

Measuring Electron Diffusion and  
Constraining the Neutral Current  $\pi^0$   
Background for Single-Photon Events in  
MicroBooNE

A Dissertation Presented for the  
Doctor of Philosophy  
Degree

The University of Tennessee, Knoxville

Andrew J. Mogan

May 2021

© by Andrew J. Mogan, 2021  
All Rights Reserved.

*For Mom and Dad, for all their support over the years.*

# Acknowledgments

Where to even start? The past six years have been a wild ride: living in three different states, moving to work at a national lab, travelling to conferences across the country, a year-long global pandemic, and of course, the culmination of my research into this thesis. I couldn't have made it this far without the support of innumerable mentors, friends, and family.

First, I of course need to thank my advisor, Sowjanya Gollapinni. It was truly a stroke of good fortune that she happened to come to UTK right as I passed my qualifying exam. In addition to the usual advisor duties of guiding my graduate research, she also provided a healthy working environment, which I consider to be invaluable. I remember being told that I wouldn't be able to work a standard nine-to-five in graduate school, but I'm happy to report this to be false (with only a handful of exceptions). In fact, I firmly believe I've made it this far not *in spite* of working such "short" hours, but *because* I've been allowed to maintain a healthy mental state. For this and all of her advice and support over the years, I can't thank Sowjanya enough.

In addition to my on-paper advisor, in my mind I've had two other "co-advisors" during my time at Fermilab: Adam Lister and Mark Ross-Lonergan. These poor, unfortunate souls were shouldered with the unenviable task of mentoring a know-nothing student like me. Thanks to Adam for the continued support on the diffusion analysis, even after moving to a different experiment. We'll always maintain that the diffusion group is the *original* DL group. Once I joined the single-photon group, Mark was then forced to share the burden of teaching me such basics as git and statistical fitting methods, both of which still mystify me to this day. Both Adam and Mark showed the patience of a saint as I continuously asked stupid questions and made dumb mistakes, and I thank them for their constant patience and mentorship.



One of the best parts of working on-site at Fermilab (well, until that whole “global pandemic” thing, anyway) was the access to a swath of experts in the field. Thanks to Michelle Stancari and Bruce Baller for propelling the diffusion analysis toward an actual publication; to Wes Ketchum for explaining makefiles and other computing-related wizardry; to Joseph Zennamo for (loudly) explaining various aspects of our simulation; to Stephen Gardiner for dumbing nuclear physics down to my level; and to Maya Wospakrik, both for explaining our calibration methods and for supporting the diffusion analysis.

Believe it or not, there is, in fact, life outside of work, and I need to thank the many friends I’ve made at Fermilab for keeping me sane: Colton, for introducing me to *Magic: the Gathering* (my wallet, however, does not thank you); Andy and Rhiannon (“Rhiandy,” as they’re properly known), for the wild parties at Site 56 and for explaining the intricacies of British culture to an uncultured American; Davio and Michelle, for subjecting me to such visual nightmares as *Zardoz* and *Dead Alive*; Samantha, for seemingly being the only other person who can (almost) match my level of sass; Lauren, for knowing literally everything, always; Afro, for her unique form of what can only be described as “Greek energy”; Ivan, for being one of the few people around here who will actually play *Super Smash Bros.* with me; Kathryn, solely because I think she would hate getting a sappy shout-out here, and I think that’s funny; Avinay, for near-perfect attendance at the virtual socials (which were huge part of keeping me sane during quarantine); Vinnie B, for being the most “sus” liar I’ve ever seen in *Among Us*; Ohana, for the delicious Brazilian food and for hosting my post-defense celebration; and of course Gray, for putting up with living 10 feet away from me for three years, and for often being the only one to appreciate my incessant *Spongebob* references.

Perhaps even more shocking than the thought of life outside work is that of life outside of the Fermilab group. Here, I thank all of my non-physicist (or at least, non-neutrino physicist) friends for giving me a much-needed break from these insufferable nerds: Cody, for the longest-running friendship of my life (has it really been *fourteen* years?) and for carrying me through *Monster Hunter*; Candace, for her bubbly energy and hilarious color commentary while Cody and I play games; Zac and Mariah, for giving me a much-needed excuse to escape to my hometown during tough times (Etcetera coffeehouse, anyone?) and for all the delicious cooking; Caleb and Hannah, for helping me properly ring in these past

few New Years; Jesse, for the wild late-night adventures in Knoxville (I still feel bad about waking your neighbor up at 5 a.m.); and Michael and Chloe (“Michloe”), for providing me a bed and delicious food when I visit Knoxville, for introducing me to *Game Grumps*, and for the best wedding experience of my life to date (who could forget the tissue handoff of the century?).

And finally, what acknowledgements would be complete without thanking my loving and supportive family? Thanks to Mom and Dad for raising a weird, awkward kid into a slightly-less weird, slightly-less awkward adult, for the financial support through college, and for helping me move all of my furniture (which they bought me) to Illinois; to Rachel, for continuing to play online games with me and for the anime recommendations (I swear I’ll get around to *Little Witch Academia* someday); and to Rebecca, for all the fancy bourbon and for bringing Jake, the cutest pup ever, to all of our family gatherings.

# Abstract

Liquid Argon Time Projection Chambers (LArTPCs) are a rising technology in the field of experimental neutrino physics. LArTPCs use ionization electrons and scintillation light to reconstruct neutrino interactions with exceptional calorimetric and position resolution capabilities. Here, I present two analyses conducted in the MicroBooNE LArTPC at Fermilab: a measurement of the longitudinal electron diffusion coefficient,  $D_L$ , in the MicroBooNE detector and a constraint of the systematic uncertainty on MicroBooNE's single-photon analysis due to the dominant neutral current (NC)  $\pi^0$  background. Longitudinal electron diffusion modifies the spatial and timing resolution of the detector, and measuring it will help correct for these effects. Furthermore, current measurements of  $D_L$  in liquid argon are sparse and in tension with one another, making the MicroBooNE measurement especially valuable. We report a measurement of  $3.74^{+0.28}_{-0.29}$  cm<sup>2</sup>/s. MicroBooNE is searching for single-photon events as a potential explanation for the MiniBooNE low-energy excess (LEE) of electron neutrino-like events, which has been interpreted as evidence for low-mass sterile neutrinos. However, this search is overwhelmed by a large NC  $\pi^0$  background. By performing a sideband selection of NC  $\pi^0$  events, we apply a data-driven rate constraint to the single-photon analysis to reduce the systematic uncertainties. At present, this constraint improves the single-photon analysis' median sensitivity to the LEE-like signal from  $0.9\sigma$  to  $1.5\sigma$ . This sensitivity is expected to improve significantly as more data become available. Both of these measurements will not only benefit MicroBooNE, but also inform future LArTPC experiments.

# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Overview of Neutrino Physics</b>	<b>4</b>
2.1	Detecting the Undetectable . . . . .	4
2.2	The Standard Model . . . . .	5
2.3	Neutrino Oscillations . . . . .	7
2.4	Neutrino Interactions . . . . .	10
2.4.1	Quasielastic Interactions . . . . .	12
2.4.2	Resonant Interactions . . . . .	12
2.4.3	Other Interactions . . . . .	14
2.4.4	Summary . . . . .	15
2.5	Open Questions in Neutrino Physics . . . . .	16
<b>3</b>	<b>Neutrino Experiments</b>	<b>19</b>
3.1	Solar Neutrino Experiments . . . . .	20
3.2	Reactor Neutrino Experiments . . . . .	21
3.3	Atmospheric Neutrino Experiments . . . . .	21
3.4	Accelerator-Based Neutrino Experiments . . . . .	22
3.5	Short-Baseline Accelerator Anomalies . . . . .	24
<b>4</b>	<b>The MicroBooNE Experiment</b>	<b>30</b>
4.1	Booster Neutrino Beam at Fermilab . . . . .	31
4.2	The MicroBooNE Detector . . . . .	35

4.3	Data Acquisition Readout Electronics . . . . .	38
4.4	Ionization Signal . . . . .	40
4.5	Light Collection . . . . .	42
4.6	UV Laser System . . . . .	44
4.7	Cosmic Ray Tagger System . . . . .	44
4.8	Data-Taking Triggers . . . . .	47
4.9	Detector Operations . . . . .	48
<b>5</b>	<b>Simulation and Reconstruction in MicroBooNE</b>	<b>51</b>
5.1	Particle Generation and Propagation . . . . .	51
5.2	Signal Processing . . . . .	53
5.3	Reconstruction of Particle Objects . . . . .	56
5.3.1	Optical Signal Reconstruction . . . . .	56
5.3.2	Ionization Signal Reconstruction . . . . .	57
5.4	Calorimetry and Calibration . . . . .	61
<b>6</b>	<b>Measurement of Longitudinal Electron Diffusion in MicroBooNE</b>	<b>68</b>
6.1	Introduction . . . . .	69
6.2	Method . . . . .	69
6.2.1	Track Selection . . . . .	74
6.2.2	Waveform Selection . . . . .	75
6.2.3	Extraction of $D_L$ . . . . .	80
6.3	Measurement of Longitudinal Electron Diffusion . . . . .	80
6.3.1	Method Validation on Simulated Samples . . . . .	80
6.3.2	Measurement using CRT Data . . . . .	82
6.4	Evaluation of Systematic Uncertainties . . . . .	83
6.4.1	Transverse Diffusion . . . . .	83
6.4.2	Drift Velocity . . . . .	87
6.4.3	Detector Response Function . . . . .	88
6.4.4	Waveform Summation Method . . . . .	90
6.4.5	Summary and Other Systematic Uncertainties . . . . .	91

6.5	Discussion and Conclusions . . . . .	93
<b>7</b>	<b>The MicroBooNE Single-Photon Analysis</b>	<b>95</b>
7.1	NC $\Delta$ Resonance . . . . .	95
7.2	Analysis Overview . . . . .	98
7.2.1	Analysis Inputs . . . . .	99
7.2.2	Shower Energy Correction . . . . .	101
7.2.3	Notes on Nomenclature . . . . .	103
7.3	Boosted Decision Trees . . . . .	104
7.4	Single-Photon Selection . . . . .	109
<b>8</b>	<b>Constraining the NC <math>\pi^0</math> Background for MicroBooNE's Single-Photon Selection</b>	<b>115</b>
8.1	Inputs and Topological Selection . . . . .	116
8.2	Pre-Selection Cuts . . . . .	117
8.3	BNB Boosted Decision Tree . . . . .	120
8.4	Final Selections . . . . .	123
8.4.1	Breakdown of Final-Selection Backgrounds . . . . .	126
8.5	$\pi^0$ mass fit . . . . .	130
8.6	Run-to-Run Comparisons . . . . .	133
8.7	Systematic Uncertainties . . . . .	133
8.8	Constraint Results and Sensitivity Projection . . . . .	138
8.9	Summary . . . . .	144
<b>9</b>	<b>Conclusions</b>	<b>146</b>
	<b>Bibliography</b>	<b>149</b>
	<b>Appendix</b>	<b>163</b>
A	Diffusion World Data Summary Plot Details . . . . .	164
B	Potential for Tagging $t_0$ Using Diffusion . . . . .	166
C	NC $\pi^0$ Data Filters . . . . .	171

	C.1	$2\gamma 1p$ Filter	171
	C.2	$2\gamma 0p$ Filter	173
D		NC $\pi^0$ BDT Training Variables	177
	D.1	$2\gamma 1p$ Training Variables	177
	D.2	$2\gamma 0p$ Training Variables	177
<b>Vita</b>			<b>182</b>

# List of Tables

3.1	Summary of best-fit three-flavor neutrino oscillation parameters . . . . .	25
4.1	Booster Neutrino Beam Systematics . . . . .	33
4.2	Data-taking run summary . . . . .	50
5.1	List of nuclear models used in GENIE v3. . . . .	55
5.2	Input values to the modified box model for recombination . . . . .	65
6.1	Selection efficiencies after each selection requirement and number of selected tracks for the longitudinal diffusion measurement . . . . .	76
6.2	Summary of the measured values of $D_L$ from the MicroBooNE data and simulation . . . . .	85
6.3	Results of $\sigma_t^2$ vs. $t$ fits for simulated muon particle gun samples with $D_T$ varied. . . . .	89
6.4	Results of a study of waveform summation for the diffusion measurement . . . . .	92
6.5	Summary of systematic uncertainties on the $D_L$ measurement . . . . .	92
7.1	Topological efficiencies for the single-photon and NC $\pi^0$ selections . . . . .	102
7.2	Summary of simulation and data samples used in the single photon and NC $\pi^0$ analyses . . . . .	102
7.3	BDT response cut efficiencies for the $1\gamma$ selections . . . . .	110
8.1	Truth-level signal definition requirements for the NC $\pi^0$ selection . . . . .	119
8.2	Pre-selection signal efficiencies for the NC $\pi^0$ selection . . . . .	119
8.3	Number of training and testing events for the NC $\pi^0$ selection BDT . . . . .	122
8.4	Training variables and importance for the NC $\pi^0$ selection . . . . .	124



8.5	Breakdown of interaction types in the NC $\pi^0$ selections . . . . .	127
8.6	Truth-level breakdown of backgrounds in the final $2\gamma 1p$ and $2\gamma 0p$ selections .	131
8.7	Gaussian and linear fit parameters for the $2\gamma 1p$ and $2\gamma 0p$ combined-run selections . . . . .	131
8.8	Gaussian-plus-linear fit parameters for the $2\gamma 1p$ and $2\gamma 0p$ Run 1 and Run 3 selections . . . . .	134
C.1	NC $\pi^0$ $2\gamma 1p$ filter passing rates . . . . .	174
C.2	NC $\pi^0$ $2\gamma 0p$ filter passing rates . . . . .	176

# List of Figures

2.1	Experimentally observed $\beta$ decay spectrum compared to expectation . . . . .	6
2.2	First observation of a neutrino interaction in a bubble chamber at Argonne National Lab . . . . .	6
2.3	The standard model of particle physics . . . . .	11
2.4	Diagrams of charged current and neutral current interactions . . . . .	11
2.5	Diagrams of common neutrino interaction modes at GeV-scale energies . . . .	13
2.6	Summary of neutrino-nucleus cross section data . . . . .	18
2.7	Neutrino mass hierarchy . . . . .	18
3.1	Fermilab Accelerator Complex . . . . .	25
3.2	LSND Excess . . . . .	27
3.3	MiniBooNE Low-energy Excess . . . . .	27
3.4	Illustration of Cherenkov rings as detected by MiniBooNE . . . . .	29
4.1	Diagram of the Boostern Neutrino Beamline . . . . .	33
4.2	Diagram of the magnetic focusing horn used in the BNB . . . . .	33
4.3	BNB flux composition . . . . .	34
4.4	Schematic of the MicroBooNE detector . . . . .	34
4.5	LArTPC detection . . . . .	36
4.6	MicroBooNE event display . . . . .	37
4.7	MicroBooNE readout electronics chain . . . . .	41
4.8	Readout wire response functions . . . . .	41
4.9	MicroBooNE PMT array . . . . .	43
4.10	Cartoon of scintillation light in liquid argon . . . . .	43

4.11	Diagram of the UV laser system in MicroBooNE . . . . .	45
4.12	MicroBooNE cosmic ray tagger (CRT) . . . . .	46
4.13	MicroBooNE collected protons-on-target (POT) . . . . .	50
5.1	Sample deconvolved waveform . . . . .	55
5.2	Simulated charge distribution for varied $\theta_{xz}$ . . . . .	59
5.3	Pandora 2D clustering and vertex finding . . . . .	59
5.4	Pandora flowchart . . . . .	63
5.5	$dQ/dx$ non-uniformities in the collection plane . . . . .	63
5.6	$dQ/dx$ distribution for a high-purity crossing muon sample . . . . .	65
5.7	Shower $dE/dx$ from the NuMI CC inclusive analysis . . . . .	67
5.8	$dE/dx$ vs. residual range for typical CC $\nu_\mu$ interactions . . . . .	67
6.1	Diagram of the MicroBooNE coordinate system and wire planes . . . . .	70
6.2	Visualization of the impact of $D_L$ on signal waveforms as a function of drift time . . . . .	71
6.3	Summary of world data for longitudinal electron diffusion in liquid argon . . .	72
6.4	Collection-plane view of a track that passes the diffusion track selection . . .	76
6.5	Track length distributions for tracks selected in the diffusion analysis . . . .	77
6.6	$yz$ -position distribution of reconstructed space points corresponding to se- lected hits on the collection plane . . . . .	79
6.7	Distribution of hit widths vs. drift time . . . . .	79
6.8	Illustration of the waveform summation technique employed in the diffusion analysis . . . . .	81
6.9	Sample summed waveform with Gaussian fit . . . . .	81
6.10	Plots of $\sigma_t^2$ versus $t$ for simulated muon samples . . . . .	84
6.11	$\sigma_t^2$ versus drift time using MicroBooNE CRT-tagged data . . . . .	84
6.12	Area-normalized comparisons of summed waveforms at three drift times . . .	85
6.13	Illustration of the impact of $D_T$ on the $D_L$ measurement . . . . .	86
6.14	2D distribution of the percent variation of the drift velocity relative to the average drift velocity near the anode . . . . .	89

6.15	Results of the study of the waveform summation technique for the diffusion measurement . . . . .	92
6.16	Comparison of the MicroBooNE $D_L$ result with world data . . . . .	94
7.1	NC $\Delta$ decay diagrams . . . . .	97
7.2	T2K upper limit on NC $\Delta$ radiative production rate . . . . .	97
7.3	2D distributions of reconstructed shower energy vs. true shower energy . . .	106
7.4	Example of a decision tree . . . . .	106
7.5	Visualization of the AdaBoost algorithm . . . . .	108
7.6	Examples of undertraining and overtraining in machine learning . . . . .	108
7.7	Single-photon pre-selection distribution . . . . .	110
7.8	BDT response distributions for the $1\gamma 0p$ selection . . . . .	112
7.9	BDT response distributions for the $1\gamma 1p$ selection . . . . .	113
7.10	Final selection distributions for the $1\gamma 1p$ and $1\gamma 0p$ selections . . . . .	114
8.1	Data/MC distributions of the $\pi^0$ invariant mass at the pre-selection stage . .	121
8.2	MC-only and data/MC distributions of the track truncated mean $dE/dx$ at the pre-selection stage . . . . .	121
8.3	BDT response distributions for the two NC $\pi^0$ selections . . . . .	122
8.4	Track calorimetry distributions used in the $2\gamma 1p$ BDT . . . . .	127
8.5	Data/MC comparisons for the $2\gamma 1p$ final selection with flux and cross-section uncertainties . . . . .	128
8.6	Data/MC comparisons for the $2\gamma 0p$ final selection with flux and cross-section uncertainties . . . . .	132
8.7	Reconstructed $\pi^0$ mass distributions with Gaussian-plus-linear fit . . . . .	134
8.8	Reconstructed $2\gamma 1p$ $\pi^0$ invariant mass distributions for Runs 1 and 3 individually with a Gaussian-plus-linear fit . . . . .	135
8.9	Reconstructed $2\gamma 0p$ $\pi^0$ invariant mass distributions for Runs 1 and 3 individually	135
8.10	Track mean truncated $dE/dx$ with detector systematic uncertainties . . . . .	139
8.11	Final fractional covariance and correlation matrices for the $1\gamma$ and $2\gamma$ selections	139
8.12	Final selection fit distributions for the $1\gamma$ and $2\gamma$ selections . . . . .	140

8.13	Reduction of full flux, cross-section, and detector systematic errors in the $1\gamma$ selections due to the $2\gamma$ constraint . . . . .	141
8.14	Reduction of flux and cross-section systematic errors in the $1\gamma$ selections due to the $2\gamma$ constraint . . . . .	141
8.15	Sensitivity of the single-photon analysis to an LEE signal before and after the $2\gamma$ constraint . . . . .	143
8.16	NC $\pi^0$ event displays . . . . .	145
B.1	One-dimensional comparisons of the area-normalized hit RMS distribution on the three wire planes for drift times of $45\ \mu\text{s}$ , $1150\ \mu\text{s}$ , and $2254\ \mu\text{s}$ . . . . .	168
B.2	Distribution of hit times for different slices in hit RMS on the collection plane.	169
B.3	Distribution of the width of the hit RMS distribution over the mean of that distribution for the U, V and Y planes. . . . .	170
B.4	Comparison of data and simulation for the hit RMS distribution in the center of the TPC, around drift time = $1150\ \mu\text{s}$ . . . . .	170
C.5	NC $\pi^0\ 2\gamma 1p$ filter efficiency as a function of true neutrino energy . . . . .	174
C.6	NC $\pi^0\ 2\gamma 0p$ filter efficiency as a function of true neutrino energy . . . . .	176
D.7	Training variables for $2\gamma 1p$ . . . . .	178
D.8	Training variables for $2\gamma 0p$ . . . . .	180

# List of Abbreviations

<b>ADC</b>	Analog-to-Digital Converter
<b>ANL</b>	Argonne National Lab
<b>ASIC</b>	Application-Specific Integrated Circuit
<b>BDT</b>	Boosted Decision Tree
<b>BF</b>	Best Fit
<b>BNB</b>	Booster Neutrino Beam
<b>BNL</b>	Brookhaven National Lab
<b>BSM</b>	Beyond Standard Model (Physics)
<b>CC</b>	Charged Current
<b>CM</b>	Center of Mass
<b>CMOS</b>	Complimentary Metal-Oxide Semiconductor
<b>CNO</b>	Carbon-Nitrogen-Oxygen (Cycle)
<b>CP</b>	Charge-Parity (Symmetry)
<b>CR</b>	Cosmic Ray
<b>CRT</b>	Cosmic Ray Tagger
<b>DAQ</b>	Data Aquisition System
<b>DIS</b>	Deep Inelastic Scattering
<b>DUNE</b>	Deep Underground Neutrino Experiment
<b>E-field</b>	Electric Field
<b>EM</b>	Electromagnetic
<b>EXT</b>	External (Cosmic Data)
<b>FEB</b>	Front-End Board
<b>FEMB</b>	Front-End Motherboard

**Fermilab** Fermi National Accelerator Lab  
**FPGA** Field-Programmable Gate Array  
**FSI** Final-State Interactions  
**GENIE** Generates Events for Neutrino Interaction Experiments  
**GoF** Goodness-of-Fit  
**ICARUS** Imaging Cosmic And Rare Underground Signals (Experiment)  
**LArSoft** Liquid Argon Software  
**LArTF** Liquid Argon Testing Facility  
**LArTPC** Liquid Argon Time Projection Chamber  
**LEE** Low-Energy Excess (anomaly from MiniBooNE)  
**LFG** Local Fermi Gas  
**Linac** Linear Accelerator  
**LSND** Liquid Scintillator Neutrino Detector  
**MC** Monte Carlo  
**MEC** Meson Exchange Current  
**MicroBooNE** Micro Booster Neutrino Experiment  
**MiniBooNE** Mini Booster Neutrino Experiment  
**MSW** Mikheyev-Smirnov-Wolfenstein (Effect)  
**NC** Neutral Current  
**NCE** Neutral Current Elastic  
**NuMI** Neutrinos from the Main Injector  
**PE** Photoelectron  
**PDG** Particle Data Group  
**PFO** Particle Flow Object  
**PID** Particle Identification  
**PMNS** Pontecorvo-Maki-Nakagawa-Sakata (Matrix)  
**PMT** Photomultiplier Tube  
**POT** Protons On Target  
**QE** Quasi-Elastic  
**RES** Resonant (Scattering)

**ROI** Region of Interest  
**SBN** Short Baseline Neutrino Program  
**SBND** Short Baseline Near Detector  
**SCE** Space Charge Effect  
**SiPM** Silicon Photomultiplier  
**SM** Standard Model (of Particle Physics)  
**SNO** Sudbury Neutrino Observatory  
**SSV** Second Shower Veto  
**TPB** Tetraphenyl Butadiene  
**TPC** Time Projection Chamber  
**UV** Ultraviolet  
**VUV** Vacuum Ultraviolet  
**WC** Wirecell  
**XGBoost** eXtreme Gradient Boosting  
**XS** Cross Section



# Chapter 1

## Introduction

The measurement and characterization of neutrinos have proved challenging to physicists since their prediction in 1936. These electrically neutral, nearly-massless particles interact only via the weak and gravitational forces, and so are exceptionally difficult to detect. Originally proposed as a hypothetical particle to explain the then-anomalous  $\beta$  decay energy spectrum, neutrinos have raised a multitude of questions over the past century: what are their absolute masses? is the neutrino its own antiparticle? how many neutrino flavors are there? can neutrinos open the way to new, exotic particle physics? Indeed, since the experimental discovery of neutrinos in the 1950s, they appear to have raised more questions than answers. We know that neutrinos come in (at least) three flavors: electron, muon, and tau. We also know that neutrinos propagate as mass eigenstates, which are themselves superpositions of the flavor states. We know the mass squared differences,  $\Delta m^2$ , between the three known states, but not the mass ordering. Beyond this, neutrinos remain largely mysterious.

These elusive particles have motivated the need for increasingly large, sensitive neutrino detectors. While there are a multitude of such detectors, the focus of this thesis will be the liquid argon time projection chamber (LArTPC), a novel detector technology with excellent position resolution and calorimetric capabilities. The past several decades have seen this technology rapidly develop, and LArTPCs are now the preferred technology for current and next-generation neutrino detectors. The Short Baseline Neutrino (SBN) program at Fermi National Accelerator Laboratory (Fermilab)—containing the MicroBooNE,

SBND, and ICARUS detectors—and the upcoming Deep Underground Neutrino Experiment (DUNE) all utilize LArTPC technology. With LArTPC detectors being the clear favorite of next-generation neutrino experiments, a proper understanding of this technology is vital to the future of the field.

The MicroBooNE experiment is a LArTPC situated along Fermilab’s Booster Neutrino Beamline (BNB) as part of the SBN program. MicroBooNE’s primary physics goal is to investigate the MiniBooNE low-energy excess (LEE) anomaly, but it will also provide valuable neutrino-nucleus cross section measurements and LArTPC detector physics measurements. It is the first hundred-tonne-scale LArTPC and the longest-running LArTPC experiment in the world as of this writing. In addition to solving the puzzle of the MiniBooNE LEE, MicroBooNE will also pave the way for future LArTPC experiments.

This thesis presents two MicroBooNE analyses: a measurement of the longitudinal electron diffusion coefficient,  $D_L$ , and a selection of neutral current (NC)  $\pi^0$ s for a data-driven rate constraint on MicroBooNE’s search for single-photon events. LArTPC detection (shown schematically in Figure 4.5) depends on a proper understanding of electron transport properties as they travel through the liquid argon. The diffusion of these electrons along the drift coordinate smears the collected signal as the electrons drift, and the rate of longitudinal electron diffusion can be measured by examining this smearing as a function of drift time. Only two measurements of  $D_L$  exist in liquid argon as of this writing, and those measurements are in tension with one another. Furthermore, the leading theoretical calculation is ill-defined at electric field strengths of  $O(100)$  V/cm, the regime in which all of the aforementioned experiments live. This motivates the MicroBooNE measurement of  $D_L$ , which will be discussed in Chapter 6. In addition to the  $D_L$  measurement, this thesis also presents a selection of NC resonant  $\pi^0$ s. This selection is primarily motivated by the large NC  $\pi^0$  background in MicroBooNE’s single-photon event search, an analysis which seeks to investigate whether the MiniBooNE LEE is photon-like in nature. We apply the NC  $\pi^0$  selection as a data-driven rate constraint to reduce the systematic errors in the single-photon selection. However, even outside the context of the single-photon analysis, NC  $\pi^0$ s are still of broad interest to the field of neutrino physics. NC  $\pi^0$ s form a large background in  $\nu_\mu \rightarrow \nu_e$

oscillation searches, making a measurement of the NC  $\pi^0$  production rate beneficial to the field as a whole.

The remainder of this thesis is organized as follows: Chapter 2 presents a brief overview of neutrino physics, including historical context and the basic mathematical formalism. Chapter 3 gives an overview of neutrino experiments. Chapter 4 then focuses on the MicroBooNE experiment, while Chapter 5 describes the simulation and reconstruction tools used in MicroBooNE. Chapter 6 presents the  $D_L$  measurement. Chapter 7 gives an overview of the MicroBooNE single-photon analysis, which contextualizes the NC  $\pi^0$  selection presented in Chapter 8. Finally, the conclusions are given in Chapter 9.

# Chapter 2

## Overview of Neutrino Physics

This chapter provides a brief overview of neutrino physics, including historical context and the basic mathematical formalism of neutrino oscillations. Section 2.1 gives a brief historical overview of the discovery of the neutrino, which leads into the discussion of the Standard Model in Section 2.2. Section 2.3 discusses neutrino oscillations, while Section 2.4 discusses the most common neutrino-nucleus interaction modes. Finally, Section 2.5 discusses some of the major open questions in the field of neutrino physics.

### 2.1 Detecting the Undetectable

Neutrinos were originally proposed by Wolfgang Pauli as a seemingly desperate attempt to rescue conservation of energy. Experiments measuring the energy spectrum observed in  $\beta$  decay found a curious result: the supposed two-body decay—which should have resulted in a single characteristic energy of the two outgoing decay products—was found to have a broad energy spectrum, as shown in Figure 2.1. So puzzling was this anomaly that Niels Bohr was reportedly prepared to abandon the conservation of energy. Pauli himself famously lamented that he had “done a terrible thing” by proposing a “particle that cannot be detected.” Thankfully, Pauli’s proposed “neutron” (which, of course, later became known as the neutrino after James Chadwick discovered what is now known as the neutron) was detected in 1956 by Reines and Cowan [1]. This experiment utilized the  $\beta$  capture method proposed by Wang in 1942 [2]. In this method, antineutrinos interact with protons as

$$\bar{\nu}_e + p \rightarrow n + e^+ \quad (2.1)$$

where  $\bar{\nu}_e$  is an antielectron neutrino,  $p$  is a proton,  $n$  is a neutron, and  $e^+$  a positron. The positron quickly annihilates with an electron, producing photons ( $\gamma$ ), while the neutron capture on a nucleus also produces a photon. The coincident observation of these two photon emission processes confirmed the existence of neutrinos, and resulted in a Nobel Prize award in 1995.

Although neutrinos are now an accepted part of the standard model of particle physics, they have provided physicists with no shortage of anomalies over the past century. Further discussion of experimental anomalies and their current status can be found in Chapter 3.

## 2.2 The Standard Model

The Standard Model (SM) of particle physics [3, 4, 5] describes the known fundamental forces and elementary particles that govern the physical laws of our universe. Each fundamental force—gravitation, electromagnetic, strong nuclear, and weak nuclear—is mediated by a force carrier. Photons mediate the electromagnetic force, gluons carry the strong nuclear force, the massive  $W^\pm$  and  $Z$  bosons carry the weak nuclear force, and gravitation is believed to be mediated by the graviton. While not a force carrier itself, the Higgs boson is responsible for particle masses via the Higgs mechanism. The SM also contains two classes of elementary particles: quarks and leptons. Each of these classes is further split into three generations. For quarks, the up ( $u$ ) and down ( $d$ ) form the first generation, the charm ( $c$ ) and strange ( $s$ ) the second, and the top ( $t$ ) and bottom ( $b$ ) the third. The lepton generations consist of a charged lepton and its corresponding neutrino flavor state: the electron ( $e$ ) and electron neutrino ( $\nu_e$ ), the muon ( $\mu$ ) and muon neutrino ( $\nu_\mu$ ), and finally the tau ( $\tau$ ) and tau neutrino ( $\nu_\tau$ ). Figure 2.3 shows the current state of the standard model.

All SM particles have a corresponding antiparticle which has the same mass but opposite electric charge and spin. Particles and antiparticles were once believed to be symmetric under charge-parity (CP) exchange, but the discovery of CP violation [6] in neutral kaon decays showed that this was only an approximate symmetry. When a particle meets its antiparticle,

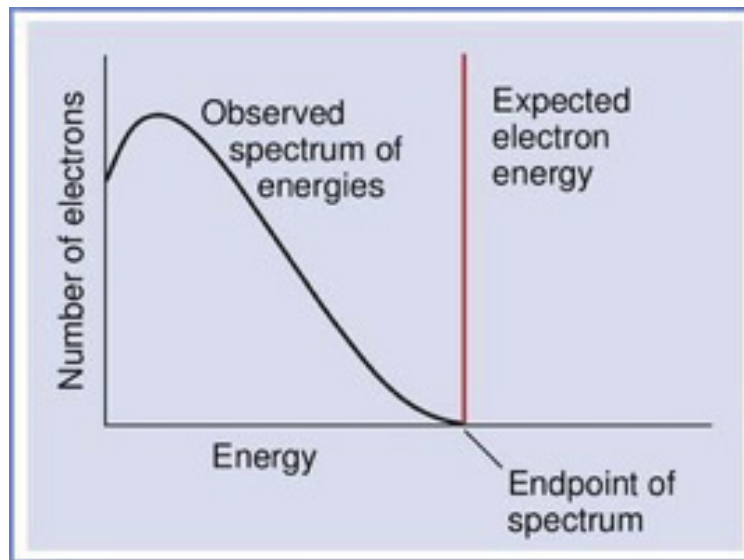


Figure 2.1: Experimentally observed  $\beta$  decay spectrum (black line) compared to expectation (vertical red line).

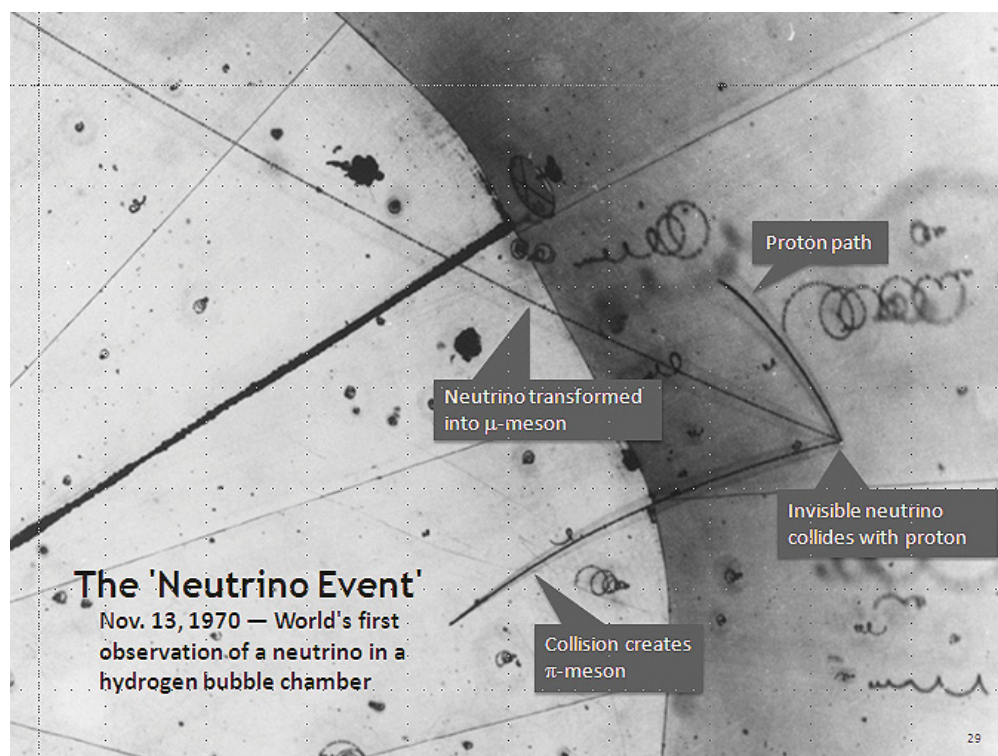


Figure 2.2: First observation of a neutrino interaction in a bubble chamber at Argonne National Lab.

they annihilate and release energy in the form of photons. All matter in the known universe is comprised of matter instead of antimatter, a phenomenon known as the *matter-antimatter asymmetry*. Explaining this asymmetry remains an open question in physics.

All SM particles can be classified as either fermions or bosons. Fermions have half-integer spin and obey Fermi-Dirac statistics. Quarks and leptons are fermions, as are baryons (heavier particles composed of three quarks, such as protons). Bosons, on the other hand, have integer spin and obey Bose-Einstein statistics. All four force carriers and the Higgs are bosons, as are mesons (heavier particles composed of a quark-antiquark pair).

While the SM is widely accepted in the particle physics community, it remains incomplete. For one, the graviton has yet to be experimentally confirmed. The SM also offers no explanation of dark matter or dark energy, the combination of which is believed to be responsible for 95% of the matter in the universe. Of particular interest to this thesis is the phenomenon of neutrino oscillations, in which neutrinos oscillate between the different flavor eigenstates as a function of time. This necessarily implies that neutrinos have non-zero mass, but the SM assumes massless neutrinos. Neutrino oscillations have been confirmed by numerous experiments [7, 8], making them the first clear indication of beyond-standard-model (BSM) physics.

## 2.3 Neutrino Oscillations

Neutrinos propagate as *mass eigenstates* (denoted  $\nu_k$ , with  $k = 1, 2, 3$ ), but are detected as *flavor eigenstates* (denoted  $\nu_\alpha$ , with  $\alpha = e, \mu, \tau$ ). Each flavor state is a superposition of the three mass states, as described by

$$|\nu_\alpha\rangle = \sum_k U_{\alpha k}^* |\nu_k\rangle \quad (2.2)$$

where the unitary matrix  $U$  is known as the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [9, 10], given by

$$\begin{bmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{bmatrix} = \begin{bmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{bmatrix} \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{bmatrix}, \quad (2.3)$$

in the case of three-flavor oscillations. The PMNS matrix can also be written in terms of three flavor-space mixing angles  $\theta_{12}$ ,  $\theta_{23}$ , and  $\theta_{13}$ , and a charge-parity (CP)-violating phase,  $\delta_{CP}$ :

$$U = \begin{bmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{CP}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{CP}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{CP}} & s_{23}c_{13} \\ -s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{CP}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{CP}} & c_{23}c_{13} \end{bmatrix}. \quad (2.4)$$

In this notation,  $c_{ij} = \cos(\theta_{ij})$  and  $s_{ij} = \sin(\theta_{ij})$ . A full description of neutrino oscillations therefore requires accurate measurements of each of the oscillation parameters. (Experimental measurements of these parameters will be discussed in Chapter 3.) The mass states,  $|\nu_k\rangle$ , evolve in time as plane waves:

$$|\nu_k(t)\rangle = e^{-iE_k t} |\nu_k\rangle. \quad (2.5)$$

Combining Equations 2.2 and 2.5 then yields

$$|\nu_\alpha(t)\rangle = \sum_k U_{\alpha k}^* e^{-iE_k t} |\nu_k\rangle. \quad (2.6)$$

Similar to Equation 2.2, the mass state  $|\nu_k\rangle$  can be expressed as a superposition of flavor states as

$$|\nu_k\rangle = \sum_\alpha U_{\alpha k} |\nu_\alpha\rangle \quad (2.7)$$

where we've used the fact that  $UU^\dagger = 1$ . We can then substitute 2.7 into 2.6 to obtain



$$|\nu_\alpha(t)\rangle = \sum_{\beta=e,\mu,\tau} \left( \sum_k U_{\alpha k}^* e^{-iE_k t} U_{\beta k} \right) |\nu_\beta\rangle. \quad (2.8)$$

Finally, the transition probability as a function of time is given by

$$P_{\alpha \rightarrow \beta}(t) = |\langle \nu_\beta | \nu_\alpha \rangle|^2 = \sum_{k,j} U_{\alpha k}^* U_{\beta k} U_{\alpha j} U_{\beta j}^* e^{-i(E_k - E_j)t}. \quad (2.9)$$

Most neutrino experiments use an ultrarelativistic neutrino beam, meaning that  $t \sim L$ . Moreover, neutrinos are (approximately) massless with energy given by  $E_k = \sqrt{p_k^2 + m_k^2}$ . This can be expanded to first order as  $E_k \simeq p_k + \frac{m_k^2}{2p_k} \approx E + \frac{m_k^2}{2E}$ . Then

$$E_k - E_j \simeq \frac{\Delta m_{kj}^2}{2E}. \quad (2.10)$$

Substituting this into 2.9 and noting that  $t \sim L$  gives

$$P(\nu_{\alpha \rightarrow \beta}) = \sum_{j,k} U_{\alpha k}^* U_{\beta k} U_{\alpha j} U_{\beta j}^* e^{-i \frac{\Delta m_{jk}^2 L}{2E}}. \quad (2.11)$$

Equation 2.11 describes the neutrino oscillation probability as a function of time in terms of the measurable matrix elements of  $U$ , the *mass splitting* ( $\Delta m^2$ ), the distance of flight ( $L$ ), and the neutrino energy ( $E$ ). We see that  $\Delta m^2$  controls the frequency of oscillation, while the magnitudes of the mixing angles contained within  $U$  control the oscillation amplitude. This expression of the oscillation probability is especially useful for oscillation experiments, which can tune the ratio  $L/E$  to suit their purposes. Equation 2.11 also shows that experiments are not sensitive to the absolute neutrino mass (which would significantly complicate things) but instead only to the mass splitting between two mass states. For most practical applications, the transition probability equation can be further simplified by considering the two-flavor oscillation approximation

$$P(\nu_{\alpha \rightarrow \beta}) = \sin^2(2\theta_{kj}) \sin^2 \left( \frac{\Delta m_{kj}^2 L}{2E} \right). \quad (2.12)$$

This approximation is valid because experimental measurements have shown that 1)  $\Delta m_{12}^2 \ll \Delta m_{13}^2$  and 2)  $\theta_{13}$  is small (relative to the other mixing angles). Most experiments

only consider two-flavor oscillations such as  $\nu_\mu \rightarrow \nu_e$ , making this approximation especially useful.

## 2.4 Neutrino Interactions

Neutrinos interact only via the weak and gravitational forces, but the gravitational effect is difficult to measure. To study neutrinos, then, we must focus on their weak force interactions. As discussed in Section 2.2, the weak force is mediated by the  $W$  and  $Z$  bosons. Neutrino interactions involving the  $W^\pm$  are known as *charged current* (CC) interactions, while those involving the  $Z$  are known as *neutral current* (NC). CC neutrino interactions are identified by the corresponding outgoing lepton. In a CC  $\nu_e$  interaction, we expect an electron in the final state, and so on for  $\nu_\mu$  and  $\nu_\tau$  interactions. In NC interactions, the neutrino appears in both the initial and final state and the exchange of a  $Z$  boson mediates the transfer of energy. More generally, CC and NC interactions proceed as

$$\nu_\ell + A \rightarrow \ell^- + X \text{ (CC)} \quad (2.13)$$

$$\nu_\ell + A \rightarrow \nu_\ell + X \text{ (NC)} \quad (2.14)$$

where  $\nu_\ell$  represents a neutrino of flavor  $\ell = e, \mu, \tau$ ,  $A$  is the target nucleus,  $\ell$  is the charged outgoing lepton (in CC interactions only), and  $X$  represents final-state particles.<sup>1</sup> Figure 2.4 shows examples of typical CC and NC neutrino interactions. In addition to CC and NC, GeV-scale neutrino interactions (the focus of this thesis) are further subdivided into distinct interaction types depending on the incident neutrino energy and how it interacts with a target nucleus. For our purposes, the most relevant neutrino interaction modes are the quasielastic (QE) and resonant modes. We discuss these, as well as other relevant interaction modes, in the following sections.

---

<sup>1</sup>Note that the corresponding processes with an antineutrino and a positively-charged lepton are also valid.

## Standard Model of Elementary Particles

three generations of matter (fermions)						interactions / force carriers (bosons)	
I		II		III			
mass $\approx 2.2 \text{ MeV}/c^2$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0	$\approx 124.97 \text{ GeV}/c^2$
charge $\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0
spin $\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0
<b>u</b> up		<b>c</b> charm		<b>t</b> top		<b>g</b> gluon	<b>H</b> higgs
$\approx 4.7 \text{ MeV}/c^2$		$\approx 96 \text{ MeV}/c^2$		$\approx 173.1 \text{ GeV}/c^2$		0	
$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	0	
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
<b>d</b> down		<b>s</b> strange		<b>b</b> bottom		<b><math>\gamma</math></b> photon	
$\approx 0.511 \text{ MeV}/c^2$		$\approx 105.66 \text{ MeV}/c^2$		$\approx 1.7768 \text{ GeV}/c^2$		$\approx 91.19 \text{ GeV}/c^2$	
-1	-1	-1	-1	-1	-1	0	
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
<b>e</b> electron		<b><math>\mu</math></b> muon		<b><math>\tau</math></b> tau		<b>Z</b> Z boson	
$< 1.0 \text{ eV}/c^2$	0	$< 0.17 \text{ MeV}/c^2$	0	$< 18.2 \text{ MeV}/c^2$	0	$\approx 80.39 \text{ GeV}/c^2$	
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\pm 1$	
<b><math>\nu_e</math></b> electron neutrino		<b><math>\nu_\mu</math></b> muon neutrino		<b><math>\nu_\tau</math></b> tau neutrino		<b>W</b> W boson	

Figure 2.3: The standard model of particle physics.

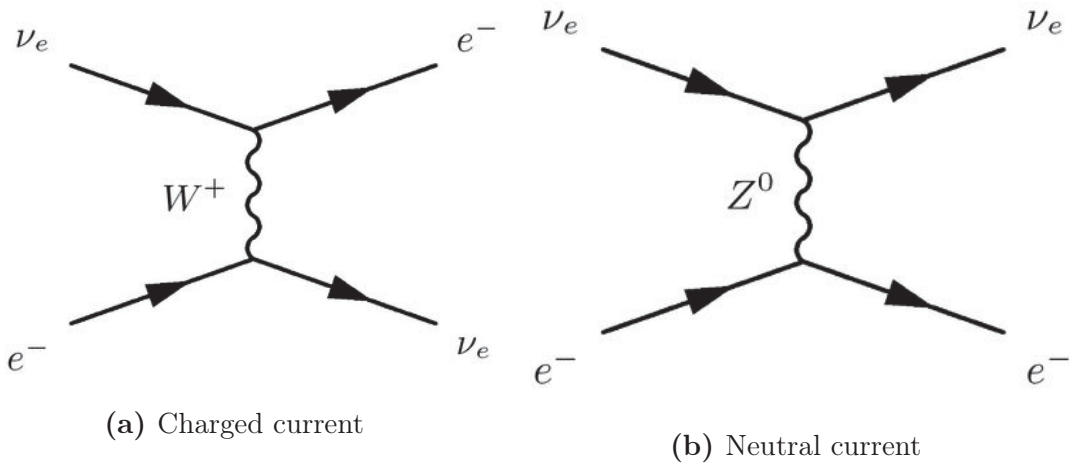


Figure 2.4: Diagrams of typical CC and NC  $\nu_e$  interactions.

### 2.4.1 Quasielastic Interactions

For  $\sim 1$  GeV neutrino energies, the most common interaction mode is the quasielastic (QE) mode. In QE interactions, an incoming neutrino interacts with the nucleus, modifying the quark flavor content of a nucleon in the process, as shown in Figure 2.5a. A more general charged current quasielastic (CCQE) interaction is given by

$$\nu_\ell + n \rightarrow \ell^- + p \quad (2.15)$$

$$\bar{\nu}_\ell + p \rightarrow \ell^+ + n \quad (2.16)$$

where  $n$  and  $p$  represent a neutron and proton, respectively. For NC elastic (NCE) interactions, a neutrino transfers energy to a single nucleon via  $Z$  boson exchange, so the interaction proceeds as

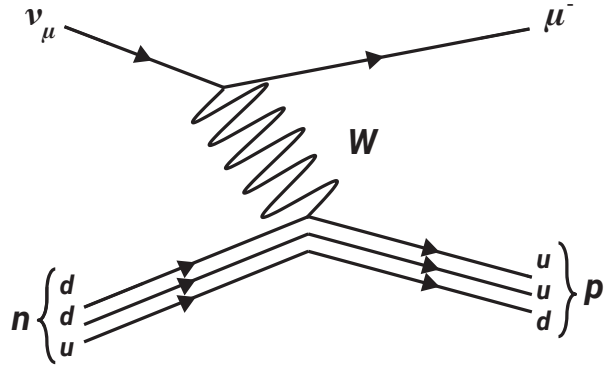
$$\nu_\ell + N \rightarrow \nu_\ell + N \quad (2.17)$$

where  $N$  represents a nucleon. Note that NCE interactions also produce a nuclear recoil which is not shown in Equation 2.17. CCQE interactions are significantly easier to recognize than NCE interactions due to the presence of an outgoing charged lepton. For NCE, the signal is generally a single neutron or proton in the final state, which can be difficult to distinguish from background. NCE interactions will not be considered further in this work.

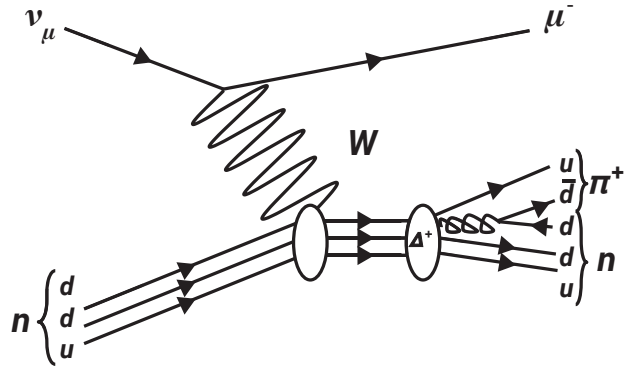
Most simulated event generators use the Llewellyn-Smith model [11] for CCQE interactions, although the more recent Nieves model [12] appears to offer better agreement between simulation and data for some experiments with larger target nuclei [13].

### 2.4.2 Resonant Interactions

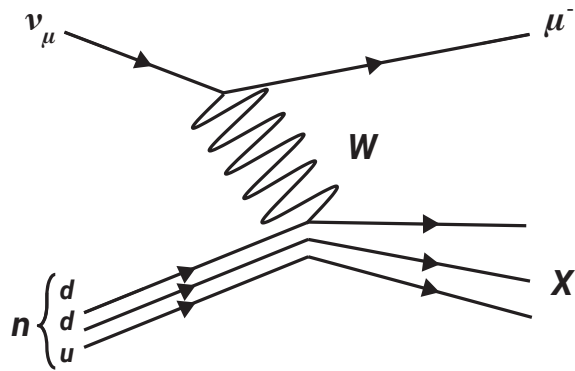
When a neutrino strikes a nucleon with sufficient energy, the interaction may result in a nucleon *resonance*, as shown in Figure 2.5b. These resonances quickly decay, generally producing a nucleon and some number of mesons (often pions) in the final state. Resonant pion production is the dominant pion production mode for GeV-scale energies, and these



(a) CCQE



(b) Resonant



(c) DIS

Figure 2.5: Diagrams of common neutrino interaction modes at GeV-scale energies. Note that these examples show charged current interactions.

pions can be a significant background for neutrino oscillation experiments (see Chapter 3). In fact, resonant NC  $\pi^0$  production constitutes one of the largest background sources for  $\nu_\mu \rightarrow \nu_e$  oscillation searches—if only one of the two photons from  $\pi^0$  decay is detected, it can mimic the single-electron signal sought by  $\nu_e$  appearance searches [14]. At energies of  $\sim 1$  GeV, the dominant resonant production is the  $\Delta(1232)$ , which is the first excited state of a nucleon. Higher-order resonances can occur as well, although these are far less common in this energy regime.

Historically, the most commonly used resonant model in simulation has been the Rein-Sehgal model [15]. The more recent Berger-Sehgal [16] model improves on the older model by including experimental pion scattering data and leptonic mass correction terms.

### 2.4.3 Other Interactions

CCQE and resonant interactions are the most common neutrino interaction types at GeV-scale energies, which is the regime most relevant to the work described in this thesis. However, even at lower energies, other interaction processes may contribute subdominantly. Deep inelastic scattering (DIS, shown in Figure 2.5c) occurs when a high-energy neutrino interacts directly with a constituent quark, breaking apart the nucleon and resulting in a hadronic shower. DIS dominates for energies above  $\sim 10$  GeV, much higher than the GeV-scale energy we consider in this work.

In addition to the resonance mode discussed in the previous section, pions can also be produced in *coherent* scattering. In coherent scattering, a low-energy neutrino interacts inelastically with the nucleus as a whole, producing either a charged or neutral pion. As with resonant production, coherent pion production is modeled using either the Berger-Sehgal or Rein-Sehgal models.

Finally, neutrino interactions which resemble CCQE can also arise from “two-particle two-hole” (2p2h) processes. In these processes, a neutrino scatters on a pair of nucleons which are interacting with each other via meson-exchange currents (MEC). Older versions of simulated event generators do not account for MEC interactions. Recent versions of the widely-used GENIE event generator [17] include a model of CC 2p2h interactions based on

the work of Nieves *et al.* [12]. For NC, the only 2p2h treatment available in GENIE is the empirical Dytman model described in Reference [18].

#### 2.4.4 Summary

As discussed in the previous sections, challenges arise when measuring and predicting neutrino-nucleus cross sections for heavier nuclei. Historical constraints of nuclear form factors come from scattering experiments using hydrogen or deuterium as the target, but more recent experiments use nuclei such as carbon, iron, lead, or argon. In addition to uncertainties on nuclear form factors, final-state interactions (FSI) also complicate these predictions. In large nuclei, final-state particles must travel through the nucleus before being emitted in a detectable state. While traversing the nucleus, particles (particularly protons and pions) can re-interact within the nucleus before exiting. Pions, for example, can charge exchange (converting a neutral pion to a charged pion or vice versa), or be absorbed by the nucleus. FSI are difficult to model, and therefore greatly increase the systematic uncertainty on cross section measurements. Most event generators model intranuclear interactions using a cascade model, in which particles interact with a series of individual nucleons. The interaction probability with each nucleon is governed by a mean free path, which depends on the nucleon cross section. Due to the computational challenges involved in performing a full cascade simulation, most generators simplify this into an *effective* cascade model which uses a single effective nuclear cross section.

Modern experimental cross section measurements using heavier nuclei show broad agreement between prediction and data in the CCQE mode, although discrepancies exist in the most forward-scattering regions [19, 20]. Significant tensions between data and prediction exist for resonant pion production [20]. More experimental data is necessary to resolve these tensions.

Figure 2.6, taken from Reference [21], summarizes experimental world data for the three major interaction modes: CCQE, resonant, and DIS. Figure 2.6 also shows the predicted total cross section. The data shown here come from a variety of experiments using different target nuclei, and have been collected over several decades.

## 2.5 Open Questions in Neutrino Physics

Despite the abundance of neutrinos in nature, many open questions remain about their basic properties. For example, we do not yet know whether neutrinos are Dirac or Majorana fermions. This question essentially boils down to whether the neutrino is its own antiparticle, in which case it would be a Majorana particle. To answer this question, some experiments are searching for a process known as *neutrinoless double-beta decay* [23]. Double beta decay is a rare process that has been observed in atoms (such as Germanium and Selenium) in which regular beta decay is energetically suppressed, while double beta decay is allowed. In every observed case to date, this process released two neutrinos, as expected. However, if the neutrino is its own antiparticle, there is a small probability that the two neutrinos released in double beta decay will annihilate, and the two outgoing electrons will carry away the entirety of the decay energy. In principle, this would result in an energy spike characteristic of the atom in question. To date, neutrinoless double-beta decay has not been observed, but increasingly sensitive experiments are under development [24].

Another open question is the absolute mass of each mass eigenstate. The current best upper bound on the effective mass (i.e., the sum of the masses of the three mass states) comes from the KATRIN experiment [25], who place the upper bound at 1.1 eV at a 90% confidence level. While the absolute masses of each mass eigenstate remain unknown, the mass splittings have been measured with reasonable precision (see Section 3.4). However, there is an ambiguity in the ordering of the masses. This ordering is known as the *neutrino mass hierarchy* (or *mass ordering*). The *normal* hierarchy has  $m_1 < m_2 < m_3$ , while the *inverted* hierarchy has  $m_3 < m_1 < m_2$ . This is depicted in Figure 2.7. We know that  $m_1 < m_2$  from examining solar neutrino oscillation probabilities. Neutrinos produced in the Sun have their oscillation parameters modified by matter effects, known as the Mikheyev-Smirnov-Wolfenstein (MSW) effect [26, 27, 28]. In particular, solar  $\nu_e$ s interact with the electrons in solar matter via CC interactions, and this introduces a dependence on the sign of  $\Delta m_{21}^2$  (see Equation 5 of Reference [29]), allowing us to determine the ordering of those two states. As for  $m_3$ , current neutrino experiments are not sensitive to the small matter effects caused by neutrino propagation through the Earth, meaning that we cannot yet determine



whether  $m_3$  is the heaviest state or the lightest. Future oscillation experiments are expected to resolve the neutrino mass hierarchy.

In addition to the mass splittings, other ambiguities remain in certain oscillation parameters. The three mixing angles in Equation 2.4 have been measured with reasonable precision, but there remains an ambiguity of the octant in which  $\theta_{23}$  lies. This angle has been measured to be close to  $45^\circ$ , but it is currently unclear whether the true value is greater than  $45^\circ$ , less than  $45^\circ$ , or exactly  $45^\circ$ . In the latter case, we would say that the mixing is *maximal*, implying that the  $\nu_\mu$  and  $\nu_\tau$  components of  $\nu_3$  would be equal and providing evidence for a previously unknown symmetry in the neutrino sector [30]. Furthermore, some neutrino oscillation models relate the  $\theta_{23}$  octant to the mass hierarchy [31]. The determination of the  $\theta_{23}$  octant is therefore important for understanding neutrino oscillations.

