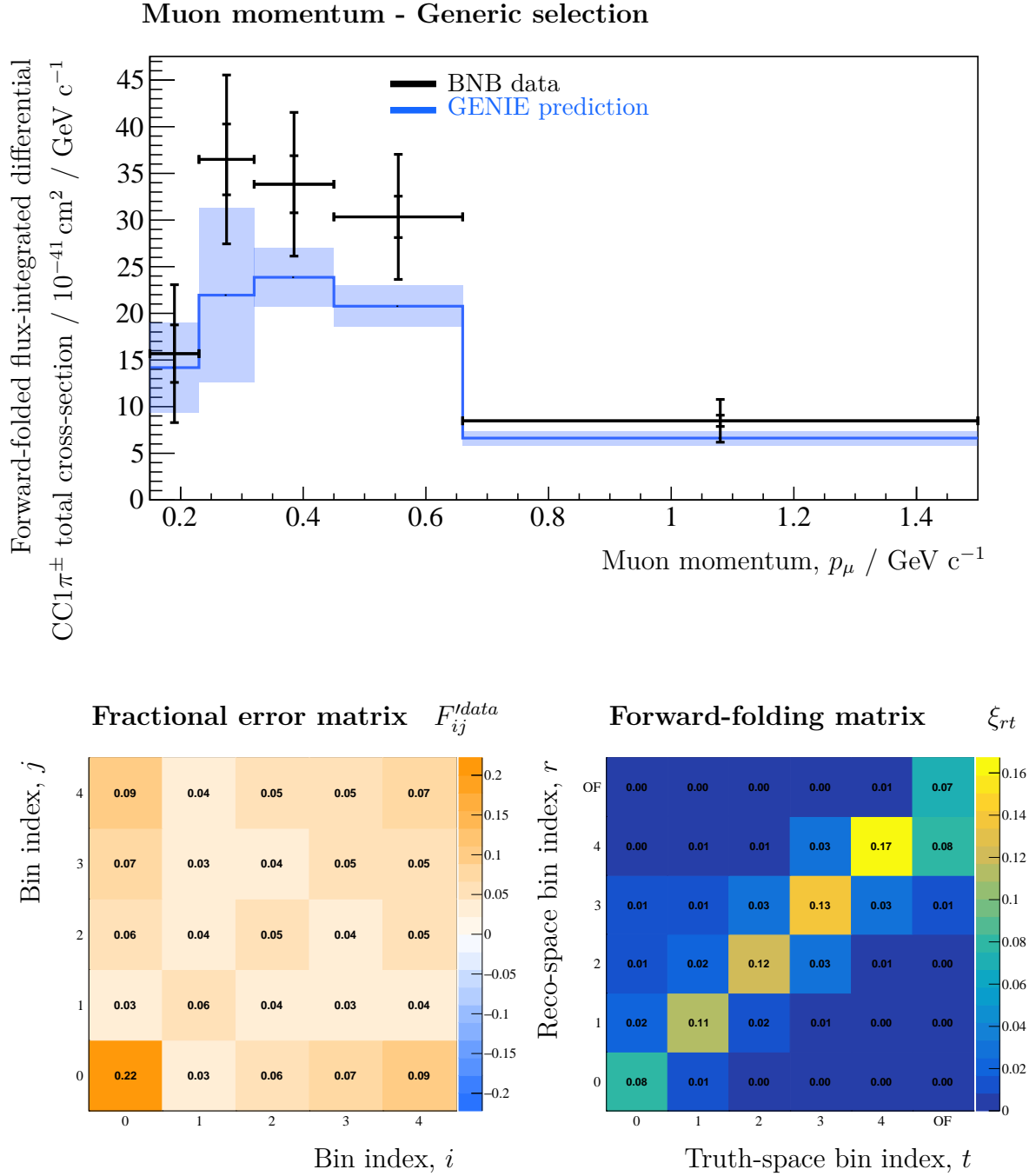
Measurement	Selection	Figure	$\chi^2$	$\nu$	$\chi^2/\nu$	p-value
Proton multiplicity, $N_p$	Generic	7.5	0.935	3	0.312	0.817
Muon momentum, $p_\mu$	Generic	7.6	2.587	5	0.517	0.763
Muon angle, $\cos(\theta_\mu)$	Generic	7.7	4.123	11	0.375	0.966
Pion momentum, $p_\pi$	Golden	7.8	3.600	4	0.900	0.463
Pion angle, $\cos(\theta_\pi)$	Generic	7.9	2.755	7	0.394	0.907
Muon-pion angle, $\theta_{\mu\pi}$	Generic	7.10	2.183	7	0.312	0.949

**Table 7.7:** Comparisons between the measured BNB data cross-sections and the forward-folded GENIE predictions. For each cross-section measurement, the  $\chi^2$  test statistic is calculated and the number of degrees of freedom  $\nu$  are supplied. The generic selection is primarily used as it offers higher statistics and fewer selection biases than the golden selection. However, the golden selection is superior for the pion momentum measurement as the momentum of non-golden pions can not be accurately measured by the methods used in this analysis.

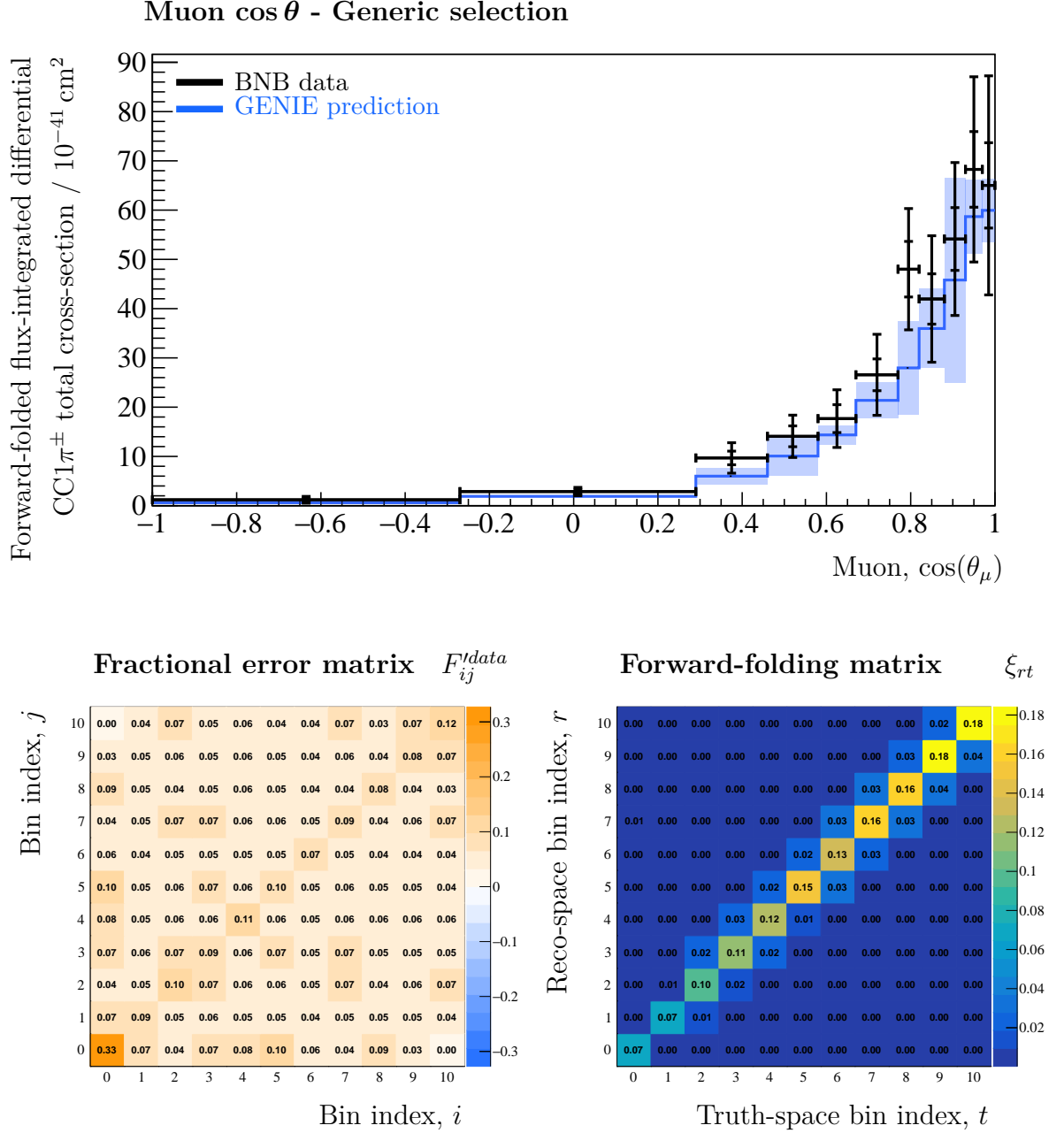




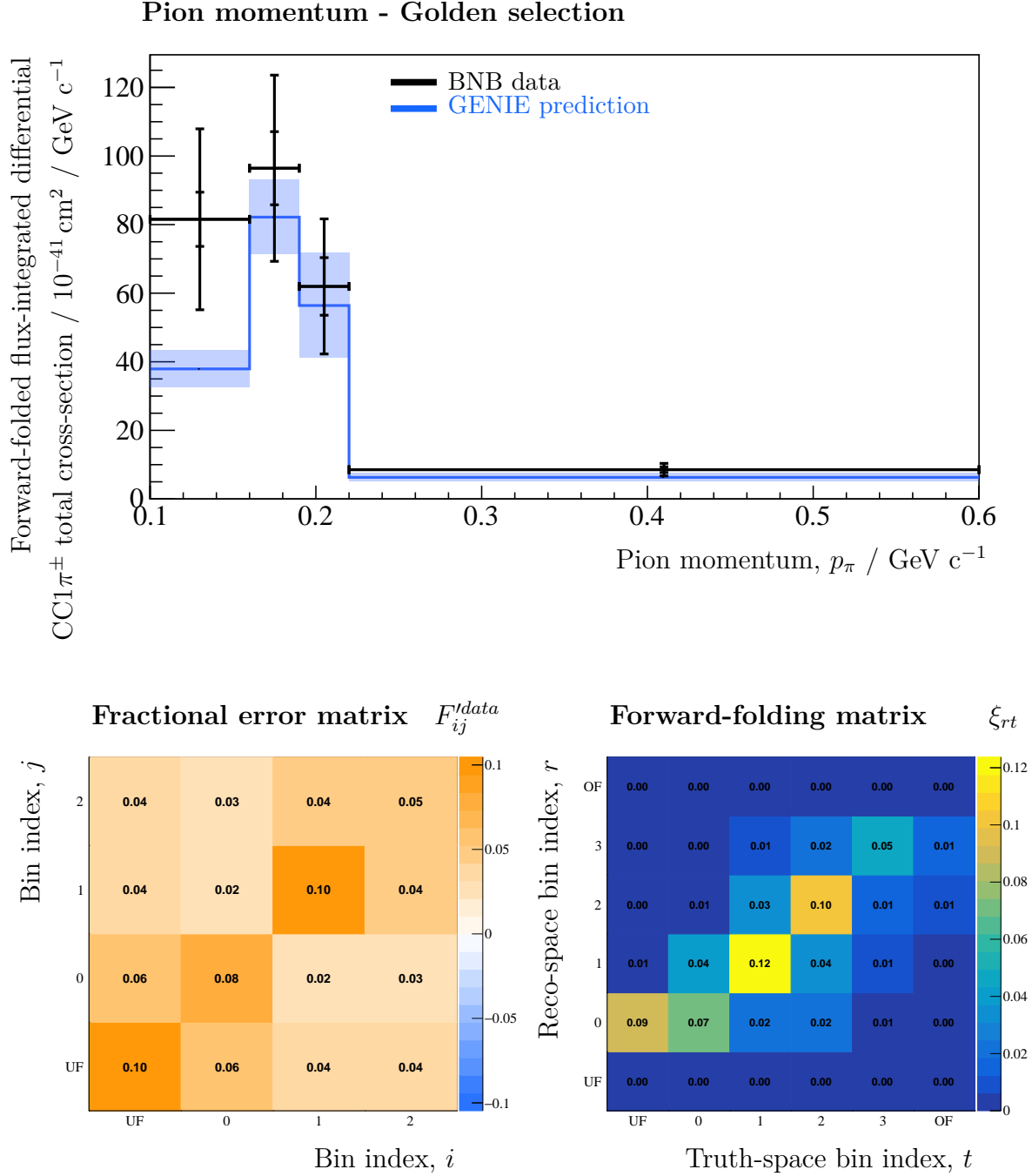
**Figure 7.5:** The total forward-folded flux-integrated cross-section for the proton-exclusive channels,  $N_p = 0, 1$  and  $N_p > 1$  (shown top). The black data points show the cross-section extracted using BNB data. The inner error bars indicate the statistical uncertainty, and the outer show the total uncertainty from the diagonals of the error matrix (shown bottom-left). The blue line shows the GENIE prediction that has been forward folded using the bottom-right matrix. The blue bands indicate the uncertainty on the prediction, which includes the uncertainties on the forward-folding matrix.