Finally, there remain large uncertainties in the determination of the CP-violating phase  $\delta_{CP}$ . This may manifest, for example, as an asymmetric oscillation probability between  $\nu_e \rightarrow \nu_\mu$  and  $\bar{\nu}_e \rightarrow \bar{\nu}_\mu$  oscillations. At present, CP violation has been observed in the quark sector, but not the leptonic sector. CP violation may provide an explanation for the matter-antimatter asymmetry of the universe. The quark sector observations of CP violation are not sufficient to explain the asymmetry, making the leptonic sector an attractive avenue through which to explain this.

Open questions in neutrino physics relating to oscillation parameters require precise experimental measurements. In the next chapter, we discuss the major types of neutrino experiments and the current best-fit values of the neutrino oscillation parameters.

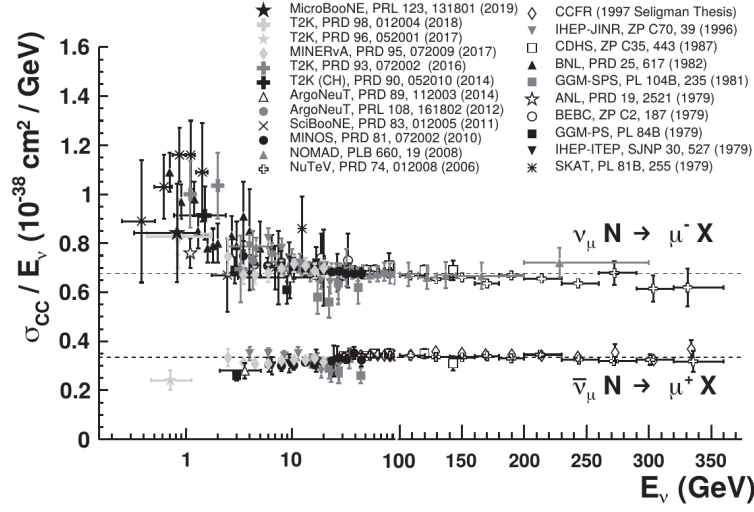


Figure 2.6: Summary of accumulated neutrino-nucleus cross section data as a function of neutrino energy [22]. The data come from a wide variety of experiments on various target nuclei.

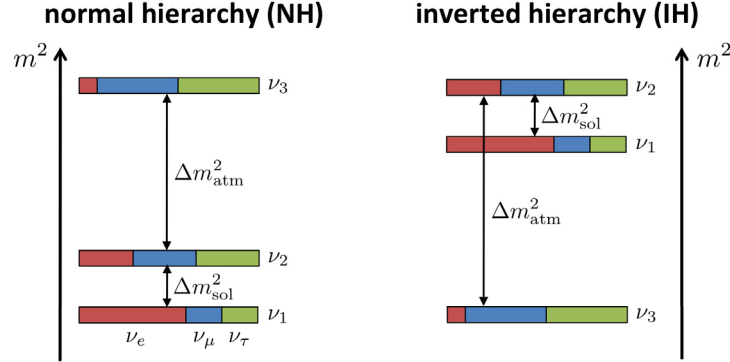


Figure 2.7: Depiction of the normal (left) and inverted (right) neutrino mass hierarchies. Each mass eigenstate (denoted as  $\nu_1$ ,  $\nu_2$ , and  $\nu_3$ ) is a superposition of three flavor states.  $\Delta m_{atm}^2$  and  $\Delta m_{sol}^2$  represent the atmospheric and solar neutrino mass splittings, respectively.

## Chapter 3

# Neutrino Experiments

Neutrino experiments broadly fall into four categories: solar experiments, which seek to detect neutrinos emitted by the Sun during the fusion of helium into hydrogen; atmospheric experiments, which measure neutrinos produced during the collision of cosmic rays with nuclei in the Earth’s atmosphere; reactor experiments, which detect  $\bar{\nu}_e$  emitted from nuclear reactors during  $\beta$  decay from unstable fission fragments; and finally, accelerator-based experiments, which use particle accelerators to produce a neutrino beam from proton beam collisions with a fixed target. While the focus of this thesis is accelerator neutrinos, we pause here to briefly discuss the other experiment types in order to contextualize the accelerator results. In each of the following sections, the discussion of global fits comes primarily from References [32] and [33] and the best-fit values of neutrino mixing parameters come from the Particle Data Group (PDG) [22].

Most experiments assume the standard three-flavor neutrino oscillation model discussed in Section 2.3, but recent anomalies point toward the existence of *sterile* neutrino states which only interact via the gravitational force. Such sterile flavor states would be a major discovery if confirmed, as the standard model currently assumes only three active neutrino states.

Sections 3.1–3.4 discuss each of the aforementioned experiment types. Section 3.5 discusses the MiniBooNE low-energy excess anomaly, the primary motivation for the MicroBooNE experiment.

### 3.1 Solar Neutrino Experiments

The Sun produces neutrinos during thermonuclear reactions, primarily through the  $pp$  (proton-proton) chain and Carbon-Nitrogen-Oxygen (CNO) cycle. Neutrinos from the  $pp$  chains are low-energy, about 400 keV. However, other solar neutrino production modes can produce neutrinos with energy  $O(10)$  MeV. Solar neutrino experiments were the first to prove the existence of neutrino oscillations: in 1965, the Homestake experiment [34] measured the solar neutrino flux by examining  $\text{Cl}^{37}$   $\beta$  decay via the  $\nu_e + \text{Cl}^{37} \rightarrow \text{Ar}^{37} + e^-$  reaction. The result was a substantial deficit of detected  $\nu_e$ s, roughly one third of the prediction at the time. This became known as the *solar neutrino problem*. Throughout the late 20th century, other experiments observed a similar deficit [35, 36]. In 1968, Pontecorvo proposed the possibility of neutrino oscillations [10] to explain this apparent deficit. In this scenario, many of the  $\nu_e$ s produced in the Sun oscillate into different flavor states by the time they reach the Earth, making them undetectable by the aforementioned  $\text{Cl}^{37}$  decay channel. Finally, in 2002, the Sudbury Neutrino Observatory (SNO) experiment confirmed neutrino oscillations by measuring both the total neutrino flux and the  $\nu_e$  flux. Thus, neutrinos were confirmed to oscillate as a function of distance [7], with the beyond-standard-model implication that neutrinos masses are non-zero.

Solar experiments are particularly well-suited to measure the oscillation parameters  $\Delta m_{21}^2$  and  $\theta_{12}$ . Indeed,  $\theta_{12}$  is often referred to as the “solar” mixing angle. Recent global fit analyses include results from all past and present solar experiments, particularly Homestake, GALLEX/GNO, Borexino, and Super-Kamiokande. They also include results from the KamLAND reactor experiment, whose long baseline (distance the neutrino travels) of 180 km makes it sensitive to  $\Delta m_{21}^2$ . The current  $1\sigma$  best-fit values for each of these parameters are  $\Delta m_{21}^2 = 7.42_{-0.20}^{+0.21} \times 10^{-5} \text{ eV}^2$  and  $\theta_{12} = 33.44_{-0.74}^{+0.77}$  degrees for the normal hierarchy, and  $\Delta m_{21}^2 = 7.42_{-0.20}^{+0.21} \times 10^{-5} \text{ eV}^2$  and  $\theta_{12} = 33.45_{-0.75}^{+0.78}$  for the inverted hierarchy.

## 3.2 Reactor Neutrino Experiments

Reactor neutrino experiments measure the incidental neutrino flux produced in nuclear reactors during  $\beta$  decay. Here, the unstable isotopes  $\text{U}^{235}$ ,  $\text{U}^{238}$ ,  $\text{Pu}^{239}$ , and  $\text{Pu}^{241}$  decay into fission products (such as  $\text{Xe}^{140}$  and  $\text{Sr}^{94}$ ) which then produce neutrinos via inverse  $\beta$  decay. The average energy of the fission reaction is about 200 MeV, but most of this energy is retained as heat in the nucleus. The resultant  $\bar{\nu}_e$ s are therefore generally low energy,  $O(1)$  MeV. Unlike other types of neutrino experiments, reactor experiments only measure one flavor:  $\bar{\nu}_e$ . Reactor experiments are generally sensitive to  $\theta_{13}$  and  $\Delta m_{31}^2$ . Analogous to  $\theta_{12}$ 's moniker as the solar mixing angle,  $\theta_{13}$  is often referred to as the “reactor” mixing angle. In addition to KamLAND, reactor data primarily come from the RENO, Daya Bay, and Double Chooz<sup>1</sup> experiments. The current  $1\sigma$  best-fit parameters are  $\Delta m_{31}^2 = 2.517_{-0.028}^{+0.026} \times 10^{-5} \text{ eV}^2$  and  $\theta_{13} = 8.57_{-0.12}^{+0.12}$  degrees for the normal hierarchy, and  $\Delta m_{31}^2 = 2.498_{-0.028}^{+0.028} \times 10^{-5} \text{ eV}^2$  and  $\theta_{13} = 8.60_{-0.12}^{+0.12}$  for the inverted hierarchy.

An open question in reactor experiments pertains to the so-called “reactor anomaly,” in which the measured  $\bar{\nu}_e$  fluxes are consistently lower than predicted [37]. It remains an open question whether this is due to mis-modeling of neutrino flux predictions, or whether this points to the existence of one or more *sterile* neutrino state which do not interact via the weak force. The potential existence of sterile neutrino oscillations will be discussed further in Section 3.5.

## 3.3 Atmospheric Neutrino Experiments

When cosmic rays collide with air nuclei in the atmosphere, they produce hadronic showers which eventually result in an atmospheric neutrino flux. Most commonly, charged pions decay into muons and  $\nu_\mu$ s, and the muons further decay into electron,  $\nu_\mu$ s, and  $\bar{\nu}_e$ s. Atmospheric neutrino experiments detect the neutrinos produced in these interactions. The

---

<sup>1</sup>The global analysis in Reference [33] only uses RENO and Daya Bay. The best fit parameters are still reasonably consistent with Reference [32].

range of energies for atmospheric neutrinos is large—anywhere from MeV-scale to  $10^9$  GeV—but modern detection technology can only detect neutrinos up to about 100 TeV. For oscillation experiments, only energies up to about 100 GeV are considered.

While the Super-Kamiokande experiment was the first to definitively confirm neutrino oscillations [8], early indications of this phenomenon were seen in atmospheric experiments during the late 20th century. Experiments such as Kamiokande [38], IMB [39], and SOUDAN 2 [40] measured the ratio of muon-like to electron-like events (which provides a handle on the ratio of  $\nu_\mu$ -like events to  $\nu_e$ -like events) in both data and simulation, with the expectation that the ratio be unity if atmospheric neutrino fluxes were accurately predicted. While the number of  $\nu_e$  events generally met expectation, the number of observed  $\nu_\mu$  events was far below, resulting in measured ratios less than unity. It was later discovered that this was due  $\nu_\mu \rightarrow \nu_\tau$  oscillations, with a large mixing angle  $\theta_{23}$  (known now as the “atmospheric” mixing angle).

Global fits to atmospheric neutrino data mostly use Super-Kamiokande [41], IceCube [42], and the IceCube DeepCore upgrade [43]. The current  $1\sigma$  best-fit value for the atmospheric mixing angle is  $\theta_{23} = 49.2^{+0.9}_{-1.2}$  for the normal hierarchy and  $\theta_{23} = 49.3^{+0.9}_{-1.1}$  for the inverted hierarchy.

### 3.4 Accelerator-Based Neutrino Experiments

The final type of neutrino experiment, and the primary focus of this thesis, is the accelerator-based neutrino experiment. Accelerator-based experiments rely on a neutrino beam incident on some target nucleus contained within the detector. Here, the idea is to accelerate protons (usually stripped from Hydrogen gas) to near-light speed and smash them into a target such as beryllium, graphite, or water. This collision produces secondary particles—mostly pions and kaons—which are focused using magnetic focusing horns before finally decaying into  $\nu_\mu$  (or  $\bar{\nu}_\mu$ , depending on the direction of the current in the focusing horn—see Section 4.1). The remaining secondaries are collected in a beam dump, leaving a pure beam of (anti)muon neutrinos. These neutrinos then reach a detector where their interactions with the target nucleus allow for reconstruction of particle interactions. Accelerator neutrino energies vary

depending on the initial energy of the proton beam, but generally range from  $O(100)$  MeV to  $O(100)$  GeV. Since the neutrinos come from proton-nucleus collisions, the amount of data collected by accelerator experiments is generally reported in terms of protons-on-target (POT). Naturally, the number of POT varies between experiments, but is generally  $> 10^{20}$  POT.

As an example, Figure 3.1 shows an overview of the accelerator complex at Fermilab. The accelerator complex provides two neutrino beams: the Booster Neutrino Beam (BNB) and Neutrinos at the Main Injector (NuMI) [44]. The BNB takes protons accelerated up to 8 GeV by the Booster ring (red ring in Figure 3.1) which are then impinged upon a beryllium target. NuMI, on the other hand, takes protons accelerated up to 400 MeV from the Linear Accelerator (Linac, not shown in Figure 3.1) which are then further accelerated to 120 GeV by the Main Injector (blue ring in Figure 3.1).

Accelerator experiments are subdivided into *long-baseline* and *short-baseline* experiments. Equation 2.11, which describes the neutrino oscillation probability,  $P(\nu_{\alpha \rightarrow \beta})$ , contains two experimentally tunable parameters:  $L$ , the distance the neutrinos travel, and  $E$ , the neutrino energy. The ratio of these parameters,  $L/E$ , characterizes an experiment as either long ( $L/E \lesssim 10^3$  km/GeV) or short ( $L/E \lesssim 1$  km/GeV) baseline. Long-baseline experiments use two detectors, one placed close to the neutrino source (aptly named the *near* detector), and another detector usually hundreds of kilometers away from the source (the *far* detector). The two detectors allow for observation of both neutrino *appearance* and *disappearance*. In appearance experiments, a neutrino of flavor  $\beta$  appears in  $\nu_{\alpha} \rightarrow \nu_{\beta}$  oscillations. Disappearance searches start with a beam of  $\nu_{\beta}$  and observe  $\nu_{\alpha}$ . Short-baseline experiments, on the other hand, generally feature only one detector, usually within one kilometer of the neutrino source. Due to their short baseline, these experiments are not well-suited to oscillation searches under the three-flavor oscillation model. They are, however, applicable to searches for more massive (eV-scale) sterile neutrino states, which will be further discussed in the next section.

Global fits to oscillation parameters use data from the long-baseline NO $\nu$ A [45], T2K [46], and MINOS [47] experiments.<sup>2</sup> These experiments are sensitive to  $\Delta m_{31}^2$ ,  $\theta_{23}$ ,  $\theta_{13}$ ,

---

<sup>2</sup>Reference [33] also uses K2K [48] data.

the CP-violating phase  $\delta_{CP}$ , and the neutrino mass hierarchy. The most recent results from T2K and NO $\nu$ A introduce an additional tension in the  $\delta_{CP}$  parameter, reducing the sensitivity relative to previous global fits [49] under the normal hierarchy hypothesis. While this particular tension is not present in the inverted hierarchy, other experimental data show a slight preference for the normal ordering. Future experimental data, particularly from the upcoming long-baseline Deep Underground Neutrino Experiment (DUNE) and Hyper-K, are needed to resolve this tension.

Table 3.1 summarizes the  $1\sigma$  best-fit values for the neutrino oscillation parameters. In general, measurements of the three-flavor oscillation parameters are well-constrained, with the notable exception of  $\delta_{CP}$ . Accurate measurements of this phase may help explain the matter-antimatter asymmetry of the universe. Future measurements from DUNE and Hyper-K are expected to constrain this uncertainty.

Up until this point, we have focused on three-flavor oscillations predicted by the standard model. However, some experiments have observed anomalous results under the standard three-flavor model. Such anomalies may indicate additional, non-active (sterile) neutrino states or previously unknown or mis-modeled background sources. Different sterile neutrino models assume different numbers of sterile states. For example, models that assume one sterile neutrino state are referred to as 3+1 models (one sterile state in addition to the three known mass states). In the next section, we discuss the MiniBooNE low-energy excess anomaly, one of the primary indications for the existence of sterile neutrino states, and the main motivation for the MicroBooNE experiment.

### 3.5 Short-Baseline Accelerator Anomalies

In 2001, the Liquid Scintillator Neutrino Detector (LSND, shown schematically in Figure 3.2a) experiment reported a significant excess of  $\bar{\nu}_e$  appearance events [50], shown in Figure 3.2b. As a short-baseline experiment ( $L = 30$  m,  $E_\nu \sim 50$  MeV), LSND should have been insensitive to  $\nu_\mu \rightarrow \nu_e$  oscillations under the standard three-flavor model. However, an additional mass-squared splitting of  $\Delta m_{41}^2 \sim 1$  eV<sup>2</sup>—several orders of magnitude larger than either of the three-flavor mass-squared splittings—could increase the flavor oscillation



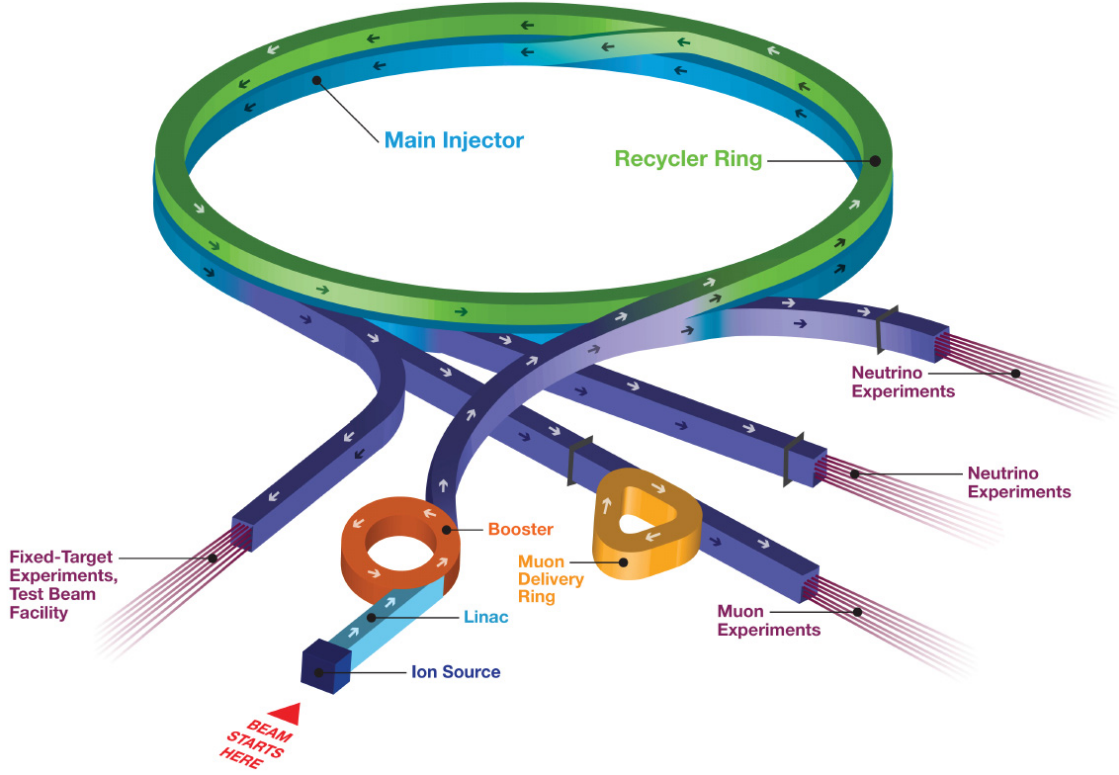


Figure 3.1: Cartoon depiction of the Fermilab accelerator complex.

Table 3.1: Summary of best-fit (BF) three-flavor oscillation parameters, along with  $3\sigma$  ranges. Values taken from Reference [32].

Parameter	Normal Hierarchy		Inverted Hierarchy	
	$1\sigma$ BF	$3\sigma$ range	$1\sigma$ BF	$3\sigma$ range
$\theta_{12}$ ( $^\circ$ )	$33.44^{+0.77}_{-0.74}$	$31.27 - 35.86$	$33.45^{+0.78}_{-0.75}$	$31.27 - 35.87$
$\theta_{23}$ ( $^\circ$ )	$49.2^{+0.9}_{-1.2}$	$40.1 - 51.7$	$49.3^{+0.9}_{-1.1}$	$40.3 - 51.8$
$\theta_{13}$ ( $^\circ$ )	$8.57^{+0.12}_{-0.12}$	$8.20 - 8.93$	$8.60^{+0.12}_{-0.12}$	$8.24 - 8.96$
$\delta_{CP}$ ( $^\circ$ )	$197^{+27}_{-24}$	$120 - 369$	$282^{+26}_{-30}$	$193 - 352$
$\Delta m_{21}^2$ ( $10^{-5}$ eV $^2$ )	$7.42^{+0.21}_{-0.20}$	$6.82 - 8.04$	$7.42^{+0.21}_{-0.20}$	$6.82 - 8.04$
$\Delta m_{31}^2$ ( $10^{-3}$ eV $^2$ )	$2.517^{+0.026}_{-0.027}$	$2.435 - 2.598$	$2.498^{+0.028}_{-0.028}$	$2.414 - 2.518$

frequency such that additional *sterile* neutrino states appear at short-baseline values. Indeed, LSND concluded that the observed excess was consistent with a  $\Delta m^2 > 0.4 \text{ eV}^2$ . In order to explain this excess, the MiniBooNE experiment was constructed at Fermilab along the Booster Neutrino Beamline (BNB). MiniBooNE began taking data in 2002, accruing over  $18.75 \times 10^{20}$  protons-on-target (POT) worth of data. Similarly to LSND, MiniBooNE’s short-baseline values of  $L = 541 \text{ m}$  and an average  $E_\nu$  of 700 MeV should make it insensitive to  $\nu_e$  appearance. Although designed to explain the observed LSND excess, MiniBooNE itself saw a  $4.8\sigma$  excess [51] of  $\nu_e$ -like events in the region of  $E_\nu \lesssim 600 \text{ MeV}$ . This is shown in Figure 3.3b. The observed data are significantly higher than both the predicted background and a best-fit 3+1 oscillation model (dashed line in Figure 3.3b). This anomaly is referred to as the MiniBooNE low-energy excess (LEE).

MiniBooNE was a mineral oil Cherenkov detector, a type of detector which has difficulty distinguishing photons from electrons. When charged particles travel faster than the speed of light in a medium, they emit Cherenkov radiation, similar to a sonic boom created when an object travels faster than the speed of sound through air. Cherenkov radiation is emitted in a cone of light, which is then detected by a photodetection system in a Cherenkov detector, usually an array of photomultiplier tubes (PMTs). The Cherenkov radiation then manifests as a ring of activity in the PMTs. Minimally-ionizing muon tracks are indicative of  $\nu_\mu$  interactions, while electromagnetic showers indicate either  $\nu_e$  or  $\pi^0$  events. Because muons travel in a long, straight path, the resulting Cherenkov ring is distinctly sharp. Showers, on the other hand, create “fuzzy” rings due to the cascade of particles. Electron showers from  $\nu_e$  interactions create a single fuzzy ring, while two-photon events from a  $\pi^0 \rightarrow \gamma\gamma$  decay will appear as two overlapping fuzzy rings. Figure 3.4 shows illustrations of each of these types of Cherenkov rings. This presents a problem in distinguishing electrons from photons: in cases where only one of the two  $\pi^0$  showers is detected—due, for example, to one shower exiting the detector—the resultant ring is identical to that of an electron shower. Similarly, if the two rings overlap completely, this mimics an electron ring. Thus, Cherenkov detectors show considerable difficulty in distinguishing photons from electrons. The MiniBooNE LEE must therefore be either electron-like or photon-like in nature. The former case would indicate an excess of  $\nu_e$  appearance events at a short baseline, providing evidence for sterile neutrino

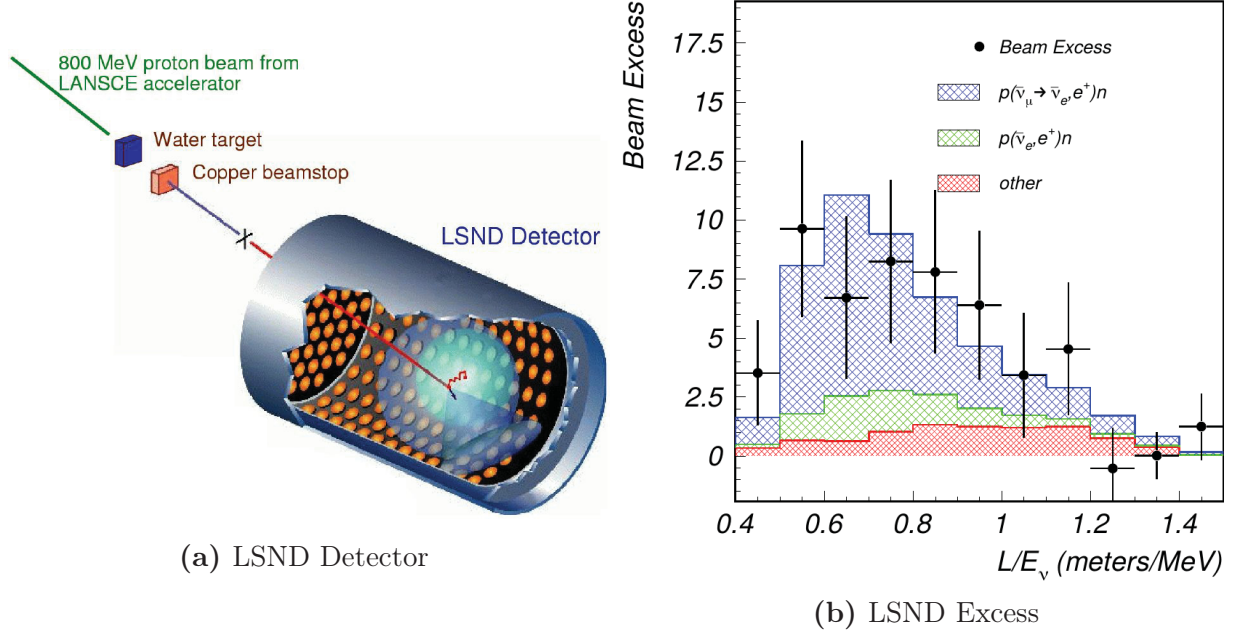


Figure 3.2: (a) Schematic of the LSND detector and (b) excess of electron antineutrino events observed by the LSND experiment in terms of  $L/E$ . The red and green histograms represent the expected background, while the blue histogram assumes an additional neutrino oscillation with  $\Delta m^2 \sim 1 \text{ eV}^2$ .

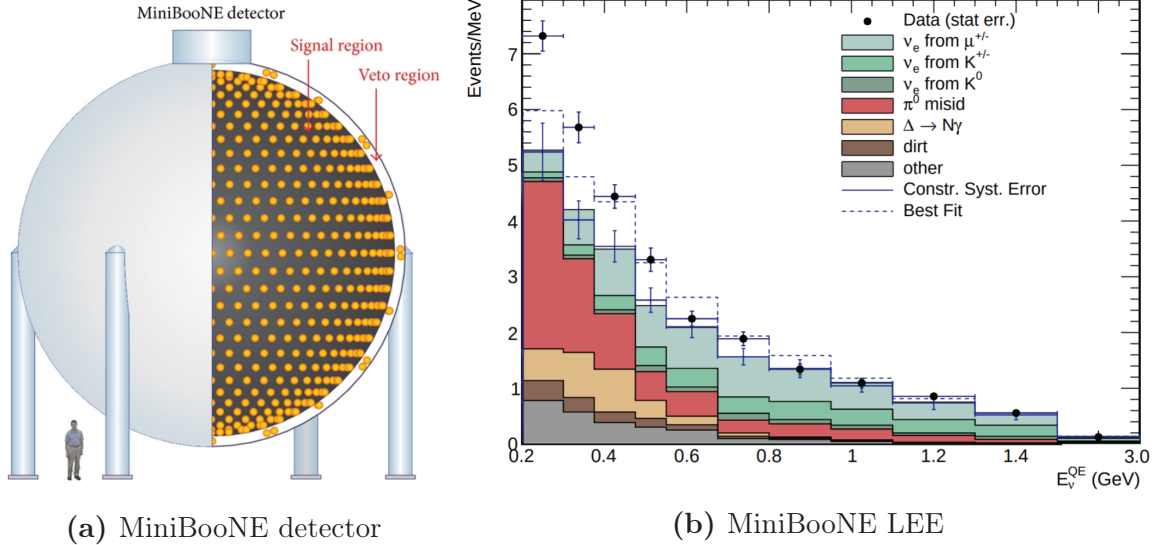


Figure 3.3: (a) Schematic of the MiniBooNE detector and (b) observed excess of low-energy electron neutrino-like events in MiniBooNE. The colored histograms represent the predicted backgrounds, while the black dots show the measured data. Systematic uncertainties (vertical lines on the background histogram) primarily come from uncertainties on the neutrino flux and cross sections and are constrained using in-situ cross-section measurements. The dashed line represents the best-fit assuming a 3+1 oscillation model.

oscillations; the latter implies a previously unknown or mis-modeled background source. As of this writing, the source of the MiniBooNE LEE remains an open question in experimental neutrino physics.

To investigate the MiniBooNE LEE, the MicroBooNE experiment was built along Fermilab's Booster Neutrino Beamline, the same beamline MiniBooNE used. Using the liquid argon time projection chamber (LArTPC) technology, MicroBooNE can distinguish electron showers from photon showers, and will provide resolution to the LEE anomaly. In the next chapter, we discuss the MicroBooNE experiment, including the relevant neutrino beams and detector subsystems.

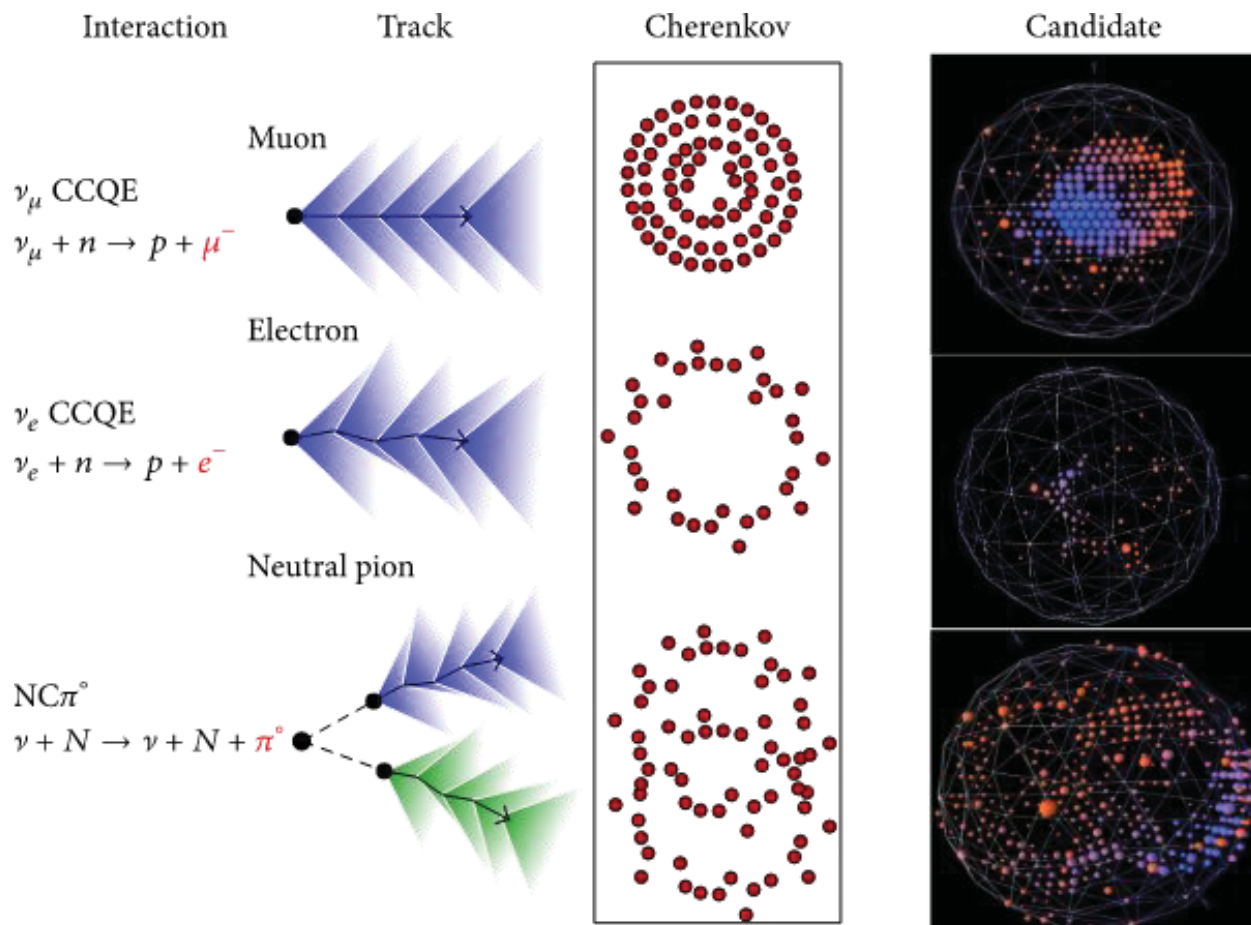


Figure 3.4: Illustration of Cherenkov rings as detected by MiniBooNE.

## Chapter 4

# The MicroBooNE Experiment

MicroBooNE, the successor to MiniBooNE, is a liquid argon time projection chamber (LArTPC) operating along Fermilab’s Booster Neutrino Beamline (BNB) since 2015. It is the first detector in Fermilab’s planned Short-Baseline Neutrino (SBN) program [52], which will also include the upcoming Short-Baseline Near Detector (SBND) and the recently commissioned ICARUS experiment, both of which are also LArTPCs. LArTPCs are one of the preferred technologies for current and future neutrino oscillation experiments due to their exceptional position and calorimetric resolution capabilities. LArTPCs can accurately measure the energy deposition per unit length,  $dE/dx$ , of reconstructed particle objects, which allows for accurate particle identification. In particular, the  $dE/dx$  profile for electromagnetic showers in the active volume can distinguish electron showers from photon showers, which will aid in answering the question of whether MiniBooNE’s LEE is electron-like or photon-like. While investigation of the LEE is MicroBooNE’s primary physics goal, it will also serve the neutrino community by providing high-statistics neutrino-argon cross section measurements, along with LArTPC detector physics measurements and calibration techniques. These measurements will inform future LArTPC experiments such as SBND, ICARUS, and the upcoming Deep Underground Neutrino Experiment (DUNE).

MicroBooNE receives neutrinos from both the BNB and NuMI (see Section 3.4) beamlines. The BNB is on-axis relative to MicroBooNE, meaning that the beam direction is parallel to the  $z$  (forward) direction in the MicroBooNE detector. NuMI, on the other hand, is an off-axis neutrino beam. This can be used for independent measurements such as

studying the normalization of CC and NC  $\pi^0$  production. One such measurement includes the recent  $\nu_e + \bar{\nu}_e$  CC inclusive cross section on argon [53]. Data from the NuMI beam also provide an excellent opportunity to simultaneously measure the electron-neutrino cross section due to its higher electron-neutrino flux component. This sample is also very useful in developing and testing  $\nu_e$  reconstruction methods. While NuMI provides an additional neutrino source for various cross-checks, the work performed in this thesis does not use NuMI data. We therefore choose to focus on the BNB.

Section 4.1 describes the design and operation of the BNB, which provides a pure beam of  $\nu_\mu$  to the MicroBooNE detector. The remaining sections discuss the MicroBooNE detector itself (Section 4.2), the major detector subsystems (Sections 4.3–4.8), and the current status of the MicroBooNE detector 4.9.

## 4.1 Booster Neutrino Beam at Fermilab

MicroBooNE sits along Fermilab’s Booster Neutrino Beamline, the same beamline used by the MiniBooNE experiment. A diagram of the beamline is shown in Figure 4.1.  $H^-$  ions are first accelerated through the Fermilab Linear Accelerator (Linac) to a kinetic energy of 400 MeV [54]. These  $H^-$  ions then pass through carbon foil, stripping the electrons and leaving bare protons. These protons are then injected into the Booster, a 474 m circumference synchrotron operating at 15 Hz. The Booster accelerates the protons to 8 GeV momentum and sorts them into *beam spills* of  $4 \times 10^{12}$  protons spaced apart by a  $1.6 \mu s$  time window, commonly referred to as the *beam window*. The spills are then sent to a target hall containing a beryllium target.

The proton-beryllium (p-Be) collision produces secondary particles (see Section 3.4) which are then focused toward the beamline by a magnetic focusing horn pulsed at  $\pm 174$  kA. A positive current produces a beam composed mostly of neutrinos (“neutrino mode”), while a negative current leads to a beam of anti-neutrinos (“anti-neutrino mode”). A schematic of the focusing horn is shown in Figure 4.2. In neutrino mode, the current runs along the beam direction in the inner conductor before being redirected up and backward along the outer conductor. This produces a magnetic field perpendicular to the beam direction which focuses



same-charged particles along the beam direction and defocuses oppositely-charged particles. The focused particles then travel through a 50 m cylindrical decay pipe filled with air, in which they decay into the neutrinos which comprise the beam. Any remaining backgrounds stop either in the concrete at the end of the decay pipe (the “beam dump”), or in the dirt between the decay pipe and the detector. The neutrinos that reach the detector comprise the beam.

Secondary particles in the decay pipe generally decay into  $\nu_\mu$ , but some contributions from other neutrino flavors also reach the detector. Figure 4.3 shows the predicted composition of the BNB while running in neutrino mode. In this mode, the beam is predicted to be  $\sim 93.6\%$   $\nu_\mu$ , with the next largest contribution coming from  $\bar{\nu}_\mu$  (5.86%), especially at energies below 200 MeV;  $\nu_e$  (0.52%) and  $\bar{\nu}_e$  (0.05%) contributions are orders of magnitude below  $\nu_\mu$  components. The  $\nu_\mu$ s are mainly produced via  $\pi^+ \rightarrow \mu^+ + \nu_\mu$ , which has a branching ratio of  $\sim 99.98\%$ . The  $\bar{\nu}_\mu$  component generally comes from highly-energetic, forward-going  $\pi^-$ , which fail to be completely defocused by the magnetic horn.  $\nu_e$ s are produced via  $\pi^+ \rightarrow e^+ + \nu_e$ , a process which is suppressed by a factor of  $10^{-4}$  relative to the  $\pi^+ \rightarrow \mu^+ + \nu_\mu$  process due to helicity conservation. Finally, the  $\bar{\nu}_e$ s come from semileptonic decays of  $K_L^0$ .

The BNB has been operational for nearly twenty years, and as such, is one of the most well-understood neutrino beams in the world. Still, due to the difficulty of predicting neutrino fluxes, sizeable systematic uncertainties exist in the flux prediction. The dominant source of systematic uncertainty in the BNB flux prediction comes from the uncertainty in hadron production following the p-Be collision. Other subleading contributions come from uncertainties in the proton delivery rate, horn current modeling, and the the total interaction cross section for p-Be collisions. MiniBooNE performed a simulation-based constraint on the total flux uncertainty [54]. MicroBooNE, following a similar procedure, obtained updated systematic uncertainty values [56], which are listed in Table 4.1. The work in this thesis is focused on the  $\nu_\mu$  component of the beam, where the dominant systematic is  $\pi^+$  production.



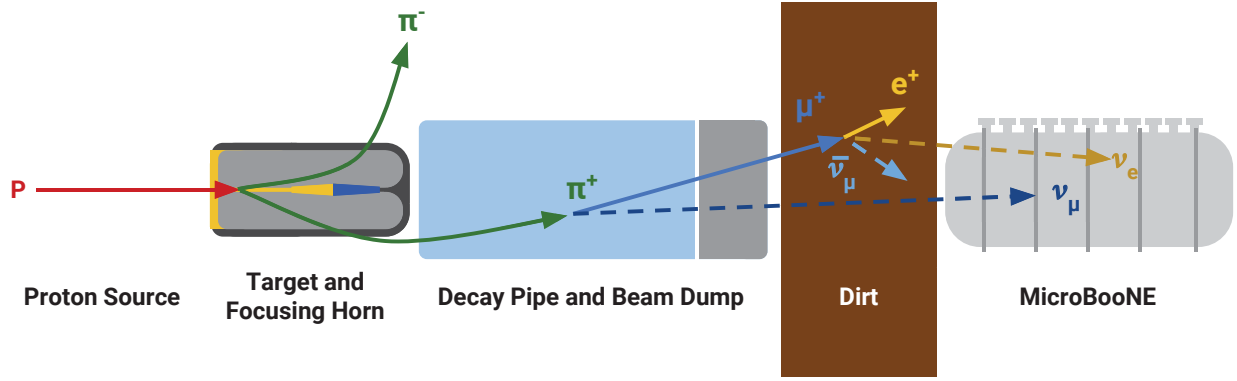


Figure 4.1: Diagram of the Booster Neutrino Beamline.

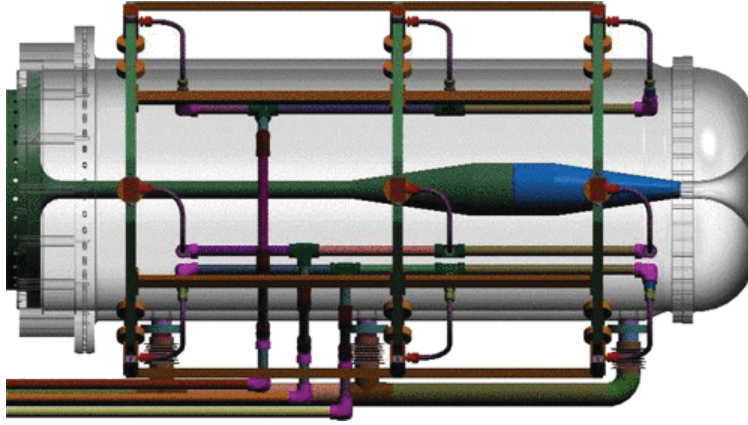


Figure 4.2: Diagram of the magnetic focusing horn used in the BNB, taken from Reference [54]. In neutrino mode, the positive 174 kA current flows from left to right along the inside of the aluminum conductors (gray), then from right to left along the outside. The inner conductor components are shown in blue and green in the center.

Table 4.1: Contributions to the total systematic uncertainty on the BNB flux. Hadron production uncertainties are the dominant source for each  $\nu$  flavor. Here, “other” includes horn current modeling and the total p-Be interaction cross-section.

Systematic	$\nu_\mu/\%$	$\bar{\nu}_\mu/\%$	$\nu_e/\%$	$\bar{\nu}_e/\%$
Proton delivery	2.0	2.0	2.0	2.0
$\pi^+$	11.7	1.0	10.7	0.03
$\pi^-$	0.0	11.6	0.0	3.0
$K^+$	0.2	0.1	2.0	0.1
$K^-$	0.0	0.4	0.0	3.0
$K_L^0$	0.0	0.3	2.3	21.4
Other	3.9	6.6	3.2	5.3
Total	12.5	13.5	11.7	22.6

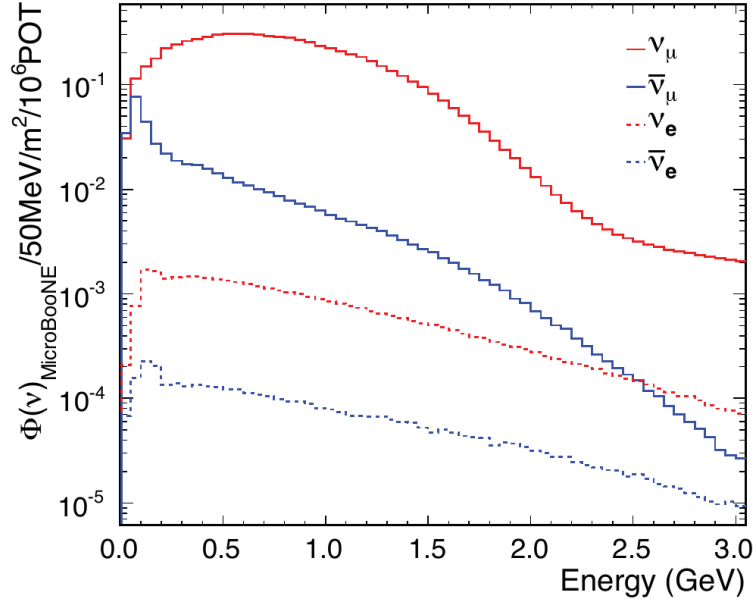


Figure 4.3: Predicted neutrino-mode BNB composition received by MicroBooNE, taken from Reference [55]. The average  $\nu_\mu$  energy is  $\sim 700$  MeV.

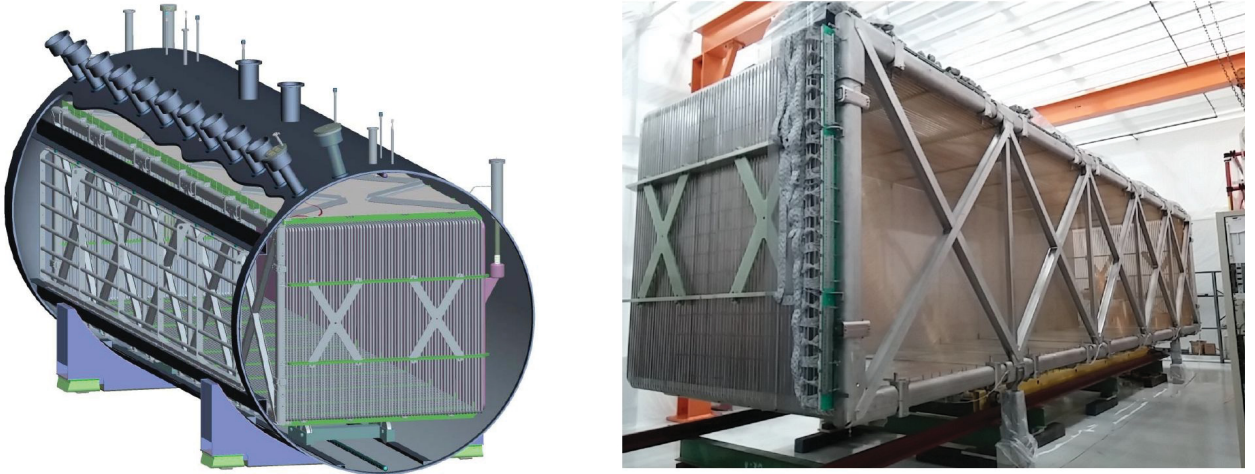


Figure 4.4: (Left) schematic view of the MicroBooNE detector and (right) picture of the TPC as viewed through the anode plane. Note that the cylindrical cyrostat is not shown in the picture on the right.

## 4.2 The MicroBooNE Detector

470 m downstream of the BNB sits the MicroBooNE detector, a 170 tonne (90 tonne active volume) LArTPC shown in Figure 4.4. The detector active volume dimensions are 2.56 m (drift direction,  $x$ ), 2.16 m (vertical direction,  $y$ ), and 10.36 m (beam direction,  $z$ ). Anode and cathode plane arrays establish an electric field (E-field) within the TPC volume. The cathode plane is held at a constant -70 kV with a field cage composed of 64 2.54 cm diameter stainless steel tubes that uniformly steps down the voltage in 1 kV increments to ground at the anode. The anode consists of three readout wire planes, labeled as U, V, and Y. The U and V planes contain 2400 wires each and are oriented at  $\pm 60^\circ$  relative to the vertical, while the collection plane contains 3456 wires oriented vertically, making for a total of 8256 wires. Each wire is made of copper-plated stainless steel. They are 150  $\mu\text{m}$  thick, with 3 mm separation between each wire and wire plane. Behind the anode is an array of 32 8-inch photomultiplier tubes (PMTs) for light collection and amplification (not shown in Figures 4.4 or 4.5). The TPC is housed in a cylindrical cryostat 12.3 m long and 3.81 m in diameter insulated with 41 cm of polyurethane foam. The cryostat maintains a temperature of 89 K.

LArTPCs seek to reconstruct particle interactions, or “events,” by analyzing signals induced on the anode wires. When a neutrino interacts with the argon, it can produce charged particles. As charged particles traverse the detector volume, they ionize the argon atoms. The resultant ionization electrons are then drifted toward the anode plane under the influence of the applied E-field. As electrons travel through the anode, charge is induced on U and V (“induction”) planes, and the electrons are subsequently collected on the Y (“collection”) plane. These signals induced on the readout wires are used to reconstruct neutrinos events, as shown in Figure 4.5. By analyzing the induced signal on each plane, we can reconstruct accurate 3D *event displays* of particle interactions. A sample collection-plane event display is shown in Figure 4.6.

In order for ionization electrons to travel unimpeded to the anode wires, the argon in the TPC must be kept pure. Electronegative contaminants such as  $\text{O}_2$  and  $\text{H}_2\text{O}$  can reduce the electron drift lifetime, thereby attenuating the collected signal. The MicroBooNE

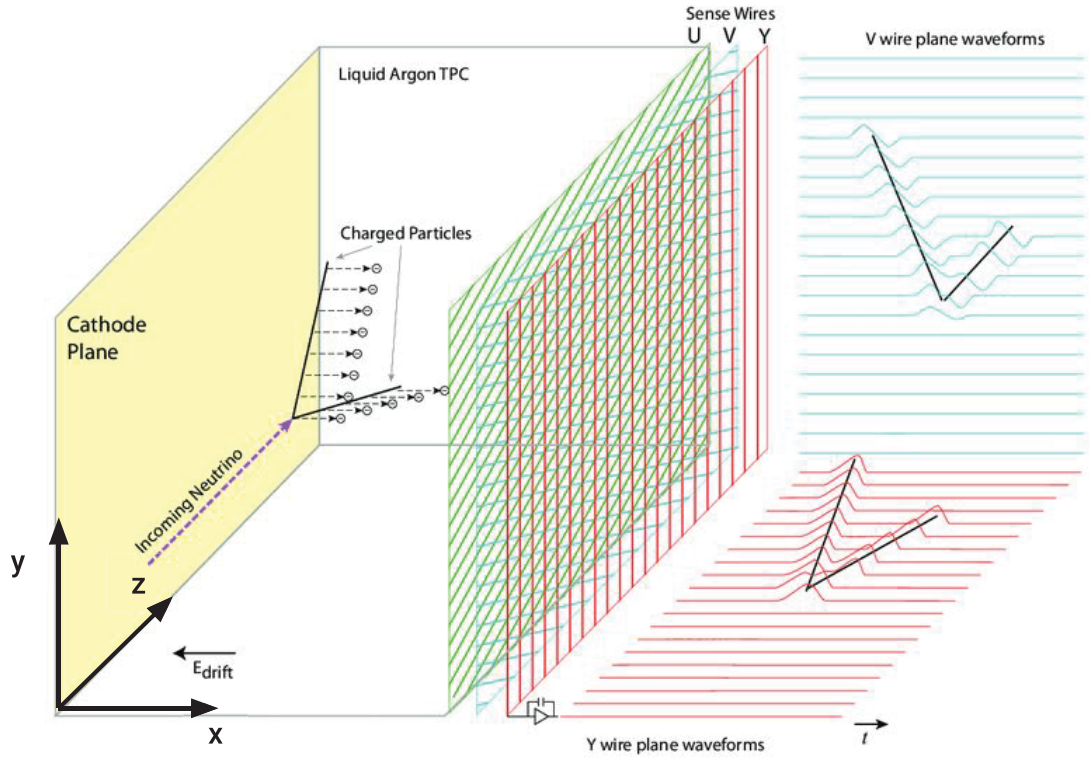
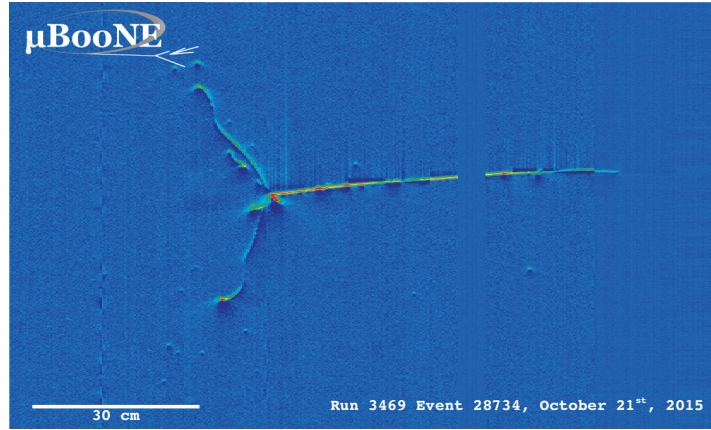
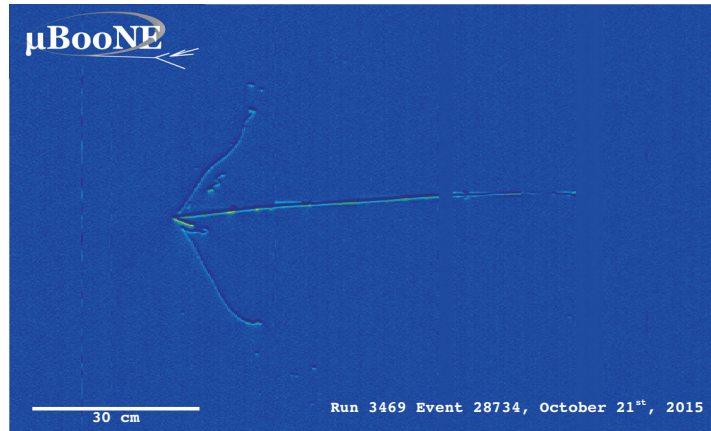


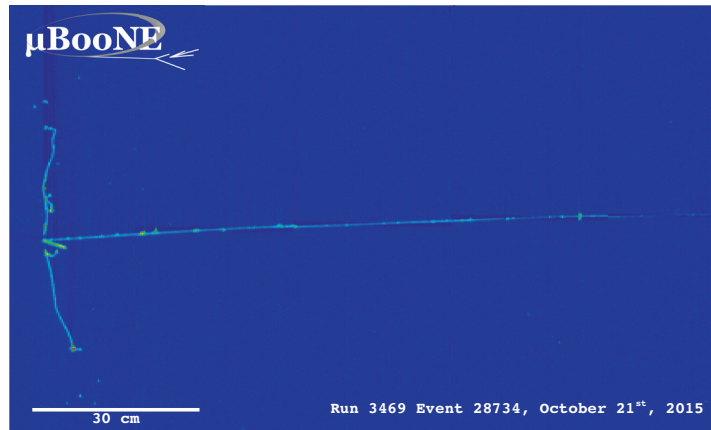
Figure 4.5: Diagram of LArTPC detection. Ionization electrons are drifted toward three readout wire planes via an electric field and the signals are used to reconstruct the neutrino event. Also shown is the MicroBooNE coordinate system in which  $x$  is the electric field direction,  $y$  is the vertical direction, and  $z$  is the beam direction.



(a) U plane



(b) V plane



(c) Y plane

Figure 4.6: Sample MicroBooNE event display showing a candidate  $\nu_\mu$  interaction from on-beam data as viewed on each wire plane. The  $y$  axis corresponds to drift time, the  $x$  axis to wire number. Event displays on the collection (Y) plane can be thought of as a top-down view of the candidate neutrino interaction.



purification subsystem consists of two pumps that circulate the argon through filters designed to remove  $\text{O}_2$  and  $\text{H}_2\text{O}$  impurities [57].

MicroBooNE has been taking data since August 2015, with over  $7.5 \times 10^{20}$  POT. Since March 2017, the detector has been taking data with a cosmic ray tagger (CRT) system, accumulating  $2.5 \times 10^{20}$  POT with the CRTs.

### 4.3 Data Acquisition Readout Electronics

In order to convert the LArTPC signals into a useful data format, MicroBooNE employs readout electronics which amplify, shape, and digitize the data. These electronics are split into *warm* and *cold* components. The cold electronics consist of complementary metal-oxide semiconductor (CMOS) application-specific integrated circuits (ASICs). These analog front-end ASICs contain a signal pre-amplifier, shaper, and driver, and are mounted on front-end motherboards (FEMBs) placed in the liquid argon near the anode-plane wires. This proximity to the wires, along with the low liquid argon temperature, reduces the impact of electronics noise.

From the ASICs, induction signals are transmitted through twisted-pair copper cables to a warm flange, and then to an intermediate amplifier line-driver, which accounts for signal losses as the signals are finally carried through additional twisted-pair copper wires to data acquisition (DAQ) machines. The DAQ contains an analog-to-digital converter (ADC), which digitizes the signals at a rate of 2 MHz; this sampling rate defines one time “tick” as  $0.5 \mu\text{s}$ . The DAQ readout window of 9600 ticks (or 4.8 ms) corresponds to roughly twice the maximum ionization electron drift time of 2.3 ms. Finally, the digitized signals are passed to a field-programmable gate array (FPGA) for data processing and reduction. A schematic of the MicroBooNE readout electronics chain, taken from Reference [58], is shown in Figure 4.7.

The ASICs can operate at one of four gain settings (4.7 mV/fC, 7.8 mV/fC, 14 mV/fC, or 25 mV/fC) and four peaking times ( $0.5 \mu\text{s}$ ,  $1.0 \mu\text{s}$ ,  $2.0 \mu\text{s}$ , or  $3.0 \mu\text{s}$ ), where the peaking time is defined as the time difference between 5% of the signal peak and the magnitude of the signal peak value. The choice of gain setting impacts the signal pulse heights, while

the peaking time affects pulse widths. The gain setting must be chosen so as to provide a high signal-to-noise ratio for minimally-ionizing particle tracks (such as cosmic muons), while avoiding channel saturation from highly-ionizing sources (such as stopping protons); the chosen gain value is 14 mV/fC. Meanwhile, the peaking time must be optimized so as to accurately record the induced signals without introducing additional noise. Considering the MicroBooNE sampling frequency of 2 MHz, one may assume that a peaking time of 1  $\mu$ s is the obvious choice. However, due to the relatively slow drift time of ionization electrons (2.3 ms at the maximum drift distance), signal waveforms are broadened due to a) the longer drift time through the anode wires and b) signal-attenuating effects such as electron diffusion. Thus, we choose a peaking time of 2  $\mu$ s, as little information is lost relative to a 1  $\mu$ s peaking time, and the longer peaking time reduces inherent noise [58].

Post-installation studies into the readout electronics have revealed a number of misconfigured, shorted, and dead wire regions in the detector. Misconfigured regions are those in which the raw data read from the ASICs is consistent with the factory default gain and peaking time settings of 4.7 mV/fC and 1  $\mu$ s, respectively. These channels—of which there are 224, all in the first induction plane toward the downstream end of the TPC—have higher inherent noise and a lower signal-to-background ratio than properly configured channels. Signals from these channels are corrected using an offline noise filter. In addition to the misconfigured region, the first induction plane also contains a region of shorted wires, believed to be the result of direct contact between many U-plane wires and a single V-plane wire. This results in some electrons being collected on the U plane instead of the Y plane in this so-called “U-short region.” The collection plane signals in this region therefore show a lower signal amplitude than those in other regions. Finally, the collection plane contains a region of functionally dead wires in which signals are almost entirely collected on the V plane. In the MicroBooNE coordinate system, this dead region corresponds to  $z$ -positions ranging from roughly 675 to 775 cm.

## 4.4 Ionization Signal

For each MeV of energy deposited by a charged particle in the MicroBooNE active volume, approximately  $4 \times 10^4$  ionization electrons are produced. These electrons drift toward the anode plane under the influence of an applied electric field, with a maximum drift time of 2.3 ms. As the electrons drift through each of the three anode wire planes, they induce signals on nearby wires, manifesting as a bipolar response on the induction planes (a positive response as the electrons approach the plane, and a negative response as they leave) and a unipolar response on the collection plane; see Figure 4.8. These response functions are convolutions of a *field response* and an *electronics response*. The field response describes the charge induced on one anode-plane wire by a single ionization electron, while the electronics response describes the impact on the signal waveform due to shaping and amplification during signal readout [59]. Each plane maintains a constant bias voltage to ensure transparency to drifting electrons. The V and Y planes are shielded by the voltages on adjacent planes, but the U plane can see ionization clouds from farther away, broadening the response function on that plane.

As ionization electrons drift through the detector volume toward the anode plane, several physical effects can modify or distort the shape of the electron clouds, which in turn modifies the signal. For example, local E-field distortions caused by  $\text{Ar}^+$  ions which drift slowly (relative to ionization electrons) toward the cathode plane; we refer to this as the space charge effect (SCE) [60]. SCE impact the spatial resolution of the detector, resulting in reconstructed tracks that may appear bent or bowed. These local distortions also alter the local drift velocity of the ionization electrons.

Ionization electrons may also recombine with argon ions, reducing the collected charge. Electron-ion recombination, or simply “recombination,” depends on the local density of ionization electrons and the local E-field strength. In addition to recombining with  $\text{Ar}^+$  ions, ionization electrons can also attach to electronegative contaminants such as  $\text{O}_2$  and  $\text{H}_2\text{O}$ , attenuating the collected signal. The electron lifetime—which describes the expected drift time of ionization electrons and serves as a measure of argon purity—has been measured



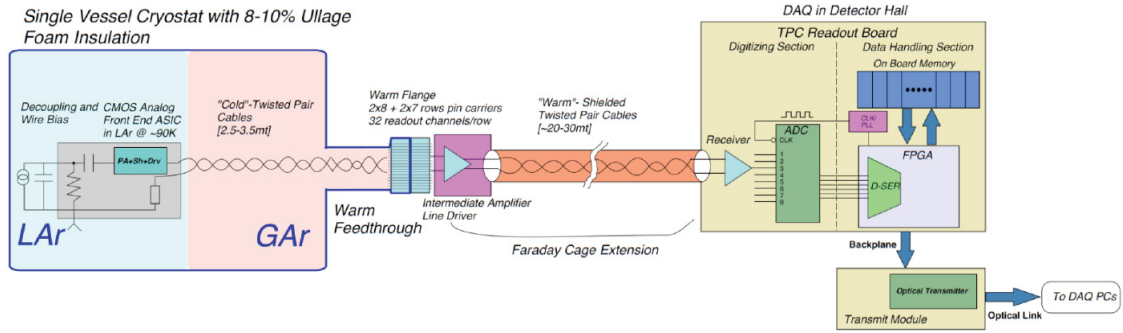


Figure 4.7: Schematic of the MicroBooNE readout electronics chain, taken from Reference [58].

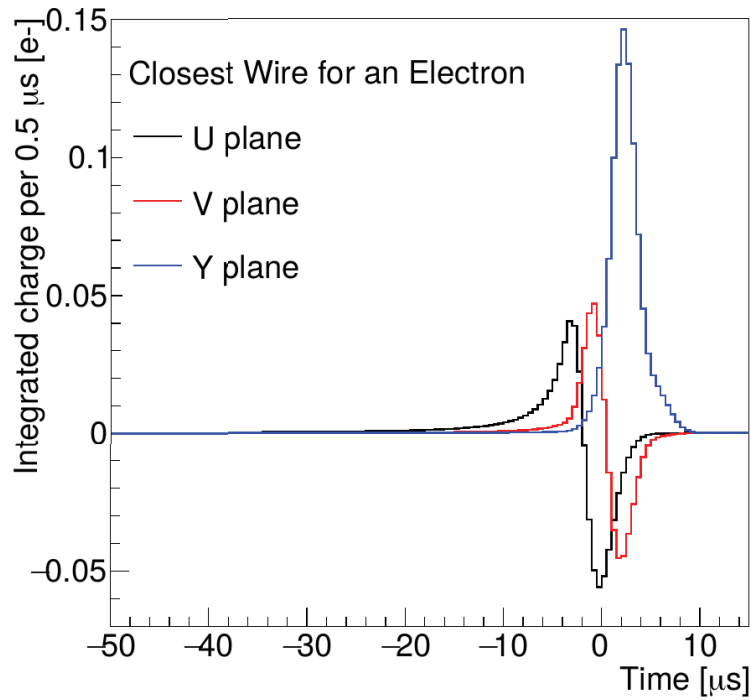


Figure 4.8: Average time-domain response on the induction (black and red) and collection (blue) planes.

to be 18 ms in MicroBooNE [61], significantly greater than the maximum drift time of 2.3 ms.

Finally, electron diffusion acts to spread the ionization clouds as a function of drift time. The transverse component (in the plane parallel to the anode wire plane) spreads charge to neighboring wires, while the longitudinal component (perpendicular to the wire plane) widens signal pulses in time. We discuss electron diffusion in detail in Chapter 6.

## 4.5 Light Collection

Liquid argon is a bright scintillator, producing  $\mathcal{O}(10^4)$  photons per MeV of deposited energy at the nominal E-field of 273 V/cm. Additionally, argon is transparent to its own scintillation light, making light collection an efficient method of determining the precise time an event occurred, denoted as  $t_0$ . The MicroBooNE light collection system consists of 32 8-inch Hamamatsu 5912-02MOD PMTs arranged behind the collection plane, shown in Figure 4.9. The PMTs are most efficient at detecting light with wavelengths between 350 and 450 nm; however, argon ionization light peaks in the vacuum ultraviolet (VUV) spectrum at 128 nm. Thus, each PMT is coated with wavelength-shifting Tetraphenyl Butadiene (TPB) to bring the ionization light into the visible spectrum of the PMTs.<sup>1</sup>

Scintillation light in argon is produced when excited diargon atoms (dimers) radiatively decay following excitation [62]. There are two methods by which dimers form in liquid argon: self-trapping and recombination luminescence. In self-trapping, a charged particle excites an argon atom, which then “traps” a ground state argon atom, forming a dimer. In recombination luminescence, an ionized argon atom combines with an ionization electron and a ground state argon atom to form the dimer. In both cases, the excited dimer can form in either a singlet or triplet state, with the singlet (triplet) state having a decay time of 6 ns (1.6  $\mu$ s). These processes are shown in Figure 4.10. Scintillation light originating from the decay of the singlet state is referred to as *prompt* light, while the slower light from the triplet state decay is called *late* light.

---

<sup>1</sup>The PMT array also contains 4 light guide paddles, but these are intended for R&D for future, larger LArTPCs, since the PMT-plus-plate system is not scaleable.

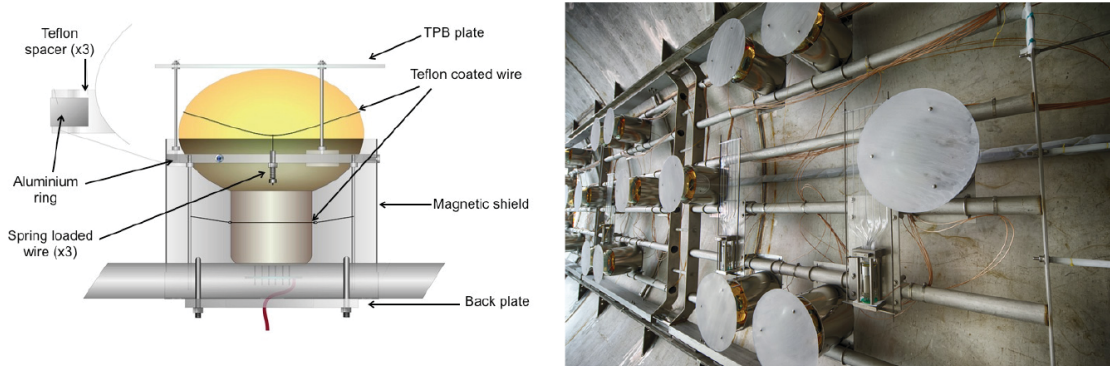
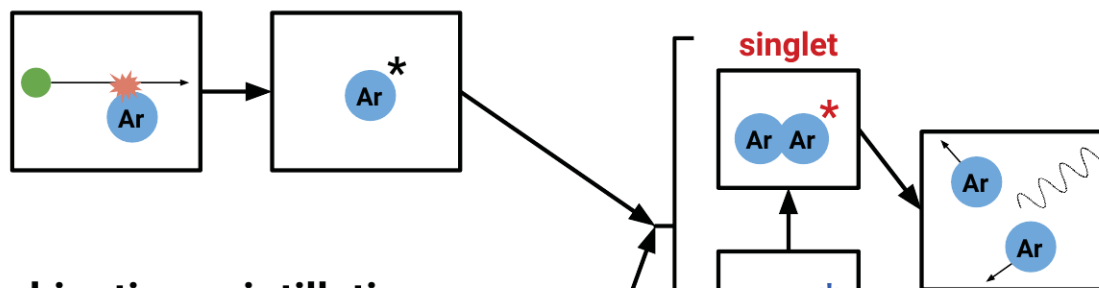


Figure 4.9: (Left) PMT diagram and (right) picture of the PMT array mounted behind the anode plane in the MicroBooNE TPC, take from Reference [57].

## Self-trapped scintillation



## Recombination scintillation

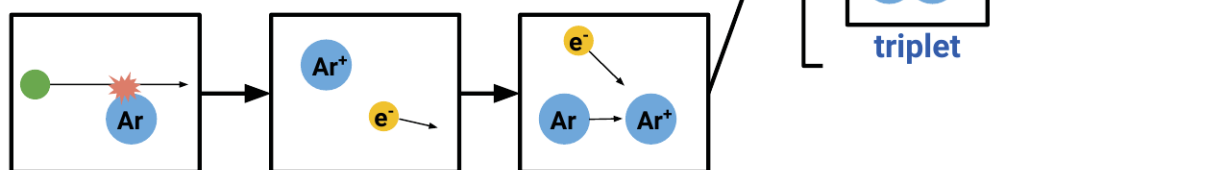


Figure 4.10: Cartoon of scintillation light production processes in liquid argon. Image credit: Reference [63].

## 4.6 UV Laser System

As discussed in Section 4.4, MicroBooNE’s position resolution is impacted by SCE, primarily caused by slow-drifting  $\text{Ar}^+$  ions. These ions distort the local E-field in their vicinity, leading to non-uniformities in both the electric field and the drift velocity. To study this effect, and to obtain a correction map for E-field and drift velocity values, MicroBooNE employs an ultraviolet (UV) laser system [64], shown diagrammatically in Figure 4.11. Before entering the cryostat, the beam is directed, attenuated, and sized in a dark laser box. Two mirrors (labeled M1 and M2 in Figure 4.11) align the beam, while an attenuator controls the beam energy and an aperture controls the size. M1 transmits IR light while reflecting green and UV; the infrared (IR) light is collected by the first beam dump (BD1). M2 then transmits green light (collected by BD2) and reflects the UV light to a third mirror, M3, which directs the beam to the 2.5 m feedthrough. At the end of the feedthrough is a rotatable cold mirror submerged in the liquid argon, which can be used to control the beam angle in the TPC.

The UV laser system has been used to measure local drift velocity and E-field variations, as well as the average drift velocity of ionization electrons. Because the angle of the beam is a known quantity, we can compare the reconstructed track to the “true” laser track to observe track distortions as a function of TPC position. The result of this is a *distortion map*, which can then be used to obtain a *correction map* to correct the local drift velocity and E-field values. Details of obtaining these maps can be found in Reference [64].

## 4.7 Cosmic Ray Tagger System

As a surface-level detector, MicroBooNE is exposed to a high rate of cosmic ray flux. In a given 4.8 ms readout, we expect about 20 cosmic muons to enter the detector. In order to tag these cosmic tracks and reduce backgrounds, MicroBooNE installed a series of Cosmic Ray Tagger (CRT) planes outside the top, bottom, anode, and cathode faces of the detector. Figure 4.12 shows the placement of the CRT planes relative to the detector, along with a simulation of cosmic muons crossing the CRT planes. The design and construction of the CRT system is described in detail in Reference [65].

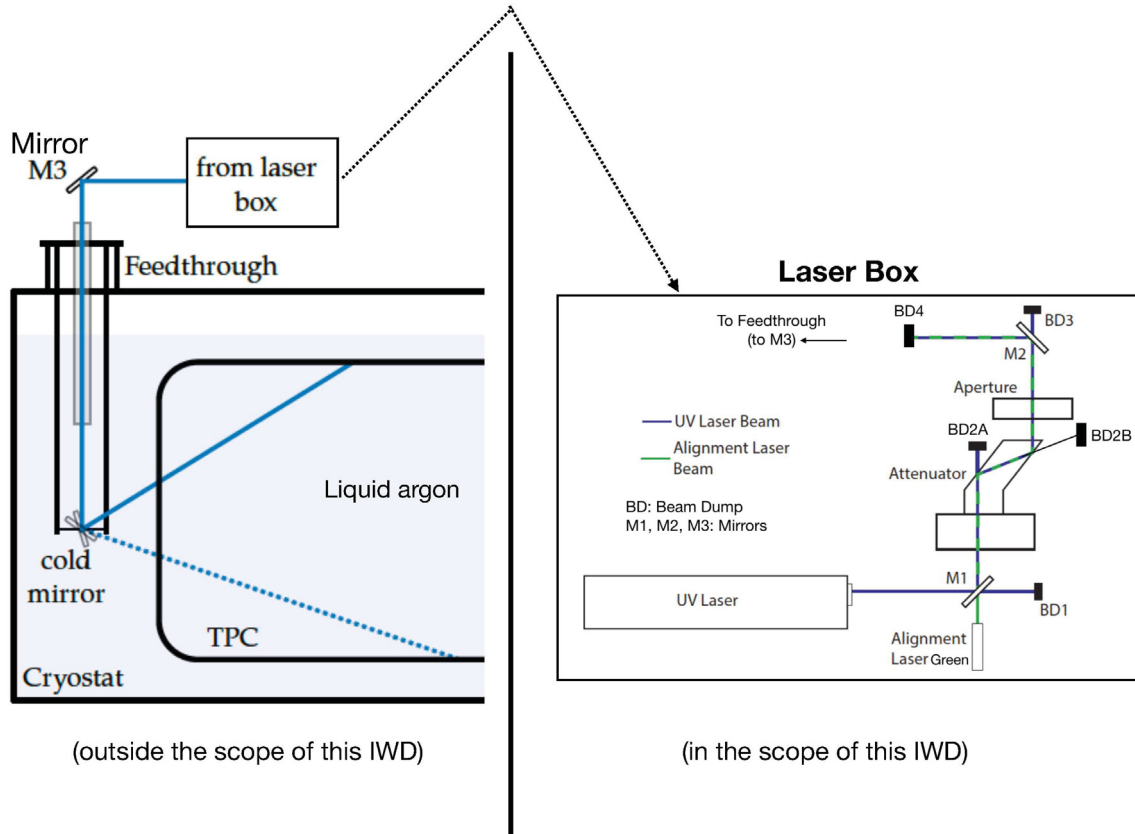


Figure 4.11: Diagram of the UV laser system in MicroBooNE. The 266 nm laser is aligned by two mirrors (M1 and M2), and passed through an attenuator to control the beam energy, and an aperture to control beam size. The laser then enters the TPC after being directed by a third mirror (M3). Finally, the rotatable cold mirror reflects the beam within the TPC volume.

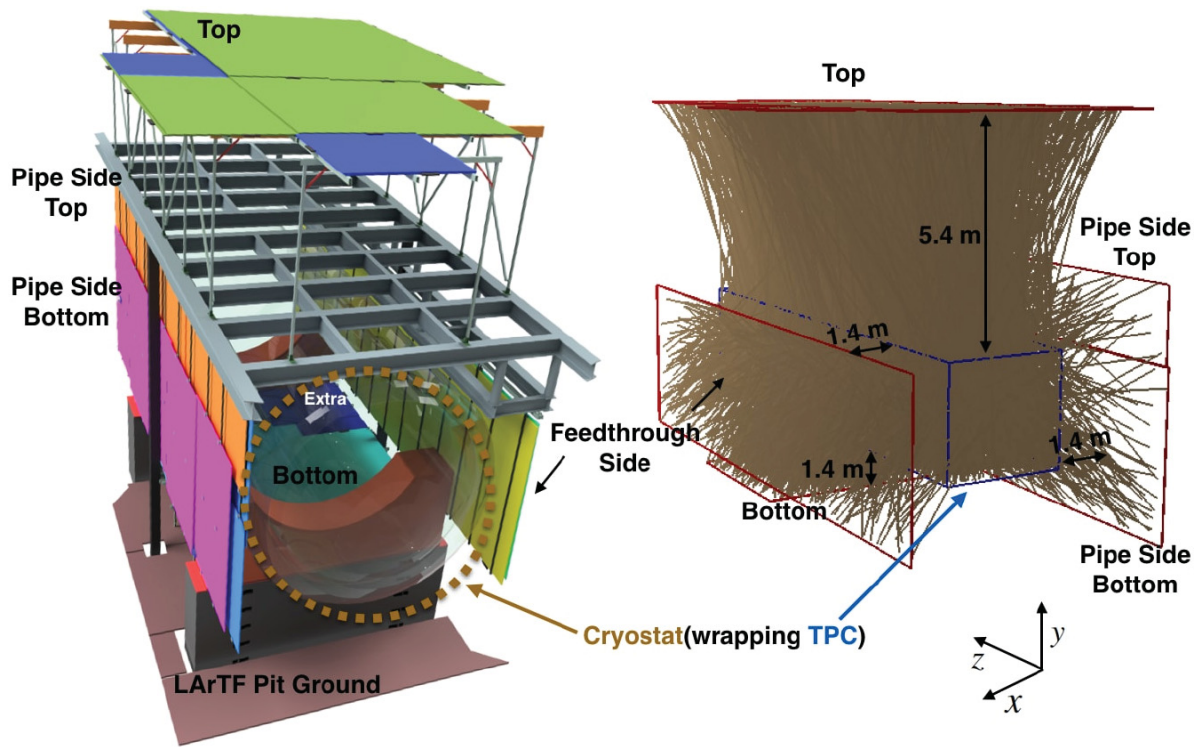


Figure 4.12: Overview of the CRT layout in MicroBooNE. The left shows a cartoon of where the CRT planes are placed, while the right shows a simulation of cosmic muons (brown lines) passing through the CRT planes. Note that “feedthrough” and “pipe” side refer to the anode and cathode side, respectively.

Each CRT plane is comprised of several CRT modules (denoted as individual rectangles in the left side of Figure 4.12) arranged into top and bottom layers. These modules are themselves composed of 16  $10.8 \times 2$  cm plastic scintillating strips. When crossing cosmic muon tracks interact with the scintillator, they produce scintillation light, which is collected by silicon photomultipliers (SiPMs) in the scintillation strips. This signal is then digitized and read out by specialized front-end boards (FEBs). Using this signal, along with the start and end points of the reconstructed cosmic track, we can determine the precise time that a cosmic muon entered the detector, denoted by  $t_0$ . Tracks with a known  $t_0$  are said to be  *$t_0$ -tagged*.

Due to space constraints in the Liquid Argon Test Facility (LArTF)—which houses the MicroBooNE detector—there are no CRT planes at the upstream or downstream ends of the detector. Additionally, the top plane is 5.4 m from the top face of the detector in order to accommodate detector electronic racks. Thus, the CRTs attain a maximum solid angle coverage of 85%.

Due to budget constraints, the CRT system was not installed until March 2017, meaning that the first two data-taking runs lack CRT information.

## 4.8 Data-Taking Triggers

If every event in MicroBooNE were recorded, that collected data would amount to about 13 TB per day. Obviously, this is not sustainable in the long term. To reduce this sizeable data load, MicroBooNE employs two data-taking triggers to determine whether an event should be recorded or not.

The first of these triggers is a *hardware* trigger. Because the BNB is delivered in discrete “spills,” MicroBooNE can leverage timing information from the accelerator division to know when to expect a spill to reach the detector; this is the purpose of the hardware trigger. When the hardware trigger fires, the TPC readout stream opens up a 4.8 ms readout window, while the PMT stream opens up a  $23.4 \mu\text{s}$  readout. The hardware trigger efficiency is 99.8%.

Even after applying the hardware trigger, however, only about 1 in 600 spills will produce a neutrino interaction. To further slim down the collected data, a *software* trigger in the

PMT stream checks for optical activity exceeding a certain threshold. Here, the threshold is 6.5 effective photoelectrons (PE),<sup>2</sup> where the number of PE is determined by comparing the maximum optical waveform value with a baseline value. This trigger is “online,” meaning that the decision to keep the data is made after recording the TPC and PMT streams. The software trigger rejects about 97% of spills, increasing the signal to background ratio from 1 in 600 to roughly 1 in 6.

In addition to the hardware and software triggers for on-beam events, MicroBooNE also uses an *external* trigger to collect background (cosmic) data. As a surface-level detector, MicroBooNE has no shortage of cosmic events. Thus, the software trigger is applied to the external sample as well. Data collected from the external trigger is often referred to as “EXT” or off-beam. When generating simulation samples for analyzers, MicroBooNE takes GENIE-generated particle interactions and overlays EXT data on top to accurately reflect the impact of the large cosmic background. These samples are commonly referred to as MC-Overlays. Details of the MicroBooNE simulation and reconstruction will be discussed in the following chapter.

## 4.9 Detector Operations

MicroBooNE has been collecting data since August 2015, with over  $1.3 \times 10^{21}$  protons-on-target (POT) collected to date. However, the software trigger (described in Section 4.8) was not implemented until February 2016, meaning that data collected before this date is generally not used. The first run period (“Run 1”) is therefore considered to start with the implementation of the software trigger. The work in this thesis will only consider data collected during the first three runs of data taking; this corresponds to the time period from February 2016 to March 2018, during which a total of  $\sim 7 \times 10^{20}$  POT was collected. Figure 4.13 shows the cumulative collected POT during these dates.

For the purpose of minimizing potential bias in evaluating the cause of the MiniBooNE LEE, most MicroBooNE data is “blinded,” meaning that it is not available for analyzers to use. Instead, analyzers generally use a small “unblinded” sample of Run 1 data corresponding

---

<sup>2</sup>Here, an “effective” PE refers to a recorded optical waveform consistent with a single PE.



to approximately  $5 \times 10^{19}$  POT (roughly 5% of the expected final dataset). This is often referred to as simply the “5e19” sample. As noted in Section 4.7, the CRT was not operational until Run 3, although installation began during Run 2. To allow analyzers to use CRT data, and to check for cross-run differences, a small sample of Run 3 data collected with the CRT is unblinded as well. To date, MicroBooNE has performed five full runs of data-taking, with scheduled shutdowns and maintenance performed in between runs. The work in this thesis only uses data from Runs 1–3, which corresponds to about two-thirds of the total expected dataset from all five runs. Table 4.2 summarizes the five data-taking runs.

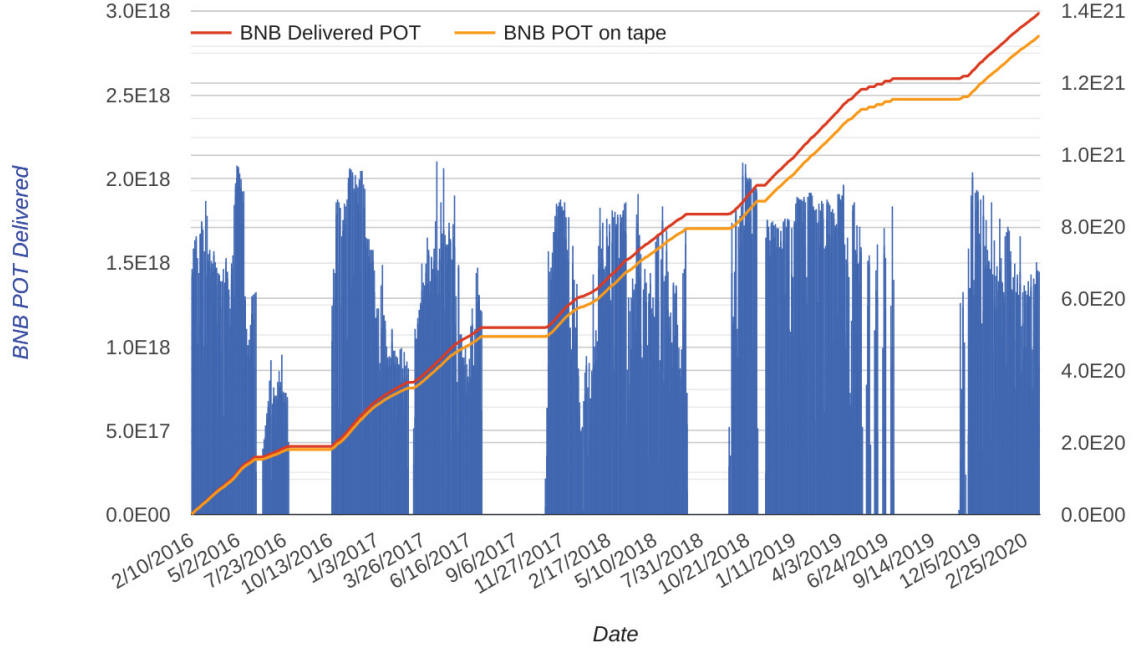


Figure 4.13: Cumulative POT collected by MicroBooNE during Runs 1–5, excluding the period before the software trigger was applied. The total POT delivered across all five runs with the software trigger is  $1.39 \times 10^{21}$ , while the POT written to tape is  $1.33 \times 10^{21}$ .

Table 4.2: Summary of the five data-taking run periods in MicroBooNE. Data from Runs 4 and 5 are not used in this thesis, but are noted here for completeness. The low-purity data in Run 4 were caused by a delivery of impure argon to the detector. Note that the date format used here is mm/dd/yy.

Run Period	Start Date	End Date	POT Delivered	Notes
Run 1	02/10/16	07/29/16	$1.90 \times 10^{20}$	Software trigger applied
Run 2	10/15/16	07/07/17	$3.30 \times 10^{20}$	CRT planes installed
Run 3	10/27/17	07/06/18	$3.15 \times 10^{20}$	CRT operational
Run 4	09/17/18	07/06/19	$3.76 \times 10^{20}$	Includes low-purity data
Run 5	11/05/19	03/21/20	$1.79 \times 10^{20}$	

# Chapter 5

## Simulation and Reconstruction in MicroBooNE

This chapter describes the simulation and reconstruction algorithms used in MicroBooNE, both for generating Monte Carlo (MC) samples and for reconstructing data. These algorithms are implemented into the Liquid Argon Software (LArSoft) [66] framework, a common framework used by multiple LArTPC experiments, including future experiments such as DUNE. Section 5.1 describes the simulated generation of particles from neutrino-nucleus interactions and their propagation through liquid argon. Section 5.2 discusses the detector response modeling, while Section 5.3 explains how we reconstruct neutrino interactions based on the measured TPC signals. Finally, calibration methods applied to MC and data are described in Section 5.4.

### 5.1 Particle Generation and Propagation

As discussed in Section 4.1, neutrinos in MicroBooNE are produced from the decay of secondary mesons resulting from proton-beryllium collisions. To simulate this, we use the GEANT4 [67] framework for both the primary proton-target interaction and the decay of secondary particles as they travel through the air-filled decay pipe. This results in neutrino momentum and spatial position distributions at the upstream end of the TPC, which are subsequently used to configure the GENIE event generator [68] to simulate neutrino-nucleus

interactions in the active TPC volume. MicroBooNE uses GENIE v3, which includes the nuclear model set listed in Table 5.1.

The GENIE v3 nuclear model treats nucleons as a non-interacting Fermi gas, and is therefore referred to as the local Fermi gas (LFG) model [69]. Quasielastic (QE) interactions and the meson exchange current (MEC) channel—which accounts for correlated nucleon pairs—are modeled using the Nieves model [12]. This model accounts for random phase approximation (RPA) screening and multinucleon effects in charged current channels. As of this writing, no equivalent for neutral current channels exists, so GENIE uses an empirical model instead. Both resonance and coherent pion production are modeled according to Berger and Sehgal [16], an updated version of the classic Rein-Sehgal resonance model [15]. Relative to the Rein-Sehgal model, Berger-Sehgal includes lepton mass effects and updated pion scattering data. Deep inelastic scattering is modeled using the Bodek-Yang model [70]. Final-state hadron production (or hadronization) uses the Andreopoulos, Gallagher, Kehayias, and Yang (AGKY) model, which transitions from the Koba-Nielsen-Olesen model [71] at low hadron invariant mass to the PYTHIA model at higher invariant mass [72]. Finally, final-state interactions (FSI) are modeled according to an effective cascade model (denoted hA) [73], which treats FSI as a single interaction parametrized by a mean free path and interaction probabilities, rather than multiple intranuclear interactions.

The output of GENIE neutrino-nucleus scattering simulation is a set of final-state particles that exit the nucleus. These particles are then propagated through the liquid argon using a LArSoft implementation of the GEANT4 framework [67], known as LArG4. GEANT4 simulates the traversal of particles through a given medium based on the mean free path and interaction probability of the particle in that medium. This proceeds in discrete steps within the defined detector volume, resulting in simulated energy depositions at each step. These depositions are then drifted toward the anode plane, where the number of remaining electrons is calculated based on the simulated recombination model and the measured electron lifetime. The position of the electron cloud is calculated from a Gaussian probability distribution, and the cloud is then assigned to the nearest readout wire. In addition to the information provided from GEANT4, MicroBooNE applies spatial corrections to depositions to account for the space charge effect (SCE, see Section 4.4). This information

is then passed to the detector simulation, which handles the modeling of the detector response function described in the next section.

## 5.2 Signal Processing

The MicroBooNE detector simulation utilizes the novel Wirecell (WC) framework, described in extensive detail in References [59] and [74]; much of the information in this section is derived from those sources.

Once the ionization electrons reach the anode wire plane, the process of unfolding the “true” signal from the measured signal begins. Recall from Section 4.4 that the detector response function is modeled as a convolution of a field response—which describes the signal on a wire due to a single ionization electron, as modeled by Ramo’s theorem [75]—and an electronics response, which describes the effect of the readout electronics (see Section 4.3). The ultimate goal of MicroBooNE’s signal processing is to recover the arrival time distribution of an ionization electron cloud, the integral of which provides the total collected charge. To do so, we deconvolve the measured signal with the modeled detector response and electronics noise. In the case of a simple one-dimensional deconvolution, the true deconvolved frequency-space signal,  $S(\omega)$ , can be modeled as [59]

$$S(\omega) = \frac{M(\omega)}{R(\omega)}F(\omega), \quad (5.1)$$

where  $M$  is the measured signal,  $R$  is the detector response, and  $F$  is a Gaussian low-pass noise filter used to mitigate the impact of electronics noise. To simplify the computation, the field response used in the deconvolution is calculated for the central wire and 10 neighboring wires on either side, for a total of 21 wire response calculations. Each individual wire response is averaged over six equally-spaced drift paths within 1.5 mm (half of one wire pitch) of the wire. The electronics response is the same for all channels. The software filter,  $F$ , is a “Wiener-inspired” filter based on the classic Wiener filter [76]. The modifications introduced to the standard Wiener filter account for variations in the TPC signal due to event topology and the increased impact of electronics noise due to MicroBooNE’s “long” (4.8 ms) readout window.

The MicroBooNE deconvolution is two-dimensional, applied in both the time and wire dimensions. The time-dimension deconvolution accounts for the drifting of ionization electrons through the anode plane, while the wire dimension corrects for charge induced on neighboring wires in the vicinity of the central wire. To expedite this process, and to save on computational resources, the deconvolution is applied to a Region of Interest (ROI) around the signal peak. This ROI-finding method also mitigates the impact of low-frequency noise, as noisy regions outside of the signal are generally ignored. The result is known as a *deconvolved waveform*, an example of which is shown in Figure 5.1.

Following the extraction of the arrival time distribution of the ionization electron cloud, the WC framework then calculates a TPC drift simulation based on the SCE-corrected input from GEANT4. Recall from the previous section that GEANT4 simulates energy depositions of a charged particle traveling through the liquid argon, and then assigns each deposition to a wire with a given position distribution and number of arriving electrons, after which we apply SCE corrections. The WC drift simulation then accounts for drift-dependent effects that modify the shape of the electron cloud, such as electron diffusion, attenuation, etc.

In addition to drift-dependent effects, the event topology can significantly impact the extracted signal. The track angle in the  $xz$ -plane, denoted  $\theta_{xz}$ , affects the width of the extracted charge distribution. This effect is especially pronounced for  $\theta_{xz} > 60^\circ$ , as shown in Figure 5.2. Tracks with a large  $\theta_{xz}$  result in energy depositions to appear “stretched” in time on neighboring wires, increasing the charge bias and resolution. The induction planes are especially sensitive to this effect due to their bipolar response functions. Similarly,  $\theta_{yz}$ , the angle with respect to the beam direction in the  $yz$ -plane, affects the height of extracted pulses. At high  $\theta_{yz}$ , more ionization electron clouds fall on or near the central wire, increasing the signal amplitude. Charge bias and resolution studies accounting for both of these topological effects have been performed, and the signal processing framework has been extensively validated using MicroBooNE data.

Table 5.1: List of nuclear models used in GENIE v3.

Interaction	Model
Nuclear model	Local Fermi Gas
Quasielastic scattering	Nieves w/ dipole axial FF
CC MEC	Nieves
NC MEC	Empirical
NC elastic	Ahrens
Resonance	Berger Sehgal
Coherent pion production	Berger Sehgal
Deep inelastic scattering	Bodek-Yang
Hadronization	AGKY
Final-state interactions	hA2018

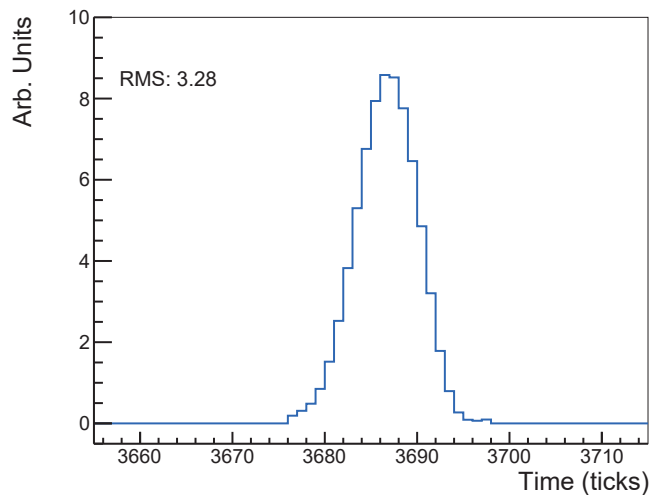


Figure 5.1: Sample deconvolved waveform recovered using MicroBooNE’s 2D deconvolution. Only an ROI around the signal peak is preserved.

## 5.3 Reconstruction of Particle Objects

The previous section outlined the detector response simulation, the goal of which is to obtain the arrival time distribution of ionization electron clouds. From here, we begin the process of reconstructing the particle interaction of interest, or “event.” In addition to ionization signal, the reconstruction also utilizes PMT information to aid in background rejection, particularly those due to cosmic rays. There are three primary reconstruction algorithms used by MicroBooNE analyses: the Pandora pattern recognition algorithm [77], a deep neural network algorithm known as Deep Learning (DL) [78], and the novel Wire-Cell (WC) reconstruction algorithm [79] (not to be confused with the WC drift simulation discussed in the previous section). The DL reconstruction is an image-based algorithm that seeks to identify particles based on individual pixels in the event display. The WC reconstruction is a tomographic algorithm that uses charge information collected on the wire planes to construct 3D images from the 2D charge distribution on each individual plane. While the DL and WC algorithms have both produced significant physics results in MicroBooNE, the work in this thesis relies on the Pandora pattern recognition algorithm. Sections 5.3.1 and 5.3.2 describe Pandora in more detail.

### 5.3.1 Optical Signal Reconstruction

The goal of optical reconstruction is to combine optical activity recorded in the PMTs into a reconstructed object known as a *flash*. The timing of the flash can then be used to identify candidate neutrino interactions through the process of *flash matching*.

The PMTs record optical activity as raw waveforms. First, these waveforms are examined for optical pulses, defined as optical activity that exceeds some threshold value. A baseline-finding algorithm calculates a rolling mean for each time tick outside the pulse region (assuming that this is a signal region), and interpolates the baseline within the pulse region. Waveforms exceeding the threshold are stored as *optical hits*, which contain relevant information such as the waveform height, width, and area. A combination of multiple time-coincident optical hits forms a reconstructed flash. These flashes aid in neutrino interaction identification through the process of *flash matching*. For each PMT, the reconstructed optical



activity is compared to the predicted activity for each candidate interaction in the event. The interaction whose optical activity most closely matches expectation (based on a  $\chi^2$  test) is then “matched” to the reconstructed flash. This flash-matching technique reduces backgrounds by an order of magnitude, significantly improving MicroBooNE’s ability to distinguish neutrino interactions against a large cosmic background [80].

### 5.3.2 Ionization Signal Reconstruction

Once deconvolved waveforms (such as those shown in Figure 5.1) are obtained, the reconstruction algorithm fits a Gaussian to these waveforms. This Gaussian fit is known as a *hit* (not to be confused with the optical hits described in the previous section). Hits contain key information about the underlying waveform, particularly the peak value and the width. These hits are then input to Pandora, which groups hits into *clusters*, and associates those clusters with a particle object. These objects are further grouped into hierarchies of parent and daughter particles, known as Particle Flow Objects (PFOs).

In LArTPCs, the two primary reconstructed objects are particle *tracks* and electromagnetic (EM) *showers*. Tracks tend to manifest as relatively straight and narrow lines traversing the TPC volume, such as the cosmic muons shown in Figure 4.6. At MicroBooNE energies, the most common track-producing particles are muons, charged pions, and protons. Showers, on the other hand, produce a cascade of particles in a roughly conical shape, such as the one shown just below the interaction vertex in Figure 4.6. EM showers in MicroBooNE are generally produced by either electrons or photons. Electrons shower immediately when produced, and thus the showers are usually attached to the interaction vertex. Photons, on the other hand, propagate invisibly over some distance before converting to an  $e^+/e^-$  pair, which then creates the shower. Photon showers therefore show a distinct gap between the interaction vertex and the shower starting point, such as the aforementioned shower shown in Figure 4.6.

The first step in identifying candidate neutrino interactions is the process of cosmic rejection. Pandora first identifies “unambiguous” cosmic rays (CR), usually by identifying clusters which either a) enter the TPC outside of the known beam timing window or b) both enter and exit the detector, meaning that both the start and endpoints pass through the

top and bottom of the TPC. Tracks identified as unambiguous cosmics are then removed, along with any associated daughter delta rays, and their hits are no longer considered. In cases where an unambiguous cosmic overlaps a track not easily identified as cosmic, the hits of the unambiguous cosmic are preserved. For example, if a cosmic track overlaps with a neutrino-induced muon tracks, then the hits of both tracks are kept. At the end of cosmic removal, ambiguous clusters (those that aren't clearly of cosmic origin) of hits are divided into candidate neutrino *slices*.

After removing easily-identifiable cosmics, Pandora then runs its neutrino hypothesis algorithms over the remaining hits in order to identify candidate neutrino interactions. The cosmic hypothesis algorithms are also re-run over each slice in order to compare the cosmic and neutrino hypotheses for each slice. The neutrino algorithms first group hits into 2D clusters on each wire plane. This process utilizes timing information, which is common to all three planes. New clusters are formed any time a cluster shows a significant deviation in its direction (shown in Figure 5.3a), which increases cluster *purity* while sacrificing cluster *completeness*. Here, purity is defined as the number of hits in a cluster that are truly associated with the correct underlying particle object (e.g., the number of proton hits associated to a true proton), while completeness is defined as the number of associated hits relative to the total number for that particle. Cluster merging algorithms subsequently improve the completeness while maintaining purity.

From here, Pandora begins the process of finding a 3D interaction vertex that forms a consistent grouping. To do so, Pandora places an interaction vertex at each possible cluster start and end point, and then evaluates how “neutrino-like” the interaction looks with a vertex in that position. An example of the vertex finding is shown in Figure 5.3b. The next step is to match the clusters between the views that correspond to the same particle trajectory using an iterative process. During this process, clusters that cause ambiguities are identified. For example, an induction-plane cluster may be matched to two collection-plane clusters, resulting in an ambiguity. The 2D clusters are then split or merged using this new information, and the 3D reconstruction begins again. This process is repeated until all ambiguities are removed and a single, consistent 3D neutrino interaction is reconstructed. This process takes place in each neutrino slice, and each slice is then evaluated under both

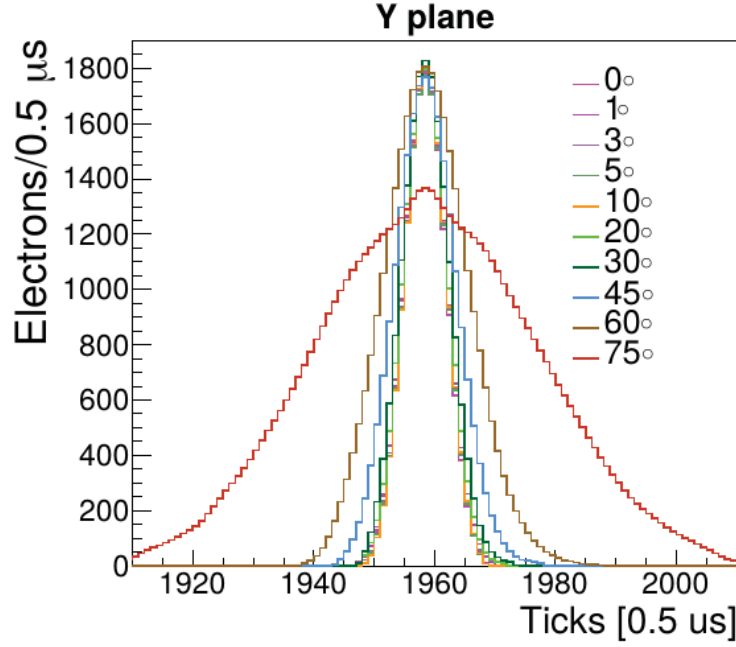


Figure 5.2: Simulated charge distribution for varied  $\theta_{xz}$  values on the collection plane, with  $\theta_{yz}$  fixed at  $0^\circ$ . As  $\theta_{xz}$  increases, the charge distribution is smeared, particularly for tracks above  $\theta_{xz} > 60^\circ$ .

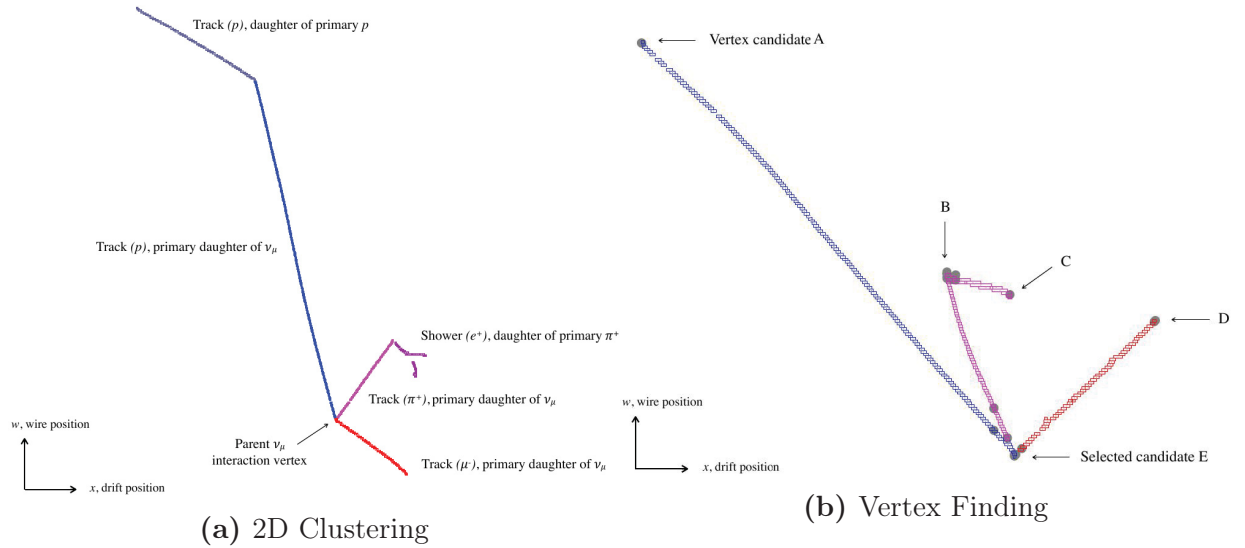


Figure 5.3: Example of the Pandora (a) 2D clustering and (b) vertex finding algorithms on a typical simulated CC  $\nu_\mu$  interaction. Each visible identified particle is shown in a different color.

the neutrino and CR hypotheses and assigned a neutrino slice score, where a higher score indicates a more neutrino-like interaction, and a lower score indicates a CR-like interaction.

Figure 5.4 shows a flowchart summarizing the Pandora pattern recognition process. The result of this process is a candidate neutrino slice (the one with the highest slice score) and a set of reconstructed objects within that slice which are scored as either track-like or shower-like. Downstream reconstruction in LArSoft uses the track and shower scores to identify objects as either tracks or showers.

The performance of Pandora reconstruction varies depending on several factors. The number of final-state particles, particle momenta, event topology, and the interaction mode can all impact Pandora’s reconstruction *efficiency* [81]. Here, efficiency is defined as the number of simulated particles that are matched to one reconstructed particle. We can also define the *correct event fraction* as the percentage of events in which all simulated particles are matched to one reconstructed particle. For example, in simulated CC  $1\mu 1p$  events, the muon reconstruction efficiency is 95.8% and the proton efficiency is 87.3%, leading to a correct event fraction of 83.6%.

Other, more complex event topologies can prove more challenging to reconstruct correctly. Reconstructing two showers coming from  $\pi^0 \rightarrow \gamma\gamma$  decay proves especially difficult for several reasons. In highly asymmetric  $\pi^0$  decays, the subleading (less energetic) photon may not be reconstructed due to being too low-energy. For events with a high-momentum, forward-going  $\pi^0$ , the two photon showers may overlap and be reconstructed as a single shower. Finally, either photon may have a large conversion distance, defined as the distance traveled by the photon before pair-converting. Photons which convert far from the interaction vertex may not be correctly associated with that vertex. For simulated CC resonant  $\pi^0$  events with one muon and one proton in the final state, the leading shower reconstruction efficiency is 88.0%, while the subleading shower efficiency is 66.4%. The muon and proton efficiencies are similar to the CC  $1\mu 1p$  case described earlier, leading to a total correct event fraction of 49.9%.

## 5.4 Calorimetry and Calibration

Although the algorithms described in the previous sections have been extensively studied and validated using MicroBooNE data, imperfections in the reconstructed objects are inevitable. In particular, the collected charge per unit length,  $dQ/dx$ , and the energy deposition per unit length,  $dE/dx$  (which is extracted from  $dQ/dx$ ), show significant differences between data and MC. These quantities are vital to any analysis in a LArTPC.  $dE/dx$  describes the energy loss of a particle as it travels through the liquid argon in the TPC, and is an invaluable tool in particle identification (PID). Recall from Section 3.5 that the MiniBooNE low-energy excess (LEE) is believed to be either electron-like or photon-like, as Cherenkov detectors have difficulty distinguishing electrons from photons. In LArTPCs, the primary way to distinguish electron showers from photon showers is by their measured  $dE/dx$  profile. Since photon showers are created by the conversion of a photon into an  $e^+/e^-$  pair, the measured  $dE/dx$  is expected to be twice that of a lone electron shower. Proper calibration of the extracted  $dE/dx$  is therefore vital to MicroBooNE’s primary physics goal of explaining the LEE.

There are several potential causes of the known differences in  $dQ/dx$  and  $dE/dx$  in data vs. MC. As discussed in Section 4.4, a multitude of physical processes modify ionization electron clouds as they drift in the electric field. Electron-ion recombination, diffusion, local E-field distortions caused by SCE, etc. can all significantly impact the measured arrival time distribution of the electrons, from which we extract  $dQ/dx$ . We also discussed in Section 4.3 that many of the TPC readout wires are either misconfigured, shorted, or dead, all of which affect the measured charge on the collection plane. Finally, detector conditions are known to vary over time,<sup>1</sup> particularly the argon purity, which in turn impacts the electron drift lifetime. The combined result of these effects is a non-uniformity in the collected charge as a function of TPC position. Figure 5.5 shows a collection-plane view of these non-uniformities in terms of the  $yz$ -position of the deposited charge. Recall from Section 4.3 that a significant number of wires in the U plane (first induction plane) are known to be shorted in the region  $z < 400$  cm, leading some charge to be collected on the U plane instead of the collection

---

<sup>1</sup>“Time” here meaning months or years, not to be confused with the  $O(\text{ms})$  readout window.

plane; this is shown in the left boxed region of Figure 5.5. The boxed region on the right (around  $z \sim 700$  cm) shows the collection-plane dead wire region, where practically no charge is collected. The purpose of  $dQ/dx$  calibration is to correct these non-uniformities, as well as the smaller non-uniformities observed throughout the TPC volume. Once we obtain a calibrated  $dQ/dx$  calibration, we can then obtain  $dE/dx$ . This can be summarized as

$$\frac{dQ}{dx} \xrightarrow[\text{correction}]{\text{Uniformity}} \left( \frac{dQ}{dx} \right)_{\text{cal}} \xrightarrow[\text{correction}]{\text{Recombination}} \left( \frac{dE}{dx} \right)_{\text{cal}}. \quad (5.2)$$

The  $dQ/dx$  calibration consists of two parts: a  $yz$  spatial correction and an electron lifetime correction. The  $yz$  correction addresses the impact of transverse diffusion and shorted/dead wires, while the lifetime correction accounts for longitudinal diffusion, electron attenuation, and other drift-dependent effects. We calculate many lifetime correction factors corresponding to different days of data-taking, so these corrections also account for temporal variations in detector conditions, such as fluctuations in argon purity. To derive each of these correction factors, we use a sample of *crossing* cosmic muon tracks, meaning that the track crosses both the anode and the cathode. These tracks traverse the entire drift distance and are uniformly ionizing, allowing for study of drift-dependent effects. Furthermore, their known start and end positions allow for extraction of the precise arrival time of the track,  $t_0$ , which allows for a more precise calibration. Tracks that are either perpendicular to or parallel with the collection plane wires are discarded, as these lead to poor signal reconstruction.

To obtain E-field corrections, we use the measured E-field map derived using the UV laser system [64]. This map provides local E-field values in 10 cm voxels in the detector volume, and we obtain correction values by comparing the local value to the nominal value of 273 V/cm. The derivation of  $yz$  correction factors is similar: we first split the  $yz$  plane into  $5 \times 5$  cm bins, and calculate the median  $dQ/dx$  in each bin. This median value is then compared to a global value derived using all reconstructed hits from the crossing muon sample. The ratio of the local  $dQ/dx$  to the global  $dQ/dx$  forms the  $yz$  correction factor. For the lifetime correction, we split the drift coordinate into 10 cm bins, and extract the median  $dQ/dx$  for each bin. We then plot  $dQ/dx$  vs.  $x$  and extract the electron lifetime,  $\tau$ , from the slope of this fit. Electron attenuation is modeled as an exponential process, so the lifetime is

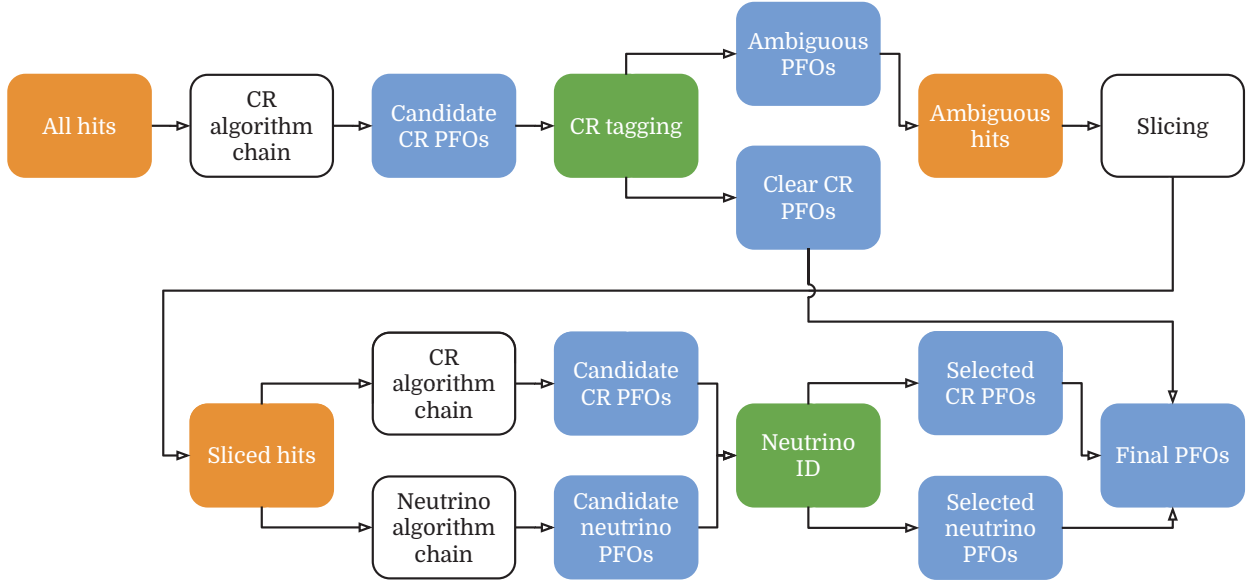


Figure 5.4: Flowchart of the Pandora pattern recognition reconstruction process. The top row shows the process of identifying unambiguous cosmic rays (CR), while the bottom row includes 2D clustering, 3D vertex reconstruction, and neutrino slice score assignment.

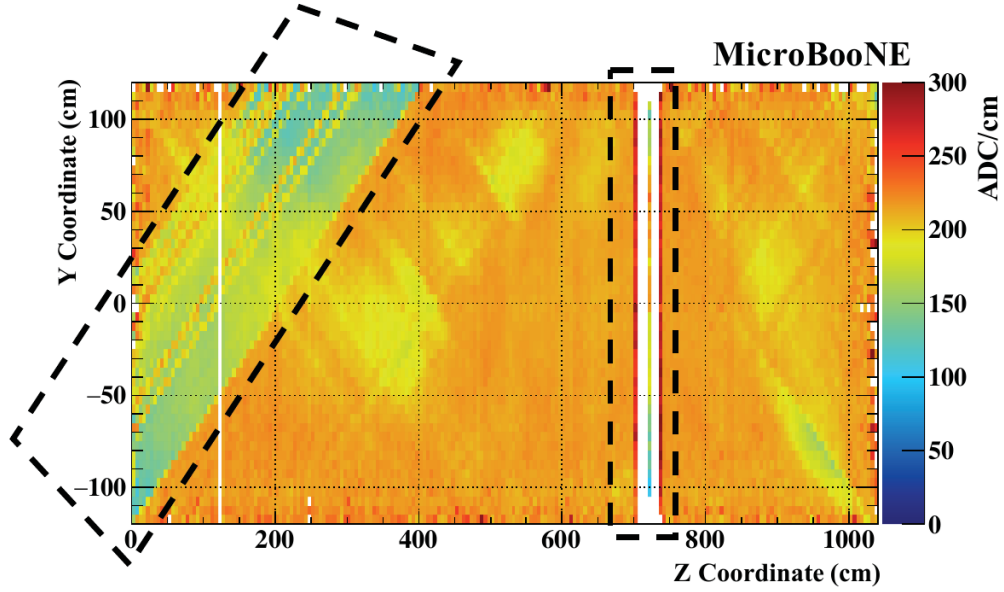


Figure 5.5: Extracted  $dQ/dx$  in the collection plane as a function of  $yz$ -position. The boxed region on the left shows the U-shaped region, while the box on the right shows the collection plane dead wire region.

extracted from an exponential fit to the  $dQ/dx$  distribution. In practice, however, the slope is very small due to the high argon purity in MicroBooNE, as shown in Figure 5.6. This procedure is performed for each day of data-taking, resulting in many lifetime measurements corresponding to different dates, and we store these values in a lifetime database. These values are then used to correct drift-dependent effects in the measured  $dQ/dx$ . Note that these corrections are applied sequentially, meaning that the  $yz$  correction is input to the lifetime correction, and the E-field correction is input to the  $yz$  correction.