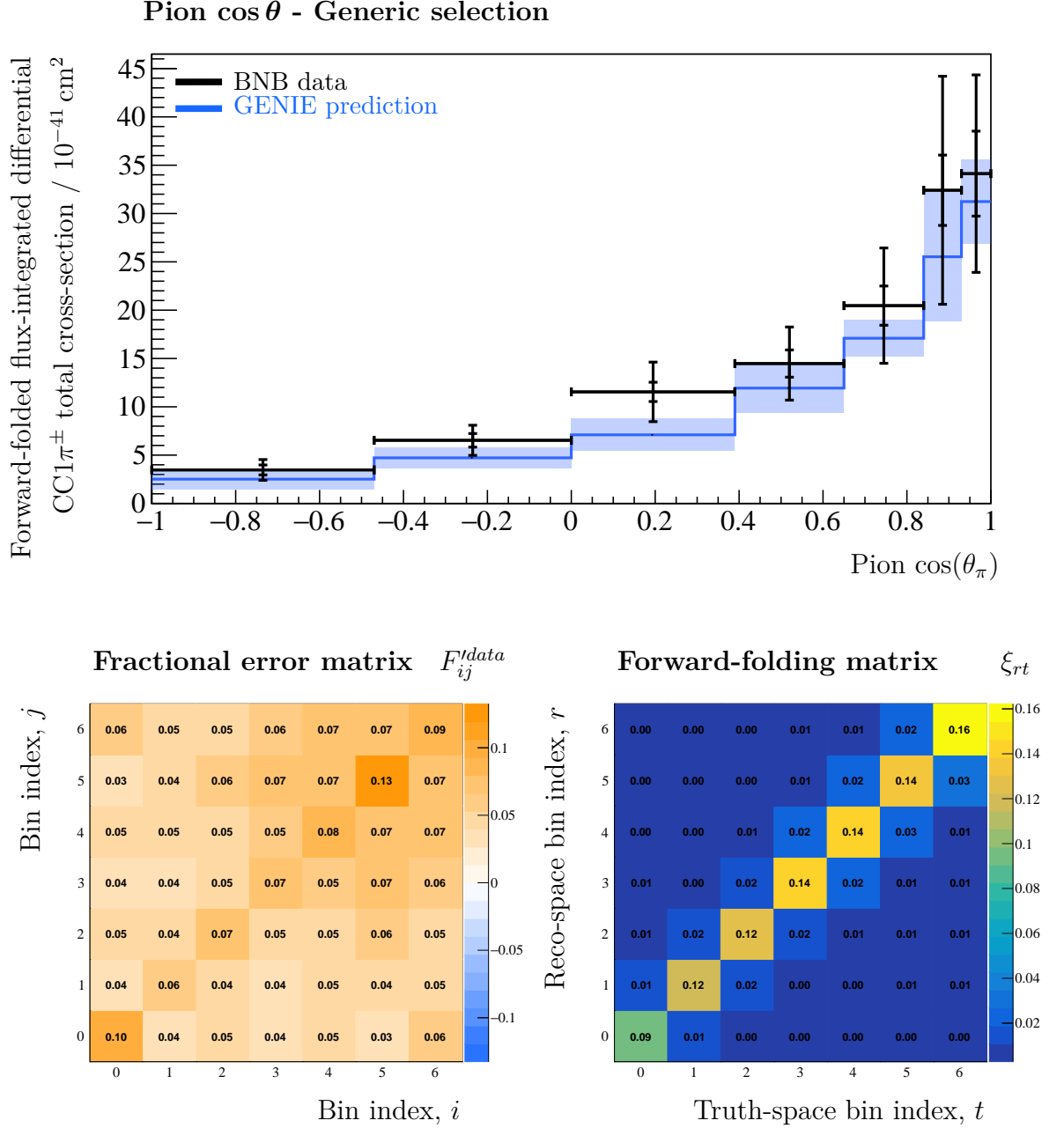
**Figure 7.6:** The forward-folded flux-integrated differential cross-section with respect to the muon momentum. The black data points show the cross-section extracted using BNB data. The inner error bars indicate the statistical uncertainty, and the outer show the total uncertainty from the diagonals of the error matrix (shown bottom-left). The blue line shows the GENIE prediction that has been forward folded using the bottom-right matrix. The blue bands indicate the uncertainty on the prediction, which includes the uncertainties on the forward-folding matrix.



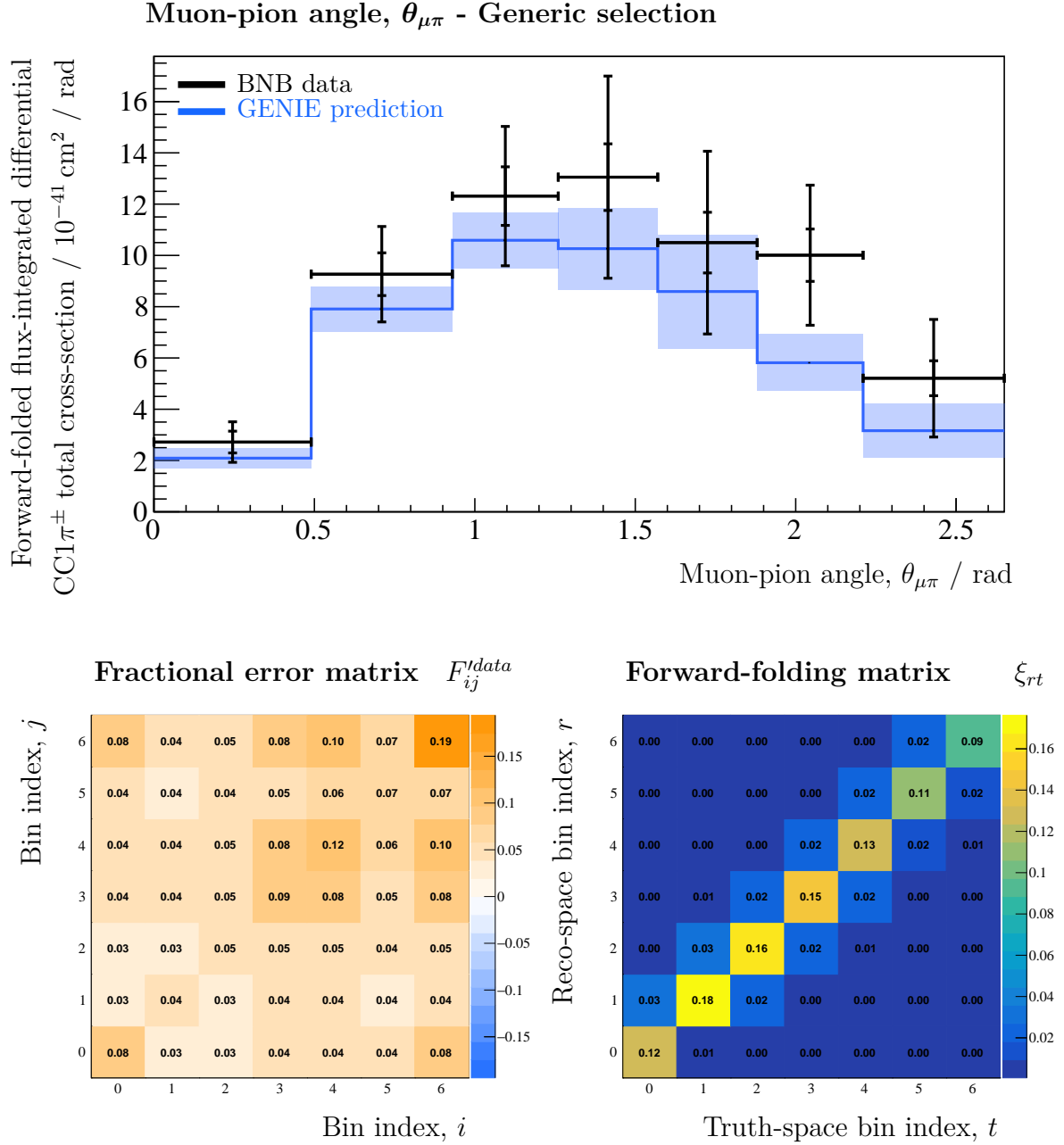
**Figure 7.7:** The forward-folded flux-integrated differential cross-section with respect to the muon  $\cos \theta$ . The black data points show the cross-section extracted using BNB data. The inner error bars indicate the statistical uncertainty, and the outer show the total uncertainty from the diagonals of the error matrix (shown bottom-left). The blue line shows the GENIE prediction that has been forward folded using the bottom-right matrix. The blue bands indicate the uncertainty on the prediction, which includes the uncertainties on the forward-folding matrix.



**Figure 7.8:** The forward-folded flux-integrated differential cross-section with respect to the pion momentum. The black data points show the cross-section extracted using BNB data. The inner error bars indicate the statistical uncertainty, and the outer show the total uncertainty from the diagonals of the error matrix (shown bottom-left). The blue line shows the GENIE prediction that has been forward folded using the bottom-right matrix. The blue bands indicate the uncertainty on the prediction, which includes the uncertainties on the forward-folding matrix.



**Figure 7.9:** The forward-folded flux-integrated differential cross-section with respect to the pion  $\cos \theta$ . The black data points show the cross-section extracted using BNB data. The inner error bars indicate the statistical uncertainty, and the outer show the total uncertainty from the diagonals of the error matrix (shown bottom-left). The blue line shows the GENIE prediction that has been forward folded using the bottom-right matrix. The blue bands indicate the uncertainty on the prediction, which includes the uncertainties on the forward-folding matrix.



**Figure 7.10:** The forward-folded flux-integrated differential cross-section with respect to the muon-pion angle,  $\theta_{\mu\pi}$ . The black data points show the cross-section extracted using BNB data. The inner error bars indicate the statistical uncertainty, and the outer show the total uncertainty from the diagonals of the error matrix (shown bottom-left). The blue line shows the GENIE prediction that has been forward folded using the bottom-right matrix. The blue bands indicate the uncertainty on the prediction, which includes the uncertainties on the forward-folding matrix.

## 7.5 Summary and discussion

### 7.5.1 Total cross-section

To summarise, the total CC1 $\pi^\pm$  flux-integrated forward-folded cross-section has been found using data from MicroBooNE's first run and is measured to be:

$$\begin{aligned} \text{Cross-section: } \sigma'^{data} &= [22.4 \pm 0.9(\text{stat.}) \pm 5.2(\text{syst.})] \times 10^{-41} \text{ cm}^2, \\ \text{Efficiency: } \epsilon &= [18.8 \pm 1.3] \%. \end{aligned}$$

This result is consistent with the forward-folded GENIE prediction of:

$$\text{GENIE: } \sigma'^{pred} = [17.2 \pm 1.3] \times 10^{-41} \text{ cm}^2.$$

The leading source of uncertainty (16.8%) on the measured cross-section,  $\sigma'^{data}$ , is due to the rate of predicted backgrounds. These backgrounds are estimated to constitute  $\sim 49\%$  of the selected sample and are primarily neutrino interactions with the  $1\mu + Xp$  topology ( $X \geq 0$ ) where a proton is misidentified. Additionally, the uncertainty due to the flux prediction (14%) is of a similar magnitude which is primarily due to uncertainties on the rate of hadron production in proton-beryllium collisions. The uncertainties on the efficiency,  $\epsilon$ , have been propagated to the forward-folded GENIE prediction and are primarily due to detector modelling. The leading sources are due to recombination and the Space-Charge Effect, which distort the topological and calorimetric features used by the particle-identification BDTs.

In addition to this proton-inclusive measurement, Figure 7.5 shows the world's first measurement of the proton-exclusive CC1 $\pi^\pm$  flux-integrated forward-folded cross-section for proton multiplicities,  $N'_p = 0, 1$  and  $N'_p > 1$ , with momenta above 300 MeV  $c^{-1}$ . This result is possible due to MicroBooNE's high spatial resolution and the reconstruction capabilities of the Pandora pattern-recognition software package. The estimated selection efficiency is not strongly dependent on the true proton multiplicity at  $\sim 19\%$  ( $N_p = 0, 1$ ), and  $\sim 15\%$  ( $N_p > 1$ ) — shown explicitly in Figure A.9. Additionally, the reconstructed and true proton multiplicities are equal in  $\sim 75\%$  of events (as indicated by the forward-folding matrix:  $\xi_{tt}/\sum_r \xi_{rt}$ ). As in the proton-inclusive case, the leading source of uncertainty is due to the predicted background rate which is largest in the  $N'_p = 0$  bin where  $1\mu + 1p$  is mistaken for the  $1\mu + 1\pi^\pm$  topology. Again, this result is consistent with the forward-folded GENIE prediction.

### 7.5.2 Differential cross-sections

The CC1 $\pi^\pm$  flux-integrated forward-folded differential cross-sections have been extracted with respect to:

- The muon and pion momenta,  $p_\mu$  &  $p_\pi$ , shown in Figure 7.6 and Figure 7.8. Note that the region  $p_\mu < 150 \text{ MeV c}^{-1}$  is excluded from the signal definition.
- The angle between the beam direction and the directions of the muon and pion,  $\theta_\mu$  &  $\theta_\pi$ , shown in Figure 7.7 and Figure 7.9.
- The muon-pion opening angle,  $\theta_{\mu\pi}$ , shown in Figure 7.10. Note that the region  $\theta_{\mu\pi} > 2.65 \text{ rad}$  is excluded from the signal definition.

The generic selection is used for all measurements other than the pion momentum (discussed below) for which the golden selection is used. Other than in the regions excluded from the signal definition, the efficiency of the generic selection is largely independent of the momenta and directions of the final-state particles<sup>1</sup>, with a slight selection preference toward forward-going ( $\cos(\theta) > 0.5$ ) particles. The angular variables,  $\cos(\theta_\mu)$ ,  $\cos(\theta_\pi)$  and  $\theta_{\mu\pi}$  are well-estimated and have small off-diagonal smearing matrix elements. The angular bins were selected such that the standard deviation of the reconstructed values within a given true bin is at most half the bin width. The resolution of the muon momentum reconstruction is greatest at higher momenta.

The measurement of the differential cross-section with respect to the pion momentum, shown in Figure 7.8, is a world's first using argon as the nuclear target. This measurement is challenging as the pion commonly undergoes secondary interactions in which energy is lost. To mitigate these effects, the golden selection is utilised which aims to enhance the fraction of selected events with a pion that stops in the detector without scattering. The momentum of these *golden pions* can be estimated by the distance travelled before stopping. The independence of the efficiency of the golden selection with respect to the momenta and directions of the final-state particles largely follows that of the generic selection. However, as the likelihood of a secondary interaction increases with the pion's momentum, this selection (by construction) preferentially selects pions with low-momenta ( $\lesssim 300 \text{ MeV c}^{-1}$ ) which tend to be less forward-going.

---

<sup>1</sup>The efficiency is indicated by the sum of the columns of the forward-folding matrix,  $\epsilon_t = \sum_r \xi_{rt}$ , and is given explicitly in Appendix A.



The resolution with which the pion momentum can be measured is limited by the efficacy of the golden selection. The fraction of selected events with a non-golden pion is estimated to be  $\sim 35\%$ . In these cases, the range-based pion momentum measurement is not accurate leading to the off-diagonal elements of the forward-folding matrix shown in Figure 7.8. Additionally, the entries in the lowest reconstructed pion momentum bin ( $100 \text{ MeV c}^{-1} < p'_\pi < 160 \text{ MeV c}^{-1}$ ) have significant contributions from events with a true pion momentum below  $100 \text{ MeV c}^{-1}$  for the reasons illustrated by Figure 7.3.

### 7.5.3 Future improvements

In future iterations of this analysis, the uncertainties on the data can be reduced in the following ways:

- The statistical uncertainty can be reduced by including other runs of MicroBooNE data. As shown in Figure 3.4, MicroBooNE has taken data for five run periods. However, due to availability at the time of analysis, only run-1 data was included in this thesis.
- The primary source of systematic uncertainty is due to the background prediction which can be reduced by including a *side-band* measurement. In particular, the leading  $1\mu + Xp$  background can be isolated by selecting events with one “non-proton”, as shown in Figure 6.19<sup>2</sup>. A measurement of these events can be used to constrain the uncertainties on the background prediction.

The uncertainties on the smearing matrices are primarily due to the detector modelling. However, in future iterations of this analysis, the uncertainty due to the GENIE model prediction could be further reduced by employing a different binning scheme in truth-space and reco-space. The binning presented in this analysis is the same in both spaces and is driven by the available data statistics and the resolution of the reconstruction. However, one could in principle choose finer bins in truth-space (giving a non-square forward-folding matrix), or indeed profile the smearing and efficiency over multiple true kinematic variables. The choice of binning in truth-space is ultimately limited by the available MC statistics. This approach is discussed in [107].

<sup>2</sup>Specifically, it is desirable to obtain a sample of  $1\mu + Xp$  whose distribution (with respect to the kinematics of the muon and proton) is very similar to that of the  $1\mu + Xp$  backgrounds of the generic CC1 $\pi^\pm$  selection. Recent studies by another MicroBooNE member have shown that such a sample can be obtained by restricting the events with one “non-proton” to those where a proton candidate has an appropriately low proton-BDT response. In this case, the selected  $1\mu + Xp$  events would have a “pion-like” proton of the kind likely to pass the CC1 $\pi^\pm$  selection.

It is the opinion of the author that future improvements to the resolution of the pion momentum measurement would most likely arise from advancements in the reconstruction of secondary interactions. As illustrated by Figure 6.7, the signal for a golden pion is a straight particle track with a Bragg-peak followed by a low-energy electron shower (produced in the decay of the pion,  $\pi \rightarrow \mu \rightarrow e$ , where neutrinos are not shown)<sup>3</sup>. If this signal was reliably identified by the reconstruction then the identification of golden pions could be improved.

At present, secondary interactions (such as the scattering or decay of the pion) are commonly missed by Pandora’s pattern-recognition algorithms. One reason this can occur is if the primary pion track is mistaken as a shower spine, in which case the secondary particles would be merged as shower branches. Alternately, scatters of the primary pion track through small angles may be insufficient to cause Pandora to split a cluster. If a pion decay is missed, the fitted track trajectory will encompass both the primary pion and the decay electron. The resulting reconstructed energy-loss curve will typically not have the expected Bragg-peak signature of a stopping pion. In this analysis, the calorimetric variables used by the golden-pion BDT are sensitive to the differences between golden and non-golden pions when the end-state secondary interaction (e.g. decay, absorption, charge-exchange) is correctly identified, whereas the topological variables are sensitive to secondary interactions that may be missed by the pattern-recognition. Improved identification of secondary interactions during the pattern-recognition stage would improve the performance of the calorimetric features at the analysis-level. This should allow for a better separation between golden and non-golden pions and hence a better pion momentum resolution. Dedicated work on secondary interaction vertex finding would benefit this and other analyses and is under consideration by Pandora developers.

---

<sup>3</sup>This signal also arises for muons that decay at rest, however this background is largely irreducible.

# Chapter 8

## Conclusions

In the 91 years since Pauli’s letter [8], proposing the existence of an elusive new particle, neutrinos from a wide range of sources have been measured by numerous experiments around the globe. The unexpected observation of neutrino flavour oscillations implies that these particles have mass, in contrast to the assumptions of the Standard Model. In general, these observations are well-explained by oscillations between the three known active neutrino flavours,  $\nu_e$ ,  $\nu_\mu$ ,  $\nu_\tau$ . Many of the parameters that govern these oscillations have been identified and the measurement of the remaining parameters is one of the primary goals of the next generation of neutrino oscillation experiments. However, existing anomalous results [3] from the LSND and MiniBooNE experiments are in tension with this three-neutrino paradigm and are currently being confronted by the MicroBooNE experiment.

The focus of this thesis is MicroBooNE’s capacity to probe the complex interactions of neutrinos on large nuclear targets. It is critical for the success of future neutrino oscillation experiments that these interactions are well understood. As discussed in Chapter 2, neutrino-nucleus interactions can occur via a wide range of processes and result in numerous final state topologies. Although various theoretical models of these interaction processes exist, more experimental data is required to validate, tune, and constrain their uncertainties. The MicroBooNE detector is a Liquid Argon Time Projection Chamber which is capable of imaging neutrino interactions with millimetre-scale resolution, as described in Chapter 3. Consequently, it is ideally suited to measure neutrino-nucleus interactions. This technology will also be used by future short and long-baseline neutrino oscillation experiments, which will build upon many of the analysis procedures developed by MicroBooNE.

One key component of any such analysis is pattern recognition. The Pandora project [95], described in Chapter 4, was developed to facilitate the development and deployment of pattern-recognition algorithms and those created for MicroBooNE [96] will also serve future LArTPC experiments. In particular, Pandora’s consolidated approach (developed in part by the author of this thesis) allows for neutrino interactions to be reconstructed in dense cosmic-ray environments. Pandora’s cosmic-ray tagging algorithm was developed by the author of this thesis and is described in Chapter 5. Cosmic rays are the primary background to neutrino interactions for surface-based LArTPC experiments and so this stage is critical to the success of Pandora-based analyses using MicroBooNE and other detectors. The approach presented in this thesis is capable of identifying 45.8% of reconstructed particles that represent cosmic rays, with a neutrino-induced particle misidentification rate of 1.7%.

The primary research work presented in this thesis regards the measurement of neutrino-argon interactions in MicroBooNE that result in the  $CC1\pi^\pm$  topology:  $1\mu^\pm + 1\pi^\pm + N_p p + N_n n$ , where the nucleon multiplicities  $N_p$  and  $N_n$  can take any value including zero. This topology is of particular importance for future neutrino oscillation experiments and limited neutrino-argon data is currently available. A Pandora-based selection of these interactions is presented in Chapter 6 which is built upon a CC-inclusive pre-selection developed by MicroBooNE that makes use of the aforementioned CR-tagging algorithm. The  $CC1\pi^\pm$  selection is capable of removing 96.8% of backgrounds that pass the pre-selection leaving a sample of 1857 events from MicroBooNE’s first run with an exposure of  $1.455 \times 10^{20}$  POT. Of the events selected, 52.4% are predicted to have the  $CC1\pi^\pm$  topology. In addition, a sub-sample of these events is identified which has an enhanced fraction (65.5%) of *golden* pions that stop in the detector without undergoing any secondary interactions.

Using this sample of events, the  $CC1\pi^\pm$  flux-integrated forward-folded cross-section on argon was measured and is presented in Chapter 7. The total forward-folded cross-section was found to be  $[22.4 \pm 0.9(\text{stat.}) \pm 5.2(\text{syst.})] \times 10^{-41} \text{ cm}^2$ , with an efficiency  $[18.8 \pm 1.3] \%$  and is consistent with the forward-folded GENIE [90] prediction of  $[17.2 \pm 1.3] \times 10^{-41} \text{ cm}^2$  with a p-value of 0.33. In addition the proton-exclusive total cross-section was found for proton multiplicities of  $N_p = 0, 1$  and  $N_p > 1$  with momenta above  $300 \text{ MeV } c^{-1}$ , constituting a world’s first.

Finally, the  $CC1\pi^\pm$  differential cross-section was measured with respect to the momenta and directions of the muon and pion. The measurement with respect to the pion momentum on argon is also a world’s first and was facilitated by the aforementioned

golden-enhanced sub-sample. These measurements are important for neutrino oscillation experiments as cross section modelling is, at present, one of their dominant sources of systematic uncertainty [28]. New cross section data for specific exclusive topologies, such as the measurements presented in this thesis, will be used to constrain these uncertainties and help to determine if existing simulations of neutrino interactions are sufficiently well modelled, or if major changes are required for future analyses. In particular, the DUNE experiment will measure neutrinos with energies where the  $\text{CC}1\pi^\pm$  topology is common [28]. Given the difficulty to differentiate muons and pions in LArTPC detectors, the techniques developed in this thesis can be used by future experiments to successfully identify  $\text{CC}1\pi^\pm$  using this technology.

Further improvements to this analysis that were not possible to complete on the timescale of this thesis are already underway by other MicroBooNE collaborators. In particular, a  $1\mu + Xp$  background control sample will be employed to constrain the primary source of uncertainty and data from other runs will be added before the measurement is published. In addition, MicroBooNE data will be compared against predictions made by multiple generators (in addition to GENIE) and made available for these generators to tune their models.



# Colophon

This thesis was made in  $\text{\LaTeX}$  using the “hepthesis” class [113].