The correction factors described above result in a corrected  $dQ/dx$  distribution. To extract the calibrated  $dE/dx$ , we use the modified box model [82] for recombination,

$$\left(\frac{dE}{dx}\right)_{\text{cal}} = \frac{\exp\left[\frac{\left(\frac{dQ}{dx}\right)_{\text{cal}}}{C_{\text{cal}}} \frac{\beta' W_{\text{ion}}}{\rho\epsilon}\right] - \alpha}{\frac{\beta'}{\rho\epsilon}}, \quad (5.3)$$

where  $(dQ/dx)_{\text{cal}}$  is the calibrated  $dQ/dx$ ,  $W_{\text{ion}}$  is the work function in argon,  $\epsilon$  is the electric field,<sup>2</sup>  $\rho$  is the argon density at the nominal operating pressure of 124 kPa, and  $\alpha$  and  $\beta'$  are measurable recombination parameters. The work function, electric field, and argon density are known quantities, while  $\alpha$  and  $\beta'$  have been measured in argon by the ArgoNeuT experiment [82]. Finally,  $C_{\text{cal}}$  is a calibration constant used to convert the ADC value obtained during signal readout to the number of electrons. Table 5.2 lists the values of each variable. The details of measuring the calibration constant can be found in Ref. [83]. Note that this quantity is different for data and MC.

With these known values, we can readily obtain the calibrated  $dE/dx$  distribution, an invaluable tool in PID. To illustrate this, Figure 5.7 shows the leading shower  $dE/dx$  from the NuMI CC inclusive analysis [53]. When calculating shower  $dE/dx$ , we take the first 4 cm of the shower and calculate the average  $dE/dx$  as though it were a track. This method accounts for the spurious nature of shower topologies, which can be difficult to fully reconstruct [84]. Recall from Section 3.5 that one of MicroBooNE's primary physics goals is to investigate the MiniBooNE LEE, which requires separation of electrons from photons. Figure 5.7 shows

---

<sup>2</sup>Note that we're only using  $\epsilon$  to denote the electric field to avoid confusion with the  $E$  in  $dE/dx$ . This will not necessarily be the case in the remainder of this thesis.



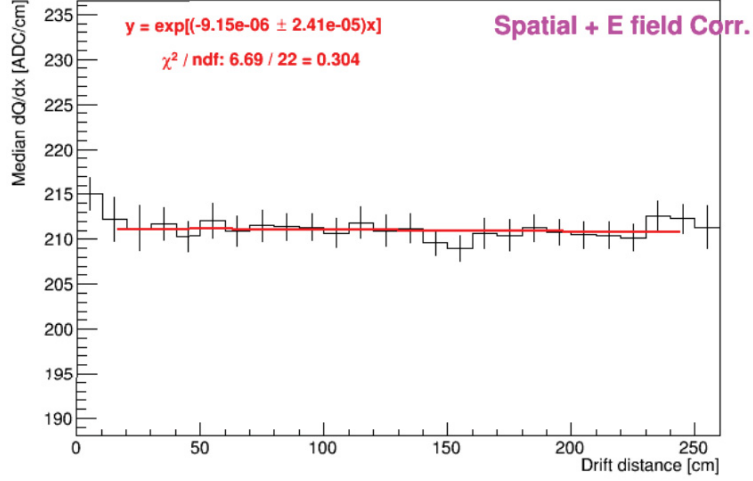


Figure 5.6:  $dQ/dx$  vs.  $x$  distribution for a high-purity sample, dated June 23, 2018. Spatial and E-field corrections have been applied. The flat slope indicates a high electron lifetime.

Table 5.2: Input values to the modified box model [82] when calculating  $dE/dx$  from the corrected  $dQ/dx$ .  $e$  denotes the number of electrons.

Parameter	Value
MC $C_{\text{cal}}$ (ADC/ $e$ )	$(5.077 \pm 0.001) \times 10^{-3}$
Data $C_{\text{cal}}$ (ADC/ $e$ )	$(4.113 \pm 0.011) \times 10^{-3}$
$W_{\text{ion}}$ (MeV/ $e$ )	$23.6 \times 10^{-6}$
$\epsilon$ (kV/cm)	0.273
$\rho$ (g/cm <sup>3</sup> )	1.38
$\alpha$ (dimensionless)	0.93
$\beta'$ (kV/cm)(g/cm <sup>2</sup> )/MeV	0.212

that  $\nu_e$ -induced electron showers peak at 2 MeV/cm, while photon-like showers from  $\pi^0$  decay peak at 4 MeV/cm due to their conversion into an  $e^+/e^-$  pair.

In addition to shower  $dE/dx$ , Figure 5.8 shows the distribution of track  $dE/dx$  vs. residual range (defined as the distance from a point along the track to the track endpoint) for a BNB-induced inclusive CC  $\nu_\mu$  sample. The colored curves show the theoretical expectation based on the Bethe-Bloch formula. By examining the  $dE/dx$  vs. residual range distribution, we can reasonably distinguish between protons and muons.<sup>3</sup> In MicroBooNE, multiple PID algorithms exist which use the  $dE/dx$  information—as well as other calorimetric, geometric, and kinematic variables—to formulate a test statistic which quantifies the likelihood of a particle being a particular species. For example, the PID-a algorithm [85] parametrizes the  $dE/dx$  distribution as an exponential of the form  $aR^b$ , where  $R$  is the residual range and  $a$  and  $b$  are fit parameters. The ArgoNeuT collaboration found  $b = -0.42$  to be a reasonable approximation [85], and so the extracted fit  $a$  can be used to determine the particle species. In MicroBooNE, the PID-a algorithm was found to have poor agreement between data and MC [63], leading to the development of other algorithms. The most recent log-likelihood-based PID algorithm improves on older version by accounting for the various detector effects listed in Section 4.4 and the event topology [86], which can significantly impact charge reconstruction and therefore  $dE/dx$ . This novel algorithm is over 80% efficient at identifying proton tracks using charge information from all three anode wire planes.

---

<sup>3</sup>Differentiating muons and charged pions is much trickier. Although not shown in Figure 5.8, the pion curve is quite similar to the muon curve making the separation of the two difficult.

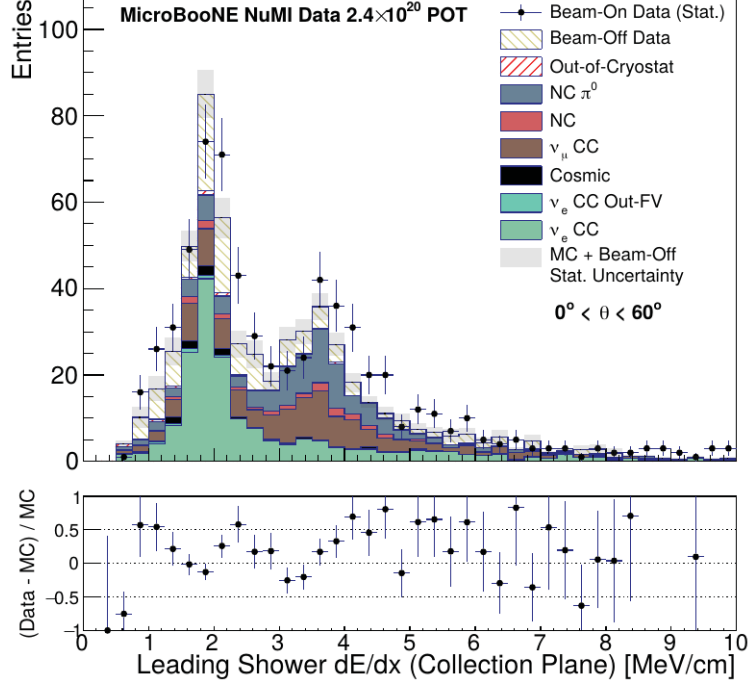


Figure 5.7: Distribution of the most energetic (leading) shower  $dE/dx$  from the NuMI CC inclusive analysis at MicroBooNE [53].  $\nu_e$ -induced electron shower peak at 2 MeV/cm, while photon showers from  $\pi^0$  decays peak around 4 MeV/cm.

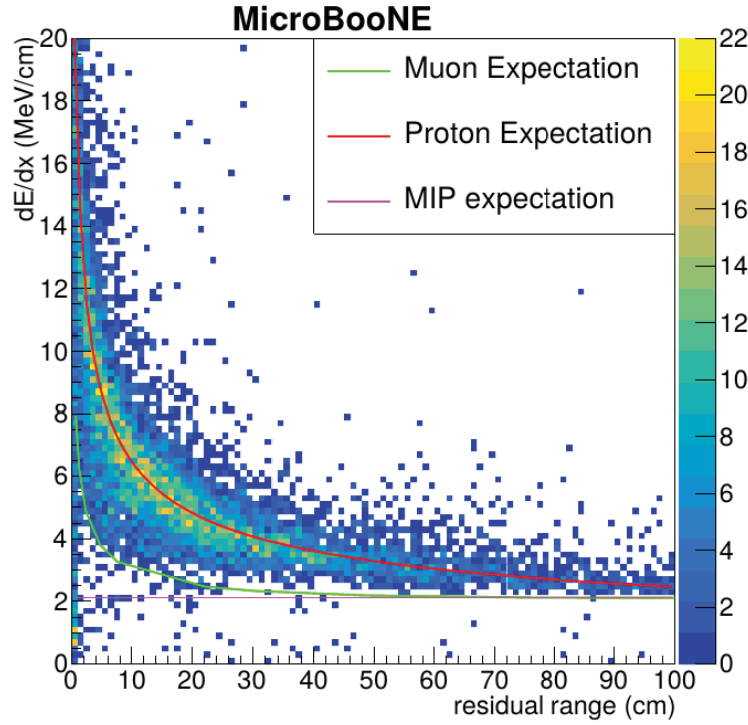


Figure 5.8:  $dE/dx$  vs. residual range distributions for an inclusive CC  $\nu_\mu$  sample. The colored lines represent the theoretical expectation from the Bethe-Bloch formula [22].

## Chapter 6

# Measurement of Longitudinal Electron Diffusion in MicroBooNE

*Statement of disclosure:* This chapter contains an edited version of a soon-to-be published journal article [87]. While this article lists the entire MicroBooNE collaboration as the author (this is standard MicroBooNE procedure), I contributed the majority of the material in the article along with my colleague Adam Lister. The version in this thesis has been edited to remove redundant introductory material, particularly information on the MicroBooNE detector which was covered in Chapter 4. I’ve also made slight changes to wording throughout and modified the format to be consistent with the remainder of this thesis.

In Section 4.4, we discussed some of the physical processes that alter extracted LArTPC signals as a function of drift time, including the space charge effect (SCE), electron-ion recombination, and ionization electron diffusion. This chapter focuses on the measurement of the longitudinal component of electron diffusion,  $D_L$ , in MicroBooNE. Section 6.1 provides the motivation of the analysis and current world data on  $D_L$  measurements in liquid argon. Section 6.2 describes the analysis method. Section 6.3 provides the central value measurement result, while Section 6.4 discusses systematic uncertainties. Finally, Section 6.5 discusses our result in the context of the current world data and theoretical prediction.

## 6.1 Introduction

Electron diffusion is non-isotropic under the influence of an electric field [88, 89, 90] and is split into components which are transverse and longitudinal to the E-field. The transverse component,  $D_T$ , impacts the spatial resolution of a given LArTPC in the plane parallel to the readout wire plane (the  $yz$ -plane in the MicroBooNE coordinate system, shown in the top half of Figure 6.1). Similarly, the longitudinal component,  $D_L$ , impacts the spatial resolution along the drift coordinate (perpendicular to the wire plane as shown in the bottom half of Figure 6.1) broadening the signal waveforms as a function of drift time as shown visually in Figure 6.2. For particles near the anode, where the drift time is low, the signal waveform is relatively tall and narrow. As the drift time increases, the pulses become shorter and broader.

Few measurements of  $D_L$  currently exist in liquid argon. In 1994, the ICARUS collaboration reported measurements of  $D_L$  in a three-ton test stand at E-fields ranging from 100 to 350 V/cm using a three-ton LArTPC with a maximum drift distance of 42 cm [89]. A more recent but preliminary measurement using the ICARUS T600 detector is reported in Reference [91]. Li et al. from Brookhaven National Lab (BNL) reported measurements between 100 and 2000 V/cm in 2015 using a laser-pulsed gold photocathode with drift distances ranging from 5 to 60 mm [90]. The ICARUS results show good agreement with the prediction of Atrazhev and Timoshkin [88], while the results of Li et al. are systematically higher than both. Figure 6.3 summarizes the current world data for  $D_L$  measurements.

This work describes the measurement of longitudinal ionization electron diffusion in the MicroBooNE detector using cosmic-ray muons tagged by the CRT system (see Section 4.7). The CRT data used in this measurement were collected between October 27, 2017 and March 13, 2018.

## 6.2 Method

To first order, the relationship between the time-width of a signal pulse at a given time  $t$ ,  $\sigma_t(t)$ , and  $D_L$  can be parametrized [89, 90] as

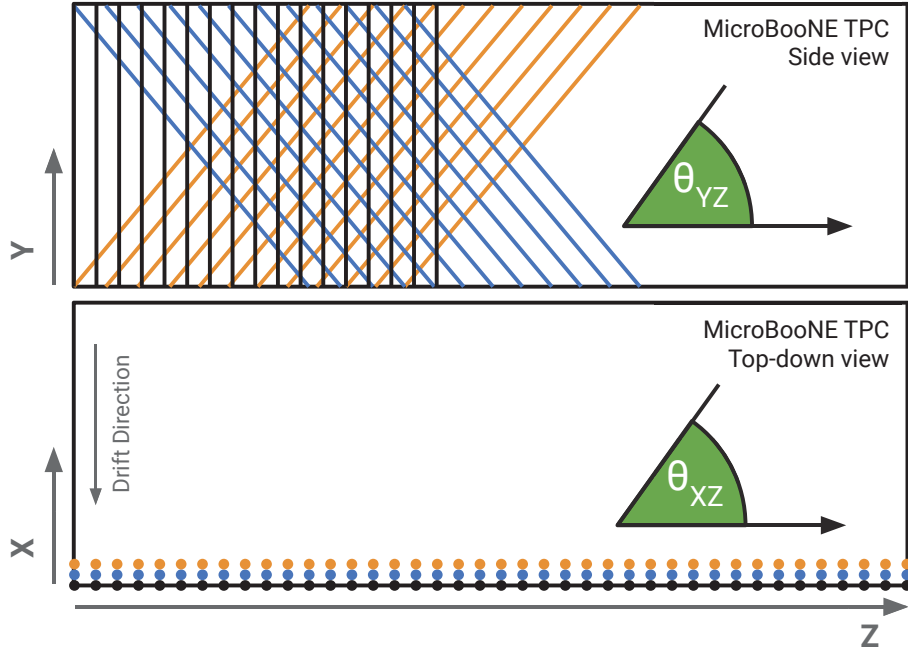


Figure 6.1: Diagram of the MicroBooNE coordinate system and wire planes. The beam travels along the  $z$ -direction, while ionization electrons drift in the decreasing  $x$ -direction.  $y$  denotes the vertical direction. The angles  $\theta_{xz}$  and  $\theta_{yz}$  denote the angle of a reconstructed object (i.e., track or shower) with respect to the beam direction in the  $xz$ - and  $yz$ -planes respectively. The top half shows a side view of the TPC through the anode plane, while the bottom half shows a top-down view. The colored lines (dots) on the top (bottom) half represent the three readout wire planes.

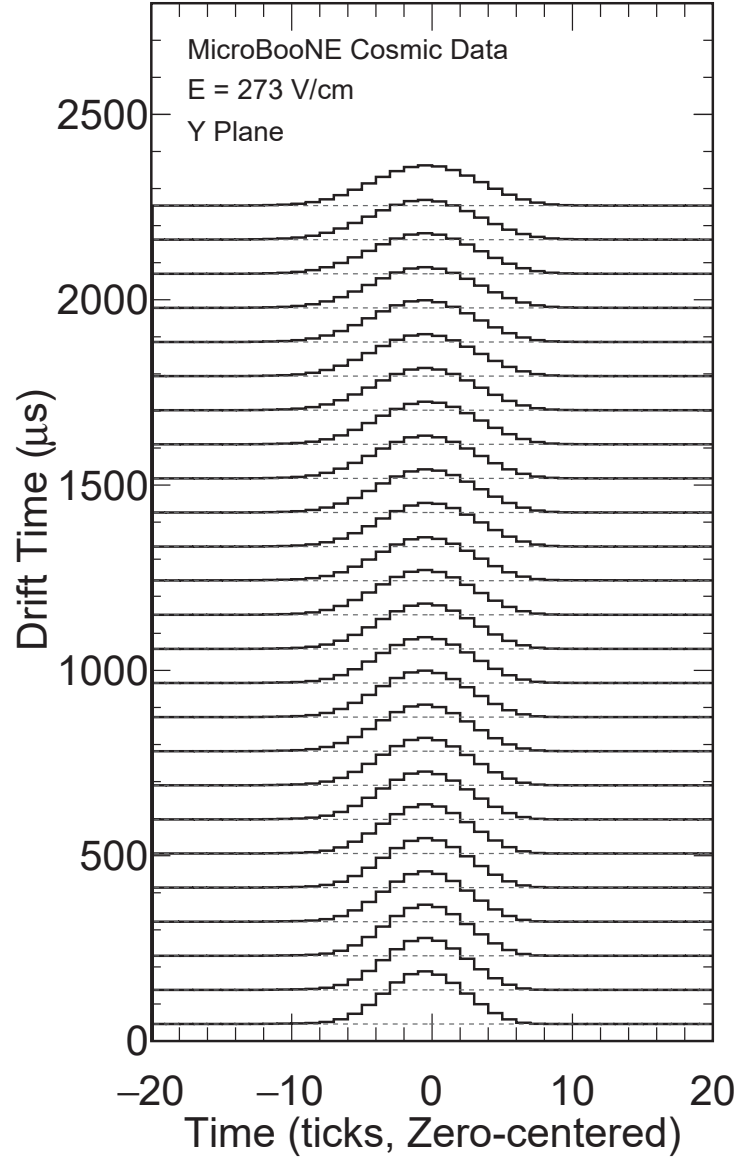


Figure 6.2: Visualization of the impact of  $D_L$  on signal waveforms as a function of drift time. The waveform peak times have been shifted in order to align with one another. One time tick is equivalent to  $0.5 \mu\text{s}$ . Each waveform displays the deconvolved ADC count, arbitrarily scaled.

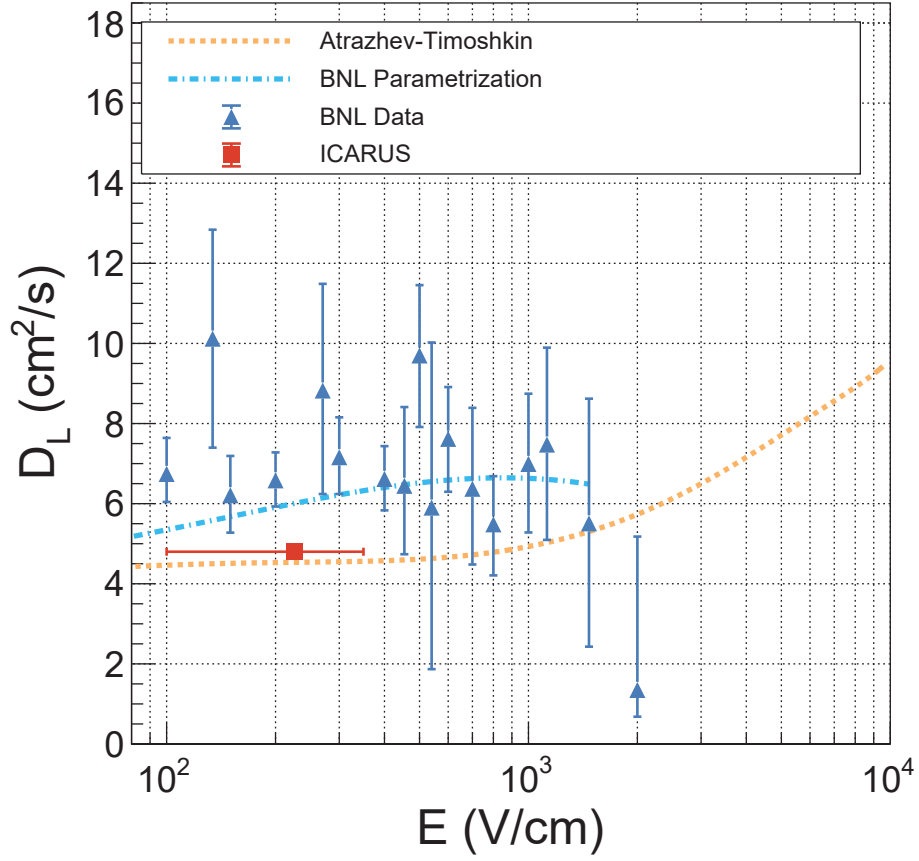


Figure 6.3: Summary of world data for longitudinal electron diffusion in liquid argon. The orange-dashed curve shows the theory prediction [88], the blue dot-dashed curve shows the parametrization from Li et al. [90], and the red and dark blue points show the ICARUS [89] and Li et al. measurements, respectively. Details of this plot can be found in Appendix A. Note that the ICARUS error bars ( $\pm 0.2 \text{ cm}^2/\text{s}$ ) are obscured by the data point.



$$\sigma_t^2(t) \simeq \sigma_t^2(0) + \left( \frac{2D_L}{v_d^2} \right) t, \quad (6.1)$$

where  $v_d$  is the drift velocity and  $\sigma_t^2(0)$  is added to account for the Gaussian noise filter used during waveform deconvolution [59] which enforces a minimum width for the pulses. The expected minimum width is  $\sigma_t^2(0) \sim 1.96 \mu s^2$ . Because equation 6.1 is an approximation,  $D_L$  is actually an *effective* diffusion coefficient that contains a small contribution from transverse diffusion (see section 6.4). Equation 6.1 assumes a constant  $v_d$ . However, due to the abundant cosmic ray flux in MicroBooNE caused by its location near the surface, the electric field varies as a function of position in the detector due to SCE. This means that  $v_d$  also changes throughout the detector volume. MicroBooNE has measured the values of  $v_d$  as a function of TPC position using electric field maps determined using UV laser data [92]. Because equation 6.1 captures the size of the electron cloud at the point of measurement, it is important to use the value of the drift velocity at the location of that measurement. Specifically, the signal processing removes the electronics response and the field shaping and returns a measured time distribution that corresponds to the arrival time of the electrons at  $x = 0$  (the first induction plane) convoluted with a Gaussian low-pass filter function that removes high frequency noise. Thus, the mean drift velocity at  $x = 0$ ,  $v_d = 1.076 \text{ mm}/\mu s$ , is used for the measurement of  $D_L$  in the MicroBooNE data. When measuring  $D_L$  from our simulation samples (see section 6.3.1), we use the nominal simulated  $v_d$  value of  $1.098 \text{ mm}/\mu s$ , since the simulated signal deconvolution assumes the ionization electrons drift at this velocity across the volume.

Although the MicroBooNE E-field varies as a function of position within  $273.9_{-8\%}^{+12\%} \text{ V/cm}$  [60, 92] due to space charge effects, equation 6.1 assumes that the value of  $D_L$  is constant. Figure 6.3 shows that, within MicroBooNE's E-field variations, the current world data and theoretical expectations for  $D_L$  are consistent with an assumption of a constant  $D_L$  value in the region of the MicroBooNE E-field. The MicroBooNE nominal  $D_L$  simulation value is extracted from the parametrization of Li et al. (blue-dashed curve in Figure 6.3) at  $E = 273.9 \text{ V/cm}$  and corresponds to a  $D_L$  value of  $6.40 \text{ cm}^2/\text{s}$ .

Due to the linear relationship between the squared-pulse-width in time and ionization electron drift time (Equation 6.1), it suffices to perform a linear fit of  $\sigma_t^2$  versus  $t$  and

extract  $D_L$  from the slope. The widths of waveforms (“pulse widths”) are sensitive to more effects than just longitudinal diffusion. Transverse diffusion, the detector response modeling, collinear delta ray production, and the angle of the reconstructed track can all significantly impact the measured time width of the pulse. To minimize the additional broadening from such effects, we place a strict set of requirements on tracks reconstructed from the MicroBooNE data.

### 6.2.1 Track Selection

To measure  $D_L$ , we use cosmic muons tagged by MicroBooNE’s CRT. Using the signals read out from the CRT system along with the start and end points of the reconstructed cosmic track, we can determine the precise drift time ( $t_0$ ) that a cosmic muon entered the detector. This allows us to use  $t_0$  as the track start time to determine the drift time of the waveforms used in the final measurement. Tracks with a known  $t_0$  are said to be  *$t_0$ -tagged*. For CRT-tagged tracks with length greater than 50 cm, the  $t_0$ -tagging efficiency is 56.6%. For this analysis, we require that tracks must

- have a reconstructed length greater than 50 cm;
- be through-going, meaning that both the start and end points must be within 5 cm of any TPC wall;
- have  $|\theta_{xz}| < 6^\circ$  and  $|\theta_{yz}| < 40^\circ$  (figure 6.1); and
- have an average track deflection of less than 6 cm.

The track length requirement ensures track reconstruction quality and reduces potential track mis-identification of shorter tracks or shower-like objects. We require through-going tracks as an additional reconstruction quality check. The strict angular selection is designed to mitigate additional pulse width broadening due to the combined effects of track angle,  $D_T$ , and the detector response modeling (see Figure 9 of Reference [59]) particularly in the  $xz$ -plane. As  $\theta_{xz}$  increases, so does the intrinsic spread in  $x$  of the ionization position distribution. A stringent  $\theta_{xz}$  requirement therefore mitigates this effect while providing a

sufficient number of waveforms to perform the analysis.  $\theta_{yz}$ , on the other hand, impacts pulse height rather than pulse width, so we choose a looser requirement for that angle. Finally, as a measure of track straightness, we use the average deflection defined as the average transverse distance between each point along the track and a straight line connecting the track start and end points. Track angles are determined using the track starting direction, but, in some cases, the track can significantly deviate from this starting direction. This requirement therefore ensures that tracks remain relatively forward-going. An event display of a selected track is shown in Figure 6.4.

Track length distributions at each stage of the selection are shown in Figure 6.5, while the selection efficiencies and number of selected tracks are shown in Table 6.1. The requirements on the track angle are the least efficient, reducing the number of selected tracks by two orders of magnitude. The final selection contains  $\sim 70,000$  tracks and each track can have hundreds of waveforms. This provides an ample number of waveforms to perform the analysis.

## 6.2.2 Waveform Selection

The pulse widths in this analysis are extracted from deconvolved waveforms, low-level data products which attempt to recover a “true” signal by deconvolving the raw signal measured at the anode wires with the detector response. Recall from Section 4.4 that the MicroBooNE detector response is modeled as a convolution of a *field* response and an *electronics* response. The field response describes the charge induced on one anode-plane wire by a single ionization electron, while the electronics response describes the impact on the signal waveform due to shaping and amplification during signal readout [74, 58]. The deconvolution process also applies a Gaussian low-pass noise filter to mitigate the effects of electronics noise [58]. The MicroBooNE deconvolution is two-dimensional, applied in both time and wire space. The result of this deconvolution is a signal waveform with a distinct region of interest preserved around signal peaks that exceed a predefined threshold value.

As with the reconstructed tracks, we place a set of requirements on the reconstructed hits to ensure waveform quality. While the final  $D_L$  measurement uses deconvolved waveforms rather than reconstructed hits, hit information is easily accessible and can be used as a proxy

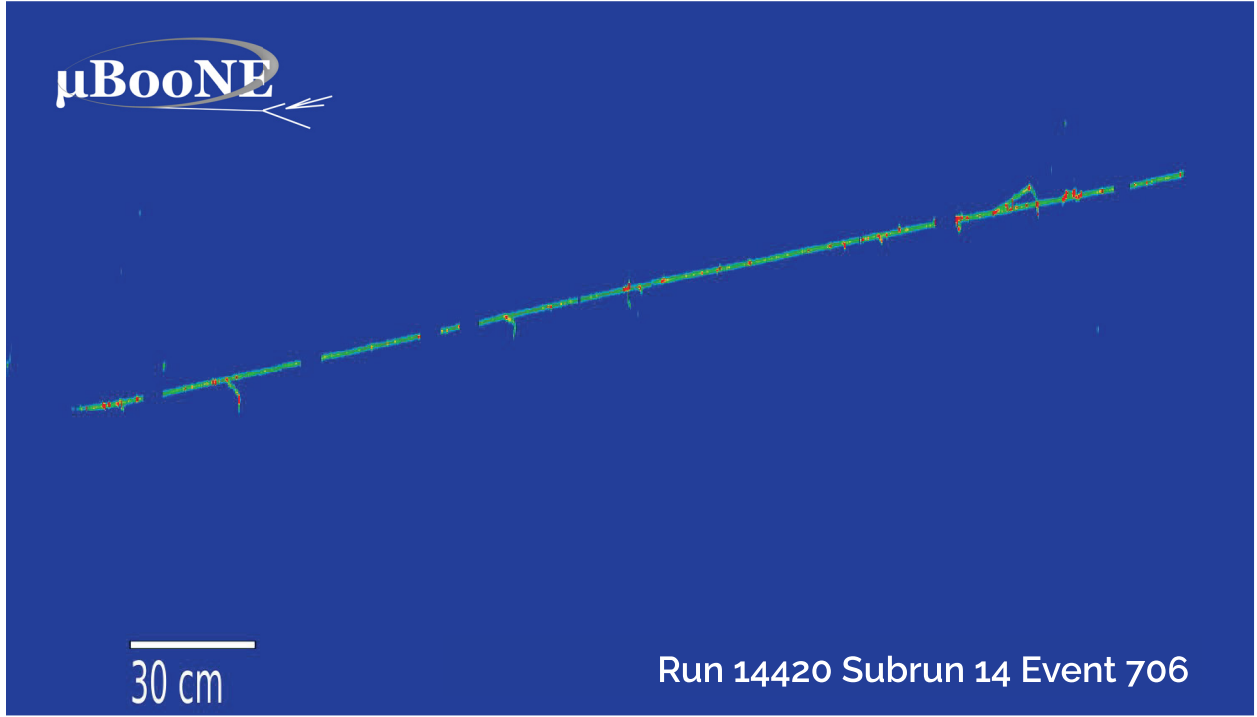


Figure 6.4: Collection-plane view of a track that passes the diffusion track selection requirements outlined in section 6.2.1. The horizontal axis is the wire number (increasing from left to right), while the vertical axis is the drift time (increasing from bottom to top).

Table 6.1: Selection efficiencies after each selection requirement and number of selected tracks. Relative efficiencies are calculated relative to the number of tracks at the previous stage of the selection.

Selection Requirement	No. Tracks	Relative Efficiency	Absolute Efficiency
Total tracks	$5.27 \times 10^7$	100%	100%
Length > 50 cm	$2.27 \times 10^7$	43.1%	43.1%
$t_0$ -tagged	$1.28 \times 10^7$	56.4%	24.3%
Through-going	$1.25 \times 10^7$	97.7%	23.7%
Angular requirement	79,896	0.64%	0.15%
Deflection requirement	71,698	89.7%	0.14%

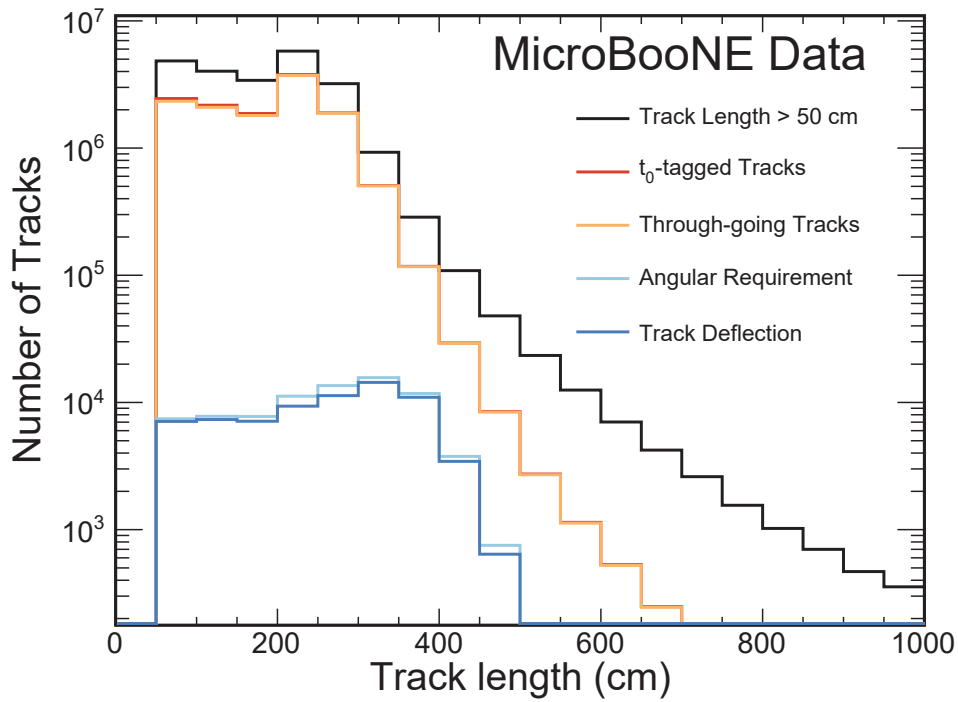


Figure 6.5: Track length distributions at each stage of the track selection. The peak around 230 cm in the orange and black curves corresponds to the height of the TPC since most CRT tracks traverse the detector top-to-bottom.

for the shape of the underlying waveform. We require reconstructed waveforms for which the reconstructed hits have

- been fit to a single Gaussian distribution;
- a goodness-of-fit (GoF)  $< 1.1$ ; and
- a  $z$ -position between  $400 \text{ cm} < z < 675 \text{ cm}$  or  $775 < z < 951 \text{ cm}$ .

Requiring the waveform to have been fit to a single Gaussian distribution removes hits that are contaminated with other charge depositions, particularly those due to delta ray production along the reconstructed track. The hit GoF test ensures that the waveform shape is reasonably Gaussian; we model electron diffusion as a Gaussian process, and the deconvolution uses a Gaussian noise filter. We expect the waveforms to follow this shape as well. Finally, we apply a hit fiducial volume along the  $z$ -direction. The first induction plane in MicroBooNE is known to have a region of shorted wires in the upstream half of the TPC [59, 58]; requiring hit positions to be at least 400 cm from the upstream end of the TPC removes this region from consideration. The downstream portion of the detector volume is impacted by SCE [92], so we remove that region as well. Finally, we ignore the region between 675 and 775 cm in  $z$  to avoid a region of dead wires in the collection plane. Figure 6.6 shows the  $yz$ -position distribution of reconstructed space points corresponding to the selected hits. The waveform fiducial volume removes slightly more than half of the TPC volume with most of the selected waveforms coming from  $z > 800 \text{ cm}$  due to the detector geometry combined with the requirement that reconstructed tracks have a shallow  $\theta_{yz}$  and be through-going.

In addition to the criteria listed above, we place an additional requirement that the hit width of each individual waveform be representative of the pulse width distribution in its corresponding bin of drift time. To do so, we reject all waveforms whose hit widths fall outside of a one standard deviation region around the median value in that drift bin as shown in Figure 6.7. The dark blue regions in Figure 6.7a show that many hit widths differ significantly from the median value in that drift bin largely due to effects such as unresolved delta rays, misreconstruction, and the statistical nature of diffusion. This leads to long “tails”

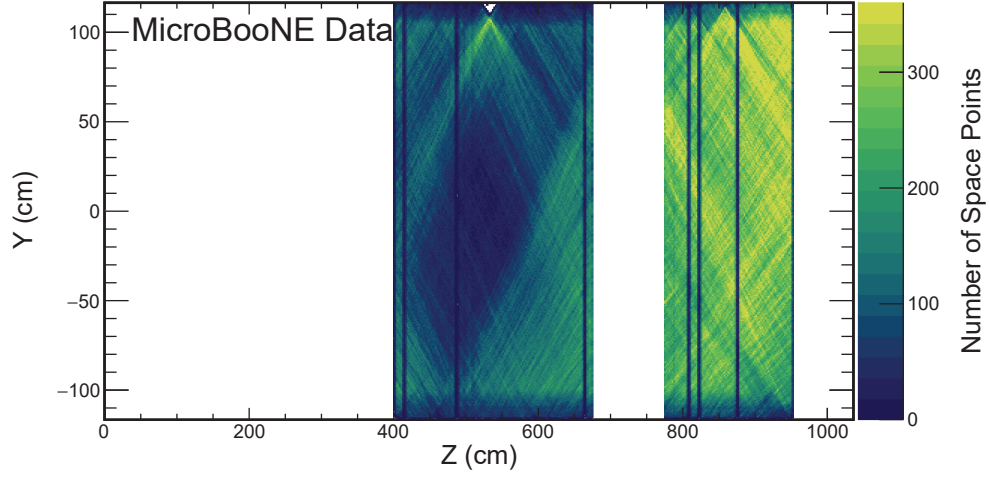


Figure 6.6:  $yz$ -position distribution of reconstructed space points corresponding to selected hits on the collection plane. The majority of the selected hits are in the downstream portion of the detector due to geometric effects along with the track selection. The empty region at the top of the detector around  $z \sim 550$  cm is due to the overlap of two dead regions on the two induction planes. 3D reconstructed objects such as space points require charge to have been measured on at least two wire planes.

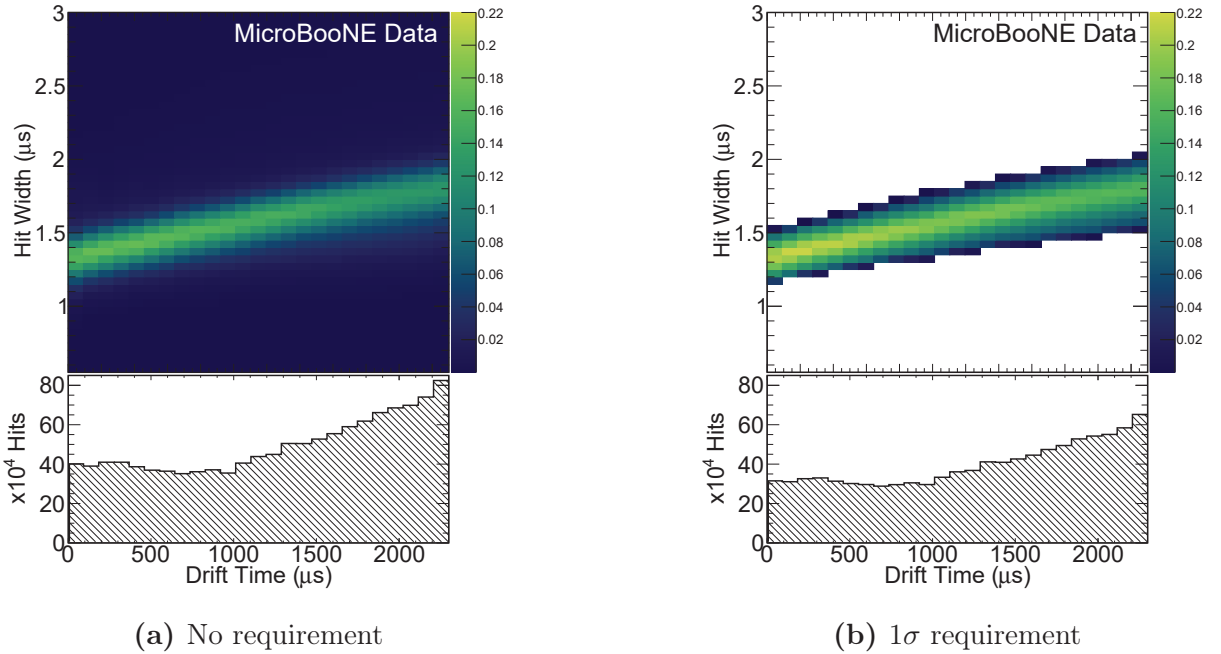


Figure 6.7: (a) Distribution of hit widths vs. drift time. (b) The same distribution after requiring hits be within one standard deviation of the median value in each drift bin as described in the text. Each bin of drift time has been area normalized in the two dimensional histogram so that the structure is more visible. The bottom histograms show the number of hits collected in each bin of drift time.

in the distribution of  $\sigma_t^2$  in each bin which could bias the width of the resultant summed waveforms in each bin.

### 6.2.3 Extraction of $D_L$

The electron drift time in MicroBooNE ranges from 0 to 2300  $\mu\text{s}$ , which we split into 25 bins. At the nominal drift velocity of 1.098 mm/ $\mu\text{s}$ , each bin corresponds to roughly 10 cm of drift distance. Within each of the 25 drift time bins, we employ a waveform summation technique to obtain a single representative waveform of that bin. To account for time offsets between waveforms we iteratively shift each additional waveform from -5 ticks to +5 ticks relative to the center of the summed waveform and choose the configuration which minimizes the hit width (RMS) of the resultant summed waveform. An example of this process is shown in Figure 6.8, and a sample summed waveform is shown in Figure 6.9. The summed waveform retains a Gaussian shape, without a significant additional broadening due to the waveform summation method; see section 6.4.4.

Once we have a summed waveform in each bin, we fit a Gaussian to that summed waveform, taking the standard deviation as our measure of  $\sigma_t$ , and the mean as  $t$ . We then plot  $\sigma_t^2$  vs  $t$ , and extract  $D_L$  from the slope of this fit. Figure 6.9 shows a sample summed waveform with the Gaussian fit drawn on top. It is clear that the underlying distribution is not perfectly Gaussian, but when restricted to the region around the peak of the distribution, the Gaussian functional form is a good estimator of the width of the distribution. The statistical uncertainty on  $\sigma_t$  is negligible due to the large number of waveforms used in each drift bin.

## 6.3 Measurement of Longitudinal Electron Diffusion

### 6.3.1 Method Validation on Simulated Samples

To validate the method described in section 6.2, we use simulated samples containing only a single muon. These simplified samples contain 500 events, each with exactly one muon track and no backgrounds. The muon tracks are generated precisely in-time with the beam, so



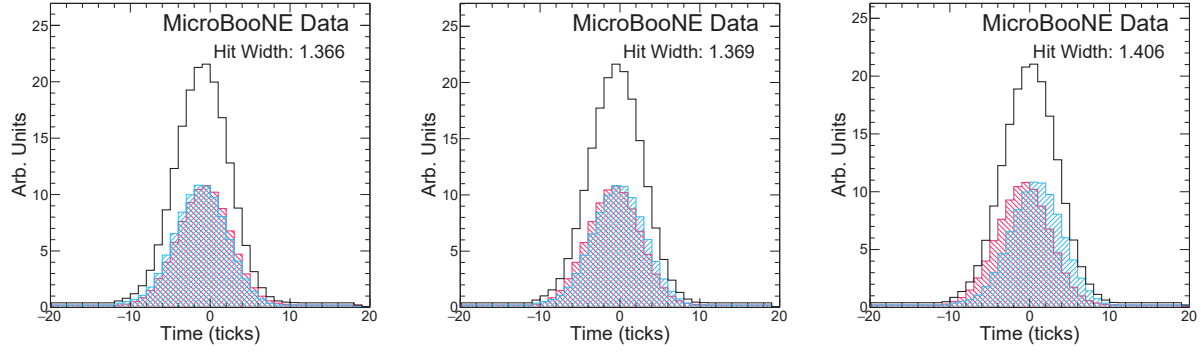


Figure 6.8: Illustration of the waveform summation technique employed in this analysis. The cyan waveform is iteratively shifted from -5 to +5 ticks in increments of one tick. At each iteration, the cyan waveform is added to the magenta waveform and the hit width of the summed waveform (black) is calculated. In this simplified example, the cyan waveform is shown shifted by -1, 0, and +1 ticks. In this case, the left-hand configuration would be selected.

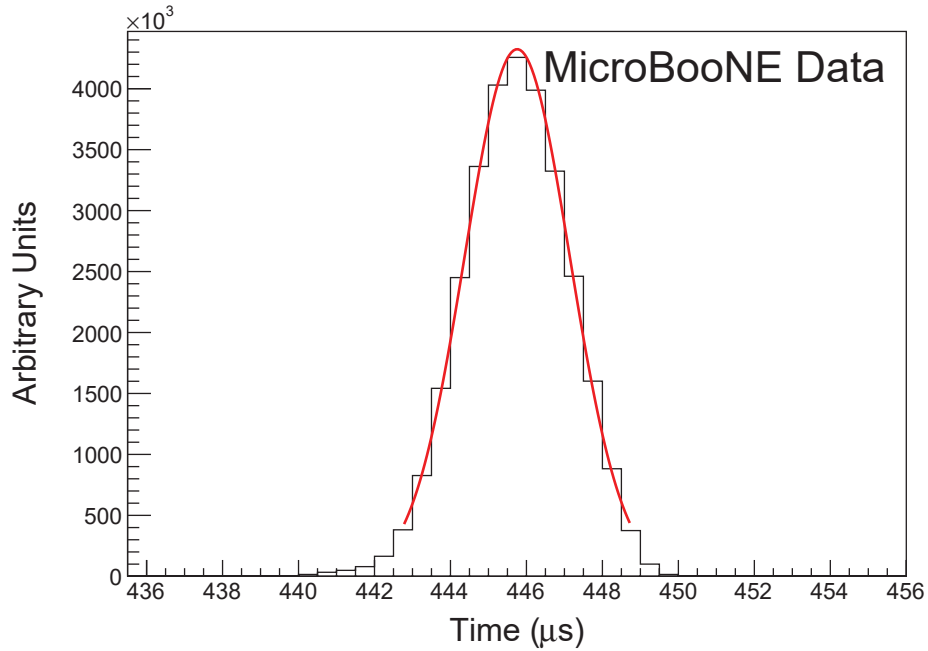


Figure 6.9: Sample summed waveform with Gaussian fit.  $\sigma_t$  is extracted from the standard deviation of the fit and  $t$  from the mean. This waveform is taken from the first drift bin on the collection plane.

there is no potential bias from  $t_0$  mis-tagging. They populate the detector volume uniformly and have a fixed momentum of 1 GeV/c, with an angular coverage of  $\theta_{xz} = \pm 6^\circ$ . Figure 6.10 shows the resultant plots of  $\sigma_t^2$  vs.  $t$  on each wire plane for simulated single muons within the angular selection values listed in section 6.2.1. For each plane, the top plot shows the linear fit and an area-normalized histogram of the number of waveforms in each bin; the bottom plot shows the fit residuals. As discussed in the previous section, each point on the plots in Figure 6.10 represents the standard deviation of a Gaussian fit to the summed waveforms in each bin of drift time. We extract the measured  $D_L$  value from equation 6.1 using the simulated drift velocity  $v_d = 1.098$  mm/ $\mu$ s. This simplified sample results in a measured  $D_L$  value of 6.30 cm<sup>2</sup>/s. Compared to the nominal (default) simulation value of 6.40 cm<sup>2</sup>/s, the measured value is well within the estimated systematic uncertainties, discussed in section 6.4. The values of  $\sigma_t^2(0)$  are extracted from the y-intercept of the linear fit, and their values are close to the expected value of  $\sigma_t^2(0) = 1.96$   $\mu$ s<sup>2</sup>. Fit errors on  $D_L$  and  $\sigma_t^2(0)$  are negligible (<1%).

### 6.3.2 Measurement using CRT Data

Figure 6.11 shows the  $\sigma_t^2$  versus drift time distribution from which we extract  $D_L$  for MicroBooNE data. When using CRT data, the distribution of waveforms peaks near the cathode because of the CRT plane geometry; the CRT plane on the cathode side is nearly twice as large as the anode-side plane. The  $D_L$  central value extracted from the slope is 3.74 cm<sup>2</sup>/s when using collection-plane waveforms. The statistical uncertainties and uncertainties from the fit are negligible. The y-intercept of 1.88  $\mu$ s<sup>2</sup> is slightly below the expected 1.96  $\mu$ s<sup>2</sup>. While Figure 6.11 shows the fit results on all three wire planes, we choose to quote the value extracted on the collection plane as our measurement. There are two primary reasons for this: 1) the induction planes are known to be more impacted by electronics noise than the collection plane, and 2) the bipolar nature of the induction plane response functions may introduce additional bias in the extracted pulse widths during deconvolution [59]. The other wire planes are used for systematic uncertainty studies as described in section 6.4.3. As a cross-check of this measurement, Figure 6.12 shows area-normalized comparisons of summed waveforms between MicroBooNE data and simulated datasets with  $D_L = 6.40$

$\text{cm}^2/\text{s}$  (MicroBooNE nominal) and  $D_L = 3.74 \text{ cm}^2/\text{s}$  (measured data value). It is clear from these comparisons that the  $D_L = 3.74 \text{ cm}^2/\text{s}$  dataset more closely matches the data waveforms, lending weight to our measurement. Table 6.2 displays a summary of the results presented in figures 6.10 and 6.11.

The extracted  $D_L$  value of  $3.74 \text{ cm}^2/\text{s}$  differs significantly from the default simulation value of  $6.40 \text{ cm}^2/\text{s}$ . Recall Figure 6.3 which shows a summary of current world data on diffusion. The blue dot-dashed curve shows the parametrization of Li et al. [90] while the orange-dashed curve shows the theory prediction of Atrazhev and Timoshkin [88]. The default simulation value was extracted from the Li et al. parametrization which is known to be systematically higher than the theory curve.

## 6.4 Evaluation of Systematic Uncertainties

This section describes studies performed to evaluate the total systematic uncertainty on the  $D_L$  measurement. While a multitude of effects could potentially bias the measurement, the largest expected systematic effects are due to transverse diffusion, drift velocity variations, and the detector response function modeling. We also considered other possible sources of systematic uncertainty but found them to be sub-dominant.

### 6.4.1 Transverse Diffusion

Recall that  $D_L$  in equation 6.1 is actually an *effective* longitudinal diffusion coefficient with residual contributions to the pulse width from  $D_T$  [90]. For tracks with non-zero  $\theta_{xz}$ , adjacent electron clouds begin to overlap as they spread in the  $yz$ -plane under the influence of  $D_T$  causing additional  $\sigma_t$  smearing. Figure 6.13 shows an illustration of this effect. The impact of  $D_T$  increases as a function of track  $\theta_{xz}$ . This motivates the strict angular requirement outlined in section 6.2.1.

To evaluate a systematic uncertainty on  $D_T$ , we generate three simulated particle gun samples using the same configuration as the sample described in section 6.3.1 except that we vary the simulated  $D_T$  value in each sample. The ratio  $D_L/D_T$  can be expressed as

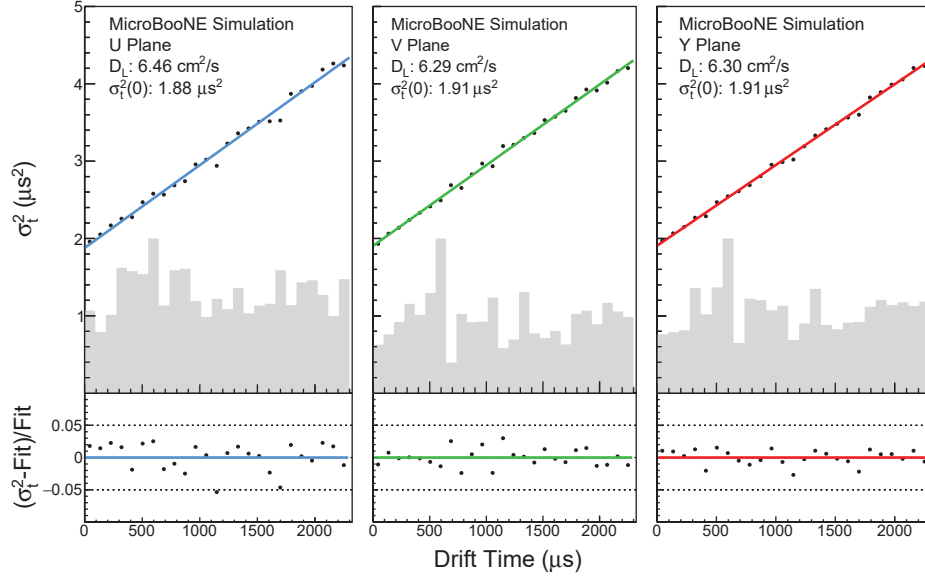


Figure 6.10: Plots of  $\sigma_t^2$  versus  $t$  for simulated muons generated within the angular selection values of  $\theta_{xz} = \pm 6^\circ$  and  $\theta_{yz} = \pm 40^\circ$ . The shaded histograms show the area-normalized distributions of the number of waveforms in each bin. The bottom plots show the fit residuals of each point. The induction planes are used only to estimate systematic uncertainties (see section 6.4.3).

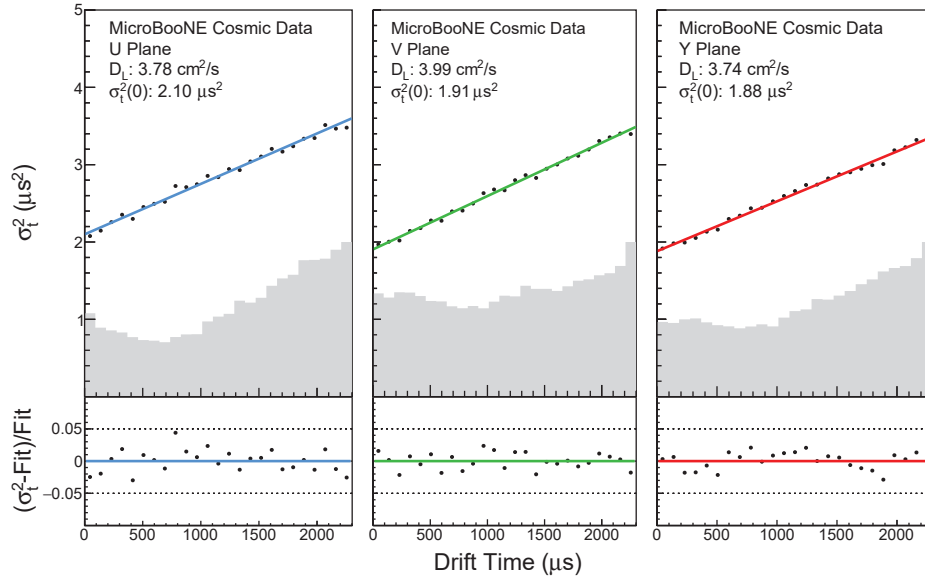


Figure 6.11:  $\sigma_t^2$  versus drift time using MicroBooNE CRT-tagged data. The shaded histograms show the area-normalized distributions of the number of waveforms in each bin. The bottom plots show the fit residuals of each point. The induction planes are used only to estimate systematic uncertainties (see section 6.4.3).

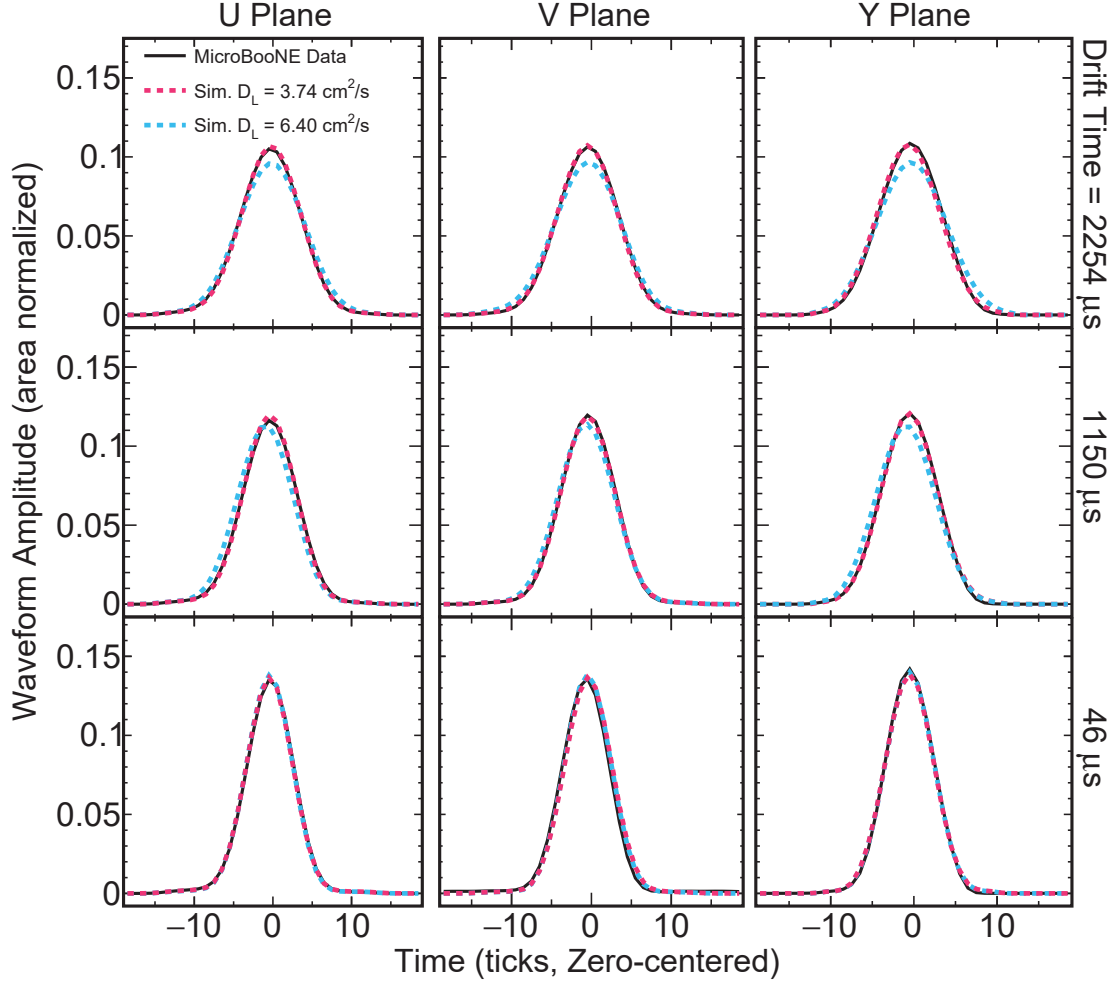


Figure 6.12: Area-normalized comparisons of summed waveforms at three drift times ( $46 \mu\text{s}$ ,  $1150 \mu\text{s}$ ,  $2254 \mu\text{s}$ ) for data, and two simulated datasets with different  $D_L$  values ( $6.40 \text{ cm}^2/\text{s}$  and  $3.74 \text{ cm}^2/\text{s}$ ). To aid in this comparison, we have drawn a smooth line through each bin's contents, rather than showing the original digitized data.

Table 6.2: Summary of the measured values of  $D_L$  from the MicroBooNE data and simulation. The value extracted on the Y plane constitutes our final measurement. The induction planes are used only to estimate systematic uncertainties (see section 6.4.3).

Sample	Measured $D_L$ Value ( $\text{cm}^2/\text{s}$ )		
	U Plane	V Plane	Y Plane
Simulation ( $D_L = 6.4 \text{ cm}^2/\text{s}$ )	6.46	6.29	6.30
Data	3.78	3.99	3.74

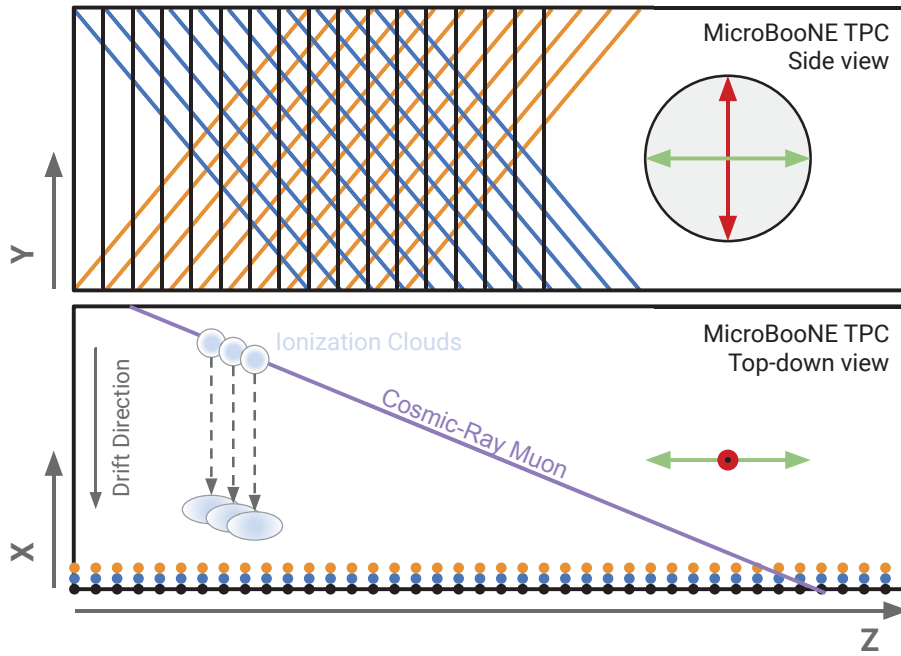


Figure 6.13: Illustration of the impact of  $D_T$  on the  $D_L$  measurement.  $D_T$  causes electron clouds (light blue gradient) to spread in the  $yz$ -plane as a function of drift distance. For tracks with non-zero  $\theta_{xz}$  (purple line), the  $z$ -direction spread (green arrows) causes adjacent electron clouds to overlap as they drift toward the anode plane. The  $y$ -direction spread (red arrow) does not significantly impact  $D_L$ .

$$\frac{D_L}{D_T} = 1 + \frac{E}{\mu(E)} \frac{\partial \mu(E)}{\partial E}, \quad (6.2)$$

where  $\mu(E)$  is the electron mobility as a function of electric field strength [90]. Figure 10 of Reference [90] shows that, at the MicroBooNE E-field of 273.9 V/cm,  $\partial \mu(E)/\partial E$  is approximately constant, and thus  $D_L/D_T$  is constant. We choose  $D_T$  variation values of 4.8 cm<sup>2</sup>/s (down), 5.7 cm<sup>2</sup>/s (central value), and 7.2 cm<sup>2</sup>/s (up). These values are scaled from the nominal simulated MicroBooNE  $D_T$  value and uncertainties, which were designed using the Atrazhev-Timoshkin theory [88] and the available world data [89, 90].

Table 6.3 shows the results of running these  $D_T$ -varied samples through the  $D_L$  analysis. The measured  $D_L$  central values and  $\sigma_0^2$  values show virtually no change when varying  $D_T$ . We attribute this to the two-dimensional nature of the MicroBooNE deconvolution—which deconvolves the signal in both time and wire space, mitigating the impact of charge spread to neighboring wires—and our stringent requirement on the value of  $\theta_{xz}$ . We conclude that the uncertainty on  $D_T$  does not contribute to the systematic uncertainty on the  $D_L$  measurement.

### 6.4.2 Drift Velocity

Equation 6.1 shows that  $D_L$  is proportional to  $v_d^2$ , meaning that any uncertainty in  $v_d$  could lead to a sizeable systematic uncertainty on  $D_L$ . MicroBooNE has measured the drift velocity across the active volume of the detector using UV laser and cosmic data [60, 92]. Across the anode plane, the drift velocity is not constant due to edge effects near the field cage. To extract  $D_L$  from the measured slope in Figure 6.11 (using Equation 6.1), we use  $v_d = 1.076$  mm/ $\mu$ s, the average value of the measured drift velocity across the anode plane.

To evaluate a systematic uncertainty on the measurement from the drift velocity, we take  $1\sigma$  variations of  $v_d$  near the anode and recalculate  $D_L$  using these varied  $v_d$  values. Figure 6.14 shows a 2D map of the percent variation of  $v_d$  with respect to  $v_d = 1.076$  mm/ $\mu$ s in a  $yz$ -slice near the anode. The drift velocity values in each bin come from the UV laser data map which was calculated using data from a dedicated calibration run in Summer 2016. Here, we ignore any bins that fall outside the waveform fiducial volume (see section 6.2.2). The maximum  $v_d$  variation is approximately 3% in the region near  $z = 400$  cm where

$y < 0$ . However, Figure 6.6 shows that our selected waveforms fall mostly in the region where  $z > 800$  cm. In this region, the drift velocity map shows that the  $v_d$  variations are sub-percent level.

Additional sources of uncertainty on  $v_d$  include the statistical and systematic uncertainties on the drift velocity map and cosmic ray flux variations over time. Reference [92] shows that the uncertainties in the drift velocity map are dominated by statistical errors, but those errors are sub-percent level in our region of interest. The drift velocity map was calculated using laser data during the Summer of 2016, while the CRT data used in this analysis was taken between October 2017 and March 2018. Time variations of the SCE were studied in Reference [60] and found to be small compared to the absolute scale of the effect. We therefore conclude that variations in SCE due to cosmic ray flux variations are already accounted for in the drift velocity map.

Considering that variations of  $v_d$  in our region of interest are sub-percent level and that other potential sources of uncertainty are small, we choose to apply a  $\pm 2\%$  variation to  $v_d$ . Varying the anode  $v_d$  up and down by 2% yields variation values of 1.098 mm/ $\mu s$  and 1.055 mm/ $\mu s$ , respectively. This difference covers any impact caused by cosmic ray flux variation and statistical uncertainties in the drift velocity map. Re-calculating the  $D_L$  value shown in Figure 6.11 using these variation values, we obtain an asymmetric drift velocity systematic uncertainty of +3.9%, -4.1%.

### 6.4.3 Detector Response Function

Equation 5.1 shows that the MicroBooNE 2D deconvolution depends on the detector response function,  $R(\omega)$ , as part of the deconvolution kernel. The response function has been validated on each of the three wire planes using MicroBooNE data [59], but small uncertainties on the width of the field response function<sup>1</sup> can have a significant impact on the width of deconvolved waveforms which in turn impacts  $D_L$ . While the final  $D_L$  measurement uses only collection-plane waveforms, we can perform the measurement on each of the three wire planes as shown in figures 6.10 and 6.11. Since the response function on each plane was tuned independently of the others, we expect some difference in the extracted  $D_L$  on each plane.

---

<sup>1</sup>Recall that the response function is itself a convolution of a field response and electronics response.



Table 6.3: Results of  $\sigma_t^2$  vs.  $t$  fits for simulated muon particle gun samples with  $D_T$  varied.

Simulated $D_T$ (cm <sup>2</sup> /s)	Measured $D_L$ (cm <sup>2</sup> /s)	Measured $\sigma_0^2$ ( $\mu$ s <sup>2</sup> )
4.80	6.26	1.96
5.70	6.26	1.97
7.20	6.25	1.98

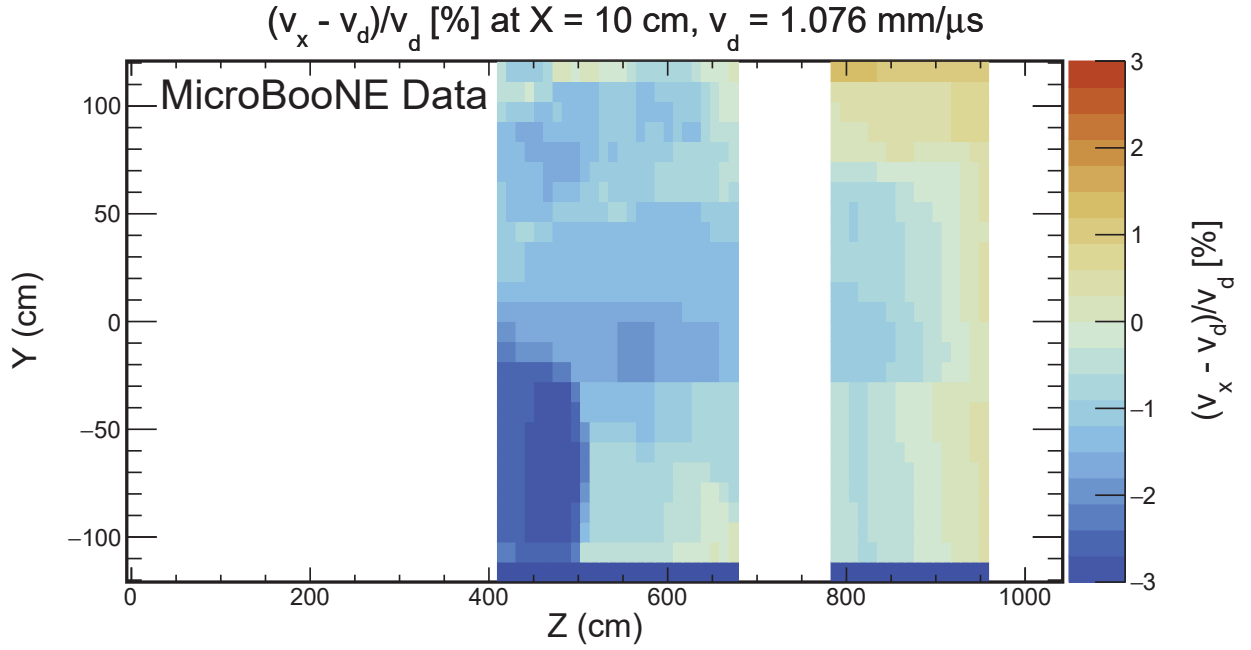


Figure 6.14: 2D distribution of the percent variation of the drift velocity relative to the average drift velocity near the anode,  $v_d = 1.076$  mm/ $\mu$ s, using the UV laser data map. Here, we've applied the waveform fiducial volume described in section 6.2.2

The difference in the measured  $D_L$  serves as a conservative estimate of the uncertainty of the wire response tuning method.

Table 6.2 shows that the maximum cross-plane difference in  $D_L$  is 6.5% corresponding to the difference between the V and Y planes. We therefore take 6.5% as the systematic uncertainty on the response function modeling.

#### 6.4.4 Waveform Summation Method

The waveform summation technique described in section 6.2.2 may introduce additional broadening in the summed waveform. When aligning two waveforms, we can only shift them by integer tick values meaning that the peaks may be misaligned by as much as half a tick. We mitigate this smearing by taking the configuration which minimizes the resultant hit width, but there may still be some residual broadening.

To check the impact of this effect and whether the impact is drift-dependent, we perform a study in which we sum 1000 idealized Gaussian waveforms under different conditions. We start by generating an initial Gaussian whose mean and standard deviation resemble those of waveforms from particle interactions near the anode; here, we chose “anode-like” values of  $\mu = 891.5$  ticks<sup>2</sup> and  $\sigma = 1.42$  ticks.<sup>3</sup> To simulate the impact of misalignment, we also apply a random shift drawn from a uniform distribution between  $-0.5$  and  $+0.5$  ticks to the mean of this initial Gaussian. In the control case, we simply add this waveform to itself 1000 times using our waveform summation technique. Then, to simulate the effect of adding misaligned waveforms, we instead add 1000 waveforms with the same  $\sigma$  as the initial generated Gaussian, but whose means have been shifted randomly between  $-0.5$  and  $+0.5$  ticks. Any difference in the extracted  $\mu$  and  $\sigma$  is attributed to the summation technique. We then repeat this study using “cathode-like” waveform values of  $\mu = 5123.4$  ticks and  $\sigma = 3.80$  ticks. Figure 6.15 and Table 6.4 summarize the results of this study. While  $\sigma$  does increase slightly in each case, the broadening is consistent at both the anode and the cathode. This may impact our extracted  $\sigma_t^2(0)$  but not  $D_L$ . We repeated this study multiple times to account for different

---

<sup>2</sup>MicroBooNE TPC waveforms are recorded beginning 800 ticks before the trigger time, so the position of the anode is at 800 ticks.

<sup>3</sup> $\sigma$  here should not be confused with  $\sigma_t$ , the time width of measured signal pulses.

random shifts in the mean of the initial Gaussian and found no significant change in the results, including for cases where the initial Gaussian was shifted by the maximum allowed value ( $\pm 0.5$  ticks). We conclude that the waveform summation technique does not introduce a sizeable systematic uncertainty to the  $D_L$  measurement.

#### 6.4.5 Summary and Other Systematic Uncertainties

Other systematic uncertainties that may impact the  $D_L$  measurement include microphysics effects that are either drift-dependent or field-dependent, particularly SCE, electron-ion recombination, and electron attenuation. For SCE, the size of the electron cloud when it arrives at the anode wire plane depends only on the amount of time that has elapsed since the electrons were ionized. We measure this time directly by using the  $t_0$  extracted from CRT information meaning that the measurement is not biased by the presence of space charge. Thus, the measured slope of the line in Figure 6.11 has no systematic uncertainty due to space charge. The strength of electron-ion recombination changes with the electric field, but, for MicroBooNE E-field fluctuations, this effect is small [93]. Moreover, the impact of the recombination systematic uncertainty on collected charge is much smaller than the impact of statistical Landau fluctuations in the density of ionization electron clouds. As for electron attenuation, the measured electron lifetime in MicroBooNE [61] is 18 ms. The maximum drift for a single electron is 2.3 ms meaning that charge attenuation in MicroBooNE is minimal, and this is due to the extremely high argon purity in the TPC. We conclude that both electron recombination and attenuation do not contribute to the systematic uncertainty on  $D_L$ .

Table 6.5 summarizes the  $D_L$  systematic uncertainties. The two dominant systematic uncertainties come from the uncertainties on the response function modeling and the drift velocity. We have considered many other potential sources of systematic uncertainties but found them to be sub-dominant. We assume that the individual systematic uncertainties are uncorrelated and add them in quadrature to obtain the total systematic uncertainty of  $+7.6\%$ ,  $-7.7\%$ . This results in our final measurement from the MicroBooNE data of  $D_L = 3.74^{+0.28}_{-0.29} \text{ cm}^2/\text{s}$ .

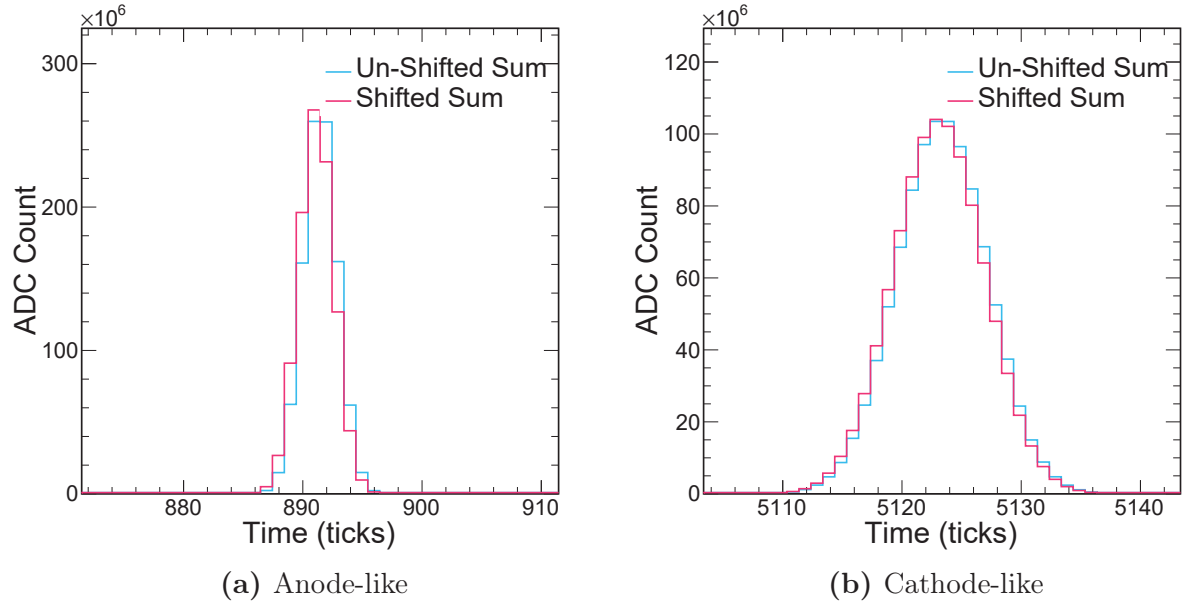


Figure 6.15: Results of the study of the waveform summation technique for anode-like (left) and cathode-like (right) Gaussians.

Table 6.4: Results of a study of waveform summation. “Un-shifted” denotes the control case, in which we add the same Gaussian to itself 1000 times, while “Shifted” denotes the case in which each added waveform has its mean randomly shifted before addition.

	$\mu$ (ticks)	$\sigma$ (ticks)
<b>Anode-like</b>		
Un-shifted	891.5	1.44
Shifted	891.1	1.48
<b>Cathode-like</b>		
Un-shifted	5123.4	3.80
Shifted	5123.1	3.83

Table 6.5: Summary of systematic uncertainties on the  $D_L$  measurement. The total uncertainty assumes that the systematic uncertainties are uncorrelated.

Systematic	Value
Response Function	6.5%
Drift Velocity	+3.9%, -4.1%
$D_T$	< 1%
Waveform Summation	< 1%
Noise and microphysics	< 1%
<b>Total</b>	+7.6%, -7.7%

## 6.5 Discussion and Conclusions

We report a measurement of the effective longitudinal electron diffusion coefficient of  $D_L = 3.74^{+0.28}_{-0.29}$  cm<sup>2</sup>/s at an E-field of 273.9 V/cm. This represents the first measurement in a large-scale (90 tonne) LArTPC. Figure 6.16 shows the measured  $D_L$  value in MicroBooNE as it compares to the Li et al. parametrization, the theory curve, and the available data from ICARUS and Li et al. The vertical error bars correspond to systematic uncertainties on  $D_L$ , while the horizontal error bars account for the maximum E-field variation values of  $273.9^{+12\%}_{-8\%}$ . The MicroBooNE  $D_L$  value sits slightly below the theory curve even when including systematic uncertainties, but it should be noted that this curve is ill-defined for E-fields greater than zero and below  $\sim 1$  kV/cm. We used an interpolation in that region, the details of which are described in Appendix A. Our measurement is in better agreement with the ICARUS measurement and the Atrazhev-Timoshkin prediction than the measurement and parametrization of Li et al.

At present, the cause of tension among  $D_L$  measurements is unknown. Li et al. performed their measurements using a gridded drift cell, similar to historical measurements performed in gaseous media [94, 95], with a maximum drift distance of 60 mm. They note the possibility of underestimating the impact of Coulomb repulsion among the drifting electrons, which they calculated using an approximate model described in Reference [96]. Based on their calculations, Li et al. chose not to apply a correction for this effect. ICARUS, however, concluded that this effect contributes significantly to their measured value when using the same model [89]. We apply no such correction in this work. Further measurements in LArTPCs are needed in order to resolve this tension.

One potential application of the  $D_L$  measurement is to  $t_0$  tag small energy depositions to investigate MeV-scale physics in LArTPCs. The potential for using diffusion to  $t_0$ -tag single waveforms has been investigated and is presented in Appendix B.

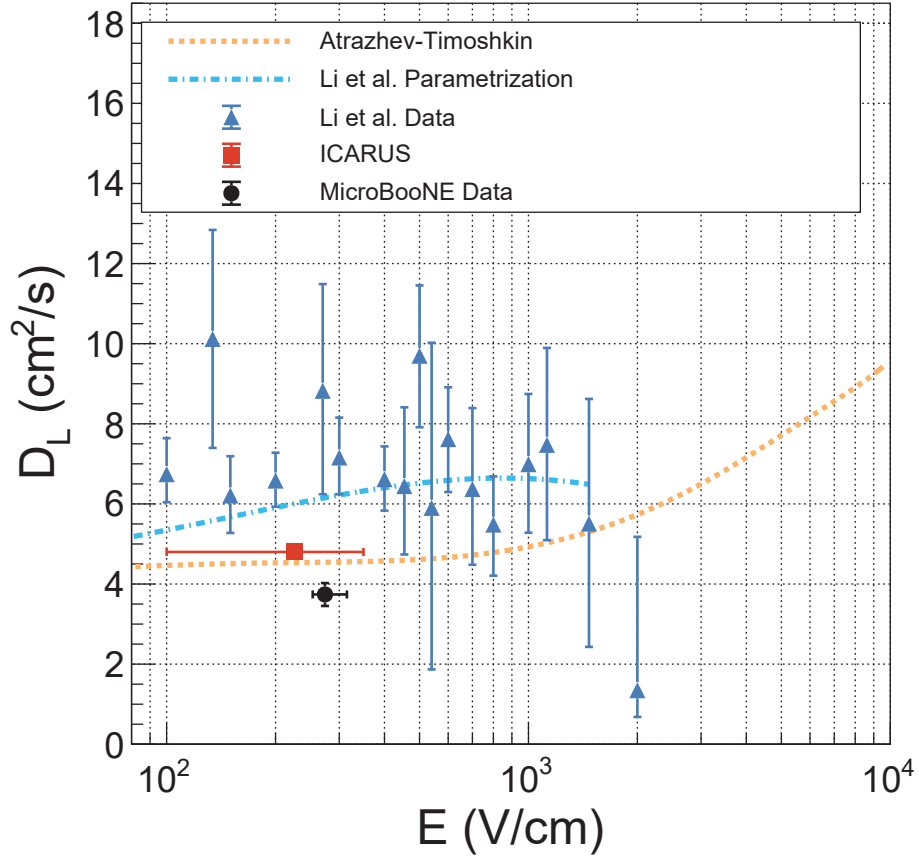


Figure 6.16: Comparison of the MicroBooNE result with world data [89, 90] for  $D_L$ , along with the Atrazhev-Timoshkin theory curve [88] and the parametrization of Li et al. [90]. Note that the theory curve is not well-defined between 0 and  $\sim 1000$  V/cm, so we use a fourth-order polynomial interpolation between 10 and 1200 V/cm; see Appendix A for details.

# Chapter 7

## The MicroBooNE Single-Photon Analysis

This chapter presents the current status of the MicroBooNE search for single-photon events as they pertain to the photon-like hypothesis of the MiniBooNE low-energy excess (LEE, see Section 3.5). Although my work is specifically in the NC  $\pi^0$  sideband selection, we discuss the single photon selection here in order to contextualize the NC  $\pi^0$  selection. Furthermore, the two analyses share a similar framework, so all discussion presented on the single-photon analysis is directly relevant to the NC  $\pi^0$  analysis. Section 7.1 introduces the NC  $\Delta$  resonance, the radiative decay of which is the dominant contributor to single-photon events in MicroBooNE. Section 7.2 presents an overview of the common elements between the single photon and NC  $\pi^0$  analyses, including common analysis inputs (Section 7.2.1) and an overview of Boosted Decision Trees (BDTs, Section 7.3). Finally, Section 7.4 briefly discusses the current status of the single-photon selection.

### 7.1 NC $\Delta$ Resonance

Recall from Section 4.1 that the BNB flux peaks at around 700 MeV. At these neutrino energies, the most common source of single-photon events is the production of an NC  $\Delta$  resonance followed by  $\Delta$  radiative decay. A diagram of this process is shown in Figure 7.1a. Although this process is the leading contributor of single-photon events, the far more

common decay mode of the  $\Delta$  is NC resonant  $\pi^0$  production, shown in Figure 7.1b. The branching fraction of resonant  $\pi^0$  production is 99.4%, while the radiative decay mode is inferred to be between 0.55 and 0.65% [22]. However,  $\Delta$  radiative decay has never been observed in neutrino scattering. Experimental measurements come from electron scattering experiments [97], and the branching ratios for neutrino-induced resonance decay have been estimated using the electron scattering data.

To date, two experimental searches have been performed for NC  $\Delta$  radiative decay in neutrino scattering. The NOMAD experiment [98] searched for neutrino-induced single-photon events (not necessarily CC or NC) from the 25 GeV Super Proton Synchrotron (SPS) beam. They reported a null measurement of an excess of single-photon events, and placed an upper bound of  $4 \times 10^{-4}$  single-photon events per CC  $\nu_\mu$  interaction. The T2K experiment searched for NC single-photon events using neutrinos from the J-PARC beam at an average energy of 600 MeV [99]. T2K also reported a null result, and placed an upper bound on the NC  $\Delta$  radiative cross section at  $0.114 \times 10^{-38} \text{ cm}^2$  with a 90% confidence level.

As a potential explanation for the MiniBooNE LEE, MicroBooNE is performing a similar search for NC  $\Delta$  radiative decays in neutrino-nucleus scattering. The single photon emitted during this process can mimic the single-electron signal in  $\nu_\mu \rightarrow \nu_e$  oscillation searches. The lack of experimental observation of this process points to a potential for a mis-modeled or underestimated interaction rate. Previous work on MicroBooNE has shown that, in order to explain the MiniBooNE LEE, the NC  $\Delta$  radiative decay standard model (SM) rate must be scaled up by a factor of three [100]. Figure 7.2 (modified from Reference [99]) shows the T2K upper limit along with the J-PARC neutrino flux, the theoretical prediction of Wang *et al.* [101], and the NOMAD measurement. Figure 7.2 also shows the theoretical calculation scaled up by a factor of three (green line) to show the scaling required to explain the MiniBooNE LEE. Even assuming this enhanced NC  $\Delta$  radiative production rate, the T2K upper limit cannot definitely exclude this hypothesis at  $\sim 1$  GeV-scale energies.

A substantial hurdle in measuring NC  $\Delta$  radiative decay is the large background from NC resonant  $\pi^0$  decays. As previously mentioned, the resonant  $\pi^0$  production mode is far more common; we expect roughly 200 NC resonant  $\pi^0$  events for each NC  $\Delta$  radiative decay. NC  $\pi^0$  events in which one photon isn't detected—possibly due to exiting the detector, being too



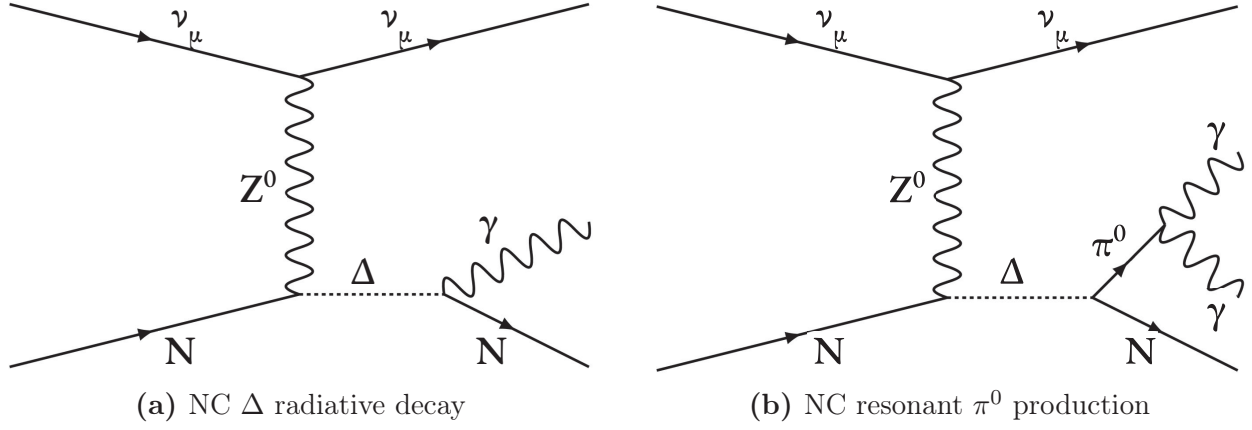


Figure 7.1: Diagrams of (a) NC  $\Delta$  radiative decay and (b) NC resonant  $\pi^0$  production.

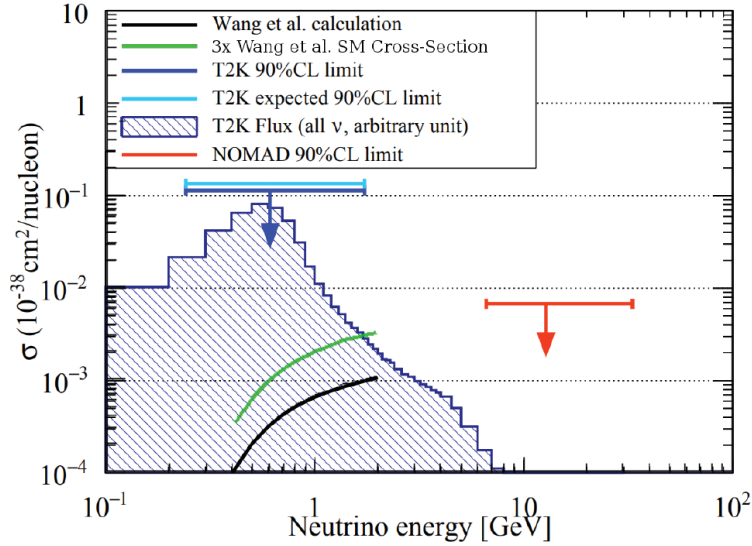


Figure 7.2: Current world data on neutrino-induced NC  $\Delta$  radiative decay [99]. The blue lines show the T2K upper bound (expectation from MC in cyan, actual measurement in dark blue), the red line shows the NOMAD measurement, and the blue histogram (arbitrarily scaled) shows the J-PARC neutrino flux prediction. Also shown are the theoretical calculation of Wang *et al.* [101] (black line) and this same calculation scaled up by a factor of three (green line) to represent the scaling required to explain the MiniBooNE LEE [100].

low-energy, or being mis-reconstructed—can mimic the NC  $\Delta$  radiative signal. To constrain this background, most experiments measure the NC resonant  $\pi^0$  production rate *in situ*. In particular, MiniBooNE constrained their systematic uncertainties on photon backgrounds using a large sample of NC  $\pi^0$  events from their NC  $\pi^0$  cross section measurement on CH<sub>2</sub> [14]. In MicroBooNE, we seek to perform a similar constraint, which will be covered in detail in the next chapter.

## 7.2 Analysis Overview

To begin, it is prudent to emphasize that when we talk about the MicroBooNE single-photon selection, what we are actually referring to is a search for the radiative decay of the  $\Delta(1232)$ . At the average BNB energy of  $\sim 700$  MeV, the  $\Delta(1232)$  is the dominant nucleon resonance. Other sources of single-photon production—such as higher-order  $\Delta$  resonances, coherent single-photon production, and MEC interactions (see Section 2.4.3)—have been considered in [102], but are expected to be subdominant to incoherent  $\Delta(1232)$  radiative decay. Although the  $\Delta$  radiative decay is the dominant source of single-photon events, we still expect to see only a few hundred  $\Delta$  radiative decay events in MicroBooNE, even when assuming a factor of three enhancement in the SM prediction. We therefore simply refer to the  $\Delta$  radiative decay search as the single photon selection.

As shown in Figure 7.1a, the  $\Delta$  decays into a single photon and a nucleon. The final-state nucleon may be either a proton or a neutron or, in some highly-energetic interactions, there may be multiple final-state nucleons. If a proton exits the nucleus, we expect a relatively short, highly-ionizing particle track to be reconstructed. Neutrons, on the other hand, propagate invisibly in LArTPCs, and therefore cannot be reconstructed as a track. We therefore target the one-shower, one-track topology ( $1\gamma 1p$ ) and the one-shower, zero-track topology ( $1\gamma 0p$ ). Final states with more than one reconstructed track are not considered in this analysis.

In tandem with the single-photon selection, we’ve developed a targeted NC  $\pi^0$  sideband selection. In this case, we target the  $2\gamma 1p$  and  $2\gamma 0p$  topologies in order to obtain mutually exclusive, independent samples of NC  $\pi^0$  events. We expect two reconstructed

electromagnetic (EM) showers corresponding to two photons from  $\pi^0 \rightarrow \gamma\gamma$  decay. This sample then provides an *in situ* constraint on the correlated interaction uncertainties in the single-photon selection.

The efficiencies for each topological selection are shown in Table 7.1. The left column describes various simulated samples used. Here, BNB All refers to all on-beam backgrounds, NC  $\pi^0$  is the resonant  $\pi^0$  background, NC  $\Delta_{\text{Rad}}$  refers to the  $\Delta$  radiative signal, BNB  $\nu_e$  refers specifically to the on-beam background contribution from  $\nu_e$  contamination in the neutrino beam, BNB External is the off-beam cosmic background, and Dirt refers to events which occur outside the TPC and scatter in. The second column shows the percentage of events for each sample that contain a candidate neutrino vertex (see Section 5.3.2). This efficiency is lowest for off-beam backgrounds due to Pandora’s cosmic rejection algorithms. The remaining columns show the topological reconstruction efficiency for each of the selection topologies described in the preceding paragraphs. These efficiencies are generally low due to the challenge of reconstructing EM showers, particularly for the  $2\gamma$  cases. We choose to proceed with these rigid topological selections in order to obtain independent, mutually-exclusive selections for each topology.

Once the signal topologies have been identified and selected, we then apply a series of conservative “pre-selection” cuts. These cuts vary depending on the signal topology, but generally consist of, for example, reconstructed shower energy thresholds, minimum shower conversion distances, etc. The pre-selection cuts are not intended to reject a significant number of backgrounds, but instead are utilized to remove events that may be mis-reconstructed. For example, low-energy showers may be reconstructed as tracks, so we apply shower energy thresholds. The process of background rejection begins in the next stage of the analysis in which we apply tailored Boosted Decision Trees (BDTs) to separate signal-like events from background-like events. Section 7.3 will give a brief overview of BDTs, while Section 7.4 and Chapter 8 will discuss the specific implementation used in each analysis.

### 7.2.1 Analysis Inputs

Both the single photon and NC  $\pi^0$  selections rely on simulated samples generated using the GENIE v3 model predictions described in Section 5.1. We use these GENIE predictions

to generate dedicated samples of various types of signal and backgrounds for each analysis. Both the  $1\gamma$  and  $2\gamma$  selections use the same set of simulation samples. The simulated sample names and definitions are as follows:

- **SM NC  $\Delta$  Radiative**: True NC  $\Delta$  radiative decay according to the standard model prediction (i.e., no factor-of-three enhancement)
- **NC  $\pi^0$  Non-Coherent**: NC  $\pi^0$  events with one final-state  $\pi^0$
- **CC  $\nu_\mu$   $\pi^0$** : CC  $\nu_\mu$ <sup>1</sup>  $\pi^0$  events with one final-state  $\pi^0$
- **BNB Other**: On-beam backgrounds which do not fall into any other category listed here
- **CC  $\nu_e/\bar{\nu}_e$** : Events which contain a  $\nu_e/\bar{\nu}_e$  interaction from  $\nu_e$  contamination in the beam, also referred to as “intrinsic”  $\nu_e$ s
- **Dirt (Outside TPC)**: Events which occur outside the TPC and scatter into the active volume
- **Cosmic Data Overlays**: Off-beam cosmic data overlaid on top of GENIE-generated neutrino interactions used to emulate the effect of cosmic contamination in on-beam events

For both the  $1\gamma$  and  $2\gamma$  selections, we denote a corresponding signal sample (NC  $\Delta$  radiative decay for  $1\gamma$ , NC  $\pi^0$  for  $2\gamma$ ) and refer to the others as backgrounds. These simulation samples are common to each topology, but when performing data-to-MC comparisons, the on-beam data samples are different for the  $1\gamma$  and  $2\gamma$  selections. The reason for this is that MicroBooNE is performing a blind analysis for LEE-like signals, including the NC  $\Delta$  radiative decay signal and  $\nu_\mu \rightarrow \nu_e$  oscillation signals. Any LEE search is therefore restricted to a small subset of the total collected data. This “unblinded” dataset corresponds to approximately  $5 \times 10^{19}$  POT and is often simply referred to as the “5e19” sample. The 5e19 sample comes entirely from Run 1 of data-taking, so a smaller subset of Run 3 data is also unblinded

---

<sup>1</sup>Note that the  $\nu_\mu$  CC  $\pi^0$  sample is the only one that was generated without  $\bar{\nu}_\mu$ . This difference has no significant impact on any of the selections.

corresponding to roughly  $0.73 \times 10^{19}$  POT. On the other hand, the NC  $\pi^0$  analysis is blind to both  $\Delta$  radiative and  $\nu_\mu \rightarrow \nu_e$  oscillation signals and is therefore allowed to analyze a larger dataset. This is done through the implementation of data filters targeting the  $2\gamma 1p$  and  $2\gamma 0p$  topologies. Appendix C describes the data filters implemented in the  $2\gamma 1p$  and  $2\gamma 0p$  selections in detail. These filters are applied to Runs 1, 2 and 3 of data-taking, and correspond to about  $5.8 \times 10^{20}$  POT each.

The simulation and data samples used in each analysis are summarized in Table 7.2.<sup>2</sup> The BNB, dirt, and filtered data samples contain events from Runs 1–3, while the NC  $\Delta$  radiative, NC  $\pi^0$ ,  $\nu_\mu$  CC  $\pi^0$ , and  $\nu_e$  samples contain events only from Runs 1 and 3. Unless stated otherwise, all data-to-MC comparisons shown in this section correspond to a selection using a combination of Runs 1, 2, and 3. Note that, when comparing data to MC, the MC samples are all scaled to match the POT of the corresponding data sample. For example, when looking at  $2\gamma 1p$  filtered data, the simulated NC  $\pi^0$  sample is scaled by a factor of  $5.8e20/5.0e21 \simeq 0.08$ , and so on for the other simulated samples.

### 7.2.2 Shower Energy Correction

In general, the reconstructed shower energy in MicroBooNE is systematically below the true shower energy due to lossy effects in the detector volume, namely mis-clustering and thresholding. Showers are clustered via Pandora, but this often fails to correctly cluster the entire shower, leading to a partially reconstructed shower. Furthermore, some shower hits fall below our energy threshold and therefore aren't reconstructed. On average, we expect shower energy losses of  $\sim 20\%$  due to these effects [84].

To account for this, we derive a shower energy correction factor. Using Run 1 NC  $\pi^0$  production overlay samples, we begin by plotting reconstructed photon energy vs. true photon energy for all showers with at least 30 MeV of simulated and reconstructed energy; see Fig. 7.3a. As expected, the reconstructed energy is systematically below the simulated energy. We then perform a linear fit to the most probable value (MPV) in variable-sized slices of true shower energy. The width of each true shower energy slice is hand-tuned to

---

<sup>2</sup>The slight difference in POT between the two filters is due to grid processing inefficiencies; we processed the  $2\gamma 0p$  filter several months after the  $1p$  filter.

Table 7.1: Topological efficiencies for the  $1\gamma 0p$ ,  $1\gamma 1p$ ,  $2\gamma 0p$ , and  $2\gamma 1p$  selections. The sample definitions (left column) are described in the text. “ $\nu$  Candidate” refers to the percentage of events in each sample which contain a candidate neutrino vertex.

Sample	$\nu$ Candidate	$1\gamma 0p$	$1\gamma 1p$	$2\gamma 0p$	$2\gamma 1p$
BNB All	43.0%	0.7%	2.9 %	0.4%	0.8%
NC $\pi^0$	41.9%	3.8%	7.1 %	3.6%	5.3%
NC $\Delta_{\text{Rad}}$ (All)	62.5%	12.8%	17.5 %	3.0%	3.8%
NC $\Delta_{\text{Rad}}$ ( $1\gamma 1p$ Signal)	72.5%	9.63%	28.9 %	2.3%	4.1%
NC $\Delta_{\text{Rad}}$ ( $1\gamma 0p$ Signal)	64.3%	20.6%	11.6 %	4.3%	4.0%
BNB $\nu_e$	79.7%	6.5%	16.2%	2.9 %	6.1%
BNB External	15.0%	0.37%	1.68%	0.13%	0.27%
Dirt	22.8%	0.49%	1.77%	0.14%	0.24%

Table 7.2: Summary of simulation and data samples used in the single photon and NC  $\pi^0$  analyses. Here, we show number of events and POT equivalent for each sample. Note that for filtered  $2\gamma$  on-beam data, the number of events only shows those that pass the filter, while the quoted POT for MC samples corresponds to all generated events.

Sample	No. Events	POT
Simulation		
SM NC $\Delta$ Rad.	39k	1.9e23
NC $\pi^0$	191k	5.0e21
$\nu_\mu$ CC $\pi^0$	280k	9.8e21
BNB Other	2.7M	3.3e21
Intrinsic $\nu_e/\bar{\nu}_e$	42k	2.5e22
Dirt (Outside TPC)	511k	1.6e21
Data		
Unblinded 5e19 data	157k	4.1e19
Open Run 3 data	28k	0.7e19
$2\gamma 1p$ Filtered On-Beam Data	4,609	5.8e20
$2\gamma 0p$ Filtered On-Beam Data	3,223	5.9e20
Cosmic Data	7.7M	—

account for lower statistics at high energies and to provide a reasonable fit to the underlying 2D distribution. We then take the equation of this linear fit, and derive a correction factor that corrects the line to  $y = x$ . The linear fit equation is

$$E_{reco} = (0.83 \pm 0.02)E_{true} + (-8.15 \pm 3.79) \text{ MeV}, \quad (7.1)$$

where  $E_{reco}$  and  $E_{true}$  correspond to the reconstructed and true shower energies, respectively. This yields the correction equation

$$E_{corr} = (1.21 \pm 0.03)E_{reco} - (-9.88 \pm -4.86) \text{ MeV}. \quad (7.2)$$

As expected, this represents an approximately 20% correction. Figure 7.3b shows the corrected shower energy vs. true energy. The effect of the correction on the  $\pi^0$  invariant mass can be seen in later sections such as Figure 8.5a, where the result is a distribution whose peak more closely aligns with the expected  $\pi^0$  mass of 135 MeV, while preserving data/MC agreement. This correction is applied to all  $1\gamma$  and  $2\gamma$  selection whenever we are discussing shower energies, or higher order quantities such as invariant masses, which are functions of shower energies.

### 7.2.3 Notes on Nomenclature

Certain  $\pi^0$  kinematic variables, such as the invariant mass and opening angle between the photons, depend on both the energy and direction of the showers. When showing plots in which we’ve applied the shower energy correction from Section 7.2.2, we label the plot as “corrected.” For shower directions, there are actually two possible definitions. One is the default direction as reconstructed by Pandora. Pandora reconstructs shower directions based on the direction of the EM cluster as a whole. In general, this direction should point back to the neutrino interaction vertex, but this isn’t always the case. For this analysis, we choose to use what we call the “implied” shower direction, defined as a unit vector pointing straight from the neutrino interaction vertex to the shower start point. We have generally found that these implied shower directions provide better data-to-MC agreement and have therefore chosen to use this definition of shower direction when reconstructing, for example, the  $\pi^0$

invariant mass. A  $\pi^0$  mass peak with both the shower energy correction and the implied shower directions will therefore be labeled as the “corrected implied  $\pi^0$  invariant mass.” The same is true for  $\pi^0$  momentum, opening angle between the showers, etc.

Finally, it is important to note that the term “signal” has some ambiguity in the following sections and chapter. For the  $1\gamma$  selections, the signal is a factor-of-three enhancement to the standard model rate of  $\Delta$  radiative decay. This is the signal enhancement required to explain the MiniBooNE LEE. For the NC  $\pi^0$  selection, the signal is defined as a neutral current interaction with a single  $\pi^0$  in the final state (see Chapter 8 for more details). Note that, by this definition, non-resonant  $\pi^0$  production is also considered signal. We expect non-resonant  $\pi^0$  production to account for less than 20% of the final selection (see Table 8.5 in Section 8.4).

## 7.3 Boosted Decision Trees

Before proceeding to the details of the single photon and NC  $\pi^0$  analysis details, it is instructive to pause here to discuss Boosted Decision Trees (BDTs). This type of machine learning algorithm is central to both analyses, and understanding the basic principles of machine learning is important for understanding the work presented in later sections.

For our purposes, we use BDTs to sort data as either signal-like or background-like. This is known in machine learning as a *classification problem*. Furthermore, the work described here is considered *supervised* learning because the user is able to define the BDT input and monitor the corresponding output. Here, the input to the BDTs is a dataset in which all members have a known classification. For example, we may hand the BDT a sample of MC-generated NC  $\pi^0$  events and classify this as the signal sample. We can similarly input other MC-generated neutrino interactions as background. We then define a vector of *training variables* which are chosen to allow the BDT to learn to distinguish signal from background, where signal and background are defined in the BDT configuration. In addition to the training variables, we also split the input data into *training* and *testing* samples. These samples are independent, mutually exclusive subsets of the full input dataset. The training sample is used to train the BDT to distinguish signal from background, while the testing



sample provides a cross-check of training performance. Ideally, the BDT performance should be similar between the training and testing samples.

A common method of classifying data is through the use of a decision tree. A decision tree is a flowchart-like structure in which a decision is made based on a test performed at each node (or branch). As a simple example, Figure 7.4 [103] shows a decision tree for a person deciding whether to walk to work or take the bus. The most important test, shown at the top (“root”) node, is whether it’s raining outside. If it’s raining, the person takes the bus, and the decision is made. If, on the other hand, the weather is sunny or cloudy, the decision is then further split by how much time the person has (in the case of sunny weather) or how hungry they are (in the case of cloudy weather). Note that the test performed at each node is binary. Even when asking how much time the person has (a continuous quantity), the decision must be made based on some binary split. In this case, the decision point is whether the person has more or fewer than 30 minutes to get to work.

For a realistic classification problem, a single decision tree is insufficient. More robust methods use many decision trees and combine the results into a final decision. Such algorithms can be broadly categorized as *bagging* and *boosting*. In bagging, each tree makes an independent decision in parallel with the other trees. In boosting, decisions are made sequentially, with errors in each tree informing the decision of subsequent trees. In this work, we focus on a particular type of boosting algorithm known as AdaBoost [104] (short for “adaptive boosting”). In the AdaBoost algorithm, misclassified samples from one tree are assigned a larger weight in subsequent trees such that the later trees learn from this misclassification. This is shown visually in Figure 7.5 [105]. Here, the first tree misclassifies two blue circles as red. The second tree then assigns a larger weight to these misclassified samples (denoted by the size of the circles in Figure 7.5) in order to learn from this mistake. This process repeats until all samples are correctly classified, as shown in the rightmost plot of Figure 7.5.

Both the single photon and NC  $\pi^0$  analyses use the eXtreme Gradient Boosting (XGBoost) algorithm [106] to implement BDTs. XGBoost uses a modified version of the AdaBoost algorithm, with the primary difference being how it calculates optimal split points in a decision tree. When deciding where to split a decision tree, most algorithms implement

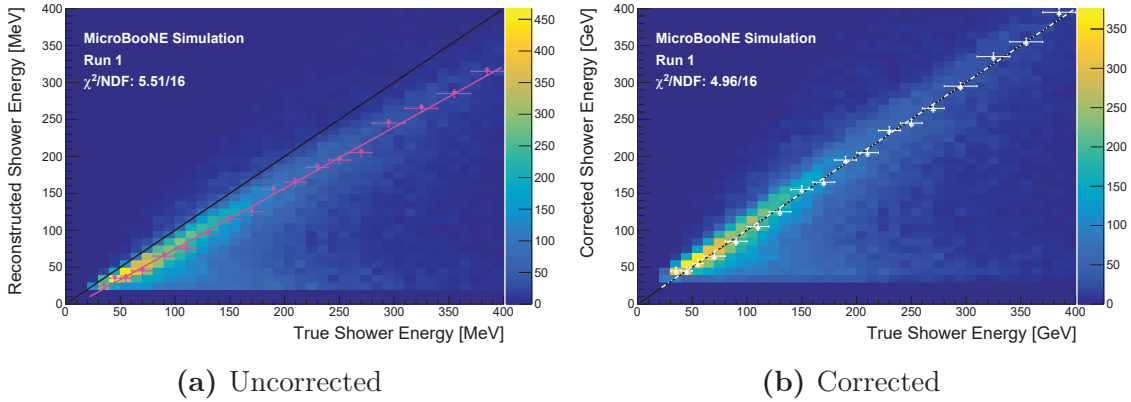


Figure 7.3: (a) 2D distribution of reconstructed shower energy vs. true energy for reconstructed showers with at least 30 MeV, taken from the MC sample of true NC  $\pi^0$  events. The points represent the most probable value in each slice of true energy. The slice width (shown by the horizontal error bars) is hand-tuned to account for the decreasing statistics at high true energy values. (b) Corrected shower energy vs. true shower energy.

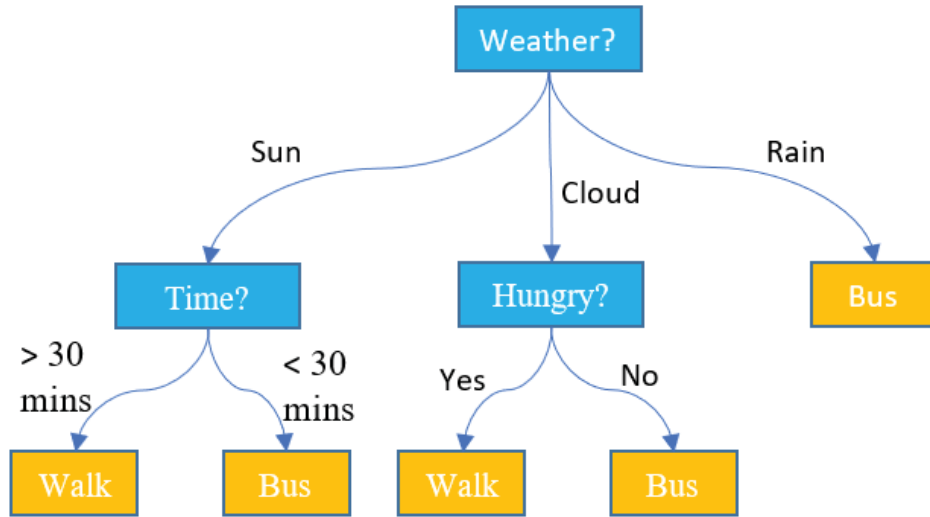


Figure 7.4: Example of a decision tree in which a person commuting to work decides whether to walk or take the bus, taken from Reference [103].

some form of purity metric, where the purity is defined as the percentage of “signal” relative to the total number of entries in each leaf. XGBoost uses a more sophisticated algorithm which attempts to minimize the gradient of the training errors. Reference [106] describes this algorithm in more detail.

One potential drawback of the AdaBoost algorithm and, by extension, its implementation in XGBoost, is the possibility of overtraining or undertraining the BDT (also called overfitting and underfitting). When a BDT is overtrained, it begins to interpret statistical noise in the training sample as a true feature of the underlying data. This leads to a high *variance*<sup>3</sup> in the BDT. Here, variance is defined as the degree to which the BDT prediction fluctuates when given different input data. An overtrained BDT will apply its knowledge of statistical noise in one training set to another, leading to a substantially different prediction in the latter set. On the other hand, an undertrained BDT will perform similarly across input datasets, but will not accurately separate signal from background. In other words, an undertrained BDT has low variance but high *bias*, defined as how closely the BDT prediction matches the optimal prediction. Visual examples of undertrained, overtrained, and optimally trained BDT models are shown in Figure 7.6.

When configuring a BDT, one must account for this *bias-variance tradeoff* in order to obtain a balanced model that accurately describes the data without picking up on statistical noise. In order to ensure a balanced model, a BDT must be configured with a number of *hyperparameters*. Hyperparameters refer specifically to those parameters which configure the BDT learning process, not to be confused with more general parameters relating to the input data. A brief description of the relevant XGBoost hyperparameters are listed below.

- **Number of trees:** the number of decision trees in the BDT. Too few trees leads to undertraining, while too many trees can lead to overtraining.
- **Maximum tree depth:** the largest number of decision nodes in any single line of the decision tree. For example, the tree shown in Figure 7.4 has a maximum depth of 3.

---

<sup>3</sup>Note that variance here should not be confused with the statistical term referring to the square of the standard deviation,  $\sigma^2$ .

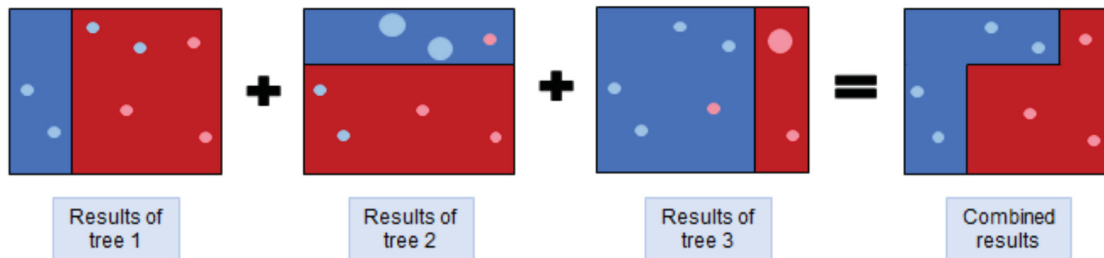


Figure 7.5: Visualization of the AdaBoost algorithm, taken from Reference [105].

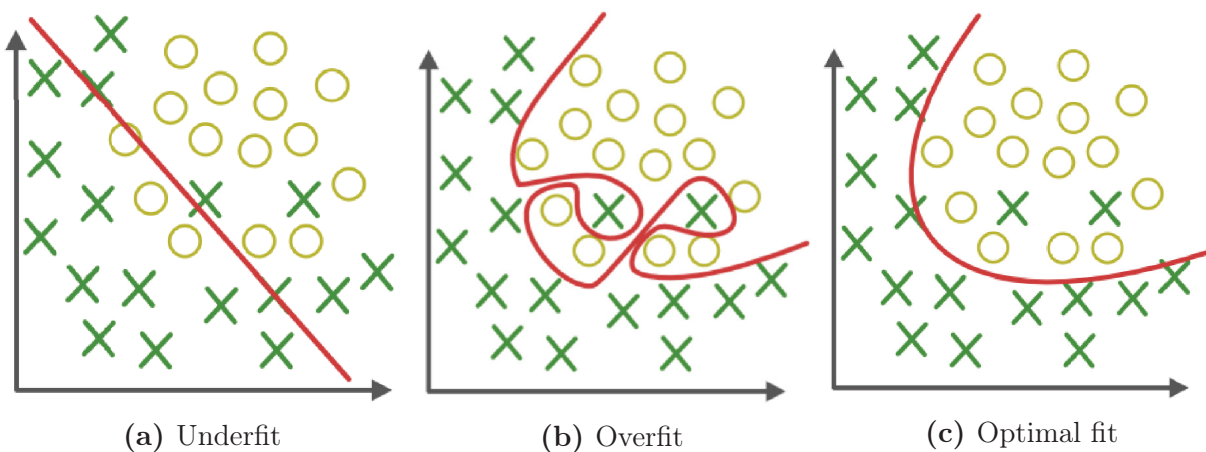


Figure 7.6: Examples of undertraining and overtraining in machine learning. Here, the BDT model (red line) tries to separate X's from O's. In (a), the model underfits the data, leading to high bias and low variance. In (b), the model overfits the data, leading to low bias and high variance. Finally, in (c), the model is balanced between bias and variance.