# Bibliography

- [1] LSND Collaboration, A. Aguilar *et al.*, *Evidence for neutrino oscillations from the observation of  $\bar{\nu}_e$  appearance in a  $\bar{\nu}_\mu$  beam*, Phys. Rev. D **64**, 112007 (2001).
- [2] MiniBooNE, A. A. Aguilar-Arevalo *et al.*, *A Search for Electron Antineutrino Appearance at the  $\Delta m^2 \sim 1$  eV<sup>2</sup> Scale*, Phys. Rev. Lett. **103**, 111801 (2009), arXiv:0904.1958.
- [3] J. M. Conrad, W. C. Louis, and M. H. Shaevitz, *The LSND and MiniBooNE Oscillation Searches at High  $\Delta m^2$* , Ann. Rev. Nucl. Part. Sci. **63**, 45 (2013), arXiv:1306.6494.
- [4] MicroBooNE, R. Acciarri *et al.*, *Design and Construction of the MicroBooNE Detector*, JINST **12**, P02017 (2017), arXiv:1612.05824.
- [5] Weizmann Institute of Science, *The Scanners' Story*, (2014), <https://wiswander.weizmann.ac.il/space-physics/scannersOnline>, accessed 30-Apr-2021.
- [6] MicroBooNE, P. Abratenko *et al.*, *Measurement of the atmospheric muon rate with the MicroBooNE Liquid Argon TPC*, JINST **16**, P04004 (2021), arXiv:2012.14324.
- [7] D. Caratelli, *Study of Electromagnetic Interactions in the MicroBooNE Liquid Argon Time Projection Chamber*, PhD thesis, Columbia U., 2018.
- [8] W. Pauli, *Dear radioactive ladies and gentlemen*, Phys. Today **31N9**, 27 (1978).
- [9] C. D. Ellis, W. A. Wooster, and E. Rutherford, *The average energy of disintegration of radium E*, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **117**, 109 (1927), <https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1927.0168>.
- [10] F. L. Wilson, *Fermi's Theory of Beta Decay*, American Journal of Physics **36**, 1150 (1968), <https://doi.org/10.1119/1.1974382>.

- [11] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, *Detection of the Free Neutrino: a Confirmation*, Science **124**, 103 (1956), <http://science.sciencemag.org/content/124/3212/103.full.pdf>.
- [12] G. Danby *et al.*, *Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos*, Phys. Rev. Lett. **9**, 36 (1962).
- [13] M. L. Perl *et al.*, *Evidence for Anomalous Lepton Production in  $e^+ - e^-$  Annihilation*, Phys. Rev. Lett. **35**, 1489 (1975).
- [14] DONUT, K. Kodama *et al.*, *Observation of tau neutrino interactions*, Phys. Lett. **B504**, 218 (2001), arXiv:hep-ex/0012035.
- [15] D. DeCamp *et al.*, *Determination of the number of light neutrino species*, Physics Letters B **231**, 519 (1989).
- [16] R. Davis, D. S. Harmer, and K. C. Hoffman, *Search for Neutrinos from the Sun*, Phys. Rev. Lett. **20**, 1205 (1968).
- [17] K. S. Hirata *et al.*, *Observation of  $^8\text{B}$  solar neutrinos in the Kamiokande-II detector*, Phys. Rev. Lett. **63**, 16 (1989).
- [18] P. Anselmann *et al.*, *Solar neutrinos observed by GALLEX at Gran Sasso*, Physics Letters B **285**, 376 (1992).
- [19] J. Abdurashitov *et al.*, *Results from SAGE (The Russian-American gallium solar neutrino experiment)*, Physics Letters B **328**, 234 (1994).
- [20] Y. Fukuda *et al.*, *Measurements of the Solar Neutrino Flux from Super-Kamiokandes First 300 Days*, Physical Review Letters **81**, 11581162 (1998).
- [21] L. Bonolis, *Bruno Pontecorvo: From slow neutrons to oscillating neutrinos*, American Journal of Physics - AMER J PHYS **73**, 487 (2005).
- [22] SNO, Q. R. Ahmad *et al.*, *Measurement of the rate of  $\nu_e + d \rightarrow p + p + e^-$  interactions produced by  $^8\text{B}$  solar neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **87**, 071301 (2001), arXiv:nucl-ex/0106015.
- [23] M. Nakahata *et al.*, *Atmospheric Neutrino Background and Pion Nuclear Effect for KAMIOKA Nucleon Decay Experiment*, Journal of the Physical Society of Japan **55**, 3786 (1986), <https://doi.org/10.1143/JPSJ.55.3786>.
- [24] Super-Kamiokande Collaboration, Y. Fukuda *et al.*, *Evidence for Oscillation of*

- Atmospheric Neutrinos*, Phys. Rev. Lett. **81**, 1562 (1998).
- [25] A. Strumia, *Baryogenesis via leptogenesis*, in *Les Houches Summer School on Theoretical Physics: Session 84: Particle Physics Beyond the Standard Model*, 2006, hep-ph/0608347.
- [26] T2K Collaboration, *Measurements of neutrino oscillation in appearance and disappearance channels by the T2K experiment with  $6.6 \times 10^{20}$  protons on target*, Phys. Rev. D **91**, 072010 (2015).
- [27] T2K Collaboration, *Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations*, Nature **580**, 339 (2020).
- [28] DUNE, R. Acciarri *et al.*, *Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE): Conceptual Design Report, Volume 2: The Physics Program for DUNE at LBNF*, (2015), arXiv:1512.06148.
- [29] M. Thomson, *Modern particle physics*, Cambridge University Press, 2013.
- [30] R. Samanta, M. Chakraborty, and A. Ghosal, *Evaluation of the Majorana phases of a general Majorana neutrino mass matrix: Testability of hierarchical flavour models*, Nuclear Physics B **904**, 86 (2016).
- [31] P. F. de Salas *et al.*, *2020 Global reassessment of the neutrino oscillation picture*, (2020), arXiv:2006.11237.
- [32] M. Sajjad Athar, S. K. Singh, *The Physics of Neutrino Interactions*, Cambridge University Press, 2020.
- [33] Particle Data Group, *Review of Particle Physics*, Progress of Theoretical and Experimental Physics **2020** (2020), <https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf>, 083C01.
- [34] J. A. Formaggio and G. P. Zeller, *From eV to EeV: Neutrino Cross Sections Across Energy Scales*, Rev. Mod. Phys. **84**, 1307 (2012), arXiv:1305.7513.
- [35] C. Llewellyn Smith, *Neutrino reactions at accelerator energies*, Physics Reports **3**, 261 (1972).
- [36] D. Rein and L. M. Sehgal, *Neutrino-excitation of baryon resonances and single pion production*, Annals of Physics **133**, 79 (1981).
- [37] GENIE Collaboration, J. Tena-Vidal *et al.*, *Neutrino-nucleon cross-section model*

- tuning in GENIE v3*, Phys. Rev. D **104**, 072009 (2021).
- [38] R. P. Feynman, M. Kislinger, and F. Ravndal, *Current Matrix Elements from a Relativistic Quark Model*, Phys. Rev. D **3**, 2706 (1971).
- [39] K. S. Kuzmin, V. V. Lyubushkin, and V. A. Naumov, *Lepton polarization in neutrino nucleon interactions*, Mod. Phys. Lett. A **19**, 2815 (2004), arXiv:hep-ph/0312107.
- [40] C. Berger and L. M. Sehgal, *Lepton mass effects in single pion production by neutrinos*, Phys. Rev. D **76**, 113004 (2007), arXiv:0709.4378.
- [41] K. M. Graczyk and J. T. Sobczyk, *Lepton mass effects in weak charged current single pion production*, Phys. Rev. D **77**, 053003 (2008).
- [42] MiniBooNE, J. A. Nowak, *Four Momentum Transfer Discrepancy in the Charged Current  $\pi^+$  Production in the MiniBooNE: Data vs. Theory*, AIP Conf. Proc. **1189**, 243 (2009), arXiv:0909.3659.
- [43] A. Bodek and U.-k. Yang, *Axial and Vector Structure Functions for Electron- and Neutrino- Nucleon Scattering Cross Sections at all  $Q^2$  using Effective Leading order Parton Distribution Functions*, (2010), arXiv:1011.6592.
- [44] C. Andreopoulos *et al.*, *The GENIE Neutrino Monte Carlo Generator: Physics and User Manual*, (2015), arXiv:1510.05494.
- [45] A. A. Aguilar-Arevalo *et al.*, *First measurement of the muon neutrino charged current quasielastic double differential cross section*, Physical Review D **81** (2010).
- [46] J. Nieves, J. E. Amaro, and M. Valverde, *Inclusive quasi-elastic neutrino reactions*, Phys. Rev. C **70**, 055503 (2004), arXiv:nucl-th/0408005, [Erratum: Phys.Rev.C **72**, 019902 (2005)].
- [47] J. Nieves, I. Ruiz Simo, and M. J. Vicente Vacas, *Inclusive Charged-Current Neutrino-Nucleus Reactions*, Phys. Rev. C **83**, 045501 (2011), arXiv:1102.2777.
- [48] M. Martini, *Two Particle-Two Hole Excitations in Charged Current Quasielastic Neutrino-Nucleus Interactions*, J. Phys. Conf. Ser. **408**, 012041 (2013), arXiv:1110.5895.
- [49] MicroBooNE Collaboration, *Neutrino Interaction Model and Uncertainties for MicroBooNE Analyses - Public note 1074*, (2020),

- <https://microboone.fnal.gov/wp-content/uploads/MICROBOONE-NOTE-1074-PUB.pdf>.
- [50] S. J. Barish *et al.*, *Observation of Single Pion Production by a Weak Neutral Current*, Phys. Rev. Lett. **33**, 448 (1974).
  - [51] M. Derrick *et al.*, *Study of single-pion production by weak neutral currents in low-energy  $\nu d$  interactions*, Phys. Rev. D **23**, 569 (1981).
  - [52] ArgoNeuT, R. Acciarri *et al.*, *First Measurement of Neutrino and Antineutrino Coherent Charged Pion Production on Argon*, Phys. Rev. Lett. **113**, 261801 (2014), arXiv:1408.0598, [Erratum: Phys.Rev.Lett. 114, 039901 (2015)].
  - [53] K2K, M. Hasegawa *et al.*, *Search for coherent charged pion production in neutrino-carbon interactions*, Phys. Rev. Lett. **95**, 252301 (2005), arXiv:hep-ex/0506008.
  - [54] MINERvA, A. Higuera *et al.*, *Measurement of Coherent Production of  $\pi^\pm$  in Neutrino and Antineutrino Beams on Carbon from  $E_\nu$  of 1.5 to 20 GeV*, Phys. Rev. Lett. **113**, 261802 (2014), arXiv:1409.3835.
  - [55] MINERvA, A. Mislivec *et al.*, *Measurement of total and differential cross sections of neutrino and antineutrino coherent  $\pi^\pm$  production on carbon*, Phys. Rev. D **97**, 032014 (2018), arXiv:1711.01178.
  - [56] SciBooNE, K. Hiraide *et al.*, *Search for Charged Current Coherent Pion Production on Carbon in a Few-GeV Neutrino Beam*, Phys. Rev. D **78**, 112004 (2008), arXiv:0811.0369.
  - [57] T2K, K. Abe *et al.*, *Measurement of Coherent  $\pi^+$  Production in Low Energy Neutrino-Carbon Scattering*, Phys. Rev. Lett. **117**, 192501 (2016), arXiv:1604.04406.
  - [58] C. Anderson *et al.*, *The ArgoNeuT detector in the NuMI low-energy beam line at Fermilab*, Journal of Instrumentation **7**, P10019P10019 (2012).
  - [59] L. Aliaga *et al.*, *Design, calibration, and performance of the MINERvA detector*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **743**, 130159 (2014).
  - [60] P. Adamson *et al.*, *The NuMI neutrino beam*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **806**, 279306 (2016).