- **Gamma:** the minimum information gain required to cause additional splits in the decision tree. A larger gamma will lead to a more conservative algorithm, reducing the possibility of overtraining. We explicitly spell this out as “gamma” in order to avoid confusion with the  $\gamma$  symbol used to denote photons in this chapter.
- **Eta:** the BDT learning rate. A smaller value increases model complexity and training time, but may lead to overtraining if set too small.

We will discuss these parameters as they relate to the NC  $\pi^0$  analysis in Chapter 8.

## 7.4 Single-Photon Selection

In this section, we give a brief overview of the MicroBooNE single photon selection. My work is primarily focused on the NC  $\pi^0$  selection but, given that the NC  $\pi^0$  selection is used to constrain the errors on the single photon selection, it is important to review the single photon selection in order to contextualize the NC  $\pi^0$  analysis.

As discussed in Section 7.2, the single-photon selection begins with a topological selection. Here, the topologies considered are the  $1\gamma 1p$  (one shower, one track) and  $1\gamma 0p$  (one shower, zero tracks). The efficiencies of these topological selections (relative to all generated MC events) can be found in Table 7.1. These reconstructed objects are then passed through a series of conservative pre-selection cuts, such as a shower energy threshold of 40 MeV and a 5 cm fiducial volume for the reconstructed vertex. In the  $1\gamma 1p$  case, additional pre-selection cuts are placed on the track, including a track containment cut (meaning that the start and end points of the track are within 5 cm of any TPC wall). Figure 7.7 shows a sample pre-selection distribution for the  $1\gamma 1p$  selection at the pre-selection stage. Here, the  $\Delta$  invariant mass is reconstructed assuming that the event is a true  $\Delta \rightarrow N\gamma$  event (as opposed to an NC  $\pi^0$  event in which one photon was not detected). The distribution shows overall reasonable agreement between data and MC. However, the NC  $\Delta$  radiative signal (yellow) is not even visible at this stage due to the overwhelming number of background events.

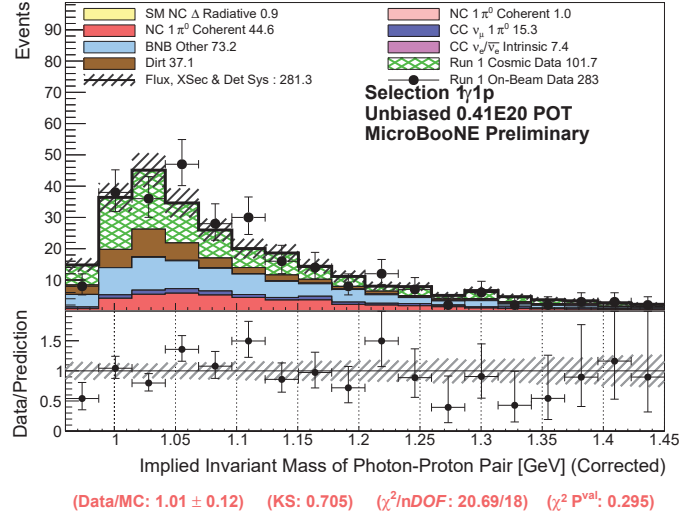


Figure 7.7: Distribution of the implied  $\Delta$  invariant mass at the pre-selection stage for the  $1\gamma 1p$  selection. The calculation of the  $\Delta$  invariant mass uses the shower energy correction described in Section 7.2.2 and the implied shower directions discussed in Section 7.2.3.

Table 7.3: BDT response cut values and signal efficiencies for the  $1\gamma$  selections. The signal efficiency is defined here as the percentage of signal events remaining after the BDT response cut relative to the number of signal events before the cut.

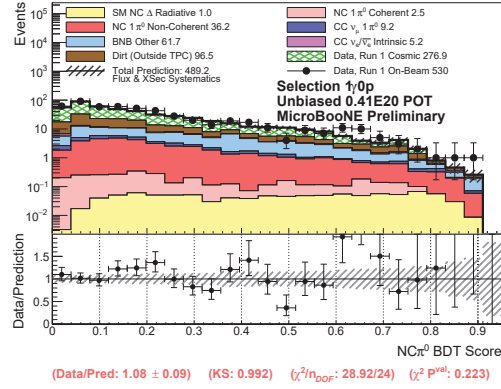
BDT	Cut Position	Signal Efficiency
$1\gamma 1p$		
NC $\pi^0$	0.467	14.6%
Cosmic	0.953	18.3%
BNB Other	0.985	12.1%
$\nu_e$	0.747	12.9%
SSV	0.709	17.2%
$1\gamma 0p$		
NC $\pi^0$	0.429	47.4%
Cosmic	0.988	55.3%
BNB Other	0.893	69.6%

The  $1\gamma 1p$  selection rejects these backgrounds by implementing five separately-trained BDTs. Each BDT targets a specific type of background event: NC  $\pi^0$ ,  $\nu_e$ , cosmic (off-beam), other BNB backgrounds, and a second shower veto (SSV). The SSV further targets NC  $\pi^0$  backgrounds by searching for shower-like clusters that were not associated with the neutrino slice during Pandora’s reconstruction. Each of these BDTs uses a set of training variables specifically tuned to reject the specific background in question. The  $1\gamma 0p$  selection, however, does not use a dedicated  $\nu_e$  or SSV BDT. The  $\nu_e$  BDT was found to be redundant, with all events rejected by the  $\nu_e$  BDT also being rejected by the BNB BDT. The  $\nu_e$  BDT was therefore removed. As for the SSV, without access to a reconstructed track, Pandora’s placement of the vertex is less precise than when a track is present. In this case, the vertex is usually placed at the shower start point. This means that the SSV has difficulty finding additional showers which point back to the interaction vertex, even when a second shower truly is present. The  $1\gamma 0p$  selection therefore only uses the NC  $\pi^0$ , cosmic, and BNB BDTs for background rejection. Figures 7.8 and 7.9 show the  $1\gamma 0p$  and  $1\gamma 1p$  BDT responses, respectively, while Table 7.3 summarizes the BDT response cut values and signal efficiencies.

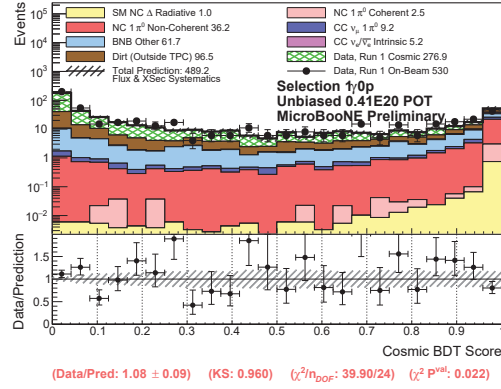
Final selection distributions for the two single-photon topologies are shown in Figure 7.10. Due to the low statistics of selected data events from the unblinded 5e19 sample,<sup>4</sup> here we show MC-only distributions scaled to  $6.91 \times 10^{20}$  POT, the total POT collected during Runs 1–3. We see that the overwhelming majority of background events are rejected by the BDTs. However, the  $\Delta$  radiative signal is still a relatively small portion of each selection. In both cases, NC  $\pi^0$  events comprise the single largest remaining background. In the  $1\gamma 1p$  selection, 87.9% of backgrounds are NC  $\pi^0$ s, while in the  $1\gamma 0p$  selection, 46.0% of the backgrounds are NC  $\pi^0$ s. Furthermore, the large correlated interaction uncertainties caused by this NC  $\pi^0$  background contribute significantly to the error bars shown in Figure 7.10. These large uncertainties in turn reduce the sensitivity of the single-photon selection to the  $\Delta$  radiative LEE signal.

---

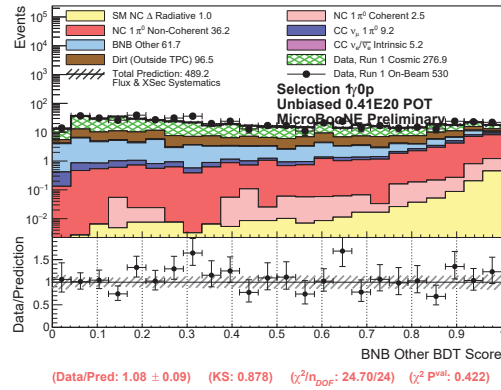
<sup>4</sup>As of this writing, only one event in the unblinded 5e19 sample passes the  $1\gamma 1p$  selection, and only seven pass the  $1\gamma 0p$ .



(a) NC  $\pi^0$



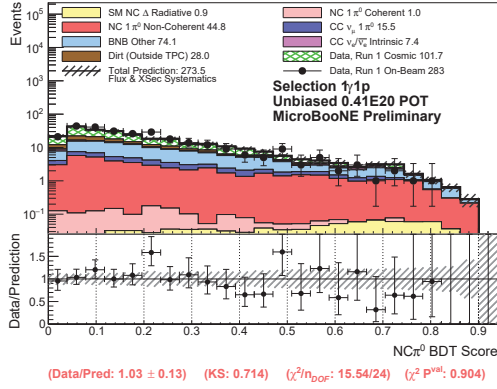
(b) Cosmic



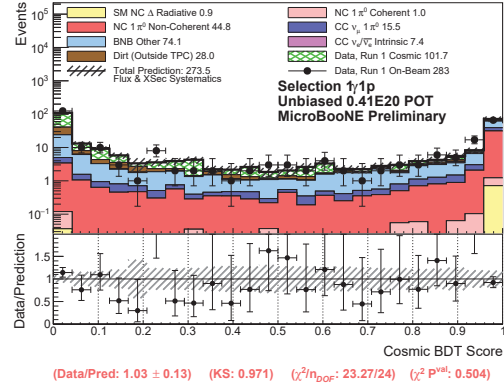
(c) BNB Other

Figure 7.8: BDT response (score) distributions for the  $1\gamma 0p$  selection.

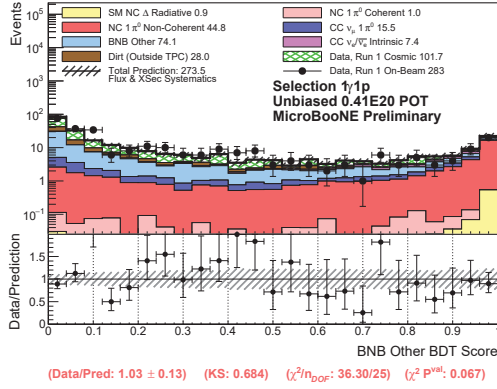




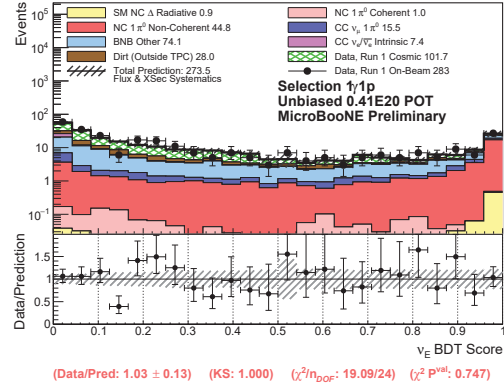
(a) NC  $\pi^0$



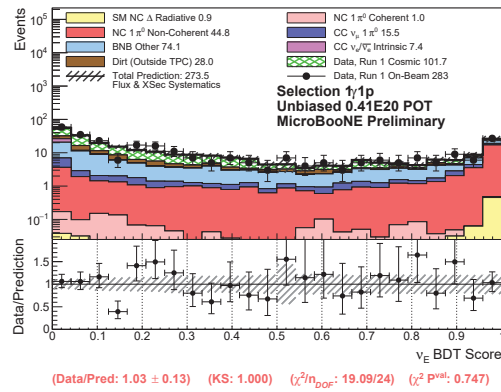
(b) Cosmic



(c) BNB Other

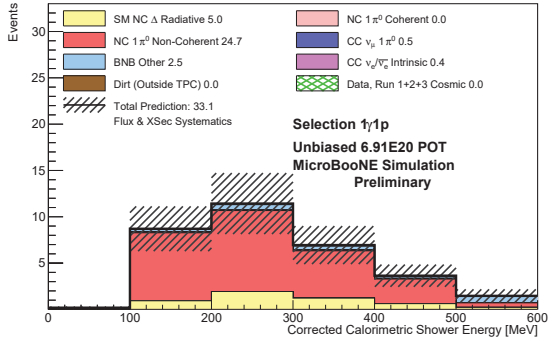


(d)  $\nu_e$

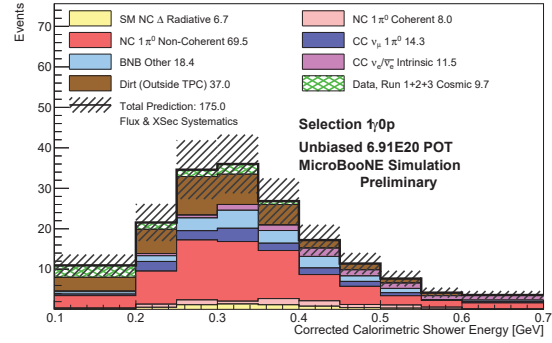


(e) SSV

Figure 7.9: BDT response (score) distributions for the 1 $\gamma$ 1p selection.



(a)  $1\gamma 1p$



(b)  $1\gamma 0p$

Figure 7.10: Final selection distributions for the  $1\gamma 1p$  and  $1\gamma 0p$  selections. Note that these are MC-only predictions scaled to  $6.91 \times 10^{20}$  POT, the total expected POT for Runs 1–3. Even after applying the background rejection BDTs, a large NC  $\pi^0$  background remains in both selections.

## Chapter 8

# Constraining the NC $\pi^0$ Background for MicroBooNE’s Single-Photon Selection

As described in Section 7.1, Neutral Current (NC)  $\pi^0$  events with a  $\pi^0$  decaying into two photons (and where one of the photons is not reconstructed) comprise the dominant background in the search for NC  $\Delta$  radiative decay (single-photon) events. Recall that the  $\Delta$  (1232 MeV) resonance has two primary decay modes:  $N\pi$ , with a branching ratio of 99.4%, and  $N\gamma$ , with a branching ratio between 0.55% and 0.65% [22]. In order to reduce the systematic uncertainty on the single-photon analysis, we adapt the single-photon analysis framework to select a sample of well-reconstructed two-EM-shower NC  $\pi^0$  events, which we subsequently use to constrain the rate of NC  $\pi^0$  misidentified backgrounds in the final single-photon sample selection. Section 8.1 describes the analysis inputs and topological selection. Section 8.2 then describes the conservative “pre-selection” cuts we apply to minimize poorly reconstructed events prior to BDT training. The BDT used in this selection is described in Section 8.3, while Section 8.4 shows the final selection obtained using this BDT. Section 8.5 discusses fits to the reconstructed  $\pi^0$  mass peak, followed by comparisons of mass peak fits for individual runs in Section 8.6. Section 8.7 discusses systematic uncertainties, while Section 8.8 shows the results of the NC  $\pi^0$  constraint. Finally, Section 8.9 provides a summary and discussion of the NC  $\pi^0$  analysis.

## 8.1 Inputs and Topological Selection

The MC and data input samples are described in Section 7.2.1 and Table 7.2. As discussed in Section 7.2.1, MicroBooNE is pursuing a blind analysis strategy for its low-energy excess (LEE) results. However, analyzers may be approved to use on-beam data filters which haven't been shown to maintain blindness to LEE-like signals. For the NC  $\pi^0$  analysis, we use the NC  $\pi^0$  filtered samples (described in Appendix C) in order to examine a larger dataset than the available open data (the “5e19” sample).

The NC  $\pi^0$  analysis framework is an adapted version of the single-photon framework. At truth level, we define our signal as a true neutral current interaction with exactly one  $\pi^0$  in the final state. We then identify neutrino interactions whose topologies are consistent with reconstructable objects resulting from NC resonant pion production. We define our signal topology as neutrino interactions that contain two reconstructed EM showers and either one or zero reconstructed particle tracks (the  $2\gamma 1p$  and  $2\gamma 0p$  topologies, respectively). Each EM shower corresponds to the pair conversion of a photon from  $\pi^0$  decay. These photons propagate invisibly through the detector and subsequently convert to  $e^+/e^-$  pairs, which manifest as showers in the detector volume. The particle track requirement is intended to include events in which a proton exits the nucleus, although even in cases where one or more protons exit the nucleus, they may not be reconstructed, either due to being too low-energy or being misreconstructed as a shower-like object. Protons tend to produce short, highly-ionizing tracks at the interaction vertex, while neutrons propagate invisibly.

Even for events with two photons, most are reconstructed with only one shower. The subleading shower is often misreconstructed or missed altogether, particularly those that are low-energy or have a high conversion distance. Low-energy showers may be misidentified as small tracks, while showers that have a large conversion distance—and are therefore far away from the vertex—may not be associated with the correct Pandora neutrino slice. While the requirement of two reconstructed showers is restrictive, requiring two showers serves two purposes: 1) it allows us to reconstruct important  $\pi^0$  kinematic variables, including the  $\pi^0$  invariant mass, momentum, and the opening angle between the two photons, and 2) it keeps the NC  $\pi^0$  selection mutually exclusive from the single photon selection.

In order to estimate the number of events that could pass the NC  $\pi^0$  selection, we begin by applying a truth-level signal definition to estimate the percentage of reconstructable events. Table 8.1 shows the signal definition for each topology, along with the passing rate and number of events (scaled to  $5.85 \times 10^{20}$  POT) that satisfy each portion of the definition. The event numbers are extracted from our combined NC  $\pi^0$  sample, which contains events generated in the TPC. We choose a 20 MeV true energy threshold for the track and showers, as objects below these thresholds are difficult to reconstruct correctly. While most showers pass the energy thresholds, fewer than 40% of true NC  $\pi^0$  events pass the  $2\gamma 1p$  proton energy threshold. We expect a neutron to exit the nucleus in about 55% of NC  $\pi^0$  events, and even in cases where a proton exits, it may not be reconstructed correctly. Of the events that pass the  $2\gamma 1p$  signal definition, 14.6% are reconstructed with the  $2\gamma 1p$  topology.<sup>1</sup>

The  $2\gamma 0p$  signal definition is identical to that of the  $2\gamma 1p$  case, except that we define the signal to have zero final-state protons with energy greater than 20 MeV. While this definition does allow for events with true low-energy protons exiting the nucleus, protons with less than 20 MeV energy will likely not be reconstructed, and thus the event would be reconstructed with a  $2\gamma 0p$  topology. Relative to the number of events that pass the  $2\gamma 0p$  signal definition shown in Table 8.1, the  $2\gamma 0p$  topological efficiency is 12.3%.

## 8.2 Pre-Selection Cuts

Following the topological selection, we apply a series of conservative pre-selection cuts. While these cuts do reduce the number of background events, their primary purpose is to prevent the BDT from training on misreconstructed events, which could bias the training. We apply the following requirements to *both* topologies:

- **Reconstructed neutrino vertex > 5 cm from any TPC wall**
- **Reconstructed leading shower energy > 30 MeV**
- **Reconstructed subleading shower energy > 20 MeV**

---

<sup>1</sup>Note that this number is higher than the one quoted for the  $2\gamma 1p$  topological reconstruction efficiency quoted in Table 7.1. This is because the number here is quoted relative to the truth-level signal definition, while the number in Table 7.1 is relative to all events.

and the following additional requirements *only to the  $2\gamma 1p$  topology*:

- **Shower conversion distances  $> 1$  cm**
- **3D Distance between the vertex and the track starting point  $< 10$  cm**

We take the vertex position cut as our definition of a fiducial volume. This reduces the number of events in which the track or one of the showers exits the TPC, while also reducing dirt and cosmic backgrounds. This fiducial volume is chosen to maintain consistency with the  $\pi^0$  filter fiducial volume cut (see Appendix C for details about the  $2\gamma$   $\pi^0$  filters). The shower energy thresholds help us avoid low-energy shower-like objects that may be poorly reconstructed. The  $2\gamma 1p$  conversion distance cut is defined as the distance between the reconstructed shower start point and the vertex. We expect photons to propagate some non-zero distance before pair converting, which distinguishes photon-like showers from electron-like showers, which tend to be attached to the vertex. Finally, for events with a proton track, we expect the vertex to be reconstructed at the track starting position; events where the vertex is more than 10 cm away from this point are likely misreconstructed. The signal efficiency for each of these cuts is listed in Table 8.2.

Figure 8.1 shows data/MC comparisons of the  $\pi^0$  invariant mass at the pre-selection stage for each topology, with the shower energy correction (discussed in Section 7.2.2) applied. Here, each sample contributing to the stacked histogram (listed in Section 7.2.1) is scaled to  $\sim 6\text{e}20$  POT to match the total on-beam data collected from the corresponding data filter applied to Runs 1–3. The hatched error bands on the stacked histogram include uncertainties on the neutrino beam flux and the cross-section uncertainties from the GENIE model set used.

Data/MC plots also list the ratio of the number of data events to the number of MC events, labelled “Data/MC”, along with the binned Kolmogorov-Smirnov (KS) test statistic [107], the  $\chi^2$  per number of degrees of freedom ( $n_{\text{DOF}}$ ), and the  $\chi^2$  probability. Note that the MC error bands include flux and cross-section systematics, but not detector systematics (this will be discussed further in Section 8.7). At the pre-selection stage, all distributions show reasonable agreement between data and MC. The signal distributions peak near the expected  $\pi^0$  mass value of 135 MeV, but the signal purity at this stage is relatively low,

Table 8.1: Truth-level signal definition cuts and expected event rates for the  $2\gamma 1p$  and  $2\gamma 0p$  signal topologies. The two signal definitions are identical except for proton energy threshold. The fractions shown here are cumulative, not individual, and are calculated relative to the generated NC  $\pi^0$  MC samples, scaled to  $5.85\text{e}20$  POT to match the filtered data POT from Runs 1–3.

Definition	Signal Fraction	No. Events in $5.85\text{e}20$ POT
True NC w/ one $\pi^0$ in final state in TPC	100.0%	11,958
True neutrino vertex in 5 cm fiducial volume	78.0%	9,327
True leading photon energy $> 20$ MeV	77.1%	9,217
True subleading photon energy $> 20$ MeV	75.5%	9,030
No. protons w/ kinetic energy $> 20$ MeV		
One ( $2\gamma 1p$ )	34.2%	4,092
Zero ( $2\gamma 0p$ )	24.7%	2,949
<b>Total</b>	<b>34.2% (<math>2\gamma 1p</math>)</b> <b>24.7% (<math>2\gamma 0p</math>)</b>	<b>9,234 (<math>2\gamma 1p</math>)</b> <b>6,654 (<math>2\gamma 0p</math>)</b>

Table 8.2: Cumulative pre-selection cut signal efficiencies, relative to the number of events that satisfy the topological definition.

Cut	$2\gamma 1p$ Signal Eff.	$2\gamma 0p$ Signal Eff.
5 cm fiducial volume	92.1%	90.5%
Leading shower energy $> 30$ MeV	91.0%	89.9%
Subleading shower energy $> 20$ MeV	82.7%	82.1%
Leading shower conversion distance $> 1$ cm	72.0%	—
Subleading shower conversion distance $> 1$ cm	65.9%	—
Distance from vertex to track start $< 10$ cm	63.6%	—
<b>Total</b>	<b>63.6%</b>	<b>82.1%</b>

roughly 20%. We also observe a greater fraction of coherent signal NC  $\pi^0$  events in the  $0p$  selection. Coherent interactions are less likely to produce an outgoing proton track and are therefore more likely to be reconstructed with the  $2\gamma 0p$  topology.

### 8.3 BNB Boosted Decision Tree

To aid in background rejection, we employ a single BDT trained on MC BNB neutrino interactions via the same framework as the single-photon analysis. As discussed in Section 7.3, this BDT uses the XGBoost implementation of the AdaBoost algorithm [106, 104]. For the NC  $\pi^0$  search, we utilize a variety of variables that aid in rejecting on-beam backgrounds whose topology can mimic that of our signal topology. The BDT takes as input MC distributions of kinematic and geometric variables for true signal and background events. As an example, Figure 8.2 shows both the MC separation and data/MC agreement for the track mean truncated  $dE/dx$ . One of the most powerful BDT variables, the track  $dE/dx$  shows both excellent separation between signal and background in MC and reasonably good data/MC agreement at the pre-selection stage (note that Figure 8.2 does not include detector systematic uncertainties, which cover the apparent shift near 2 MeV/cm; see Figure 8.10). By training on these distributions and noting the correlations between them, the BDT assigns a score—known as the *BDT response*—to each event. Signal-like events tend to accrue at higher BDT response values, while background-like events do the opposite. Figure 8.3 shows the BDT response distribution for each selection. By placing a cut on this score, we can reject the majority of background-like events while maintaining signal efficiency, the result of which is a reasonably pure selection of signal events. For the NC  $\pi^0$  selections, we choose a cut value to maximize signal efficiency times purity in the final selection. The cut value is 0.854 for the  $2\gamma 1p$  selection and 0.950 for  $2\gamma 0p$ .

For both selections, we choose training variables based on the level of data/MC agreement at the pre-selection stage and the separation power between signal and background in MC. The number of training and testing events for both selections is shown in Table 8.3. We currently utilize 10 training variables in each selection. While there is some overlap in the chosen variables between the two selections, the  $0p$  training must rely on more detailed



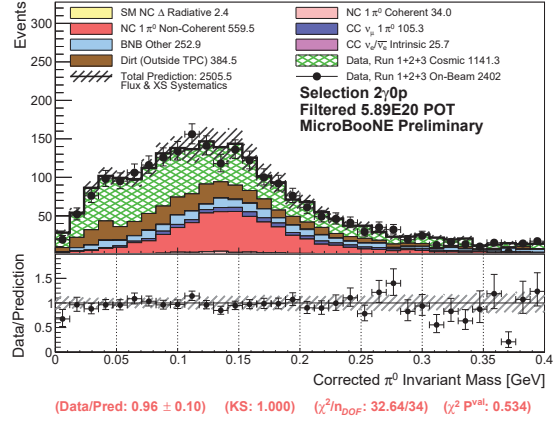
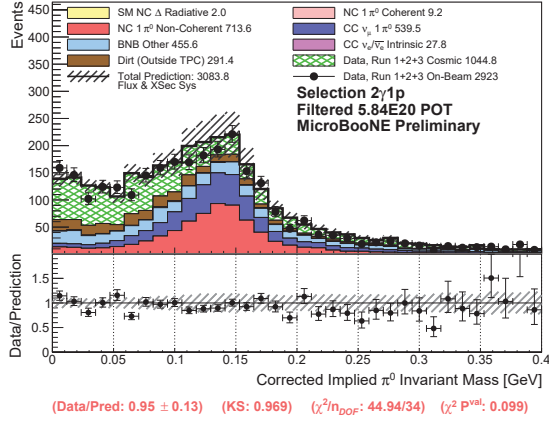


Figure 8.1: Data/MC comparisons of the reconstructed  $\pi^0$  invariant mass at the pre-selection stage for both signal topologies. The MC error bands include flux and cross-section systematics.

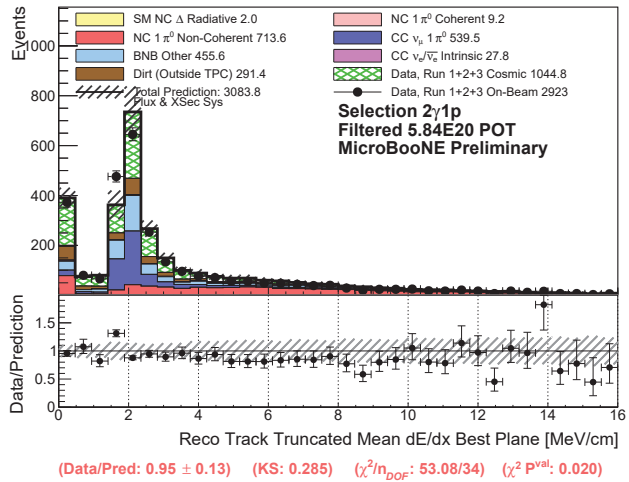
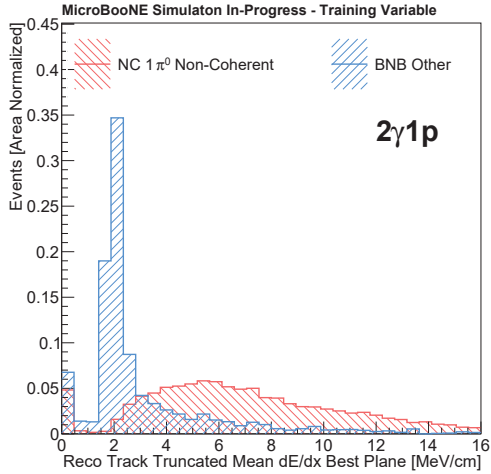


Figure 8.2: (a) MC-only distribution of track mean truncated  $dE/dx$ , separated between signal (red) and BNB backgrounds (blue). (b) Data/MC distribution of the same variable.

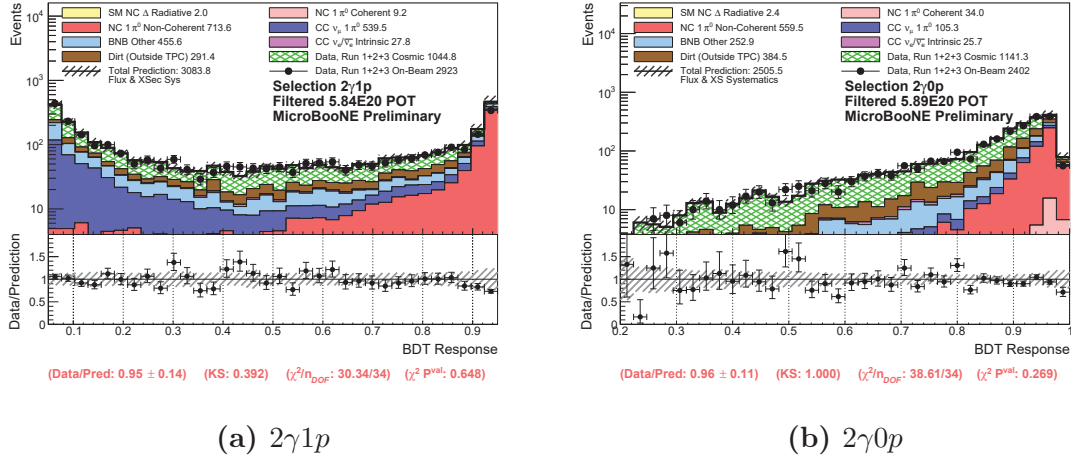


Figure 8.3: BDT response distributions for the two NC  $\pi^0$  selections. To maximize efficiency times purity in the final selection, we place a cut on the  $2\gamma 1p$  distribution at 0.854 and on the  $2\gamma 0p$  distribution at 0.950.

Table 8.3: Number of training and testing events for each selection.

Sample	Signal	Background
$2\gamma 1p$		
Training	15,725	2,187
Testing	3,149	3,016
$2\gamma 0p$		
Training	13,178	924
Testing	2,558	1,208

shower kinematics due to the absence of a reconstructed track. Table 8.4 lists the training variables for each selection and their *importance*, defined here as the relative BDT gain of the variable. Variables with a higher importance tend to reject more backgrounds compared to those with lower importance. In the  $1p$  case, the track  $dE/dx$  (Figure 8.2) is by far the most important variable due to its large separation power between signal and background. Other track-specific variables also rank highly in terms of importance. However, in the  $0p$  case, all variables have similar importance values; in this case, the BDT cuts on each variable at roughly the same rate. Data/MC distributions of all training variables at the pre-selection stage can be found in Appendix D.1 ( $2\gamma 1p$ ) and D.2 ( $2\gamma 0p$ ).

The BDT response for  $2\gamma 0p$  (Figure 8.3b) shows a significantly different shape than the  $2\gamma 1p$  response. In the  $1p$  case, the presence of a reconstructed track allows for variables such as calorimetry and track length to aid in rejecting  $CC\text{-}\pi^0$ -like events, while the  $0p$  case has to rely solely on shower information. However, photon-like showers tend to look similar across different interaction types, leading to worse separation power in the  $0p$  BDT than in the  $1p$  case. This leads to the “hill-like” structure of the  $0p$  response, where the  $1p$  response has clear peaks at high and low values. This also explains why the  $0p$  response distribution requires a higher cut value in order to achieve maximum efficiency times purity.

## 8.4 Final Selections

Final selections for each topology are obtained by placing a cut on the BDT responses shown in Figure 8.3. The events with BDT response greater than the chosen cut value comprise the final selection. After applying a cut at 0.854 in Figure 8.3a, we arrive at our final  $2\gamma 1p$  selection. The final selection is 69.9% efficient (relative to the pre-selection) and 63.5% pure in NC  $1\pi^0$  events. Data/MC comparisons of several important kinematic quantities for the final selection are shown in Figure 8.5. Each distribution that depends on reconstructed shower energy is calculated using what we call *implied* shower directions—that is, a unit vector pointing straight from the shower start point to the vertex (see Section 7.2.3 for details)—and the shower energy correction described in Section 7.2.2. Distributions which use implied shower directions are labeled as “Implied” in the x-axis label. Compared to the

Table 8.4: Training variables for each selection and their importance, where importance is defined as the relative BDT gain.

Variable	Importance
$2\gamma 1p$	
Track dE/dx	0.43
Track displacement	0.17
Ratio of track front-half to end-half dE/dx	0.11
Track $\theta_{yz}$	0.06
Leading shower energy	0.05
Leading shower impact parameter	0.04
Subleading shower conversion distance	0.04
Track end distance to wall	0.04
Leading shower conversion distance	0.03
Subleading shower impact parameter	0.03
$2\gamma 0p$	
Leading shower conversion distance	0.15
Leading shower energy	0.12
Subleading shower conversion distance	0.11
Pandora neutrino slice score	0.10
Leading shower $\theta_{yz}$	0.09
Subleading shower energy	0.09
Subleading shower ratio of length/energy	0.09
Subleading shower impact parameter	0.09
Leading shower impact parameter	0.09
Leading shower ratio of length/energy	0.08

pre-selection stage (Figure 8.1), we see an  $\sim 80\%$  reduction in the number of each category of background events, while maintaining a signal efficiency of nearly 70%.

Figure 8.5a shows the  $\pi^0$  invariant mass of the final selection. The  $2\gamma 1p$  final selection shows a  $\sim 20\%$  deficit in data relative to the MC prediction. However, this is covered by flux and cross-section uncertainties, the combination of which is roughly a 20% effect. A Gaussian-plus-linear fit to the data points (shown in Figure 8.7) gives a mean of  $138.9 \pm 2.1$  MeV with a width of  $31.7 \pm 2.4$  MeV. Figure 8.5b shows the  $\pi^0$  momentum. The reconstructed cosine of the center-of-mass (CM) decay angle—defined as the angle between the lab-frame  $\pi^0$  momentum direction and the nearest photon in the CM frame—is shown in Figure 8.5c. In theory, this quantity should give a flat distribution for signal events, but in reality we see some tapering off at high  $\cos(\theta_{cm})$  corresponding to more asymmetric  $\pi^0$  decays. When reconstructing asymmetric  $\pi^0$  decay events, we are more likely to miss the subleading photon shower due to its low energy. The opening angle between the photon showers (again, calculated using implied shower directions) is shown in Figure 8.5d. Figures 8.5e and 8.5f show the  $dE/dx$  for each shower, which is calculated using a Kalman fitter algorithm [108]. Finally, Figures 8.5g and 8.5h show the shower conversion distances. These distributions generally show the expected exponential shape, except for the first bin, where the shower conversion distance pre-selection cut of 1 cm reduces the number of events in that bin. Note that, aside from shower conversion distances, the BDT doesn't train on any of these variables.

Additionally, Figure 8.4 shows final distributions of the two track calorimetry variables used in the  $2\gamma 1p$  BDT, namely the track mean truncated  $dE/dx$  (8.4a) and the ratio of the track start half  $dE/dx$  to the track end half  $dE/dx$  (8.4b). At the final selection, the BDT cuts almost all events with a mean  $dE/dx$  near 2.0 MeV/cm. The track  $dE/dx$  ratio of the start and end half separates tracks with a clear Bragg peak—which is expected for proton tracks—from MIPs. This distribution shows a slight deficit in data (relative to the MC prediction) in the proton-like signal peak region. However, the p-value is still greater than 0.05 (although just barely), and these plots do not include detector systematics. Table 8.4 shows that these two variables provide some of the largest gain in the selection, with the mean

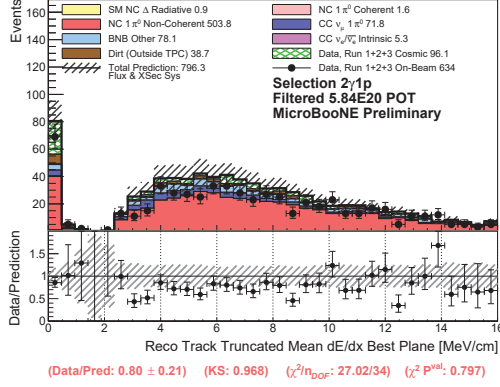
truncated  $dE/dx$  providing the largest relative gain by far. These variables demonstrate the power of track calorimetry to identify proton-like tracks.

Final selection distributions for the  $2\gamma 0p$  selection are shown in Figure 8.6. Unlike in the  $1p$  case, the data/MC normalization difference is less than 10%. The  $0p$  selection is 54.8% efficient (relative to the pre-selection) and 59.6% pure in signal events. Both the efficiency and purity are lower in this selection than in the  $2\gamma 1p$  case. This is largely due to the absence of a track which, as previously discussed, provides powerful separation between signal and background events through the track calorimetry. A Gaussian-plus-linear fit to the data points in the invariant mass distribution gives a mean of  $143.3 \pm 3.2$  MeV and a width of  $47.9 \pm 4.9$  MeV. In general, the final selection distributions show good agreement between data and MC. One notable exception is the  $2\gamma 0p$  leading shower  $dE/dx$ , which shows discrepancies around 6 MeV/cm. However, nearly all other bins are covered by systematics uncertainties, and the few bins that deviate significantly are in a region with fewer than 10 events per bin. Note that none of the distributions in Figure 8.6 are BDT training variables.

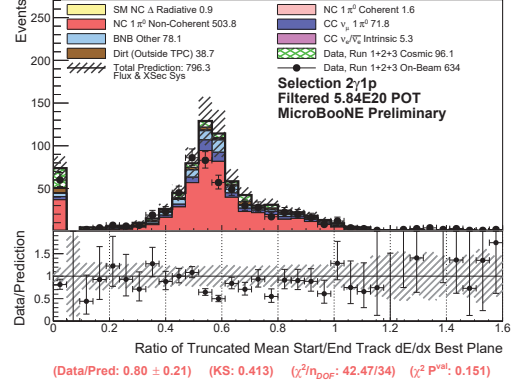
Table 8.5 breaks down the signal events in the final selections in terms of interaction type. As expected for MicroBooNE energies, the majority of selected NC  $\pi^0$  events are resonant. Deep inelastic events comprise  $\sim 10\%$  of each selection, while quasielastic and MEC events each account for  $\sim 1\%$  or less. The primary difference between the  $1p$  and  $0p$  cases is the significant increase in coherent NC  $\pi^0$  events for the  $0p$  selection. As mentioned previously, coherent events are less likely to produce a final-state proton, making them more likely to fall into the  $2\gamma 0p$  selection.

### 8.4.1 Breakdown of Final-Selection Backgrounds

While the purity of the final selection is quite good, a significant number of CC  $\pi^0$  and other BNB backgrounds survive. In order to understand why these backgrounds appear in the final selection, we break them down categorically. BNB Other backgrounds are broken down into the following categories:



(a) Track truncated mean  $dE/dx$



(b) Ratio of track Start/End  $dE/dx$

Figure 8.4: Track calorimetry distributions used in the  $2\gamma 1p$  BDT: (a) the track truncated  $dE/dx$ , and (b) the ratio of the track start half  $dE/dx$  to the track end half  $dE/dx$ . The latter provides a measure of whether a track has a Bragg peak, which is expected for proton-like tracks.

Table 8.5: Breakdown of interaction types in the NC  $\pi^0$  selections, both at the pre-selection stage and final selection.

	Resonant	DIS	QE	Coherent	MEC
$2\gamma 1p$					
Pre-Selection	81.3%	16.3%	1.3%	1.31%	0.06%
Final Selection	85.2%	13.2%	1.2%	0.28%	0.07%
$2\gamma 0p$					
Pre-Selection	79.1%	14.9%	0.52%	5.5%	0.02%
Final Selection	79.2%	13.5%	0.45%	6.8%	0.00%

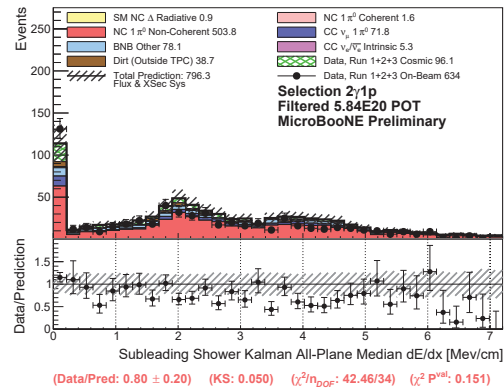
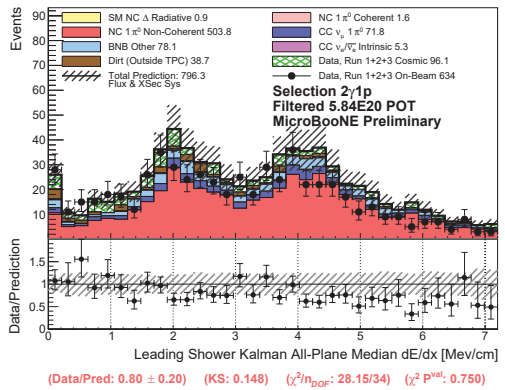
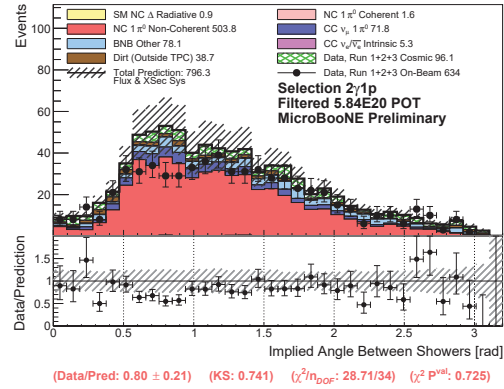
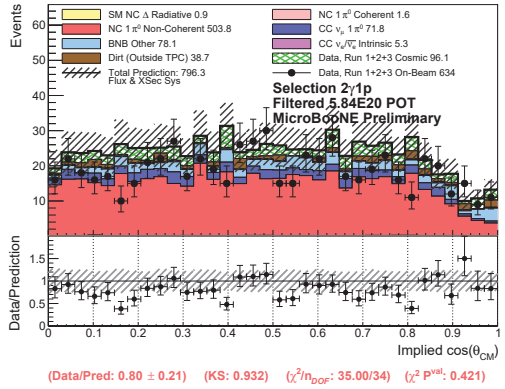
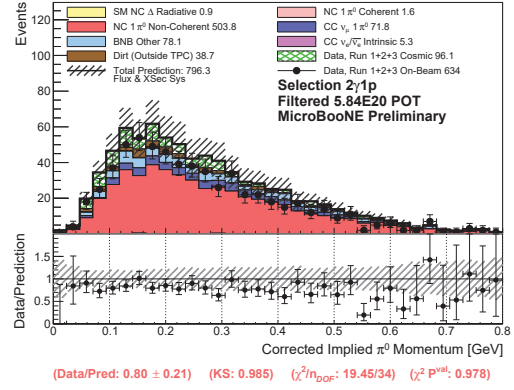
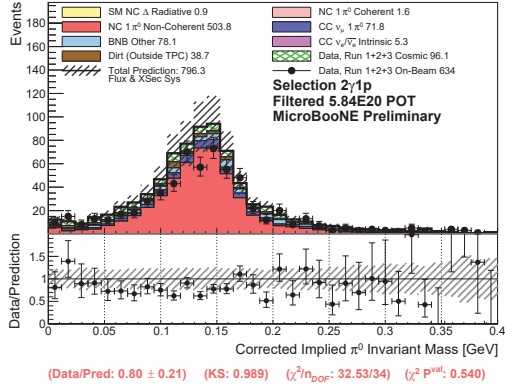


Figure 8.5: Data/MC comparisons for the  $2\gamma 1p$  final selection with flux and cross-section uncertainties.



- **$\pi^0$  Charge Exchange:** Events with at least one  $\pi^\pm$  exiting the nucleus and no exiting  $\pi^0$ , but both reconstructed showers come from photons resulting from  $\pi^0$  decay, implying a  $\pi^0$  underwent charge exchange in the nucleus
- **CC Multi- $\pi^0$ :** CC events with more than one  $\pi^0$  in the final state
- **CC Other:** CC events that don't fall into the above categories
- **NC Multi- $\pi^0$ :** NC events with more than one final-state  $\pi^0$
- **NC Other:** NC backgrounds with no final-state  $\pi^0$ s or EM showers from  $\pi^0$  decay
- **$\eta$  Decay:** Events in which the photon showers came from an  $\eta$  decay instead of a  $\pi^0$
- **Overlay:** Events in which at least one of the reconstructed objects contains 80% or more contamination from cosmic data overlays
- **Other:** Events that don't clearly fall into any of the categories above, likely due to misreconstruction.

CC  $1\pi^0$  events in the  $2\gamma 1p$  selection are categorized according to the identity of the reconstructed objects:

- **Proton track:** Events in which both showers come from photons resulting from  $\pi^0$  decay, but the reconstructed track matches to a proton
- **Muon track:** Events in which both showers come from photons resulting from  $\pi^0$  decay and the track matches to a muon, as expected for CC  $\pi^0$  events
- **Shower mis-ID:** Events in which either of the reconstructed showers is not a true shower, but a misreconstructed object instead
- **Overlay:** Same as in the BNB Other case
- **Other:** Events that don't clearly fall into any of the above categories

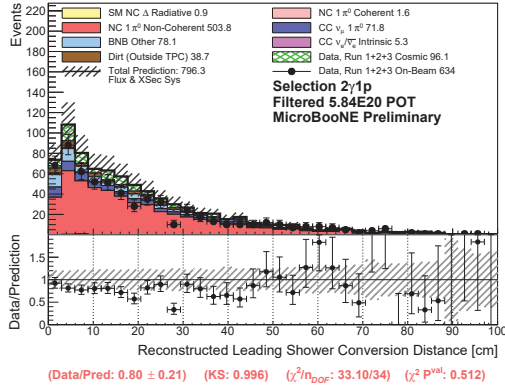
Table 8.6 shows the percentages of each category that appear in the BNB Other for both selections, and CC  $\pi^0$  backgrounds for the  $2\gamma 1p$  selection. For BNB Other backgrounds, the single largest background source in both selections is overlay contamination. We also see a significant portion of NC Multi- $\pi^0$  events, as well as “Other” events, likely either misreconstruction or some as-yet-unconsidered category.

The majority of selected CC  $\pi^0$  events in the  $1p$  selection are those in which the reconstructed track matches to a proton instead of a muon. In these cases, the muon falls outside the Pandora neutrino slice, and the remaining reconstructed objects precisely match what we expect from an NC  $\pi^0$ , making these difficult to distinguish from true NC  $\pi^0$  decay. We also see a large number of CC  $\pi^0$  in which one or both of the reconstructed shower objects is not truly an EM shower, but instead some misreconstructed object, likely a short track.

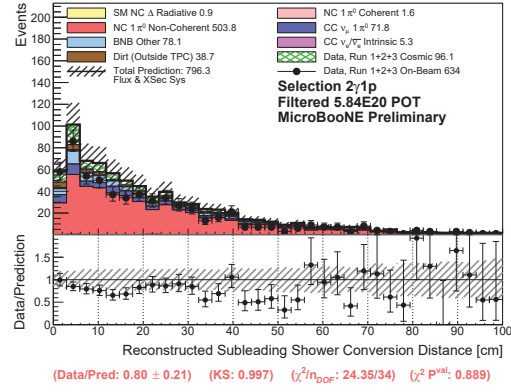
## 8.5 $\pi^0$ mass fit

Given the purity of the  $2\gamma$  selections in terms of true  $\pi^0$  events, we use the reconstructed  $\pi^0$  invariant mass distribution to cross-check our ability to accurately and precisely reconstruct calorimetric information. To evaluate the  $\pi^0$  mass reconstruction, we use a Gaussian-plus-linear fit to the selected data. The  $\pi^0$  mass distributions with this fit are shown in Figure 8.7. We chose a Gaussian-plus-linear fit functional form in order to account for non-Gaussian backgrounds in the final selection. The Gaussian and linear parameters extracted from the fits are summarized in Table 8.7. The  $\pi^0$  mass extracted from the Gaussian portion of the fit in the NC  $\pi^0$  selection is consistent with the expected value in the  $2\gamma 1p$  selection, though the  $2\gamma 0p$  value is slightly higher than expected. The fitted resolutions are 32.2% for the  $2\gamma 1p$  selection and 35.6% for the  $2\gamma 0p$  selection.

This result builds confidence that well-reconstructed  $2\gamma$  samples are available and can thus be used to constrain the dominant background to the single-photon selections when searching for a low-energy excess.



(g) continued  $\gamma_1$  Conversion Distance



(h) continued  $\gamma_2$  Conversion Distance

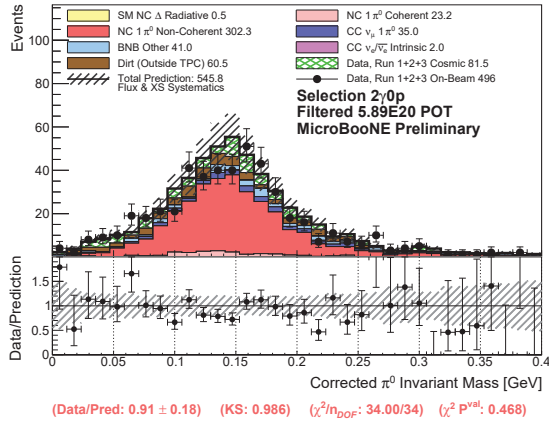
Figure 8.5 continued: Data/MC comparisons for the  $2\gamma 1p$  final selection with flux and cross-section uncertainties.

Table 8.6: Truth-level breakdown of BNB Other backgrounds in the final  $2\gamma 1p$  and  $2\gamma 0p$  selections, as well as CC  $\pi^0$  backgrounds in the final  $2\gamma 1p$  selection.

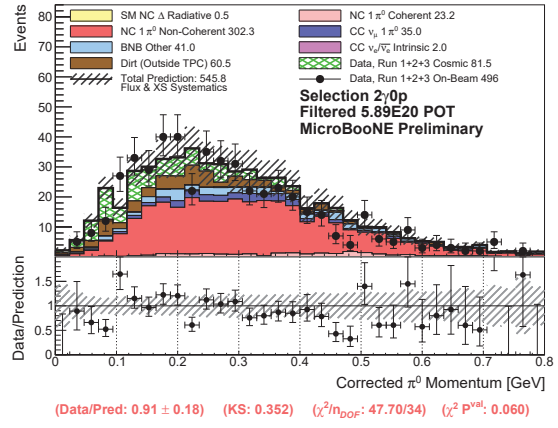
Background	$2\gamma 1p$ %	$2\gamma 0p$ %
BNB Other		
$\pi^0$ Charge Exchange	10.9	8.7
CC Multi- $\pi^0$	3.5	5.5
CC Other	16.9	13.0
NC Multi- $\pi^0$	20.7	14.5
NC Other	5.3	12.6
$\eta$	6.0	7.3
Overlay	23.9	23.3
Other	13.5	15.2
CC $\pi^0$		
Proton track	49.6	—
Muon track	11.5	—
Shower Mis-ID	31.4	—
Overlay	2.2	—
Other	5.4	—

Table 8.7: Gaussian and linear fit parameters for the  $2\gamma 1p$  and  $2\gamma 0p$  combined-run selections, using the  $2\gamma 1p$  and  $2\gamma 0p$  filtered data sets.

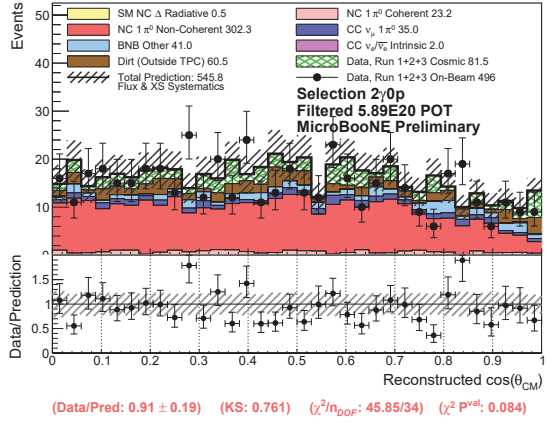
Selection	Gaussian Mean (MeV)	Gaussian Width (MeV)	Linear y-intercept	Linear slope
$2\gamma 1p$	$138.9 \pm 2.1$	$31.7 \pm 2.4$	$12.9 \pm 1.7$	$-30.6 \pm 5.1$
$2\gamma 0p$	$143.3 \pm 3.2$	$47.9 \pm 4.9$	$4.7 \pm 2.4$	$-9.2 \pm 6.7$



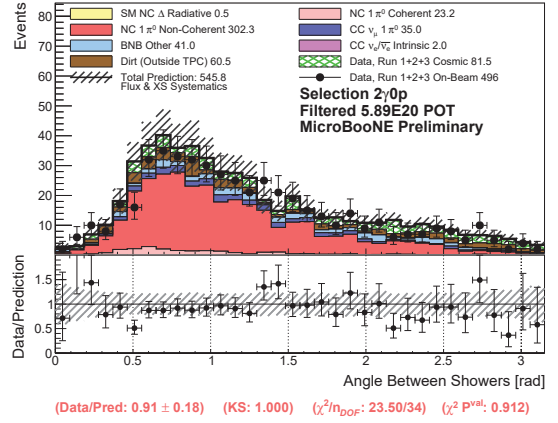
(a)  $\pi^0$  Invariant Mass



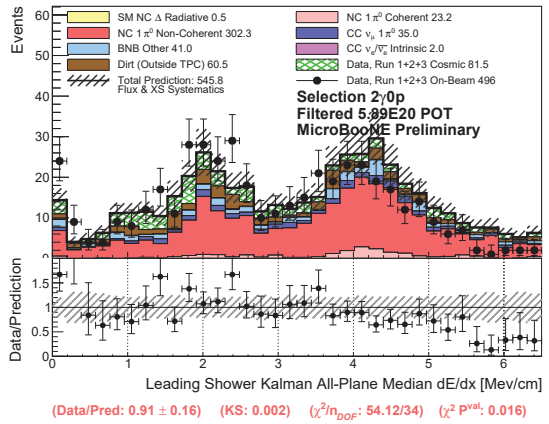
(b)  $\pi^0$  Momentum



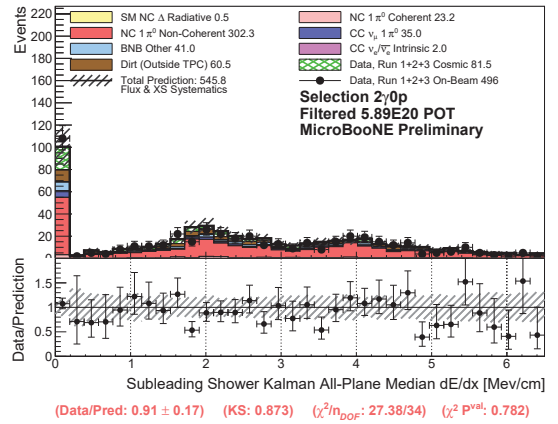
(c)  $\cos(\theta_{cm})$



(d) Angle Between Showers



(e)  $\gamma_1$  dE/dx



(f)  $\gamma_2$  dE/dx

Figure 8.6: Data/MC comparisons for the  $2\gamma 0p$  final selection.

## 8.6 Run-to-Run Comparisons

The NC  $\pi^0$  filters (see Appendix C) allow us to compare data/MC distributions between Runs 1 and 3<sup>2</sup> to ensure consistency. There are known detector response differences between Runs 1 and 3. However, those are primarily attributed to a decrease in light yield [109], which we use minimally in our analysis (light information is only used at trigger and Pandora reconstruction stage, not for topological selection stages and onward). To investigate the impact of cross-run differences on the NC  $\pi^0$  selections, we compare the reconstructed  $\pi^0$  mass distributions from Runs 1 and 3 for both the  $2\gamma 1p$  and  $2\gamma 0p$  selections. These are shown in Figures 8.8 and 8.9. Table 8.8 summarizes the results of Gaussian fits applied to each of these distributions. The extracted fit parameters for each selection are consistent (within uncertainties) across runs. We conclude that run-to-run variations are a minimal effect for this analysis.