- [61] ArgoNeuT, R. Acciarri *et al.*, *First measurement of the cross section for  $\nu_\mu$  and  $\bar{\nu}_\mu$  induced single charged pion production on argon using ArgoNeuT*, Phys. Rev. D **98**, 052002 (2018), arXiv:1804.10294.
- [62] MINERvA, B. Eberly *et al.*, *Charged Pion Production in  $\nu_\mu$  Interactions on Hydrocarbon at  $\langle E_\nu \rangle = 4.0$  GeV*, Phys. Rev. D **92**, 092008 (2015), arXiv:1406.6415.
- [63] MINERvA, C. McGivern *et al.*, *Cross sections for  $\nu_\mu$  and  $\bar{\nu}_\mu$  induced pion production on hydrocarbon in the few-GeV region using MINERvA*, Phys. Rev. D **94**, 052005 (2016), arXiv:1606.07127.
- [64] MINERvA, T. Le *et al.*, *Measurement of  $\bar{\nu}_\mu$  Charged-Current Single  $\pi^-$  Production on Hydrocarbon in the Few-GeV Region using MINERvA*, Phys. Rev. D **100**, 052008 (2019), arXiv:1906.08300.
- [65] MiniBooNE, A. Aguilar-Arevalo *et al.*, *Measurement of Neutrino-Induced Charged-Current Charged Pion Production Cross Sections on Mineral Oil at  $E_\nu \sim 1$  GeV*, Phys. Rev. D **83**, 052007 (2011), arXiv:1011.3572.
- [66] K2K, A. Rodriguez *et al.*, *Measurement of single charged pion production in the charged-current interactions of neutrinos in a 1.3-GeV wide band beam*, Phys. Rev. D **78**, 032003 (2008), arXiv:0805.0186.
- [67] T2K, K. Abe *et al.*, *Measurement of the muon neutrino charged-current single  $\pi^+$  production on hydrocarbon using the T2K off-axis near detector ND280*, Phys. Rev. D **101**, 012007 (2020), arXiv:1909.03936.
- [68] T2K, K. Abe *et al.*, *First measurement of the muon neutrino charged current single pion production cross section on water with the T2K near detector*, Phys. Rev. D **95**, 012010 (2017), arXiv:1605.07964.
- [69] MiniBooNE, A. Aguilar-Arevalo *et al.*, *Measurement of the  $\nu_\mu$  charged current  $\pi^+$  to quasi-elastic cross section ratio on mineral oil in a 0.8-GeV neutrino beam*, Phys. Rev. Lett. **103**, 081801 (2009), arXiv:0904.3159.
- [70] MINERvA collaboration, [https://minerva.fnal.gov/wp-content/uploads/2017/03/Updated\\_1pi\\_data.pdf](https://minerva.fnal.gov/wp-content/uploads/2017/03/Updated_1pi_data.pdf), 2017, Online, accessed 14-Dec-2021.
- [71] DUNE Collaboration, *Long-baseline neutrino oscillation physics potential of the DUNE experiment*.
- [72] MicroBooNE, LAr1-ND, ICARUS-WA104, M. Antonello *et al.*, *A Proposal for*

- a Three Detector Short-Baseline Neutrino Oscillation Program in the Fermilab Booster Neutrino Beam*, (2015), arXiv:1503.01520.
- [73] P. Adamson *et al.*, *The NuMI Neutrino Beam*, Nucl. Instrum. Meth. **A806**, 279 (2016), arXiv:1507.06690.
- [74] MiniBooNE, A. A. Aguilar-Arevalo *et al.*, *Significant Excess of ElectronLike Events in the MiniBooNE Short-Baseline Neutrino Experiment*, Phys. Rev. Lett. **121**, 221801 (2018), arXiv:1805.12028.
- [75] P. A. Machado, O. Palamara, and D. W. Schmitz, *The Short-Baseline Neutrino Program at Fermilab*, Ann. Rev. Nucl. Part. Sci. **69** (2019), arXiv:1903.04608.
- [76] Google maps, Maxar Technologies, U.S Geological Survey, USDA Farm Service Agency, <https://www.google.com/maps/place/Fermilab/@41.8279164,-88.2665482,398a,35y,1.16h,61.28t/data=!3m1!1e3!4m5!3m4!1s0x880efc06a65b1cc3:0x1ad8682a94d33d69!8m2!3d41.840676!4d-88.2793933>, 2019, Online, accessed 14-Oct-2019.
- [77] MiniBooNE, A. A. Aguilar-Arevalo *et al.*, *The Neutrino Flux prediction at Mini-BooNE*, Phys. Rev. **D79**, 072002 (2009), arXiv:0806.1449.
- [78] ICARUS, S. Amoruso *et al.*, *Study of electron recombination in liquid argon with the ICARUS TPC*, Nucl. Instrum. Meth. **A523**, 275 (2004).
- [79] Y. Li *et al.*, *Measurement of Longitudinal Electron Diffusion in Liquid Argon*, Nucl. Instrum. Meth. **A816**, 160 (2016), arXiv:1508.07059.
- [80] B. J. P. Jones, *Sterile Neutrinos in Cold Climates*, <http://inspirehep.net/record/1389805/>, 2015, Online, accessed 19-Nov-2019.
- [81] MicroBooNE, R. Acciarri *et al.*, *Noise Characterization and Filtering in the MicroBooNE Liquid Argon TPC*, JINST **12**, P08003 (2017), arXiv:1705.07341.
- [82] MicroBooNE, C. Adams *et al.*, *Design and construction of the MicroBooNE Cosmic Ray Tagger system*, JINST **14**, P04004 (2019), arXiv:1901.02862.
- [83] D. Kaleko, *PMT Triggering and Readout for the MicroBooNE Experiment*, JINST **8**, C09009 (2013), arXiv:1308.3446.
- [84] MicroBooNE, *MICROBOONE-NOTE-1018-PUB: Study of Space Charge Effects in MicroBooNE*, <http://microboone.fnal.gov/wp-content/uploads/>

- MICROBOONE-NOTE-1018-PUB.pdf, 2016, Online, accessed 5-Dec-2019.
- [85] MicroBooNE, C. Adams *et al.*, *Calibration of the charge and energy loss per unit length of the MicroBooNE liquid argon time projection chamber using muons and protons*, JINST **15**, P03022 (2020), arXiv:1907.11736.
- [86] MicroBooNE Collaboration, *Booster Neutrino Flux Prediction at MicroBooNE - Public note 1031*, (2018),  
<https://microboone.fnal.gov/wp-content/uploads/MICROBOONE-NOTE-1031-PUB.pdf>.
- [87] J. R. Sanford and C. L. Wang, *Empirical formulas for particle production in P-Be collision between 10 and 35 BeV/c*, 1967,  
<https://www.bnl.gov/isd/documents/86527.pdf>.
- [88] HARP, Catanesi, T., Radicioni, E., Edgecock, R. et al., *Measurement of the production cross-section of positive pions in the collision of 8.9 GeV/c protons on beryllium*, Eur. Phys. J. C **52** (2007).
- [89] S. Agostinelli *et al.*, *Geant4 - a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506**, 250 (2003).
- [90] C. Andreopoulos *et al.*, *The GENIE Neutrino Monte Carlo Generator*, Nucl. Instrum. Meth. A **614**, 87 (2010), arXiv:0905.2517.
- [91] R. Gran, J. Nieves, F. Sanchez, and M. Vicente Vacas, *Neutrino-nucleus quasi-elastic and  $2p2h$  interactions up to 10 GeV*, Phys. Rev. D **88**, 113007 (2013), arXiv:1307.8105.
- [92] T2K Collaboration, Abe, K. and others, *Measurement of double-differential muon neutrino charged-current interactions on  $C_8H_8$  without pions in the final state using the T2K off-axis beam*, Phys. Rev. D **93**, 112012 (2016).
- [93] MicroBooNE, C. Adams *et al.*, *Ionization electron signal processing in single phase LArTPCs. Part I. Algorithm Description and quantitative evaluation with MicroBooNE simulation*, JINST **13**, P07006 (2018), arXiv:1802.08709.
- [94] MicroBooNE Collaboration, *Novel Approach for Evaluating Detector Systematics in the MicroBooNE LArTPC - Public note 1075*, (2020),  
<https://microboone.fnal.gov/wp-content/uploads/>



- MICROBOONE-NOTE-1075-PUB.pdf.
- [95] J. S. Marshall and M. A. Thomson, *The Pandora Software Development Kit for Pattern Recognition*, Eur. Phys. J. **C75**, 439 (2015), arXiv:1506.05348.
  - [96] MicroBooNE, 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. **C78**, 82 (2018), arXiv:1708.03135.
  - [97] MicroBooNE, C. Adams *et al.*, *Ionization electron signal processing in single phase LArTPCs. Part II. Data/simulation comparison and performance in MicroBooNE*, JINST **13**, P07007 (2018), arXiv:1804.02583.
  - [98] W. Van De Pontseele, *Search for Electron Neutrino Anomalies with the MicroBooNE Detector*, PhD thesis, 2020.
  - [99] MicroBooNE, P. Abratenko *et al.*, *Measurement of Space Charge Effects in the MicroBooNE LArTPC Using Cosmic Muons*, (2020), arXiv:2008.09765.
  - [100] MicroBooNE, C. Adams *et al.*, *A method to determine the electric field of liquid argon time projection chambers using a UV laser system and its application in MicroBooNE*, JINST **15**, P07010 (2020), arXiv:1910.01430.
  - [101] ArgoNeuT, R. Acciarri *et al.*, *A Study of Electron Recombination Using Highly Ionizing Particles in the ArgoNeuT Liquid Argon TPC*, JINST **8**, P08005 (2013), arXiv:1306.1712.
  - [102] A. Lister, *Constraint of systematic uncertainties in an electron neutrino search using muon neutrinos at MicroBooNE*, PhD thesis, Lancaster University, 2019.
  - [103] A. Smith, *Measurement of the muon neutrino charged current single pion cross section on argon with the MicroBooNE detector*, 2021, MicroBooNE technical note (document database number 33809).
  - [104] MicroBooNE, P. Abratenko *et al.*, *Determination of muon momentum in the MicroBooNE LArTPC using an improved model of multiple Coulomb scattering*, JINST **12**, P10010 (2017), arXiv:1703.06187.
  - [105] R. D. Cousins, S. J. May, and Y. Sun, *Should unfolded histograms be used to test hypotheses?*, (2016), arXiv:1607.07038.
  - [106] MicroBooNE, P. Abratenko *et al.*, *First Measurement of Inclusive Muon Neutrino*

- Charged Current Differential Cross Sections on Argon at  $E_\nu \sim 0.8$  GeV with the MicroBooNE Detector*, Phys. Rev. Lett. **123**, 131801 (2019), arXiv:1905.09694.
- [107] L. Koch, *A response-matrix-centred approach to presenting cross-section measurements*, Journal of Instrumentation **14**, P09013 (2019).
- [108] M. Koutras, *On the Generalized Noncentral Chi-Squared Distribution Induced by an Elliptical Gamma Law*, Biometrika **73**, 528 (1986).
- [109] W. He and N. Prabhu, *Jacobi Decomposition and Eigenvalues of Symmetric Matrices*, pp. 527–550.
- [110] B. P. Roe, *Statistical errors in Monte Carlo estimates of systematic errors*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **570**, 159 (2007).
- [111] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, Ann. Statist. **7**, 1 (1979).
- [112] KASCADE, D. Heck, *Extensive air shower simulations with CORSIKA and the influence of high-energy hadronic interaction models*, in *30th International Symposium on Multiparticle Dynamics*, pp. 252–259, 2000, astro-ph/0103073.
- [113] A. Buckley, *The hepthesis L<sup>A</sup>T<sub>E</sub>X class*.

# Appendix A

## CC1 $\pi^\pm$ cross-section binning selection

### A.1 Muon kinematics

Table A.1 and Figure A.1 show the binning scheme used for the muon momentum. A threshold momentum of  $150 \text{ MeV c}^{-1}$  is applied to the signal definition - which corresponds to the 20 cm muon-length cut in the CC-inclusive selection. Any events with a reconstructed muon momentum below this threshold are not considered in the cross-section calculation, and any events with a true muon momentum below the threshold are considered as background. The cross-section is not extracted for events with a reconstructed muon momentum above the uppermost bin-edge ( $1.5 \text{ GeV c}^{-1}$ ), but these events are still considered signal and do contribute as an overflow bin in the smearing matrix. Of the true CC1 $\pi^\pm$  events passing the generic selection, 1.3% are taken as background on account of their muon momentum. Table A.2, Figure A.2, Table A.3 and Figure A.3 show the binning scheme used for the muon  $\cos(\theta)$  and  $\phi$  respectively. No phase-space restrictions are explicitly made on the muon angles.

Bin	Edges / GeV c <sup>-1</sup>
Background	0 → 0.15
0	0.15 → 0.23
1	0.23 → 0.32
2	0.32 → 0.45
3	0.45 → 0.66
4	0.66 → 1.5
Overflow	1.5 → $\infty$