## 8.7 Systematic Uncertainties

Systematics uncertainties for the single-photon and NC  $\pi^0$  selections are broadly divided into three categories: flux, cross-section (abbreviated “XS” in plot labels), and detector systematic uncertainties. The flux uncertainties encapsulate the uncertainties in the expected event rate due to the uncertainty on the flux of the neutrino beam. Cross section systematics include uncertainties on various parameters used in the GENIE nuclear model set. Finally, detector systematic uncertainties account for variations in the detector conditions as a function of time or location within the TPC. The specific method for evaluating each of these uncertainties differs slightly, but the basic idea is the same for all three: determine reasonable  $1\sigma$  variations for various parameters, run those variations through the analysis (or the relevant part of the analysis, at least), and see how the results are impacted.

Flux and cross-section uncertainties are said to be *reweightable*, meaning that, rather than vary each individual parameter and propagate uncertainties through the entire analysis

---

<sup>2</sup>The filtered data sets also allow us to look at a large portion of Run 2 data. However, our current MC analysis samples do not contain sufficient Run 2 MC events to allow for high-statistics comparisons. For now, we choose to focus on Runs 1 and 3.

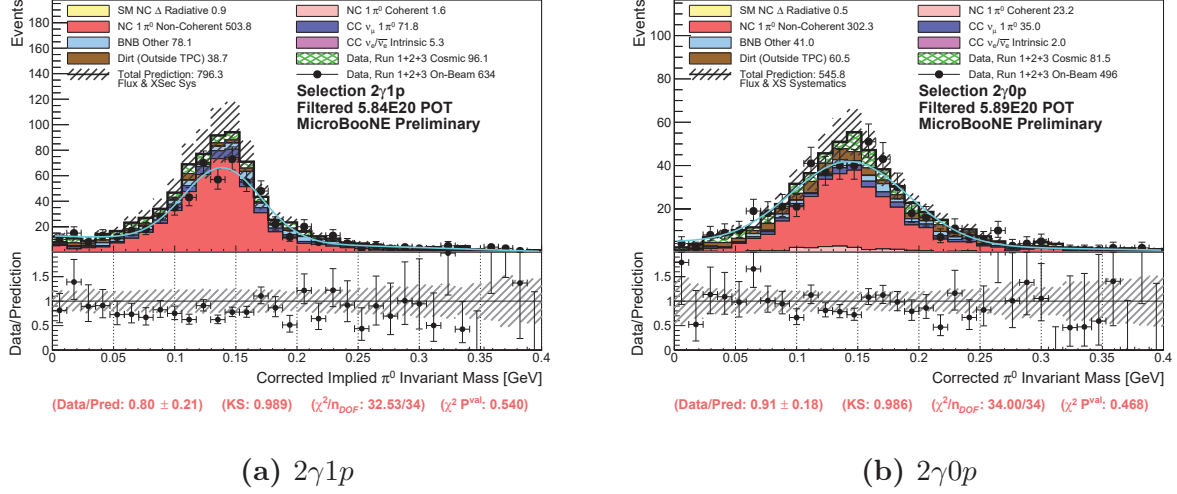
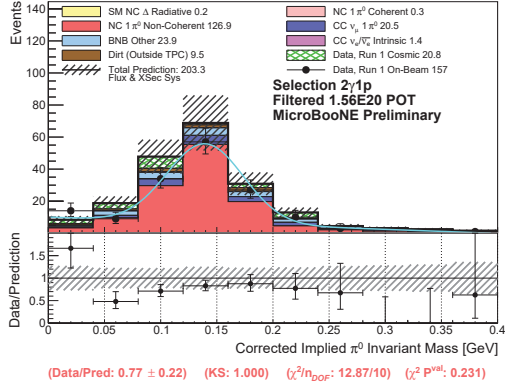


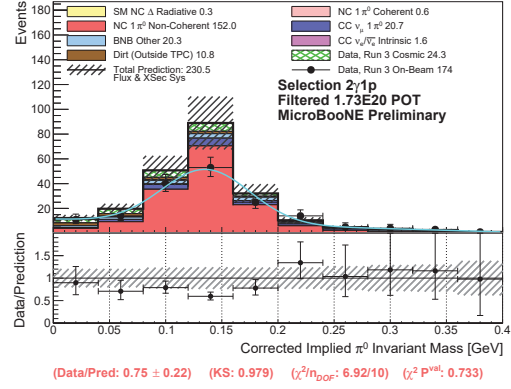
Figure 8.7: Reconstructed  $\pi^0$  mass distributions for (a)  $2\gamma 1p$  and (b)  $2\gamma 0p$  with a Gaussian-plus-linear fit to the data points (cyan line).

Table 8.8: Gaussian-plus-linear fit parameters for the  $2\gamma 1p$  and  $2\gamma 0p$  Run 1 and Run 3 selections.

	Gaussian Mean (MeV)	Gaussian Width (MeV)	Linear y-intercept	Linear Slope
<b><math>2\gamma 1p</math></b>				
Run 1	$139.5 \pm 4.2$	$34.1 \pm 4.3$	$10.1 \pm 2.2$	$-25.5 \pm 5.9$
Run 3	$135.8 \pm 5.1$	$39.4 \pm 6.1$	$11.4 \pm 3.7$	$-26.0 \pm 11.0$
<b><math>2\gamma 0p</math></b>				
Run 1	$138.6 \pm 8.5$	$42.0 \pm 9.2$	$7.7 \pm 5.0$	$-21.3 \pm 16.8$
Run 3	$132.5 \pm 6.5$	$54.1 \pm 8.9$	$-0.27 \pm 5.0$	$4.7 \pm 14.2$

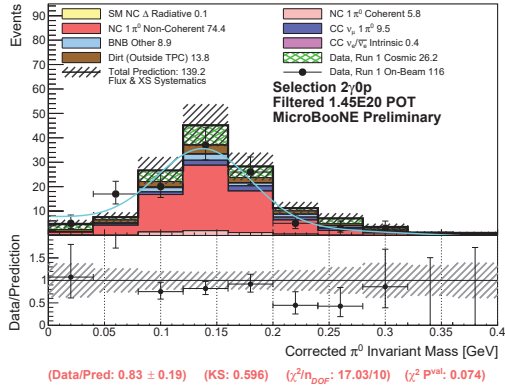


(a) Run 1

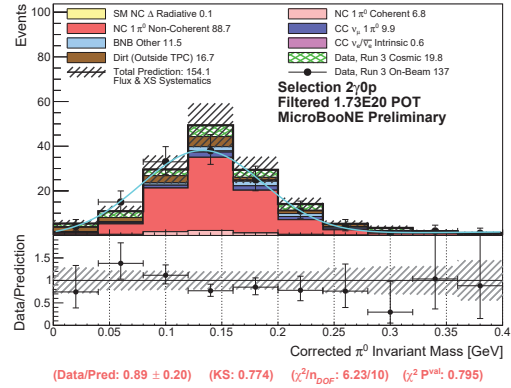


(b) Run 3

Figure 8.8: Reconstructed  $2\gamma 1p \pi^0$  invariant mass distributions for Runs 1 and 3 individually with a Gaussian-plus-linear fit, using NC  $\pi^0$  filtered data. The correction factor from Equation 7.2 is applied to each shower energy.



(a) Run 1



(b) Run 3

Figure 8.9: Reconstructed  $2\gamma 0p \pi^0$  invariant mass distributions for Runs 1 and 3 individually, using NC  $\pi^0$  filtered data. The correction factor from Equation 7.2 is applied to each shower energy.

chain, we can apply known “weights” to simulated events in order to calculate the impact of these variations. This drastically reduces computation time. For a given physics parameter  $P$ , we can vary  $P$  as

$$P \rightarrow P' \left( 1 + x_P \frac{\delta P}{P} \right) \quad (8.1)$$

where  $\delta P$  is the estimated standard deviation of  $P$  and  $x_P$  varies between  $-1$  and  $+1$ .  $x_P$  is sometimes referred to as a “knob” (or dial) that tweaks the weight of a simulated event. When this knob is set to zero,  $P = P'$ , meaning that the event weight has not changed. To calculate systematic uncertainties on reweightable parameters, we apply these weights within the known  $\pm 1\sigma$  variations and examine the impact on our analysis. Note that GENIE’s reweighting framework also accounts for correlations between different parameters. The complete list of interaction parameters and their uncertainties can be found in Reference [68].

The BNB neutrino flux has been extensively studied by both MiniBooNE [54] and MicroBooNE [110]. The primary sources of flux uncertainty are discussed in Section 4.1 and listed in Table 4.1. Sources of uncertainty in the beam flux include hadron production rates (particularly  $\pi^\pm$ ), horn current modeling, and proton delivery to the beryllium target. Hadron production uncertainties are constrained using world data [111] but are still the dominant contributor to the total uncertainty on  $\nu_\mu$  and  $\nu_e$  flux. Combined with the horn current modeling and proton delivery uncertainties (both percent-level), the total  $\nu_\mu$  ( $\nu_e$ ) flux uncertainty is 12.5% (11.7%).

As previously noted, the full list of cross-section parameters used in the GENIE model set and their uncertainties can be found in Reference [68]. For this analysis, the largest relevant uncertainty is on the neutral current resonant axial mass form factor. This form factor is used in the calculation of the NC resonance production rate in the Berger-Sehgal resonance model [16], and its uncertainty is  $\pm 20\%$ . Nuclear form factors are generally constrained using data from the Argonne National Lab (ANL) and Brookhaven National Lab (BNL) bubble chamber experiments [112, 113]. However, as we can see in the case of the NC resonant axial mass, large uncertainties remain in many cross-section parameters.



In contrast to flux and cross-section systematic uncertainties, detector systematics are not reweightable. To evaluate detector systematic uncertainties, MicroBooNE has developed a novel wire modification framework in which deconvolved waveforms are individually modified to account for differences in the pulse height and width between data and MC. By examining the differences in waveforms in data vs. MC, we obtain a ratio function that allows us to modify each bin in a deconvolved waveform. For example, we obtain an  $x$ -dependent correction function by examining reconstructed hit widths as a function of  $x$ . This correction accounts for various drift-dependent effects such as longitudinal electron diffusion and electron attenuation. We then re-run the Pandora reconstruction over these modified events to obtain a varied selection. This points to a key difference between reweightable (flux and cross section) and non-reweightable (detector) systematics: reweighting an event modifies the underlying distribution, but not the reconstruction; a reweighted event can't be added to or removed from the selection. Wire modification, however, may cause a reconstructed object to be classified differently. In other words, it's possible that, after wire modification, an object previously classified as shower-like may now be considered track-like, or vice versa. Using this framework, we create multiple detector systematic samples in MC with key detector response variables varied by  $\pm 1\sigma$ . These variables include the aforementioned  $x$  correction, a transverse ( $yz$ ) correction, a correction for the space charge effect, and PMT light yield corrections. These variations are then propagated through the selection to obtain the full detector systematic uncertainty. Due to a lack of available MC detector variation samples as of this writing, all detector systematic uncertainties are assumed to be uncorrelated.

Using these variation methods for flux, cross-section, and detector systematics, we construct a covariance matrix,  $M$ , according to

$$M_{ij} = \frac{1}{N} \sum_{n=1}^N (P_i - V_{i,n}^k)(P_j - V_{j,n}^k). \quad (8.2)$$

where  $P$  is the central value prediction,  $V$  is the varied distribution,  $k$  denotes the particular source of systematic uncertainty, and  $i$  and  $j$  are bin numbers. We consider up to  $N = 1000$  varied distributions for each source of systematic uncertainty. Figure 8.11 shows the corresponding collapsed fractional covariance and correlation matrices. Each block in the

covariance/correlation matrices corresponds to one of the four signal topologies. The strength of the resultant constraint on correlated interaction uncertainties depends on the level of correlation between the corresponding  $2\gamma$  and  $1\gamma$  samples. From Figure 8.11b, we can see that the correlation coefficients between the  $1\gamma 1p$  and  $2\gamma 1p$  selections are generally around 60–80%, while the  $0p$  correlations are around 60–70% in the most relevant bins. We therefore expect a reasonably strong constraint from the  $2\gamma$  samples. We apply this conditional constraint using a block diagonal matrix formulation,

$$Cov(X) = \Sigma = \begin{pmatrix} \Sigma_{ij}^{1\gamma,1\gamma} & \Sigma_{ib}^{1\gamma,2\gamma} \\ \Sigma_{aj}^{2\gamma,1\gamma} & \Sigma_{ab}^{2\gamma,2\gamma} \end{pmatrix} \quad (8.3)$$

where each  $\Sigma$  represents a block covariance matrix and  $i$  and  $j$  ( $a$  and  $b$ ) run from 1 to the number of  $1\gamma$  ( $2\gamma$ ) bins. Equation 8.3 shows a simplified  $2 \times 2$  version of the full  $4 \times 4$  block matrix shown in Figure 8.11. The constrained covariance matrix for the  $1\gamma 1p$  and  $1\gamma 0p$  selections is then calculated via

$$\Sigma^{1\gamma,1\gamma \text{ constrained}} = \Sigma^{1\gamma,1\gamma} - \Sigma^{1\gamma,2\gamma} (\Sigma^{2\gamma,2\gamma})^{-1} \Sigma^{2\gamma,1\gamma}. \quad (8.4)$$

The constrained error bars for the  $1\gamma$  final selections can then be obtained by taking the square root of the diagonal elements of  $\Sigma^{1\gamma,1\gamma \text{ constrained}}$ .

We perform the side-by-side fits using the final-selection distributions shown in Figure 8.12. Due to the large detector systematic uncertainties per bin, the  $1\gamma 1p$  final selection is treated here as a single-bin counting experiment. This is consistent with the overall goal of measuring the NC  $\Delta$  radiative decay rate as opposed to a full cross-section measurement.

## 8.8 Constraint Results and Sensitivity Projection

Figure 8.13 shows the reduction in the full systematic error bars in the  $1\gamma$  selection due to the  $2\gamma$  constraint. The  $1\gamma 1p$  error bar is reduced by roughly a factor of two, while the  $1\gamma 0p$  uncertainty is reduced by  $O(10)\%$  per bin. The weaker constraint in the  $0p$  sample is due to the weaker correlations between the  $1\gamma 0p$  and  $2\gamma 0p$  selections (compared to the correlations between the  $1p$  samples), as can be seen in Figure 8.11b. To illustrate the impact

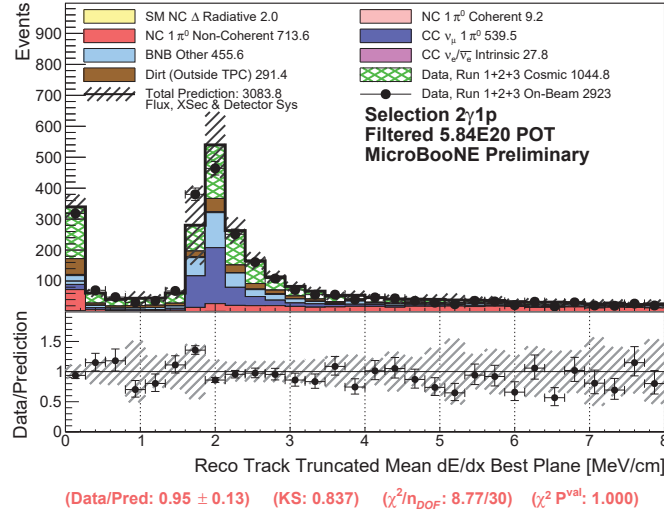


Figure 8.10: Track mean truncated  $dE/dx$  with detector systematic uncertainties. The addition of detector systematics sufficiently covers the shift in data relative to MC around 2 MeV/cm.

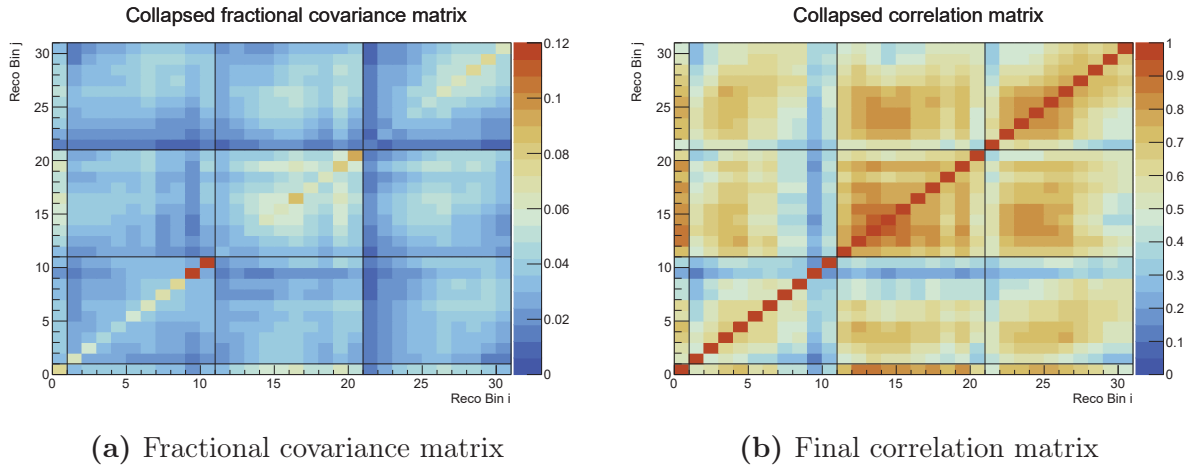


Figure 8.11: Final fractional covariance matrix (left) and correlation matrix (right) for (from left to right) the  $1\gamma 1p$ ,  $1\gamma 0p$ ,  $2\gamma 1p$  and  $2\gamma 0p$  combined fit, with full flux, cross section, and detector systematics included.

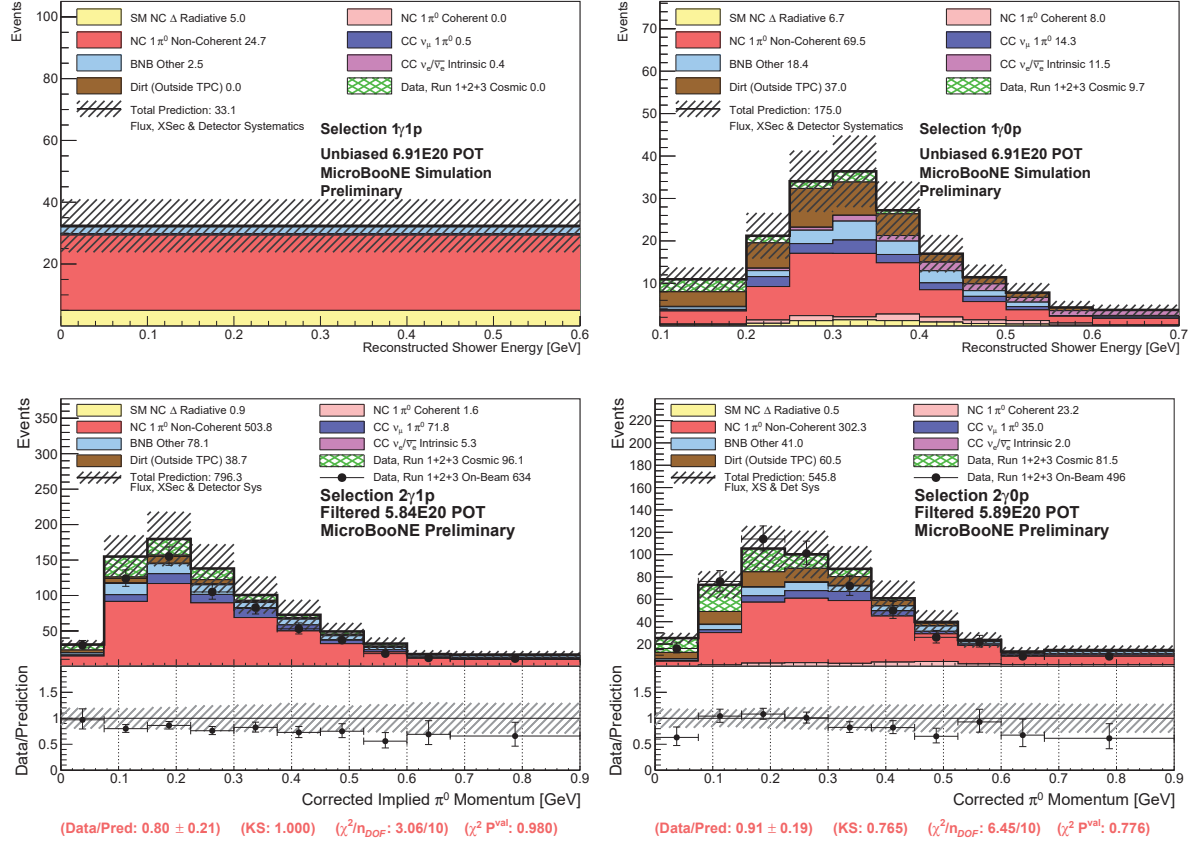


Figure 8.12: Final selection distributions for 1 $\gamma$  (top), 2 $\gamma$  (bottom), 0 $p$  (left) and 1 $p$  (right) topologies. The 1 $\gamma$  distributions show predictions scaled to  $6.9 \times 10^{20}$  POT, whereas the 2 $\gamma$  distributions correspond to the currently available filtered data (see Appendix C).

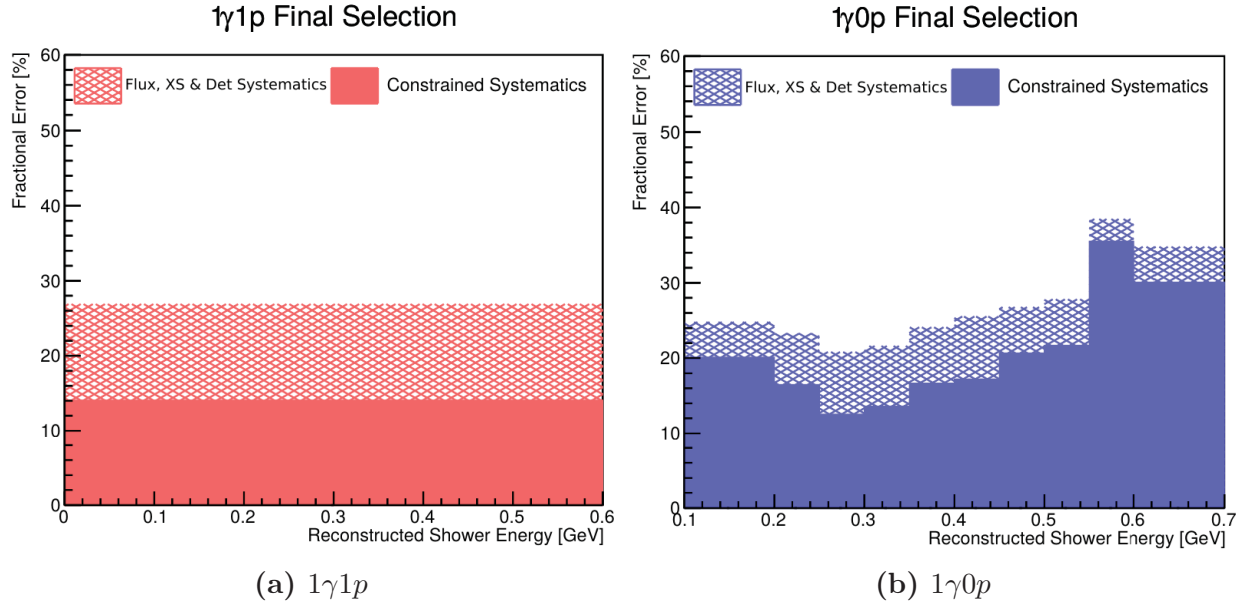


Figure 8.13: Reduction of full flux, cross-section, and detector systematic errors in the 1 $\gamma$ 1p (left) and 1 $\gamma$ 0p (right) selections due to the 2 $\gamma$  constraint.

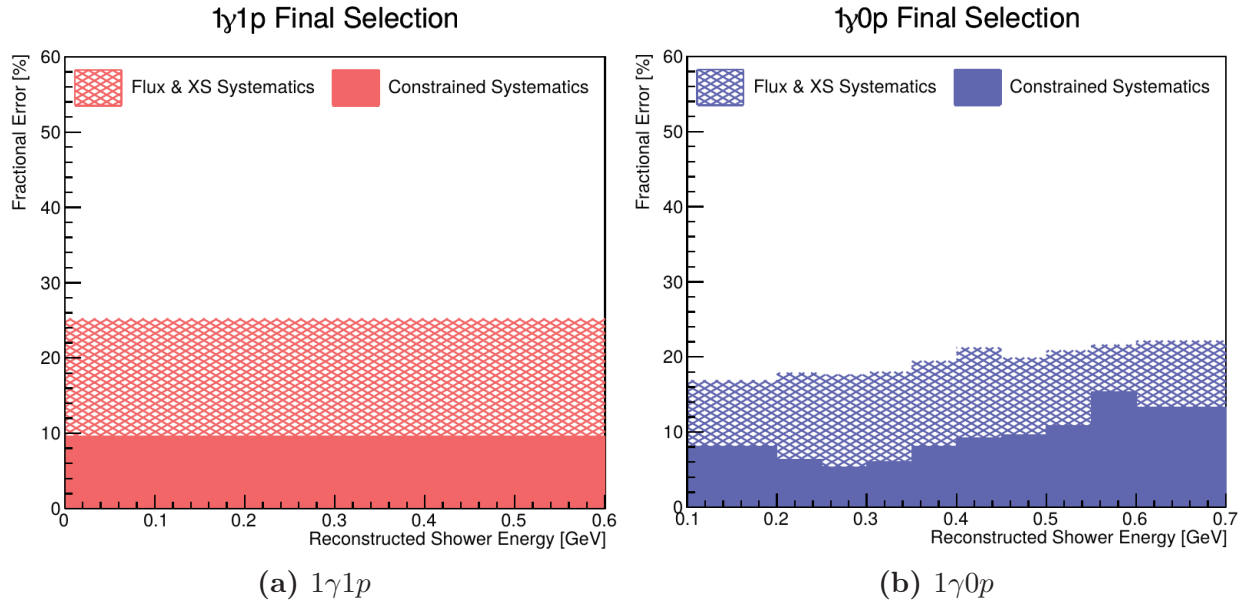


Figure 8.14: Reduction of flux and cross-section systematic errors in the 1 $\gamma$ 1p (left) and 1 $\gamma$ 0p (right) selections due to the 2 $\gamma$  constraint. In the absence of uncorrelated detector systematics, the 2 $\gamma$  constraint is significantly more powerful.

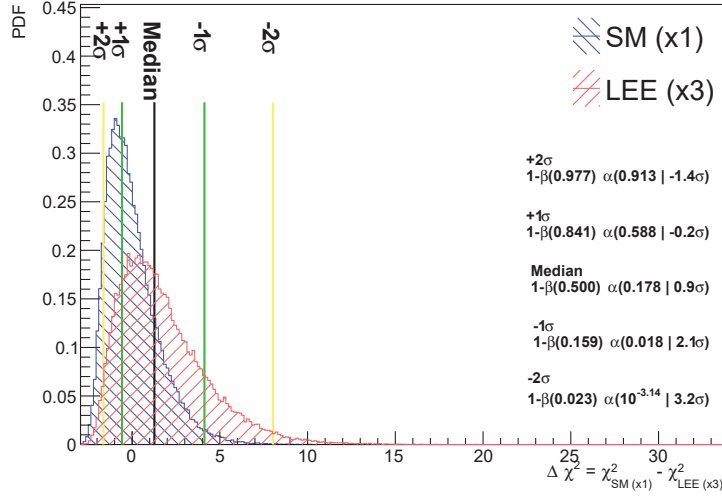
of uncorrelated detector systematics, Figure 8.14 shows the impact of the  $2\gamma$  constraint when considering only flux and cross-section uncertainties. In this case, the  $1\gamma 1p$  systematic error is reduced by nearly a factor of three, while the  $1\gamma 0p$  uncertainties see a factor of 2–3 reduction in most bins. Naturally, the inclusion of detector systematics will always increase the uncertainty, but the lack of correlations between detector systematic samples significantly reduces the power of the constraint.

As stated in previous sections, the purpose of the  $2\gamma$  selections and constraint is to improve the sensitivity of the single-photon analysis to the MiniBooNE LEE signal. We define the photon-like LEE signal as a factor-of-three enhancement to the predicted SM rate of  $\Delta \rightarrow N\gamma$  production. In this case, the null hypothesis would be a measurement consistent with the SM prediction. We then define a two-hypothesis test where  $\mathcal{H}_{\text{SM}}$  corresponds to the SM prediction and  $\mathcal{H}_{\text{LEE}}$  corresponds to a factor-of-three enhancement. These hypotheses contain no free parameters, so the sensitivity evaluation is calculated using simulated pseudo-experiments through the SBNFit framework [114]. For each hypothesis, SBNFit pulls varied data points  $D_i$  by considering the full systematic variations (Figure 8.11) and constructs a varied dataset  $D$ . We then calculate the combined Neyman-Pearson (CNP)  $\chi^2$  [115] for each hypothesis and the  $\Delta\chi^2$  between them,

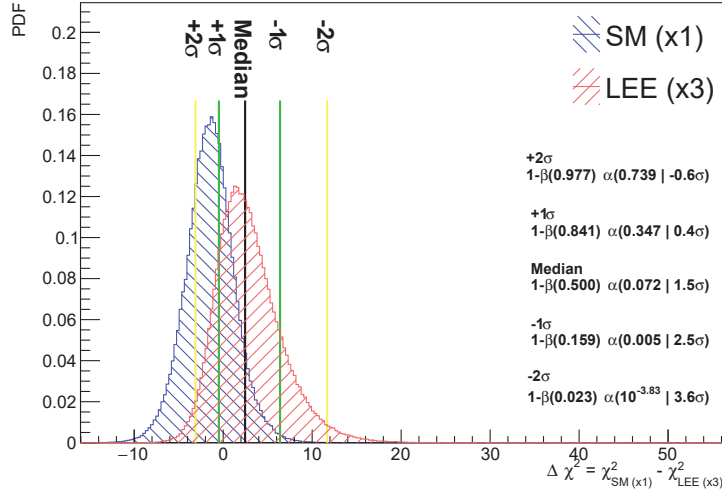
$$\Delta\chi^2 = \chi_{\text{CNP}}^2(D, H_0) - \chi_{\text{CNP}}^2(D, H_1). \quad (8.5)$$

We then plot the  $\Delta\chi^2$  probability distributions assuming each hypothesis is true and can then calculate the p-value,  $\alpha$ , of rejecting the null hypothesis. These distributions, along with the p-value calculations, are shown in Figure 8.15. Here, the SM  $\Delta$  radiative decay rate is taken as the null hypothesis (blue histograms), and we calculate the p-value based on the power of the analysis to reject this hypothesis in favor of the LEE rate (red histograms). Figure 8.15a shows these distributions before applying the  $2\gamma$  constraint, while Figure 8.15b shows the increased sensitivity after applying the NC  $\pi^0$  constraint. Before the constraint, the median sensitivity is  $0.9\sigma$ , which increases to  $1.5\sigma$  after the constraint.

While the power of the NC  $\pi^0$  constraint is apparent from Figure 8.15, the current median sensitivity calculation of  $1.5\sigma$  is too low to reliably reject the null hypothesis. As



(a) Before  $2\gamma$  constraint



(b) After  $2\gamma$  constraint

Figure 8.15: Sensitivity of the single-photon analysis to an LEE signal before and after the  $2\gamma$  constraint

of this writing, the single-photon analysis is largely limited by low data statistics. As the MicroBooNE experiment moves toward unblinding its full dataset, we expect this projected sensitivity to improve significantly.

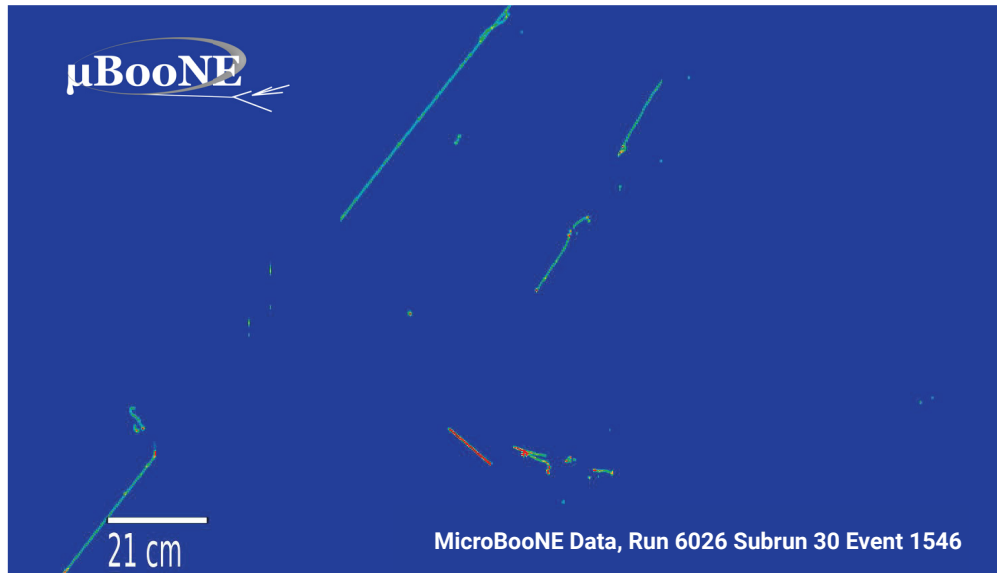
## 8.9 Summary

The selection presented in this section provides a high-statistics NC  $\pi^0$  sample that can be used to constrain  $1\gamma$  systematic uncertainties caused by the dominant NC  $\pi^0$  background. Final selected data-to-MC comparison plots show good agreement within flux and cross-section systematic uncertainties, although we observe an overall data deficit in the  $2\gamma 1p$  selection. Furthermore, the final selection distributions show the expected shape for a  $\pi^0$  sample, both in data and MC, in several reconstructed kinematic variables. The  $\pi^0$  mass peaks for both  $2\gamma 1p$  and  $2\gamma 0p$  align with the expected value of 135 MeV, and the shower conversion distances show the expected exponential shape, with photon-like showers generally having a gap between the vertex and the conversion point.

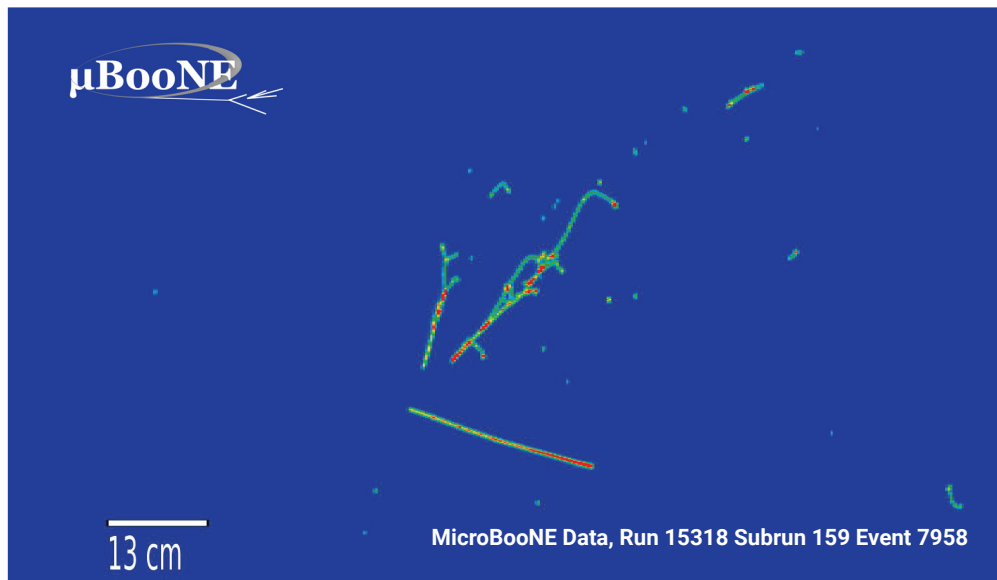
For the  $1p$  selection, track calorimetry verifies that the selected signal events have proton-like tracks, as expected for NC resonant pion production in which a proton exits the nucleus. The ratio of the track  $dE/dx$  at the start half and end half shows a deficit at the signal peak, but this is in line with the general  $2\gamma 1p$  normalization difference, which itself is covered by systematic errors. Both selections are reasonably pure in signal NC  $\pi^0$  events, with the  $2\gamma 0p$  selection being somewhat less efficient due to the less powerful signal/background separation power of the  $0p$  BDT. The clear signal and background peaks of the  $1p$  BDT response (Figure 8.3a) demonstrate the power of track calorimetry in separating events with proton-like tracks from those with muon-like tracks.

Finally, Figure 8.16 shows two event displays of selected NC  $\pi^0$  candidate events from MicroBooNE on-beam data. These events each contain one short, highly-ionizing track with two cleanly reconstructed EM showers pointing back to the interaction vertex. For the event shown in Figure 8.16a, the reconstructed leading and subleading shower energies are 332 MeV and 98 MeV, respectively, corresponding to an invariant mass of 158.2 MeV. For the event shown in Figure 8.16b, the reconstructed invariant mass is 146.2 MeV.





(a) Candidate NC  $\pi^0$  interaction from Run 1



(b) Candidate NC  $\pi^0$  interaction from Run 3

Figure 8.16: Event displays for NC  $\pi^0$  candidate events from (a) Run 1 and (b) Run 3. Both displays show the interaction as viewed by the collection plane.

# Chapter 9

## Conclusions

This thesis presents my work on two novel analyses in liquid argon time projection chambers: a measurement of the longitudinal electron diffusion coefficient,  $D_L$ , in a hundred-tonne-scale LArTPC and a data-driven rate constraint on the NC  $\pi^0$  background for the NC  $\Delta$  radiative decay search. These measurements are not only important for MicroBooNE, but will also benefit future LArTPC experiments.

During my time at Fermilab, I worked as the co-lead analyzer on The  $D_L$  measurement, which represents the first of its kind at medium electric field strength in a large-scale LArTPC. Our reported measurement of  $D_L = 3.74_{-0.29}^{+0.28}$  cm<sup>2</sup>/s is in tension with the measurement of Li et al. but is more consistent with the ICARUS data and the theory calculation of Atrazhev and Timoshkin. In this E-field region, the theory curve is ill-defined and there are few experimental measurements. This tension therefore points to the need for more measurements in this E-field region. The measurements of Li et al. and ICARUS were performed in smaller test-stand detectors in which the electric field strength could be varied, but the MicroBooNE value applies only to the nominal E-field strength of 273 V/cm. One potential avenue for a future study would be to perform measurements in MicroBooNE at various E-fields. This would require a dedicated detector run in which the cathode voltage is turned up or down, but such a run has not been performed in MicroBooNE as of this writing. Thanks to the coverage and tagging efficiency of the CRT system, it may be feasible to accrue enough high-quality cosmic muon tracks to perform the analysis after a relatively short period of running. The data collected for the measurement in this thesis were collected during a

five-month period with  $\sim 70,000$  tracks passing the selection requirements. Considering that each track can have hundreds of waveforms, a measurement could be performed with only a few hundred tracks, provided the spatial coverage is sufficient. This would provide more data at intermediate field strengths, further benefiting future LArTPC experiments.

In addition to my work on the  $D_L$  measurement, I developed a BDT-based NC  $\pi^0$  selection in MicroBooNE. The NC  $\pi^0$  selection is the largest of its kind on argon to date, with 634 on-beam data events selected in the  $2\gamma 1p$  topology with 63.5% purity and 496 events in the  $2\gamma 0p$  with 59.6% purity. The reconstructed  $\pi^0$  invariant mass distributions peak near the expected value of 135 MeV, lending credibility to the selection. This selection demonstrates the power of BDTs in rejecting backgrounds based on reconstructed quantities. BDTs are becoming increasingly popular in physics analyses, particularly for rare event searches. Furthermore, BDTs are highly configurable and relatively transparent. As machine learning algorithms become more popular, BDTs will likely become an increasingly common method of selecting signal events in various experiments.

While an NC  $\pi^0$  selection on argon is interesting in its own right, my selection is specifically designed for a data-driven rate constraint applied to the systematic uncertainty on the single photon selection. MicroBooNE's primary physics goal is to explain the MiniBooNE LEE, which is only possible with a high enough sensitivity to accept or reject the photon-like hypothesis of the LEE. The single photon analysis in MicroBooNE is searching for the  $\Delta$  radiative decay as a candidate for the photon-like hypothesis, but these events are overwhelmed by a large NC  $\pi^0$  background. The constraint provided by this selection improves the sensitivity of the single-photon analysis to the LEE signal from  $0.9\sigma$  to  $1.5\sigma$ . At present, this analysis is largely statistics-limited, so this sensitivity is expected to improve significantly as more data become available to analyzers. In addition, the NC  $\pi^0$  selection can be readily adapted to obtain valuable neutrino-argon cross section measurements. As of this writing, the NC  $\pi^0$  selection is being adapted to a flux-averaged cross section measurement, and could evolve into a differential cross section measurement in the near future.

The future of Fermilab's SBN program depends on LArTPC technology, with multiple upcoming experiments utilizing this detector technology. With the SBN program ramping up in the near future, precision measurements and constraints on systematic uncertainties

are crucial for SBN’s LArTPC experiments. The Short-Baseline Near Detector (SBND) is planned for commissioning soon, while the ICARUS T600 detector currently commissioning. In addition to the SBN LArTPC experiments, the ProtoDUNE experiment (the 400-ton DUNE prototype detector) will be entering its Phase II in 2022, and the Deep Underground Neutrino Experiment (DUNE) is planned to begin collecting its first physics data in the late 2020s. All of these experiments use LArTPC technology similar to MicroBooNE. The analyses presented in this thesis will benefit these future experiments by constraining large uncertainties. The  $D_L$  measurement points to the need for additional measurements at E-fields of  $O(100)$  V/cm, while the NC  $\pi^0$  selection can provide valuable cross-section data to constrain large interaction uncertainties on heavy nuclei. In the near future, we may have answers to such questions as the matter-antimatter asymmetry of the universe, the neutrino mass hierarchy, and potentially beyond-standard-model physics informed by neutrino experiments. It is truly an exciting time for the field of neutrino physics.

# Bibliography

- [1] Frederick Reines, CL Cowan Jr, FB Harrison, AD McGuire, and HW Kruse. Detection of the free antineutrino. *Physical Review*, 117(1):159, 1960. [4](#)
- [2] Kan Chang Wang. A suggestion on the detection of the neutrino. *Physical Review*, 61(1-2):97, 1942. [4](#)
- [3] Sheldon L Glashow. Partial-symmetries of weak interactions. *Nuclear physics*, 22(4):579–588, 1961. [5](#)
- [4] Steven Weinberg. A model of leptons. *Physical review letters*, 19(21):1264, 1967. [5](#)
- [5] A Salam. Proceedings of the 8th nobel symposium on elementary particle theory, relativistic groups and analyticity. *Stockholm, Sweden*, pages 367–377, 1969. [5](#)
- [6] James H Christenson, Jeremiah W Cronin, Val L Fitch, and René Turlay. Evidence for the  $2\pi$  decay of the  $K^0$  meson. *Physical Review Letters*, 13(4):138, 1964. [5](#)
- [7] Q Retal Ahmad, RC Allen, TC Andersen, JD Anglin, JC Barton, EW Beier, M Bercovitch, J Bigu, SD Biller, RA Black, et al. Direct evidence for neutrino flavor transformation from neutral-current interactions in the sudbury neutrino observatory. *Physical review letters*, 89(1):011301, 2002. [7](#), [20](#)
- [8] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998. [7](#), [22](#)
- [9] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. Remarks on the unified model of elementary particles. *Progress of Theoretical Physics*, 28(5):870–880, 1962. [7](#)

- [10] Bruno Pontecorvo. Neutrino experiments and the problem of conservation of leptonic charge. *Sov. Phys. JETP*, 26(984-988):165, 1968. [7](#), [20](#)
- [11] C HL Smith. Neutrino reactions at accelerator energies. Technical report, Stanford Univ., Calif., 1972. [12](#)
- [12] J. Nieves, Jose Enrique Amaro, and M. Valverde. Inclusive quasi-elastic neutrino reactions. *Phys. Rev. C*, 70:055503, 2004. [Erratum: *Phys.Rev.C* 72, 019902 (2005)]. [12](#), [15](#), [52](#)
- [13] J. Nieves, I. Ruiz Simo, and M.J. Vicente Vacas. The nucleon axial mass and the miniboone quasielastic neutrino–nucleus scattering problem. *Physics Letters B*, 707(1):72–75, 2012. [12](#)
- [14] Alexis A Aguilar-Arevalo, CE Anderson, AO Bazarko, SJ Brice, BC Brown, L Bugel, J Cao, L Coney, JM Conrad, DC Cox, et al. Measurement of  $\nu \mu$  and  $\nu^- \mu$  induced neutral current single  $\pi^0$  production cross sections on mineral oil at e  $\nu$  o (1 gev). *Physical Review D*, 81(1):013005, 2010. [14](#), [98](#)
- [15] Dieter Rein and Lalit M. Sehgal. Neutrino Excitation of Baryon Resonances and Single Pion Production. *Annals Phys.*, 133:79–153, 1981. [14](#), [52](#)
- [16] Ch Berger and LM Sehgal. Lepton mass effects in single pion production by neutrinos. *Physical Review D*, 76(11):113004, 2007. [14](#), [52](#), [136](#)
- [17] C. Andreopoulos, A. Bell, D. Bhattacharya, F. Cavanna, J. Dobson, S. Dytman, H. Gallagher, P. Guzowski, R. Hatcher, P. Kehayias, A. Meregaglia, D. Naples, G. Pearce, A. Rubbia, M. Whalley, and T. Yang. The genie neutrino monte carlo generator. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 614(1):87–104, 2010. [14](#)
- [18] Teppei Katori. Meson exchange current (mec) models in neutrino interaction generators, 2013. [15](#)

- [19] P Abratenko, C Adams, M Alrashed, R An, J Anthony, J Asaadi, A Ashkenazi, M Auger, S Balasubramanian, B Baller, et al. First measurement of inclusive muon neutrino charged current differential cross sections on argon at  $e \nu$  0.8 gev with the microboone detector. *Physical review letters*, 123(13):131801, 2019. [15](#)
- [20] Kendall Mahn, Chris Marshall, and Callum Wilkinson. Progress in measurements of 0.1–10 gev neutrino–nucleus scattering and anticipated results from future experiments. *Annual Review of Nuclear and Particle Science*, 68:105–129, 2018. [15](#)
- [21] Joseph A Formaggio and GP Zeller. From ev to eev: Neutrino cross sections across energy scales. *Reviews of Modern Physics*, 84(3):1307, 2012. [15](#)
- [22] P.A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020. [18](#), [19](#), [67](#), [96](#), [115](#)
- [23] D. Q. Adams et al. Improved Limit on Neutrinoless Double-Beta Decay in  $^{130}\text{Te}$  with CUORE. *Phys. Rev. Lett.*, 124(12):122501, 2020. [16](#)
- [24] Andrea Pocar, nEXO Collaboration, et al. The nexo detector: design overview. In *Journal of Physics: Conference Series*, volume 1468, page 012131. IOP Publishing, 2020. [16](#)
- [25] Max Aker, K Altenmüller, M Arenz, M Babutzka, J Barrett, S Bauer, M Beck, A Beglarian, J Behrens, T Bergmann, et al. Improved upper limit on the neutrino mass from a direct kinematic method by katrin. *Physical review letters*, 123(22):221802, 2019. [16](#)
- [26] Lincoln Wolfenstein. Neutrino oscillations in matter. *Physical Review D*, 17(9):2369, 1978. [16](#)
- [27] SP Mikheev and A Yu Smirnov. Resonance enhancement of oscillations in matter and solar neutrino spectroscopy. *Soviet Journal of Nuclear Physics*, 42(6):913–917, 1985. [16](#)

- [28] SP Mikheyev and A Yu Smirnov. Resonant amplification of  $\nu$  oscillations in matter and solar-neutrino spectroscopy. *Il Nuovo Cimento C*, 9(1):17–26, 1986. [16](#)
- [29] X Qian and P Vogel. Neutrino mass hierarchy. *Progress in Particle and Nuclear Physics*, 83:1–30, 2015. [16](#)
- [30] P Adamson, L Aliaga, D Ambrose, N Anfimov, A Antoshkin, E Arrieta-Diaz, K Augsten, A Aurisano, C Backhouse, M Baird, et al. Measurement of the neutrino mixing angle  $\theta_{23}$  in nova. *Physical review letters*, 118(15):151802, 2017. [17](#)
- [31] AE Cárcamo Hernández, Sergey Kovalenko, José WF Valle, and Carlos A Vaquera-Araujo. Neutrino predictions from a left-right symmetric flavored extension of the standard model. *Journal of High Energy Physics*, 2019(2):1–24, 2019. [17](#)
- [32] Ivan Esteban, MC Gonzalez-Garcia, Michele Maltoni, Thomas Schwetz, and Albert Zhou. The fate of hints: updated global analysis of three-flavor neutrino oscillations. *Journal of High Energy Physics*, 2020(9):1–22, 2020. [19](#), [21](#), [25](#)
- [33] PF de Salas, DV Forero, S Gariazzo, P Martínez-Miravé, O Mena, CA Ternes, M Tórtola, and JWF Valle. 2020 global reassessment of the neutrino oscillation picture. *arXiv preprint arXiv:2006.11237*, 2020. [19](#), [21](#), [23](#)
- [34] Raymond Davis. Solar neutrinos. ii. experimental. *Phys. Rev. Lett.*, 12:303–305, Mar 1964. [20](#)
- [35] KS Hirata, K Inoue, T Ishida, T Kajita, K Kihara, M Nakahata, K Nakamura, S Ohara, N Sato, Y Suzuki, et al. Real-time, directional measurement of  $\sim 8$  solar neutrinos in the kamiokande ii detector. *Physical Review D*, 44(8):2241, 1991. [20](#)
- [36] Wolfgang Hampel, J Handt, G Heusser, J Kiko, T Kirsten, M Laubenstein, E Pernicka, W Rau, M Wojcik, Yu Zakharov, et al. Gallex solar neutrino observations: Results for gallex iv. *Physics Letters B*, 447(1-2):127–133, 1999. [20](#)
- [37] G. Mention, M. Fechner, Th. Lasserre, Th. A. Mueller, D. Lhuillier, M. Cribier, and A. Letourneau. Reactor antineutrino anomaly. *Phys. Rev. D*, 83:073006, Apr 2011. [21](#)



- [38] KS Hirata, T Kajita, M Koshiba, M Nakahata, S Ohara, Y Oyama, N Sato, A Suzuki, M Takita, Y Totsuka, et al. Experimental study of the atmospheric neutrino flux. *Physics Letters B*, 205(2-3):416–420, 1988. [22](#)
- [39] D Casper, R Becker-Szendy, CB Bratton, DR Cady, R Claus, ST Dye, W Gajewski, M Goldhaber, TJ Haines, PG Halverson, et al. Measurement of atmospheric neutrino composition with the imb-3 detector. *Physical Review Letters*, 66(20):2561, 1991. [22](#)
- [40] WWM Allison, GJ Alner, DS Ayres, WL Barrett, C Bode, PM Border, CB Brooks, JH Cobb, DJA Cockerill, RJ Cotton, et al. Measurement of the atmospheric neutrino flavour composition in soudan 2. *Physics Letters B*, 391(3-4):491–500, 1997. [22](#)
- [41] K. Abe et al. Atmospheric neutrino oscillation analysis with external constraints in super-kamiokande i-iv. *Phys. Rev. D*, 97:072001, Apr 2018. [22](#)
- [42] M. G. Aartsen et al. Measurement of atmospheric neutrino oscillations at 6–56 gev with icecube deepcore. *Phys. Rev. Lett.*, 120:071801, Feb 2018. [22](#)
- [43] M. G. Aartsen et al. Measurement of atmospheric tau neutrino appearance with icecube deepcore. *Phys. Rev. D*, 99:032007, Feb 2019. [22](#)
- [44] P Adamson, K Anderson, M Andrews, R Andrews, I Anghel, D Augustine, A Aurisano, S Avvakumov, DS Ayres, B Baller, et al. The numi neutrino beam. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 806:279–306, 2016. [23](#)
- [45] M. A. Acero et al. First measurement of neutrino oscillation parameters using neutrinos and antineutrinos by nova. *Phys. Rev. Lett.*, 123:151803, Oct 2019. [23](#)
- [46] K. Abe et al. Search for electron antineutrino appearance in a long-baseline muon antineutrino beam. *Phys. Rev. Lett.*, 124:161802, Apr 2020. [23](#)
- [47] P. Adamson et al. Combined analysis of  $\nu_\mu$  disappearance and  $\nu_\mu \rightarrow \nu_e$  appearance in minos using accelerator and atmospheric neutrinos. *Phys. Rev. Lett.*, 112:191801, May 2014. [23](#)

- [48] M. H. Ahn et al. Measurement of neutrino oscillation by the k2k experiment. *Phys. Rev. D*, 74:072003, Oct 2006. [23](#)
- [49] Ivan Esteban, MC Gonzalez-Garcia, Michele Maltoni, Ivan Martinez-Soler, and Thomas Schwetz. Updated fit to three neutrino mixing: exploring the accelerator-reactor complementarity. *Journal of High Energy Physics*, 2017(1):87, 2017. [24](#)
- [50] A Aguilar, LB Auerbach, RL Burman, DO Caldwell, ED Church, AK Cochran, JB Donahue, A Fazely, GT Garvey, RM Gunasingha, et al. Evidence for neutrino oscillations from the observation of  $\nu$  e appearance in a  $\nu$   $\mu$  beam. *Physical Review D*, 64(11):112007, 2001. [24](#)
- [51] AA Aguilar-Arevalo, BC Brown, JM Conrad, R Dharmapalan, A Diaz, Z Djurcic, DA Finley, R Ford, GT Garvey, S Gollapinni, et al. Updated miniboone neutrino oscillation results with increased data and new background studies. *arXiv preprint arXiv:2006.16883*, 2020. [26](#)
- [52] R Acciarri, C Adams, R An, C Andreopoulos, AM Ankowski, M Antonello, J Asaadi, W Badgett, L Bagby, B Baibussinov, et al. A proposal for a three detector short-baseline neutrino oscillation program in the fermilab booster neutrino beam. *arXiv preprint arXiv:1503.01520*, 2015. [30](#)
- [53] P Abratenko, M Alrashed, R An, J Anthony, J Asaadi, A Ashkenazi, S Balasubramanian, B Baller, C Barnes, G Barr, et al. Measurement of the flux-averaged inclusive charged-current electron neutrino and antineutrino cross section on argon using the numi beam and the microboone detector. *arXiv preprint arXiv:2101.04228*, 2021. [31](#), [64](#), [67](#)
- [54] AA Aguilar-Arevalo, CE Anderson, AO Bazarko, Stephen J Brice, BC Brown, L Bugel, J Cao, L Coney, JM Conrad, DC Cox, et al. Neutrino flux prediction at miniboone. *Physical Review D*, 79(7):072002, 2009. [31](#), [32](#), [33](#), [136](#)

- [55] R Acciarri, C Adams, R An, C Andreopoulos, AM Ankowski, M Antonello, J Asaadi, W Badgett, L Bagby, B Baibussinov, et al. A proposal for a three detector short-baseline neutrino oscillation program in the fermilab booster neutrino beam. *arXiv preprint arXiv:1503.01520*, 2015. [34](#)
- [56] The MicroBooNE Collaboration. MICROBOONE-NOTE-1031-PUB: Booster Neutrino Flux Prediction at MicroBooNE. <https://microboone.fnal.gov/wp-content/uploads/MICROBOONE-NOTE-1031-PUB.pdf>, 2018. [32](#)
- [57] R Acciarri, C Adams, R An, A Aparicio, S Aponte, J Asaadi, M Auger, N Ayoub, L Bagby, B Baller, et al. Design and construction of the microboone detector. *Journal of Instrumentation*, 12(02):P02017, 2017. [38](#), [43](#)
- [58] R. Acciarri et al. Noise Characterization and Filtering in the MicroBooNE Liquid Argon TPC. *JINST*, 12(08):P08003, 2017. [38](#), [39](#), [41](#), [75](#), [78](#)
- [59] C Adams, R An, J Anthony, J Asaadi, M Auger, L Bagby, S Balasubramanian, B Baller, C Barnes, G Barr, et al. Ionization electron signal processing in single phase lartpcs. part i. algorithm description and quantitative evaluation with microboone simulation. *Journal of Instrumentation*, 13(07):P07006, 2018. [40](#), [53](#), [73](#), [74](#), [78](#), [82](#), [88](#)
- [60] P Abratenko, M Alrashed, R An, J Anthony, J Asaadi, A Ashkenazi, S Balasubramanian, B Baller, C Barnes, G Barr, et al. Measurement of space charge effects in the microboone lartpc using cosmic muons. *arXiv preprint arXiv:2008.09765*, 2020. [40](#), [73](#), [87](#), [88](#)
- [61] Varuna Meddage. Electron attenuation measurement using cosmic ray muons at the microboone lartpc. *arXiv preprint arXiv:1710.00396*, 2017. [42](#), [91](#)
- [62] Benjamin J. P. Jones. *Sterile Neutrinos in Cold Climates*. PhD thesis, MIT, 2015. [42](#)
- [63] Adam Lister. *Constraint of Systematic Uncertainties in an Electron Neutrino Search Using Muon Neutrinos at MicroBooNE*. PhD thesis, Lancaster U., 2019. [43](#), [66](#)

- [64] C Adams, M Alrashed, R An, J Anthony, J Asaadi, A Ashkenazi, S Balasubramanian, B Baller, C Barnes, G Barr, 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. *Journal of Instrumentation*, 15(07):P07010, 2020. [44](#), [62](#)
- [65] C Adams, M Alrashed, R An, J Anthony, J Asaadi, A Ashkenazi, M Auger, S Balasubramanian, B Baller, C Barnes, et al. Design and construction of the microboone cosmic ray tagger system. *Journal of Instrumentation*, 14(04):P04004, 2019. [44](#)
- [66] E.L. Snider and G. Petrillo. LArSoft: toolkit for simulation, reconstruction and analysis of liquid argon TPC neutrino detectors. *Journal of Physics: Conference Series*, 898:042057, oct 2017. [51](#)
- [67] Sea Agostinelli, John Allison, K al Amako, John Apostolakis, H Araujo, P Arce, M Asai, D Axen, S Banerjee, G 2 Barrand, et al. Geant4—a simulation toolkit. *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003. [51](#), [52](#)
- [68] Costas Andreopoulos, Christopher Barry, Steve Dytman, Hugh Gallagher, Tomasz Golan, Robert Hatcher, Gabriel Perdue, and Julia Yarba. The genie neutrino monte carlo generator: physics and user manual. *arXiv preprint arXiv:1510.05494*, 2015. [51](#), [136](#)
- [69] R.C. Carrasco and E. Oset. Interaction of Real Photons With Nuclei From 100-MeV to 500-MeV. *Nucl. Phys. A*, 536:445–508, 1992. [52](#)
- [70] Arie Bodek and UK Yang. Modeling deep inelastic cross sections in the few gev region. *Nuclear Physics B-Proceedings Supplements*, 112(1-3):70–76, 2002. [52](#)
- [71] Z. Koba, H.B. Nielsen, and P. Olesen. Scaling of multiplicity distributions in high energy hadron collisions. *Nuclear Physics B*, 40:317–334, 1972. [52](#)

- [72] T Yang, C Andreopoulos, H Gallagher, K Hofmann, and P Kehayias. A hadronization model for few-gev neutrino interactions. *The European Physical Journal C*, 63(1):1–10, 2009. [52](#)
- [73] A. Papadopolou, A. Ashkenazi, S. Gardiner, M. Betancourt, S. Dytman, L. B. Weinstein, E. Piasetzky, F. Hauenstein, M. Khachatryan, S. Dolan, G. Megias, and O. Hen. Inclusive electron scattering and the genie neutrino event generator, 2020. [52](#)
- [74] C Adams, R An, J Anthony, J Asaadi, M Auger, S Balasubramanian, B Baller, C Barnes, G Barr, M Bass, et al. Ionization electron signal processing in single phase lartpcs. part ii. data/simulation comparison and performance in microboone. *Journal of Instrumentation*, 13(07):P07007, 2018. [53](#), [75](#)
- [75] S. Ramo. Currents induced by electron motion. *Proceedings of the IRE*, 27(9):584–585, 1939. [53](#)
- [76] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. MIT press, 1950. [53](#)
- [77] J. S. Marshall and M. A. Thomson. The Pandora Software Development Kit for Pattern Recognition. *Eur. Phys. J.*, C75(9):439, 2015. [56](#)
- [78] MicroBooNE Collaboration, C Adams, M Alrashed, R An, J Anthony, J Asaadi, A Ashkenazi, M Auger, S Balasubramanian, B Baller, et al. Deep neural network for pixel-level electromagnetic particle identification in the microboone liquid argon time projection chamber. *Physical Review D*, 99(9):092001, 2019. [56](#)
- [79] P Abratenko, M Alrashed, R An, J Anthony, J Asaadi, A Ashkenazi, S Balasubramanian, B Baller, C Barnes, G Barr, et al. Neutrino event selection in the microboone liquid argon time projection chamber using wire-cell 3-d imaging, clustering, and charge-light matching. *arXiv preprint arXiv:2011.01375*, 2020. [56](#)
- [80] D Caratelli. Neutrino identification with scintillation light in microboone. *Journal of Instrumentation*, 15(03):C03023, 2020. [57](#)

- [81] R Acciarri, C Adams, R An, J Anthony, J Asaadi, M Auger, L Bagby, S Balasubramanian, B Baller, C Barnes, et al. The pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the microboone detector. *The European Physical Journal C*, 78(1):82, 2018. [60](#)
- [82] R Acciarri, C Adams, J Asaadi, B Baller, T Bolton, C Bromberg, F Cavanna, E Church, D Edmunds, A Ereditato, et al. A study of electron recombination using highly ionizing particles in the argoneut liquid argon tpc. *Journal of Instrumentation*, 8(08):P08005, 2013. [64](#), [65](#)
- [83] C Adams, M Alrashed, R An, J Anthony, J Asaadi, A Ashkenazi, S Balasubramanian, B Baller, C Barnes, G Barr, et al. Calibration of the charge and energy loss per unit length of the microboone liquid argon time projection chamber using muons and protons. *Journal of Instrumentation*, 15(03):P03022, 2020. [64](#)
- [84] David Caratelli. *Study of Electromagnetic Interactions in the MicroBooNE Liquid Argon Time Projection Chamber*. PhD thesis, Columbia University, 2018. [64](#), [101](#)
- [85] R Acciarri, C Adams, J Asaadi, B Baller, T Bolton, C Bromberg, F Cavanna, E Church, D Edmunds, A Ereditato, et al. A study of electron recombination using highly ionizing particles in the argoneut liquid argon tpc. *Journal of Instrumentation*, 8(08):P08005, 2013. [66](#)
- [86] The MicroBooNE Collaboration. Log-likelihood ratio PID paper presentation at the biweekly meeting. *MicroBooNE docdb-34063*, 2021. [66](#)
- [87] P. Abratenko, R. An, J. Anthony, J. Asaadi, A. Ashkenazi, S. Balasubramanian, B. Baller, C. Barnes, G. Barr, et al. Measurement of the longitudinal diffusion of ionization electrons in the microboone detector, 2021. [68](#)
- [88] V. M. Atrazhev and I. V. Timoshkin. Transport of electrons in atomic liquids in high electric fields. *Trans. Dielectrics and Electrical Insulation*, 5(3):450–457, 1998. [69](#), [72](#), [83](#), [87](#), [94](#), [164](#)

- [89] P. Cennini et al. Performance of a 3-ton liquid argon time projection chamber. *Nucl. Instrum. Meth.*, A345:230–243, 1994. [69](#), [72](#), [87](#), [93](#), [94](#), [164](#), [165](#)
- [90] Yichen Li, Thomas Tsang, Craig Thorn, Xin Qian, Milind Diwan, Jyoti Joshi, Steve Kettell, William Morse, Triveni Rao, James Stewart, et al. Measurement of longitudinal electron diffusion in liquid argon. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 816:160–170, 2016. [69](#), [72](#), [83](#), [87](#), [94](#)
- [91] Marta Torti. Electron diffusion measurements in the ICARUS T600 detector. *J. Phys. Conf. Ser.*, 888(1):012060, 2017. [69](#)
- [92] C Adams, M Alrashed, R An, J Anthony, J Asaadi, A Ashkenazi, S Balasubramanian, B Baller, C Barnes, G Barr, 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. *arXiv preprint arXiv:1910.01430*, 2019. [73](#), [78](#), [87](#), [88](#)
- [93] 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(03):P03022, 2020. [91](#)
- [94] SR Hunter, JG Carter, and LG Christophorou. Electron transport measurements in methane using an improved pulsed townsend technique. *Journal of Applied Physics*, 60(1):24–35, 1986. [93](#)
- [95] Hiroki Kusano, José A. Matias-Lopes, Mitsuhiro Miyajima, Eido Shibamura, and Nobuyuki Hasebe. Electron mobility and longitudinal diffusion coefficient in high-density gaseous xenon. *Japanese Journal of Applied Physics*, 51:116301, oct 2012. [93](#)
- [96] Eido Shibamura, Tan Takahashi, Shinzou Kubota, and Tadayoshi Doke. Ratio of diffusion coefficient to mobility for electrons in liquid argon. *Phys. Rev.*, A20(6):2547, 1979. [93](#)

- [97] J. S. O’Connell et al. Electromagnetic Excitation of the  $\Delta$  Resonance in Nuclei. *Phys. Rev. C*, 35:1063, 1987. [96](#)
- [98] C.T. Kullenberg et al. A search for single photon events in neutrino interactions. *Physics Letters B*, 706(4):268–275, 2012. [96](#)
- [99] K Abe, R Akutsu, A Ali, C Andreopoulos, L Anthony, M Antonova, S Aoki, A Ariga, Y Ashida, Y Awataguchi, et al. Search for neutral-current induced single photon production at the nd280 near detector in t2k. *Journal of Physics G: Nuclear and Particle Physics*, 46(8):08LT01, 2019. [96](#), [97](#)
- [100] The MicroBooNE Collaboration. MICROBOONE-NOTE-1043-PUB: MicroBooNE low-energy excess signal prediction from unfolding MiniBooNE Monte-Carlo and data. *MicroBooNE docdb-15587*, 2018. [96](#), [97](#)
- [101] E. Wang, L. Alvarez-Ruso, and J. Nieves. Photon emission in neutral-current interactions at intermediate energies. *Phys. Rev. C*, 89:015503, Jan 2014. [96](#), [97](#)
- [102] En Wang, Luis Alvarez-Ruso, and Juan Nieves. Single photon events from neutral current interactions at minibooNE. *Physics Letters B*, 740:16–22, 2015. [98](#)
- [103] Nikita Sharma. Understanding the mathematics behind decision trees. <https://heartbeat.fritz.ai/understanding-the-mathematics-behind-decision-trees-22d86d55906>. Accessed: 2021-03-05. [105](#), [106](#)
- [104] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997. [105](#), [120](#)
- [105] Julia Nikulski. The ultimate guide to adaboost, random forests and xgboost. <https://towardsdatascience.com/the-ultimate-guide-to-adaboost-random-forests-and-xgboost-7f9327061c4f>. Accessed: 2021-03-05. [105](#), [108](#)



- [106] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016. [105](#), [107](#), [120](#)
- [107] Frank J Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78, 1951. [118](#)
- [108] Rudolf Frühwirth. Application of kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2-3):444–450, 1987. [125](#)
- [109] Patrick Green. Lidine 2019: "scintillation light response, reconstruction, and calibration in microboone with four years of data. MicroBooNE DocDB: 25039, 2019. [133](#)
- [110] MicroBooNE Collaboration. Booster neutrino flux predication at microboone. MicroBooNE Public Note: 1031, 2018. [136](#)
- [111] M Apollonio, A Artamonov, A Bagulya, G Barr, A Blondel, F Bobisut, M Bogomilov, M Bonesini, C Booth, S Borghi, et al. Forward production of charged pions with incident protons on nuclear targets at the cern proton synchrotron. *Physical Review C*, 80(3):035208, 2009. [136](#)
- [112] GM Radecky, VE Barnes, DD Carmony, AF Garfinkel, M Derrick, E Fernandez, L Hyman, G Levman, D Koetke, B Musgrave, et al. Study of single-pion production by weak charged currents in low-energy  $\nu$  d interactions. *Physical Review D*, 25(5):1161, 1982. [136](#)
- [113] NJ Baker, PL Connolly, SA Kahn, MJ Murtagh, RB Palmer, NP Samios, and M Tanaka. Total cross sections for  $\nu \mu$  n and  $\nu \mu$  p charged-current interactions in the 7-foot bubble chamber. *Physical Review D*, 25(3):617, 1982. [136](#)
- [114] Davio Cianci, Andy Furmanski, Georgia Karagiorgi, and Mark Ross-Lonergan. Prospects of light sterile neutrino oscillation and c p violation searches at the fermilab short baseline neutrino facility. *Physical Review D*, 96(5):055001, 2017. [142](#)

- [115] Xiangpan Ji, Wenqiang Gu, Xin Qian, Hanyu Wei, and Chao Zhang. Combined neyman–pearson chi-square: An improved approximation to the poisson-likelihood chi-square. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 961:163677, 2020. [142](#)
- [116] Yichen Li et al. A 20-Liter Test Stand with Gas Purification for Liquid Argon Research. *JINST*, 11(06):T06001, 2016. [164](#), [165](#)
- [117] Gregory H Wannier. Motion of gaseous ions in strong electric fields. *Bell System Technical Journal*, 32(1):170–254, 1953. [164](#)
- [118] Robert Edward Robson. A thermodynamic treatment of anisotropic diffusion in an electric field. *Australian Journal of Physics*, 25(6):685–694, 1972. [164](#)
- [119] Thomas Karl Warburton. *Simulations and Data analysis for the 35 ton Liquid Argon detector as a prototype for the DUNE experiment*. PhD thesis, Sheffield U., 2017. [166](#)

# Appendix

## A Diffusion World Data Summary Plot Details

*Statement of disclosure:* This appendix contains an edited version of an appendix from a soon-to-be published journal article **CITE ARXIV**. While this article lists the entire MicroBooNE collaboration as the author (this is standard MicroBooNE procedure), I contributed the majority of the material in the article along with my colleague Adam Lister. The version in this thesis has been edited with slight changes to wording throughout and modified to match the format of the remainder of this thesis.

This Appendix describes the production of the  $D_L$  world data summary plot (Figure 6.16). The Atrazhev-Timoshkin theory calculation [88] and the Li et al. data [116] are presented in terms of the effective longitudinal electron energy,  $\epsilon_L$ , while the MicroBooNE and ICARUS [89] results are in terms of  $D_L$ . To convert between  $\epsilon_L$  and  $D_L$ , we use the Einstein-Smoluchowski relation [117, 118],

$$D_L = \frac{\mu(E)\epsilon_L}{e}, \quad (1)$$

where  $\mu(E)$  is the electron mobility as a function of electric field and  $e$  is the electron charge.

The Atrazhev-Timoshkin theory calculation is parametrized in terms of the effective longitudinal and transverse electron energies,  $\epsilon_L$  and  $\epsilon_T$ , respectively. They note that, for E-fields above  $10^3$  V/cm,

$$\epsilon_L = 0.5\epsilon_T, \quad (2)$$

where  $\epsilon_T$  is given by

$$\epsilon_T = 0.8T(E/E_h). \quad (3)$$

Here,  $E_h$  is the boundary field strength above which electrons are considered “hot.” For fields below this value,  $\epsilon_L = T$ . There is no description of  $\epsilon_L$  for intermediate fields (i.e., above 0 and below  $10^3$  V/cm), so we interpolate in this region. We use a fourth order polynomial fit between  $T = 7.67 \times 10^{-3}$  eV (89 K) and the region above  $E = 1200$  V/cm (which follows Equation 3). The resultant functional form is given by

$$\begin{aligned}
\epsilon_L = & 7.67 \times 10^{-3} + 1.39 \times 10^{-5} E \\
& + 2.19 \times 10^{-9} E^2 - 2.69 \times 10^{-13} E^3 \\
& + 1.15 \times 10^{17} E^4.
\end{aligned}$$

The data of Li et al. are estimated using graph clicking software. The functional form of the parametrization is provided in [116]. This reference also provides a parametrization of  $\mu(E)$ , which is shown in Figure 10 of Reference [116] to have excellent agreement with world data. Finally, the ICARUS data point is taken directly from [89], which reports an average  $D_L$  value of  $D_L = 4.8 \pm 0.2 \text{ cm}^2/\text{s}$  for E-field values of 100, 150, 250, and 350 V/cm.

## B Potential for Tagging $t_0$ Using Diffusion

*Statement of disclosure:* This appendix contains an edited version of an appendix from a soon-to-be published journal article **CITE ARXIV**. While this article lists the entire MicroBooNE collaboration as the author (this is standard MicroBooNE procedure), I contributed the majority of the material in the article along with my colleague Adam Lister. The version in this thesis has been edited with slight changes to wording throughout and modified to match the format of the remainder of this thesis.

The potential for  $t_0$ -tagging using diffusion has been investigated in reference [119] where many hits along a single track are considered in order to reconstruct a  $t_0$  for that track. Recently, this method has gained some attention in the context of  $t_0$ -tagging individual energy depositions. The feasibility of performing  $t_0$ -tagging for individual energy depositions using this method is dependent on the spread of the hit RMS values for each drift time. This is shown in figure 6.7a, where each bin in drift time has a wide range of allowed hit widths. In addition, comparisons of the hit RMS distributions at drift times of 45  $\mu\text{s}$ , 1150  $\mu\text{s}$ , and 2254  $\mu\text{s}$  are shown in figure B.1. Each of these plots uses the nominal angular selection of this analysis,  $\theta_{xz} < 6^\circ$ , meaning this should be comparable to a point source. Figure B.1 shows that the spread in the hit RMS is relatively wide on all three planes. In order to boost the success rate of tagging the  $t_0$  of individual energy depositions, one may imagine performing charge matching across planes in order to obtain three hits rather than one; however, statistical fluctuations in electron transport are likely much larger than any plane-to-plane differences that might be present in a given LArTPC. The collection plane has the narrowest hit RMS distributions and therefore should be the most promising for  $t_0$  tagging individual waveforms, and so we focus the rest of this appendix there.

The ability to  $t_0$ -tag a single energy deposition accurately relies on each hit RMS value corresponding to a tight distribution of possible drift times. The wider the distribution of possible drift times, the less accurately the  $t_0$  can be measured. Figure B.2 shows the distribution of hit times on the collection plane for hits in 0.1  $\mu\text{s}$  bins of hit width from zero to the maximum drift time in the MicroBooNE TPC, 2300  $\mu\text{s}$ . For each bin of hit RMS, the range of drift times spans the entire 0-2300  $\mu\text{s}$  region. To make a more quantitative

statement, we fit a Gaussian functional form around the peak of the  $1.5 \mu\text{s} < \text{Hit RMS} < 1.6 \mu\text{s}$  plot, from which we can estimate a  $1\sigma$  uncertainty of approximately  $\pm 560 \mu\text{s}$ . However, we caution that the distribution is relatively non-Gaussian and this should be taken as a lower bound on the resolution. We also note that the resolution is likely larger for hits with larger drift times because the hit width is proportional to  $\sqrt{t}$  and the width changes more slowly for longer drifts. The first 10-kTon module of the DUNE Far Detector is planned to have a drift distance of 3.6 m with a drift field of 500 V/cm resulting in a maximum drift time of 2.25 ms. The data presented in this work cover this region of drift time and the field dependence of  $D_L$  is negligible (figure 6.16), making this measurement relevant for the DUNE far detector.

It is clear that even using the collection plane, which is expected to out-perform the induction planes, there remain significant hurdles to overcome. The measured central value of  $D_L$  combined with statistical fluctuations from the diffusion process means that  $t_0$  tagging of individual energy depositions using hit RMS alone will result in poor time resolution. Combination of the hit RMS with other variables has not been investigated in this work. Application of this technique to charged particle tracks which are reconstructed from energy depositions on many readout channels remains an intriguing possibility, as the statistical fluctuations will average out as the number of hits increases.

To aid in making predictions for future long-drift detectors, we make the observation that for drift times above  $\sim 1000 \mu\text{s}$  the ratio of the width to the hit RMS distribution with the mean of the hit RMS distributions is approximately constant at 0.056 (figure B.3). This relationship does not appear to hold for the induction planes. We provide this extrapolation for use with other LArTPCs, but we emphasize that this is not a substitution for a full analysis with a dedicated simulation. Such an endeavor demands more precision from simulations than has been required to date. For example, we have noted that the distribution of the hit RMS for a given drift time tends to be narrower in our simulations than in our data (figure B.4). Any attempt to  $t_0$  tag single energy depositions using diffusion would need to tune the simulation to the data with great care.

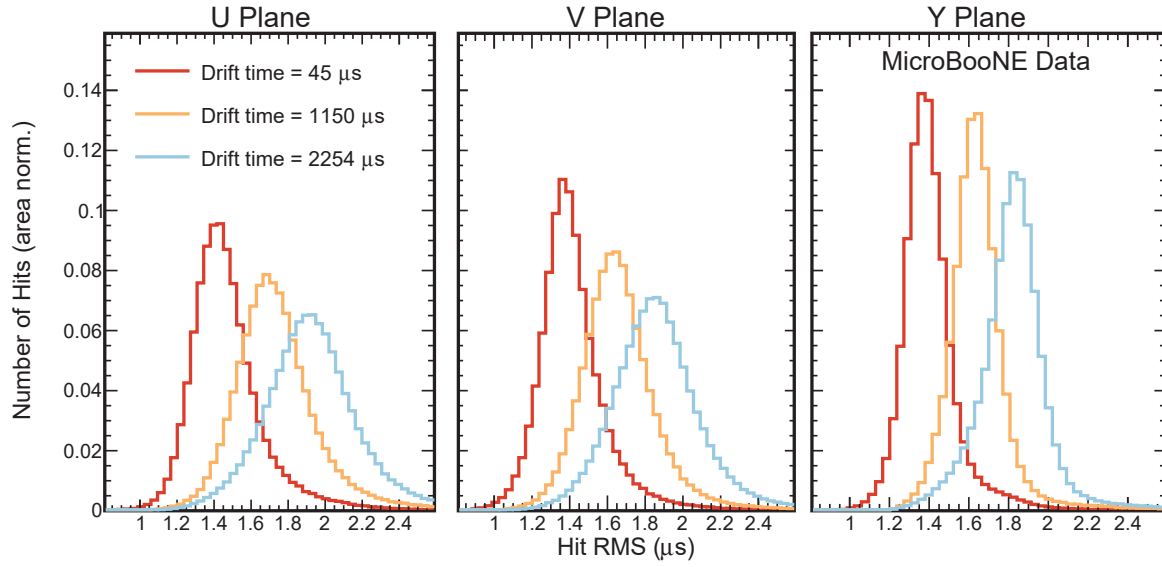


Figure B.1: One-dimensional comparisons of the area-normalized hit RMS distribution on the three wire planes for drift times of 45  $\mu\text{s}$ , 1150  $\mu\text{s}$ , and 2254  $\mu\text{s}$ .



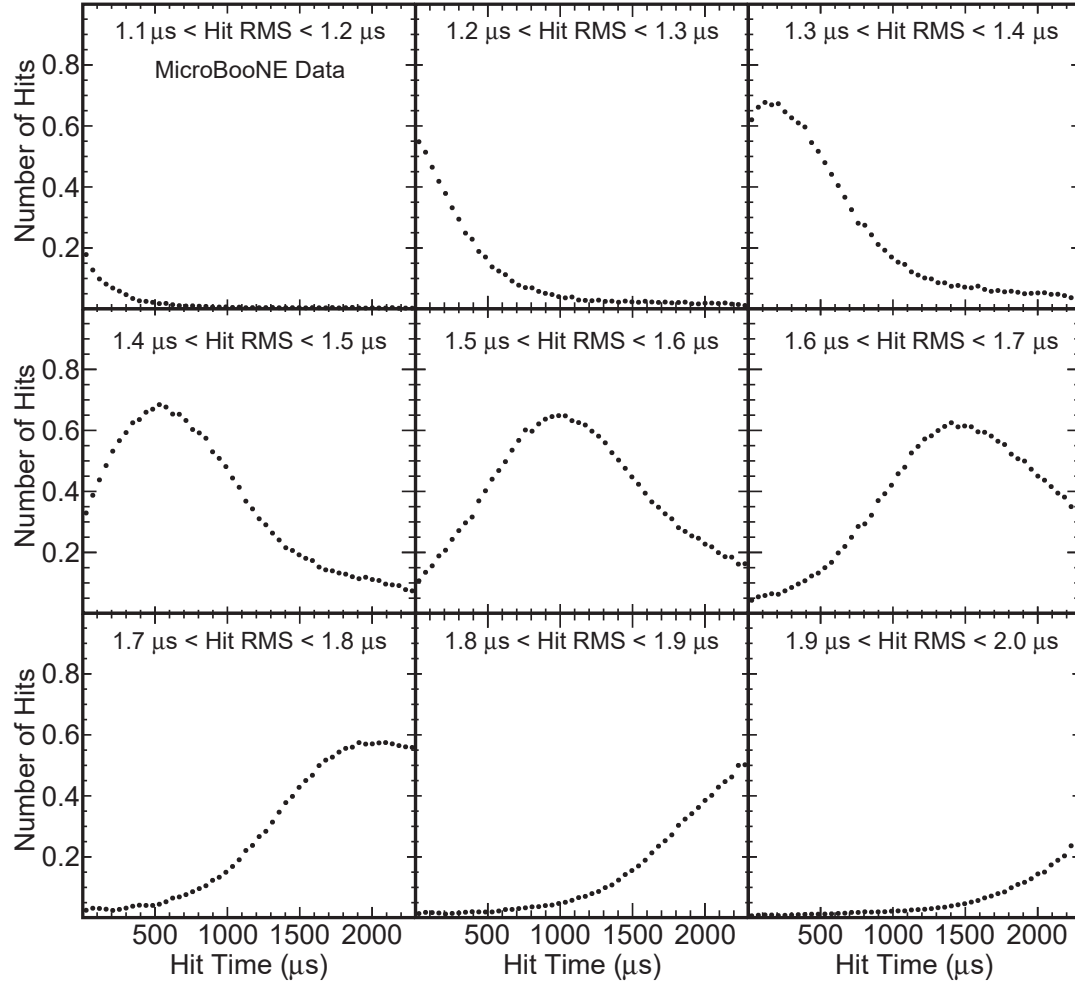


Figure B.2: Distribution of hit times for different slices in hit RMS on the collection plane.

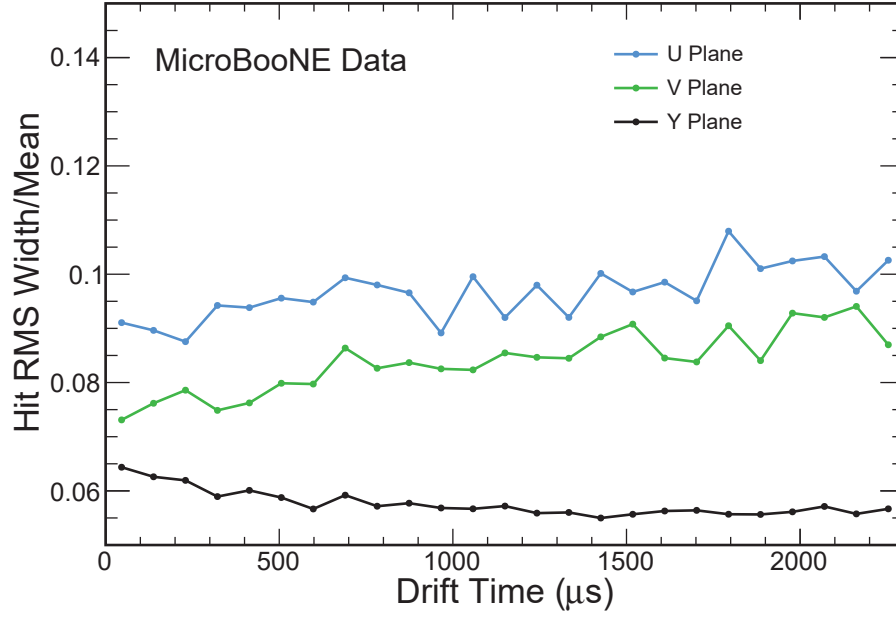


Figure B.3: Distribution of the width of the hit RMS distribution over the mean of that distribution for the U, V and Y planes.

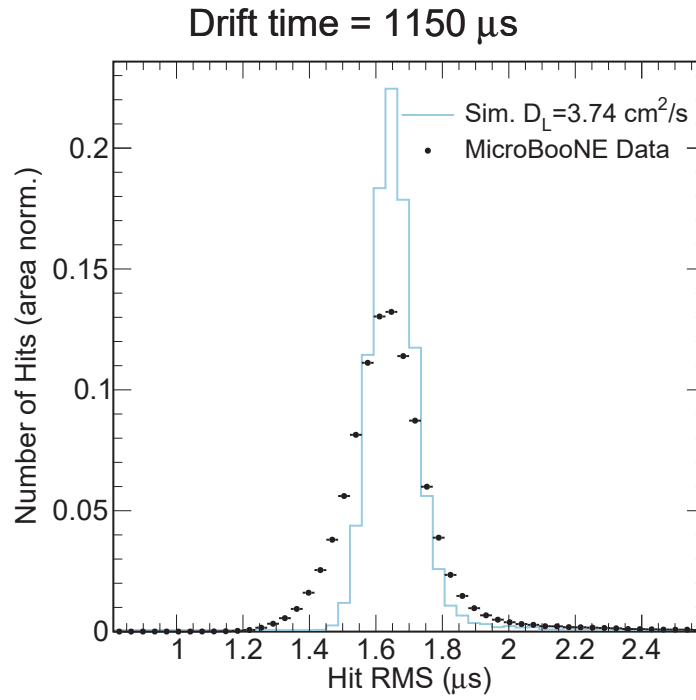


Figure B.4: Comparison of data and simulation for the hit RMS distribution in the center of the TPC, around drift time = 1150 μs.

## C NC $\pi^0$ Data Filters

In order to analyze events from across Runs 1-3 while satisfying the MicroBooNE blindness criteria, we employ two data filters, one for each of the  $2\gamma$  topologies (see Section 8). The MicroBooNE blindness criteria states that a data filter must not select more than 45  $\nu_e$  or  $\Delta$  radiative LEE events per  $13.2\text{e}20$  POT according to the GENIE v3 prediction. The blindness criteria ensures that no data filter accidentally examines the blinded LEE-like data. This section describes the cuts used in each filter, along with the predicted passing rates of our signal and the relevant LEE backgrounds. All data-to-MC comparisons shown in Section 8 use the filtered data from Runs 1, 2 and 3 as the on-beam data sample. To date, this corresponds to  $\sim 6 \times 10^{20}$  POT for each topology. The following two sections come with the caveat that the filter blindness studies were performed using older versions of the corresponding samples, hence the total number of events in each sample and exact signal definition passing rates differ from those listed in Sections 7.2 and 8. However, because these older studies led to the approval of the filters, we choose to show as they were at the time of approval.

### C.1 $2\gamma 1p$ Filter

The  $2\gamma 1p$  filter consists of three primary cuts:

- The reconstructed event topology consists of two showers and one track
- The conversion distances for both showers must be greater than 1 cm
- The reconstructed neutrino vertex must be at least 5 cm away from any TPC wall

Note that these cuts are a subset of the pre-selection and topological cuts used in the NC  $\pi^0$  analysis. Figure C.5 shows the filter passing rates of various samples as a function of true neutrino energy. The samples are divided into the following definitions:<sup>1</sup>

- **Signal NC  $\pi^0$**

---

<sup>1</sup>Note that the  $2\gamma 1p$  filter studies were performed at a time where we used a 10 cm fiducial volume and a 40 MeV proton kinetic energy threshold. This has no significant impact on the filter performance or blindness tests.

- True NC event with exactly one  $\pi^0$  in the final state
  - One proton in the final state with range-based kinetic energy  $> 40$  MeV
  - Two photon showers originating from  $\pi^0$  decay, each with  $> 20$  MeV energy
  - True neutrino vertex at least 10 cm from any TPC wall
- **Other NC  $\pi^0$ :**
    - True NC event with exactly one  $\pi^0$  in the final state
    - Fails some combination of the fiducial volume and/or energy thresholds listed under “Signal NC  $\pi^0$ .”
- **Intrinsic  $\nu_e$ :**
    - True CC  $\nu_e$  or  $\bar{\nu}_e$
    - True neutrino energy between 50 MeV and 1.5 GeV
- **NC  $\Delta$  radiative:**
    - True NC  $\Delta$  radiative decay
    - No  $\pi^0$  in the final state
    - One photon decay from  $\Delta$  with energy  $> 20$  MeV
    - For the 1g1p (1g0p) topology, one (zero) proton(s) in the final state with range-based kinetic energy  $> 40$  MeV
    - True neutrino vertex within 10 cm of any TPC wall

The efficiencies of each cut, along with the number of selected events as a function of true neutrino energy, can be found in Table C.1. Here, the “Definition” column shows the number of MC events that pass the above sample definitions, scaled to  $13.2\text{e}20$  POT;<sup>2</sup> The “Topology” column then shows the number of events that pass the topological definition of the NC  $\pi^0$  filter, i.e., two showers and one track; and the “Filter” column shows how many

---

<sup>2</sup>While the expected final dataset is now  $\sim 12.25 \times 10^{20}$  POT, the MicroBooNE blindness criteria are still defined relative to  $13.2\text{e}20$  POT, so we use this number for blindness studies. This applies to the  $2\gamma 0p$  case as well.

events pass the conversion distance and fiducial volume cuts listed earlier in this section. The MicroBooNE blindness criteria states that a filter must not select more than 45  $\nu_e$  or  $\Delta$  radiative LEE events per 13.2e20 POT. The filter satisfies these conditions.

## C.2 $2\gamma 0p$ Filter

The  $2\gamma 0p$  filter uses three primary cuts:

- The reconstructed event topology consists of two showers and zero tracks
- The leading shower energy must be greater than 30 MeV
- The reconstructed neutrino vertex must be at least 5 cm away from any TPC wall

As in the  $1p$  filter, we require a specific topology with a vertex within a 5 cm fiducial volume. However, instead of cutting on shower conversion distances, we instead cut on the leading shower energy. In events with no reconstructed tracks, Pandora generally places the vertex at the leading shower starting position. A conversion distance cut on the leading shower would therefore remove the majority of signal events. Furthermore, the lack of a track makes determining shower directions difficult, as we don't have a handle on how far each photon pair traveled before converting. We therefore choose to use the leading shower energy as a conservative cut for removing poorly-reconstructed or low-energy events.

Figure C.6 shows the filter passing rates as a function of true neutrino energy. Here, the samples are divided into a more inclusive and streamlined set of definitions relative to the  $2\gamma 1p$  case:

- **NC  $\pi^0$** 
  - True NC event with exactly one  $\pi^0$  in the final state
  - Two photon showers originating from  $\pi^0$  decay, each with  $> 20$  MeV energy
  - True neutrino vertex at least 5 cm from any TPC wall
- **Intrinsic  $\nu_e$ :**
  - True CC  $\nu_e$  or  $\bar{\nu}_e$

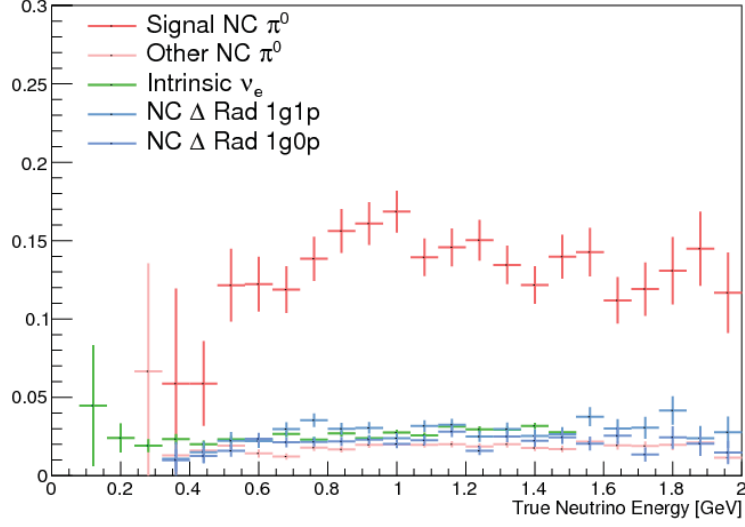


Figure C.5: NC  $\pi^0$   $2\gamma 1p$  filter efficiency as a function of true neutrino energy for various samples.

Table C.1: NC  $\pi^0$   $2\gamma 1p$  filter passing rates for various samples, scaled to  $13.2e20$  POT. Passing rate percentages are defined relative to the “Definition” column. The meaning of each column title is described in the text.

Sample	Definition	Topology	Filter
Signal NC $\pi^0$	8,318 (100%)	1,305 (15.7%)	1,126 (13.5%)
Other NC $\pi^0$	42,415 (100%)	1,343 (3.2%)	781 (1.8%)
Intrinsic $\nu_e$	1,195 (100%)	63 (5.3%)	32 (2.7%)
$\Delta$ rad 1g1p	87 (100%)	4 (4.6%)	3 (3.4%)
$\Delta$ rad 1g0p	102 (100%)	4 (3.9%)	2 (2.0%)

- True neutrino energy between 50 MeV and 1.5 GeV
- **NC  $\Delta$  radiative:**
  - True NC  $\Delta$  radiative decay
  - No  $\pi^0$  in the final state
  - One photon decay from  $\Delta$  with energy  $> 20$  MeV
  - True neutrino vertex within 5 cm of any TPC wall

The efficiencies of each cut, along with the number of selected events as a function of true neutrino energy, can be found in Table C.2. As before, the “Definition” column shows the number of MC events that pass the above sample definitions, scaled to  $13.2 \times 10^{20}$  POT; the “Topology” column then shows the number of events that pass the topological definition of the NC  $\pi^0$  filter, i.e., two showers and no tracks; and the “Filter” column shows how many events pass the shower energy and fiducial volume cuts listed earlier in this section. The  $2\gamma 0p$  filter satisfies the blindness criteria.

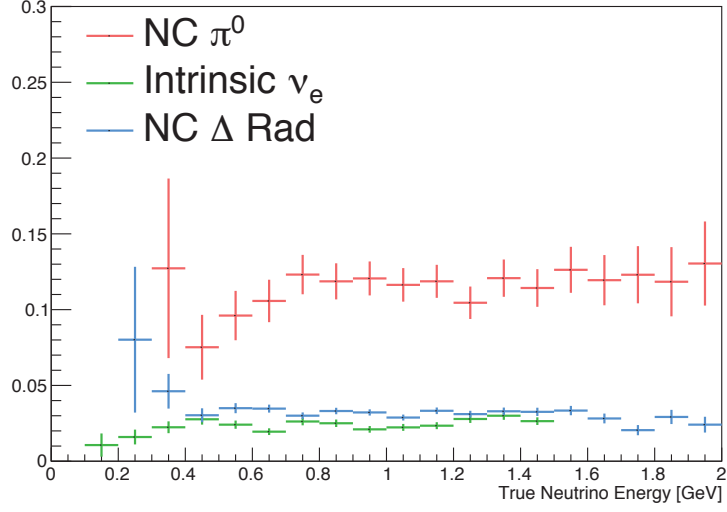


Figure C.6: NC  $\pi^0$   $2\gamma 0p$  filter efficiency as a function of true neutrino energy for signal and relevant LEE backgrounds.

Table C.2: NC  $\pi^0$   $2\gamma 0p$  filter passing rates for various samples, scaled to  $13.2e20$  POT. Passing rate percentages are defined relative to the “Definition” column. The meaning of each column title is described in the text.

Sample	Definition	Topology	Filter
NC $\pi^0$	8,806 (100%)	1,008 (11.4%)	987 (11.2%)
Intrinsic $\nu_e$	1,479 (100%)	40 (2.7%)	36 (2.4%)
$\Delta$ rad	227 (100%)	8 (3.5%)	7 (3.1%)



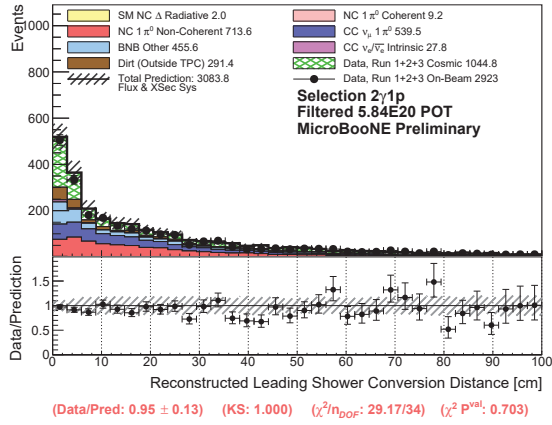
## D NC $\pi^0$ BDT Training Variables

### D.1 $2\gamma 1p$ Training Variables

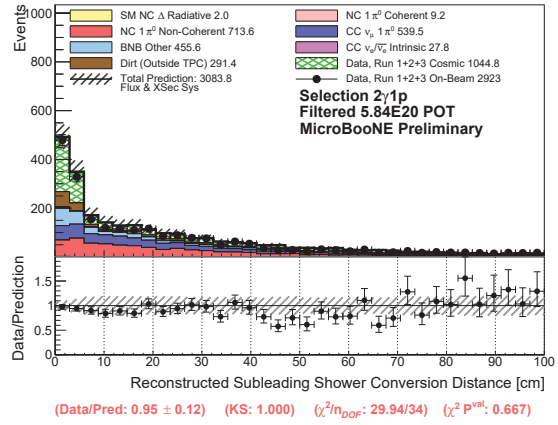
Figure D.7 shows the data-to-MC comparisons for all training variables in the  $2\gamma 1p$  selection at the pre-selection stage. Flux and cross-section uncertainties are included. At this stage, the only variable with a  $\chi^2$  p-value  $< 0.05$  is the track mean truncated dE/dx. This is due to a shift near the peak at 2 MeV/cm, where we expect minimally-ionizing particles (MIPs) such as muons to peak. In all other regions of track dE/dx, the data/MC agreement is quite good. This shift is covered by detector systematic uncertainties, as shown in Figure 8.10. We also note that, at the final selection, the data/MC agreement improves significantly even without detector systematics, as the BDT cuts nearly all events near the MIP-like peak. This is shown in Figure 8.4a. All other variables show good agreement between data and MC at the pre-selection stage, as evidenced by the p-values and the  $\chi^2/n_{\text{DOF}}$ .

### D.2 $2\gamma 0p$ Training Variables

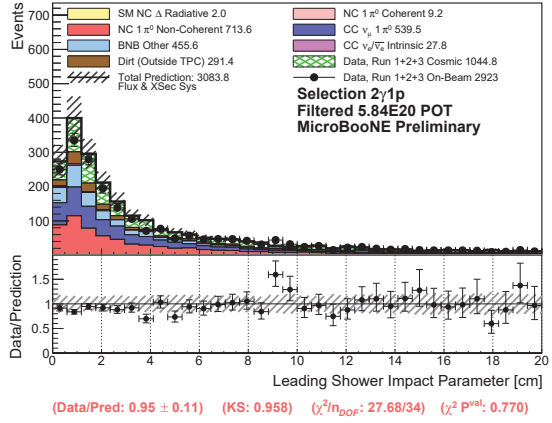
Figure D.8 shows the data-to-MC comparisons for all training variables in the  $2\gamma 0p$  selection at the pre-selection stage. Flux and cross-section uncertainties are included. All variables show good agreement between data and MC at the pre-selection stage, as evidenced by the p-values and the  $\chi^2/n_{\text{DOF}}$ .



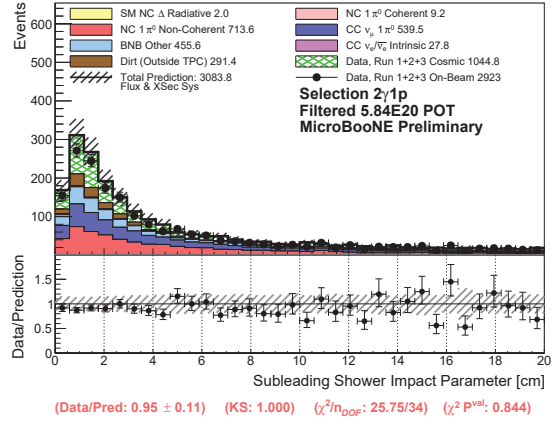
(a) Leading conversion distance



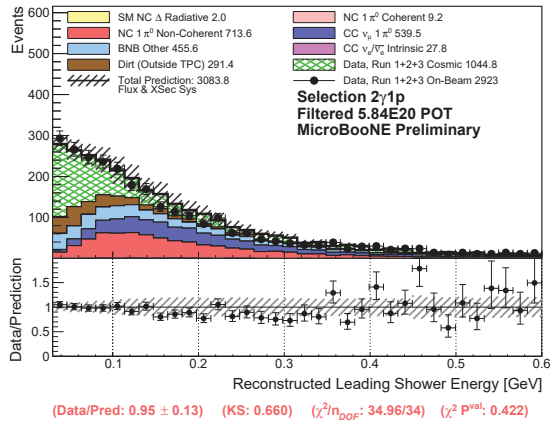
(b) Subleading conversion distance



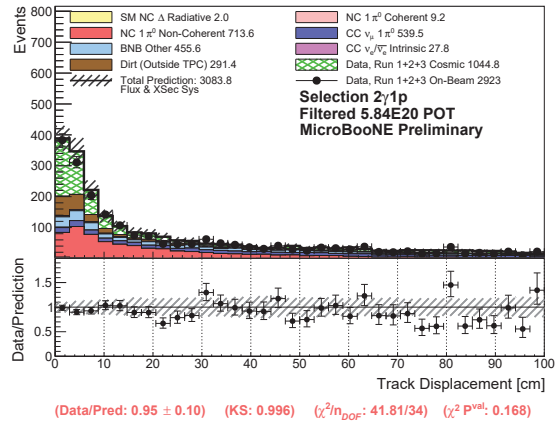
(c) Leading shower impact parameter



(d) Subleading shower impact parameter

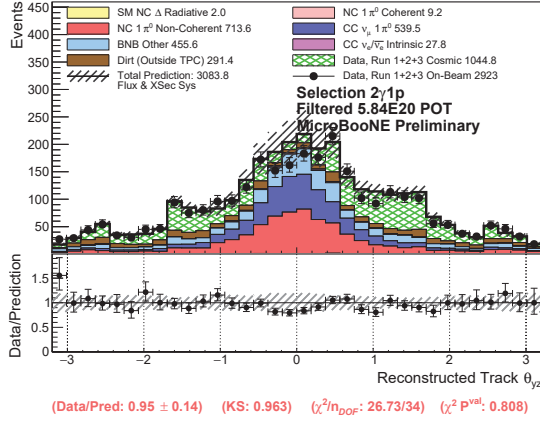


(e) Leading shower energy

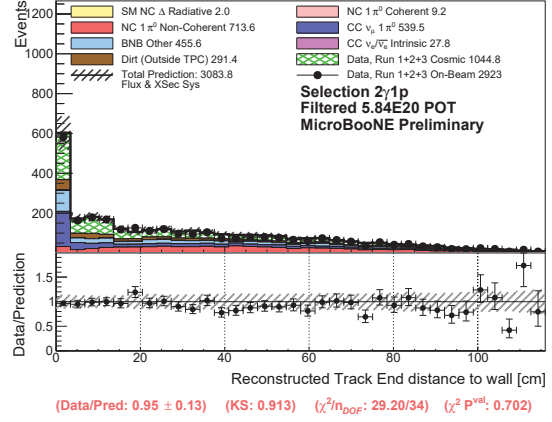


(f) Track length

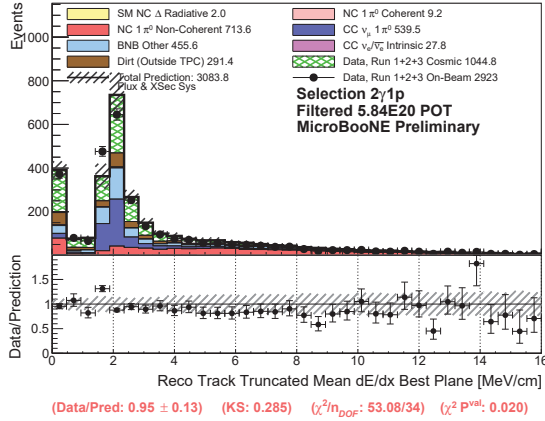
Figure D.7: Training variables for 2γ1p.



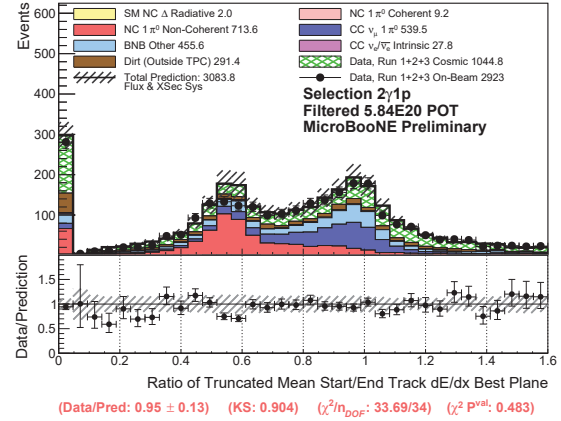
(g) continued Track  $\theta_{yz}$



(h) continued Distance from track end point to nearest TPC wall

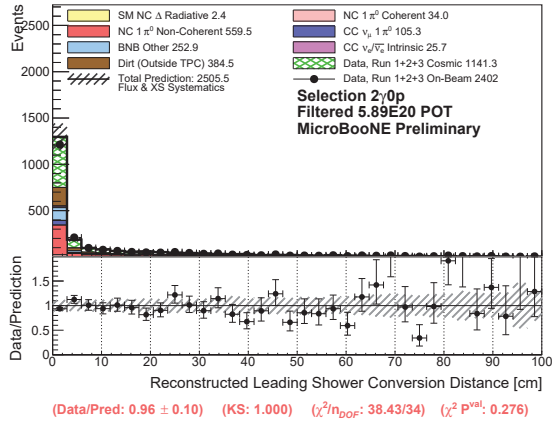


(i) continued Track mean truncated dE/dx

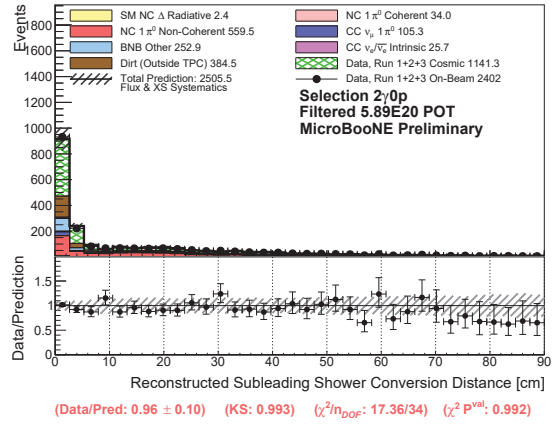


(j) continued Ratio of track start/end dE/dx

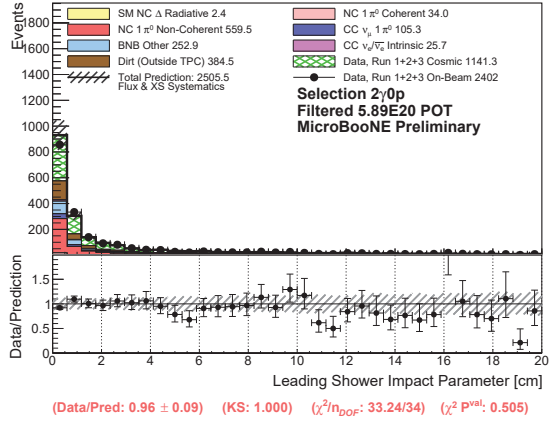
Figure D.7 continued: Training variables for 2 $\gamma$ 1p.



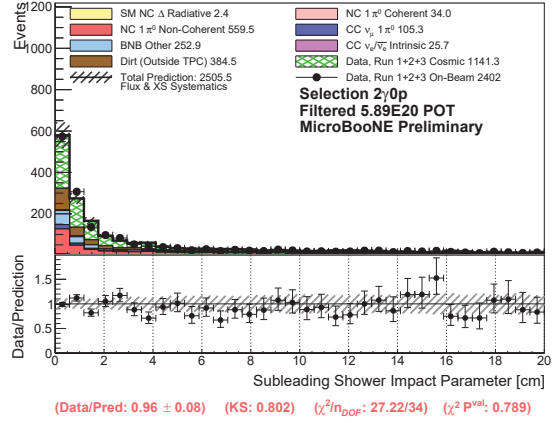
(a) Leading conversion distance



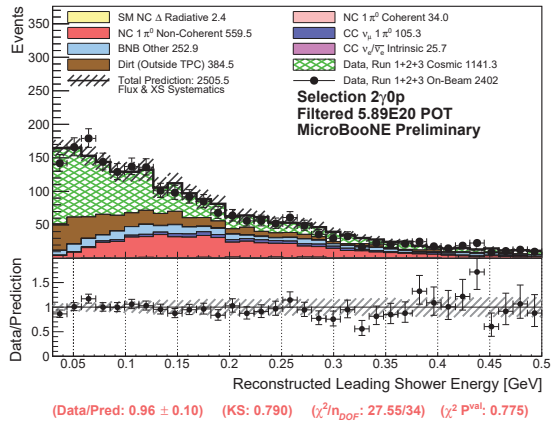
(b) Subleading conversion distance



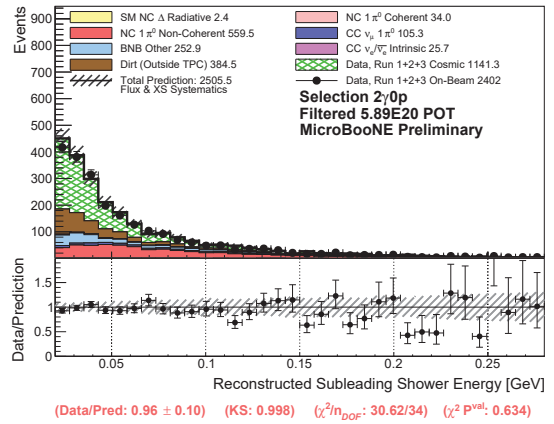
(c) Leading shower impact parameter



(d) Subleading shower impact parameter

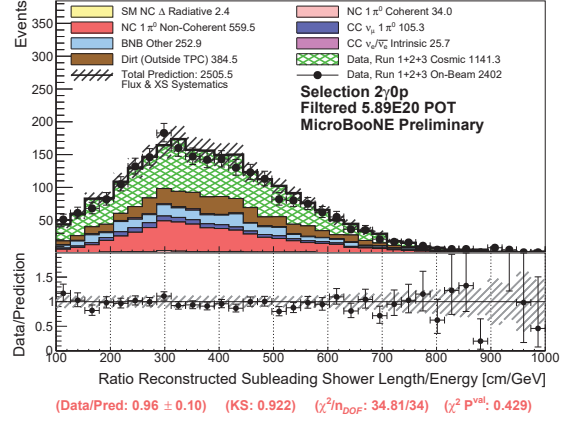
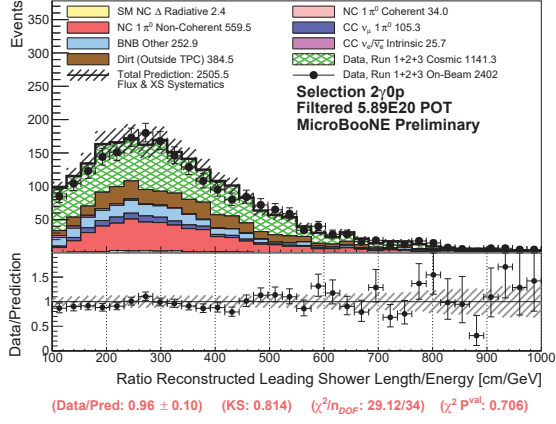


(e) Leading shower energy



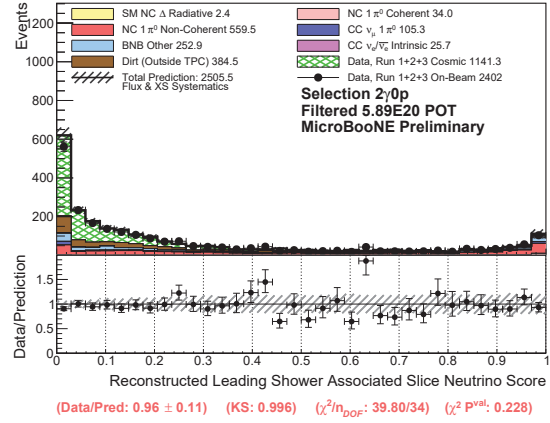
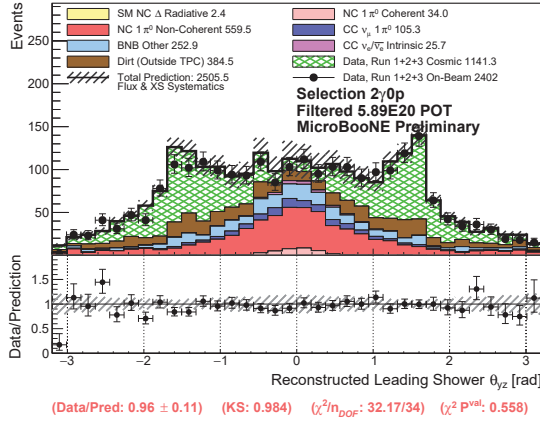
(f) Subleading shower energy

Figure D.8: Training variables for  $2\gamma 0p$ .



(g) continued Ratio of leading shower length/en-  
ergy

(h) continued Ratio of subleading shower  
length/energy



(i) continued Leading shower  $\theta_{yz}$

(j) continued Leading shower neutrino slice score

Figure D.8 continued: Training variables for  $2\gamma 0p$ .

# Vita

Andrew Mogan [REDACTED]

[REDACTED]. He then double majored in physics and mathematics at Murray State University before moving to the University of Tennessee, Knoxville (UTK) to continue his physics education. After graduating from UTK, he then moved to a postdoctoral research associate position at Colorado State University.