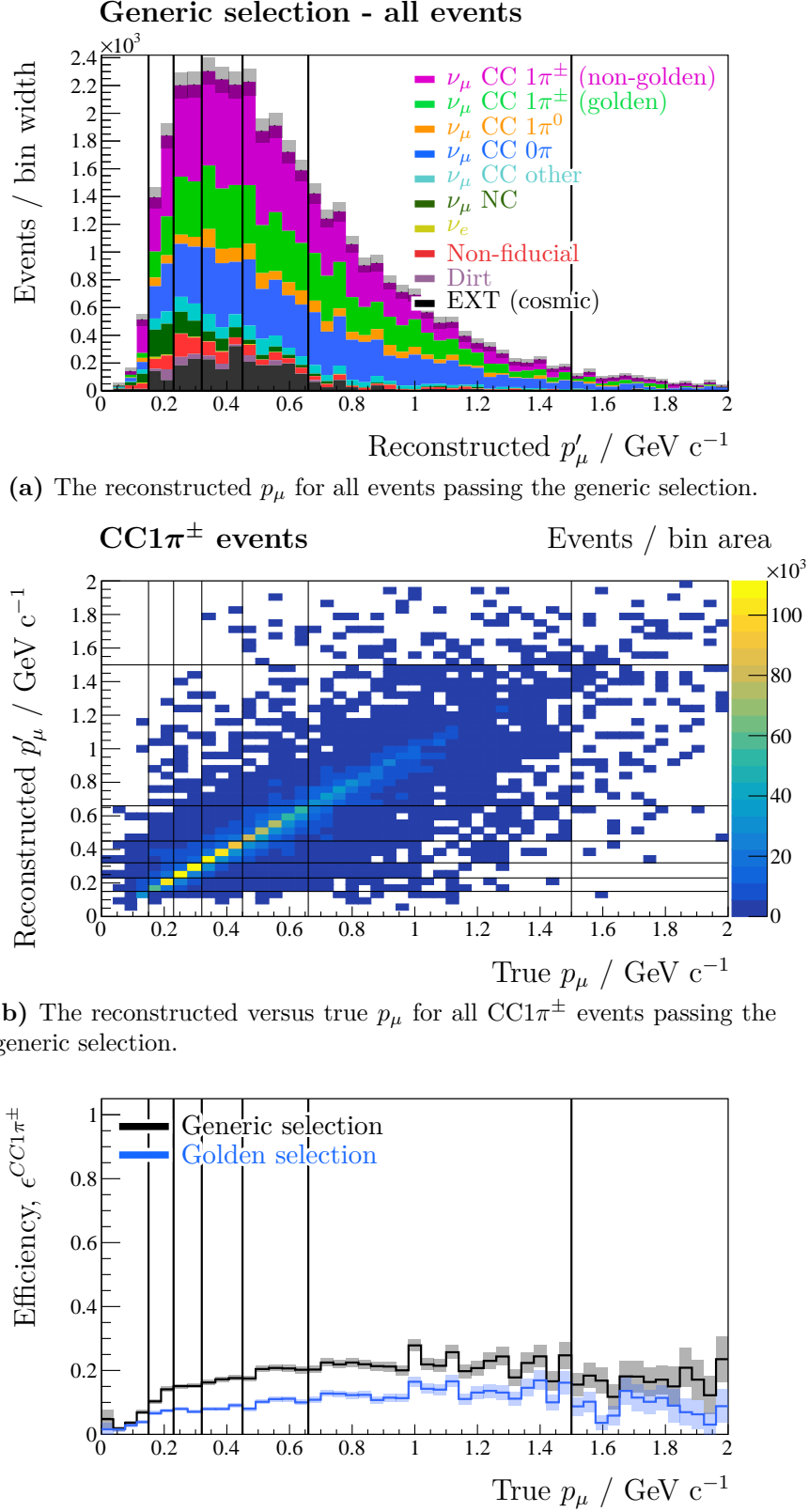
**Table A.1:** The muon momentum bin edges

Bin	Edges	Bin	Edges
0	-1.00 → -0.27	6	0.77 → 0.82
1	-0.27 → 0.29	7	0.82 → 0.88
2	0.29 → 0.46	8	0.88 → 0.93
3	0.46 → 0.58	9	0.93 → 0.97
4	0.58 → 0.67	10	0.97 → 1.00
5	0.67 → 0.77		

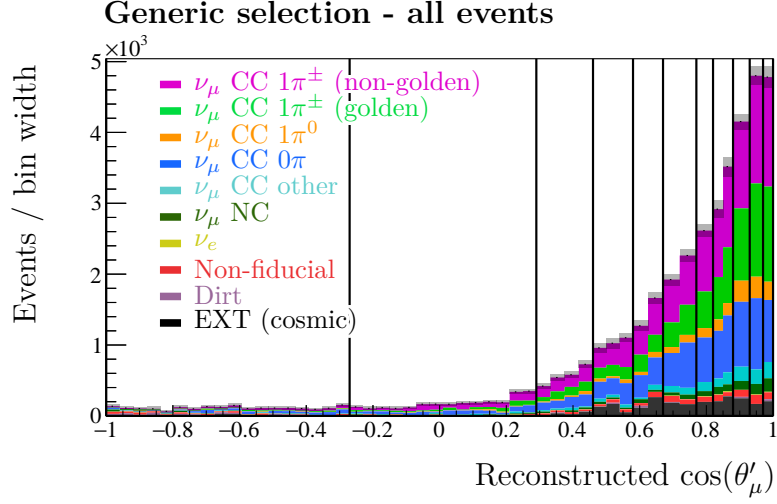
**Table A.2:** The muon  $\cos(\theta)$  bin edges

Bin	Edges / rad
0	-3.14 → -2.72
$\vdots$	
$i$	$\left(\frac{2i}{15} - 1\right)\pi \rightarrow \left(\frac{2(i+1)}{15} - 1\right)\pi$
$\vdots$	
14	2.72 → 3.14

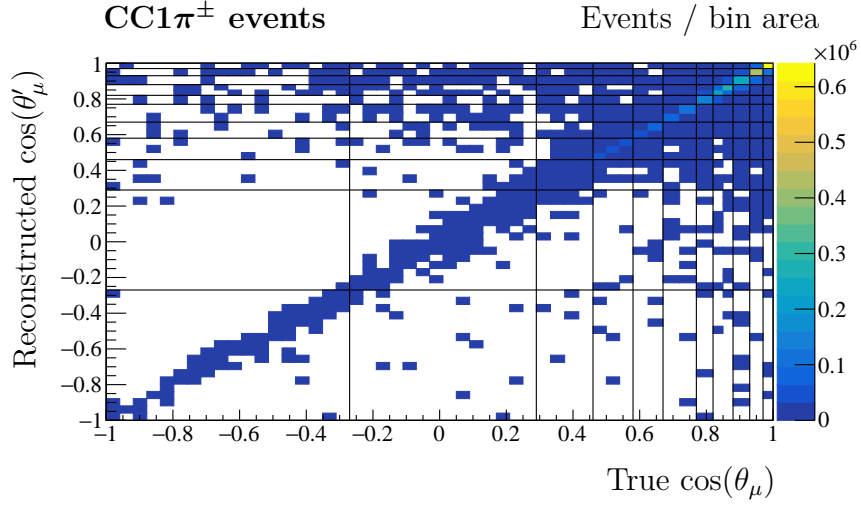
**Table A.3:** The muon  $\phi$  bin edges



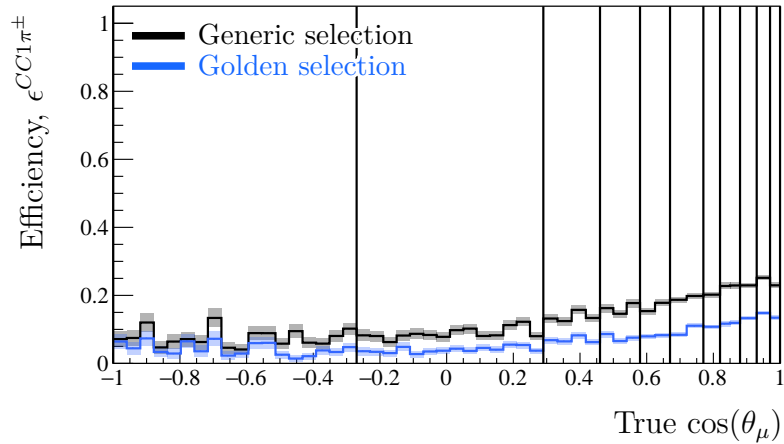
**Figure A.1:** The edges of the muon momentum bins in which the cross-section is extracted, shown as solid black lines.



(a) The reconstructed  $\cos(\theta_\mu)$  for all events passing the generic selection.



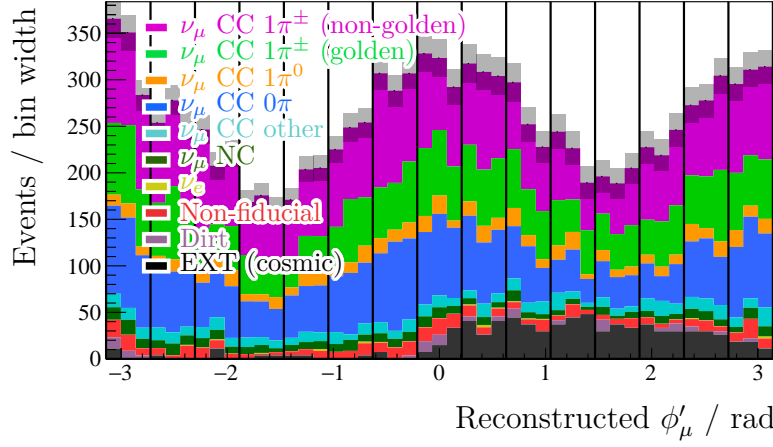
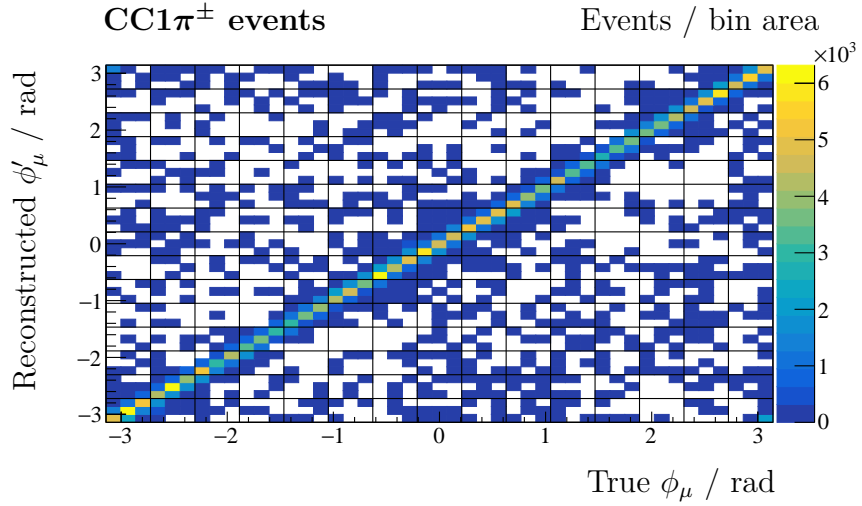
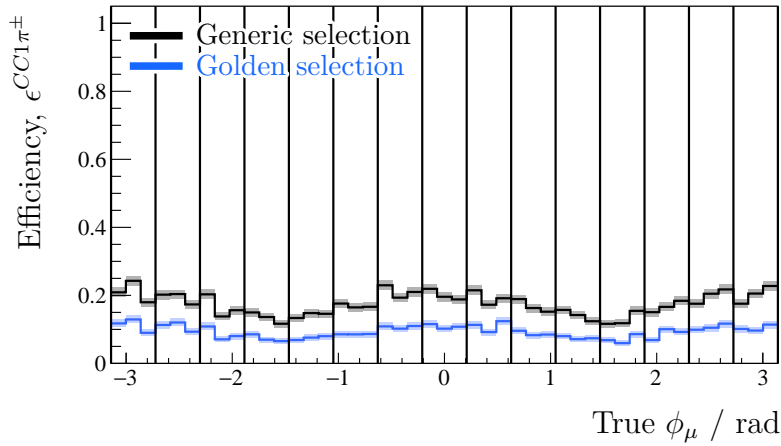
(b) The reconstructed versus true  $\cos(\theta_\mu)$  for all CC1 $\pi^\pm$  events passing the generic selection.



(c) The selection efficiency as a function of the true  $\cos(\theta_\mu)$ . The statistical uncertainty is indicated by the semitransparent bands.

**Figure A.2:** The edges of the muon  $\cos(\theta)$  bins in which the cross-section is extracted, shown as solid black lines.

## Generic selection - all events

(a) The reconstructed  $\phi_\mu$  for all events passing the generic selection.(b) The reconstructed versus true  $\phi_\mu$  for all CC1 $\pi^\pm$  events passing the generic selection.(c) The selection efficiency as a function of the true  $\phi_\mu$ . The statistical uncertainty is indicated by the semitransparent bands.**Figure A.3:** The edges of the muon  $\phi$  bins in which the cross-section is extracted, shown as solid black lines.

## A.2 Pion kinematics

Table A.4, Figure A.4 and Figure A.5 show the binning scheme used for the pion momentum. The cross-section is not extracted for events with a reconstructed pion momentum below  $100 \text{ MeV c}^{-1}$  or above  $600 \text{ MeV c}^{-1}$  however these events are still considered as signal and contribute as underflow and overflows bin in the smearing matrix. The differences between Figure A.4 (generic selection) and Figure A.5 (golden selection) show that, as desired, the golden selection acts to remove events with a non-golden pion for which the momentum reconstruction is inaccurate ( $p'_\pi < p_\pi$ ). The off-diagonal entries below  $\sim 150 \text{ MeV c}^{-1}$  (for which  $p'_\pi > p_\pi$ ), are explained in Section 7.4.1.

Table A.5, Figure A.6, Table A.6 and Figure A.7 show the binning scheme used for the pion  $\cos(\theta)$  and  $\phi$  respectively. No phase-space restrictions are explicitly made on the pion angles.

Bin	Edges / $\text{GeV c}^{-1}$
Underflow	0 $\rightarrow$ 0.10
0	0.10 $\rightarrow$ 0.16
1	0.16 $\rightarrow$ 0.19
2	0.19 $\rightarrow$ 0.22
3	0.22 $\rightarrow$ 0.60
Overflow	0.60 $\rightarrow$ $\infty$

**Table A.4:** The pion momentum bin edges

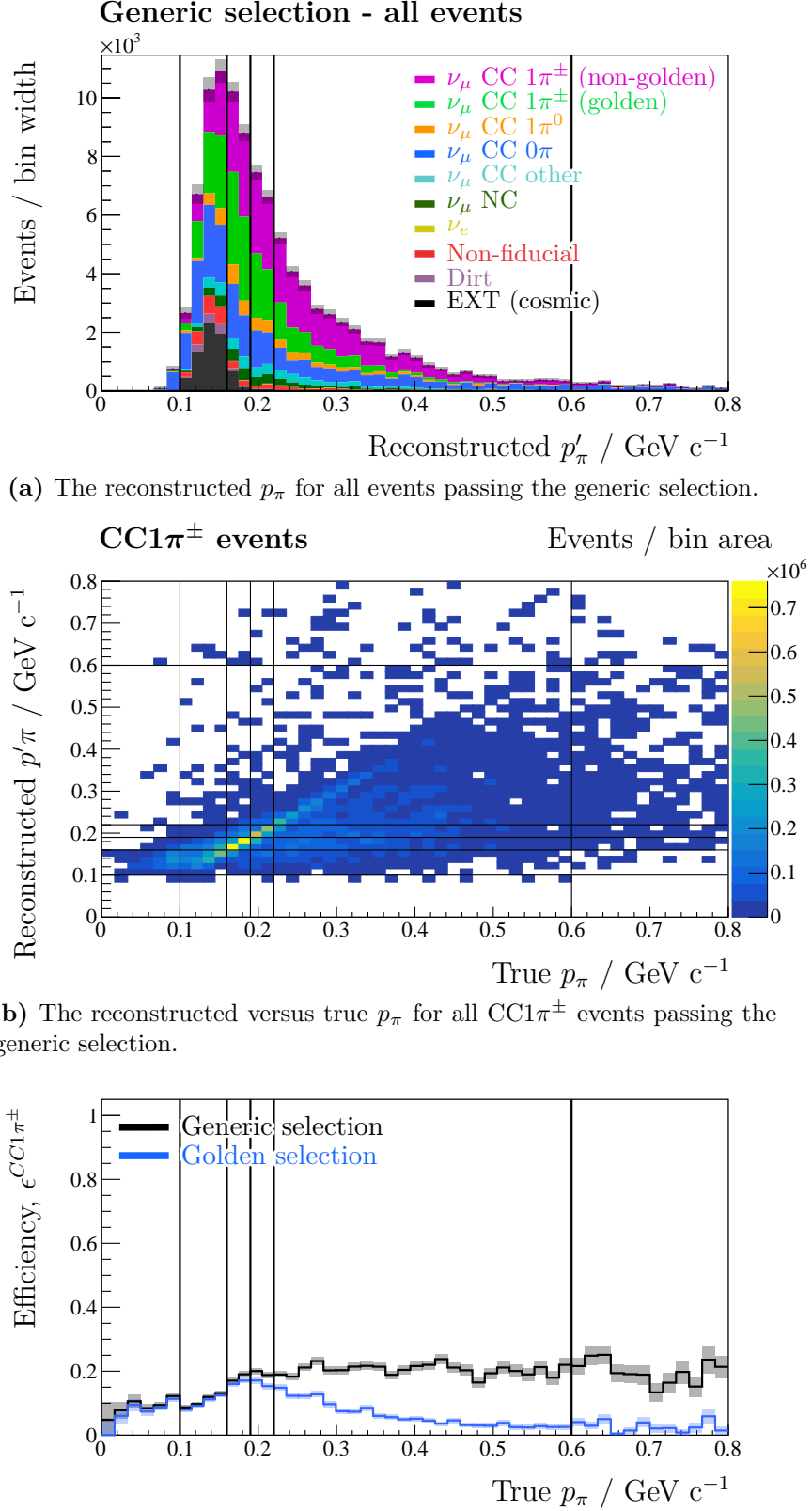
Bin	Edges
0	-1.00 $\rightarrow$ -0.47
1	-0.47 $\rightarrow$ 0.00
2	0.00 $\rightarrow$ 0.39
3	0.39 $\rightarrow$ 0.65
4	0.65 $\rightarrow$ 0.84
5	0.84 $\rightarrow$ 0.93
6	0.93 $\rightarrow$ 1.00

**Table A.5:** The pion  $\cos(\theta)$  bin edges

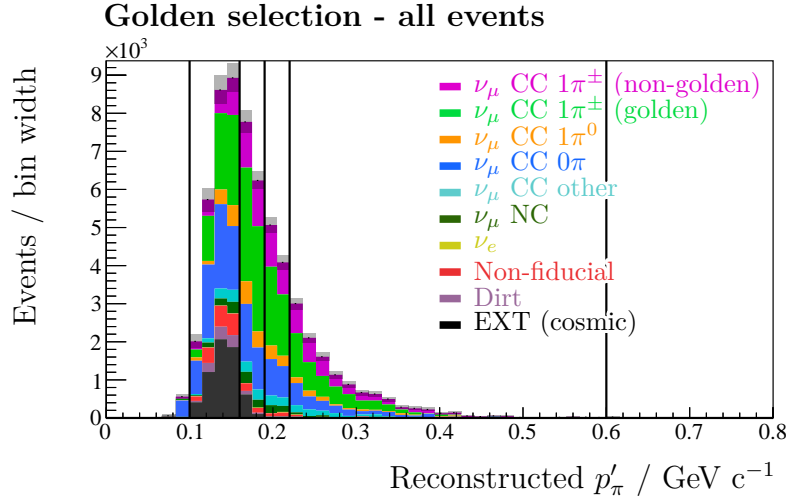
Bin	Edges / rad
0	-3.14 $\rightarrow$ -2.51
$\vdots$	
$i$	$\left(\frac{2i}{10} - 1\right) \pi \rightarrow \left(\frac{2(i+1)}{10} - 1\right) \pi$
$\vdots$	
9	2.51 $\rightarrow$ 3.14

**Table A.6:** The pion  $\phi$  bin edges

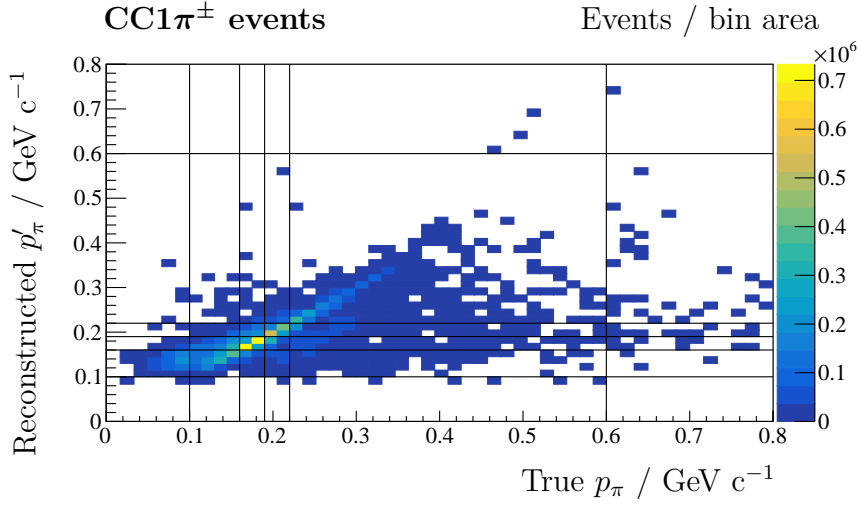




**Figure A.4:** The edges of the pion momentum bins in which the cross-section is extracted, shown as solid black lines.

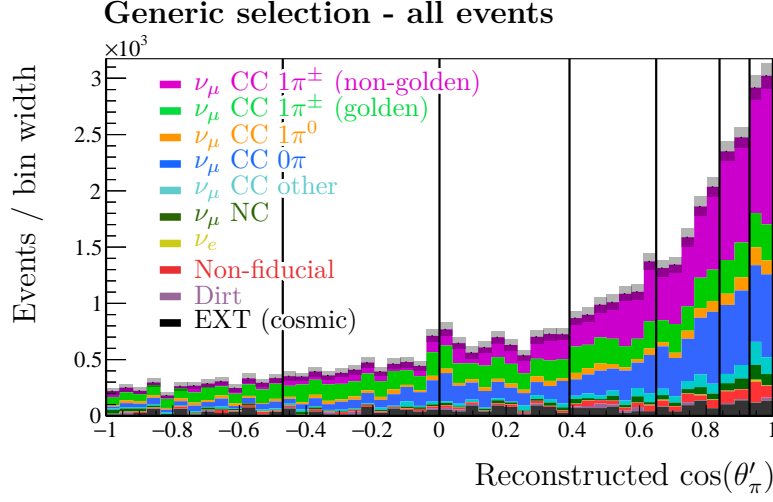
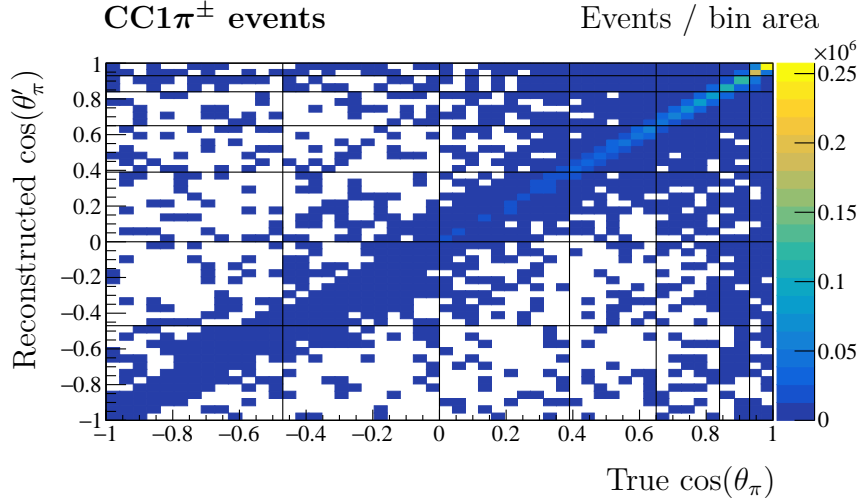
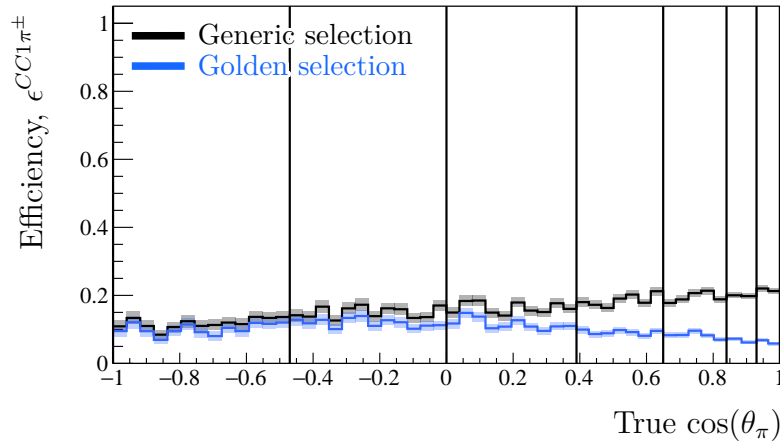


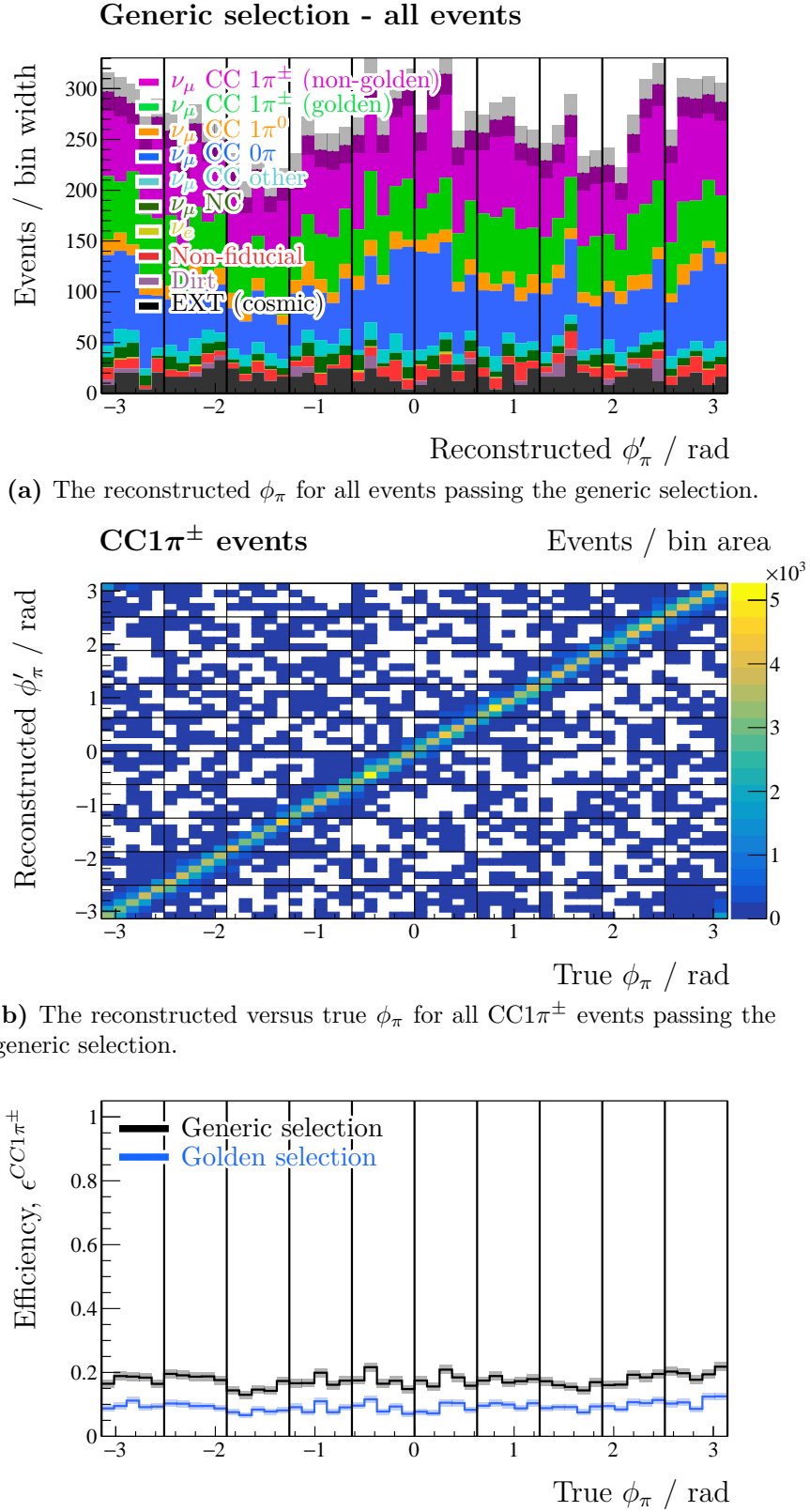
(a) The reconstructed  $p_\pi$  for all events passing the golden selection.



(b) The reconstructed versus true  $p_\pi$  for all CC1 $\pi^\pm$  events passing the golden selection.

**Figure A.5:** The edges of the pion momentum bins in which the cross-section is extracted, shown as solid black lines.

(a) The reconstructed  $\cos(\theta_\pi)$  for all events passing the generic selection.(b) The reconstructed versus true  $\cos(\theta_\pi)$  for all CC1 $\pi^\pm$  events passing the generic selection.(c) The selection efficiency as a function of the true  $\cos(\theta_\pi)$ . The statistical uncertainty is indicated by the semitransparent bands.**Figure A.6:** The edges of the pion  $\cos(\theta)$  bins in which the cross-section is extracted, shown as solid black lines.



**Figure A.7:** The edges of the pion  $\phi$  bins in which the cross-section is extracted, shown as solid black lines.

### A.3 Muon-pion opening angle

Figure A.8 shows the binning scheme used for the muon-pion opening angle,  $\theta_{\mu\pi}$ . Events in which a reconstructed  $\theta'_{\mu\pi} > 2.65$  are explicitly rejected by the opening angle cut, and so the cross-section is not extracted in this region. Any events with a true opening angle  $\theta_{\mu\pi} > 2.65$  that pass the selection are considered as a background. Of the true CC1 $\pi^\pm$  events passing the generic selection, 1.3% are taken as background on account of their opening angle.

Bin	Edges / rad
0	0.00 $\rightarrow$ 0.49
1	0.49 $\rightarrow$ 0.93
2	0.93 $\rightarrow$ 1.26
3	1.26 $\rightarrow$ 1.57
4	1.57 $\rightarrow$ 1.88
5	1.88 $\rightarrow$ 2.21
6	2.21 $\rightarrow$ 2.65
Background	2.65 $\rightarrow$ 3.14

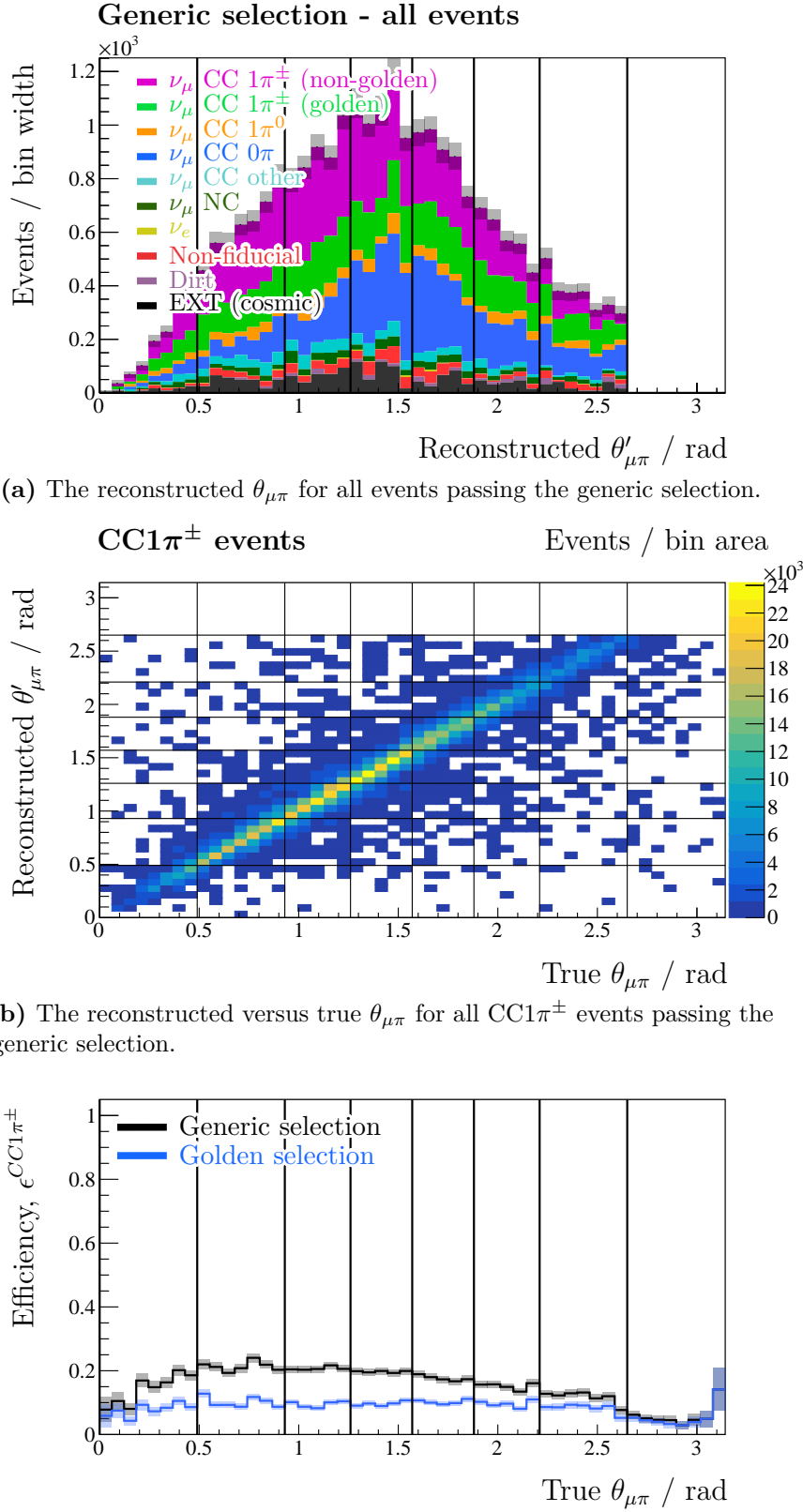
**Table A.7:** The muon-pion opening angle,  $\theta_{\mu\pi}$ , bin edges

### A.4 Proton multiplicity

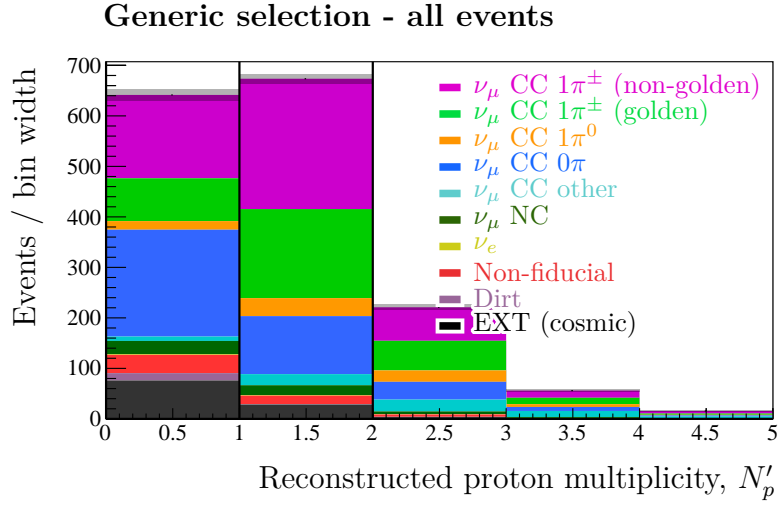
Figure A.9 shows the binning scheme used for the proton multiplicity,  $N_p$ . As described in Section 7.1.4, the proton-exclusive cross-sections are extracted in reconstructed-space in a similar fashion to the continuous kinematic variables. A true momentum threshold of 300 MeV  $c^{-1}$  is applied when counting the true proton multiplicity  $N_p$ .

Bin	Value(s)
0	0
1	1
2	$\geq 2$

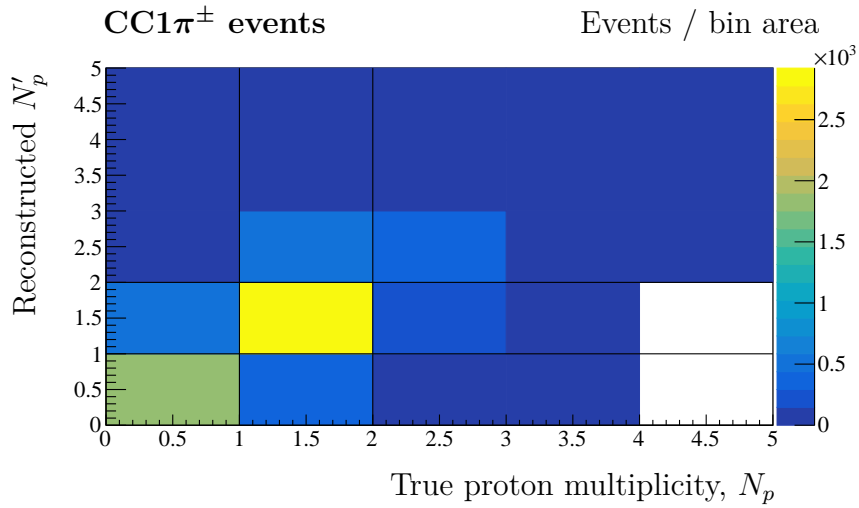
**Table A.8:** The proton multiplicity bin edges



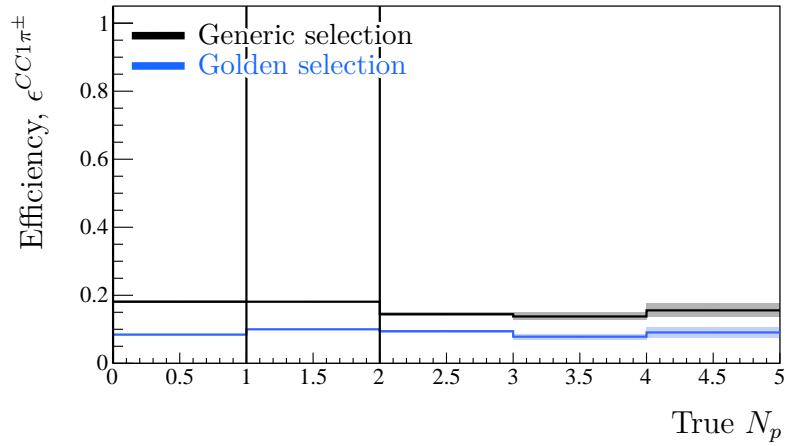
**Figure A.8:** The edges of the muon-pion angle bins in which the cross-section is extracted, shown as solid black lines.



(a) The reconstructed proton multiplicity for all events passing the generic selection.



(b) The reconstructed versus true  $N_p$  for all CC1 $\pi^\pm$  events passing the generic selection.



(c) The selection efficiency as a function of the true  $N_p$ . The statistical uncertainty is indicated by the semitransparent bands.

**Figure A.9:** The edges of the pion  $\cos(\theta)$  bins in which the cross-section is extracted, shown as solid black lines.